*Article*

# De-Biased Graphical Lasso for High-Frequency Data

**Yuta Koike**[ID]

Mathematics and Informatics Center and Graduate School of Mathematical Sciences, The University of Tokyo, 3-8-1 Komaba, Meguro-ku, Tokyo 153-8914, Japan; kyuta@ms.u-tokyo.ac.jp

check for updates

**Abstract:** This paper develops a new statistical inference theory for the precision matrix of high-frequency data in a high-dimensional setting. The focus is not only on point estimation but also on interval estimation and hypothesis testing for entries of the precision matrix. To accomplish this purpose, we establish an abstract asymptotic theory for the weighted graphical Lasso and its de-biased version without specifying the form of the initial covariance estimator. We also extend the scope of the theory to the case that a known factor structure is present in the data. The developed theory is applied to the concrete situation where we can use the realized covariance matrix as the initial covariance estimator, and we obtain a feasible asymptotic distribution theory to construct (simultaneous) confidence intervals and (multiple) testing procedures for entries of the precision matrix.

**Keywords:** asymptotic mixed normality; factor model; high-dimensions; Malliavin calculus; precision matrix; sparsity

## 1. Introduction

In high-frequency financial econometrics, covariance matrix estimation of asset returns has been extensively studied in the past two decades. High-frequency financial data are commonly modeled as a discretely observed semimartingale for which the quadratic covariation matrix plays the role of the covariance matrix, so their treatments are often different from those in a standard i.i.d. setting. In recent years, motivated by application to portfolio allocation and risk management in a large-scale asset universe, the high-dimensionality problem has attracted much attention in this area. Since the 2000s, great progress has been made in high-dimensional covariance estimation from i.i.d. data, so researchers are naturally led to apply the techniques developed therein to the context of high-frequency data. For example, Wang and Zou [1] have applied the entry-wise shrinkage methods considered in [2,3] to estimating the covariance matrix of high-frequency data which are asynchronously observed with noise. See also [4–7] for further developments in this approach. In the meantime, it is well-recognized that the *factor structure* is an important ingredient both theoretically and empirically for financial data. In the context of high-dimensional covariance estimation from high-frequency data, this perspective was first taken into account by Fan et al. [8] and subsequently built up by, among others, [9–11]. Other common methods used in i.i.d. settings have also been investigated in the literature of high-frequency financial econometrics. Hautsch et al. [12] and Morimoto and Nagata [13] formally apply eigenvalue regularization methods based on random matrix theory to high-frequency data. Lam et al. [14] accommodate the non-linear shrinkage estimator of [15] to a high-frequency data setting with the help of the spectral distribution theory for the realized covariance matrix developed in [16]. Brownlees et al. [17] employ the $\ell_1$-penalized Gaussian MLE, which is known as the *graphical Lasso*, to estimate the precision matrix (the inverse of the covariance matrix) of high-frequency data. The last approach is closely related to the methodology we will focus on. Despite the recent advances in this topic as above, most studies in this area focus only on *point estimation* of covariance and precision matrices, and there are little work about *interval estimation* and *hypothesis testing* for these objects. A few

exceptions are [18–20]. The first two articles are concerned with continuous-time factor models: Kong and Liu [18] propose a test for the constancy of the factor loading matrix, while Pelger [19] assumes constant loadings and develops an asymptotic distribution theory to make inference for the factors and loadings. Meanwhile, Koike [20] establishes a high-dimensional central limit theorem for the realized covariance matrix which allows us to construct simultaneous confidence regions or carry out multiple testing for entries of the high-dimensional covariance matrix of high-frequency data. The aim of this study is to develop such a statistical inference theory for the *precision matrix* of high-frequency data. This is naturally motivated by the fact that the precision matrix of asset returns plays an important role in mean-variance analysis of portfolio allocation (see e.g., [21], Chapter 5). We accomplish this purpose by imposing a sparsity assumption on the precision matrix. Such an assumption has a clear interpretation in connection with *Gaussian graphical models*: For a Gaussian random vector $\xi = (\xi_1, \dots, \xi_d)^\top$ with covariance matrix $\Sigma$, $\xi_i$ and $\xi_j$ are conditionally independent given the other components if and only if the $(i, j)$-th entry of $\Sigma^{-1}$ is equal to 0, so the sparsity of $\Sigma^{-1}$ is interpreted as the sparsity of the edge structure of the *conditional independence graph* associated with $\xi$. We refer to Chapter 13 of [22] and references therein for more details on graphical models. This standpoint also makes it interesting to estimate the precision matrix of financial data in view of the recent attention to financial network analysis such as [23].

Statistical inference for high-dimensional sparse precision matrices has been actively studied in the recent literature, and various methodologies have ever been proposed; see [24] for an overview. Among others, this paper studies (a weighted version of) the de-biased (or de-sparsified) graphical Lasso in the context of high-frequency data. The de-biased graphical Lasso was introduced in Janková and van de Geer [25] and its theoretical property was investigated in the i.i.d. case. In this paper, we consider its weighted version discussed in [24] because of its theoretically preferable behavior due to its adaptive nature (see Remarks 1 and 2). Compared to the i.i.d. case, we need to handle a new theoretical difficulty in the application to high-frequency data, which is caused by the non-ergodic nature of the problem, i.e., the precision matrix of high-frequency data is generally stochastic and not (stochastically) independent of the observation data. In our context, the precision matrix appears in the coefficients of the linear approximation of the de-biased estimator (see Lemma 1), so it spoils the martingale structure of the linear approximation which we usually have in the i.i.d. case. In a low-dimensional setting, this issue is typically resolved by the concept of *stable convergence* (see e.g., [26]), but the applicability of this approach is questionable in our setting due to the high-dimensionality (see pages 1451–1452 of [20] for a discussion). Instead, we rely on the recent high-dimensional central limit theory of [20] to establish the asymptotic distribution theory for the de-biased estimator, where we settle the above difficulty with the help of Malliavin calculus.

The graphical Lasso is an example of penalized estimation methods. We shall mention that penalized estimation has recently become an active research topic in the setting of asymptotic statistics for stochastic processes. For example, penalized quasi-likelihood estimation for stochastic processes has been developed in the fixed-dimensional setting by [27–30], while estimation for linearly parameterized high-dimensional diffusion models has been studied in [31,32]. Compared to these articles, this paper is novel in the respect that we develop an *asymptotic distribution theory* in a *high-dimensional setting*.

The rest of this paper is organized as follows. In Section 2 we develop an abstract asymptotic theory for the weighted graphical Lasso based on a generic estimator for the quadratic covariation matrix of a high-dimensional semimartingale. This allows us to flexibly apply the developed theory to various settings arising in high-frequency financial econometrics. In Section 3 we extend the scope of the theory to a situation where a known factor structure is present in data and a sparsity assumption is imposed on the precision matrix of the residual process rather than that of the original process. In Section 4, we apply the abstract theory developed in Section 3 to a concrete setting where we observe the process at equidistant times without jumps and noise. Section 5 conducts a Monte Carlo study to assess the finite sample performance of the asymptotic theory, while Section 6 performs a simple real data analysis for illustration. All the technical proofs are collected in the Appendices A–D.

**Notation 1.** *Throughout the paper, we assume $d \geq 2$. $\top$ stands for the transpose of a matrix. For a vector $x \in \mathbb{R}^d$, we write the i-th component of x as $x^i$ for $i = 1, \ldots, d$. For two vectors $x, y \in \mathbb{R}^d$, the statement $x \leq y$ means $x^i \leq y^i$ for all $i = 1, \ldots, d$. The identity matrix of size d is denoted by $\mathsf{E}_d$. We write $\mathbb{R}^{l \times k}$ for the set of all $l \times k$ matrices. $\mathcal{S}_d$ denotes the set of all $d \times d$ symmetric matrices. $\mathcal{S}_d^+$ denotes the set of all $d \times d$ positive semidefinite matrices. $\mathcal{S}_d^{++}$ denotes the set of all $d \times d$ positive definite matrices. For a $l \times k$ matrix A, the $(i, j)$-th entry of A is denoted by $A^{ij}$. Also, $A^{i\cdot}$ and $A^{\cdot j}$ denote the i-th row vector and the j-th column vector, respectively (both are regarded as column vectors). We write $\mathrm{vec}(A)$ for the vectorization of A:*

$$\mathrm{vec}(A) := (A^{11}, \ldots, A^{l1}, A^{12}, \ldots, A^{l2}, \ldots, A^{1k}, \ldots, A^{lk})^\top \in \mathbb{R}^{lk}.$$

*For every $w \in [1, \infty]$, we set*

$$\|A\|_{\ell_w} := \begin{cases} \left\{ \sum_{i=1}^{l} \sum_{j=1}^{k} |A^{ij}|^w \right\}^{1/w} & \text{if } w < \infty, \\ \max_{1 \leq i \leq l} \max_{1 \leq j \leq k} |A^{ij}| & \text{if } w = \infty. \end{cases}$$

*Also, we write $\|\|A\|\|_w$ for the $\ell_w$-operator norm of A:*

$$\|\|A\|\|_w := \sup\{\|Ax\|_{\ell_w} : x \in \mathbb{R}^k, \|x\|_{\ell_w} = 1\}.$$

*It is well-known that $\|\|A\|\|_1 = \max_{1 \leq j \leq k} \sum_{i=1}^{l} |A^{ij}|$ and $\|\|A\|\|_\infty = \max_{1 \leq i \leq l} \sum_{j=1}^{k} |A^{ij}|$. When $l = k$, $\mathrm{diag}(A)$ denotes the diagonal matrix with the same diagonal entries as A, and we set $A^- := A - \mathrm{diag}(A)$. If A is symmetric, we denote by $\Lambda_{\max}(A)$ and $\Lambda_{\min}(A)$ the maximum and minimum eigenvalues of A, respectively. For two matrices A and B, $A \otimes B$ denotes their Kronecker product. When A and B has the same size, we write $A \circ B$ for their Hadamard product.*

*For a random variable $\xi$ and $p \in (0, \infty]$, $\|\xi\|_p$ denotes the $L^p$-norm of $\xi$. For a l-dimensional semimartingale $X = (X_t)_{t \in [0,1]}$ and a k-dimensional semimartingale $Y = (Y_t)_{t \in [0,1]}$, we define $\Sigma_{XY} := [X, Y]_1 := ([X^i, Y^j]_1)_{1 \leq i \leq l, 1 \leq j \leq k}$. We write $\Sigma_X = \Sigma_{XX}$ for short. If $\Sigma_X$ is a.s. invertible, we write $\Theta_X := \Sigma_X^{-1}$.*

## 2. Estimators and Abstract Results

Given a stochastic basis $\mathcal{B} = (\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in [0,1]}, P)$, we consider a $d$-dimensional semimartingale $Y = (Y_t)_{t \in [0,1]}$ defined there. We assume $\Sigma_Y = [Y, Y]_1$ is a.s. invertible. In this paper, we consider the asymptotic theory such that the dimension $d$ possibly depends on a parameter $n \in \mathbb{N}$ so that $d = d_n \to \infty$ as $n \to \infty$. Consequently, both $\mathcal{B}$ and $Y$ may also depend on $n$. However, following the custom of the literature, we omit the indices $n$ from these objects and many other ones appearing below.

Our aim is to estimate the precision matrix $\Theta_Y = \Sigma_Y^{-1}$ when we have an estimator $\hat{\Sigma}_n$ for $\Sigma_Y$; as a corollary, we can also estimate $\Sigma_Y$ itself. We assume that $\hat{\Sigma}_n$ is an $\mathcal{S}_d^+$-valued random variable all of whose diagonal entries are a.s. positive, but we do not specify the form of $\hat{\Sigma}_n$ because the asymptotic theory developed in this section depends on the property of $\hat{\Sigma}_n$ rather than their construction. This is convenient because construction of the estimator depends heavily on observation schemes for $Y$ (with or without noise, synchronous or not, continuous or discontinuous and so on; see [33] for details). In Section 4 we illustrate how we apply the abstract theory developed in this and the next sections to a concrete situation.

We use the *weighted graphical Lasso* to estimate $\Theta_Y$ (cf. [24]). The weighted graphical Lasso estimator $\hat{\Theta}_\lambda$ with penalty parameter $\lambda > 0$ based on $\hat{\Sigma}_n$ is defined by

$$\hat{\Theta}_\lambda := \arg \min_{\Theta \in \mathcal{S}_d^{++}} \left\{ \mathrm{tr}\left(\Theta \hat{\Sigma}_n\right) - \log \det\left(\Theta\right) + \lambda \sum_{i \neq j} \hat{V}_n^{ii} \hat{V}_n^{jj} \left| \Theta^{ij} \right| \right\}, \tag{1}$$

where $\hat{V}_n := \mathrm{diag}(\hat{\Sigma}_n)^{\frac{1}{2}}$. According to the proof of [34] (Lemma 1), the optimization problem in Equation (1) has the unique solution when $\lambda > 0$ and $\hat{\Sigma}_n$ is positive semidefinite and all the diagonal

entries of $\hat{\Sigma}_n$ are positive, so $\hat{\Theta}_\lambda$ is a.s. defined in our setting. In the following we allow $\lambda$ to be a random variable because we typically select $\lambda$ in a data-driven way.

To analyze the theoretical property of $\hat{\Theta}_\lambda$, it is convenient to consider the graphical Lasso estimator $\hat{K}_\lambda$ based on the correlation matrix estimator $\hat{R}_n := \hat{V}_n^{-1}\hat{\Sigma}_n\hat{V}_n^{-1}$ as follows:

$$\hat{K}_\lambda := \arg\min_{K \in \mathcal{S}_d^{++}} \left\{ \operatorname{tr}\left(K\hat{R}_n\right) - \log\det\left(K\right) + \lambda \left\|K^-\right\|_{\ell_1} \right\}. \tag{2}$$

We can easily check $\hat{\Theta}_\lambda = \hat{V}_n^{-1}\hat{K}_\lambda\hat{V}_n^{-1}$.

**Remark 1.** *As pointed out in Rothman et al. [35] and Janková and van de Geer [24], the graphical Lasso based on correlation matrices is theoretically preferable to that based on covariance matrices (so the weighted graphical Lasso is also preferable). In particular, we do not need to impose the so-called irrepresentability condition on $\Sigma_Y$ to derive the theoretical properties of our estimators, which contrasts with Brownlees et al. [17] (see Assumption 2 in [17]). See also Remark 2 for an additional discussion.*

We introduce some notation related to the sparsity assumptions we will impose on $\Theta_Y$. Let $A \in \mathcal{S}_d$. For $j = 1,\ldots,d$, we set $\mathfrak{D}_j(A) := \{i : A^{ij} \neq 0, i \neq j\}$ and $\mathfrak{d}_j(A) := \#\mathfrak{D}_j(A)$. Then we define $\mathfrak{d}(A) := \max_{1 \leq j \leq d} \mathfrak{d}_j(A)$. We also define $S(A) := \bigcup_{j=1}^d \mathfrak{D}_j(A) = \{(i,j) : A^{ij} \neq 0, i \neq j\}$ and $s(A) := \#S(A)$. These quantities have a clear interpretation when the matrix $A$ represents the edge structure of some graph so that $A^{ij} \neq 0$ is equivalent to the presence of an edge between vertices $i$ and $j$ for $i \neq j$; in this case, $\mathfrak{d}_j(A)$ is the number of edges adjacent to vertex $j$ (which is called the *degree* of vertex $j$) and $s(A)$ is the total number of edges contained in the graph.

To derive our asymptotic results, we will impose the following structural assumptions on $\Sigma_Y$.

[A1] $\Lambda_{\max}(\Sigma_Y) + 1/\Lambda_{\min}(\Sigma_Y) = O_p(1)$ as $n \to \infty$.
[A2] $s(\Theta_Y) = O_p(s_n)$ as $n \to \infty$ for some sequence $s_n \in [1,\infty)$, $n = 1,2,\ldots$.
[A3] $\mathfrak{d}(\Theta_Y) = O_p(\mathfrak{d}_n)$ as $n \to \infty$ for some sequence $\mathfrak{d}_n \in [1,\infty)$, $n = 1,2,\ldots$.

[A1] is standard in the literature; see e.g., Condition A1 in [24]. [A2] states that the sparsity of $\Theta_Y$ is controlled by the deterministic sequence $s_n$; we will require the growth rate of $s_n$ to be moderate. [A3] is another sparsity assumption on $\Theta_Y$. It is weaker than [A2] in the sense that it always holds true with $\mathfrak{d}_n = s_n$ under [A2]. However, we can generally take $\mathfrak{d}_n$ smaller than $s_n$.

*2.1. Consistency*

Set $V_Y := \operatorname{diag}(\Sigma_Y)^{\frac{1}{2}}$, $R_Y := V_Y^{-1}\Sigma_Y V_Y^{-1}$ and $K_Y := R_Y^{-1}$.

**Proposition 1.** *Assume [A1]–[A2]. Let $(\lambda_n)_{n=1}^\infty$ be a sequence of positive-valued random variables satisfying the following conditions:*

[B1] $\lambda_n^{-1}\|\hat{\Sigma}_n - \Sigma_Y\|_{\ell_\infty} \to^p 0$ *as* $n \to \infty$.
[B2] $s_n\lambda_n \to^p 0$ *as* $n \to \infty$.

*Then we have*

$$\lambda_n^{-1}\|\hat{K}_{\lambda_n} - K_Y\|_{\ell_2} = O_p(\sqrt{s_n}), \qquad \lambda_n^{-1}\|\|\hat{K}_{\lambda_n} - K_Y\|\|_w = O_p(s_n) \tag{3}$$

*and*

$$\lambda_n^{-1}\|\|\hat{\Theta}_{\lambda_n} - \Theta_Y\|\|_w = O_p(s_n), \qquad \lambda_n^{-1}\|\|\hat{\Theta}_{\lambda_n}^{-1} - \Sigma_Y\|\|_2 = O_p(s_n) \tag{4}$$

*as* $n \to \infty$ *for any* $w \in [1,\infty]$.

Proposition 1 is essentially a rephrasing of Theorem 14.1.3 in [24]. To get a better convergence rate in Proposition 1, we should choose $\lambda_n$ as small as possible, where a lower bound of $\lambda_n$ is determined

by the convergence rate of $\hat{\Sigma}_n$ in the $\ell_\infty$-norm by [B1]. One typically derives this convergence rate by establishing entry-wise concentration inequalities for $\hat{\Sigma}_n$. Such inequalities have already been established for various covariance estimators used in high-frequency financial econometrics; see Theorems 1–2 and Lemma 3 in [36], Theorem 1 in [4], Theorem 1 in [37], and Theorem 2 in [17] for example. We however note that $\hat{\Sigma}_n$ should be positive semidefinite to ensure that the graphical Lasso has the unique solution. This property is not necessarily ensured by many covariance estimators used in this area. In this regard, we mention that pre-averaging and realized kernel estimators have versions to ensure this property, for which relevant bounds are available in [6] (Theorem 2) and [11] (Lemma 1).

**Remark 2** (Comparison to Brownlees et al. [17]). *Compared with [17] (Theorem 1), Proposition 1 has two major theoretical improvements. First, Proposition 1 does not assume the so-called irrepresentability condition, which is imposed in [17] (Theorem 1) as Assumption 2. In fact, under the assumptions of Proposition 1, the unweighted graphical Lasso estimator adopted in [17] would have the convergence rate $(s_n + d)\lambda_n$ (rather than $s_n\lambda_n$ in our case) to estimate $\Theta_Y$ in the norm $\|\|\cdot\|\|_w$, in view of [24] (Theorem 14.1.2). This means that we need to select $\lambda_n$ so that $d\lambda_n \to^p 0$ as $n \to \infty$ to ensure the consistency, which is much stronger than the corresponding assumption [B2] in our setting. Since $\lambda_n$ typically converges to 0 no faster than $1/\sqrt{n}$ with n being the sample size (cf. Section 4), the condition $d\lambda_n \to^p 0$ excludes high-dimensional settings such that $d \gg n$.*

*Second, Proposition 1 gives consistency in the $\ell_w$-operator norm for all $w \in [1, \infty]$, while [17] (Theorem 1) only shows consistency in the $\ell_\infty$-norm. We shall remark that consistency in matrix operator norms is important in application. For example, the consistency of $\hat{\Theta}_{\lambda_n}$ in the $\ell_2$-operator norm implies that eigenvalues of $\hat{\Theta}_{\lambda_n}$ consistently estimate the corresponding eigenvalues of $\Theta_Y$. Also, the consistency in the $\ell_\infty$-operator norm ensures $\|\hat{\Theta}_{\lambda_n}x - \Theta_Y x\|_{\ell_\infty} \to^p 0$ as $n \to \infty$ for any $x \in \mathbb{R}^d$ such that $\|x\|_{\ell_\infty} = O(1)$. This result is important for portfolio allocation because the weight vector for the global minimum variance portfolio is given by $\Theta_Y\mathbf{1}/\mathbf{1}^\top\Theta_Y\mathbf{1}$ when assets have covariance matrix $\Sigma_Y$, where $\mathbf{1} = (1,\ldots,1)^\top \in \mathbb{R}^d$; see e.g., [21] (Section 5.2).*

*On the other hand, unlike [17] (Theorem 1), we do not show selection consistency (i.e., $P(S(\hat{\Theta}_{\lambda_n}) = S(\Theta_Y)) \to 1$ as $n \to \infty$) under our assumptions. Indeed, in the linear regression setting, it is known that an irrepresentability type condition is necessary for the selection consistency of the Lasso; see [22] (Section 7.5.3) for more details. This suggests that our estimator would not have oracle property in the sense of [38] in general. However, we shall remark that the asymptotic mixed normality of the de-biased estimator stated below can be used to construct an estimator with selection consistency via thresholding as in e.g., [39] (Section 3.1) and [40] (Section 4.2). See Corollary 2 and the subsequent discussion for details.*

*2.2. Asymptotic Mixed Normality*

The following lemma states that $\hat{\Theta}_{\lambda_n} - \Theta_Y$ is asymptotically linear in $\hat{\Sigma}_n - \Sigma_Y$ after bias correction when $\Theta_Y$ is sufficiently sparse.

**Lemma 1.** *Suppose that the assumptions of Proposition 1 and [A3] are satisfied. Then we have*

$$\lambda_n^{-2}\|\hat{\Theta}_{\lambda_n} - \Theta_Y - \Gamma_n + \Theta_Y(\hat{\Sigma}_n - \Sigma_Y)\Theta_Y\|_{\ell_\infty} = O_p(s_n\sqrt{\mathfrak{d}_n})$$

*as $n \to \infty$, where $\Gamma_n := -(\hat{\Theta}_{\lambda_n} - \hat{\Theta}_{\lambda_n}\hat{\Sigma}_n\hat{\Theta}_{\lambda_n})$.*

Lemma 1 is an almost straightforward consequence of Equation (4) and the Karush–Kuhn–Tucker (KKT) conditions for the optimization problem in Equation (1). As a consequence of this lemma, we obtain the following result, which states that the "de-biased" weighted graphical Lasso estimator $\hat{\Theta}_{\lambda_n} - \Theta_Y - \Gamma_n$ inherits the asymptotic mixed normality of $\hat{\Sigma}_n$.

**Proposition 2.** *Suppose that the assumptions of Lemma 1 are satisfied. For every $n \in \mathbb{N}$, let $a_n > 0$, $\mathfrak{C}_n$ be a $d^2 \times d^2$ positive semidefinite random matrix and $J_n$ be an $m \times d^2$ random matrix, where $m = m_n$ may depend on $n$. Assume $a_n \|\|\|J_n\|\|\|_\infty \lambda_n^2 s_n \sqrt{\mathfrak{d}_n \log(m+1)} \to^p 0$ as $n \to \infty$. Assume also that*

$$\lim_{n \to \infty} \sup_{y \in \mathbb{R}^m} \left| P\left( a_n \widetilde{J}_n \operatorname{vec}\left(\hat{\Sigma}_n - \Sigma_Y\right) \le y \right) - P\left( \widetilde{J}_n \mathfrak{C}_n^{1/2} \zeta_n \le y \right) \right| = 0 \tag{5}$$

*and*

$$\lim_{b \downarrow 0} \limsup_{n \to \infty} P(\min \operatorname{diag}(\widetilde{J}_n \mathfrak{C}_n \widetilde{J}_n^\top) < b) = 0 \tag{6}$$

*as $n \to \infty$, where $\widetilde{J}_n := -J_n(\Theta_Y \otimes \Theta_Y)$ and $\zeta_n$ is a $d^2$-dimensional standard Gaussian vector independent of $\mathcal{F}$, which is defined on an extension of the probability space $(\Omega, \mathcal{F}, P)$ if necessary. Then,*

$$\lim_{n \to \infty} \sup_{y \in \mathbb{R}^m} \left| P\left( a_n J_n \operatorname{vec}\left(\hat{\Theta}_{\lambda_n} - \Gamma_n - \Theta_Y\right) \le y \right) - P\left( \widetilde{J}_n \mathfrak{C}_n^{1/2} \zeta_n \le y \right) \right| = 0.$$

In a standard i.i.d. setting such that $\Theta_Y$ is non-random, we can usually verify Equation (5) by classical Lindeberg's central limit theorem when $m = 1$ and $J_n$ is non-random because $a_n \widetilde{J}_n \operatorname{vec}(\hat{\Sigma}_n - \Sigma_Y)$ can be written as a sum of independent random variables; see the proof of [25] (Theorem 1) for example. By contrast, $\Theta_Y$ is generally random and not independent of $\hat{\Sigma}_n - \Sigma_Y$ in our setting, so $a_n \widetilde{J}_n \operatorname{vec}(\hat{\Sigma}_n - \Sigma_Y)$ may not be a martingale even if $\operatorname{vec}(\hat{\Sigma}_n - \Sigma_Y)$ is a martingale. In the case that $d$ is fixed, we typically resolve this issue by proving *stable* convergence in law of $\operatorname{vec}(\hat{\Sigma}_n - \Sigma_Y)$; see e.g., [26] for details. However, extension of this approach to the case that $d \to \infty$ as $n \to \infty$ is far from trivial as discussed at the beginning of [20] (Section 3). For this reason, [20] gives a result to directly establish Equation (5) type convergence in a high-dimensional setting. This result will be used in Section 4 to apply our abstract theory to a more concrete setting.

**Remark 3.** *Proposition 2 also allows $m$ to diverge as $n \to \infty$, which is necessary when we need to derive an asymptotic approximation of the joint distribution of $\operatorname{vec}(\hat{\Theta}_{\lambda_n} - \Gamma_n - \Theta_Y)$. Such an approximation can be used to make simultaneous inference for entries of $\Theta_Y$; see [40] for example.*

## 3. Factor Structure

In financial applications, it is often important to take account of the factor structure of asset prices. In fact, many empirical studies have documented the existence of common factors in financial markets (e.g., [41] (Section 6.5)). Also, factor models play a dominant role in asset pricing theory (cf. [21] (Chapter 9)). When common factors are present across asset returns, the precision matrix cannot be sparse because all pairs of the assets are partially correlated given other assets through the common factors. Therefore, in such a situation, it is common practice to impose a sparsity assumption on the precision matrix of the residual process which is obtained after removing the co-movements induced by the factors (see e.g., [17] (Section 4.2) and [42] (Section 4.2)). In this section, we accommodate the theory developed in Section 2 to such an application.

Specifically, suppose that we have an $r$-dimensional known factor process $X$, and consider the following continuous-time factor model:

$$Y = \beta X + Z. \tag{7}$$

Here, $\beta$ is a non-random $d \times r$ matrix and $Z$ is a $d$-dimensional semimartingale such that $[Z, X]_1 = 0$. $\beta$ and $Z$ represent the factor loading matrix and residual process of the model, respectively. This model is widely used in high-frequency financial econometrics; see [8,9,11] in the context of high-dimensional covariance matrix estimation. One restriction of the model Equation (7) is that the factor loading $\beta$ is assumed to be constant, but there is empirical evidence that $\beta$ may be regarded as constant in short time intervals (one week or less); see [18,43] for instance.

**Remark 4.** *The number of factors r possibly depends on n and (slowly) diverges as $n \to \infty$. Also, $\beta$ may depend on n.*

We are interested in estimating $\Sigma_Y$ based on observation data for $X$ and $Y$ while taking account of the factor structure given by Equation (7). Suppose that we have generic estimators $\hat{\Sigma}_{Y,n}, \hat{\Sigma}_{X,n}$ and $\hat{\Sigma}_{YX,n}$ for $\Sigma_Y, \Sigma_X$ and $\Sigma_{YX}$, respectively. $\hat{\Sigma}_{Y,n}, \hat{\Sigma}_{X,n}$ and $\hat{\Sigma}_{YX,n}$ are assumed to be random variables taking values in $\mathcal{S}_d, \mathcal{S}_r^+$ and $\mathbb{R}^{d \times r}$, respectively. Now, by assumption we have

$$\Sigma_Y = \beta \Sigma_X \beta^\top + \Sigma_Z. \tag{8}$$

Assume $\Sigma_X$ is a.s. invertible. Then $\beta$ can be written as $\beta = \Sigma_{YX}\Sigma_X^{-1}$. Therefore, we can naturally estimate $\beta$ by $\hat{\beta}_n := \hat{\Sigma}_{YX,n}\hat{\Sigma}_{X,n}^{-1}$, provided that $\hat{\Sigma}_{X,n}$ is invertible. In practical applications, the invertibility of $\hat{\Sigma}_{X,n}$ is usually not problematic because the number of factors $r$ is sufficiently small compared to the sample size. However, it is theoretically convenient to (formally) define $\hat{\beta}_n$ in the case that $\hat{\Sigma}_{X,n}$ is singular. For this reason, we take an $\mathcal{S}_d^{++}$-valued random variable $\hat{\Sigma}_{X,n}^\dagger$ such that $\hat{\Sigma}_{X,n}^\dagger = \hat{\Sigma}_{X,n}^{-1}$ on the event where $\hat{\Sigma}_{X,n}$ is invertible, and redefine $\hat{\beta}_n$ as $\hat{\beta}_n := \hat{\Sigma}_{YX,n}\hat{\Sigma}_{X,n}^\dagger$. This does not affect the asymptotic properties of our estimators because $\hat{\Sigma}_{X,n}$ is asymptotically invertible under our assumptions we will impose. Now, from Equation (8), $\Sigma_Z$ is estimated by

$$\hat{\Sigma}_{Z,n} := \hat{\Sigma}_{Y,n} - \hat{\beta}_n \hat{\Sigma}_{X,n} \hat{\beta}_n^\top. \tag{9}$$

Since $\hat{\Sigma}_{Z,n}$ might be a poor estimator for $\Sigma_Z$ because $d$ can be extremely large in our setting, we apply the weighted graphical Lasso to $\hat{\Sigma}_{Z,n}$ in order to estimate $\Sigma_Z$. Specifically, we construct the weighted graphical Lasso estimator $\hat{\Theta}_{Z,\lambda}$ based on $\hat{\Sigma}_{Z,n}$ as follows:

$$\hat{\Theta}_{Z,\lambda} = \arg \min_{\Theta \in \mathcal{S}_d^{++}} \left\{ \operatorname{tr}\left(\Theta \hat{\Sigma}_{Z,n}\right) - \log \det\left(\Theta\right) + \lambda \sum_{i \neq j} \sqrt{\hat{\Sigma}_{Z,n}^{ii} \hat{\Sigma}_{Z,n}^{jj}} \left|\Theta^{ij}\right| \right\}. \tag{10}$$

Then $\Sigma_Z$ is estimated by the inverse of $\hat{\Theta}_{Z,\lambda}$. Hence our final estimator for $\Sigma_Y$ is constructed as

$$\hat{\Sigma}_{Y,\lambda} := \hat{\beta}_n \hat{\Sigma}_{X,n} \hat{\beta}_n^\top + \hat{\Theta}_{Z,\lambda}^{-1}. \tag{11}$$

**Remark 5.** *Although we will impose the assumptions which guarantee that the optimization problem in Equation (10) asymptotically has the unique solution with probability 1, it may have no solution for a fixed n. Thus, we formally define $\hat{\Theta}_{Z,\lambda}$ as an $\mathcal{S}_d^{++}$-valued random variable such that $\hat{\Theta}_{Z,\lambda}$ is defined by Equation (10) on the event where the optimization problem in Equation (10) has the unique solution.*

**Remark 6** (Positive definiteness of $\hat{\Sigma}_{Y,\lambda}$). *Since $\hat{\Theta}_{Z,\lambda}^{-1}$ is positive definite by construction, $\hat{\Sigma}_{Y,\lambda}$ is positive definite (note that we assume $\hat{\Sigma}_{X,n}$ is positive semidefinite).*

We will impose the following structural assumptions on the model:

[C1] $\|\Sigma_Y\|_{\ell_\infty} = O_p(1)$ and $\|\beta\|_{\ell_\infty} = O(1)$ as $n \to \infty$.
[C2] $\Lambda_{\max}(\Sigma_Z) + 1/\Lambda_{\min}(\Sigma_Z) = O_p(1)$ as $n \to \infty$.
[C3] $\|\Sigma_X\|_{\ell_\infty} + 1/\Lambda_{\min}(\Sigma_X) = O_p(1)$ as $n \to \infty$.
[C4] $s(\Theta_Z) = O_p(s_n)$ as $n \to \infty$ for some sequence $s_n \in [1, \infty)$, $n = 1, 2, \dots$.
[C5] $\mathfrak{d}(\Theta_Z) = O_p(\mathfrak{d}_n)$ as $n \to \infty$ for some sequence $\mathfrak{d}_n \in [1, \infty)$, $n = 1, 2, \dots$.
[C6] There is a positive definite $d \times d$ matrix $\boldsymbol{B}$ such that $\||d^{-1}\beta^\top\beta - \boldsymbol{B}\||_2 \to 0$ and $\Lambda_{\min}(\boldsymbol{B})^{-1} = O(1)$ as $n \to \infty$.

[C1]–[C3] are natural structural assumptions on the model and standard in the literature; see e.g., Assumptions 2.1 and 3.3 in [44]. [C4]–[C5] are sparsity assumptions on the precision matrix of the residual process and necessary for our application of the (weighted) graphical Lasso. [C6] requires the

factors to have non-negligible impact on almost all assets and is also standard in the context of covariance matrix estimation based on a factor model; see e.g., Assumption 3.5 in [44] and Assumption 6 in [8].

The following result establishes the consistency of the residual precision matrix estimator $\hat{\Theta}_{Z,\lambda}$.

**Proposition 3.** *Assume [C1]–[C4]. Let $(\lambda_n)_{n=1}^{\infty}$ be a sequence of positive-valued random variables satisfying the following conditions:*

[D1] $\lambda_n^{-1}\|\hat{\Sigma}_{X,n} - \Sigma_X\|_{\ell_\infty} \to^p 0$, $\lambda_n^{-1}\|\hat{\Sigma}_{YX,n} - \beta\hat{\Sigma}_{X,n}\|_{\ell_\infty} \to^p 0$ *and* $\lambda_n^{-1}\|\check{\Sigma}_{Z,n} - \Sigma_Z\|_{\ell_\infty} \to^p 0$ *as* $n \to \infty$, *where* $\check{\Sigma}_{Z,n} := \hat{\Sigma}_{Y,n} - \hat{\Sigma}_{YX,n}\beta^\top - \beta\hat{\Sigma}_{YX,n}^\top + \beta\hat{\Sigma}_{X,n}\beta^\top$.

[D2] $(s_n + r)\lambda_n \to^p 0$ *as* $n \to \infty$.

[D3] $P(\overline{\Sigma}_n \in \mathcal{S}_d^+) \to 1$ *as* $n \to \infty$, *where*

$$\overline{\Sigma}_n := \begin{pmatrix} \hat{\Sigma}_{X,n} & \hat{\Sigma}_{YX,n}^\top \\ \hat{\Sigma}_{YX,n} & \hat{\Sigma}_{Y,n} \end{pmatrix}.$$

*Then* $\lambda_n^{-1}\||\hat{\Theta}_{Z,\lambda_n} - \Theta_Z\||_w = O_p(s_n)$ *and* $\lambda_n^{-1}\||\hat{\Theta}_{Z,\lambda_n}^{-1} - \Sigma_Z\||_2 = O_p(s_n)$ *as* $n \to \infty$ *for any* $w \in [1,\infty]$.

**Remark 7.** *(a) Since* $\Sigma_{ZX} = \Sigma_{YX} - \beta\Sigma_X$ *and* $\Sigma_Z = \Sigma_Y - \Sigma_{YX}\beta^\top - \beta\Sigma_{XY} + \beta\Sigma_X\beta^\top$, $\hat{\Sigma}_{YX,n}$ *and* $\check{\Sigma}_{Z,n}$ *are seen as natural estimators for* $\Sigma_{ZX}(= 0)$ *and* $\Sigma_Z$ *respectively if* $\beta$ *were known. In this sense, [D1] is a natural extension of [B1]. In particular, if* $r = O(1)$ *as* $n \to \infty$, *[D1] follows from the convergences* $\lambda_n^{-1}\|\hat{\Sigma}_{X,n} - \Sigma_X\|_{\ell_\infty} \to^p 0$, $\lambda_n^{-1}\|\hat{\Sigma}_{YX,n} - \Sigma_{YX}\|_{\ell_\infty} \to^p 0$ *and* $\lambda_n^{-1}\|\hat{\Sigma}_{Y,n} - \Sigma_Y\|_{\ell_\infty} \to^p 0$ *under [C1], which are typically derived from entry-wise concentration inequalities for* $\hat{\Sigma}_{X,n}$, $\hat{\Sigma}_{YX,n}$ *and* $\hat{\Sigma}_{Y,n}$.

*(b) [D3] ensures that* $\hat{\Sigma}_{Z,n}$ *is asymptotically positive semidefinite. This is necessary for guaranteeing that the optimization problem in Equation (10) asymptotically has the unique solution with probability 1.*

From Proposition 3 we can also derive the convergence rates for the estimators $\hat{\Sigma}_{Z,\lambda_n}$ and $\hat{\Sigma}_{Z,\lambda_n}^{-1}$ in appropriate norms, which may be seen as counterparts of Theorems 1–2 in [8].

**Proposition 4.** *Under the assumptions of Proposition 3,* $\lambda_n^{-1}\|\hat{\Sigma}_{Z,\lambda_n} - \Sigma_Z\|_{\ell_\infty} = O_p(s_n + r^2)$ *as* $n \to \infty$

**Proposition 5.** *Under the assumptions of Proposition 3, we additionally assume [C5]–[C6]. Then,* $\lambda_n^{-1}\||\hat{\Sigma}_{Y,\lambda_n}^{-1} - \Sigma_Y^{-1}\||_2 = O_p(s_n + r)$ *and* $\lambda_n^{-1}\||\hat{\Sigma}_{Y,\lambda_n}^{-1} - \Sigma_Y^{-1}\||_\infty = O_p(r^{3/2}\mathfrak{d}_n(s_n + r))$ *as* $n \to \infty$.

Next we present the high-dimensional asymptotic mixed normality of the de-biased version of $\hat{\Theta}_{Z,\lambda}$.

**Proposition 6.** *Suppose that the assumptions of Proposition 3 and [C5] are satisfied. For every* $n \in \mathbb{N}$, *let* $a_n > 0$, $\mathfrak{C}_n$ *be a* $d^2 \times d^2$ *positive semidefinite random matrix and* $J_n$ *be an* $m \times d^2$ *random matrix, where* $m = m_n$ *may depend on* $n$. *Assume* $a_n\||J_n\||_\infty\lambda_n^2 s_n\sqrt{\mathfrak{d}_n\log(m+1)} \to^p 0$ *as* $n \to \infty$. *Assume also that*

$$\lim_{n\to\infty} \sup_{y\in\mathbb{R}^m} \left| P\left(a_n\widetilde{J}_{Z,n}\operatorname{vec}\left(\check{\Sigma}_{Z,n} - \Sigma_Z\right) \le y\right) - P\left(\widetilde{J}_{Z,n}\mathfrak{C}_n^{1/2}\zeta_n \le y\right) \right| = 0 \tag{12}$$

*and*

$$\lim_{b\downarrow 0}\limsup_{n\to\infty} P(\min\operatorname{diag}(\widetilde{J}_{Z,n}\mathfrak{C}_n\widetilde{J}_{Z,n}^\top) < b) = 0 \tag{13}$$

*as* $n \to \infty$, *where* $\widetilde{J}_{Z,n} := -J_n(\Theta_Z \otimes \Theta_Z)$ *and* $\zeta_n$ *is a* $d^2$-*dimensional standard Gaussian vector independent of* $\mathcal{F}$, *which is defined on an extension of the probability space* $(\Omega, \mathcal{F}, P)$ *if necessary. Then,*

$$\lim_{n\to\infty} \sup_{y\in\mathbb{R}^m} \left| P\left(a_n J_n\operatorname{vec}\left(\hat{\Theta}_{Z,\lambda_n} - \Gamma_{Z,n} - \Theta_Z\right) \le y\right) - P\left(\widetilde{J}_{Z,n}\mathfrak{C}_n^{1/2}\zeta_n \le y\right) \right| = 0,$$

*where* $\Gamma_{Z,n} := -(\hat{\Theta}_{Z,\lambda_n} - \hat{\Theta}_{Z,\lambda_n}\hat{\Sigma}_{Z,n}\hat{\Theta}_{Z,\lambda_n})$.

**Remark 8.** *It is worth mentioning that condition Equation (12) is stated for $\check{\Sigma}_{Z,n}$ rather than $\hat{\Sigma}_{Z,n}$. In other words, for deriving the asymptotic distribution, we do not need to take account of the effect of plugging $\hat{\beta}_n$ into $\beta$, at least in the first order. This is thanks to Lemma A11.*

Although it is generally difficult to derive the asymptotic mixed normality of (the de-biased version of) $\hat{\Sigma}_{Y,\lambda_n}^{-1}$, this is possible when $d$ is sufficiently large. In fact, in such a situation, the entry-wise behavior of $\Sigma_Y^{-1}$ is dominated by $\Theta_Z$ as described by the following lemma:

**Lemma 2.** *Under the assumptions of Proposition 5, $\|\hat{\Sigma}_{Y,\lambda_n}^{-1} - \hat{\Theta}_{Z,\lambda_n}\|_{\ell_\infty} = O_p(r\mathfrak{d}_n/d)$ and $\|\Sigma_Y^{-1} - \Theta_Z\|_{\ell_\infty} = O_p(r\mathfrak{d}_n/d)$ as $n \to \infty$.*

Consequently, we obtain the following result.

**Proposition 7.** *Suppose that the assumptions of Proposition 6 and [C6] are satisfied. Suppose also $a_n \||J_n|\|_\infty r\mathfrak{d}_n \sqrt{\log(m+1)}/d \to 0$ as $n \to \infty$. Then we have*

$$\lim_{n \to \infty} \sup_{y \in \mathbb{R}^m} \left| P\left( a_n J_n \operatorname{vec}\left( \hat{\Sigma}_{Y,\lambda_n}^{-1} - \Gamma_{Z,n} - \Sigma_Y^{-1} \right) \leq y \right) - P\left( \tilde{J}_{Z,n} \mathfrak{C}_n^{1/2} \zeta_n \leq y \right) \right| = 0.$$

## 4. Application to Realized Covariance Matrix

In this section, we apply the abstract theory developed above to the simplest situation where the processes have no jumps and are observed at equidistant times without noise. Specifically, we consider the continuous-time factor model Equation (7) and assume that both $Y$ and $X$ are observed at equidistant time points $h/n$, $h = 0, 1, \ldots, n$. In this case, $\Sigma_Y = [Y, Y]_1$ is naturally estimated by the *realized covariance matrix*:

$$\hat{\Sigma}_{Y,n} := \widehat{[Y, Y]}_1^n := \sum_{h=1}^n (Y_{h/n} - Y_{(h-1)/n})(Y_{h/n} - Y_{(h-1)/n})^\top. \tag{14}$$

Analogously, we define $\hat{\Sigma}_{X,n} := \widehat{[X, X]}_1^n$ and $\hat{\Sigma}_{YX,n} := \widehat{[Y, X]}_1^n$. In addition, we assume that $Z$ and $X$ are respectively $d$-dimensional and $r$-dimensional continuous Itô semimartingales given by

$$Z_t = Z_0 + \int_0^t \mu_s ds + \int_0^t \sigma_s dW_s, \qquad X_t = X_0 + \int_0^t \tilde{\mu}_s ds + \int_0^t \tilde{\sigma}_s dW_s,$$

where $\mu = (\mu_s)_{s \in [0,1]}$ and $\tilde{\mu} = (\tilde{\mu}_s)_{s \in [0,1]}$ are respectively $d$-dimensional and $r$-dimensional $(\mathcal{F}_t)$-progressively measurable processes, $\sigma = (\sigma_s)_{s \in [0,1]}$ and $\tilde{\sigma} = (\tilde{\sigma}_s)_{s \in [0,1]}$ are respectively $\mathbb{R}^{d \times d'}$-valued and $\mathbb{R}^{r \times d'}$-valued $(\mathcal{F}_t)$-progressively measurable processes, and $W = (W_s)_{s \in [0,1]}$ is a $d'$-dimensional standard $(\mathcal{F}_t)$-Wiener process. To apply the convergence rate results to this setting, we impose the following assumptions:

[E1] For all $n, \nu \in \mathbb{N}$, we have an event $\Omega_n(\nu) \in \mathcal{F}$ and $(\mathcal{F}_t)$-progressively measurable processes $\mu(\nu) = (\mu(\nu)_s)_{s \in [0,1]}$, $\tilde{\mu}(\nu) = (\tilde{\mu}(\nu)_s)_{s \in [0,1]}$, $\sigma(\nu) = (\sigma(\nu)_s)_{s \in [0,1]}$ and $\tilde{\sigma}(\nu) = (\tilde{\sigma}(\nu)_s)_{s \in [0,1]}$ which take values in $\mathbb{R}^d$, $\mathbb{R}^r$, $\mathbb{R}^{d \times d'}$ and $\mathbb{R}^{r \times d'}$, respectively, and they satisfy the following conditions:

(i) $\lim_{\nu \to \infty} \limsup_{n \to \infty} P(\Omega_n(\nu)^c) = 0$.
(ii) $\mu = \mu(\nu)$, $\tilde{\mu} = \tilde{\mu}(\nu)$, $\sigma = \sigma(\nu)$ and $\tilde{\sigma} = \tilde{\sigma}(\nu)$ on $\Omega_n(\nu)$ for all $\nu \in \mathbb{N}$.
(iii) For all $\nu \in \mathbb{N}$, there is a constant $C_\nu > 0$ such that

$$\sup_{n \in \mathbb{N}} \sup_{0 \leq t \leq 1} \sup_{\omega \in \Omega} \left( \|\mu(\nu)_t(\omega)\|_{\ell_\infty} + \|\tilde{\mu}(\nu)_t(\omega)\|_{\ell_\infty} + \|c(\nu)_t(\omega)\|_{\ell_\infty} + \|\tilde{c}(\nu)_t(\omega)\|_{\ell_\infty} \right) \leq C_\nu,$$

where $c(\nu)_t := \sigma(\nu)_t \sigma(\nu)_t^\top$ and $\tilde{c}(\nu)_t := \tilde{\sigma}(\nu)_t \tilde{\sigma}(\nu)_t^\top$.

[E2]  $r = O(d)$ and $(\log d)/\sqrt{n} \to 0$ as $n \to \infty$.

[E1] is a local boundedness assumption on the coefficient processes and typical in the literature: For example, [E1] is satisfied when $\mu, \widetilde{\mu}, \sigma$ and $\widetilde{\sigma}$ are all bounded by some locally bounded process independent of $n$. This latter condition is imposed in [8], among others. [E2] restricts the growth rates of $d$ and $r$. It is indeed an adaptation of [D1] to the present setting.

**Theorem 1.** *Assume [C1]–[C4] and [E1]–[E2]. Let $\lambda_n$ be a sequence of positive-valued random variables such that $\lambda_n^{-1}\sqrt{(\log d)/n} \to^p 0$ and $(s_n + r)\lambda_n \to^p 0$ as $n \to \infty$. Then $\lambda_n^{-1}\|\|\hat{\Theta}_{Z,\lambda_n} - \Theta_Z\|\|_w = O_p(s_n)$, $\lambda_n^{-1}\|\|\hat{\Theta}_{Z,\lambda_n}^{-1} - \Sigma_Z\|\|_2 = O_p(s_n)$ and $\lambda_n^{-1}\|\hat{\Sigma}_{Y,\lambda_n} - \Sigma_Y\|_{\ell_\infty} = O_p(s_n + r^2)$ as $n \to \infty$ for any $w \in [1,\infty]$. Moreover, if we additionally assume [C5]–[C6], then $\lambda_n^{-1}\|\|\hat{\Sigma}_{Y,\lambda_n}^{-1} - \Sigma_Y^{-1}\|\|_2 = O_p(s_n + r)$ and $\lambda_n^{-1}\|\|\hat{\Sigma}_{Y,\lambda_n}^{-1} - \Sigma_Y^{-1}\|\|_\infty = O_p(r^{3/2}\mathfrak{d}_n(s_n + r))$ as $n \to \infty$.*

**Remark 9** (Optimal convergence rate). *From Theorem 1, the convergence rate of $\hat{\Theta}_{Z,\lambda_n}$ to $\Theta_Z$ in the $\ell_w$-operator norm for any $w \in [1,\infty]$ can be arbitrarily close to $s_n\sqrt{(\log d)/n}$, which is similar to that in a standard i.i.d. setting (cf. Theorem 14.1.3 in [24]). On the other hand, in the Gaussian i.i.d. setting without factor structure, the minimax optimal rate for this problem is known to be $\mathfrak{d}(\Theta_Z)\sqrt{(\log d)/n}$ (see [45] (Theorem 1.1) and [46] (Theorem 5)), which can be faster than $s_n\sqrt{(\log d)/n}$. In a standard i.i.d. setting, this rate can be attained by using a node-wise penalized regression (see e.g., [46] (Section 3.1)), so it would be interesting to study the convergence rate of such a method in our setting. We leave it to future research. In the meantime, such a method does not ensure the positive definiteness of the estimated precision matrix in general, so our estimator would be preferable for some practical applications such as portfolio allocation.*

Next we derive the asymptotic mixed normality of the de-biased estimator in the present setting. As announced, we accomplish this purpose with the help of Malliavin calculus. In the following we will freely use standard concepts and notation from Malliavin calculus. We refer to [47,48] (Chapter 1) for detailed treatments of this subject.

We consider the Malliavin calculus with respect to $W$. For any real number $p \geq 1$ and any integer $k \geq 1$, $\mathbb{D}_{k,p}$ denotes the stochastic Sobolev space of random variables which are $k$ times differentiable in the Malliavin sense and the derivatives up to order $k$ have finite moments of order $p$. If $F \in \mathbb{D}_{k,p}$, we denote by $D^k F$ the $k$th Malliavin derivative of $F$, which is a random variable taking values in $L^2([0,1]^k; (\mathbb{R}^{d'})^{\otimes k})$. Here, we identify the space $(\mathbb{R}^{d'})^{\otimes k}$ with the set of all $d'$-dimensional $k$-way arrays, i.e., real-valued functions on $\{1, \ldots, d'\}^k$. Since $D^k F$ is a random function on $[0,1]^k$, we can consider the value $D^k F(t_1, \ldots, t_k)$ evaluated at $(t_1, \ldots, t_k) \in [0,1]^k$. We denote this value by $D_{t_1,\ldots,t_k} F$. Moreover, since $D_{t_1,\ldots,t_k} F$ takes values in $(\mathbb{R}^{d'})^{\otimes k}$, we can consider the value $D_{t_1,\ldots,t_k} F(a_1, \ldots, a_k)$ evaluated at $(a_1, \ldots, a_k) \in \{1, \ldots, d'\}^k$. This value is denoted by $D_{t_1,\ldots,t_k}^{(a_1,\ldots,a_k)} F$. We remark that the variable $D_{t_1,\ldots,t_k} F$ is defined only a.e. on $[0,1]^k \times \Omega$ with respect to the measure $\mathrm{Leb}_k \times P$, where $\mathrm{Leb}_k$ denotes the Lebesgue measure on $[0,1]^k$. Therefore, if $D_{t_1,\ldots,t_k} F$ satisfies some property a.e. on $[0,1]^k \times \Omega$ with respect to $\mathrm{Leb}_k \times P$, by convention we will always take a version of $D_{t_1,\ldots,t_k} F$ satisfying that property everywhere on $[0,1]^k \times \Omega$ if necessary. We set $\mathbb{D}_{k,\infty} := \bigcap_{p=1}^{\infty} \mathbb{D}_{k,p}$. We denote by $\mathbb{D}_{k,\infty}(\mathbb{R}^d)$ the space of all $d$-dimensional random variables $F$ such that $F^i \in \mathbb{D}_{k,\infty}$ for every $i = 1, \ldots, d$. The space $\mathbb{D}_{k,\infty}(\mathbb{R}^{d \times r})$ is defined in an analogous way. Finally, for any $(\mathbb{R}^{d'})^{\otimes k}$-valued random variable $F$ and $p \in (0,\infty]$, we set

$$\|F\|_{p,\ell_2} := \left\|\sqrt{\sum_{a_1,\ldots,a_k=1}^{d'} F(a_1, \ldots, a_k)^2}\right\|_p.$$

We also need to define some variables related to the "asymptotic" covariance matrices of the estimators. We define $d^2 \times d^2$ random matrix $\mathfrak{C}_n$ by

$$\mathfrak{C}_n^{(i-1)d+j,(k-1)d+l} :=$$

$$n \sum_{h=1}^{n} \left\{ \left( \int_{(h-1)/n}^{h/n} c_s^{ik} ds \right) \left( \int_{(h-1)/n}^{h/n} c_s^{jl} ds \right) + \left( \int_{(h-1)/n}^{h/n} c_s^{il} ds \right) \left( \int_{(h-1)/n}^{h/n} c_s^{jk} ds \right) \right\},$$

$$i, j, k, l = 1, \ldots, d,$$

where $c_s := \sigma_s \sigma_s^\top$. Then we set $\mathfrak{V}_n := (\Theta_Z \otimes \Theta_Z) \mathfrak{C}_n (\Theta_Z \otimes \Theta_Z)$ and $\mathfrak{S}_n := \mathrm{diag}(\mathfrak{V}_n)^{1/2}$. In addition, under [E1], we define $\mathfrak{C}_n(\nu)$ similarly to $\mathfrak{C}_n$ with replacing $\sigma$ by $\sigma(\nu)$. $\mathfrak{C}_n$ and $\mathfrak{V}_n$ play roles of the asymptotic covariance matrices of $\check{\Sigma}_{Z,n}$ and $\hat{\Theta}_{Z,\lambda_n}$, respectively.

We impose the following assumptions on the model.

[F1] We have [E1] and $\Sigma_Z(\nu) := \int_0^1 c(\nu)_t dt$ is a.s. invertible for all $n, \nu \in \mathbb{N}$. Moreover, for all $n, \nu \in \mathbb{N}$ and $t \in [0,1]$, $\mu(\nu)_t \in \mathbb{D}_{1,\infty}(\mathbb{R}^d)$, $\sigma(\nu)_t \in \mathbb{D}_{2,\infty}(\mathbb{R}^{d \times r})$ and

$$\sup_{n \in \mathbb{N}} \max_{1 \leq i \leq d} \sup_{0 \leq s,t \leq 1} \|D_s \mu(\nu)_t^i\|_{\infty, \ell_2} < \infty, \tag{15}$$

$$\sup_{n \in \mathbb{N}} \max_{1 \leq i \leq d} \left( \sup_{0 \leq s,t \leq 1} \|D_s \sigma(\nu)_t^{i\cdot}\|_{\infty, \ell_2} + \sup_{0 \leq s,t,u \leq 1} \|D_{s,t} \sigma(\nu)_u^{i\cdot}\|_{\infty, \ell_2} \right) < \infty, \tag{16}$$

$$\sup_{n \in \mathbb{N}} \left( \max_{1 \leq i \leq d} \|\Theta_Z(\nu)^{ii}\|_\infty + \max_{1 \leq k \leq d^2} \|1/\mathfrak{V}_n(\nu)^{kk}\|_\infty \right) < \infty, \tag{17}$$

where $\Theta_Z(\nu) := \Sigma_Z(\nu)^{-1}$ and $\mathfrak{V}_n(\nu) := (\Theta_Z(\nu) \otimes \Theta_Z(\nu)) \mathfrak{C}_n(\nu) (\Theta_Z(\nu) \otimes \Theta_Z(\nu))$.

[F2] The $d \times d$ matrix $Q_Z := (1_{\{\Theta_Z^{ij} \neq 0\}})_{1 \leq i,j \leq d}$ is non-random and $\mathfrak{d}(Q_Z) = O(1)$ as $n \to \infty$.

[F3] $r = O(d)$ and $(\log d)^{13}/n \to 0$ as $n \to \infty$.

We give a few remarks on these assumptions. First, [F1] imposes the (local) Malliavin differentiability on the coefficient processes of the residual process $Z$ and the local boundedness on their Malliavin derivatives. Such an assumption is necessary for the application of the high-dimensional mixed normal limit theorem of [20] to our setting (see Lemma A16). Please note that we do not need to impose this type of assumption on the factor process $X$. We also remark that analogous assumptions are sometimes used in the literature of high-frequency financial econometrics even in low-dimensional settings; see e.g., [49,50]. Second, [F2] is clearly understood when we consider a Gaussian graphical model associated with $\Sigma_Z$: The non-randomness of $Q_Z$ implies that the edge structure of this Gaussian graphical model is determined in a non-random manner (by conditioning, it is indeed sufficient that the edge structure is determined independently of the driving Wiener process $W$). Also, we remark that the condition $\mathfrak{d}(Q_Z) = O(1)$ is equivalent to [C5] with $\mathfrak{d}_n = 1$. It is seemingly possible to relax this condition so that it allows a diverging sequence $\mathfrak{d}_n$ as long as $\mathfrak{d}_n(\log d)^\kappa/n \to 0$ for an appropriate constant $\kappa > 0$. However, to determine the precise value of $\kappa$, we need to carefully revise the proof of Lemma A16 so that it allows the quantity inside $\sup_{n \in \mathbb{N}}$ in (A7) to diverge as $n \to \infty$. To avoid such an additional complexity, we restrict our attention to the case of $\mathfrak{d}_n = 1$. Third, the condition $(\log d)^{13}/n \to 0$ in [F3] is used again for applying the high-dimensional CLT of [20].

Now we are ready to state our result. Let $\mathcal{A}^{\mathrm{re}}(d^2)$ be the set of all hyperrectangles in $\mathbb{R}^{d^2}$, i.e., $\mathcal{A}^{\mathrm{re}}(d^2)$ consists of all sets $A$ of the form $A = \{x \in \mathbb{R}^{d^2} : a_j \leq x^j \leq b_j \text{ for all } j = 1, \ldots, d^2\}$ for some $-\infty \leq a_j \leq b_j \leq \infty$, $j = 1, \ldots, d^2$.

**Theorem 2.** *Assume [C1]–[C4] and [F1]–[F3]. Let $\lambda_n$ be a sequence of positive-valued random variables such that $\lambda_n^{-1}\sqrt{(\log d)/n} \to^p 0$, $(s_n + r)\lambda_n \to^p 0$ and $\lambda_n^2 s_n \sqrt{n \log d} \to^p 0$ as $n \to \infty$. Then we have*

$$\sup_{A \in \mathcal{A}^{\mathrm{re}}(d^2)} \left| P\left( \sqrt{n}\, \mathrm{vec}(\hat{\Theta}_{Z,\lambda_n} - \Gamma_{Z,n} - \Theta_Z) \in A \right) - P\left( \mathfrak{V}_n^{1/2} \zeta_n \in A \right) \right| \to 0 \tag{18}$$

*and*

$$\sup_{A \in \mathcal{A}^{\mathrm{re}}(d^2)} \left| P\left( \sqrt{n}\, \mathfrak{S}_n^{-1} \mathrm{vec}(\hat{\Theta}_{Z,\lambda_n} - \Gamma_{Z,n} - \Theta_Z) \in A \right) - P\left( \mathfrak{S}_n^{-1} \mathfrak{V}_n^{1/2} \zeta_n \in A \right) \right| \to 0 \tag{19}$$

*as $n \to \infty$.*

**Remark 10.** *$\lambda_n$ is typically chosen of order close to $\sqrt{\log d / n}$ as possible, so $\lambda_n^2 s_n \sqrt{n \log d} \to^p 0$ is almost equivalent to $s_n (\log d)^{\frac{3}{2}} / \sqrt{n} \to 0$. This is stronger than the condition $s_n (\log d) / \sqrt{n} \to 0$ which is used to derive the asymptotic normality of the de-biased weighted graphical Lasso estimator in [24] (Theorem 14.1.6) (note that we assume $\mathfrak{d}(\Theta_Z) = O_p(1)$). This is because Theorem 2 derives the approximations of the joint distributions of the de-biased estimator and its Studentization, while [24] (Theorem 14.1.6) focuses only on approximation of their marginal distributions.*

Theorem 2 is statistically infeasible in the sense that $\mathfrak{V}_n$ is unobservable. Thus, we need to estimate it from the data. Since $\Theta_Z$ is naturally estimated by $\hat{\Theta}_{Z,\lambda_n}$, we construct an estimator for $\mathfrak{C}_n$. Define the $d^2$-dimensional random vectors $\hat{\chi}_h$ by

$$\hat{\chi}_h := \mathrm{vec}\left[ (\hat{Z}_{h/n} - \hat{Z}_{(h-1)/n})(\hat{Z}_{h/n} - \hat{Z}_{(h-1)/n})^\top \right], \qquad h = 1, \ldots, n,$$

where $\hat{Z}_{h/n} := Y_{h/n} - \hat{\beta} X_{h/n}$. Then we set

$$\hat{\mathfrak{C}}_n := n \sum_{h=1}^{n} \hat{\chi}_h \hat{\chi}_h^\top - \frac{n}{2} \sum_{h=1}^{n-1} \left( \hat{\chi}_h \hat{\chi}_{h+1}^\top + \hat{\chi}_{h+1} \hat{\chi}_h^\top \right).$$

**Lemma 3.** *Suppose that the assumptions of Theorem 2 are satisfied. Suppose also $r^2 (\log d) / n = O(1)$ as $n \to \infty$ and that there is a constant $\gamma \in (0, \frac{1}{2}]$ such that*

$$\sup_{0 < t \leq 1 - \frac{1}{n}} \left\| \max_{1 \leq i,j \leq d} \left| c(v)_{t+\frac{1}{n}}^{ij} - c(v)_t^{ij} \right| \right\|_2 = O(n^{-\gamma}) \tag{20}$$

*as $n \to \infty$ for all $v \in \mathbb{N}$. Then, $\|\hat{\mathfrak{C}}_n - \mathfrak{C}_n\|_{\ell_\infty} = O_p(r(\log d)^{5/2} / \sqrt{n} + n^{-\gamma})$ as $n \to \infty$.*

Let us set $\hat{\mathfrak{V}}_n := (\hat{\Theta}_{Z,\lambda_n} \otimes \hat{\Theta}_{Z,\lambda_n}) \hat{\mathfrak{C}}_n (\hat{\Theta}_{Z,\lambda_n} \otimes \hat{\Theta}_{Z,\lambda_n})$ and $\hat{\mathfrak{S}}_n := \mathrm{diag}(\hat{\mathfrak{V}}_n)^{1/2}$.

**Corollary 1.** *Under the assumptions of Lemma 3, we have the following results:*

(a) *Assume $s_n \lambda_n \log d \to^p 0$ and $r(\log d)^{7/2} / \sqrt{n} + n^{-\gamma} \log d \to 0$ as $n \to \infty$. Then,*

$$\lim_{n \to \infty} \sup_{A \in \mathcal{A}^{\mathrm{re}}(d^2)} \left| P\left( \sqrt{n} \hat{\mathfrak{S}}_n^{-1} \mathrm{vec}(\hat{\Theta}_{Z,\lambda_n} - \Gamma_{Z,n} - \Theta_Z) \in A \right) - P\left( \mathfrak{S}_n^{-1} \mathfrak{V}_n^{1/2} \zeta_n \in A \right) \right| = 0.$$

(b) *Assume $s_n \lambda_n (\log d)^2 \to^p 0$ and $r(\log d)^{9/2} / \sqrt{n} + n^{-\gamma} (\log d)^2 \to 0$ as $n \to \infty$. Then,*

$$\sup_{A \in \mathcal{A}^{\mathrm{re}}(d^2)} \left| P\left( \hat{\mathfrak{V}}_n^{1/2} \zeta_n \in A \mid \mathcal{F} \right) - P\left( \mathfrak{V}_n^{1/2} \zeta_n \in A \mid \mathcal{F} \right) \right| \to^p 0,$$

$$\sup_{A \in \mathcal{A}^{\mathrm{re}}(d^2)} \left| P\left( \hat{\mathfrak{S}}_n^{-1} \hat{\mathfrak{V}}_n^{1/2} \zeta_n \in A \mid \mathcal{F} \right) - P\left( \mathfrak{S}_n^{-1} \mathfrak{V}_n^{1/2} \zeta_n \in A \mid \mathcal{F} \right) \right| \to^p 0$$

*as $n \to \infty$.*

Corollary 1(a) particularly implies that

$$\lim_{n \to \infty} \max_{1 \leq i,j \leq d} \sup_{x \in \mathbb{R}} \left| P\left( \frac{\sqrt{n} \left( \hat{\Theta}_{Z,\lambda_n}^{ij} - \Gamma_{Z,n}^{ij} - \Theta_Z^{ij} \right)}{\hat{\mathfrak{s}}_n^{ij}} \leq x \right) - \Phi(x) \right| = 0, \tag{21}$$

where $\hat{\mathfrak{s}}_n^{ij} := \hat{\mathfrak{S}}_n^{(i-1)d+j,(i-1)d+j}$ and $\Phi$ is the standard normal distribution function. This result can be used to construct entry-wise confidence intervals for $\Theta_Z$. Meanwhile, combining Corollary 1(b) with [20] (Proposition 3.2), we can estimate the quantiles of $\max_{k \in \mathcal{K}} (\mathfrak{V}_n^{1/2} \zeta_n)^k$ and $\max_{k \in \mathcal{K}} (\mathfrak{S}_n^{-1} \mathfrak{V}_n^{1/2} \zeta_n)^k$ for a given set of indices $\mathcal{K} \subset \{1, \ldots, d^2\}$ by simulation. Such a result can be used to construct simultaneous confidence intervals and control the family-wise error rate in multiple testing for entries of $\Theta_Z$; see Sections 2.3–2.4 of [51] for details.

As announced, another application of our result is to construct an estimator with selection consistency via thresholding. This is carried out by using the following result:

**Corollary 2.** *Let* $\alpha_n \in (0,1)$ $(n = 1, 2, \ldots)$ *satisfy* $\alpha_n \to \alpha$ *and* $-\log \alpha_n = O(\log d)$ *as* $n \to \infty$ *for some* $\alpha \in [0,1)$. *Define* $\mathfrak{c}_n := \Phi^{-1}\left(1 - \frac{\alpha_n}{d(d-1)}\right)$ *and*

$$\hat{S}_n(\Theta_Z) := \left\{ (i,j) : i \neq j \text{ and } \frac{\sqrt{n}|\hat{\Theta}_{Z,\lambda_n}^{ij} - \Gamma_{Z,n}^{ij}|}{\hat{\mathfrak{s}}_n^{ij}} > \mathfrak{c}_n \right\}.$$

*Then, under the assumptions of Corollary 1(a), we have*

$$\liminf_{n \to \infty} P(\hat{S}_n(\Theta_Z) = S(\Theta_Z)) \geq 1 - 2\alpha,$$

*provided that* $\sqrt{n/\log d} \min_{(i,j) \in S(\Theta_Z)} |\Theta_Z^{ij}| \to^p \infty$ *as* $n \to \infty$.

Please note that the last condition is satisfied if $\min_{(i,j) \in S(\Theta_Z)} |\Theta_Z^{ij}|$ is bounded away from zero because $\sqrt{n/\log d} \to \infty$ under our assumptions. Taking the sequence $\alpha_n$ so that $\alpha = 0$ in Corollary 2, we can asymptotically recover the support of $\Theta_Z$. In this case, if we define $\widetilde{\Theta}_{Z,\lambda_n} = (\widetilde{\Theta}_{Z,\lambda_n}^{ij})_{1 \leq i,j \leq d}$ by

$$\widetilde{\Theta}_{Z,\lambda_n}^{ij} = \begin{cases} \hat{\Theta}_{Z,\lambda_n}^{ij} - \Gamma_{Z,n}^{ij} & \text{if } i = j \text{ or } (i,j) \in \hat{S}_n(\Theta_Z), \\ 0 & \text{otherwise}, \end{cases}$$

$\widetilde{\Theta}_{Z,\lambda_n}$ will be oracle in the sense of [38]. However, we note that the estimator $\widetilde{\Theta}_{Z,\lambda_n}$ would not be continuous in data, so it would not satisfy the third desirable property in [38] (p. 1349). To construct an oracle estimator for $\Theta_Z$ which is continuous in data, we will need to consider a non-concave penalized estimator as in [52]. This is left to future research.

## 5. Simulation Study

### 5.1. Implementation

To implement the proposed estimation procedure, we need to solve the optimization problem in Equation (10). Among many existing algorithms to solve this problem, we employ the GLASSOFAST algorithm of [53], which is an improved implementation of the popular GLASSO algorithm of [54] and implemented in the R package **glassoFast**.

The remaining problem is how to select the penalty parameter $\lambda$. Following [17,55], we select it by minimizing the following formally defined Bayesian information criterion (BIC):

$$\text{BIC}(\lambda) := n \left\{ \text{tr}\left(\hat{\Theta}_{Z,\lambda} \hat{\Sigma}_{Z,n}\right) - \log \det\left(\hat{\Theta}_{Z,\lambda}\right) \right\} + (\log n) \sum_{i \leq j} \mathbb{1}_{\left\{\hat{\Theta}_{Z,\lambda}^{ij} \neq 0\right\}}.$$

The minimization is carried out by grid search. The grid $\{\lambda_1, \ldots, \lambda_m\}$ is constructed analogously to the R package **glmnet** (see Section 2.5 of [56] for details): First, as the maximum value $\lambda_{\max}$ of the grid, we take the smallest value for which all the off-diagonal entries of $\hat{\Theta}_{Z,\lambda_{\max}}$ are zero: In our case, $\lambda_{\max}$ is set to the maximum modulus of the off-diagonal entries of $\hat{\Sigma}_{Z,n}$ (cf. [57]

(Corollary 1)). Next, we take a constant $\varepsilon > 0$ and set $\lambda_{\min} := \varepsilon\lambda_{\max}$ as the minimum value of the grid. Finally, we construct the values $\lambda_1, \ldots, \lambda_m$ increasing from $\lambda_{\min}$ to $\lambda_{\max}$ on the log scale:

$$\lambda_i = \exp\left(\log(\lambda_{\min}) + \frac{i-1}{m-1}\log(\lambda_{\max}/\lambda_{\min})\right), \qquad i = 1, \ldots, m.$$

We use $\varepsilon = \sqrt{(\log d)/n}$ and $m = 10$ in our experiments (the computation procedure of the weighted graphical Lasso described here is implemented in the R package **yuima** as the function `cce.factor` with the option `regularize="glasso"` since version 1.9.2).

*5.2. Simulation Design*

We basically follow the setting of [8]. We simulate the model (7) with the following specification: For the factor process $X$, we set $r = 3$ and

$$dX_t^j = \mu_j dt + \sqrt{v_t^j}dW_t^j, \qquad dv_t^j = \kappa_j(\theta_j - v_t^j)dt + \eta_j\sqrt{v_t^j}\left(\rho_j dW_t^j + \sqrt{1-\rho_j^2}d\widetilde{W}_t^j\right), \qquad j = 1, 2, 3,$$

where $W^1, W^2, W^3, \widetilde{W}^1, \widetilde{W}^2, \widetilde{W}^3$ are independent standard Wiener processes. We set $\kappa = (3, 4, 5), \theta = (0.09, 0.04, 0.06), \eta = (0.3, 0.4, 0.3), \rho = (-0.6, -0.4, -0.25)$ and $\mu = (0.05, 0.03, 0.02)$. The initial value $v_0^j$ is drawn from the stationary distribution of the process $(v_t^j)_{t\in[0,1]}$, i.e., the gamma distribution with shape $2\kappa_j\theta_j/\eta_j^2$ and rate $2\kappa_j/\eta_j^2$. The entries of the loading $\beta$ are independently drawn as $\beta^{i1} \overset{i.i.d.}{\sim} \mathcal{U}[0.25, 2.25]$ and $\beta^{i2}, \beta^{i3} \overset{i.i.d.}{\sim} \mathcal{U}[-0.5, 0.5]$ ($\mathcal{U}[a, b]$ denotes the uniform distribution on $[a, b]$). Finally, as the residual process $Z$, we take a $d$-dimensional Wiener process with covariance matrix $Q$. We consider the following two designs for $Q$:

**Design 1**   $Q$ is a block diagonal matrix with 10 blocks of size $(d/10) \times (d/10)$. Each block has diagonal entries independently generated from $\mathcal{U}[0.2, 0.5]$ and a constant correlation of 0.25.

**Design 2**   We simulate a Chung-Lu random graph $\mathcal{G}$ and set $Q := (\mathsf{E}_d + D - A)^{-1}$, where $D$ and $A$ are respectively the degree and adjacent matrices of the random graph $\mathcal{G}$. Formally, given a weight vector $w \in \mathbb{R}^d$ with $w \geq 0$, $A$ is defined as a $d \times d$ symmetric random matrix such that all the diagonal entries of $A$ are equal to 0 and the off-diagonal upper triangular entries are generated by independent Bernoulli variables so that $P(A^{ij} = 1) = 1 - P(A^{ij} = 0) = w^i w^j / \sum_{k=1}^d w^k$ for $i < j$. Then, $D$ is defined as the diagonal matrix such that the $j$-th diagonal entry of $D$ is given by $\mathfrak{d}_j(A) = \sum_{i=1}^d A^{ij}$. The weight vector $w$ is specified as follows: For every $i = 1, \ldots, d$, we set $w^i := c\left((i + i_0 - 1)/d\right)^{-1/(\alpha-1)}$ with $i_0 := d(c/w_M)^{\alpha-1}$ and $c := (\alpha-2)/(\alpha-1)$, where we use $\alpha = 2.5$ and $w_M = \lfloor d^{0.45}\rfloor$.

Design 1 is the same one as in [8]. Design 2 is motivated by the recent work of Barigozzi et al. [42], which reports that several characteristics of the residual precision matrix of the S&P 500 assets exhibit power-law behaviors and they are well-described by the power-law partial correlation network model proposed in [42]; the specification in Design 2 is the same one as in the simulation study of [42].

We observe the processes $Y$ and $X$ at the equidistant times $h/n$, $h = 0, 1, \ldots, n$. We set $d = 500$ and vary $n$ as $n \in \{78, 130, 195, 390, 780\}$. We run 10,000 Monte Carlo iterations for each experiment.

*5.3. Results*

We begin by assessing the estimation accuracy of the proposed estimator in various norms. For comparison, we consider the following 5 different methods to estimate $\Sigma_Y$:

RC          We simply use the realized covariance matrix $\widehat{[Y, Y]}_1^n$ defined by Equation (14) to estimate $\Sigma_Y$.

glasso     We estimate $\Sigma_Y^{-1}$ by the (unweighted) graphical Lasso based on $\widehat{[Y, Y]}_1^n$. Then, $\Sigma_Y$ is estimated by its inverse.

wglasso    We estimate $\Sigma_Y^{-1}$ by the weighted graphical Lasso based on $\widehat{[Y,Y]}_1^n$ (i.e., the estimator defined by Equation (1) with $\hat{\Sigma}_n = \widehat{[Y,Y]}_1^n$). Then, $\Sigma_Y$ is estimated by its inverse.

f-glasso    We estimate $\Sigma_Z^{-1}$ by the (unweighted) graphical Lasso based on $\hat{\Sigma}_{Z,n}$ defined by Equation (9) with $\hat{\Sigma}_{Y,n} = \widehat{[Y,Y]}_1^n$ and $\hat{\Sigma}_{X,n} = \widehat{[X,X]}_1^n$. Then, $\Sigma_Y$ is estimated by Equation (11) with $\hat{\Theta}_{Z,\lambda}$ being the estimator so constructed.

f-wglasso    We estimate $\Sigma_Z^{-1}$ by the weighted graphical Lasso based on $\hat{\Sigma}_{Z,n}$ defined by Equation (9) with $\hat{\Sigma}_{Y,n} = \widehat{[Y,Y]}_1^n$ and $\hat{\Sigma}_{X,n} = \widehat{[X,X]}_1^n$. Then, $\Sigma_Y$ is estimated by Equation (11) with $\hat{\Theta}_{Z,\lambda}$ being the estimator so constructed.

In addition, for Design 1, we also consider the estimator proposed in [8]: Assuming that we know which entries of $\Sigma_Z$ are zero, we estimate $\Sigma_Y$ by $\hat{\beta}_n \hat{\Sigma}_{X,n} \hat{\beta}_n^\top + (\hat{\Sigma}_{Z,n}^{ij} 1_{\{\Sigma_Z^{ij} \neq 0\}})_{1 \leq i,j \leq d}$. We label this method f-thr. Since the estimates of RC and f-thr are not always regular, we use their Moore-Penrose generalized inverses to estimate $\Sigma_Y^{-1}$ when they are singular. Please note that the methods glasso and f-glasso correspond to those proposed in [17], while wglasso and f-wglasso are those proposed in this paper. We report the simulation results in Tables 1 and 2.

We first focus on the accuracy of estimating the precision matrix $\Sigma_Y^{-1}$. The tables reveal the excellent performance of graphical Lasso-based methods. In particular, they outperform f-thr in Design 1 except for the case $n = 780$ even when we ignore the factor structure of the model. Nevertheless, the tables also show apparent benefit to take the factor structure into account in constructing the graphical Lasso type estimators. When we compare the weighted graphical Lasso estimators with the unweighted versions, the weighted ones tend to outperform the unweighted ones as $n$ increases, especially when the factor structure is taken into account. This is more pronounced in Design 2. It is also worth mentioning that the estimation errors for $\Sigma_Y^{-1}$ in the method RC are greater at $n = 390,780$ than those at $n = 78,130,195$. This is presumably due to a "resonance" effect between the sample size $n$ and dimension $d$ coming from the use of the Moore-Penrose generalized inverse, which is well-known in multivariate analysis (see e.g., [58]): The estimation error for the precision matrix by the generalized inverse of the sample covariance matrix drastically increases as $n$ approaches $d$. Theoretically, this occurs because the smallest non-zero eigenvalue of the sample covariance matrix tends to 0 as $n$ approaches $d$.

Turning to the estimation accuracy for $\Sigma_Y$ in terms of the $\ell_\infty$-norm, we find little advantage to use the graphical Lasso type methods over the realized covariance matrix: f-glasso and f-wglasso tend to outperform RC at small values of $n$, but the differences of the performance become less clear as $n$ increases. From a theoretical point of view, this is not surprising because the realized covariance matrix is a consistent estimator for $\Sigma_Y$ in the $\ell_\infty$-norm with the convergence rate $\sqrt{\log d / n}$; this can be seen from e.g., Lemma A15. Meanwhile, in Design 1 f-thr performs the best in terms of estimating $\Sigma_Y$ at all the values of $n$.

Next we assess the accuracy of the mixed normal approximation for the de-biased estimator. For this purpose, we construct entry-wise confidence intervals for $\Theta_Z$ based on Equation (21) (with taking the factor structure into account) and evaluate their nominal coverages. Table 3 reports these coverages averaged over the sets $\{(i,j) : i \leq j, \Theta_Z^{ij} = 0\}$ and $\{(i,j) : i \leq j, \Theta_Z^{ij} \neq 0\}$, respectively. We see from the table that the asymptotic approximation perfectly works to construct confidence intervals for zero entries of $\Theta_Z$. By contrast, confidence intervals for non-zero entries of $\Theta_Z$ tend to be over-coverages, especially in Design 1. However, these coverage distortions tend to be moderate at larger values of $n$, which suggests that the normal approximation starts to work for relatively large sample sizes.

**Table 1.** Estimation accuracy of different methods in Design 1.

|  | $n$ | RC | glasso | wglasso | f-glasso | f-wglasso | f-thr |
|---|---|---|---|---|---|---|---|
| $\|\|\|\hat{\Sigma}_Y^{-1} - \Sigma_Y^{-1}\|\|\|_\infty$ | 78 | 22.431 | 18.857 | 19.083 | 15.122 | 15.130 | 416.197 |
|  | 130 | 26.307 | 17.931 | 17.954 | 14.353 | 14.353 | 93.242 |
|  | 195 | 45.795 | 17.447 | 17.471 | 13.923 | 13.928 | 50.605 |
|  | 390 | 722.381 | 16.687 | 16.678 | 11.306 | 10.806 | 25.335 |
|  | 780 | 423.434 | 15.965 | 15.908 | 9.387 | 8.851 | 15.227 |
| $\|\|\|\hat{\Sigma}_Y^{-1} - \Sigma_Y^{-1}\|\|\|_2$ | 78 | 6.576 | 4.270 | 4.263 | 3.419 | 3.420 | 138.442 |
|  | 130 | 6.508 | 3.654 | 3.468 | 3.193 | 3.193 | 28.384 |
|  | 195 | 6.480 | 3.381 | 3.271 | 3.094 | 3.097 | 14.307 |
|  | 390 | 203.038 | 3.009 | 3.015 | 2.133 | 2.100 | 6.446 |
|  | 780 | 93.354 | 2.788 | 2.855 | 1.782 | 1.693 | 3.562 |
| $\|\hat{\Sigma}_Y - \Sigma_Y\|_{\ell_\infty}$ | 78 | 0.361 | 0.432 | 0.441 | 0.351 | 0.351 | 0.347 |
|  | 130 | 0.279 | 0.311 | 0.296 | 0.281 | 0.281 | 0.268 |
|  | 195 | 0.227 | 0.255 | 0.250 | 0.241 | 0.241 | 0.219 |
|  | 390 | 0.160 | 0.181 | 0.189 | 0.166 | 0.169 | 0.154 |
|  | 780 | 0.112 | 0.130 | 0.143 | 0.118 | 0.119 | 0.108 |

*Note.* RC: realized covariance matrix; glasso: graphical Lasso; wglasso: weighted graphical Lasso; f-glasso: graphical Lasso with taking the factor structure into account; f-wglasso: weighted graphical Lasso with taking the factor structure into account; f-thr: location-based thresholding with taking the factor structure into account (the method of [8]). The results are based on 10,000 Monte Carlo iterations.

**Table 2.** Estimation accuracy of different methods in Design 2.

|  | $n$ | RC | glasso | wglasso | f-glasso | f-wglasso |
|---|---|---|---|---|---|---|
| $\|\|\|\hat{\Sigma}_Y^{-1} - \Sigma_Y^{-1}\|\|\|_\infty$ | 78 | 47.934 | 43.144 | 43.055 | 35.347 | 35.263 |
|  | 130 | 48.266 | 43.166 | 41.750 | 34.767 | 34.284 |
|  | 195 | 50.049 | 42.806 | 40.571 | 34.154 | 32.835 |
|  | 390 | 338.847 | 41.060 | 37.801 | 33.100 | 29.934 |
|  | 780 | 401.447 | 38.886 | 34.961 | 32.163 | 23.121 |
| $\|\|\|\hat{\Sigma}_Y^{-1} - \Sigma_Y^{-1}\|\|\|_2$ | 78 | 17.805 | 13.557 | 13.522 | 7.857 | 7.843 |
|  | 130 | 17.798 | 13.543 | 12.628 | 7.954 | 7.866 |
|  | 195 | 17.752 | 13.319 | 11.630 | 8.006 | 7.742 |
|  | 390 | 87.239 | 12.296 | 9.888 | 8.059 | 7.416 |
|  | 780 | 55.619 | 11.189 | 8.522 | 8.065 | 6.072 |
| $\|\hat{\Sigma}_Y - \Sigma_Y\|_{\ell_\infty}$ | 78 | 0.669 | 0.723 | 0.721 | 0.632 | 0.631 |
|  | 130 | 0.509 | 0.678 | 0.572 | 0.489 | 0.481 |
|  | 195 | 0.412 | 0.567 | 0.470 | 0.403 | 0.390 |
|  | 390 | 0.289 | 0.298 | 0.339 | 0.282 | 0.273 |
|  | 780 | 0.203 | 0.198 | 0.252 | 0.197 | 0.192 |

*Note.* RC: realized covariance matrix; glasso: graphical Lasso; wglasso: weighted graphical Lasso; f-glasso: graphical Lasso with taking the factor structure into account; f-wglasso: weighted graphical Lasso with taking the factor structure into account. The results are based on 10,000 Monte Carlo iterations.

**Table 3.** Average coverages of entry-wise confidence intervals.

|  |  | Design 1 | | Design 2 | |
|---|---|---|---|---|---|
|  | $n$ | 95% | 99% | 95% | 99% |
| $\Theta_Z^{ij} = 0$ | 78 | 95.21 | 99.04 | 95.22 | 99.04 |
|  | 130 | 95.13 | 99.03 | 95.13 | 99.03 |
|  | 195 | 95.09 | 99.02 | 95.09 | 99.02 |
|  | 390 | 95.04 | 99.01 | 95.05 | 99.01 |
|  | 780 | 95.02 | 99.00 | 95.02 | 99.01 |

**Table 3.** *Cont.*

|  | | Design 1 | | Design 2 | |
|---|---|---|---|---|---|
|  | $n$ | 95% | 99% | 95% | 99% |
| $\Theta_Z^{ij} \neq 0$ | 78 | 99.33 | 99.87 | 95.16 | 99.03 |
|  | 130 | 99.82 | 99.96 | 95.90 | 99.18 |
|  | 195 | 99.97 | 99.99 | 96.36 | 99.26 |
|  | 390 | 96.00 | 99.20 | 96.65 | 99.33 |
|  | 780 | 96.09 | 99.22 | 96.41 | 99.27 |

This table reports the average coverages of entry-wise confidence intervals for the residual precision matrix $\Theta_Z$ over the sets $\{(i,j) : i \leq j, \Theta_Z^{ij} = 0\}$ and $\{(i,j) : i \leq j, \Theta_Z^{ij} \neq 0\}$, respectively. The confidence intervals are constructed based on the normal approximation Equation (21). The results are based on 10,000 Monte Carlo iterations.

## 6. Empirical Application

To illustrate the applicability of the proposed method to real data analysis, we conduct a simple empirical study using high-frequency financial data. We take 1 March 2018 as the observation interval $[0, 1]$ and the log-price processes of the component stocks of the S&P 500 index as the process $Y$. In addition, as is often performed in the literature, we regard the SPDR S&P 500 ETF (SPY) as the observable factor process $X$. We use 5-minute returns to compute the estimators presented in Section 4. The dataset is provided by Bloomberg. Please note that our setting implies $d = 504$ and $n = 77$, yielding a high-dimensional setting considered in this paper (note that our dataset does not contain observations at the market opening).

The selection procedure presented in Section 5.1 suggests $\lambda_n \approx 0.272$. Then we estimate the support $S(\Theta_Z)$ of $\Theta_Z$ by the estimator $\hat{S}_n(\Theta_Z)$ with $\alpha_n = 0.05$ from Corollary 2. Figure 1 shows the partial correlation network induced by $\hat{S}_n(\Theta_Z)$, drawn by the R package **igraph**. Specifically, it depicts the undirected graph with vertices consisting of the S&P 500 component stocks and an edge set given by $\hat{S}_n(\Theta_Z)$. To illuminate the relationship between the network and sector structures, we color the vertices according to their Global Industry Classification Standard (GICS) sectors. We find there are strong interconnections in several sectors such as Consumer Staples, Energy, Real Estate and Utilities. The figure also suggests that the network would have some characteristics that are commonly observed in scale-free networks: It consists of a giant component with several hubs and a few small components. This is consistent with an observation made in [42]. Indeed, in [42] the authors have proposed a model for $\Theta_Z$ that induces a scale-free partial correlation network. According to their model, the decay of the largest eigenvalues of $\Theta_Z$ also exhibits power-law behavior. More precisely, letting $\Lambda_1 \geq \cdots \geq \Lambda_d$ be the ordered eigenvalues of $\Theta_Z$, we have $\Lambda_i \asymp i^{-\alpha}$ with some $\alpha > 0$ for moderate $i$ and large $d$. Then, it is interesting to check whether this is the case in our dataset. Figure 2 shows the log-log size-rank plot for the 50 largest eigenvalues of $\hat{\Theta}_{Z,\lambda_n}$. We see that except for the three largest eigenvalues, they clearly display power-law behavior.
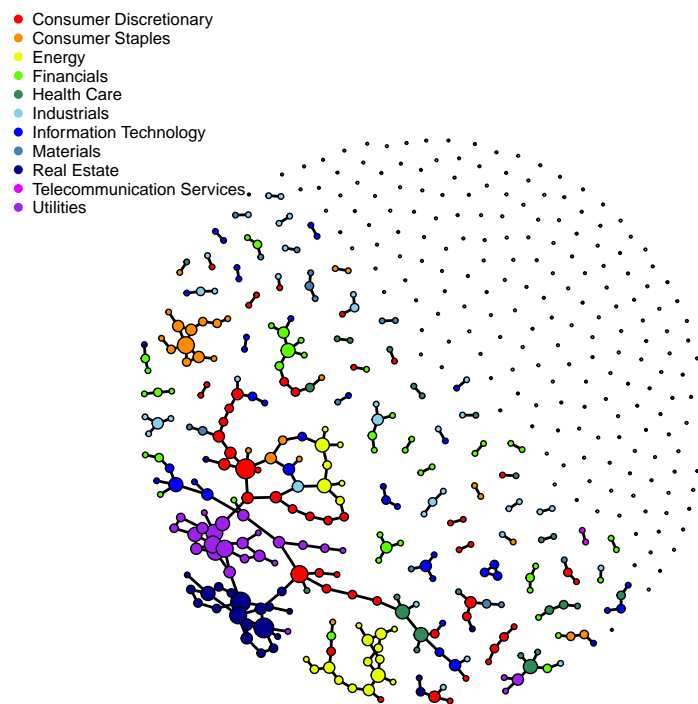
**Figure 1.** Partial correlation network of the S&P 500 component stocks on 1 March 2018.
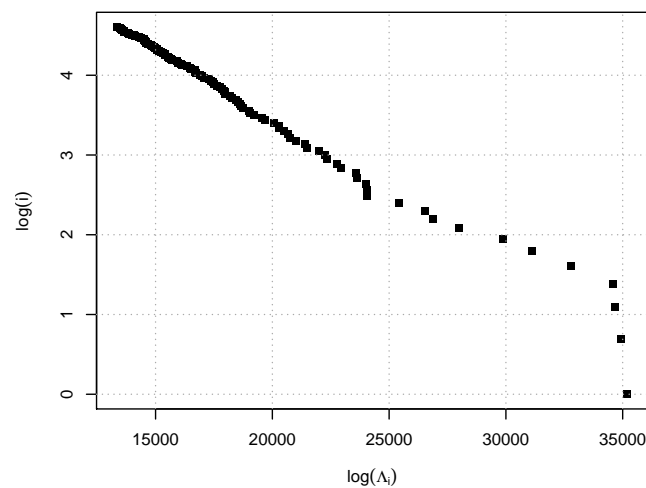


**Figure 2.** Log-log size-rank plot for the eigenvalues of the estimated residual precision matrix of the S&P 500 component stocks on 1 March 2018.

## 7. Conclusions

In this paper, we have developed a generic asymptotic theory to estimate the high-dimensional precision matrix of high-frequency data using the weighted graphical Lasso. We have shown that the consistency of the weighted graphical Lasso estimator in matrix operator norms follows from the consistency of the initial estimator in the $\ell_\infty$-norm, while the asymptotic mixed normality of its de-biased version follows from that of the initial estimator, where the asymptotic mixed normality has been formulated appropriately for the high-dimensional setting considered here. Our theory also encompasses a situation where a known factor structure is present in the data. In such a situation,

we have applied the weighted graphical Lasso to the residual process obtained after removing the effect of factors.

　　We have applied the developed theory to the concrete situation where we can use the realized covariance matrix as the initial covariance estimator. We have derived the desirable asymptotic mixed normality of the realized covariance matrix by an application of the recent high-dimensional central limit theorem obtained in [20], where Malliavin calculus resolves the main theoretical difficulties caused by the high-dimensionality. Consequently, we have obtained a feasible asymptotic distribution theory to conduct inference for entries of the precision matrix. A Monte Carlo study has shown the good finite sample performance of our asymptotic theory.

　　A natural direction for future work is to apply the developed theory to a more complex situation where the process is asynchronously observed with noise and/or jumps. To accomplish this purpose, we need to establish the high-dimensional asymptotic mixed normality of relevant covariance estimators.

**Conflicts of Interest:** The author declares no conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

| | |
|---|---|
| i.i.d. | independent and identically distributed |
| MLE | maximum likelihood estimation |
| KKT | Karush–Kuhn–Tucker |
| CLT | central limit theorem |

## Appendix A. Matrix Inequalities

　　This appendix collects some elementary (but less trivial) inequalities for matrices used in the proofs of the main results.

**Lemma A1.** *Let $A \in \mathcal{S}_d$. Then $\Lambda_{\min}(A) \leq A^{ii} \leq \Lambda_{\max}(A)$ for every $i = 1, \dots, d$.*

**Proof.** See Theorem 14 in [59] (Chapter 11). □

**Lemma A2.** *Let $A \in \mathcal{S}_d^+$ and $B \in \mathbb{R}^{d \times r}$. Then $\Lambda_{\max}(B^\top A B) \leq \Lambda_{\max}(B^\top B)\Lambda_{\max}(A)$ and $\Lambda_{\min}(B^\top A B) \geq \Lambda_{\min}(B^\top B)\Lambda_{\min}(A)$.*

**Proof.** Let $x$ be an eigenvector associated with $\Lambda_{\max}(A)$ such that $\|x\|_{\ell_2} = 1$. Then, by Theorem 4 in [59] (Chapter 11) we have $\Lambda_{\max}(B^\top A B) = x^\top B^\top A B x \leq \Lambda_{\max}(A)x^\top B^\top B x \leq \Lambda_{\max}(A)\Lambda_{\max}(B^\top B)$. Therefore, we obtain the first inequality. The second one can be shown analogously. □

**Lemma A3.** *Let $A, B \in \mathcal{S}_d$. Then $|\Lambda_{\max}(A) - \Lambda_{\max}(B)| \vee |\Lambda_{\min}(A) - \Lambda_{\min}(B)| \leq |||A - B|||_2$.*

**Proof.** Noting the identity $|||C|||_2 = \Lambda_{\max}(C) \vee (-\Lambda_{\min}(C))$ holding for any symmetric matrix $C$, the desired result follows from Weyl's inequality (cf. Corollary 4.3.15 in [60]). □

**Lemma A4.** *For any $A \in \mathcal{S}_d$, $|||A|||_1 = |||A|||_\infty \leq \sqrt{\mathfrak{d}(A)}|||A|||_2$.*

**Proof.** This is a straightforward consequence of the Schwarz inequality. □

**Lemma A5.** *Let $A, B \in \mathbb{R}^{r \times r}$. If $A$ is invertible and $|||A^{-1}(B-A)|||_w < 1$ for some $w \in [1, \infty]$, $B$ is invertible and*

$$|||B^{-1} - A^{-1}|||_w \leq \frac{|||A^{-1}|||_w |||A^{-1}(B-A)|||_w}{1 - |||A^{-1}(B-A)|||_w}.$$

**Proof.** See pages 381–382 of [60]. □

**Lemma A6.** *Let $A \in \mathcal{S}_r$ and $B, C \in \mathbb{R}^{d \times r}$. Then*

$$\|BAC^\top\|_{\ell_\infty} \leq |||A|||_2 \left( \max_{1 \leq i \leq d} \|B^{i\cdot}\|_{\ell_2} \right) \left( \max_{1 \leq j \leq d} \|C^{j\cdot}\|_{\ell_2} \right) \leq r |||A|||_2 \|B\|_{\ell_\infty} \|C\|_{\ell_\infty}.$$

**Proof.** This result has essentially been shown in [61] (Lemma A.7). Since $A$ is symmetric, there is an orthogonal matrix $U \in \mathbb{R}^{r \times r}$ such that $\Lambda := U^\top A U$ is a diagonal matrix. Now, for any $i, j \in \{1, \ldots, d\}$,

$$|(BAC^\top)^{ij}| = |(B^{i\cdot})^\top A C^{j\cdot}| = |(B^{i\cdot})^\top U \Lambda U^\top C^{j\cdot}| = \left| \sum_{k=1}^r \Lambda^{kk} ((B^{i\cdot})^\top U)^k (U^\top C^{j\cdot})^k \right|$$

$$\leq \max_{1 \leq k \leq r} |\Lambda^{kk}| \|U^\top B^{i\cdot}\|_{\ell_2} \|U C^{j\cdot}\|_{\ell_2} = |||A|||_2 \|B^{i\cdot}\|_{\ell_2} \|C^{j\cdot}\|_{\ell_2} \leq r |||A|||_2 \|B\|_{\ell_\infty} \|C\|_{\ell_\infty}.$$

This yields the desired result. □

**Lemma A7.** *Let $A, B, C \in \mathbb{R}^{d \times d}$. Then, for any $i, j = 1, \ldots, d$,*

$$|(BAC^\top)^{ij}|^2 \leq \left( \sum_{k=1}^d |B^{ik}| \right) \left( \sum_{l=1}^d |C^{il}| \right) \sum_{k,l=1}^d |B^{ik} C^{jl}| (A^{kl})^2.$$

**Proof.** This is a straightforward consequence of the Schwarz inequality. □

**Appendix B. Proofs for Section 2**

*Appendix B.1. Proof of Proposition 1*

The following result has essentially been proven in [24] and gives an estimate for the "deterministic part" of oracle inequalities for graphical Lasso type estimators.

**Proposition A1.** *Let $A_0, A \in \mathcal{S}_d$ and assume $\|A - A_0\|_{\ell_\infty} \leq \lambda_0$ for some $\lambda_0 > 0$. Assume also that there are numbers $L > 1$ and $\lambda > 0$ such that $L^{-1} \leq \Lambda_{\min}(A_0) \leq \Lambda_{\max}(A_0) \leq L$, $2\lambda_0 \leq \lambda \leq (8Lc_L)^{-1}$ and $8c_L^2 s \lambda^2 + 2c_L \lambda_0^2 \|\operatorname{diag}(A) - \operatorname{diag}(A_0)\|_{\ell_2}^2 \leq \lambda_0/(2L)$, where $s := s(B_0)$ and $c_L := 8L^2$. Set $B_0 := A_0^{-1}$. Then, for any $B \in \mathcal{S}_d^{++}$ satisfying*

$$\operatorname{tr}(BA) - \log \det(B) + \lambda \|B^-\|_{\ell_1} \leq \operatorname{tr}(B_0 A) - \log \det(B_0) + \lambda \|B_0^-\|_{\ell_1}, \tag{A1}$$

*it holds that*

$$\|B - B_0\|_{\ell_2}^2 / c_L + \lambda \|B^- - B_0^-\|_{\ell_1} \leq 8c_L^2 s \lambda^2 + 2c_L \lambda_0^2 \|\operatorname{diag}(A) - \operatorname{diag}(A_0)\|_{\ell_2}^2.$$

We first prove Proposition A1 under an additional assumption:

**Lemma A8.** *Proposition A1 holds true if we additionally have $\|B - B_0\|_{\ell_2} \leq 1/(2L)$.*

**Proof.** Set $\Delta = B - B_0$. By assumption we have $\|\Delta\|_{\ell_2} \leq 1/(2L)$, so Lemma 2 in [24] implies that

$$\mathcal{E}(\Delta) := \operatorname{tr}(\Delta A_0) - \{\log \det(\Delta + B_0) - \log \det(B_0)\}$$

is well-defined and we have

$$\mathcal{E}(\Delta) \geq c\|\Delta\|_{\ell_2}, \tag{A2}$$

where $c = c_L^{-1}$. Moreover, Equation (A1) yields

$$
\begin{aligned}
\mathcal{E}(\Delta) + \lambda\|B^-\|_{\ell_1} &= -\operatorname{tr}\left(\Delta(A - A_0)\right) + \left\{\operatorname{tr}(BA) - \log\det(B) + \lambda\|B^-\|_{\ell_1}\right\} - \operatorname{tr}(B_0 A) + \log\det(B_0) \\
&\leq -\operatorname{tr}\left(\Delta(A - A_0)\right) + \left\{\operatorname{tr}(B_0 A) - \log\det(B_0) + \lambda\|B_0^-\|_{\ell_1}\right\} - \operatorname{tr}(B_0 A) + \log\det(B_0) \quad \text{(A3)} \\
&= -\operatorname{tr}\left(\Delta(A - A_0)\right) + \lambda\|B_0^-\|_{\ell_1}.
\end{aligned}
$$

Now, note that $\operatorname{tr}(A_1 B_1) = \operatorname{tr}(A_1^- B_1^-) + \operatorname{tr}(\operatorname{diag}(A_1)\operatorname{diag}(B_1))$ and $|\operatorname{tr}(A_1 B_1)| \leq \|A_1 \circ B_1\|_{\ell_1}$ for any $A_1, B_1 \in \mathbb{R}^{d \times d}$. Thus, we infer that

$$
\begin{aligned}
|\operatorname{tr}\left(\Delta(A - A_0)\right)| &\leq \|\Delta^-\|_{\ell_1}\|A^- - A_0^-\|_{\ell_\infty} + \|\operatorname{diag}(\Delta)\|_{\ell_2}\|\operatorname{diag}(A) - \operatorname{diag}(A_0)\|_{\ell_2} \\
&\leq \lambda_0\|\Delta^-\|_{\ell_1} + \|\operatorname{diag}(A) - \operatorname{diag}(A_0)\|_{\ell_2}\|\operatorname{diag}(\Delta)\|_{\ell_2},
\end{aligned}
$$

where we use $\|A - A_0\|_{\ell_\infty} \leq \lambda_0$ in the last line. Combining this with Equations (A2) and (A3), we conclude that

$$c\|\Delta\|_{\ell_2}^2 + \lambda\|B^-\|_{\ell_1} \leq \lambda_0\|\Delta^-\|_{\ell_1} + \|\operatorname{diag}(A) - \operatorname{diag}(A_0)\|_{\ell_2}\|\operatorname{diag}(\Delta)\|_{\ell_2} + \lambda\|B_0^-\|_{\ell_1}.$$

Let $S := S(B_0)$. Also, for a subset $I$ of $\{1, \ldots, d\}^2$ and a $d \times d$ matrix $U$, we define the $d \times d$ matrix $U_I = (U_I^{ij})_{1 \leq i,j \leq d}$ by $U_I^{ij} = U^{ij} \mathbb{1}_{\{(i,j) \in I\}}$. Then, by definition and assumption, we have $\|B^-\|_{\ell_1} = \|B_S^-\|_{\ell_1} + \|B_{S^c}^-\|_{\ell_1}$, $\|\Delta^-\|_{\ell_1} = \|\Delta_S^-\|_{\ell_1} + \|B_{S^c}^-\|_{\ell_1}$, $\|B_0^-\|_{\ell_1} \leq \|\Delta_S^-\|_{\ell_1} + \|B_S^-\|_{\ell_1}$, $\lambda \geq 2\lambda_0$, so we deduce

$$c\|\Delta\|_{\ell_2}^2 + \frac{\lambda}{2}\|B_{S^c}^-\|_{\ell_1} \leq \frac{3\lambda}{2}\|\Delta_S^-\|_{\ell_1} + \|\operatorname{diag}(A) - \operatorname{diag}(A_0)\|_{\ell_2}\|\operatorname{diag}(\Delta)\|_{\ell_2}.$$

Consequently, we obtain

$$
\begin{aligned}
2c\|\Delta\|_{\ell_2}^2 + \lambda\|\Delta^-\|_{\ell_1} &= 2c\|\Delta\|_{\ell_2}^2 + \lambda(\|B_{S^c}^-\|_{\ell_1} + \|\Delta_S^-\|_{\ell_1}) \\
&\leq 4\lambda\|\Delta_S^-\|_{\ell_1} + 2\|\operatorname{diag}(A) - \operatorname{diag}(A_0)\|_{\ell_2}\|\operatorname{diag}(\Delta)\|_{\ell_2} \\
&\leq 4\lambda\sqrt{s}\|\Delta_S^-\|_{\ell_2} + 2\|\operatorname{diag}(A) - \operatorname{diag}(A_0)\|_{\ell_2}\|\operatorname{diag}(\Delta)\|_{\ell_2} \quad (\because \text{Schwarz inequality}) \\
&\leq 8s\lambda^2/c^2 + c\|\Delta_S^-\|_{\ell_2}^2/2 + 2\|\operatorname{diag}(A) - \operatorname{diag}(A_0)\|_{\ell_2}^2/c + c\|\operatorname{diag}(\Delta)\|_{\ell_2}^2/2,
\end{aligned}
$$

where we use the inequality $xy \leq (x^2 + y^2)/2$ in the last line. Since $\|\Delta\|_{\ell_2}^2 = \|\operatorname{diag}(\Delta)\|_{\ell_2}^2 + \|\Delta^-\|_{\ell_2}^2$, we conclude that

$$c\|\Delta\|_{\ell_2}^2 + \lambda\|\Delta^-\|_{\ell_1} \leq 8s\lambda^2/c^2 + 2\|\operatorname{diag}(A) - \operatorname{diag}(A_0)\|_{\ell_2}^2/c,$$

which completes the proof. □

**Proof of Proposition A1.** Thanks to Lemma A8, it suffices to prove $\|B - B_0\|_{\ell_2} \leq 1/(2L)$.

Set $\tilde{B} = \alpha B + (1 - \alpha)B_0$ with $\alpha = M/(M + \|B - B_0\|_{\ell_2})$ and $M = 1/(2L)$. By definition we have $\|\tilde{B} - B_0\|_{\ell_2} \leq M = 1/(2L)$. Moreover, Equation (A1) and the convexity of the loss function imply that

$$\operatorname{tr}(\tilde{B}A) - \log\det(\tilde{B}) + \lambda\|B^-\|_{\ell_1} \leq \operatorname{tr}(B_0 A) - \log\det(B_0) + \lambda\|B_0^-\|_{\ell_1}.$$

Therefore, we can apply Lemma A8 with replacing $B$ by $\tilde{B}$, and thus we obtain

$$\|\tilde{B} - B_0\|_{\ell_2}^2/c_L + \lambda\|\tilde{B}^- - B_0^-\|_{\ell_1} \leq 8c_L^2 s\lambda^2 + 2c_L\lambda_0^2\|\operatorname{diag}(A) - \operatorname{diag}(\Sigma)\|_{\ell_2}^2.$$

In particular, we have

$$\|\tilde{B} - B_0\|_{\ell_2}^2 \leq c_L\lambda_0/(2L) \leq 1/(16L^2),$$

so we obtain $\|\tilde{B} - B_0\|_{\ell_2} \le 1/(4L) = M/2$. By construction this yields $\|B - B_0\|_{\ell_2} \le M = 1/(2L)$, which completes the proof. $\square$

**Proof of Proposition 1.** Thanks to Lemma 7.2 of [45], it suffices to consider the case $w = 1$.

For any $L, n \in \mathbb{N}$, we define the set $\Omega_{n,L} \subset \Omega$ by

$$\Omega_{n,L} := \{\|\hat{R}_n - R_Y\|_{\ell_\infty} \le \lambda_n/2\} \cap \{L^{-1} \le \Lambda_{\min}(R_Y) \le \Lambda_{\max}(R_Y) \le L\}$$
$$\cap \{s(K_Y) \le Ls_n\} \cap \{8c_L^2 Ls_n\lambda_n > 1/(4L)\},$$

where $c_L := 8L^2$. Then we have

$$\lim_{L\to\infty} \limsup_{n\to\infty} P(\Omega_{n,L}^c) = 0.$$

In fact, noting that Lemma A1 and [A1] yield

$$\max_{1\le j\le d} \left( \Sigma_Y^{jj} + 1/\Sigma_Y^{jj} \right) = O_p(1), \tag{A4}$$

[A1]–[A2], [B1] and Lemma A2 imply that $\lambda_n^{-1}\|\hat{R}_n - R_Y\|_{\ell_\infty} = o_p(1)$, $s(K_Y) = O_p(s_n)$ and $\Lambda_{\max}(R_Y) + 1/\Lambda_{\min}(R_Y) = O_p(1)$. Finally, [B2] yields $\lim_{n\to\infty} P(8Ls_n\lambda_n > 1/(4L)) = 0$ for all $L$. Now, note that $\lambda_n \le 1/(16Lc_L^2) \le (8Lc_L)^{-1}$ on the set $\Omega_{n,L}$. Therefore, applying Proposition A1 with $\lambda := \lambda_n$ and $\lambda_0 := \lambda_n/2$, for any fixed $L$ we have

$$\|\hat{K}_{\lambda_n} - K_Y\|_{\ell_2}^2/c_L + \lambda_n\|\hat{K}_{\lambda_n}^- - K_Y^-\|_{\ell_1} \le 8c_L^2 s_n\lambda_n^2 \qquad \text{on } \Omega_{n,L}.$$

Consequently, we obtain

$$\limsup_{n\to\infty} P\left( \|\hat{K}_{\lambda_n} - K_Y\|_{\ell_2} > 64L^3\sqrt{s_n}\lambda_n \right) \le \limsup_{n\to\infty} P(\Omega_{n,L}^c)$$

and

$$\limsup_{n\to\infty} P\left( \|\hat{K}_{\lambda_n}^- - K_Y^-\|_{\ell_1} > 512L^4 s_n\lambda_n \right) \le \limsup_{n\to\infty} P(\Omega_{n,L}^c).$$

Therefore, we conclude that

$$\limsup_{M\to\infty} \limsup_{n\to\infty} P\left( \|\hat{K}_{\lambda_n} - K_Y\|_{\ell_2} > M\sqrt{s_n}\lambda_n \right) \le \limsup_{L\to\infty} \limsup_{n\to\infty} P(\Omega_{n,L}^c) = 0$$

and

$$\limsup_{M\to\infty} \limsup_{n\to\infty} P\left( \|\hat{K}_{\lambda_n}^- - K_Y^-\|_{\ell_1} > Ms_n\lambda_n \right) \le \limsup_{L\to\infty} \limsup_{n\to\infty} P(\Omega_{n,L}^c) = 0,$$

which yields $\|\hat{K}_{\lambda_n} - K_Y\|_{\ell_2} = O_p(\sqrt{s_n}\lambda_n)$ and $\|\hat{K}_{\lambda_n}^- - K_Y^-\|_{\ell_1} = O_p(s_n\lambda_n)$. In particular, we obtain the first convergence of Equation (3). Moreover, since we have

$$\|\|\hat{K}_{\lambda_n} - K_Y\|\|_1 \le \|\operatorname{diag}(\hat{K}_{\lambda_n}) - \operatorname{diag}(K_Y)\|_{\ell_\infty} + \|\hat{K}_{\lambda_n}^- - K_Y^-\|_{\ell_1}$$
$$\le \|\hat{K}_{\lambda_n} - K_Y\|_{\ell_2} + \|\hat{K}_{\lambda_n}^- - K_Y^-\|_{\ell_1},$$

we also obtain the second convergence of Equation (3).

Now we prove Equation (4). First, Equation (A4) and [B1] yield $\lambda_n^{-1}\|\|\hat{V}_n - V\|\|_1 \to^p 0$. Since $\|\|V\|\|_1 = O_p(1)$, $\|\|K_Y\|\|_1 = O_p(s_n)$ and $\lambda_n = o_p(1)$ by Equation (A4), [A2] and [B2], we obtain $\|\|\hat{V}_n\|\|_1 = O_p(1)$ and $\|\|\hat{K}_{\lambda_n}\|\|_1 = O_p(s_n)$. Since

$$\|\|\hat{\Theta}_{\lambda_n} - \Theta_Y\|\|_1 = \|\|\hat{V}_n\hat{K}_{\lambda_n}\hat{V}_n - VK_YV\|\|_1$$
$$\le \|\|\hat{V}_n - V\|\|_1\|\|\hat{K}_{\lambda_n}\|\|_1\|\|\hat{V}_n\|\|_1 + \|\|V\|\|_1\|\|\hat{K}_{\lambda_n} - K_Y\|\|_1\|\|\hat{V}_n\|\|_1 + \|\|V\|\|_1\|\|K_Y\|\|_1\|\|\hat{V}_n - V\|\|_1,$$

we obtain the first convergence of Equation (4). Next, since $\||\hat{\Theta}_{\lambda_n} - \Theta_Y\||_2 = o_p(1)$ by the above result, [A1] and Lemma A3 yield $\||\hat{\Theta}_{\lambda_n}^{-1}\||_2 = \Lambda_{\min}(\hat{\Theta}_{\lambda_n})^{-1} = O_p(1)$. Since $\||\hat{\Theta}_{\lambda_n}^{-1} - \Sigma_Y\||_2 \leq \||\hat{\Theta}_{\lambda_n}^{-1}\||_2\||\Theta_Y - \hat{\Theta}_{\lambda_n}\||_2\||\Sigma_Y\||_2$, we obtain the second convergence of Equation (4).  □

*Appendix B.2. Proof of Lemma 1*

First, by Proposition 14.4.3 of [62] there is a (not necessarily measurable) $d \times d$ random matrix $\hat{Z}_n$ such that

$$\hat{\Sigma}_n - \hat{\Theta}_{\lambda_n}^{-1} + \lambda_n \hat{V}_n \hat{Z}_n \hat{V}_n = 0, \qquad \|\hat{Z}_n\|_{\ell_\infty} \leq 1,$$

and $\hat{Z}_n^{ij} = \operatorname{sign}(\hat{\Theta}_{\lambda_n}^{ij})$ if $\hat{\Theta}_{\lambda_n}^{ij} \neq 0$. Consequently, it holds that

$$\hat{\Sigma}_n \hat{\Theta}_{\lambda_n} - \mathsf{E}_d + \lambda_n \hat{V}_n \hat{Z}_n \hat{V}_n \hat{\Theta}_{\lambda_n} = 0.$$

Therefore, we have

$$
\begin{aligned}
&\hat{\Theta}_{\lambda_n} - \Theta_Y + \Theta_Y(\hat{\Sigma}_n - \Sigma_Y)\Theta_Y \\
&= \hat{\Theta}_{\lambda_n} - \Theta_Y + \Theta_Y(\hat{\Sigma}_n - \Sigma_Y)\hat{\Theta}_{\lambda_n} - \Theta_Y(\hat{\Sigma}_n - \Sigma_Y)(\hat{\Theta}_{\lambda_n} - \Theta_Y) \\
&= \hat{\Theta}_{\lambda_n} - \Theta_Y + \Theta_Y(\mathsf{E}_d - \lambda_n \hat{V}_n \hat{Z}_n \hat{V}_n \hat{\Theta}_{\lambda_n} - \Sigma_Y \hat{\Theta}_{\lambda_n}) - \Theta_Y(\hat{\Sigma}_n - \Sigma_Y)(\hat{\Theta}_{\lambda_n} - \Theta_Y) \\
&= -\lambda_n \Theta_Y \hat{V}_n \hat{Z}_n \hat{V}_n \hat{\Theta}_{\lambda_n} - \Theta_Y(\hat{\Sigma}_n - \Sigma_Y)(\hat{\Theta}_{\lambda_n} - \Theta_Y) \\
&= \lambda_n(\hat{\Theta}_{\lambda_n} - \Theta_Y)\hat{V}_n \hat{Z}_n \hat{V}_n \hat{\Theta}_{\lambda_n} - (\hat{\Theta}_{\lambda_n} - \hat{\Theta}_{\lambda_n} \hat{\Sigma}_n \hat{\Theta}_{\lambda_n}) - \Theta_Y(\hat{\Sigma}_n - \Sigma_Y)(\hat{\Theta}_{\lambda_n} - \Theta_Y),
\end{aligned}
$$

so we obtain

$$
\begin{aligned}
&\|\hat{\Theta}_{\lambda_n} - \Theta_Y - \Gamma_n + \Theta_Y(\hat{\Sigma}_n - \Sigma_Y)\Theta_Y\|_{\ell_\infty} \\
&\leq \lambda_n \||\hat{\Theta}_{\lambda_n} - \Theta_Y\||_\infty \|\hat{V}_n \hat{Z}_n \hat{V}_n \hat{\Theta}_{\lambda_n}\|_{\ell_\infty} + \||\Theta_Y\||_\infty \|\hat{\Sigma}_n - \Sigma_Y\|_{\ell_\infty} \||\hat{\Theta}_{\lambda_n} - \Theta_Y\||_{\ell_\infty} \\
&\leq \lambda_n \||\hat{\Theta}_{\lambda_n} - \Theta_Y\||_\infty \||\hat{V}_n\||_\infty^2 \||\hat{\Theta}_{\lambda_n}\||_\infty + \||\Theta_Y\||_\infty \|\hat{\Sigma}_n - \Sigma_Y\|_{\ell_\infty} \||\hat{\Theta}_{\lambda_n} - \Theta_Y\||_{\ell_\infty}.
\end{aligned}
$$

Now the desired result follows from Proposition 1 and Lemma A4.  □

*Appendix B.3. Proof of Proposition 2*

In the light of Lemma 3.1 of [20], it is enough to prove

$$\sqrt{\log(m+1)}\left\|J_n \operatorname{vec}\left(\hat{\Theta}_{\lambda_n} - \Theta_Y - \Gamma_n\right) - \tilde{J}_n \operatorname{vec}\left(\hat{\Sigma}_n - \Sigma\right)\right\|_{\ell_\infty} \to^p 0$$

as $n \to \infty$. Please note that $\operatorname{vec}(ABC) = (C^\top \otimes A)\operatorname{vec}(B)$ for any $d \times d$ matrices $A, B, C$ (cf. Theorem 2 in [59] (Chapter 2)). Thus, we obtain the desired result once we prove

$$\sqrt{\log(m+1)}\left\|J_n\left(\hat{\Theta}_{\lambda_n} - \Theta_Y - \Gamma_n + \Theta_Y\left(\hat{\Sigma}_n - \Sigma\right)\Theta_Y\right)\right\|_{\ell_\infty} \to^p 0$$

as $n \to \infty$. This follows from Lemma 1 and the assumptions of the proposition.  □

## Appendix C. Proofs for Section 3

*Appendix C.1. Proof of Proposition 3*

Set $\Omega_n := \{\||\Sigma_X^{-1}(\hat{\Sigma}_{X,n} - \Sigma_X)\||_2 \leq 1/2\}$.

**Lemma A9.** *Under the assumptions of Proposition 3, we have the following results:*

(a) *On the event $\Omega_n$, $\hat{\Sigma}_{X,n}$ is invertible and $\||\hat{\Sigma}_{X,n}^{-1} - \Sigma_X^{-1}\||_2 \leq 2\||\Sigma_X^{-1}\||_2\||\Sigma_X^{-1}(\hat{\Sigma}_{X,n} - \Sigma_X)\||_2$.*
(b) *$\lambda_n^{-1}\||\hat{\Sigma}_{X,n} - \Sigma_X\||_2 = o_p(r)$ and $\||\hat{\Sigma}_{X,n}\||_2 = O_p(r)$ as $n \to \infty$.*

(c) $P(\Omega_n) \to 1$ as $n \to \infty$.

**Proof.** (a) is a direct consequence of Lemma A5. (b) follows from [C3], [D2] and the inequalities $\||\hat{\Sigma}_{X,n} - \Sigma_X\||_2 \le r\|\hat{\Sigma}_{X,n} - \Sigma_X\|_{\ell_\infty}$ and $\||\Sigma_X\||_2 \le r\|\Sigma_X\|_{\ell_\infty}$. (c) follows from the inequality $\||\Sigma_X^{-1}(\hat{\Sigma}_{X,n} - \Sigma_X)\||_2 \le r\||\Sigma_X^{-1}\||_2\|\hat{\Sigma}_{X,n} - \Sigma_X\|_{\ell_\infty}$. $\square$

**Lemma A10.** *Under the assumptions of Proposition 3, $\lambda_n^{-2}\|\hat{\Sigma}_{Z,n} - \check{\Sigma}_{Z,n}\|_{\ell_\infty} = o_p(r)$ as $n \to \infty$.*

**Proof.** Since $\hat{\beta}_n = \hat{\Sigma}_{YX,n}\hat{\Sigma}_{X,n}^{-1}$ on the event $\Omega_n$, we have

$$
\begin{aligned}
\hat{\Sigma}_{Z,n} - \check{\Sigma}_{Z,n} &= -\hat{\beta}_n\hat{\Sigma}_{X,n}\hat{\beta}_n^\top + \hat{\Sigma}_{YX,n}\beta^\top + \beta\hat{\Sigma}_{YX,n}^\top - \beta\hat{\Sigma}_{X,n}\beta^\top \\
&= -\hat{\Sigma}_{YX,n}(\hat{\beta}_n - \beta)^\top + \beta\hat{\Sigma}_{X,n}(\hat{\beta}_n - \beta)^\top \\
&= -(\hat{\Sigma}_{YX,n} - \beta\hat{\Sigma}_{X,n})\hat{\Sigma}_{X,n}^{-1}(\hat{\Sigma}_{YX,n} - \beta\hat{\Sigma}_{X,n})^\top \quad \text{on } \Omega_n.
\end{aligned}
$$

Therefore, Lemma A6 yields

$$
\|\hat{\Sigma}_{Z,n} - \check{\Sigma}_{Z,n}\|_{\ell_\infty} \le r\||\hat{\Sigma}_{X,n}^{-1}\||_2\|\hat{\Sigma}_{YX,n} - \beta\hat{\Sigma}_{X,n}\|_{\ell_\infty}^2 \quad \text{on } \Omega_n.
$$

Now, by [C3] and Lemma A9 we have $\||\hat{\Sigma}_{X,n}^{-1}\||_2 1_{\Omega_n} = O_p(1)$, so we obtain $\lambda_n^{-2}\|\hat{\Sigma}_{Z,n} - \check{\Sigma}_{Z,n}\|_{\ell_\infty} 1_{\Omega_n} = o_p(r)$ by [D1]. Since $P(\Omega_n) \to 1$ by Lemma A9(c), we complete the proof. $\square$

**Lemma A11.** *Under the assumptions of Proposition 3, $\lambda_n^{-1}\|\hat{\Sigma}_{Z,n} - \Sigma_Z\|_{\ell_\infty} \to^p 0$ and $P(\min_{1 \le i \le d} \hat{\Sigma}_{Z,n}^{ii} > 0) \to 1$ as $n \to \infty$.*

**Proof.** The first claim immediately follows from Lemma A10 and [D1]. The second one is a consequence of the first one, Lemma A1 and [C2]. $\square$

**Proof of Proposition 3.** Set $\mathcal{E}_n := \Omega_n \cap \{\overline{\Sigma}_n \in \mathcal{S}_d^+\} \cap \{\min_{1 \le i \le d} \hat{\Sigma}_{Z,n}^{ii} > 0\}$. From Equation (0.8.5.3) in [60], we have $\hat{\Sigma}_{Z,n} \in \mathcal{S}_d^+$ on $\mathcal{E}_n$. Hence, from the proof of [34] (Lemma 1), the optimization problem in Equation (10) has the unique solution on $\mathcal{E}_n$. Since $P(\mathcal{E}_n) \to 1$ as $n \to \infty$ by [D3] and Lemmas A9 and A11, the desired result follows once we prove $\lambda_n^{-1}\|\hat{\Sigma}_{Z,n} - \Sigma_Z\|_{\ell_\infty} \to^p 0$ as $n \to \infty$ according to Proposition 1. This has already been established in Lemma A11. $\square$

*Appendix C.2. Proof of Proposition 4*

We first establish some asymptotic properties of $\hat{\beta}_n$ which are necessary for the subsequent proofs.

**Lemma A12.** *Under the assumptions of Proposition 3, we have the following results:*

(a) $\|\beta\|_{\ell_2} = O_p(\sqrt{d})$ as $n \to \infty$.
(b) $\lambda_n^{-1}\max_{1 \le i \le d}\|\hat{\beta}_n^{i\cdot} - \beta^{i\cdot}\|_{\ell_2} = o_p(\sqrt{r})$ and $\max_{1 \le i \le d}\|\hat{\beta}_n^{i\cdot}\|_{\ell_2} = O_p(\sqrt{r})$ as $n \to \infty$.
(c) $\lambda_n^{-1}\|\hat{\beta}_n - \beta\|_{\ell_2} = o_p(\sqrt{dr})$ and $\|\hat{\beta}_n\|_{\ell_2} = O_p(\sqrt{d})$ as $n \to \infty$.
(d) $\lambda_n^{-1}\||\hat{\beta}_n - \beta\||_1 = o_p(d\sqrt{r})$ and $\||\hat{\beta}_n\||_1 = O_p(d)$ as $n \to \infty$.
(e) $\lambda_n^{-1}\||\hat{\beta}_n - \beta\||_\infty = o_p(r)$ and $\||\hat{\beta}_n\||_\infty = O_p(r)$ as $n \to \infty$.

**Proof.** (a) Since $\Sigma_X - \Lambda_{\min}(\Sigma_X)E_r$ is positive semidefinite, $\beta\Sigma_X\beta^\top - \Lambda_{\min}(\Sigma_X)\beta\beta^\top$ is also positive semidefinite. Thus, $\Sigma_Y - \Lambda_{\min}(\Sigma_X)\beta\beta^\top$ is positive definite by Equation (8). This implies that $0 \le \text{tr}(\Sigma_Y - \Lambda_{\min}(\Sigma_X)\beta\beta^\top) = \text{tr}(\Sigma_Y) - \Lambda_{\min}(\Sigma_X)\|\beta\|_{\ell_2}^2$. Since $\text{tr}(\Sigma_Y) = O_p(d)$ by [C1], we obtain $\|\beta\|_{\ell_2}^2 = O_p(d)$ by [C3].

(b) By Lemma A9, on the event $\Omega_n$, we have $\hat{\beta}_n = \hat{\Sigma}_{YX,n}\hat{\Sigma}_{X,n}^{-1}$. Hence, for every $i = 1, \ldots, d$,

$$
\begin{aligned}
\|\hat{\beta}_n^{i\cdot} - \beta^{i\cdot}\|_{\ell_2} &= \|(\hat{\Sigma}_{YX,n}^{i\cdot} - \beta\hat{\Sigma}_{X,n}^{i\cdot})\hat{\Sigma}_{X,n}^{-1}\|_{\ell_2} \le \||\hat{\Sigma}_{X,n}^{-1}\||_2\|\hat{\Sigma}_{YX,n}^{i\cdot} - \beta\hat{\Sigma}_{X,n}^{i\cdot}\|_{\ell_2} \\
&\le \sqrt{r}\||\hat{\Sigma}_{X,n}^{-1}\||_2\|\hat{\Sigma}_{YX,n} - \beta\hat{\Sigma}_{X,n}\|_{\ell_\infty} \quad \text{on } \Omega_n.
\end{aligned}
$$

Since $\||\hat{\Sigma}_{X,n}^{-1}\||_2 1_{\Omega_n} = O_p(1)$ by Lemma A9, $\lambda_n^{-1} \max_{1 \le i \le d} \|\hat{\beta}_n^{i\cdot} - \beta^{i\cdot}\|_{\ell_\infty} 1_{\Omega_n} = o_p(\sqrt{r})$ by [D1]. Since $P(\Omega_n^c) \to 0$ by Lemma A9, we obtain $\lambda_n^{-1} \max_{1 \le i \le d} \|\hat{\beta}_n^{i\cdot} - \beta^{i\cdot}\|_{\ell_2} = o_p(\sqrt{r})$. Since $\max_{1 \le i \le d} \|\beta^{i\cdot}\|_{\ell_2} \le \sqrt{r}\|\beta\|_{\ell_\infty} = O(\sqrt{r})$ by [C1], we also obtain $\max_{1 \le i \le d} \|\hat{\beta}_n^{i\cdot}\|_{\ell_\infty} = O_p(\sqrt{r})$.

(c) This follows from (a)–(b) and $r\lambda_n = o_p(1)$.

(d) This is a direct consequence of (b).

(e) This follows from (b) and the Schwarz inequality. $\square$

**Proof of Proposition 4.** Since $\|A\|_{\ell_\infty} \le \||A\||_2$ for any matrix $A$, in view of Proposition 3 it suffices to prove $\lambda_n^{-1}\|\hat{\beta}_n \hat{\Sigma}_{X,n} \hat{\beta}_n^\top - \beta \Sigma_X \beta^\top\|_{\ell_\infty} = O_p(r^2)$. By Lemma A6 we have

$$
\begin{aligned}
\|\hat{\beta}_n \hat{\Sigma}_{X,n} \hat{\beta}_n^\top - \beta \Sigma_X \beta^\top\|_{\ell_\infty} \le{}& \||\hat{\Sigma}_{X,n}\||_2 \left( \max_{1 \le i \le d} \|\hat{\beta}_n^{i\cdot} - \beta^{i\cdot}\|_{\ell_2} \right) \left( \max_{1 \le i \le d} \|\hat{\beta}_n^{i\cdot}\|_{\ell_2} \right) \\
&+ \||\hat{\Sigma}_{X,n} - \Sigma_X\||_2 \left( \max_{1 \le i \le d} \|\beta^{i\cdot}\|_{\ell_2} \right) \left( \max_{1 \le i \le d} \|\hat{\beta}_n^{i\cdot}\|_{\ell_2} \right) \\
&+ \||\Sigma_X\||_2 \left( \max_{1 \le i \le d} \|\beta^{i\cdot}\|_{\ell_2} \right) \left( \max_{1 \le i \le d} \|\hat{\beta}_n^{i\cdot} - \beta^{i\cdot}\|_{\ell_2} \right).
\end{aligned}
$$

Therefore, the desired result follows from Lemmas A9, A12(b) and assumption. $\square$

*Appendix C.3. Proof of Proposition 5*

Set $\Pi := (\Sigma_X^{-1} + \beta^\top \Sigma_Z^{-1} \beta)^{-1}$ and $\hat{\Pi}_n := (\hat{\Sigma}_{X,n}^\dagger + \hat{\beta}_n^\top \hat{\Theta}_{Z,\lambda_n} \hat{\beta}_n)^{-1}$.

**Lemma A13.** *Under the assumptions of Proposition 5, we have the following results:*

(a) $\Lambda_{\min}(\beta^\top \beta)^{-1} = O(d^{-1})$ *as* $n \to \infty$.

(b) $\||\Pi\||_2 = O_p(d^{-1})$ *as* $n \to \infty$.

(c) $\lambda_n^{-1}\||\hat{\Pi}_n - \Pi\||_2 = O_p(d^{-1}(s_n + r))$ *and* $\||\hat{\Pi}_n\||_2 = O_p(d^{-1})$ *as* $n \to \infty$.

**Proof.** (a) By Lemma A3 we have $|\Lambda_{\min}(d^{-1}\beta^\top \beta) - \Lambda_{\min}(\boldsymbol{B})| \le \||d^{-1}\beta^\top \beta - \boldsymbol{B}\||_2$. Hence the desired result follows from [C6].

(b) Since $\||\Pi\||_2 = \Lambda_{\min}(\Sigma_X^{-1} + \beta^\top \Sigma_Z^{-1} \beta)^{-1}$ and $\Sigma_X^{-1}$ is positive definite, Corollary 4.3.12 in [60] and Lemma A2 yield

$$
\||\Pi\||_2 \le \Lambda_{\min}(\beta^\top \Sigma_Z^{-1} \beta)^{-1} \le \Lambda_{\min}(\beta^\top \beta)^{-1} \Lambda_{\min}(\Sigma_Z^{-1})^{-1} = \Lambda_{\min}(\beta^\top \beta)^{-1} \Lambda_{\max}(\Sigma_Z).
$$

Thus, the desired result follows from claim (a) and [C2].

(c) First, since we have

$$
\begin{aligned}
&\|\hat{\beta}_n^\top \hat{\Theta}_{Z,\lambda_n} \hat{\beta}_n - \beta^\top \Theta_Z \beta\|_2 \\
&\le \||\hat{\beta}_n - \beta\||_2 \||\hat{\Theta}_{Z,\lambda_n}\||_2 \||\hat{\beta}_n\||_2 + \||\beta\||_2 \||\hat{\Theta}_{Z,\lambda_n} - \Theta_Z\||_2 \||\hat{\beta}_n\||_2 + \||\beta\||_2 \||\Theta_Z\||_2 \||\hat{\beta}_n - \beta\||_2,
\end{aligned}
$$

Lemma A12(a) and (c) and Proposition 3 yield $\lambda_n^{-1}\||\hat{\beta}_n^\top \hat{\Theta}_{Z,\lambda_n} \hat{\beta}_n - \beta^\top \Theta_Z \beta\||_2 = O_p(ds_n)$. Combining this with Lemma A9 and (b), we obtain $\lambda_n^{-1}\||\Pi(\hat{\Pi}_n^{-1} - \Pi^{-1})\||_2 1_{\Omega_n} = O_p(s_n + r)$. Now let us set $\Omega_{n,1} := \Omega_n \cap \{\||\Pi(\hat{\Pi}_n^{-1} - \Pi^{-1})\||_2 \le 1/2\}$. Then, using (b) and Lemmas A5 and A9(c), we obtain $\lambda_n^{-1}\||\hat{\Pi}_n - \Pi\||_2 1_{\Omega_{n,1}} = O_p(d^{-1}(s_n + r))$ and $P(\Omega_{n,1}^c) \to 0$. This completes the proof. $\square$

**Proof of Proposition 5.** By Sherman–Morrison–Woodbury formula (cf. Equation (0.7.4.1) in [60]), for any $w \in \{2, \infty\}$ we have

$$
\begin{aligned}
&\||\hat{\Sigma}_{Y,\lambda_n}^{-1} - \Sigma_Y^{-1}\||_w \\
&\le \||\hat{\Theta}_{Z,\lambda_n} - \Theta_Z\||_w + \||(\hat{\Theta}_{Z,\lambda_n} - \Theta_Z)\hat{\beta}_n \hat{\Pi}_n \hat{\beta}_n^\top \hat{\Theta}_{Z,\lambda_n}\||_w + \||\Theta_Z(\hat{\beta}_n - \beta)\hat{\Pi}_n \hat{\beta}_n^\top \hat{\Theta}_{Z,\lambda_n}\||_w \\
&\quad + \||\Theta_Z \beta(\hat{\Pi}_n - \Pi)\hat{\beta}_n^\top \hat{\Theta}_{Z,\lambda_n}\||_w + \||\Theta_Z \beta \Pi(\hat{\beta}_n - \beta)^\top \hat{\Theta}_{Z,\lambda_n}\||_w + \||\Theta_Z \beta \Pi \beta^\top (\hat{\Theta}_{Z,\lambda_n} - \Theta_Z)\||_w
\end{aligned}
$$

$$=: \Delta_1 + \Delta_2 + \Delta_3 + \Delta_4 + \Delta_5 + \Delta_6.$$

Proposition 3 yields $\lambda_n^{-1}\Delta_1 = O_p(s_n)$. Moreover, noting that $\||\Theta_Z|\|_\infty = O_p(\sqrt{\mathfrak{d}_n})$ by Lemma A4 and [C2], Proposition 3 and Lemmas A12 and A13 imply that $\lambda_n^{-1}(\Delta_2 + \Delta_6) = O_p(s_n)$, $\lambda_n^{-1}(\Delta_3 + \Delta_5) = o_p(\sqrt{r})$ and $\lambda_n^{-1}\Delta_4 = O_p(s_n + r)$ when $w = 2$ and $\lambda_n^{-1}(\Delta_2 + \Delta_6) = O_p(r^{3/2}s_n\sqrt{\mathfrak{d}_n})$, $\lambda_n^{-1}\Delta_3 = o_p(r^{3/2}\mathfrak{d}_n)$, $\lambda_n^{-1}\Delta_4 = O_p(r^{3/2}(s_n + r)\mathfrak{d}_n)$ and $\lambda_n^{-1}\Delta_5 = o_p(r^2\mathfrak{d}_n)$ when $w = \infty$. This completes the proof. $\square$

*Appendix C.4. Proof of Proposition 6*

We apply Proposition 2 to $\hat{\Sigma}_{Z,n}$. From the arguments in the proof of Proposition 3, it remains to check condition Equation (5). More precisely, we need to prove

$$\lim_{n\to\infty} \sup_{y\in\mathbb{R}^m} \left| P\left( a_n \widetilde{J}_{Z,n} \operatorname{vec}\left(\hat{\Sigma}_{Z,n} - \Sigma_Z\right) \le y \right) - P\left( \widetilde{J}_{Z,n} \mathfrak{C}_n^{1/2}\zeta_n \le y \right) \right| = 0.$$

Thanks to Lemma 3.1 in [20] and Equation (12), this claim follows once we prove $\sqrt{\log(m+1)}\|a_n\widetilde{J}_{Z,n}\operatorname{vec}(\hat{\Sigma}_{Z,n} - \Sigma_Z) - a_n\widetilde{J}_{Z,n}\operatorname{vec}(\check{\Sigma}_{Z,n} - \Sigma_Z)\|_{\ell_\infty} \to 0$. Since we have

$$\|a_n\widetilde{J}_{Z,n}\operatorname{vec}(\hat{\Sigma}_{Z,n} - \Sigma_Z) - a_n\widetilde{J}_{Z,n}\operatorname{vec}(\check{\Sigma}_{Z,n} - \Sigma_Z)\|_{\ell_\infty} \le a_n\||J_n|\|_\infty \||\Theta_Z|\|_\infty^2 \|\hat{\Sigma}_{Z,n} - \check{\Sigma}_{Z,n}\|_{\ell_\infty}$$

and $\||\Theta_Z|\|_\infty = O_p(\sqrt{\mathfrak{d}_n})$ by Lemma A4, the desired result follows from Lemma A10 and assumption. $\square$

*Appendix C.5. Proof of Lemma 2 and Proposition 7*

We use the same notation as in Section C.3. By Sherman-Morisson-Woodbury formula we have

$$\begin{aligned}
\|\hat{\Sigma}_{Y,\lambda_n}^{-1} - \Theta_{Z,\lambda_n}\|_{\ell_\infty} &\le \|\hat{\Theta}_{Z,\lambda_n}\hat{\beta}_n\hat{\Pi}_n\hat{\beta}_n^\top\hat{\Theta}_{Z,\lambda_n}\|_{\ell_\infty} \\
&\le r\|\hat{\Theta}_{Z,\lambda_n}\hat{\beta}_n\|_{\ell_\infty}^2 \||\hat{\Pi}_n|\|_2 \le r\||\hat{\Theta}_{Z,\lambda_n}|\|_\infty^2 \|\hat{\beta}_n\|_{\ell_\infty}^2 \||\hat{\Pi}_n|\|_2,
\end{aligned}$$

where the second inequality follows from Lemma A6. Since $\||\Theta_Z|\|_\infty^2 = O(\mathfrak{d}_n)$ by Lemma A4, we have $\||\hat{\Theta}_{Z,\lambda_n}|\|_\infty^2 = O_p(\mathfrak{d}_n)$ by Proposition 3. We also have $\|\hat{\beta}_n\|_{\ell_\infty} = O_p(1)$ by [C1], [D2] and Lemma A12(b). Consequently, we obtain $\|\hat{\Sigma}_{Y,\lambda_n}^{-1} - \Theta_{Z,\lambda_n}\|_{\ell_\infty} = O_p(r\mathfrak{d}_n/d)$ by Lemma A13. Similarly, we can prove $\|\Sigma_Y^{-1} - \Theta_Z\|_{\ell_\infty} = O_p(r\mathfrak{d}_n/d)$. Therefore, we complete the proof of Lemma 2.

Proposition 7 is an immediate consequence of Proposition 6, Lemma 2 and [20] (Lemma 3.1). $\square$

## Appendix D. Proofs for Section 4

*Appendix D.1. Proof of Theorem 1*

The proof relies on the following concentration inequalities for discretized quadratic covariations of continuous martingales:

**Lemma A14.** *Let $M = (M_t)_{t\in[0,1]}$ and $N = (N_t)_{t\in[0,1]}$ be two continuous martingales. Suppose that there is a constant $L > 0$ such that*

$$\left|[M,M]_t - [M,M]_s\right| \vee \left|[N,N]_t - [N,N]_s\right| \le L|t - s| \tag{A5}$$

*for all $s, t \in [0,1]$. Then, for any $\theta > 0$, there is a constant $C_{L,\theta} > 0$ which depends only on $L$ and $\theta$ such that*

$$P\left( \sqrt{n}\left|\widehat{[M,N]}_1^n - [M,N]_1\right| > x \right) \le 2\exp\left(-C_{L,\theta}x^2\right)$$

*for all $n \in \mathbb{N}$ and $x \in [0, \theta\sqrt{n}]$.*

**Remark A1.** *Similar estimates to Lemma A14 have already been obtained in the literature (see e.g., [36] (Lemma 3), [4] (Lemma 10) and [8] (Lemma A.1)). Since we use slightly different assumptions from the existing ones, we give its proof in Appendix E for the shake of completeness.*

Define the $(d + r)$-dimensional semimartingale $\bar{Z} = (\bar{Z}_t)_{t \in [0,1]}$ by $\bar{Z}_t = (Z^1_t, \ldots, Z^d_t, X^1_t, \ldots, X^r_t)^\top$.

**Lemma A15.** *Assume* [E1] *and* $\log(d + r)/\sqrt{n} \to 0$ *as* $n \to \infty$. *Then,* $\left\| \widehat{[\bar{Z}, \bar{Z}]}^n_1 - [\bar{Z}, \bar{Z}]_1 \right\|_{\ell_\infty} = O_p(\sqrt{\log(d + r)/n})$ *as* $n \to \infty$.

**Proof.** For all $n, \nu \in \mathbb{N}$ and $t \in [0, 1]$, set

$$\bar{\mu}(\nu)_t = \begin{pmatrix} \mu(\nu)_t \\ \tilde{\mu}(\nu)_t \end{pmatrix}, \qquad \bar{\sigma}(\nu)_t = \begin{pmatrix} \sigma(\nu)_t \\ \tilde{\sigma}(\nu)_t \end{pmatrix}.$$

Then we define the processes $\bar{A}(\nu) = (\bar{A}(\nu)_t)_{t \in [0,1]}$ and $\bar{M}(\nu) = (\bar{M}(\nu)_t)_{t \in [0,1]}$ by $\bar{A}(\nu)_t = \int_0^t \bar{\mu}(\nu)_s ds$ and $\bar{M}(\nu)_t = \int_0^t \bar{\sigma}(\nu)_s dW_s$. By the local property of Itô integrals (cf. [47], pp. 17–18), we have $\bar{Z} = \bar{Z}(\nu) := \bar{A}(\nu) + \bar{M}(\nu)$ on $\Omega_n(\nu)$. Hence, for every $L > 0$, it holds that

$$P \left( \left\| \widehat{[\bar{Z}, \bar{Z}]}^n_1 - [\bar{Z}, \bar{Z}]_1 \right\|_{\ell_\infty} > L\sqrt{\log(d + r)/n} \right)$$

$$\leq P \left( \left\| \widehat{[\bar{Z}(\nu), \bar{Z}(\nu)]}^n_1 - [\bar{Z}(\nu), \bar{Z}(\nu)]_1 \right\|_{\ell_\infty} > L\sqrt{\log(d + r)/n} \right) + P(\Omega_n(\nu)^c).$$

Therefore, the proof is completed once we show that

$$\lim_{L \to \infty} \limsup_{n \to \infty} P \left( \left\| \widehat{[\bar{Z}(\nu), \bar{Z}(\nu)]}^n_1 - [\bar{Z}(\nu), \bar{Z}(\nu)]_1 \right\|_{\ell_\infty} > L\sqrt{\log(d + r)/n} \right) = 0$$

for any fixed $\nu > 0$. We decompose the target quantity as

$$\widehat{[\bar{Z}(\nu), \bar{Z}(\nu)]}^n_1 - [\bar{Z}(\nu), \bar{Z}(\nu)]_1$$
$$= \left( \widehat{[\bar{M}(\nu), \bar{M}(\nu)]}^n_1 - [\bar{M}(\nu), \bar{M}(\nu)]_1 \right) + \widehat{[\bar{A}(\nu), \bar{A}(\nu)]}^n_1 + \widehat{[\bar{A}(\nu), \bar{M}(\nu)]}^n_1 + \widehat{[\bar{M}(\nu), \bar{A}(\nu)]}^n_1$$
$$=: \mathbb{I}_n + \mathbb{II}_n + \mathbb{III}_n + \mathbb{IV}_n.$$

First we consider $\mathbb{I}_n$. Since we have $|[\bar{M}(\nu)^i, \bar{M}(\nu)^i]_t - [\bar{M}(\nu)^i, \bar{M}(\nu)^i]_s| \leq C_\nu |t - s|$ for all $s, t \in [0, 1]$ and $i \in \{1, \ldots, d + r\}$ by [E1], by Lemma A14 there is a constant $C > 0$ such that

$$\max_{1 \leq i,j \leq d+r} P \left( \sqrt{n} \left| \mathbb{I}^{ij}_n \right| > x \right) \leq 2e^{-Cx^2}$$

for all $n \in \mathbb{N}$ and $x \in [0, \sqrt{n}]$. Therefore, for every $L \in [0, \sqrt{n/\log(d + r)}]$ we obtain

$$P \left( \left\| \mathbb{I}_n \right\|_{\ell_\infty} > L\sqrt{\frac{\log(d + r)}{n}} \right) \leq \sum_{i,j=1}^{d+r} P \left( \sqrt{n} \left| \mathbb{I}^{ij}_n \right| > L\sqrt{\log(d + r)} \right) \leq 2(d + r)^{2-CL^2}.$$

Hence, noting the assumption $\sqrt{n}/\log(d + r) \to \infty$, we conclude that

$$\lim_{L \to \infty} \limsup_{n \to \infty} P \left( \left\| \mathbb{I}_n \right\|_{\ell_\infty} > L\sqrt{\frac{\log(d + r)}{n}} \right) = 0.$$

Next, by [E1] we have $\|\mathbb{III}_n\|_{\ell_\infty} \leq C_v^2/n$. Therefore, we obtain $\|\mathbb{III}_n\|_{\ell_\infty} = O(n^{-1}) = O(\sqrt{\log(d+r)/n})$. Third, we consider $\mathbb{III}_n$. By the Schwarz inequality we have

$$\|\mathbb{III}_n\|_{\ell_\infty} \leq \sqrt{\|\mathbb{III}_n\|_{\ell_\infty}} \max_{1 \leq j \leq d+r} \sqrt{[\bar{M}(\widehat{v})^j, \widehat{\bar{M}(v)^j}]_1^n}.$$

From the above result we have $\sqrt{\|\mathbb{III}_n\|_{\ell_\infty}} = O(1/\sqrt{n})$. Meanwhile, using the inequality $\sqrt{x} \leq \sqrt{|x-y|} + \sqrt{y}$ holding for all $x, y \geq 0$, we have

$$\max_{1 \leq j \leq d+r} \sqrt{[\bar{M}(\widehat{v})^j, \widehat{\bar{M}(v)^j}]_1^n} \leq \sqrt{\|\mathbb{I}_n\|_{\ell_\infty}} + \max_{1 \leq j \leq d+r} \sqrt{[\bar{M}(v)^j, \bar{M}(v)^j]_1} \leq \sqrt{\|\mathbb{I}_n\|_{\ell_\infty}} + \sqrt{C_v}.$$

Hence the above result yields $\max_{1 \leq j \leq d+r} \sqrt{[\bar{M}(\widehat{v})^j, \widehat{\bar{M}(v)^j}]_1^n} = O_p(1)$. Thus, we conclude that $\|\mathbb{III}_n\|_{\ell_\infty} = O_p(1/\sqrt{n}) = O_p(\sqrt{\log(d+r)/n})$. Finally, since $\|\mathbb{IV}_n\|_{\ell_\infty} = \|\mathbb{III}_n\|_{\ell_\infty}$, we complete the proof. □

**Proof of Theorem 1.** In view of Propositions 3–5, it suffices to check [D1]. Noting that $\hat{\Sigma}_{YX,n} - \beta\hat{\Sigma}_{X,n} = \widehat{[Z, X]}_1^n$ and $\check{\Sigma}_{Z,n} = \widehat{[Z, Z]}_1^n$, [D1] immediately follows from Lemma A15. □

*Appendix D.2. Proof of Theorem 2*

Our proof relies on the following "high-dimensional" asymptotic mixed normality of the realized covariance matrix:

**Lemma A16** ([20], Theorem 4.2(b)). *Assume [F1]. For every n, let $X_n$ be an $m \times d^2$ random matrix and $Y_n$ be an $m \times d^2$ non-random matrix such that $\||Y_n\||_\infty \geq 1$, where $m = m_n$ possibly depends on n. Define $\Xi_n := Y_n \circ X_n$. Suppose that for all $n, v \in \mathbb{N}$, we have $X_n(v) \in \mathbb{D}_{2,\infty}(\mathbb{R}^{m \times d^2})$ such that $X_n = X_n(v)$ on $\Omega_n(v)$ and*

$$\lim_{b \downarrow 0} \limsup_{n \to \infty} P(\min \operatorname{diag}(\Xi_n(v)\mathfrak{C}_n(v)\Xi_n(v)^\top) < b) = 0, \tag{A6}$$

$$\sup_{n \in \mathbb{N}} \max_{1 \leq i \leq m} \max_{1 \leq j \leq d^2} \left( \|X_n(v)^{ij}\|_\infty + \sup_{0 \leq t \leq 1} \|D_t X_n(v)^{ij}\|_{\infty,\ell_2} + \sup_{0 \leq s,t \leq 1} \|D_{s,t} X_n(v)^{ij}\|_{\infty,\ell_2} \right) < \infty, \tag{A7}$$

*where $\Xi_n(v) := Y_n \circ X_n(v)$. Suppose also $\||Y_n\||_\infty^5 (\log dm)^{\frac{13}{2}} \to 0$ as $n \to \infty$. Then we have*

$$\sup_{y \in \mathbb{R}^m} \left| P\left( \Xi_n \operatorname{vec}\left( \widehat{[Z, Z]}_1^n - [Z, Z]_1 \right) \leq y \right) - P(\Xi_n \mathfrak{C}_n^{1/2} \zeta_n \leq y) \right| \to 0 \tag{A8}$$

*as $n \to \infty$.*

To apply Lemma A16 to the present setting, we prove some auxiliary results.

**Lemma A17.** *Let $A_1, A_2, B_1, B_2 \in \mathbb{R}^{d \times d}$. Then $(A_1 \otimes A_2) \circ (B_1 \otimes B_2) = (A_1 \circ B_1) \otimes (A_2 \circ B_2)$.*

**Proof.** This follows from a straightforward computation. □

**Lemma A18.** *Assume [F1]. Then, for any $n, v \in \mathbb{N}$ and $t \in [0,1]$, $c(v)_t \in \mathbb{D}_{2,\infty}(\mathbb{R}^{d \times d})$, $\mathfrak{C}_n(v) \in \mathbb{D}_{2,\infty}(\mathbb{R}^{d^2 \times d^2})$ and*

$$\sup_{n \in \mathbb{N}} \max_{1 \leq i,j \leq d} \left( \sup_{0 \leq t,u \leq 1} \|D_s c(v)_t^{ij}\|_{\infty,\ell_2} + \sup_{0 \leq t,u,v \leq 1} \|D_{u,v} c(v)_t^{ij}\|_{\infty,\ell_2} \right) < \infty,$$

$$\sup_{n\in\mathbb{N}} \max_{1\leq k,l\leq d^2} \left( \sup_{0\leq u\leq 1} \|D_u\mathfrak{C}_n(\nu)^{kl}\|_{\infty,\ell_2} + \sup_{0\leq u,v\leq 1} \|D_{u,v}\mathfrak{C}_n(\nu)^{kl}\|_{\infty,\ell_2} \right) < \infty.$$

**Proof.** This directly follows from Lemmas B.11–B.12 in [20]. $\square$

**Lemma A19.** *Assume* [F1]–[F2]. *For any* $n, \nu \in \mathbb{N}$, $\Theta_Z(\nu) \in \mathbb{D}_{2,\infty}(\mathbb{R}^{d\times d})$ *and*

$$\sup_{n\in\mathbb{N}} \max_{1\leq i,j\leq d} \left( \sup_{0\leq t\leq 1} \|D_t\Theta_Z(\nu)^{ij}\|_{\infty,\ell_2} + \sup_{0\leq s,t\leq 1} \|D_{s,t}\Theta_Z(\nu)^{ij}\|_{\infty,\ell_2} \right) < \infty.$$

**Proof.** First, by Remark 15.87 in [48] and Lemma A18, $\Sigma_Z(\nu) \in \mathbb{D}_{2,\infty}(\mathbb{R}^{d\times d})$ and $D^k\Sigma_Z(\nu) = \int_0^1 D^k c(\nu)_s ds$ for $k = 1, 2$. In particular, we have

$$\sup_{n\in\mathbb{N}} \max_{1\leq i,j\leq d} \left( \sup_{0\leq t\leq 1} \|D_t\Sigma_Z(\nu)^{ij}\|_{\infty,\ell_2} + \sup_{0\leq s,t\leq 1} \|D_{s,t}\Sigma_Z(\nu)^{ij}\|_{\infty,\ell_2} \right) < \infty \tag{A9}$$

by Lemma A18 and Equation (16). Next, by Theorem 15.78 in [48] and Theorem 4 in [59] (Chapter 8), we have $\Theta_Z(\nu) \in \mathbb{D}_{2,\infty}(\mathbb{R}^{d\times d})$ with $D_s^{(a)}\Theta_Z(\nu) = -\Theta_Z(\nu)D_s^{(a)}\Sigma_Z(\nu)\Theta_Z(\nu)$ and

$$D_{s,t}^{(a,b)}\Theta_Z(\nu) = \Theta_Z(\nu)D_t^{(b)}\Sigma_Z(\nu)\Theta_Z(\nu)D_s^{(a)}\Sigma_Z(\nu)\Theta_Z(\nu)$$
$$- \Theta_Z(\nu)D_{s,t}^{(a,b)}\Sigma_Z(\nu)\Theta_Z(\nu) + \Theta_Z(\nu)D_s^{(a)}\Sigma_Z(\nu)\Theta_Z(\nu)D_t^{(b)}\Sigma_Z(\nu)\Theta_Z(\nu)$$

for all $s, t \in [0,1]$ and $a, b \in \{1, \ldots, d'\}$. Therefore, by Lemma A7 we have

$$\|D_s\Theta_Z(\nu)^{ij}\|_{\ell_2} \leq \|\|\Theta_Z(\nu)\|\|_\infty^2 \max_{1\leq k,l\leq d} \|D_s\Sigma_Z(\nu)^{kl}\|_{\ell_2}$$

for all $i, j = 1, \ldots, d$. Then, noting that $Q_Z$ is non-random, we have $(1_{\{\Theta_Z(\nu)^{ij}\neq 0\}})_{1\leq i,j\leq d} = Q_Z$ by assumption. Therefore, we obtain $\|\|\Theta_Z(\nu)\|\|_\infty \leq \|\|Q_Z\|\|_\infty \|\Theta_Z(\nu)\|_{\ell_\infty}$. Hence, Equation (A9), [F2] and Equation (17) yield $\sup_{n\in\mathbb{N}} \max_{1\leq i,j\leq d} \sup_{0\leq t\leq 1} \|D_t\Theta_Z(\nu)^{ij}\|_{\infty,\ell_2} < \infty$. In the meantime, by Lemma A7 we also have

$$\|D_{s,t}\Theta_Z(\nu)^{ij}\|_{\ell_2}$$
$$\leq 2\sqrt{\sum_{b=1}^{d'} \left( \sum_{k=1}^d \left| \left( \Theta_Z(\nu)D_t^{(b)}\Sigma_Z(\nu)\Theta_Z(\nu) \right)^{ik} \right| \right)^2} \|\|\Theta_Z(\nu)\|\|_\infty \max_{1\leq k,l\leq d} \|D_s\Sigma_Z(\nu)^{kl}\|_{\ell_2}$$
$$+ \|\|\Theta_Z(\nu)\|\|_\infty^2 \max_{1\leq k,l\leq d} \|D_{s,t}\Sigma_Z(\nu)^{kl}\|_{\ell_2}$$
$$= 2\sqrt{\sum_{b=1}^{d'} \left( \sum_{k=1}^d \left| D_t^{(b)}\Theta_Z(\nu)^{ik} \right| \right)^2} \|\|\Theta_Z(\nu)\|\|_\infty \max_{1\leq k,l\leq d} \|D_s\Sigma_Z(\nu)^{kl}\|_{\ell_2}$$
$$+ \|\|\Theta_Z(\nu)\|\|_\infty^2 \max_{1\leq k,l\leq d} \|D_{s,t}\Sigma_Z(\nu)^{kl}\|_{\ell_2}.$$

Now, since $Q_Z$ is non-random, we have $D_t^{(b)}\Theta_Z(\nu) = Q_Z \circ D_t^{(b)}\Theta_Z(\nu)$. Therefore, the Schwarz inequality yields

$$\|D_{s,t}\Theta_Z(\nu)^{ij}\|_{\ell_2}$$
$$\leq 2\sqrt{\|\|Q_Z\|\|_\infty \sum_{k=1}^d \sum_{b=1}^{d'} \left| D_t^{(b)}\Theta_Z(\nu)^{ik} \right|^2} \|\|\Theta_Z(\nu)\|\|_\infty \max_{1\leq k,l\leq d} \|D_s\Sigma_Z(\nu)^{kl}\|_{\ell_2}$$

$$+ \|\|\Theta_Z(\nu)\|\|_\infty^2 \max_{1 \le k,l \le d} \|D_{s,t}\Sigma_Z(\nu)^{kl}\|_{\ell_2}$$

$$\le 2\|\|Q_Z\|\|_\infty \left( \max_{1 \le k,l \le d} \|D_t\Theta_Z(\nu)^{kl}\|_{\ell_2} \right) \|\|\Theta_Z(\nu)\|\|_\infty \max_{1 \le k,l \le d} \|D_s\Sigma_Z(\nu)^{kl}\|_{\ell_2}$$

$$+ \|\|\Theta_Z(\nu)\|\|_\infty^2 \max_{1 \le k,l \le d} \|D_{s,t}\Sigma_Z(\nu)^{kl}\|_{\ell_2}.$$

Consequently, we conclude $\sup_{n \in \mathbb{N}} \max_{1 \le i,j \le d} \sup_{0 \le s,t \le 1} \|D_{s,t}\Theta_Z(\nu)^{ij}\|_{\infty,\ell_2} < \infty$ by [F2], Equation (17) and the results proved above. $\square$

**Lemma A20.** *Assume* [F1]–[F2]. *For any* $n, \nu \in \mathbb{N}$, $\Theta_Z(\nu) \otimes \Theta_Z(\nu), \mathfrak{V}_n(\nu) \in \mathbb{D}_{2,\infty}(\mathbb{R}^{d^2 \times d^2})$ *and*

$$\sup_{n \in \mathbb{N}} \max_{1 \le i,j \le d^2} \left( \sup_{0 \le t \le 1} \|D_t\{\Theta_Z(\nu) \otimes \Theta_Z(\nu)\}^{ij}\|_{\infty,\ell_2} + \sup_{0 \le s,t \le 1} \|D_{s,t}\{\Theta_Z(\nu) \otimes \Theta_Z(\nu)\}^{ij}\|_{\infty,\ell_2} \right) < \infty, \quad \text{(A10)}$$

$$\sup_{n \in \mathbb{N}} \max_{1 \le i,j \le d^2} \left( \sup_{0 \le t \le 1} \|D_t\mathfrak{V}_n(\nu)^{ij}\|_{\infty,\ell_2} + \sup_{0 \le s,t \le 1} \|D_{s,t}\mathfrak{V}_n(\nu)^{ij}\|_{\infty,\ell_2} \right) < \infty. \quad \text{(A11)}$$

**Proof.** First, Corollary 15.80 in [48], Equation (17) and Lemma A19 imply that $\Theta_Z(\nu) \otimes \Theta_Z(\nu) \in \mathbb{D}_{2,\infty}(\mathbb{R}^{2d^2 \times d^2})$ and Equation (A10) holds true. Next, Corollary 15.80 in [48] and Lemma A18 imply that $\mathfrak{V}_n(\nu) \in \mathbb{D}_{2,\infty}(\mathbb{R}^{d^2 \times d^2})$ and

$$D_s^{(a)}\mathfrak{V}_n(\nu) = \{D_s^{(a)}\Theta_Z^{\otimes 2}(\nu)\}\mathfrak{C}_n(\nu)\Theta_Z^{\otimes 2}(\nu)$$
$$+ \Theta_Z^{\otimes 2}(\nu)\{D_s^{(a)}\mathfrak{C}_n(\nu)\}\Theta_Z^{\otimes 2}(\nu) + \Theta_Z^{\otimes 2}(\nu)\mathfrak{C}_n(\nu)\{D_s^{(a)}\Theta_Z^{\otimes 2}(\nu)\}$$

and

$$D_{s,t}^{(a,b)}\mathfrak{V}_n(\nu) = \{D_{s,t}^{(a,b)}\Theta_Z^{\otimes 2}(\nu)\}\mathfrak{C}_n(\nu)\Theta_Z^{\otimes 2}(\nu) + \{D_s^{(a)}\Theta_Z^{\otimes 2}(\nu)\}\{D_t^{(b)}\mathfrak{C}_n(\nu)\}\Theta_Z^{\otimes 2}(\nu)$$
$$+ \{D_s^{(a)}\Theta_Z^{\otimes 2}(\nu)\}\mathfrak{C}_n(\nu)\{D_t^{(b)}\Theta_Z^{\otimes 2}(\nu)\} + \{D_s^{(b)}\Theta_Z^{\otimes 2}(\nu)\}\{D_s^{(a)}\mathfrak{C}_n(\nu)\}\Theta_Z^{\otimes 2}(\nu)$$
$$+ \Theta_Z^{\otimes 2}(\nu)\{D_{s,t}^{(a,b)}\mathfrak{C}_n(\nu)\}\Theta_Z^{\otimes 2}(\nu) + \Theta_Z^{\otimes 2}(\nu)\{D_s^{(a)}\mathfrak{C}_n(\nu)\}\{D_t^{(b)}\Theta_Z^{\otimes 2}(\nu)\}$$
$$+ \{D_t^{(b)}\Theta_Z^{\otimes 2}(\nu)\}\mathfrak{C}_n(\nu)\{D_s^{(a)}\Theta_Z^{\otimes 2}(\nu)\} + \Theta_Z^{\otimes 2}(\nu)\{D_t^{(b)}\mathfrak{C}_n(\nu)\}\{D_s^{(a)}\Theta_Z^{\otimes 2}(\nu)\}$$
$$+ \Theta_Z^{\otimes 2}(\nu)\mathfrak{C}_n(\nu)\{D_{s,t}^{(a,b)}\Theta_Z^{\otimes 2}(\nu)\}$$

for any $s, t \in [0,1]$, where $\Theta_Z^{\otimes 2}(\nu) := \Theta_Z(\nu) \otimes \Theta_Z(\nu)$. Thus, by Lemma A7 we obtain

$$\max_{1 \le i,j \le d^2} \|D_s\mathfrak{V}_n(\nu)^{ij}\|_{\ell_2} \le 2 \max_{1 \le i \le d^2} \sqrt{\sum_{a=1}^{d'} \left( \sum_{k=1}^{d^2} \left| D_s^{(a)}\Theta_Z^{\otimes 2}(\nu)^{ik} \right| \right)^2} \|\mathfrak{C}_n(\nu)\|_{\ell_\infty} \|\|\Theta_Z^{\otimes 2}(\nu)\|\|_\infty$$
$$+ \|\|\Theta_Z^{\otimes 2}(\nu)\|\|_\infty^2 \max_{1 \le i,j \le d^2} \|D_s\mathfrak{C}_n(\nu)^{ij}\|_{\ell_2}$$

and

$$\max_{1 \le i,j \le d^2} \|D_{s,t}\mathfrak{V}_n(\nu)^{ij}\|_{\ell_2}$$

$$\le 2 \max_{1 \le i \le d^2} \sup_{0 \le s,t \le 1} \sqrt{\sum_{a,b=1}^{d'} \left( \sum_{k=1}^{d^2} \left| D_{s,t}^{(a,b)}\Theta_Z^{\otimes 2}(\nu)^{ik} \right| \right)^2} \|\mathfrak{C}_n(\nu)\|_{\ell_\infty} \|\|\Theta_Z^{\otimes 2}(\nu)\|\|_\infty$$

$$+ 4 \max_{1 \le i,j,l \le d^2} \sup_{0 \le s,t \le 1} \sqrt{\sum_{a=1}^{d'} \left( \sum_{k=1}^{d^2} \left| D_s^{(a)}\Theta_Z^{\otimes 2}(\nu)^{ik} \right| \right)^2} \|D_t\mathfrak{C}_n(\nu)^{jl}\|_{\ell_2} \|\|\Theta_Z^{\otimes 2}(\nu)\|\|_\infty$$

$$+ 2 \max_{1 \le i \le d^2} \sup_{0 \le s \le 1} \|\mathfrak{C}_n(\nu)\|_{\ell_\infty} \sum_{a=1}^{d'} \left( \sum_{k=1}^{d^2} \left| D_s^{(a)}\Theta_Z^{\otimes 2}(\nu)^{ik} \right| \right)^2$$

$$+ \max_{1\leq i,j\leq d^2} \sup_{0\leq s,t\leq 1} \||\Theta_Z^{\otimes 2}(\nu)|\|_\infty^2 \|D_{s,t}\mathfrak{C}_n(\nu)^{ij}\|_{\ell_2}.$$

Now, as pointed out in the proof of Lemma A19, we have $\Theta_Z(\nu) = Q_Z \circ \Theta_Z(\nu)$. Therefore, Lemma A17 yields $\Theta_Z^{\otimes 2}(\nu) = (Q_Z \otimes Q_Z) \circ \Theta_Z^{\otimes 2}(\nu)$. Since $Q_Z$ is non-random by [F2], we have $D_s^{(a)}\Theta_Z^{\otimes 2}(\nu) = (Q_Z \otimes Q_Z) \circ D_s^{(a)}\Theta_Z^{\otimes 2}(\nu)$. Thus, using the Schwarz inequality repeatedly, we obtain

$$\max_{1\leq i,j\leq d^2} \|D_s\mathfrak{V}_n(\nu)^{ij}\|_{\ell_2} \leq 2 \max_{1\leq ij\leq d^2} \||Q_Z|\|_\infty^2 \|D_s\Theta_Z^{\otimes 2}(\nu)^{ij}\|_{\ell_2} \|\mathfrak{C}_n(\nu)\|_{\ell_\infty} \||\Theta_Z^{\otimes 2}(\nu)|\|_\infty$$
$$+ \||\Theta_Z^{\otimes 2}(\nu)|\|_\infty^2 \max_{1\leq i,j\leq d^2} \|D_s\mathfrak{C}_n(\nu)^{ij}\|_{\ell_2}$$

and

$$\max_{1\leq i,j\leq d^2} \|D_{s,t}\mathfrak{V}_n(\nu)^{ij}\|_{\ell_2}$$
$$\leq 2 \max_{1\leq i,j\leq d^2} \sup_{0\leq s,t\leq 1} \||Q_Z|\|_\infty^2 \|D_{s,t}\Theta_Z^{\otimes 2}(\nu)^{ij}\|_{\ell_2} \|\mathfrak{C}_n(\nu)\|_{\ell_\infty} \||\Theta_Z^{\otimes 2}(\nu)|\|_\infty$$
$$+ 4 \max_{1\leq i,j,k,l\leq d^2} \sup_{0\leq s,t\leq 1} \||Q_Z|\|_\infty^2 \|D_s\Theta_Z^{\otimes 2}(\nu)^{ik}\|_{\ell_2} \|D_t\mathfrak{C}_n(\nu)^{jl}\|_{\ell_2} \||\Theta_Z^{\otimes 2}(\nu)|\|_\infty$$
$$+ 2 \max_{1\leq i,j\leq d^2} \sup_{0\leq s\leq 1} \|\mathfrak{C}_n(\nu)\|_{\ell_\infty} \||Q_Z|\|_\infty^2 \|D_s\Theta_Z^{\otimes 2}(\nu)^{ij}\|_{\ell_2}$$
$$+ \max_{1\leq i,j\leq d^2} \sup_{0\leq s,t\leq 1} \||\Theta_Z^{\otimes 2}(\nu)|\|_\infty^2 \|D_{s,t}\mathfrak{C}_n(\nu)^{ij}\|_{\ell_2}.$$

Hence we complete the proof by Lemma A18, Equation (A10) and assumption. □

**Proof of Theorem 2.** Set $U_n := \sqrt{n}\,\mathrm{vec}(\hat{\Theta}_{Z,\lambda_n} - \Gamma_{Z,n} - \Theta_Z)$. Define the $2d^2 \times d^2$ matrices $J_{n,1}$ and $J_{n,2}$ by

$$J_{n,1} = \begin{pmatrix} \mathsf{E}_{d^2} \\ -\mathsf{E}_{d^2} \end{pmatrix}, \qquad J_{n,2} = \begin{pmatrix} \mathfrak{S}_n^{-1} \\ -\mathfrak{S}_n^{-1} \end{pmatrix}.$$

Then we have

$$\sup_{A\in\mathcal{A}^{\mathrm{re}}(d^2)} \left| P\left(U_n \in A\right) - P\left(\mathfrak{V}_n^{1/2}\zeta_n \in A\right) \right| = \sup_{y\in\mathbb{R}^{2d^2}} \left| P\left(J_{n,1}U_n \leq y\right) - P\left(J_{n,1}\mathfrak{V}_n^{1/2}\zeta_n \leq y\right) \right|$$

and

$$\sup_{A\in\mathcal{A}^{\mathrm{re}}(d^2)} \left| P\left(\mathfrak{S}_n^{-1}U_n \in A\right) - P\left(\mathfrak{S}_n^{-1}\mathfrak{V}_n^{1/2}\zeta_n \in A\right) \right| = \sup_{y\in\mathbb{R}^{2d^2}} \left| P\left(J_{n,2}U_n \leq y\right) - P\left(J_{n,2}\mathfrak{V}_n^{1/2}\zeta_n \leq y\right) \right|.$$

Therefore, in view of Proposition 6, it suffices to check [D1] and Equations (12)–(13) for $J_n \in \{J_{n,1}, J_{n,2}\}$. We have already checked [D1] in the proof of Theorem 1. Meanwhile, Equation (13) immediately follows from [E1] and Equation (17). To check Equation (12), we apply Lemma A16 with $\Xi_n = -J_n(\Theta_Z \otimes \Theta_Z)$ (note that $\check{\Sigma}_{Z,n} = \widehat{[Z,Z]}_1^n$). Set

$$\Upsilon_n = \begin{pmatrix} Q_Z \otimes Q_Z \\ Q_Z \otimes Q_Z \end{pmatrix}.$$

Then we have $\Xi_n = \Upsilon_n \circ \Xi_n$ by Lemma A17. Since $\Upsilon_n$ is non-random by [F2], we can apply Lemma A16 with $X_n = \Xi_n$ once we show that for every $\nu \in \mathbb{N}$, there is an $X_n(\nu) \in \mathbb{D}_{2,\infty}(\mathbb{R}^{m\times d^2})$ such that $X_n = X_n(\nu)$ on $\Omega_n(\nu)$ and Equations (A6)–(A7) hold true. Now we separately consider the two cases.

Case 1: $J_n = J_{n,1}$. In this case, we set $X_n(\nu) := -J_{n,1}(\Theta_Z(\nu) \otimes \Theta_Z(\nu))$. By [E1] we have $X_n = X_n(\nu)$ on $\Omega_n(\nu)$, while Equations (A6)–(A7) follow from Equation (17) and Lemma A20, respectively.

Case 2: $J_n = J_{n,2}$. In this case, we set $X_n(\nu) := -J_{n,2}(\nu)(\Theta_Z(\nu) \otimes \Theta_Z(\nu))$, where

$$J_{n,2}(\nu) = \begin{pmatrix} \mathfrak{S}_n(\nu)^{-1} \\ -\mathfrak{S}_n(\nu)^{-1} \end{pmatrix}.$$

By [E1] we have $X_n = X_n(\nu)$ on $\Omega_n(\nu)$, while Equation (A6) is evident because $\Xi_n(\nu)\mathfrak{C}_n(\nu)\Xi_n(\nu)^\top$ is the identity matrix in this case. Therefore, it remains to prove Equation (A7). Noting that $\mathfrak{S}_n(\nu)$ is a diagonal matrix, Equation (A7) follows from Corollary 15.80 in [48] and Lemma A20 once we show that $\mathfrak{S}_n(\nu)^{kk} \in \mathbb{D}_{2,\infty}$ for every $k = 1, \ldots, d^2$ and

$$\sup_{n \in \mathbb{N}} \max_{1 \leq k \leq d^2} \left( \|\mathfrak{S}_n(\nu)^{kk}\|_\infty + \sup_{0 \leq t \leq 1} \|D_t \mathfrak{S}_n(\nu)^{kk}\|_{\infty,\ell_2} + \sup_{0 \leq s,t \leq 1} \|D_{s,t} \mathfrak{S}_n(\nu)^{kk}\|_{\infty,\ell_2} \right) < \infty.$$

Since we can write $\mathfrak{S}_n(\nu)^{kk} = (\mathfrak{V}_n(\nu)^{kk})^{5/2}(\mathfrak{V}_n(\nu)^{kk})^{-3}$, we obtain the desired result by combining Theorem 15.78 and Lemma 15.152 in [48] with Lemma A20. □

*Appendix D.3. Proof of Lemma 3*

We use the following notation: For a $d$-dimensional process $U = (U_t)_{t \in [0,1]}$, we set $\Delta_h^n U := U_{h/n} - U_{(h-1)/n}$, $h = 1, \ldots, n$. Also, we set $\chi_h := \text{vec}[\Delta_h^n Z (\Delta_h^n Z)^\top]$ for $h = 1, \ldots, n$ and

$$\widetilde{\mathfrak{C}}_n := n \sum_{h=1}^n \chi_h \chi_h^\top - \frac{n}{2} \sum_{h=1}^{n-1} \left( \chi_h \chi_{h+1}^\top + \chi_{h+1} \chi_h^\top \right).$$

**Lemma A21.** *Assume* [E1]. *Then* $\sum_{h=1}^n (\|\Delta_h^n Z\|_{\ell_\infty}^4 + \|\Delta_h^n X\|_{\ell_\infty}^4) = O_p(\log^2(d+r)/n)$ *as* $n \to \infty$.

**Proof.** We use the same notation as in the proof of Lemma A15. Then, we need to prove $\sum_{h=1}^n \|\Delta_h^n \bar{Z}\|_{\ell_\infty}^4 = O_p(\log^2(d+r)/n)$ as $n \to \infty$. For every $\nu \in \mathbb{N}$ and $L > 0$, we have

$$P\left( \sum_{h=1}^n \|\Delta_h^n \bar{Z}\|_{\ell_\infty}^4 > L \right) \leq P\left( \sum_{h=1}^n \|\Delta_h^n \bar{Z}(\nu)\|_{\ell_\infty}^4 > L \right) + P(\Omega_n(\nu)^c).$$

Hence it suffices to prove $\sum_{h=1}^n \|\Delta_h^n \bar{Z}(\nu)\|_{\ell_\infty}^4 = O_p(\log^2(d+r)/n)$ as $n \to \infty$ for any fixed $\nu \in \mathbb{N}$. By Lemma A23 there is a universal constant $\mathsf{c} > 0$ such that $\|\Delta_h^n \bar{M}(\nu)^j\|_p \leq \mathsf{c}\sqrt{p}\|\sqrt{\Delta_h^n[\bar{M}(\nu)^j, \bar{M}(\nu)^j]}\|_p$ for all $p \geq 2$. Thus, by [E1] we obtain $\|\Delta_h^n \bar{Z}(\nu)^j\|_p \leq C_\nu/n + \mathsf{c}\sqrt{C_\nu}\sqrt{p/n}$. Therefore, by [63] (Proposition 2.5.2), there is a constant $C' > 0$ such that $\max_{j,h} \|\Delta_h^n \bar{Z}(\nu)^j\|_{\psi_2} \leq C'/\sqrt{n}$ for all $n$, where $\|\xi\|_{\psi_2} := \inf\{\Lambda > 0 : E[\exp(|\xi|/\Lambda)] \leq 2\}$ for a random variable $\xi$. Thus, [64] (Lemma 2.2.2) implies that there is a constant $C'' > 0$ such that $\max_h \|\|\Delta_h^n \bar{Z}(\nu)\|_{\ell_\infty}\|_{\psi_2} \leq C''\sqrt{\log(d+r)/n}$ for all $n$. Thus, we obtain

$$E\left[ \sum_{h=1}^n \|\Delta_h^n \bar{Z}(\nu)\|_{\ell_\infty}^4 \right] \leq 4!^4 C'' \frac{\log^2(d+r)}{n},$$

so the desired result follows from the Markov inequality. □

**Lemma A22.** *Assume* [C1]–[C4] *and* [E1]. *Then* $\sum_{h=1}^n \|\hat{\chi}_h - \chi_h\|_{\ell_\infty}^2 = O_p(r^2(\log d)^3/n^2)$ *as* $n \to \infty$.

**Proof.** Since $\hat{Z}_{h/n} = Z_{h/n} - (\hat{\beta}_n - \beta)X_{h/n}$, we have

$$\begin{aligned} \hat{\chi}_h - \chi_h = &-\text{vec}[(\hat{\beta}_n - \beta)\Delta_h^n X (\Delta_h^n Z)^\top] - \text{vec}[\Delta_h^n Z ((\hat{\beta}_n - \beta)\Delta_h^n X)^\top] \\ &+ \text{vec}[(\hat{\beta}_n - \beta)\Delta_h^n X ((\hat{\beta}_n - \beta)\Delta_h^n X)^\top]. \end{aligned}$$

Now, since $\|\text{vec}(xy^\top)\|_{\ell_\infty} \le \|x\|_{\ell_\infty}\|y\|_{\ell_\infty}$ for any $x, y \in \mathbb{R}^d$, it holds that

$$\|\hat{\chi}_h - \chi_h\|_{\ell_\infty} \le 2\|(\hat{\beta}_n - \beta)\Delta_h^n X\|_{\ell_\infty}\|\Delta_h^n Z\|_{\ell_\infty} + \|(\hat{\beta}_n - \beta)\Delta_h^n X\|_{\ell_\infty}^2$$
$$\le 2\|\|\hat{\beta}_n - \beta\|\|_\infty \|\Delta_h^n X\|_{\ell_\infty}\|\Delta_h^n Z\|_{\ell_\infty} + \|\|\hat{\beta}_n - \beta\|\|_\infty^2 \|\Delta_h^n X\|_{\ell_\infty}^2.$$

Therefore, we obtain

$$\sum_{h=1}^n \|\hat{\chi}_h - \chi_h\|_{\ell_\infty}^2 \le 2\|\|\hat{\beta}_n - \beta\|\|_\infty^2 \sum_{h=1}^n (\|\Delta_h^n X\|_{\ell_\infty}^4 + \|\Delta_h^n Z\|_{\ell_\infty}^4) + 2\|\|\hat{\beta}_n - \beta\|\|_\infty^4 \sum_{h=1}^n \|\Delta_h^n X\|_{\ell_\infty}^4.$$

Now, noting Lemma A15, we infer that $\|\|\hat{\beta}_n - \beta\|\|_\infty = O_p(r\sqrt{(\log d)/n})$ from the proof of Lemma A12(e). Thus, we complete the proof by Lemma A21. □

**Proof of Lemma 3.** Since $\|\widetilde{\mathfrak{C}}_n - \mathfrak{C}_n\|_{\ell_\infty} = O_p((\log d)^2/\sqrt{n} + n^{-\gamma})$ by Proposition 4.1 in [20], it suffices to prove $\|\hat{\mathfrak{C}}_n - \widetilde{\mathfrak{C}}_n\|_{\ell_\infty} = O_p(r(\log d)^{5/2}/\sqrt{n})$. Since $\|\text{vec}(xy^\top)\|_{\ell_\infty} \le \|x\|_{\ell_\infty}\|y\|_{\ell_\infty}$ for any $x, y \in \mathbb{R}^d$, Lemma A21 yields $\sum_{h=1}^n \|\chi_h\|_{\ell_\infty}^2 = O_p((\log d)^2/n)$. Combining this with Lemma A22 and $r^2(\log d)/n = O(1)$, we also obtain $\sum_{h=1}^n \|\hat{\chi}_h\|_{\ell_\infty}^2 = O_p((\log d)^2/n)$. Now the desired result follows from the Schwarz inequality and Lemma A22. □

*Appendix D.4. Proof of Corollary 1*

(a) Since $\|\|\Theta_Z\|\|_\infty = O_p(1)$ by Equation (17) and [F2], we have $\|\mathfrak{C}_n\|_{\ell_\infty} + \|\mathfrak{V}_n\|_{\ell_\infty} = O_p(1)$ by [E1] and $\lambda_n^{-1}\|\|\hat{\Theta}_{Z,\lambda_n} \otimes \hat{\Theta}_{Z,\lambda_n} - \Theta_Z \otimes \Theta_Z\|\|_\infty = O_p(s_n)$ by Theorem 1. Combining this with Lemma 3 and assumption, we obtain $\|\hat{\mathfrak{V}}_n\|_{\ell_\infty} = O_p(1)$ and $(\log d)\|\hat{\mathfrak{V}}_n - \mathfrak{V}_n\|_{\ell_\infty} \to^p 0$. Noting Equation (17) and the fact that $\hat{\mathfrak{S}}_n$ is a diagonal matrix, we also obtain $\|\|\hat{\mathfrak{S}}_n^{-1}\|\|_\infty = O_p(1)$ and $(\log d)\|\|\hat{\mathfrak{S}}_n^{-1} - \mathfrak{S}_n^{-1}\|\|_\infty \to^p 0$. Since Equation (18) yields $\|\sqrt{n}\,\text{vec}(\hat{\Theta}_{Z,\lambda_n} - \Gamma_{Z,n} - \Theta_Z)\|_{\ell_\infty} = O_p(\sqrt{\log d})$, we obtain $\sqrt{\log d}\|\sqrt{n}(\hat{\mathfrak{S}}_n^{-1} - \mathfrak{S}_n)\,\text{vec}(\hat{\Theta}_{Z,\lambda_n} - \Gamma_{Z,n} - \Theta_Z)\|_{\ell_\infty} \to^p 0$. Now the desired result follows from Theorem 2 and [20] Lemma 3.1.

(b) The same argument as above implies that $(\log d)^2\|\hat{\mathfrak{V}}_n - \mathfrak{V}_n\|_{\ell_\infty} \to^p 0$ and $(\log d)^2\|\|\hat{\mathfrak{S}}_n^{-1} - \mathfrak{S}_n^{-1}\|\|_\infty \to^p 0$. Thus, the desired result follows from [20] Proposition 3.1. □

*Appendix D.5. Proof of Corollary 2*

First, we have by Corollary 1(a)

$$\limsup_{n\to\infty} P\left(\max_{1\le i<j\le d} \frac{\sqrt{n}|\hat{\Theta}_{Z,\lambda_n}^{ij} - \Gamma_{Z,n}^{ij} - \Theta_Z^{ij}|}{\hat{\mathfrak{s}}_n^{ij}} > \mathfrak{c}_n\right)$$
$$= \limsup_{n\to\infty} P\left(\max_{1\le i<j\le d} |(\mathfrak{S}_n^{-1}\mathfrak{V}_n^{1/2}\zeta_n)^{(i-1)d+j}| > \mathfrak{c}_n\right)$$
$$\le \limsup_{n\to\infty} \frac{d(d-1)}{2} \cdot 2(1 - \Phi(\mathfrak{c}_n)) = \limsup_{n\to\infty} \alpha_n = \alpha.$$

Hence we have

$$\limsup_{n\to\infty} P(\hat{S}(\Theta_Z) \not\subset S(\Theta_Z)) \le \limsup_{n\to\infty} P\left(\max_{(i,j):\Theta_Z^{ij}=0} \frac{\sqrt{n}|\hat{\Theta}_{Z,\lambda_n}^{ij} - \Gamma_{Z,n}^{ij}|}{\hat{\mathfrak{s}}_n^{ij}} > \mathfrak{c}_n\right)$$
$$\le \limsup_{n\to\infty} P\left(\max_{1\le i<j\le d} \frac{\sqrt{n}|\hat{\Theta}_{Z,\lambda_n}^{ij} - \Gamma_{Z,n}^{ij} - \Theta_Z^{ij}|}{\hat{\mathfrak{s}}_n^{ij}} > \mathfrak{c}_n\right) \le \alpha$$

and

$$\limsup_{n\to\infty} P(\hat{S}(\Theta_Z) \not\supseteq S(\Theta_Z))$$

$$\leq \limsup_{n\to\infty} P\left(\min_{(i,j)\in S(\Theta_Z)} \frac{\sqrt{n}|\hat{\Theta}^{ij}_{Z,\lambda_n} - \Gamma^{ij}_{Z,n}|}{\hat{\mathfrak{s}}^{ij}_n} \leq \mathfrak{c}_n\right)$$

$$\leq \limsup_{n\to\infty} P\left(\min_{(i,j)\in S(\Theta_Z)} \frac{\sqrt{n}|\Theta^{ij}_Z|}{\hat{\mathfrak{s}}^{ij}_n} \leq \max_{1\leq i<j\leq d} \frac{\sqrt{n}|\hat{\Theta}^{ij}_{Z,\lambda_n} - \Gamma^{ij}_{Z,n} - \Theta^{ij}_Z|}{\hat{\mathfrak{s}}^{ij}_n} + \mathfrak{c}_n\right)$$

$$\leq \limsup_{n\to\infty} P\left(\min_{(i,j)\in S(\Theta_Z)} \frac{\sqrt{n}|\Theta^{ij}_Z|}{\hat{\mathfrak{s}}^{ij}_n} \leq 2\mathfrak{c}_n\right) + \limsup_{n\to\infty} P\left(\max_{1\leq i<j\leq d} \frac{\sqrt{n}|\hat{\Theta}^{ij}_{Z,\lambda_n} - \Gamma^{ij}_{Z,n} - \Theta^{ij}_Z|}{\hat{\mathfrak{s}}^{ij}_n} > \mathfrak{c}_n\right)$$

$$\leq \limsup_{n\to\infty} P\left(\min_{(i,j)\in S(\Theta_Z)} |\Theta^{ij}_Z| \leq \frac{2\mathfrak{c}_n}{\sqrt{n}} \max_{i,j} \hat{\mathfrak{s}}^{ij}_n\right) + \alpha.$$

Now, we have $\max_{i,j} \hat{\mathfrak{s}}^{ij}_n = O_p(1)$ from the proof of Corollary 1(a). Moreover, since $\frac{\alpha_n}{d(d-1)/2} = 1 - \Phi(\mathfrak{c}_n) \leq e^{-\mathfrak{c}_n^2/2}$ by Chernoff's inequality, we have $\mathfrak{c}_n \leq \sqrt{-2\log \frac{\alpha_n}{d(d-1)/2}}$. Hence $\mathfrak{c}_n = O(\sqrt{\log d})$ as $n \to \infty$ by assumption. Consequently, we obtain

$$\limsup_{n\to\infty} P\left(\min_{(i,j)\in S(\Theta_Z)} |\Theta^{ij}_Z| \leq \frac{2\mathfrak{c}_n}{\sqrt{n}} \max_{i,j} \hat{\mathfrak{s}}^{ij}_n\right) = 0$$

by assumption. This completes the proof.

## Appendix E. Proof of Lemma A14

In this appendix we prove Lemma A14 with the help of two general martingale inequalities. The first one is the Burkholder-Davis-Gundy inequality with a sharp constant:

**Lemma A23** (Barlow and Yor [65], Proposition 4.2)**.** *There is a universal constant* $\mathsf{c} > 0$ *such that*

$$\left\|\sup_{0\leq t\leq T} |M_t|\right\|_p \leq \mathsf{c}\sqrt{p} \left\|[M,M]^{1/2}_T\right\|_p$$

*for any* $p \in [2,\infty)$ *and any continuous martingale* $M = (M_t)_{t\in[0,T]}$ *with* $M_0 = 0$.

The second one is a Bernstein-type inequality for martingales:

**Lemma A24.** *Let* $(\xi_i)^n_{i=1}$ *be a martingale difference sequence with respect to the filtration* $(\mathcal{G}_i)^n_{i=0}$*. Suppose that there are constants* $a, b > 0$ *such that* $\sum^n_{i=1} \mathrm{E}[|\xi_i|^k \mid \mathcal{G}_{i-1}] \leq k! a^{k-2} b^2/2$ *a.s. for any integer* $k \geq 2$*. Then, for any* $x \geq 0$,

$$P\left(\max_{1\leq m\leq n} \left|\sum^m_{i=1} \xi_i\right| \geq x\right) \leq 2\exp\left(-\frac{x^2}{b^2 + b\sqrt{b^2 + 2ax}}\right).$$

**Proof.** This is a special case of Pinelis [66] (Theorem 3.3). In fact, since $\mathbb{R}$ is a Hilbert space, we can apply this result with $\mathcal{X} = \mathbb{R}$ and $D = 1$ in the notation of that paper. $\square$

**Proof of Lemma A14.** For every $h = 1, \ldots, n$, set

$$\xi_{n,h} := \sqrt{n}\left\{\int^{t_h}_{t_{h-1}} (M_t - M_{t_{h-1}})dN_t + \int^{t_h}_{t_{h-1}} (N_t - N_{t_{h-1}})dM_t\right\}.$$

Itô's formula yields

$$\sqrt{n}\left(\widehat{[M,N]}_1^n - [M,N]_1\right) = \sum_{h=1}^n \xi_{n,h}.$$

Also, by assumption $(\xi_{n,h})_{h=1}^n$ is a martingale difference with respect to $(\mathcal{F}_{t_h})_{h=0}^n$. Moreover, for any integer $k \geq 2$, we have

$$E[|\xi_{n,h}|^k \mid \mathcal{F}_{t_{h-1}}]$$

$$\leq 2^{k-1}n^{k/2}\,E\left[\left|\int_{t_{h-1}}^{t_h}(M_t - M_{t_{h-1}})dN_t\right|^k + \left|\int_{t_{h-1}}^{t_h}(N_t - N_{t_{h-1}})dM_t\right|^k \mid \mathcal{F}_{t_{h-1}}\right]$$

$$\leq 2^{k-1}n^{k/2}\mathsf{c}^k k^{k/2}\,E\left[\left(\int_{t_{h-1}}^{t_h}(M_t - M_{t_{h-1}})^2 d[N,N]_t\right)^{k/2} + \left(\int_{t_{h-1}}^{t_h}(N_t - N_{t_{h-1}})^2 d[M,M]_t\right)^{k/2} \mid \mathcal{F}_{t_{h-1}}\right]$$

$$(\because \text{Lemma A23})$$

$$\leq 2^{k-1}\mathsf{c}^k k^{k/2}L^{k/2}\,E\left[\sup_{t_{h-1}<t\leq t_h}|M_t - M_{t_{h-1}}|^k + \sup_{t_{h-1}<t\leq t_h}|N_t - N_{t_{h-1}}|^k \mid \mathcal{F}_{t_{h-1}}\right] \quad (\because \text{(A5)})$$

$$\leq 2^{k-1}\mathsf{c}^{2k}k^k L^{k/2}\,E\left[([M,M]_{t_h} - [M,M]_{t_{h-1}})^{k/2} + ([N,N]_{t_h} - [N,N]_{t_{h-1}})^{k/2} \mid \mathcal{F}_{t_{h-1}}\right] \quad (\because \text{Lemma A23})$$

$$\leq 2^k \mathsf{c}^{2k}k^k \frac{L^k}{n^{k/2}} \quad (\because \text{(A5)}),$$

where $\mathsf{c} > 0$ is the universal constant appearing in Lemma A23. Thus, using Stirling's formula, we obtain

$$\sum_{h=1}^n E[|\xi_{n,h}|^k \mid \mathcal{F}_{t_{h-1}}] \leq 2^k \mathsf{c}^{2k}\frac{e^k}{\sqrt{2\pi k}}k!\frac{L^k}{n^{k/2-1}} \leq \frac{k!}{2}\left(\frac{a_0}{\sqrt{n}}\right)^{k-2}b_0^2,$$

where $a_0 := 2e\mathsf{c}^2 L$ and $b_0 := 2\sqrt{2}\mathsf{c}^2 Le/(2\pi)^{1/4}$. Hence, Lemma A24 yields

$$P\left(\left|\sum_{h=1}^n \xi_{n,h}\right| \geq x\right) \leq 2\exp\left(-\frac{x^2}{b_0^2 + b_0\sqrt{b_0^2 + 2(a_0/\sqrt{n})x}}\right)$$

for every $x \geq 0$. Consequently, when $x \in [0, \theta\sqrt{n}]$ for some $\theta > 0$, we have

$$P\left(\left|\sum_{h=1}^n \xi_{n,h}\right| \geq x\right) \leq 2\exp\left(-C_{L,\theta}x^2\right)$$

with $C_{L,\theta} := (b_0^2 + b_0\sqrt{b_0^2 + 2a_0\theta})^{-1}$. This completes the proof. $\quad\square$

## References

1. Wang, Y.; Zou, J. Vast volatility matrix estimation for high-frequency financial data. *Ann. Statist.* **2010**, *38*, 943–978. [CrossRef]
2. Bickel, P.J.; Levina, E. Covariance regularization by thresholding. *Ann. Statist.* **2008**, *36*, 2577–2604. [CrossRef]
3. Bickel, P.J.; Levina, E. Regularized estimation of large covariance matrices. *Ann. Statist.* **2008**, *36*, 199–227. [CrossRef]
4. Tao, M.; Wang, Y.; Zhou, H. Optimal sparse volatility matrix estimation for high-dimensional Itô processes with measurement errors. *Ann. Statist.* **2013**, *41*, 1816–1864. [CrossRef]
5. Tao, M.; Wang, Y.; Chen, X. Fast convergence rates in estimating large volatility matrices using high-frequency financial data. *Econom. Theory* **2013**, *29*, 838–856. [CrossRef]
6. Kim, D.; Wang, Y.; Zou, J. Asymptotic theory for large volatility matrix estimation based on high-frequency financial data. *Stoch. Process. Appl.* **2016**, *126*, 3527–3577. [CrossRef]

7. Kim, D.; Kong, X.B.; Li, C.X.; Wang, Y. Adaptive thresholding for large volatility matrix estimation based on high-frequency financial data. *J. Econom.* **2018**, *203*, 69–79. [CrossRef]

8. Fan, J.; Furger, A.; Xiu, D. Incorporating global industrial classification standard into portfolio allocation: A simple factor-based large covariance matrix estimator with high-frequency data. *J. Bus. Econom. Statist.* **2016**, *34*, 489–503. [CrossRef]

9. Aït-Sahalia, Y.; Xiu, D. Using principal component analysis to estimate a high dimensional factor model with high-frequency data. *J. Econom.* **2017**, *201*, 384–399. [CrossRef]

10. Fan, J.; Kim, D. Robust high-dimensional volatility matrix estimation for high-frequency factor model. *J. Am. Statist. Assoc.* **2018**, *113*, 1268–1283. [CrossRef]

11. Dai, C.; Lu, K.; Xiu, D. Knowing factors or factor loadings, or neither? Evaluating estimators of large covariance matrices with noisy and asynchronous data. *J. Econom.* **2019**, *208*, 43–79. [CrossRef]

12. Hautsch, N.; Kyj, L.M.; Oomen, R.C. A blocking and regularization approach to high-dimensional realized covariance estimation. *J. Appl. Econom.* **2012**, *27*, 625–645. [CrossRef]

13. Morimoto, T.; Nagata, S. Robust estimation of a high-dimensional integrated covariance matrix. *Commun. Statist. Simul. Comput.* **2017**, *46*, 1102–1112. [CrossRef]

14. Lam, C.; Feng, P.; Hu, C. Nonlinear shrinkage estimation of large integrated covariance matrices. *Biometrika* **2017**, *104*, 481–488. [CrossRef]

15. Ledoit, O.; Wolf, M. Nonlinear shrinkage estimation of large-dimensional covariance matrices. *Ann. Statist.* **2012**, *40*, 1024–1060. [CrossRef]

16. Zheng, X.; Li, Y. On the estimation of integrated covariance matrices of high dimensional diffusion processes. *Ann. Statist.* **2011**, *39*, 3121–3151. [CrossRef]

17. Brownlees, C.; Nualart, E.; Sun, Y. Realized networks. *J. Appl. Econom.* **2018**, *33*, 986–1006. [CrossRef]

18. Kong, X.B.; Liu, C. Testing against constant factor loading matrix with large panel high-frequency data. *J. Econom.* **2018**, *204*, 301–319. [CrossRef]

19. Pelger, M. Large-dimensional factor modeling based on high-frequency observations. *J. Econom.* **2019**, *208*, 23–42. [CrossRef]

20. Koike, Y. Mixed-normal limit theorems for multiple Skorohod integrals in high-dimensions, with application to realized covariance. *Electron. J. Stat.* **2019**, *13*, 1443–1522. [CrossRef]

21. Cochrane, J.H. *Asset Pricing*, revised ed.; Princeton University Press: Princeton, NJ, USA, 2005.

22. Bühlmann, P.; van de Geer, S. *Statistics for High-Dimensional Data*; Springer: Berlin/Heidelberg, Germany, 2011.

23. Acemoglu, D.; Carvalho, V.M.; Ozdaglar, A.; Tahbaz-Salehi, A. The network origins of aggregate fluctuations. *Econometrica* **2012**, *80*, 1977–2016. [CrossRef]

24. Janková, J.; van de Geer, S. Inference in high-dimensional graphical models. In *Handbook of Graphical Models*; CRC Press: Boca Raton, FL, USA, 2018; Chapter 14, pp. 325–351.

25. Janková, J.; van de Geer, S. Confidence intervals for high-dimensional inverse covariance estimation. *Electron. J. Stat.* **2015**, *9*, 1205–1229. [CrossRef]

26. Podolskij, M.; Vetter, M. Understanding limit theorems for semimartingales: A short survey. *Stat. Neerl.* **2010**, *64*, 329–351. [CrossRef]

27. De Gregorio, A.; Iacus, S.M. Adaptive LASSO-type estimation for multivariate diffusion processes. *Econom. Theory* **2012**, *28*, 838–860. [CrossRef]

28. Masuda, H.; Shimizu, Y. Moment convergence in regularized estimation under multiple and mixed-rates asymptotics. *Math. Methods Statist.* **2017**, *26*, 81–110. [CrossRef]

29. Kinoshita, Y.; Yoshida, N. Penalized quasi likelihood estimation for variable selection. *arXiv* **2019**, arXiv:1910.12871.

30. Suzuki, T.; Yoshida, N. Penalized least squares approximation methods and their applications to stochastic processes. *Jpn. J. Stat. Data Sci.* **2020**, forthcoming. [CrossRef]

31. Fujimori, K. The Dantzig selector for a linear model of diffusion processes. *Stat. Inference Stoch. Process.* **2019**, *22*, 475–498. [CrossRef]

32. Gaïffas, S.; Matulewicz, G. Sparse inference of the drift of a high-dimensional Ornstein–Uhlenbeck process. *J. Multivar. Anal.* **2019**, *169*, 1–20. [CrossRef]

33. Koike, Y.; Yoshida, N. Covariance estimation and quasi-likelihood analysis. In *Financial Mathematics, Volatility and Covariance Modelling*; Chevallier, J., Goutte, S., Guerreiro, D., Saglio, S., Sanhaji, B., Eds.; Routledge: London, UK, 2019; Volume 2, Chapter 12, pp. 308–335.

34. Duchi, J.; Gould, S.; Koller, D. Projected subgradient methods for learning sparse Gaussians. In Proceedings of the Twenty-Fourth Conference on Uncertainty in Artificial Intelligence, Helsinki, Finland, 9–12 July 2008; AUAI Press: Arlington, VA, USA, 2008; pp. 153–160.

35. Rothman, A.J.; Bickel, P.J.; Levina, E.; Zhu, J. Sparse permutation invariant covariance estimation. *Electron. J. Stat.* **2008**, *2*, 494–515. [CrossRef]

36. Fan, J.; Li, Y.; Yu, K. Vast volatility matrix estimation using high-frequency data for portfolio selection. *J. Am. Statist. Assoc.* **2012**, *107*, 412–428. [CrossRef] [PubMed]

37. Kim, D.; Wang, Y. Sparse PCA-based on high-dimensional Itô processes with measurement errors. *J. Multivar. Anal.* **2016**, *152*, 172–189. [CrossRef]

38. Fan, J.; Li, R. Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc.* **2001**, *96*, 1348–1360. [CrossRef]

39. Ren, Z.; Sun, T.; Zhang, C.H.; Zhou, H.H. Asymptotic normality and optimalities in estimation of large Gaussian graphical models. *Ann. Statist.* **2015**, *43*, 991–1026. [CrossRef]

40. Chang, J.; Qiu, Y.; Yao, Q.; Zou, T. Confidence regions for entries of a large precision matrix. *J. Econom.* **2018**, *206*, 57–82. [CrossRef]

41. Campbell, J.Y.; Lo, A.W.; MacKinlay, A.C. *The Econometrics of Financial Markets*; Princeton University Press: Princeton, NJ, USA, 1997.

42. Barigozzi, M.; Brownlees, C.; Lugosi, G. Power-law partial correlation network models. *Electron. J. Stat.* **2018**, *12*, 2905–2929. [CrossRef]

43. Reiß, M.; Todorov, V.; Tauchen, G. Nonparametric test for a constant beta between Itô semi-martingales based on high-frequency data. *Stoch. Process. Appl.* **2015**, *125*, 2955–2988. [CrossRef]

44. Fan, J.; Liao, Y.; Mincheva, M. High-dimensional covariance matrix estimation in approximate factor models. *Ann. Statist.* **2011**, *39*, 3320–3356. [CrossRef]

45. Cai, T.T.; Liu, W.; Zhou, H.H. Estimating sparse precision matrix: Optimal rates of convergence and adaptive estimation. *Ann. Statist.* **2016**, *44*, 455–488. [CrossRef]

46. Cai, T.T.; Ren, Z.; Zhou, H.H. Estimating structured high-dimensional covariance and precision matrices: Optimal rates and adaptive estimation. *Electron. J. Stat.* **2016**, *10*, 1–59. [CrossRef]

47. Nualart, D. *The Malliavin Calculus and Related Topics*, 2nd ed.; Springer: Berlin/Heidelberg, Germany, 2006.

48. Janson, S. *Gaussian Hilbert Space*; Cambridge University Press: Cambridge, UK, 1997.

49. Clément, E.; Gloter, A. Limit theorems in the Fourier transform method for the estimation of multivariate volatility. *Stoch. Process. Appl.* **2011**, *121*, 1097–1124. [CrossRef]

50. Christensen, K.; Podolskij, M.; Thamrongrat, N.; Veliyev, B. Inference from high-frequency data: A subsampling approach. *J. Econom.* **2017**, *197*, 245–272. [CrossRef]

51. Belloni, A.; Chernozhukov, V.; Chetverikov, D.; Hansen, C.; Kato, K. High-dimensional econometrics and regularized GMM. *arXiv* **2018**, arXiv:1806.01888.

52. Lam, C.; Fan, J. Sparsistency and rates of convergence in large covariance matrix estimation. *Ann. Statist.* **2009**, *37*, 4254–4278. [CrossRef] [PubMed]

53. Sustik, M.A.; Calderhead, B. *GLASSOFAST: An efficient GLASSO Implementation*; UTCS Technical Report TR-12-29; The University of Texas at Austin: Austin, TX, USA, 2012.

54. Friedman, J.; Hastie, T.; Tibshirani, R. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* **2008**, *9*, 432–441. [CrossRef] [PubMed]

55. Yuan, M.; Lin, Y. Model selection and estimation in the Gaussian graphical model. *Biometrika* **2007**, *94*, 19–35. [CrossRef]

56. Friedman, J.; Hastie, T.; Tibshirani, R. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **2010**, *33*, 1. [CrossRef]

57. Witten, D.M.; Friedman, J.H.; Simon, N. New insights and faster computations for the graphical lasso. *J. Comput. Graph. Statist.* **2011**, *20*, 892–900. [CrossRef]

58. Hoyle, D.C. Accuracy of pseudo-inverse covariance learning—A random matrix theory analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **2011**, *33*, 1470–1481. [CrossRef]

59. Magnus, J.R.; Neudecker, H. *Matrix Differential Calculus with Applications in Statistics and Econometrics*; Wiley: New York, NY, USA, 1988.

60. Horn, R.A.; Johnson, C.R. *Matrix Analysis*, 2nd ed.; Cambridge University Press: Cambridge, UK, 2013.

61. Ogihara, T. Parametric inference for nonsynchronously observed diffusion processes in the presence of market microstructure noise. *Bernoulli* **2018**, *24*, 3318–3383. [CrossRef]

62. Lange, K. *Optimization*, 2nd ed.; Springer: Berlin/Heidelberg, Germany, 2013.

63. Vershynin, R. *High-Dimensional Probability*; Cambridge University Press: Cambridge, UK, 2018.

64. Van der Vaart, A.W.; Wellner, J.A. *Weak Convergence and Empirical Processes*; Springer: Berlin/Heidelberg, Germany, 1996.

65. Barlow, M.T.; Yor, M. Semi-martingale inequalities via the Garsia–Rodemich–Rumsey lemma and application to local times. *J. Funct. Anal.* **1982**, *49*, 198–229. [CrossRef]

66. Pinelis, I. Optimum bounds for the distributions of martingales in Banach spaces. *Ann. Probab.* **1994**, *22*, 1679–1706. [CrossRef]