Incorporating Global Industrial Classification Standard Into Portfolio Allocation: A Simple Factor-Based Large Covariance Matrix Estimator With High-Frequency Data

Jianqing FAN
Department of Operations Research & Financial Engineering and Bendheim Center for Finance, Princeton University, Princeton, NJ 08544, and School of Data Science, Fudan University, Shanghai, China (*jqfan@princeton.edu*)

Alex FURGER
Department of Operations Research & Financial Engineering, Princeton University, 26 Prospect Avenue, Princeton, NJ 08544 (*afurger@princeton.edu*)

Dacheng XIU
Booth School of Business, University of Chicago 5807 S. Woodlawn Avenue, Chicago, IL 60637 (*dacheng.xiu@chicagobooth.edu*)

We document a striking block-diagonal pattern in the factor model residual covariances of the S&P 500 Equity Index constituents, after sorting the assets by their assigned Global Industry Classification Standard (GICS) codes. Cognizant of this structure, we propose combining a location-based thresholding approach based on sector inclusion with the Fama-French and SDPR sector Exchange Traded Funds (ETF's). We investigate the performance of our estimators in an out-of-sample portfolio allocation study. We find that our simple and positive-definite covariance matrix estimator yields strong empirical results under a variety of factor models and thresholding schemes. Conversely, we find that the Fama-French factor model is only suitable for covariance estimation when used in conjunction with our proposed thresholding technique. Theoretically, we provide justification for the empirical results by jointly analyzing the in-fill and diverging dimension asymptotics.

KEY WORDS: Big data; Concentration inequality; GICS; High-frequency factor model; Location-based thresholding; Low rank plus sparse; Positive-definite; Precision matrix; SDPR sector ETF's.

1. INTRODUCTION

Covariance matrix estimation has been a very important input to practical portfolio allocation since the seminal work of Markowitz (1952). In recent years, the issue of ever increasing cross-sectional dimensionality in the asset universe has plagued the application of the classical methods of covariance estimation. Asset managers have expanded their potential asset pool to include a larger variety of domestic equities, as well as foreign equity, commodities, and bonds. Allocation decisions involving thousands of potential investments are common. On the other hand, some managers and funds have chosen to reallocate their portfolios more frequently than has traditionally been common, every day, week, or month.

To guide their allocation decisions, practitioners rely on several important properties of an asset covariance matrix. (1) It should be positive definite. (2) The covariance matrix must be well-conditioned to give reasonable (and economically feasible) portfolio allocations. (3) The matrix should also take advantage of the established time locality of volatility and correlation. (4) The estimator should be easily understood so that it is transparent. Here, we propose such a covariance estimator and establish its relevant theory.

We consider the case of an asset return factor model on intraday data. Using this high-frequency factor model, we are able to simultaneously reduce the intrinsic dimensionality of the underlying estimation problem and increase the effective sample size by taking advantage of intraday return measurements. This combination enables our estimator to use many fewer days of data, while still providing a well-conditioned and stable estimate, and thus take advantage of the time locality of volatility and correlation nonparametrically. Our method also emphasizes positive definiteness and transparency as first class principles. To achieve this, we enforce an economically motivated and intuitive factor model residual correlation structure based on sector and industry group inclusion by the Global Industrial Classification Standard (GICS). This *location-based* thresholding approach preserves positive definiteness and is very effective, while remaining interpretable.

Journal of Business & Economic Statistics
October 2016, Vol. 34, No. 4
DOI: 10.1080/07350015.2015.1052458
Color versions of one or more of the figures in the article can be found online at *www.tandfonline.com/r/jbes*.

We draw several conclusions from our empirical study. First, the famous Fama-French 3-factor model (Fama and French 1993) is not sufficient to completely eliminate the high-frequency factor model residual correlations.[1] Moreover, the structure of the residuals is nondiagonal and time-varying. We demonstrate that this degrades the performance of the covariance estimate when used in conjunction with the commonly used independent idiosyncratic innovation assumption. However, after sorting the assets by their GICS codes, it becomes visually clear the remaining correlations fit suprisingly well into a block diagonal structure, where the blocks are determined by sectors.

Cognizant of that structure, we propose combining a location-based thresholding approach based on sector inclusion with the Fama-French factors. This yields a competitive covariance estimate which is both positive definite and transparent. Empirically, this model performs well in our high-dimensional portfolio allocation task. That being said, our method is not only applicable to equities. While GICS is specifically designed to classify equities, our basic insight of imposing location-based thresholding on the residuals after sorting into groups is applicable to other asset classes, such as commodities. For instance, simple groups could be meats, metals, petroleum products, etc. It is also applicable to spatial statistics problems.

Additionally, we consider a larger factor model which utilizes the widely available sector SDPR Exchange-Traded Funds (ETF's). These ETF's trade at very high frequency and therefore can serve as useful high-frequency factors. This model significantly improves the residual correlation sparsity and results in a factor model which works well under a variety of residual covariance sparsity assumptions and is stable across time periods, including during the recent financial crisis. This added robustness is valuable because it reduces the responsibility on the practitioner to alter the model over time and increases confidence during times of market stress.

Factor models have a long history in being used to model asset returns and allocate portfolios. Some of the earlier work on this topic are very insightful even in today's market. Sharpe (1963) proposed the use of a single index model for portfolio analysis to avoid estimating all the parameters in a covariance matrix. See also Lintner (1965). King (1966) conducted a latent factor analysis with 63 securities selected from six industries (tobacco products, petroleum products, metals, utilities, retail stores) and concluded that the residual covariances after removing the market communality can be explained by the comovement within industries. Ross (1976) proposed the arbitrage pricing theory, the premise of which is a linear factor model for the cross-section of asset returns. Chamberlain and Rothschild (1983) extended the theory to allow an approximate factor structure. Among many empirical models for equity returns, the Fama-French 3-factor model by Fama and French (1993) is perhaps most well-known. Their factors are explicitly constructed using portfolios formed by sorting firm characteristics. Chen, Roll, and Ross (1986) suggested using macroeconomic variables as factors, for example, inflation, output growth gap, interest rate, risk premia, and term premia.

[1]We adopt the intraday Fama-French factors constructed by Aït-Sahalia, Kalnina, and Xiu (2014), which extend the daily factors available from the webpage of Kenneth French to every 5 min.

Recently, factor models have been applied to the statistical problem of high-dimensional covariance matrix estimation. Fan, Fan, and Lv (2008) proposed an estimator based on observable factors, imposing zero cross-sectional correlations in the regression residuals. Fan, Liao, and Mincheva (2011) established the general statistical theory of approximate factor model estimators, applying high-dimensional thresholding techniques to the residual covariance matrix. In later work, Fan, Liao, and Mincheva (2013) extended their previous methodology to include the case where factors are not observed directly, but are derived from a principal component decomposition. The estimation of factor models based on principal component analysis dates back to as early as Connor and Korajczyk (1986). More recent work, including Bai (2003), Bai and Ng (2002), and Stock and Watson (2002), establishes the asymptotic theory, though their focus is not on the consistency under matrixwise norms. Bai and Shi (2011) provided a detailed review of the literature. In this article, we focus on observed factor models because they have an established history of being used in practical applications. More importantly, their performance is astonishingly good, yet they are incredibly simple to estimate.

A growing number of methods have been proposed to overcome the curse of dimensionality inherent to the covariance estimation in the presence of a large universe of assets. One class of estimators are shrinkage estimators (Ledoit and Wolf 2004a; Ledoit and Wolf 2004b; Ledoit and Wolf 2012), which effectively shrink the eigenvalues of the sample covariance matrix toward some fixed target. Alternatively, the thresholding approach is also very popular, including Bickel and Levina (2008a, 2008b) and Cai and Liu (2011), etc. Fan and Liu (2013) and Zhou, Cai, and Ren (2014) presented a very comprehensive review of this literature. The estimation strategies in this literature often involve tuning parameters, some of which are difficult to choose or interpret. By contrast, the tuning parameters in our procedure are the number of factors and the number of blocks, which are very easy to interpret.

Our article is also related to the burgeoning field of high-frequency covariance estimation, which makes use of the newly available datasets of intraday asset prices; see Andersen et al. (2001) and Barndorff-Nielsen and Shephard (2004). A recent collection of estimators have been developed, each of which attacks simultaneously the microstructure noise and asynchronous observation issues endemic to multivariate high-frequency data. Notable references include: Aït-Sahalia, Fan, and Xiu (2010), Christensen, Kinnebrock, and Podolskij (2010), Barndorff-Nielsen et al. (2011), Zhang (2011), Shephard and Xiu (2012), and Bibinger et al. (2014). This greatly increases the number of data points on which an estimate of the covariance matrix can be calculated and therefore increases the quality of the estimated matrix and the performance of the subsequent optimized portfolios.

There are very few examples of combining the high-frequency and high-dimensional regimes. Wang and Zou (2010) established the first results that take advantage of sparsity. Fan, Li, and Yu (2012) established a concentration inequality under strong assumptions and do not make use of sparsity. Tao, Wang, and Zhou (2013) found the minimax rate for high-frequency high-dimensional estimators which take advantage of sparsity under microstructure noise. See also Tao, Wang, and Chen (2013) and Tao, Wang, and Zhou (2013) for related work. The major downfall of these methods is that they either ignore sparsity or impose sparsity directly on the covariance of the assets themselves; the former are not high-dimensional results per se, and the latter is easily empirically refuted. Our approach also bridges the high-dimensional and high-frequency approaches, however only uses the very standard assumption of sparsity on the factor model residuals. On the empirical side, few papers deal with a panel as large as we do. An exception is Lunde, Shephard, and Sheppard (2014), who also discussed portfolio allocation with S&P 500 index constituents.

The structure of the rest of the article is as follows. Section 2 sets up the model and provides the assumptions. Section 3 details the econometric analysis that provides the theoretical support for our procedure. Section 4 investigates the impact of microstructure effects on our high-dimensional estimator and guides our choice of sampling frequency. Section 5 includes an empirical study that demonstrates the performance of our approach. Section 6 concludes. The appendix contains mathematical proofs.

2. MODEL SETUP AND ASSUMPTIONS

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

We observe a large intraday panel of asset and factor prices, Y and X, at $0, \Delta_n, 2\Delta_n, \ldots, t$, where Δ_n is the sampling frequency. The model that links Y to X will be the following semiparametric continuous-time factor model,

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

where Y is a d-dimensional vector process, X is a r-dimensional *observable factor* process, Z is the idiosyncratic component, and β is a constant factor loading matrix of size $d \times r$.

To complete the specification, we make the following assumptions on the dynamics of factors and idiosyncratic components.

Assumption 1. Suppose the vector of log asset prices Y follows a factor model given by (1), in which X is a continuous Itô semimartingale, that is,

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

In addition, Z_t is another Itô semimartingale satisfying

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

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

We also need an exogeneity condition for the purpose of identification:

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

This assumption is similar in spirit to the usual exogeneity condition in a standard regression setting, but it is tailored to this in-fill asymptotics framework, see, for example, Mykland and Zhang (2006) for the case of univariate X and Z. This assumption only imposes a restriction on the sample paths of X and Z, which does not rule out any long-term dependence between these two processes.

Our main object of interest is the integrated covariance matrix of Y, denoted by $\Sigma = \frac{1}{t}\int_0^t c_s ds$, where c_s is the spot covariance of Y_s. The above Assumptions 1 and 2 impose a factor structure on the covariance matrix of Y:

$$c_s = \beta e_s \beta^\intercal + g_s, \quad 0 \le s \le t,$$

which also leads to

$$\int_0^t c_s ds = \beta \left(\int_0^t e_s ds \right) \beta^\intercal + \int_0^t g_s ds.$$

Writing $\Gamma = \frac{1}{t}\int_0^t g_s ds$ and $\mathrm{E} = \frac{1}{t}\int_0^t e_s ds$, we obtain

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

Without ambiguity, we omit the dependence of Σ, E, and Γ on t, as t is a fixed number, such as 1 day, 1 week, or 1 month. This is typical in high-frequency literature, where estimation is done on the basis of small nonoverlapping intervals.

We consider a setting where dimensions increase, as the sampling frequency shrinks to zero, that is, $\Delta_n \to 0$ and $d, r \to \infty$. In such a setting, the sample covariance matrix typically behaves poorly, and is even not necessarily invertible. While we can accommodate a rapidly increasing dimensionality, the asymptotic design below reflects the limit of our analysis:

Assumption 3. $r = o(d)$, $r^4 \Delta_n \log d = o(1)$.

In other words, the sampling frequency has to be sufficiently high to accommodate the increasing dimensionality. Moreover, to estimate Σ, additional assumptions are necessary. While a large body of the statistical literature focuses on some "sparsity" assumption of Σ, we find it more empirically appealing and economically relevant to impose a block structure on Γ.

Assumption 4. Γ is a block diagonal matrix with eigenvalues bounded from above and below by some constants almost surely. The set of its nonzero entries, S, are known prior to the estimation.

This assumption generalizes the standard "strict" factor model assumption in the literature, which requires Γ to be a diagonal matrix. Our assumption thereby falls into the so-called "approximate" factor model proposed by Chamberlain and Rothschild (1983). That said, the assumption may appear very strong, in particular the part that requires prior knowledge of blocks. While it is methodologically interesting to infer the block diagonal structure, we will not advocate doing it to maintain the simplicity of the proposed estimator. Instead, we will provide an economically sensible approach that takes advantage of the GICS information on the underlying stocks and proceed to empirically validate this assumption.

To ensure that the covariance matrix is invertible, and its inverse is estimated well, we need sufficiently many blocks to control the total number of nonzero entries in Γ.

Assumption 5. $m_d r \sqrt{\Delta_n \log d} = o(1)$, where, m_d, the degree of sparsity, is defined by

$$m_d = \max_{1\le i\le d} \sum_{1\le j\le d} 1_{\{\Gamma_{ij}\neq 0\}}.$$

Assumption 5 is a typical one in the literature of large-dimensional covariance matrix estimation, which dictates the sparsity of a matrix. In our framework, Γ is sparse whereas Σ is not, since our factors are pervasive, that is, they have nonnegligible impact on almost all assets, which drive the correlations among all stocks. We follow the literature and impose such an assumption on the matrix β.[2]

Assumption 6. $\left\|d^{-1}\beta^\intercal\beta - B\right\| = o(1)$, for some positive-definite matrix B, with $\lambda_{\min}(B)$ bounded away from 0.

3. ECONOMETRIC ANALYSIS

Inspired by the techniques for the low-frequency estimation of a factor model-based covariance matrix (Fan, Fan, and Lv 2008; Fan, Liao, and Mincheva 2011), we implement a similar plug-in type estimator in the high-frequency regime.

We stack the processes Y and X into $U = (Y^\intercal, X^\intercal)^\intercal$. Therefore, we have

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

With this representation, our estimator has a simple form. To the measurements of this $(d+r)$-dimensional process, we apply the realized covariance estimator:

$$\widehat{\Pi} = \frac{1}{t}\sum_{j=1}^{[t/\Delta_n]} \left(\Delta_j^n U\right)\left(\Delta_j^n U\right)^\intercal, \text{ where}$$
$$\Delta_j^n U = U_{j\Delta_n} - U_{(j-1)\Delta_n}. \tag{2}$$

Once $\widehat{\Pi}$ is calculated, we can further construct estimators for $\widehat{\beta}$, $\widehat{\mathrm{E}}$, and $\widehat{\Gamma}$:

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

Equivalently, we can write

$$\widehat{\Gamma} = \frac{1}{t}\sum_{i=1}^{[t/\Delta_n]} \left(\Delta_i^n Y - \widehat{\beta}^\intercal\Delta_i^n X\right)\left(\Delta_i^n Y - \widehat{\beta}^\intercal\Delta_i^n X\right)^\intercal,$$

which readily shows that $\widehat{\Gamma}$ is positive-semidefinite. The current estimator of $\widehat{\Gamma}$ does not incorporate the additional information we have about the residual matrix, that is, the sparsity. To take this into account, we will allow for various specifications of the factor models through different choices of thresholding the residual covariance matrix. We will concentrate on variants of hard thresholding and will denote a hard-thresholded matrix as

$$\widehat{\Gamma}^S = \left(\widehat{\Gamma}_{ij}1_{(i,j)\in S}\right),$$

for some choice of an index set S.

The following two choices of the index set, S, will be discussed in the empirical study.

$$S_1 = \{(i, j) \text{ such that the } i\text{th and } j\text{th assets belong to the same sector}\},$$
$$S_2 = \{(i, j) \text{ such that the } i\text{th and } j\text{th assets belong to the same industry}\}.$$

For comparison purpose, we also consider the strict factor model, which diagonalizes the residual covariance matrix.

$$S_3 = \{(i, j) \text{ such that } i = j\}.$$

We now define our plug-in estimators for the covariance matrix:

$$\widehat{\Sigma}^{S_l} = \widehat{\beta}\widehat{\mathrm{E}}\widehat{\beta}^\intercal + \widehat{\Gamma}^{S_l}, \text{ where } l = 1, 2, \text{ and } 3. \tag{4}$$

Because $\widehat{\Gamma}$ is positive-semidefinite, each of its blocks on the diagonal is positive-semidefinite, which implies that $\widehat{\Gamma}^S$ is also positive-semidefinite. Moreover, if the dimension of the largest block is smaller than the number of times-series observations minus the number of factors, then with probability 1, $\widehat{\Gamma}^S$ is positive-definite. In this case, $\widehat{\Sigma}^S$ is also positive-definite, since $\widehat{\beta}^\intercal\widehat{\mathrm{E}}\widehat{\beta}$ is positive-semidefinite. This feature of our estimator is very appealing among practioners. Moreover, we will show in Theorem 2 that the minimum eigenvalue of $\widehat{\Sigma}^S$ is bounded away from 0 with probability approaching 1.

We now discuss some asymptotic properties of the proposed estimators. In particular, we are interested in the consistency of the covariance matrix $\widehat{\Sigma}^S$ under the elementwise norm and the consistency of the precision matrix $(\widehat{\Sigma}^S)^{-1}$ under the operator norm, since these two cases are particularly relevant to the risk of solutions to the portfolio allocation problem:

$$\min_w w^\intercal\widehat{\Sigma}^S w, \quad \text{subject to } \omega^\intercal 1 = 1, \|\omega\|_1 \le \gamma, \tag{5}$$

where $\|\omega\|_1 \le \gamma$ imposes an exposure constraint, see, for example, Jagannathan and Ma (2003) and Fan, Zhang, and Yu (2012). When $\gamma = 1$, the optimal portfolio allows no short-sales, that

[2]Bickel and Levina (2008b) proposed a general notion of the degree of sparsity on Σ defined as

$$\bar{m}_d = \max_{1\le i\le d}\sum_{1\le j\le d}|\Sigma_{ij}|^q.$$

We adopt the stronger notion of sparsity on Γ to obtain a simple estimator. Our assumption is more realistic for financial data in that it imposes the sparsity on the residual covariance matrix Γ instead of on the covariance matrix Σ itself.

is, all portfolio weights are nonnegative. When γ is small and binding, the optimal portfolio is sparse, that is, many weights are zero. This portfolio is empirically appealing, as it reduces transaction and monitoring costs, for example. When γ is sufficiently large so that it is no longer a binding constraint, then the optimal portfolio coincides with the global minimum risk portfolio.

As a result, when γ is small, the entrywise norm dictates the behavior of the portfolio risk, as measured by financial econometricians, since:

$$\left|\widehat{\omega}^{*\intercal}\widehat{\Sigma}^S\widehat{\omega}^* - \omega^*\Sigma\omega^*\right| \leq \left\|\widehat{\Sigma}^S - \Sigma\right\|_{\mathrm{MAX}}\gamma^2,$$

where $\widehat{\omega}^*$ is the optimal portfolio weight with respect to $\widehat{\Sigma}^S$, which yields the perceived risk, whereas ω^* yields the theoretical optimal portfolio. The next theorem establishes the desired consistency.

Theorem 1. Under Assumptions 1–4 with $S = S_1, S_2$, or S_3, we have

$$\left\|\widehat{\Sigma}^S - \Sigma\right\|_{\mathrm{MAX}} = O_p\left(\sqrt{r^4\Delta_n\log d}\right).$$

The convergence rate of the estimator under the entrywise norm is similar to the rate obtained by Fan, Liao, and Mincheva (2011) in a low-frequency time series setting. Fan, Li, and Yu (2012) considered a noise-robust sample covariance estimator in a high-frequency setting. Their convergence rate under the entrywise norm is lower due to the effect of microstructure noise.

It is worth pointing out that the proof of this result does not depend on the additional sparsity assumption we impose on Γ, except that the block diagonal structure of Γ is known a priori, since it determines how we truncate the residual covariance matrix. It also means that this result holds even if Γ is not a block diagonal matrix (or the number of blocks is 1). In this case, our estimator becomes the usual sample covariance matrix estimator. We could even obtain a better rate for the sample covariance matrix estimator by following a similar proof to Lemma A.2 (i) in the appendix:

$$\left\|\widehat{\Sigma} - \Sigma\right\|_{\mathrm{MAX}} = O_p\left(\sqrt{\Delta_n\log d}\right),$$

for some C large enough. In this regard, when γ is very small, the advantage of our estimator is not apparent.

Nonetheless, when γ is large, the performance of portfolio allocation is dictated by the precision matrix, as can be seen from the optimal allocation of the global minimum variance portfolio, for which we have a closed form: $\omega^*_{\mathrm{GMV}} = \Sigma^{-1}\mathbf{1}/(\mathbf{1}^\intercal\Sigma^{-1}\mathbf{1})$, where $\mathbf{1}$ is a $d\times 1$ vector of 1's. The theoretical risk is given by $(\mathbf{1}^\intercal\Sigma^{-1}\mathbf{1})^{-1}$. The sample covariance matrix is not invertible when d is larger than $[t/\Delta_n]$, whereas our estimator is invertible as long as the largest block size is smaller than $[t/\Delta_n]$. The next theorem describes the asymptotic behavior of the precision matrix estimation under the operator norm.

Theorem 2. Under Assumptions 1–6, we have, for $S = S_1, S_2$, and S_3,

$$\left\|(\widehat{\Sigma}^S)^{-1} - \Sigma^{-1}\right\| = O_p\left(m_d\sqrt{r^2\Delta_n\log d}\right),$$

where m_d depends on S. Moreover, $\lambda_{\min}(\widehat{\Sigma}^S) \geq \frac{1}{2}\lambda_{\min}(\Gamma)$ holds with probability approaching 1.

The convergence rate of our estimator under the operator norm is different but comparable to the large literature of the precision matrix estimation, see, for example, Bickel and Levina (2008b), Yuan (2010), Cai, Liu, and Luo (2011), and Fan, Liao, and Mincheva (2011), under different structural and sparsity assumptions. To the best of our knowledge, our article delivers the first such result on the precision matrix for high-frequency data.

This result demonstrates the advantage of our estimator. Under additional assumptions, our covariance matrix is not only invertible, its inverse (the precision matrix) also has a bounded minimum eigenvalue with probability approaching 1. This property guarantees that our covariance estimates lead to economically feasible allocations. The operator norm is relevant for portfolio allocations because of

$$\left|\left(\mathbf{1}^\intercal\left(\widehat{\Sigma}^S\right)^{-1}\mathbf{1}\right)^{-1} - \left(\mathbf{1}^\intercal\Sigma^{-1}\mathbf{1}\right)^{-1}\right| \leq 2\left(\mathbf{1}^\intercal\Sigma^{-1}\mathbf{1}\right)^{-2}$$
$$\left\|\left(\widehat{\Sigma}^S\right)^{-1} - \Sigma^{-1}\right\|,$$

which holds with probability approaching 1. Theorems 1 and 2 turn out to be a useful guide for the empirical study below.

4. MONTE CARLO SIMULATIONS

An important choice in any study using intraday returns is the appropriate sampling frequency. Specifically, the use of high-frequency data brings with it several well-known problems. Chief among them are the asynchronicity of returns and the so-called microstructure noise. As is documented in the literature, one may either use one of several methods dedicated to overcoming these difficulties or simply take a subsample of the data for which these problems are less critical.

In this section, we examine the effect of subsampling on the performance of our estimators in the presence of both asynchronous returns and microstructure noise. We sample 100 paths from a continuous-time r-factor model of d assets specified as

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

where W_j is a standard Brownian motion and B_i is a d-dimensional Brownian motion, for $i = 1, 2, \ldots, d$, and $j = 1, 2, \ldots, r$. They are mutually independent. X_j is the jth observable factor. One of the X's is deemed the market factor, so that its associated β's are positive. The covariance matrix of Z is a block diagonal matrix, denoted by Γ, so that $\Gamma_{il} = \gamma_i^\intercal\gamma_l$. We allow for time-varying $\sigma_{j,t}$ which evolves according to the following system of equations:

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

where $\widetilde{W}_j$ is a standard Brownian motion with $\mathbb{E}[dW_{j,t}d\widetilde{W}_{j,t}] = \rho_j dt$. We choose $d = 500$ and $r = 3$. In addition, $\kappa = (3,\ 4,\ 5)$, $\theta = (0.09,\ 0.04,\ 0.06)$, $\eta = (0.3, 0.4, 0.3)$,

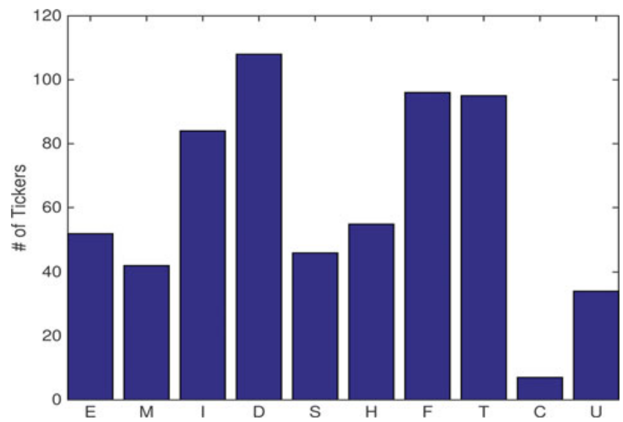

Figure 1. Sector size. The number of companies in each GICS sector on Dec. 28th, 2012: Energy (E), Materials (M), Industrials (I), Consumer Discretionary (D), Consumer Staples (S), Health Care (H), Financial (F), Information Technology (T), Telecommunication Services (C), Utilities (U).

$\rho = (-0.6, -0.4, -0.250)$, and $b = (0.05, 0.03, 0.02)$. $\beta_1 \sim \mathcal{U}[0.25, 2.25]$, and $\beta_2, \beta_3 \sim \mathcal{U}[-0.5, 0.5]$. The variances on the diagonal of Γ are uniformly generated from $[0.2, 0.5]$, with a constant within-block correlation of 0.25. In total, there are 10 blocks (of size 50×50) on the diagonal of the residual covariance matrix.

We first simulate all stocks at 1-s frequency, then contaminate and censor the data to simulate the effects of microstructure noise and asynchronous trading. As is common, we add normal random noise with mean zero and variance 0.005^2 to the simulated log prices before censoring. To censor the data, we first generate the desired number of observed prices for each of the d assets. These values are drawn from a truncated log-normal distribution so that the number of daily observations match the empirical pattern shown in Figure 1 of Lunde, Shephard, and Sheppard (2014). The log-normal distribution has parameters $\mu = 2500$ and $\sigma = 0.8$. The lower and upper truncation boundaries are 500 and 23,400, respectively. Finally, we subsample the selected number of observations within a day according to a uniform distribution. Our procedure results in a panel of transaction prices, which are contaminated by microstructure noise, and which appear to arrive according to Poisson processes. We estimate the covariance matrix based on data sampled at various frequencies using the previous-tick approach from a 21-day interval, with each day having 6.5 trading hours. We also compare with the case without microstructure noise and asynchronous observations.

In Table 1, we compare the effectiveness of subsampling in the high-dimensional setting, where many assets have very few observations. We present the errors in our two relevant norms with data which are contaminated by microstructure effects (case I) and data which are free from such effects (case II). The errors in both norms display a tradeoff between the effects of the microstructure issues and the in-fill asymptotic rates. At the fastest frequencies, we find that the microstructure problems dominate, as evidenced by the fact that the case I errors are considerably larger than the case II errors. On the other hand, at the slowest frequencies, the curse of dimensionality dominates. The sweet spot appears to be in the range between 15 and 30 min.

5. EMPIRICAL ANALYSIS

First, we describe the residual correlation structure of high-frequency returns under various factor models. We show that the high-frequency CAPM model removes some of the off-diagonal correlations leftover in the residuals, but that including the widely available sector SPDR ETF's vastly increases the sparsity of the residual correlation structure. This effect becomes pronounced when the assets are first sorted by their sector or industry group classification. This sorting induces an obvious block structure in the significant correlations of the residuals, under all the tested factor models. The block structure corresponds very well to sector groups.

Second, we will perform a classic minimum variance portfolio allocation study comparing different methods of residual thresholding, under various factor models. We will impose the exposure constraint, introduced by Fan, Zhang, and Yu (2012). Fan, Furger, and Xiu (2015) discuss the stochastic inference problem for the portfolio risk under more general constraints.

5.1 Data and Preliminaries

We include the intraday returns of all assets included in the S&P 500 index from Jan. 2004 to Dec. 2012. To strike a balance between the competing interests of using as much data as possible and alleviating the twin concerns of microstructure noise and asynchronous returns, we choose to use 15-min returns, as suggested by the above simulations. Results based on 30-min returns are similar. Each time window t is 1 month and there are 84 months in total. For each month, we have around $27 \cdot 21 = 567$ time-series observations for each of the 500 stocks. We remove overnight returns from the assets because of dividends and stock splits.

Table 1. Simulation results

	Frequency	5 s	15 s	30 s	1 min	5 min	15 min	30 min	65 min	Daily
I	$\|\widehat{\Sigma} - \Sigma\|_{\mathrm{MAX}}$	59.29	20.09	10.21	5.22	1.18	0.52	0.39	0.38	0.80
	$\|(\widehat{\Sigma}^S)^{-1} - \Sigma^{-1}\|$	6.62	6.59	6.55	6.46	5.86	5.13	4.76	25.07	*
II	$\|\widehat{\Sigma} - \Sigma\|_{\mathrm{MAX}}$	0.01	0.02	0.03	0.04	0.08	0.14	0.20	0.30	0.78
	$\|(\widehat{\Sigma}^S)^{-1} - \Sigma^{-1}\|$	0.28	0.49	0.70	1.03	2.68	6.06	11.82	36.49	*

NOTE: In this table, we report the values of $\|\widehat{\Sigma} - \Sigma\|_{\mathrm{MAX}}$ and $\|(\widehat{\Sigma}^S)^{-1} - \Sigma^{-1}\|$ for each subsampling frequency ranging from every 5 s to every day. In the case with daily sampling, there are only 21 observations for each ticker. The inverse covariance matrix estimates are ill-conditioned. Case I uses prices that are contaminated by market microstructure issues, whereas Case II is the benchmark case without using polluted prices.

For factors, we use the high-frequency returns constructed in Aït-Sahalia, Kalnina, and Xiu (2014) as our proxy for the market, small-minus-big market capitalization (SMB), and high-minus-low price-earnings ratio (HML) factors in the Fama-French 3 factor model. In addition, we use the widely available sector SDPR ETF's. These ETF's are intended to track the nine largest S&P sectors: Energy (XLE), Materials (XLB), Industrials (XLI), Consumer Discretionary (XLY), Consumer Staples (XLP), Health Care (XLV), Financial (XLF), Information Technology (XLK), Utilities (XLU). Throughout, we will make use of four factor models of increasing complexity. The "None" factor model will actually operate directly on the raw asset returns without the aid of any factor. The "CAPM" model will use the high-frequency market factor only. "CAPM + FF" will use the high-frequency market, HML, and SMB factors. Finally, "CAPM + FF + 9 IF" will use the high-frequency market, HML, and SMB factors, plus the nine- sector SDPR ETF's listed above.

We also make use of the GICS (Global Industry Classification Standard) codes from the Compustat database. These eight-digit codes are assigned to each company in the S&P 500. The code is split into four groups of two digits. Digits 1–2 describe the company's sector; digits 3–4 describe the industry group; digits 5–6 describe the industry; digits 7–8 describe the subindustry. Therefore, sorting the assets by their GICS codes results in the assets being grouped by each of the four taxonomic categories. We will find that this sorting helps elucidate an economically meaningful structure in the residual covariances.

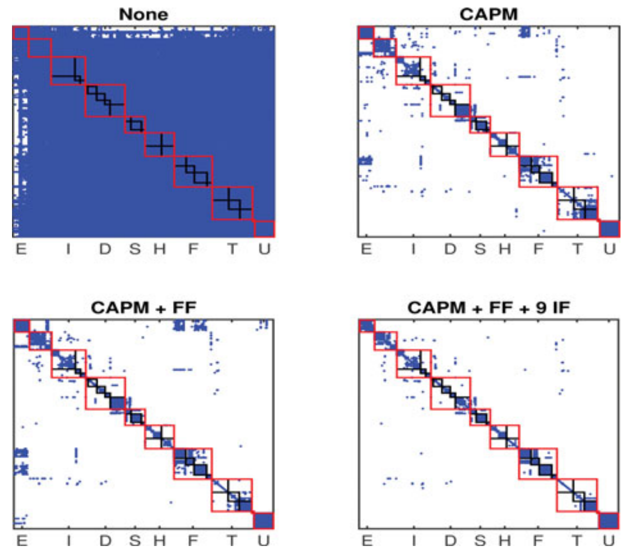

Figure 2. Pre-crisis (2004–2006) realized residual sparsity. Included blue dots are monthly residual absolute pairwise correlations that are at least 0.15 in at least 12 months over the 3-year period. The large red squares along the diagonal indicate sector blocks; the small black squares indicate industry group blocks. Correlations are calculated on the residuals after regressing raw asset returns on various factors: (top left) no factors; (top right) high-frequency market factor; (bottom left) high-frequency market factor + high-frequency SMB and HML factors; (bottom right) high-frequency market factor + high-frequency SMB and HML factors + 9 sector SPDR ETF's. Also, on the bottom of each plot, four important sectors are labeled: (E) Energy; (I) Industrials; (D) Consumer Discretionary; (S) Consumer Staples; (H) Health Care; (F) Financials; (T) Information Technology ; (U) Utilities.

5.2 Realized Residual Sparsity

In this section, we consider the realized residual sparsity of various factor models. For each model, we calculate the monthly residual covariance matrix, and then aggregate these monthly results into three economically meaningful time periods: Pre-Crisis (2004–2006), Crisis (2007–2009), Post-Crisis (2010–2012).

To calculate each monthly residual covariance matrix, we include all assets that are part of the S&P 500 over the entire month, and for which there were no missing days of data. The assets are then sorted by their GICS code at the end of the month. Finally, the covariance matrix is calculated using the estimator 3.

In Figures 2–4, we can see the realized residual sparsity patterns aggregated over the Pre-Crisis, Crisis, and Post-Crisis periods, respectively. The red (black) overlaid blocks along the diagonal correspond to sector (industry group) classifications according to the GICS codes. That is, all assets in the same square belong to the same sector (industry group). To aid in our discussion and to highlight salient features, we have also labeled several important sector blocks with their sector abbreviations. Additionally, in Table 2, we report the average R^2 of the monthly regressions, which are then averaged again across assets. Included in the table are counts of significant residual correlations that are ultimately thresholded to zero according to different imposed residual sparsity patterns. The lower the number of thresholded significant residual correlations the better the imposed sparsity pattern matches reality.

First, it is immediately obvious that there is need for some sort of parsimonious factor structure to explain the asset covariances. All three figures show that when no factors are used, the significant realized asset correlations are very nearly 100% dense. That is, one cannot reasonably impose direct sparsity on the covariances of the asset returns themselves. This refutes the empirical basis of applying the estimator of Tao, Wang, and Zhou (2013) directly on the equities. On the other hand, it is clear that not using any type of sparsity assumption, as in Fan, Li, and Yu (2012), is unlikely to give good performance when the dimensionality is as high as we consider here.

Looking more closely at the pattern of the sparsity in the three actual factor models, we see that it is highly related to the sector and industry group blocks. This is economically intuitive and a cursory analysis of the time-varying denseness of both the intra- and intersector realized residual correlations agree with the macroeconomic happenings of the time.

Regarding the intersector pattern of denseness, one sees the full 12-factor model has consistently sparse sector blocks, which are highlighted with red boxes. Interestingly, the remaining intersector correlation of the full model fits (astoundingly) well into industry group blocks, which are highlighted with black boxes. This pattern is also apparently similar throughout each of the three time periods. This can also be seen from Table 2, where the number of removed significant residual correlations in all three thresholding strategies, for the 12-factor model, is fairly constant throughout time. This is appealing from a modeling perspective because it allows a single model to be used across different time periods and economic conditions.

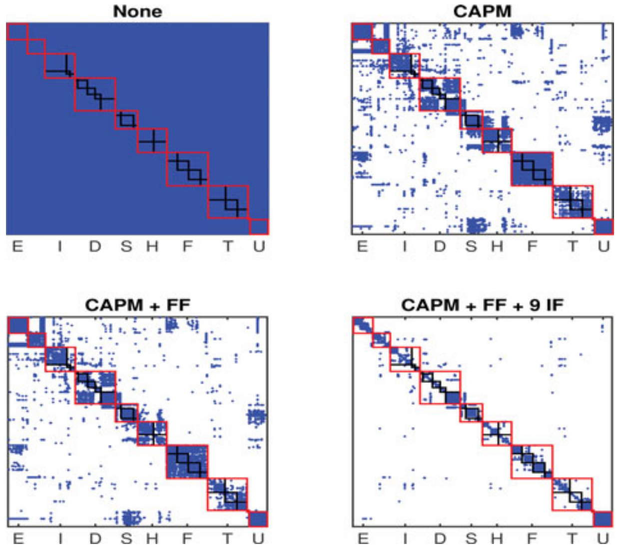

Figure 3. Crisis (2007–2009) realized residual sparsity. Included blue dots are monthly residual absolute pairwise correlations that are at least 0.15 in at least 12 months over the 3-year period. The large red squares along the diagonal indicate sector blocks; the small black squares indicate industry group blocks. Correlations are calculated on the residuals after regressing raw asset returns on various factors: (top left) no factors; (top right) high-frequency market factor; (bottom left) high-frequency market factor + high-frequency SMB and HML factors; (bottom right) high-frequency market factor + high-frequency SMB and HML factors + 9 sector SPDR ETF's. Also, on the bottom of each plot, four important sectors are labeled: (E) Energy; (I) Industrials; (D) Consumer Discretionary; (S) Consumer Staples; (H) Health Care; (F) Financials; (T) Information Technology; (U) Utilities.

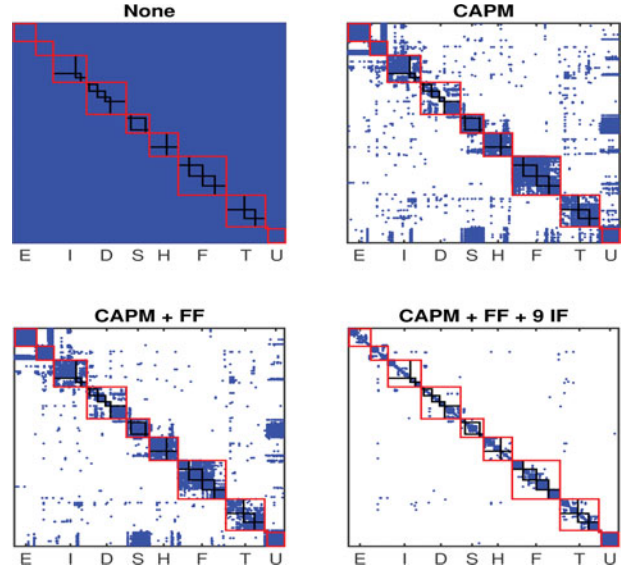

Figure 4. Post-crisis (2010–2012) realized residual sparsity. Included blue dots are monthly residual absolute pairwise correlations that are at least 0.15 in at least 12 months over the 3-year period. The large red squares along the diagonal indicate sector blocks; the small black squares indicate industry group blocks. Correlations are calculated on the residuals after regressing raw asset returns on various factors: (top left) no factors; (top right) high-frequency market factor; (bottom left) high-frequency market factor + high-frequency SMB and HML factors; (bottom right) high-frequency market factor + high-frequency SMB and HML factors + 9 sector SPDR ETF's. Also, on the bottom of each plot four important sectors are labeled: (E) Energy; (I) Industrials; (D) Consumer Discretionary; (S) Consumer Staples; (H) Health Care; (F) Financials; (T) Information Technology; (U) Utilities.

On the other hand, the CAPM and CAPM + FF models have highly time-variable denseness within the sector blocks. For instance, we see that most of the sector blocks increase in density considerably from the Pre-Crisis to the Crisis periods.

Turning now to the intrasector realized residual correlations. Again, the full model achieves a considerably more sparse estimate than the other two factor models and it maintains the same pattern throughout the total dataset. Similar to the

Table 2. Summary statistics of realized residual sparsity

		Significant correlations		
Pre-crisis	R^2	Strict	Industry group	Sector
CAPM	19.39% (6.57%)	3900	980	626
CAPM + FF	21.45% (6.53%)	3714	904	664
CAPM + FF + 9 IF	28.95% (9.00%)	2156	246	138
		Significant correlations		
Crisis	R^2	Strict	Industry group	Sector
CAPM	31.79% (7.49%)	10986	6170	2798
CAPM + FF	33.97% (7.68%)	8422	3904	1782
CAPM + FF + 9 IF	44.52% (10.09%)	2658	354	204
		Significant correlations		
Post-crisis	R^2	Strict	Industry Group	Sector
CAPM	37.12% (9.01%)	13380	6812	3198
CAPM + FF	39.00% (9.01%)	10604	4346	2274
CAPM + FF + 9 IF	48.48% (9.94%)	2560	212	110

NOTE: The results reported summarize the output of regressing raw asset returns on various factors: (CAPM) high-frequency market factor; (CAPM + FF) high-frequency market factor + high-frequency SMB and HML factors; (CAPM + FF + 9 IF) high-frequency market factor + high-frequency SMB and HML factors + 9 sector SPDR ETF's. R^2 reported is the double average of all monthly regression R^2 values over the specified time period and over all assets. The cross-sectional standard deviations of R^2 values are reported in parentheses. Nonzero's reported are the number of significant correlations shown in Figures 2–4. Strict counts number of off-diagonal significant correlations; industry group counts number of significant correlations for pairs outside of industry group blocks; sector counts number of significant correlations for pairs outside of sector blocks.

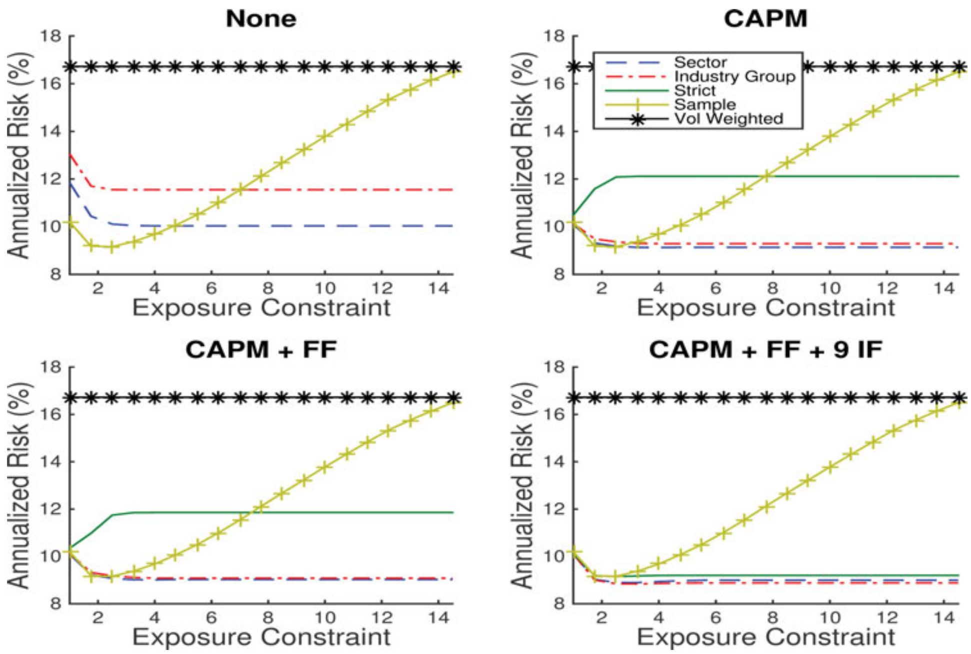

Figure 5. Out-of-sample annualized realized volatility of optimized portfolios. Shown are the annualized out-of-sample volatilities of portfolios optimized according to Equation (5) plotted against the imposed exposure constraint. $\hat{\Sigma}$ is calculated under various factor models as in Equation (4): (top left) no factors, that is sample realized volatility; (top right) high-frequency market factor; (bottom left) high-frequency market factor + high-frequency SMB and HML factors; (bottom right) high-frequency market factor + high-frequency SMB and HML factors + 9 sector SPDR ETF's. In each plot, several volatility curves are shown given by different residual thresholding techniques: (Sector) Σ^{S_1}; (Industry Group) Σ^{S_2}; (Strict) Σ^{S_3}.

intersector analysis above, the intrasector correlations of CAPM and CAPM + FF models also change through time and can be macroeconomically motivated. In the Pre-Crisis period, the correlations between the energy and financial sectors are the most predominant. This residual correlation stays prevalent in the Crisis period and, in addition, there is a new correlation between utilities and consumer staples. Finally, in the Post-Crisis period, the correlation between energy and financial sectors disappears, whereas that between Utilities and Consumer Staples stays strong.

5.3 Portfolio Study

To demonstrate the economic significance of these comparative sparsity findings, we also undertake an out-of-sample portfolio volatility minimization study. For each factor model, we rebalance monthly based on a simple volatility minimization problem with an exposure constraint, see Equation (5). This problem is chosen in lieu of more complicated models of portfolio utility, specifically, to concentrate concern on the quality of the covariance estimation. For instance, optimizing portfolio return subject to risk constraints, would also test the quality of the expected return estimation. However, it is well-known that is this is notoriously difficult. So, to not cloud the results, we use this simple portfolio optimization.

The portfolio study covers the period from Jan. 2007 to Dec. 2012. We adopt the common random walk model for forecasting future covariance matrices. Over this sample, we rebalance monthly according to Equation (5) using the current month's covariance matrix estimate. For each of the four factor models, we use one month of 15-min returns to build the factor model covariance $\hat{\Sigma} = \widehat{\Sigma}^{S_l}$ for $l = 1, 2, 3$, see Equation (4) for details. This gives (four factor models) $\cdot$ (three thresholding techniques) $= 12$ different models to be tested. The annualized out-of-sample volatility is calculated using the next month's 15-min returns. The volatility results for these 12 different models are plotted against a range of exposure constraints and are given in Figure 5.

The results show several important features. First, when applying any type of thresholding directly to the asset covariances, the results are worse than when a factor model is used. In addition, when no factor model is used, for which we expect the significant correlations to be almost completely dense, the more stringent our imposed sparsity is the worse the results. Specifically, we see that sector-based thresholding performs better than industry group thresholding which, in turn, does better than diagonalizing the covariances.[3] This is due to the model misspecification biases, created by thresholding, causing more damage than benefit, in the absence of sparsity.

In line with the work of Jagannathan and Ma (2003), we find that the sample covariance estimator shows strong performance at $c = 1$, which echoes our theory that illustrates that the entrywise error dictates the performance of the portfolio allocation at small c. Its performance degrades when the exposure constraint increases. This is because the sample covariance matrix is ill-conditioned and so increasing the exposure constraint causes more and more risky portfolios to be chosen. This also

[3]In the case of no factor model, the strict "factor model" is equivalent to the volatility weighted portfolio.

fits into our theory that at large c the operator norm of the inverse dictates the performance; under the operator norm, the sample covariance estimator is hopelessly poor. This behavior is also documented in the low-frequency regime, for example, Fan, Zhang, and Yu (2012).

Importantly, the results show that a strict factor model used in conjunction with the CAPM or CAPM + FF model yields inferior portfolios, when compared to a more lenient location-based thresholding method, namely the sector- or industry group-based approaches. It is important to note the similarity in the out-of-sample volatility of the portfolios allocated between the smaller factor models and the full model, as long as an approximate location-based factor model is used. For instance, CAPM + FF with sector thresholding performs similarly to the CAPM + FF + 9 IF model. This fits well with the findings from the previous section, that the Fama-French residuals are generally only significantly correlated inside sector blocks. Moreover, the CAPM + FF + 9 IF model performs similarly under all thresholding techniques.

These observations suggest that the CAPM and CAPM + FF models can be used for covariance estimation, but that the practitioner must pay careful attention to the sparsity of the induced residual matrix. Specifically, concerning U.S. equities, our results lead us to advocate for either an approximate Fama-French factor model with sector location-based thresholding or the larger 12-factor model. The former is much simpler and more transparent, while the latter yields a covariance matrix which is consistently strong and well-specified throughout time.

6. CONCLUSION

We introduce a simple, positive-definite covariance estimator based on high-frequency approximate factor models. This combination enables our method to take advantage of the best aspects of both the low-frequency factor model and the high-frequency data. We apply a strong residual sparsity assumption to achieve these positive aspects. However, this assumption is justified empirically and shown to be satisfied over different time periods, including throughout the recent financial crisis. This sparsity structure is found to have a strong and striking relation to sector inclusion. We conclude that the Fama-French factors are only suitable for covariance matrix estimation when used with an approximate location-based thresholding method. We also recommend the inclusion of the SPDR sector ETF's as factors.

Theoretically, we provide the dual in-fill and diverging dimensionality asymptotics for our proposed covariance and precision matrix estimators. We provide the probability error bounds for the covariance matrix estimation under the entrywise sup-norm, as well as the operator norm error bound for the precision matrix. The theory is built upon the interplay between the established concentration inequalities for Martingales and matrix norm-related calculations.

An important question for future work is whether a purely statistical factor model such as Fan, Liao, and Mincheva (2011) or Fan, Liao, and Mincheva (2013) is competitive with an estimator that takes advantage of the domain knowledge, such as what we have proposed. In addition, a thorough analysis of the value of different thresholding techniques (e.g., hard, soft, or SCAD; see Fan and Li (2001)), would help guide practitioners' choices.

APPENDIX A: MATHEMATICAL PROOFS

Lemma A.1. Suppose that U_t is a p-dimensional vector process satisfying $U_t = \int_0^t b_s ds + \int_0^t \sigma_s dW_s$, where b_t is a $p \times 1$ vector, σ_t is a $p \times q$ vector, and W_t is a $q \times 1$ standard Brownian motion. Also, we assume that $\|U_t\|_\infty \le K$, $\|b_t\|_\infty \le K$, and $\left\|\sigma_t\sigma_t^\intercal\right\|_{\mathrm{MAX}} \le K$, for some $K > 0$. Suppose that $\Delta_i^n U_j = U_{i\Delta_n, j} - U_{(i-1)\Delta_n, j}$, where $1 \le j \le p$. Then, there exists some constants C_1 and C_2, such that for any $1 \le j, m \le p$, and for any $x > 0$,

$$\mathbb{P}\left(\left|\sum_{i=0}^{[t/\Delta_n]} (\Delta_i^n U_j)(\Delta_i^n U_m) - \int_0^t (\sigma\sigma^\intercal)_{s,jm} ds\right| > x\right) \le C_1 \exp\left(-\frac{C_2 x^2}{\Delta_n}\right).$$

Proof. The proof of this Lemma is similar to the proofs of Lemma 3(i) in Fan, Li, and Yu (2012) and Lemma 10 in Tao, Wang, and Zhou (2013).

Denote $U_t^* = \int_0^t \sigma_s dW_s$, and denote for $1 \le i \le [t/\Delta_n]$, $1 \le j, m \le p$,

$$\zeta_{i,jm} = \left(\Delta_i^n U_j^*\right)\left(\Delta_i^n U_m^*\right),\ \zeta'_{i,jm} = \mathbb{E}\left(\left(\Delta_i^n U_j^*\right)\left(\Delta_i^n U_m^*\right)|\mathcal{F}_{(i-1)\Delta_n}\right), \text{ and}$$

$$\zeta''_{i,jm} = \zeta_{i,jm} - \zeta'_{i,jm},$$

then $M_t = \sum_{i=1}^{[t/\Delta_n]} \zeta''_{i,jm}$ is a continuous-time martingale. By Itô's lemma, we have

$$\begin{aligned}&\left(U_{t,j}^* - U_{s,j}^*\right)\left(U_{t,m}^* - U_{s,m}^*\right) - \int_s^t (\sigma\sigma^\intercal)_{v,jm}\, dv \\ &= \int_s^t \left(U_{v,j}^* - U_{s,j}^*\right)(\sigma_v dW_v)_m \\ &\quad + \int_s^t \left(U_{v,m}^* - U_{s,m}^*\right)(\sigma_v dW_v)_j\,.\end{aligned}$$

Therefore

$$\begin{aligned}\zeta''_{i,jm} &= \int_{(i-1)\Delta_n}^{i\Delta_n} \left(U_{s,j}^* - U_{(i-1)\Delta_n,j}^*\right)(\sigma_s dW_s)_m \\ &\quad + \int_{(i-1)\Delta_n}^{i\Delta_n} \left(U_{s,m}^* - U_{(i-1)\Delta_n,m}^*\right)(\sigma_s dW_s)_j.\end{aligned}$$

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

$$\begin{aligned}&\left|\sum_{i=1}^{[t/\Delta_n]} (\Delta_i^n U_j)(\Delta_i^n U_m) - \int_0^t (\sigma\sigma^\intercal)_{s,jm} ds\right| \\ &= \left|\sum_{i=1}^{[t/\Delta_n]} \Delta_i^n U_j^* \int_{(i-1)\Delta_n}^{i\Delta_n} b_{s,m} ds\right. \\ &\quad + \Delta_i^n U_m^* \sum_{i=1}^{[t/\Delta_n]} \int_{(i-1)\Delta_n}^{i\Delta_n} b_{s,j} ds + \sum_{i=1}^{[t/\Delta_n]} \int_{(i-1)\Delta_n}^{i\Delta_n} b_{s,m} ds \\ &\quad \left.\int_{(i-1)\Delta_n}^{i\Delta_n} b_{s,j} ds + M_t\right|.\end{aligned}$$

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

The quadratic variation of M_t is given by

$$\langle M,M\rangle_t = \Delta_n \sum_{i=1}^{[t/\Delta_n]} \int_{(i-1)\Delta_n}^{i\Delta_n} \Bigg(\Big(U^*_{s,j} - U^*_{(i-1)\Delta_n,j}\Big)^2 \sum_{l=1}^{q} \sigma^2_{s,ml} + \Big(U^*_{s,m} - U^*_{(i-1)\Delta_n,m}\Big)^2 \sum_{l=1}^{q} \sigma^2_{s,jl} + 2\Big(U^*_{s,j} - U^*_{(i-1)\Delta_n,j}\Big)\Big(U^*_{s,m} - U^*_{(i-1)\Delta_n,m}\Big) \sum_{l=1}^{q} \sigma_{s,ml}\sigma_{s,jl} \Bigg) ds.$$

Therefore by Cauchy–Schwarz inequality, we have

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

Then, by the exponential inequality for continuous Martingale, we have

$$\mathbb{P}\left(\sup_{0\le s\le t} \left| \sum_{i=1}^{[s/\Delta_n]} \zeta''_{i,jm} \right| > x \right) \le \exp\left(-\frac{x^2}{32K^3 t\Delta_n} \right). \tag{A.1}$$

In addition, by Cauchy–Schwarz inequality:

$$\mathbb{P}\left(\left| \sum_{i=1}^{[t/\Delta_n]} \Delta_i^n U_j^* \int_{(i-1)\Delta_n}^{i\Delta_n} b_{s,m} ds \right| > x \right)$$
$$\le \mathbb{P}\left(\sum_{i=1}^{[t/\Delta_n]} (\Delta_i^n U_j^*)^2 - \int_0^t (\sigma\sigma^\intercal)_{s,jj} ds > \frac{x^2}{tK\Delta_n} - \int_0^t (\sigma\sigma^\intercal)_{s,jj} ds \right)$$
$$\le \mathbb{P}\left(\left| \sum_{i=1}^{[t/\Delta_n]} (\Delta_i^n U_j^*)^2 - \int_0^t (\sigma\sigma^\intercal)_{s,jj} ds \right| > \frac{x^2}{tK\Delta_n} - tK \right)$$
$$\le \exp\left(-\frac{\left(\frac{x^2}{tK\Delta_n} - tK\right)^2}{32K^3 t\Delta_n} \right),$$

where the last inequality follows from (A.1).

Finally, notice that

$$\left| \sum_{i=1}^{[t/\Delta_n]} \int_{(i-1)\Delta_n}^{i\Delta_n} b_{s,m} ds \int_{(i-1)\Delta_n}^{i\Delta_n} b_{s,j} ds \right| \le tK^2\Delta_n,$$

we can derive

$$\mathbb{P}\left(\left| \sum_{i=1}^{[t/\Delta_n]} (\Delta_i^n U_j)(\Delta_i^n U_m) - \int_0^t (\sigma\sigma^\intercal)_{s,jm} ds \right| > 4x \right)$$
$$\le \mathbb{P}(|M_t| > x) + 2\mathbb{P}\left(\left| \sum_{i=1}^{[t/\Delta_n]} \Delta_i^n U_j^* \int_{(i-1)\Delta_n}^{i\Delta_n} b_{s,m} ds \right| > x \right) + 1_{\{tK^2\Delta_n > x\}}$$
$$\le \exp\left(-\frac{x^2}{32K^3 t\Delta_n} \right) + 2\exp\left(-\frac{\left(\frac{x^2}{tK\Delta_n} - tK\right)^2}{32K^3 t\Delta_n} \right)$$
$$\le C_1 \exp\left(-\frac{16C_2 x^2}{\Delta_n} \right),$$

where the above inequality holds if $x > (tK^2\Delta_n) \vee (tK\sqrt{\Delta_n}) \vee (\frac{tK\Delta_n}{2}\sqrt{1+\frac{4}{\Delta_n}})$, and $C_1 \ge 3$, $C_2 \le (512K^3 t)^{-1}$, which readily implies Lemma A.1. On the other hand, if x violates this bound, that is $x \le C'\sqrt{\Delta_n}$, we can choose C_1 such that $C_1 \exp(-16C_2 C'^2) \ge 1$, so that the inequality follows trivially. □

Lemma A.2. Suppose that the entrywise max norms of all the processes are bounded uniformly in $[0,t]$. Under Assumptions 1, 2, and 3, for some C_0 large enough, we have

$$\text{(i)}\quad \mathbb{P}\left(\left\|\widehat{\mathrm{E}} - \mathrm{E}\right\|_{\mathrm{MAX}} > C_0\sqrt{\Delta_n \log d} \right) = O(r^2 d^{-3}), \tag{A.2}$$

$$\text{(ii)}\quad \mathbb{P}\left(\left\|\widehat{\mathrm{E}} - \mathrm{E}\right\|_{\mathrm{F}} > C_0 r\sqrt{\Delta_n \log d} \right) = O(r^2 d^{-3}), \tag{A.3}$$

$$\text{(iii)}\quad \mathbb{P}\left(\max_{1\le j\le r, 1\le l\le d} \left| \sum_{i=1}^{[t/\Delta_n]} (\Delta_i^n X_j)(\Delta_i^n Z_l) \right| > C_0\sqrt{\Delta_n \log d} \right) = O(rd^{-2}), \tag{A.4}$$

$$\text{(iv)}\quad \mathbb{P}\left(\max_{1\le l,k\le d} \left| \sum_{i=1}^{[t/\Delta_n]} (\Delta_i^n Z_l)(\Delta_i^n Z_k) - \int_0^t g_{s,lk} ds \right| > C_0\sqrt{\Delta_n \log d} \right) = O(d^{-1}), \tag{A.5}$$

$$\text{(v)}\quad \mathbb{P}\left(\max_{1\le l\le d} \left\|\widehat{\beta}_l - \beta_l\right\| > C_0\sqrt{\Delta_n r \log d} \right) = O(rd^{-2}), \tag{A.6}$$

$$\text{(vi)}\quad \mathbb{P}\left(\left\|\widehat{\beta} - \beta\right\|_{\mathrm{F}} > C_0\sqrt{\Delta_n r d \log d} \right) = O(rd^{-2}), \tag{A.7}$$

$$\text{(vii)}\quad \mathbb{P}\left(\max_{1\le l\le d} \sum_{i=1}^{[t/\Delta_n]} \left((\beta - \widehat{\beta})^\intercal \Delta_i^n X\right)_l^2 > C_0 \Delta_n r^2 \log d \right) = O(rd^{-2}), \tag{A.8}$$

$$\text{(viii)}\quad \mathbb{P}\left(\max_{1\le l,k\le d} \left|\widehat{\Gamma}_{lk} - \Gamma_{lk}\right| \ge C_0 r\sqrt{\Delta_n \log d} \right) = O(d^{-1}). \tag{A.9}$$

Proof. (i)–(iv) By Bonferroni inequality and Lemma A.1, we can show that

$$\mathbb{P}\left(\max_{1\le j,m\le r} \left| \sum_{i=1}^{[t/\Delta_n]} (\Delta_i^n X_j)(\Delta_i^n X_m) - \int_0^t e_{s,jm} ds \right| > C_0\sqrt{\Delta_n \log d} \right) \le r^2 \times C_1 \exp\left(-\frac{C_2(C_0^2 \Delta_n \log d)}{\Delta_n} \right) = C_1 r^2 d^{-C_2 C_0^2}.$$

Similarly, as X and Z are orthogonal, we have

$$\mathbb{P}\left(\max_{1\le j\le r, 1\le l\le d} \left| \sum_{i=1}^{[t/\Delta_n]} (\Delta_i^n X_j)(\Delta_i^n Z_l) \right| > C_0\sqrt{\Delta_n \log d} \right) \le r \times d \times C_1 \exp\left(-\frac{C_2 C_0^2 \Delta_n \log d}{\Delta_n} \right) = C_1 r d^{1 - C_2 C_0^2}.$$

In addition, we have

$$\mathbb{P}\left(\max_{1\le l,k\le d} \left| \sum_{i=1}^{[t/\Delta_n]} (\Delta_i^n Z_l)(\Delta_i^n Z_k) - \int_0^t g_{s,lk} ds \right| > C_0\sqrt{\Delta_n \log d} \right) \le d^2 \times C_1 \exp\left(-\frac{C_2 C_0^2 \Delta_n \log d}{\Delta_n} \right) = C_1 d^{2 - C_2 C_0^2}.$$

The first three inequalities hold as long as C_0 is large enough such that $C_0^2 C_2 \ge 3$. Inequality (A.3) follows directly from (A.2) and

$$\left\|\widehat{\mathrm{E}} - \mathrm{E}\right\|_{\mathrm{F}} \le r \left\|\widehat{\mathrm{E}} - \mathrm{E}\right\|_{\mathrm{MAX}}.$$

(v)–(vi) Moreover, note that

$$\widehat{\beta}_l - \beta_l = \left(\sum_{i=1}^{[t/\Delta_n]} (\Delta_i^n X)(\Delta_i^n X)^\intercal \right)^{-1} \sum_{i=1}^{[t/\Delta_n]} (\Delta_i^n X)\Delta_i^n Z_l,$$

therefore under the event that

$$\mathrm{A} = \left\{ \max_{1\le j\le r, 1\le l\le d} \left| \sum_{i=1}^{[t/\Delta_n]} \Delta_i^n X_j \Delta_i^n Z_l \right| \le C_0' \sqrt{\Delta_n \log d} \right\} \cap \left\{ \lambda_{\min} \left(\sum_{i=1}^{[t/\Delta_n]} (\Delta_i^n X)(\Delta_i^n X)^\intercal \right) \ge \frac{1}{2} \lambda_{\min} \left(\int_0^t e_s ds \right) \right\}$$

we have

$$\|\widehat{\beta}_l - \beta_l\|^2 \le \frac{4}{\lambda_{\min}^2 \left(\int_0^t e_s ds \right)} \sum_{j=1}^r \left(\sum_{i=1}^{[t/\Delta_n]} \Delta_i^n X_j \Delta_i^n Z_l \right)^2 \le \frac{4r(C_0')^2 \Delta_n \log d}{\lambda_{\min}^2 \left(\int_0^t e_s ds \right)}.$$

and

$$\|\widehat{\beta} - \beta\|_{\mathrm{F}}^2 \le \frac{4r(C_0')^2 \Delta_n d \log d}{\lambda_{\min}^2 \left(\int_0^t e_s ds \right)}.$$

Therefore, it suffices to show that $\mathbb{P}(\mathrm{A}) \ge 1 - O(rd^{-2})$.

Since we assume $\lambda_{\min} \left(\int_0^t e_s ds \right)$ is bounded away from 0 and that $r = o((\Delta_n \log d)^{-1/2})$, it follows that

$$\mathbb{P}\left(\left\| \sum_{i=1}^{[t/\Delta_n]} (\Delta_i^n X)(\Delta_i^n X)^\intercal - \int_0^t e_s ds \right\| \le \frac{1}{2} \lambda_{\min} \left(\int_0^t e_s ds \right) \right) \ge \mathbb{P}\left(r \max_{1\le j,m\le r} \left| \sum_{i=1}^{[t/\Delta_n]} (\Delta_i^n X_j)(\Delta_i^n X_m) - \int_0^t e_{jm,s} ds \right| \le \frac{1}{2} \lambda_{\min} \left(\int_0^t e_s ds \right) \right) \ge 1 - O(r^2 d^{-3}),$$

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

$$\mathbb{P}\left(\lambda_{\min} \left(\sum_{i=1}^{[t/\Delta_n]} (\Delta_i^n X)(\Delta_i^n X)^\intercal \right) \ge \frac{1}{2} \lambda_{\min} \left(\int_0^t e_s ds \right) \right) \ge 1 - O(r^2 d^{-3}).$$

Combining this with (A.4), we have $\mathbb{P}(\mathrm{A}) \ge 1 - O(rd^{-2})$.

(vii) To prove (A.8), note that

$$\max_{1\le l\le d} \sum_{i=1}^{[t/\Delta_n]} \left(\sum_{j=1}^r (\beta_{lj} - \widehat{\beta}_{lj}) \Delta_i^n X_j \right)^2 \le \max_{1\le l\le d} \|\widehat{\beta}_l - \beta_l\|^2 \sum_{i=1}^{[t/\Delta_n]} \|\Delta_i^n X\|^2,$$

that by (A.2) with $C > \max_{1\le j\le r} \int_0^t e_{s,jj} ds$,

$$\mathbb{P}\left(\sum_{i=1}^{[t/\Delta_n]} \|\Delta_i^n X\|^2 \le rC \right) \ge \mathbb{P}\left(r \max_{1\le j\le r} \left| \sum_{i=1}^{[t/\Delta_n]} (\Delta_i^n X_j)^2 - \int_0^t e_{s,jj} ds \right| + r \max_{1\le j\le r} \int_0^t e_{s,jj} ds \le rC \right) \ge 1 - O(r^2 d^{-3}),$$

and (A.6), we have

$$\mathbb{P}\left(\max_{1\le l\le d} \sum_{i=1}^{[t/\Delta_n]} \left(\sum_{j=1}^r (\beta_{lj} - \widehat{\beta}_{lj}) \Delta_i^n X_j \right)^2 > C \Delta_n r^2 \log d \right) \le O(rd^{-2}).$$

(viii) Finally, note that under the event of

$$\left\{ \max_{1\le j\le d} \left| \sum_{i=1}^{[t/\Delta_n]} (\Delta_i^n Z_j)^2 - \int_0^t g_{s,jj} ds \right| \le \frac{1}{4} \max_{1\le j\le d} \int_0^t g_{s,jj} ds \right\} \cap \left\{ \max_{1\le l\le d} \sum_{i=1}^{[t/\Delta_n]} \left(\sum_{j=1}^r (\beta_{lj} - \widehat{\beta}_{lj}) \Delta_i^n X_j \right)^2 \le C_0 \Delta_n r^2 \log d \right\},$$

it follows from the Cauchy–Schwarz inequality that

$$\begin{aligned}
&\max_{1\le l,k\le d} \left| \sum_{i=1}^{[t/\Delta_n]} \left(\Delta_i^n Y_l - (\widehat{\beta} \Delta_i^n X)_l \right) \left(\Delta_i^n Y_k - (\widehat{\beta} \Delta_i^n X)_k \right) - \sum_{i=1}^{[t/\Delta_n]} \left(\Delta_i^n Z_l \right) \left(\Delta_i^n Z_k \right) \right| \\
&\le \max_{1\le l,k\le d} \left| \sum_{i=1}^{[t/\Delta_n]} \left((\widehat{\beta} - \beta) \Delta_i^n X \right)_l \left((\widehat{\beta} - \beta) \Delta_i^n X \right)_k \right| + 2 \max_{1\le l,k\le d} \left| \sum_{i=1}^{[t/\Delta_n]} \Delta_i^n Z_l \left((\widehat{\beta} - \beta) \Delta_i^n X \right)_k \right| \\
&\le \max_{1\le l\le d} \sum_{i=1}^{[t/\Delta_n]} \left((\widehat{\beta} - \beta) \Delta_i^n X \right)_l^2 + 2 \sqrt{ \max_{1\le l\le d} \sum_{i=1}^{[t/\Delta_n]} \left(\Delta_i^n Z_l \right)^2 \max_{1\le l\le d} \sum_{i=1}^{[t/\Delta_n]} \left((\widehat{\beta} - \beta) \Delta_i^n X \right)_l^2 } \\
&\le C_0 \Delta_n r^2 \log d + 2 \sqrt{ \left(\frac{5}{4} Ct \right) \left(C_0 \Delta_n r^2 \log d \right) } \\
&\le C_0' \sqrt{\Delta_n r^2 \log d},
\end{aligned}$$

as long as $C_0' \ge 3\sqrt{C_0 C t}$, where we use $\|g_s\|_{\mathrm{MAX}} \le C$ on $[0, t]$ and the fact that $\Delta_n r^2 \log d = o(1)$. Consequently, we have

$$\max_{1\le l,k\le d} \left| \sum_{i=1}^{[t/\Delta_n]} \left(\Delta_i^n Y_l - (\widehat{\beta} \Delta_i^n X)_l \right) \left(\Delta_i^n Y_k - (\widehat{\beta} \Delta_i^n X)_k \right) - \sum_{i=1}^{[t/\Delta_n]} \left(\Delta_i^n Z_l \right) \left(\Delta_i^n Z_k \right) \right| \le C_0' \sqrt{\Delta_n r^2 \log d}$$

with probability at least $1 - O(d^{-1})$, by (A.5) and (A.8). Finally, by the triangle inequality, we obtain

$$\max_{1\le l,k\le d} \left| \widehat{\Gamma}_{lk} - \Gamma_{lk} \right| \le \max_{1\le l,k\le d} \left| \sum_{i=1}^{[t/\Delta_n]} \left(\Delta_i^n Z_l \right) \left(\Delta_i^n Z_k \right) - \int_0^t g_{s,lk} ds \right| + \max_{1\le l,k\le d} \left| \sum_{i=1}^{[t/\Delta_n]} \left(\Delta_i^n Y_l - (\widehat{\beta} \Delta_i^n X)_l \right) \left(\Delta_i^n Y_k - (\widehat{\beta} \Delta_i^n X)_k \right) - \sum_{i=1}^{[t/\Delta_n]} \left(\Delta_i^n Z_l \right) \left(\Delta_i^n Z_k \right) \right|,$$

which yields the desired result by using (A.5).

□

Proof of Theorem 1. This proof follows similar steps in Fan, Liao, and Mincheva (2011) and Fan, Fan, and Lv (2008). We first establish the result under a stronger assumption that the entrywise max norms of all the processes are bounded uniformly in $[0, t]$. By straightforward calculations, we have

$$\begin{aligned}\|\widehat{\Sigma}^S - \Sigma\|_{\mathrm{MAX}} &\le \|2(\widehat{\beta} - \beta)\mathrm{E}\beta^\intercal\|_{\mathrm{MAX}} + \|\beta(\widehat{\mathrm{E}} - \mathrm{E})\beta^\intercal\|_{\mathrm{MAX}} \\ &+ \|(\widehat{\beta} - \beta)\mathrm{E}(\widehat{\beta} - \beta)^\intercal\|_{\mathrm{MAX}} + \|2\beta(\widehat{\mathrm{E}} - \mathrm{E})(\widehat{\beta} - \beta)^\intercal\|_{\mathrm{MAX}} \\ &+ \|(\widehat{\beta} - \beta)(\widehat{\mathrm{E}} - \mathrm{E})(\widehat{\beta} - \beta)^\intercal\|_{\mathrm{MAX}} + \|\widehat{\Gamma}^S - \Gamma\|_{\mathrm{MAX}} \qquad (A.10)\end{aligned}$$

We now need Lemma A.2. Under the events of

$$\begin{aligned}&\|\widehat{\mathrm{E}} - \mathrm{E}\|_{\mathrm{MAX}} \le C\sqrt{\Delta_n \log d}, \quad \lambda_{\min}\left(\sum_{i=1}^{[t/\Delta_n]} (\Delta_i^n X)(\Delta_i^n X)^\intercal\right) \\ &\ge \frac{1}{2}\lambda_{\min}\left(\int_0^t e_s ds\right), \max_{1\le l\le d} \|\widehat{\beta}_l - \beta_l\| \le C\sqrt{\Delta_n r \log d}, \\ &\max_{1\le j\le r, 1\le l\le d}\left|\sum_{i=1}^{[t/\Delta_n]} (\Delta_i^n X_j)(\Delta_i^n Z_l)\right| \le C\sqrt{\Delta_n \log d}, \\ &\|\widehat{\Gamma}^S - \Gamma\|_{\mathrm{MAX}} \le \|\widehat{\Gamma} - \Gamma\|_{\mathrm{MAX}} \le Cr\sqrt{\Delta_n \log d},\end{aligned}$$

for some C sufficiently large, and using $\|\mathrm{E}\| \le r\|\mathrm{E}\|_{\mathrm{MAX}} \le rC$, $\max_{1\le k\le d}\|\beta_k\| \le \sqrt{r}\|\beta\|_{\mathrm{MAX}}$, and $\|\widehat{\beta} - \beta\|_{\mathrm{MAX}} \le \sqrt{r}\max_{1\le l\le d}\|\widehat{\beta}_l - \beta_l\|$, we have

$$\begin{aligned}&\|2(\widehat{\beta} - \beta)\mathrm{E}\beta^\intercal\|_{\mathrm{MAX}} = 2\max_{1\le l,k\le d}\|e_l^\intercal(\widehat{\beta} - \beta)\mathrm{E}\beta^\intercal e_k\| \\ &\le 2\max_{1\le l\le d}\|\widehat{\beta}_l - \beta_l\|\,\|\mathrm{E}\|\max_{1\le k\le d}\|\beta_k\| \le C'r^2\sqrt{\Delta_n \log d}; \\ &\|\beta(\widehat{\mathrm{E}} - \mathrm{E})\beta^\intercal\|_{\mathrm{MAX}} \le r^2\|\beta\|_{\mathrm{MAX}}^2\|\widehat{\mathrm{E}} - \mathrm{E}\|_{\mathrm{MAX}} \le C'r^2\sqrt{\Delta_n \log d}; \\ &\|(\widehat{\beta} - \beta)\mathrm{E}(\widehat{\beta} - \beta)^\intercal\|_{\mathrm{MAX}} \le \max_{1\le l\le d}\|e_l^\intercal(\widehat{\beta} - \beta)\|^2\|\mathrm{E}\| \le C'\Delta_n r^2 \log d; \\ &\|2\beta(\widehat{\mathrm{E}} - \mathrm{E})(\widehat{\beta} - \beta)^\intercal\|_{\mathrm{MAX}} \le 2r^2\|\beta\|_{\mathrm{MAX}}\|\widehat{\mathrm{E}} - \mathrm{E}\|_{\mathrm{MAX}}\|\widehat{\beta} - \beta\|_{\mathrm{MAX}} \\ &\le C'r^3\Delta_n \log d; \|(\widehat{\beta} - \beta)(\widehat{\mathrm{E}} - \mathrm{E})(\widehat{\beta} - \beta)^\intercal\|_{\mathrm{MAX}} \\ &\le r^2\|\widehat{\beta} - \beta\|_{\mathrm{MAX}}^2\|\widehat{\mathrm{E}} - \mathrm{E}\|_{\mathrm{MAX}} \le C'r^4\Delta_n \log d\sqrt{\Delta_n \log d}.\end{aligned}$$

Finally, combining all these estimates, we obtain

$$\|\widehat{\Sigma}^S - \Sigma\|_{\mathrm{MAX}} \le C\sqrt{r^4\Delta_n \log d}.$$

By Lemma A.2, these events occur with probability no less than $1 - O(d^{-1})$. By the localization argument, we can weaken the boundedness assumption, hence $\|\widehat{\Sigma}^S - \Sigma\|_{\mathrm{MAX}} = O_p(\sqrt{r^4\Delta_n \log d})$. □

Lemma A.3. Suppose that the entrywise max norms of all the processes are bounded uniformly in $[0, t]$. Under Assumptions 1–5, we have

$$\text{(i) } \mathbb{P}\left(\|\widehat{\Gamma}^S - \Gamma\| > C_0 m_d r\sqrt{\Delta_n \log d}\right) = O(d^{-1}), \qquad (A.11)$$

$$\text{(ii) } \mathbb{P}\left(\lambda_{\min}\left(\widehat{\Gamma}^S\right) \ge \frac{1}{2}\lambda_{\min}(\Gamma)\right) > 1 - O(d^{-1}), \qquad (A.12)$$

$$\text{(iii) } \mathbb{P}\left(\|(\widehat{\Gamma}^S)^{-1} - \Gamma^{-1}\| > C_0 m_d r\sqrt{\Delta_n \log d}\right) = O(d^{-1}), \qquad (A.13)$$

$$\begin{aligned}\text{(iv) } &\mathbb{P}\left(\|\widehat{\beta}^\intercal(\widehat{\Gamma}^S)^{-1}\widehat{\beta} - \beta^\intercal\Gamma^{-1}\beta\| > C_0 m_d r\sqrt{\Delta_n d^2 \log d}\right) \\ &= O(d^{-1}), \qquad (A.14)\end{aligned}$$

$$\begin{aligned}\text{(v) } &\mathbb{P}\left(\|(\widehat{\mathrm{E}}^{-1} + \widehat{\beta}^\intercal(\widehat{\Gamma}^S)^{-1}\widehat{\beta}) - (\mathrm{E}^{-1} + \beta^\intercal(\Gamma^S)^{-1}\beta)\|\right. \\ &\left.> C_0 m_d r\sqrt{\Delta_n d^2 \log d}\right) = O(d^{-1}), \qquad (A.15)\end{aligned}$$

$$\text{(vi) } \mathbb{P}\left(\left\|\left(\widehat{\mathrm{E}}^{-1} + \widehat{\beta}^\intercal(\widehat{\Gamma}^S)^{-1}\widehat{\beta}\right)^{-1}\right\| > C_0 d^{-1}\right) = O(d^{-1}), \qquad (A.16)$$

$$\begin{aligned}\text{(vii) } &\mathbb{P}\left(\left\|\widehat{\beta}\left(\widehat{\mathrm{E}}^{-1} + \widehat{\beta}^\intercal(\widehat{\Gamma}^S)^{-1}\widehat{\beta}\right)^{-1}\widehat{\beta}^\intercal(\widehat{\Gamma}^S)^{-1}\right\| > C_0\right) \\ &= O(d^{-1}), \qquad (A.17)\end{aligned}$$

$$\begin{aligned}\text{(viii) } &\mathbb{P}\left(\left\|\widehat{\beta}\left(\widehat{\mathrm{E}}^{-1} + \widehat{\beta}^\intercal(\widehat{\Gamma}^S)^{-1}\widehat{\beta}\right)^{-1}\widehat{\beta}^\intercal\Gamma^{-1}\right\| > C_0\right) \\ &= O(d^{-1}). \qquad (A.18)\end{aligned}$$

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

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

Then, by (A.9), with probability no less than $O(d^{-1})$, we have

$$\|\widehat{\Gamma}^S - \Gamma\| \le m_d \max_{1\le l,k\le d}|\widehat{\Gamma}_{lk} - \Gamma_{lk}| \le C_0 m_d r\sqrt{\Delta_n \log d}.$$

As we can see, this proof does not require $m_d r\sqrt{\Delta_n \log d} = o(1)$.

(ii) By Lemma A.1 of Fan, Liao, and Mincheva (2011), we have

$$\begin{aligned}&\mathbb{P}\left(\lambda_{\min}\left(\widehat{\Gamma}^S\right) \ge 0.5\lambda_{\min}(\Gamma)\right) \ge \mathbb{P}\left(\|\widehat{\Gamma}^S - \Gamma\|\right. \\ &\left.\le 0.5\lambda_{\min}(\Gamma)\right) > 1 - O(d^{-1}),\end{aligned}$$

where we use the fact that $m_d r\sqrt{\Delta_n \log d} = o(1)$ and $\lambda_{\min}(\Gamma) > K$. This inequality warrants the invertibility of $\widehat{\Gamma}^S$.

(iii) We note that under the event of $\lambda_{\min}\left(\widehat{\Gamma}^S\right) \ge 0.5\lambda_{\min}(\Gamma)$, we have

$$\begin{aligned}&\mathbb{P}\left(\|(\widehat{\Gamma}^S)^{-1} - \Gamma^{-1}\| \le 2C\|\Gamma^{-1}\| m_d r\sqrt{\Delta_n \log d}\right) \\ &\ge \mathbb{P}\left(\|(\widehat{\Gamma}^S)^{-1} - \Gamma^{-1}\| \le 2C\|\Gamma^{-1}\|\cdot\|\widehat{\Gamma}^S - \Gamma\|, \|\widehat{\Gamma}^S - \Gamma\|\right. \\ &\left.\le m_d r\sqrt{\Delta_n \log d}\right) \ge \mathbb{P}\left(\|(\widehat{\Gamma}^S)^{-1} - \Gamma^{-1}\| \le 2C\|\Gamma^{-1}\|\cdot\|\widehat{\Gamma}^S - \Gamma\|\right) \\ &-\mathbb{P}\left(\|\widehat{\Gamma}^S - \Gamma\| > m_d r\sqrt{\Delta_n \log d}\right) \ge \mathbb{P}\left(\|\widehat{\Gamma}^S - \Gamma\| \le 0.5\lambda_{\min}(\Gamma)\right) \\ &- O(d^{-1}) \ge 1 - O(d^{-1}).\end{aligned}$$

where the third inequality follows from Lemma A.1 of Fan, Liao, and Mincheva (2011).

(iv) We note that

$$\begin{aligned}&\|\widehat{\beta}^\intercal(\widehat{\Gamma}^S)^{-1}\widehat{\beta} - \beta^\intercal\Gamma^{-1}\beta\| \le 2\|(\widehat{\beta} - \beta)^\intercal\Gamma^{-1}\beta\| \\ &+2\|(\widehat{\beta} - \beta)^\intercal\left((\widehat{\Gamma}^S)^{-1} - \Gamma^{-1}\right)\beta\| \\ &+\|\beta^\intercal\left((\widehat{\Gamma}^S)^{-1} - \Gamma^{-1}\right)\beta\| + \|(\widehat{\beta} - \beta)^\intercal\Gamma^{-1}(\widehat{\beta} - \beta)\| \\ &+\|(\widehat{\beta} - \beta)^\intercal\left((\widehat{\Gamma}^S)^{-1} - \Gamma^{-1}\right)(\widehat{\beta} - \beta)\|\end{aligned}$$

This can be verified directly using (A.7), (A.13), $\|\cdot\| \le \|\cdot\|_{\mathrm{F}}$, and $\|\beta\|_{\mathrm{F}} = O(\sqrt{d})$, which can be shown by following the same argument in Fan, Fan, and Lv (2008) and Assumption 6.

(v) By (A.3) and $\|\cdot\| \leq \|\cdot\|_F$, we have with probability at least $1 - O(r^2 d^{-3})$,

$$\|\widehat{\mathrm{E}} - \mathrm{E}\| \leq Cr\sqrt{\Delta_n \log d}.$$

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

$$\mathbb{P}\left(\|\widehat{\mathrm{E}}^{-1} - \mathrm{E}^{-1}\| < C'\|\widehat{\mathrm{E}} - \mathrm{E}\|\right) \geq \mathbb{P}\left(\|\widehat{\mathrm{E}} - \mathrm{E}\| \leq C'r\sqrt{\Delta_n \log d}\right)$$
$$\geq 1 - O(r^2 d^{-3}).$$

Combining this with (A.14), and by the triangle inequality we have

$$\mathbb{P}\left(\|\widehat{\mathrm{E}}^{-1} + \widehat{\beta}^\intercal(\widehat{\Gamma}^S)^{-1}\widehat{\beta} - \beta^\intercal\Gamma^{-1}\beta - \mathrm{E}^{-1}\| < C'r\sqrt{\Delta_n \log d}\right.$$
$$\left.+ C'm_d r\sqrt{\Delta_n d^2 \log d}\right) > 1 - O(d^{-1}).$$

Hence, the desired inequality follows from the fact that $d \times m_d \to \infty$.

(vi) By their Lemma A.1 again, we have

$$\mathbb{P}\left(\lambda_{\min}\left(\widehat{\mathrm{E}}^{-1} + \widehat{\beta}^\intercal(\widehat{\Gamma}^S)^{-1}\widehat{\beta}\right) \geq Cd\right) \geq \mathbb{P}\left(\|\widehat{\mathrm{E}}^{-1} + \widehat{\beta}^\intercal(\widehat{\Gamma}^S)^{-1}\widehat{\beta}\right.$$
$$\left. - \beta^\intercal\Gamma^{-1}\beta - \mathrm{E}^{-1}\| \leq Cd\right) > 1 - O(d^{-1}),$$

where we use the fact that $m_d r\sqrt{\Delta_n \log d} = o(1)$. This readily yields the desired inequality.

(vii) By the triangle inequality and (A.7), we have, with probability at least $1 - O(rd^{-2})$,

$$\|\widehat{\beta}\| \leq \|\widehat{\beta}\|_F \leq \|\widehat{\beta} - \beta\|_F + \|\beta\|_F \leq C\sqrt{\Delta_n r d \log d} + C\sqrt{d}$$
$$\leq C'\sqrt{d}. \tag{A.19}$$

since $\Delta_n r \log d = o(1)$.

On the other hand, with probability at least $1 - O(d^{-1})$, by (A.12), (A.16), and the above inequality, there exists some C_0 sufficiently large, such that

$$\left\|\widehat{\beta}\left(\widehat{\mathrm{E}}^{-1} + \widehat{\beta}^\intercal(\widehat{\Gamma}^S)^{-1}\widehat{\beta}\right)^{-1}\widehat{\beta}^\intercal(\widehat{\Gamma}^S)^{-1}\right\| \leq \|\widehat{\beta}\|^2 \cdot$$
$$\left\|\left(\widehat{\mathrm{E}}^{-1} + \widehat{\beta}^\intercal(\widehat{\Gamma}^S)^{-1}\widehat{\beta}\right)^{-1}\right\| \cdot \|(\widehat{\Gamma}^S)^{-1}\| \leq C_0.$$

(viii) This follows from the similar argument in (vi) and that $\lambda_{\min}(\Gamma) \geq C$. □

Proof of Theorem 2. We follow similar steps in Fan, Fan, and Lv (2008). By the localization argument, we only need to prove the result under a stronger assumption that the entrywise norms of all the processes are bounded uniformly in $[0, t]$. By the Sherman–Morrison–Woodbury formula, we have

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

We now bound each term above with probability no less than $1 - O(d^{-1})$. First of all, by (A.11)

$$L_1 \leq Cm_d r\sqrt{\Delta_n \log d}.$$

To bound L_2, we have, by (A.13) and (A.17),

$$L_2 \leq \|(\widehat{\Gamma}^S)^{-1} - \Gamma^{-1}\| \cdot \left\|\widehat{\beta}\left(\widehat{\mathrm{E}}^{-1} + \widehat{\beta}^\intercal(\widehat{\Gamma}^S)^{-1}\widehat{\beta}\right)^{-1}\widehat{\beta}^\intercal(\widehat{\Gamma}^S)^{-1}\right\|$$
$$\leq Cm_d r\sqrt{\Delta_n \log d}.$$

Similarly, L_3 can be bounded using (A.13) and (A.18).

Next, for L_4, we use (A.7), (A.16), and (A.19), $\|\cdot\| \leq \|\cdot\|_F$, and $\lambda_{\min}(\Gamma) \geq C'$,

$$L_4 \leq \|\Gamma^{-1}\|^2 \cdot \|\widehat{\beta} - \beta\| \cdot \|\widehat{\beta}\| \cdot \left\|\left(\widehat{\mathrm{E}}^{-1} + \widehat{\beta}^\intercal(\widehat{\Gamma}^S)^{-1}\widehat{\beta}\right)^{-1}\right\|$$
$$\leq C\sqrt{\Delta_n r \log d}.$$

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

Finally, we have

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

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

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

It then follows that

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

On the other hand, by Assumption 6, we have

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

where C is some constant. Thus, $\lambda_{\min}(\beta^\intercal\beta) > Cd$. Therefore $\lambda_{\min}(\beta^\intercal\Gamma^{-1}\beta) > C'd$, following from the fact that $\lambda_{\max}(\Gamma) \leq C_0$. It then implies that $\lambda_{\min}(\mathrm{E}^{-1} + \beta^\intercal\Gamma^{-1}\beta) \geq \lambda_{\min}(\beta^\intercal\Gamma^{-1}\beta) > C'd$.

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

Using (A.15) and (A.16) , we have

$$L_6 \leq Cm_d r\sqrt{\Delta_n \log d}.$$

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

$$\|(\widehat{\Sigma}^S)^{-1} - \Sigma^{-1}\| \leq C\left(m_d r\sqrt{\Delta_n \log d} + \sqrt{\Delta_n r \log d}\right),$$

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

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

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

which implies that

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

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

ACKNOWLEDGMENTS

Jianqing Fan's research was supported by NSF grants DMS-1206464 and DMS-1406266. Xiu gratefully acknowledges financial support from the FMC Faculty Scholar Fund at the University of Chicago Booth School of Business. We have benefited from valuable comments by the Associate Editor, two anonymous referees, and our discussant Dobrislav Dobrev. We are grateful to the seminar and conference participants at the University of Chicago, the 2015 Annual Conference of the Midwest Finance Association, the 2015 AMS Sectional Meeting at Michigan State University, and Two Sigma, LLC.

[Received January 2015. Revised April 2015.]

REFERENCES

Aït-Sahalia, Y., Fan, J., and Xiu, D. (2010), "High-Frequency Covariance Estimates With Noisy and Asynchronous Financial Data," *Journal of the American Statistical Association*, 105, 1504–1517. [490]

Aït-Sahalia, Y., Kalnina, I., and Xiu, D. (2014), "The Idiosyncratic Volatility Puzzle: A Reassessment at High Frequency," Discussion Paper, The University of Chicago. [490,495]

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2001), "The Distribution of Exchange Rate Realized Volatility," *Journal of the American Statistical Association*, 96, 42–55. [490]

Bai, J. (2003), "Inferential Theory for Factor models of Large Dimensions," *Econometrica*, 71, 135–171. [490]

Bai, J., and Ng, S. (2002), "Determine the Number of Factors in Approximate Factor Models," *Econometrica*, 70, 191–221. [490]

Bai, J., and Shi, S. (2011), "Estimating High Dimensional Covariance Matrices and Its Applications," *Annals of Economics and Finance*, 12, 199–215. [490]

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2011), "Multivariate Realised Kernels: Consistent Positive Semi-Definite Estimators of the Covariation of Equity Prices With Noise and Non-Synchronous Trading," *Journal of Econometrics*, 162, 149–169. [490]

Barndorff-Nielsen, O. E., and Shephard, N. (2004), "Econometric Analysis of Realized Covariation: High Frequency Based Covariance, Regression, and Correlation in Financial Economics," *Econometrica*, 72, 885–925. [490]

Bibinger, M., Hautsch, N., Malec, P., and Reiß, M. (2014), "Estimating the Quadratic Covariation Matrix From Noisy Observations: Local Method of Moments and Efficiency," *The Annals of Statistics*, 42, 1312–1346. [490]

Bickel, P. J., and Levina, E. (2008a), "Covariance Regularization by Thresholding," *Annals of Statistics*, 36, 2577–2604. [490]

——— (2008b), "Regularized Estimation of Large Covariance Matrices," *Annals of Statistics*, 36, 199–227. [490,492,493]

Cai, T., and Liu, W. (2011), "Adaptive Thresholding for Sparse Covariance Matrix Estimation," *Journal of American Statistical Association*, 106, 672–684. [490]

Cai, T. T., Liu, W., and Luo, X. (2011), "A Constrained l_1 Minimization Approach to Sparse Precision Matrix Estimation," *Journal of American Statistical Association*, 494, 594–607. [493]

Chamberlain, G., and Rothschild, M. (1983), "Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets," *Econometrica*, 51, 1281–1304. [490,491]

Chen, N. F., Roll, R., and Ross, S. A. (1986), "Economic Forces and the Stock Market," *Journal of Business*, 50, 383–403. [490]

Christensen, K., Kinnebrock, S., and Podolskij, M. (2010), "Pre-Averaging Estimators of the Ex-Post Covariance Matrix in Noisy Diffusion Models With Non-Synchronous Data," *Journal of Econometrics*, 159, 116–133. [490]

Connor, G., and Korajczyk, R. A. (1986), "Performance Measurement with the Arbitrage Pricing Theory: A New Framework for Analysis," *Journal of Financial Economics*, 15, 373–394. [490]

Fama, E. F., and French, K. R. (1993), "Common Risk Factors in the Returns on Stocks and Bonds," *Journal of Financial Economics*, 33, 3–56. [490]

Fan, J., Fan, Y., and Lv, J. (2008), "High Dimensional Covariance Matrix Estimation Using a Factor Model," *Journal Econometrics*, 147, 186–197. [490,492,500,501,502]

Fan, J., Furger, A., and Xiu, D. (2015), "Realized Volatility of Optimal Portfolios," Discussion Paper, Princeton University and the University of Chicago. [494]

Fan, J., and Li, R. (2001), "Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties," *Journal of American Statistical Association*, 96, 1348–1360. [498]

Fan, J., Li, Y., and Yu, K. (2012), "Vast Volatility Matrix Estimation Using High-Frequency Data for Portfolio Selection," *Journal of the American Statistical Association*, 107, 412–428. [490,493,495,498]

Fan, J., Liao, Y., and Mincheva, M. (2011), "High Dimensional Covariance Matrix Estimation in Approximate Factor Models," *The Annals of Statistics*, 39, 3320–3356. [490,492,493,498,500,501,502]

——— (2013), "Large Covariance Estimation by Thresholding Principal Orthogonal Complements," *Journal of the Royal Statistical Society,* Series B, 75, 603–680. [490,498]

Fan, J., and Liu, H. (2013), "Statistical Analysis of Big Data on Pharmacogenomics," *Advanced Drug Delivery Reviews*, 65, 987–1000. [490]

Fan, J., Zhang, J., and Yu, K. (2012), "Vast Portfolio Selection With Gross-Exposure Constraints," *Journal of the American Statistical Association*, 107, 592–606. [492,494,498]

Jagannathan, R., and Ma, T. (2003), "Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps," *Journal of Finance*, 58, 1651–1684. [492,497]

King, B. F. (1966), "Market and Industry Factors in Stock Price Behavior," *Journal of Business*, 39, 139–190. [490]

Ledoit, O., and Wolf, M. (2004a), "Honey, I Shrunk the Sample Covariance Matrix," *Journal of Portfolio Management*, 30, 110–119. [490]

——— (2004b), "A Well-Conditioned Estimator for Large-Dimensional Covariance Matrices," *Journal of Multivariate Analysis*, 88, 365–411. [490]

——— (2012), "Nonlinear Shrinkage Estimation of Large-Dimensional Covariance Matrices," *The Annals of Statistics*, 40, 1024–1060. [490]

Lintner, J. (1965), "Security Prices, Risk, and Maximal Gains From Diversification," *Journal of Finance*, 20, 587–615. [490]

Lunde, A., Shephard, N., and Sheppard, K. (2014), "Econometric Analysis of Vast Covariance Matrices Using Composite Realized Kernels," Discussion Paper, Aarhus University. [491,494]

Markowitz, H. (1952), "Portfolio Selection," *Journal of Finance*, 7, 77–91. [489]

Mykland, P. A., and Zhang, L. (2006), "ANOVA for Diffusions and Itô Processes," *Annals of Statistics*, 34, 1931–1963. [491]

Ross, S. A. (1976), "The Arbitrage Theory of Capital Asset Pricing," *Journal of Economic Theory*, 13, 341–360. [490]

Sharpe, W. F. (1963), "A Simplified Model for Portfolio Analysis," *Management Science*, 9, 277–293. [490]

Shephard, N., and Xiu, D. (2012), "Econometric Analysis of Multivariate Realized QML: Estimation of the Covariation of Equity Prices Under Asynchronous Trading," Discussion Paper, University of Oxford and University of Chicago. [490]

Stock, J. H., and Watson, M. W. (2002), "Forecasting using Principal Components From a Large Number of Predictors," *Journal of American Statistical Association*, 97, 1167–1179. [490]

Tao, M., Wang, Y., and Chen, X. (2013), "Fast Convergence Rates in Estimating Large Volatility Matrices Using High-Frequency Financial Data," *Econometric Theory*, 29, 838–856. [490]

Tao, M., Wang, Y., and Zhou, H. H. (2013), "Optimal Sparse Volatility Matrix Estimation for High Dimensional Itō Processes With Measurement Errors," *Annals of Statistics*, 41, 1816–1864. [490,495,498]

Wang, Y., and Zou, J. (2010), "Vast Volatility Matrix Estimation for High-Frequency Financial Data," *Annals of Statistics*, 38, 943–978. [490]

Yuan, M. (2010), "Sparse Inverse Covariance Matrix Estimation via Linear Programming," *Journal of Machine Learning Research*, 11, 2261–2286. [493]

Zhang, L. (2011), "Estimating Covariation: Epps Effect, Microstructure Noise," *Journal of Econometrics*, 160, 33–47. [490]

Zhou, H. H., Cai, T. T., and Ren, Z. (2014), "Estimating Structured High-Dimensional Covariance and Precision Matrices: Optimal Rates and Adaptive Estimation," Discussion Paper, Yale University. [490]