LARGE COVARIANCE/CORRELATION MATRIX ESTIMATION FOR LONG MEMORY TEMPORAL DATA

By Hai Shu* and Bin Nan*,†

University of Michigan

We consider the estimation of high-dimensional covariance and correlation matrices under slow-decaying temporal dependence. For generalized thresholding estimators, convergence rates are obtained and properties of sparsistency and sign-consistency are established. The impact of temporal dependence on convergence rates is also investigated. An intuitive cross-validation method is proposed for the thresholding parameter selection, which shows good performance in simulations. Convergence rates are also obtained for banding method if the covariance or correlation matrix is bandable. The considered temporal dependence has longer memory than those in the current literature and has particular implications in analyzing resting-state fMRI data for brain connectivity studies.

1. Introduction. Suppose $X_1, \ldots, X_n$ are $n$ temporally observed $p$-dimensional random vectors with mean $EX_i = \mu_p$ and covariance matrix $Var(X_i) = \Sigma$. When $p$ is fixed and $X_i, i = 1, \ldots, n$, are independent and identically distributed (i.i.d.) random vectors, the sample covariance matrix

$$
\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^{n} X_i X_i^T - \bar{X} \bar{X}^T,
$$

where $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$, is a consistent estimator of $\Sigma$

When $p$ grows with $n$, however, the sample covariance matrix (1) is inconsistent [Bai and Yin (1993)]. To overcome the inconsistency issue for i.i.d. data, several sample covariance matrix based regularization methods have been proposed, including banding [Bickel and Levina (2008a)], tapering [Cai, Zhang and Zhou (2010)], thresholding [El Karoui (2008); Bickel and Levina (2008b); Rothman, Levina and Zhu (2009); Cai and Liu (2011)] and block-thresholding [Cai and Yuan (2012)]. Alternative regularized approaches include Cholesky-based method [Huang et al. (2006); Rothman, Huang et al. (2009a); Cai and Yuan (2012)], and various types of shrinkage methods [Cai and Yuan (2012)].

Keywords and phrases: Large covariance matrix, large correlation matrix, consistency, generalized thresholding, temporal dependence, long memory process, functional connectivity, brain image.

*Partly supported by NIH grant R01-AG036802.
†Partly supported by NSF grants DMS-1007590 and DMS-1407142.
Levina and Zhu (2010)], penalized pseudo-likelihood method [Lam and Fan (2009)], and sparse matrix transform [Cao, Bachega and Bouman (2011)]. Consistent correlation matrix estimates can be obtained similarly from i.i.d. data [Jiang (2003); El Karoui (2008)].

Recently, researchers become interested in the large covariance or correlation matrix estimation from temporally dependent data. Such research is particularly useful in analyzing the functional magnetic resonance imaging (fMRI) data to assess the brain functional connectivity [Power et al. (2011)], where the number of brain nodes (voxels or regions of interest) \( p \) can be much bigger than the number of temporally dependent images \( n \). The temporal dependence for stochastic processes or time series is traditionally dealt with by imposing the so-called strong mixing conditions [Bradley (2005)]. To overcome the difficulties in computing strong mixing coefficients and verifying strong mixing conditions, Wu (2005) introduced a new type of dependence measure, the functional dependence measure, and recently applied it to the hard thresholding estimator of large covariance matrix [Chen, Xu and Wu (2013)]. But the functional dependence measure is still difficult to understand and to interpret. It is straightforward to describe the temporal dependence directly by using cross-covariance or cross-correlation [Brockwell and Davis (1991)]. By imposing certain weak dependence conditions directly on the cross-covariance matrix of samples \( X_i, i = 1, \ldots, n \), Bhattacharjee and Bose (2014) extended the banding and tapering regularization methods for covariance matrix. Following Bhattacharjee and Bose (2014), we consider a family of cross-covariance or cross-correlation matrix with weaker conditions that better describe the resting state fMRI (rfMRI) data for brain connectivity studies.

A univariate stationary process is called a long memory process if its autocorrelation \( \rho(t) \sim C t^{-\alpha} \) as \( t \to \infty \) with constants \( C \neq 0 \) and \( \alpha > 0 \) [Brockwell and Davis (1991)]. The notation \( x_t \sim y_t \) means that \( x_t/y_t \to 1 \) as \( t \to \infty \). This power rate of decay for long memory processes is much slower than the exponential rate. An important example of long memory processes is the stationary and invertible autoregressive fractionally integrated moving average model with \( 0 < \alpha < 2 \) [Hosking (1981)]. We use a generalized form of such a power decaying structure to the cross-covariance or cross-correlation matrix of multivariate time series, and simply call it long memory temporal dependence. The weak dependence decaying as temporal distance in Bhattacharjee and Bose (2014) does not cover the long memory processes when \( 0 < \alpha \leq 3 \), and the short-range temporal dependence assumption of Chen, Xu and Wu (2013) excludes the case with \( 0 < \alpha \leq 1 \). Later we argue that the rfMRI data do not meet their restrictive temporal dependence.
conditions well, but satisfy our long memory dependence that allows any \( \alpha > 0 \) (see Figure 1(a)).

Note that the estimation of large correlation matrix was not considered by either Chen, Xu and Wu (2013) or Bhattacharjee and Bose (2014), which is a more interesting problem in, for example, the aforementioned study of brain functional connectivity. Moreover, both Chen, Xu and Wu (2013) and Bhattacharjee and Bose (2014) assumed that the mean \( \mu_p = (\mu_{pi})_{1 \leq i \leq p} \) is known, hence its estimating was not needed. But \( \mu_p \) is often unknown in practice and need to be estimated. Although the sample mean \( \bar{X}_i = \frac{1}{n} \sum_{j=1}^{n} X_{ij} \) entry-wise converges to \( \mu_{pi} \) in probability or even almost surely under some dependence conditions [Hamilton (1994), pp. 186-189; Hu, Rosalsky and Volodin (2008)], extra care will still be needed when true mean is replaced by sample mean in the estimation of covariance. We consider unknown \( \mu_p \) in this article.

Also note that the estimation of large correlation matrix of \( X_i \) is considered in a recent work by Zhou (2014). However, her method requires that the \( p \) time series have the same temporal decaying rate, which is rather restrictive and often violated (see Figure 1(b) for an example of the rfMRI data).

In this article, we mainly focus on the generalized thresholding method because it does not require the spatial bandable structure of the covariance or correlation matrix, which rarely holds for 3D brain image after vectorization. An intuitive cross-validation method is proposed for thresholding parameter selection. We show that generalized thresholding for data with long memory dependence keeps the sparsistency and sign-consistency originally developed for i.i.d. data [(Rothman, Levina and Zhu (2009))]. Consistency results of banding estimation are also provided for bandable matrices.

The article is organized as follows. In Section 2, we introduce a general model for temporal dependence which includes the long memory temporal dependence as a special case, and argue that the considered long memory temporal dependence best describes the rfMRI data comparing to the dependence conditions considered in the current literature [Chen, Xu and Wu (2013); Bhattacharjee and Bose (2014); Zhou (2014)]. In Section 3, we first provide the main theoretical results for the proposed general model, which can be applied to a broad range of temporal dependence structures, then apply the general results to the long memory temporal dependence. For generalized thresholding estimator, an intuitive cross-validation method is proposed for choosing the thresholding parameter in Section 4, and its performance is evaluated by simulations in Section 5. Section 6 contains the proofs of the theoretical results.
2. Temporal dependence. We start with a brief introduction of useful notation. For any real matrix $M = (M_{ij})_{p \times n}$, we consider the following norms [see e.g. Golub and Van Loan (1996)]:

$$\|M\|_2 = \sqrt{\lambda_{\text{max}}(M^T M)} \quad \text{(spectral norm)},$$

where $\lambda_{\text{max}}$ denotes the largest eigenvalue;

$$\|M\|_F = \sqrt{\sum_{i=1}^{p} \sum_{j=1}^{n} M^2_{ij}} \quad \text{(Frobenius norm)};$$

$$\|M\|_1 = \max_{1 \leq j \leq n} \sum_{i=1}^{p} |M_{ij}|;$$

and

$$\|M\|_{\text{max}} = \max_{1 \leq i \leq p} |M_{ij}|.$$

We have the inequality

$$\|M\|_{\text{max}}^2 \leq \|M\|_2^2 \leq \|M\|_F^2 \leq n \|M\|_1 \|M\|_{\text{max}}.$$

If $M$ is a symmetric matrix, define $\lambda_{\text{max}}(M) = \lambda_1(M) \geq \cdots \geq \lambda_p(M) = \lambda_{\text{min}}(M)$. Then we have

$$\|M\|_2 = \max_{1 \leq i \leq p} |\lambda_i(M)|,$$

and

$$\|M\|_2 \leq \|M\|_1.$$

We write $x_n \asymp y_n$ if there exists some constant $C > 1$ such that $C^{-1} \leq \liminf y_n/x_n \leq \limsup y_n/x_n \leq C$. Define $\lfloor x \rfloor$ and $\lceil x \rceil$ to be the smallest integer $\geq x$ and the largest integer $\leq x$, respectively. Denote $X \overset{d}{=} Y$ if $X$ and $Y$ have the same distribution. Let $\mathbf{1}(A)$ be the indicator function of event $A$, $(x)_+ = x1(x \geq 0)$, and $\text{sign}(x) = 1(x \geq 0) - 1(x < 0)$. Define $\text{vec}(M) = \text{vec}\{M_{ij} : 1 \leq i \leq p, 1 \leq j \leq n\} = (M_{11}^T, M_{12}^T, \ldots, M_{n1}^T)^T$, where $M_j, 1 \leq j \leq n$, are the columns of $M$. Denote $p_n = \max\{p, n\}$. If without further clarification, a constant is independent of $n$ and $p$. We allow $p \to \infty$ but require $n^{-1} \log p = o(1)$ (see Lemma 2), thus we only use $n$ to be the asymptotic leading parameter in this article for simplicity.
2.1. A general model for temporal dependence. Let \( X_{p \times n} = (X_{ij})_{p \times n} \) be the data matrix, \( \Sigma = (\sigma_{ij})_{p \times p} = (\text{Cov}(X_{ik}, X_{jk}))_{p \times p} \) be the covariance matrix that is the same for all \( 1 \leq k \leq n \), and \( R = (\rho_{ij})_{p \times p} \) be the corresponding correlation matrix. Let \( X_1, \ldots, X_n \) be the columns of \( X_{p \times n} \). The cross-covariance between \( X_i \) and \( X_j \) is defined as

\[
\text{Cov}(X_i, X_j) = \left( \text{Cov}(X_{ik}, X_{jk}) \right)_{p \times p} = \left( \theta_{ij}^{kl} \sigma_{kl} \right)_{p \times p} = \Theta_{ij}^{kl} \cdot \Sigma,
\]

where * is the entrywise product of two matrices and \( \Theta_{ij}^{kl} = (\theta_{ij}^{kl})_{p \times p} \). It is easy to see that \( (\theta_{kk}^{ij})_{1 \leq i, j \leq n} \) are the autocorrelations of time series \( \{X_{kt}\}_{t=1}^{n} \), and \( (\theta_{kl}^{ij})_{1 \leq i, j \leq n}, k \neq l \), indicates the decaying rate of the temporal dependence between \( \{X_{ki}\}_{i=1}^{n} \) and \( \{X_{lj}\}_{j=1}^{n} \). Thus, we call \( (\theta_{kk}^{ij})_{1 \leq i, j \leq n} \) the within-rate and \( (\theta_{kl}^{ij})_{1 \leq i, j \leq n}, k \neq l \), the cross-rate.

Let \( \text{vec}(X_{p \times n}) = (X_1^T, X_2^T, \ldots, X_n^T)^T \sim \mathcal{N}_{np}(I_n \otimes \mu_p, \Delta) \), where \( I_n = (1, 1, \ldots, 1)^T \) with size \( n \), \( \otimes \) is the Kronecker product, \( \mu_p = E(X_i) \), and the \( np \times np \) matrix

\[
\Delta = \begin{bmatrix}
\Sigma & \Theta^{12} \cdot \Sigma & \Theta^{13} \cdot \Sigma & \cdots & \Theta^{1n} \cdot \Sigma \\
\Theta^{21} \cdot \Sigma & \Sigma & \Theta^{23} \cdot \Sigma & \cdots & \Theta^{2n} \cdot \Sigma \\
\Theta^{31} \cdot \Sigma & \Theta^{32} \cdot \Sigma & \Sigma & \cdots & \Theta^{3n} \cdot \Sigma \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\Theta^{n1} \cdot \Sigma & \Theta^{n2} \cdot \Sigma & \Theta^{n3} \cdot \Sigma & \cdots & \Sigma
\end{bmatrix}.
\]

Define the following \( np \times np \) rate matrix

\[
\Gamma = \begin{bmatrix}
\mathbf{J}_{p \times p} & \Theta^{12} & \Theta^{13} & \cdots & \Theta^{1n} \\
\Theta^{21} & \mathbf{J}_{p \times p} & \Theta^{23} & \cdots & \Theta^{2n} \\
\Theta^{31} & \Theta^{32} & \mathbf{J}_{p \times p} & \cdots & \Theta^{3n} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\Theta^{n1} & \Theta^{n2} & \Theta^{n3} & \cdots & \mathbf{J}_{p \times p}
\end{bmatrix},
\]

where \( \mathbf{J}_{p \times p} \) is the \( p \times p \) matrix of all ones. Then we consider the following general model for temporal dependence

\[\mathcal{A}(f(n, p), g(n, p)) = \left\{ \Gamma : \max_{1 \leq j \leq n} \sum_{i \in \{1 \leq i \leq n : \ |i-j|=k \cdot f, \ k=1, \ldots, \lfloor n/f \rfloor \}} \| \Theta_{ij} \|_{\max} \leq g(n, p) \right\},\]
where \( f(n,p) \) is a positive integer-valued function satisfying \( 1 \leq f \leq n \). For notational simplicity we sometimes drop the dependence of \( f, g \) on \( n,p \) without causing any confusion.

### 2.2. long memory temporal dependence

We call \( X_{p \times n} \) with long memory temporal dependence if its rate matrix satisfies

\[
\Gamma \in \mathcal{B}(C, \alpha) = \{ \Gamma : \| \Theta^{ij} \|_{\max} \leq C|i-j|^{-\alpha} \text{ for any } i \neq j \}
\]

with some constants \( C, \alpha > 0 \). In the following we show that such a dependence structure belongs to the general dependence family given in Subsection 2.1.

Let

\[
\Psi = \max_{1 \leq j \leq n} \sum_{i \in \{1 \leq i \leq n: |i-j|=kf, k=1, ..., \lfloor n/f \rfloor \}} \| \Theta^{ij} \|_{\max}.
\]

For \( X_{p \times n} \) with the long memory temporal dependence, we have for any \( 1 \leq f \leq n \),

\[
\Psi/(2C) \leq \sum_{i=1}^{\lfloor n/f \rfloor} \left( 1 + \int_{1}^{\lfloor n/f \rfloor} y^{-\alpha} \, dy \right) / f^\alpha
\]

\[
= \left\{ \begin{array}{ll}
  \left( 1 + \frac{\lfloor n/f \rfloor - 1}{1-\alpha} \right) / f^\alpha & \leq \frac{(n/f)^{1-\alpha} - \alpha}{(1-\alpha)f^\alpha}, \quad \alpha \neq 1, \\
  \left( 1 + \log \frac{\lfloor n/f \rfloor}{1} \right) / f \leq f^{-1}(1 + \log(n/f)), \quad \alpha = 1.
\end{array} \right.
\]

Let

\[
g = 2C \times \left\{ \begin{array}{ll}
  \frac{(n/f)^{1-\alpha} - \alpha}{(1-\alpha)f^\alpha}, & \alpha \neq 1, \\
  f^{-1}(1 + \log(n/f)), & \alpha = 1.
\end{array} \right.
\]

Thus, we have shown there exist \( f \) and \( g \) such that \( \mathcal{B}(C, \alpha) \subset \mathcal{A}(f,g) \).

### 2.3. Comparison to existing models

For banding and tapering estimators, Bhattacharjee and Bose (2014) considered following weak dependence based on temporal distance: for any \( n \geq 1 \),

\[
\Gamma \in \mathcal{A}_n(a_n) = \left\{ \Gamma : \max_{a_n \leq |i-j| \leq n} \| \Theta^{ij} \|_{\max} = O(n^{-2}a_n) \right\},
\]
where \(a_n \sqrt{n^{-1} \log p} = o(1)\) and \(\{a_n\}_{n \geq 1}\) is a non-decreasing sequence of non-negative integers. Here \(a_n \sqrt{n^{-1} \log p} = o(1)\) implies \(a_n = o(n^{1/2})\). Thus,

\[
\|\Theta^{ij} : |i - j| = a_n\|_{\text{max}} \leq \max_{a_n \leq |i - j| \leq n} \|\Theta^{ij}\|_{\text{max}} = O(n^{-2} a_n) = o(a_n^{-3}).
\]

Hence we see that for \(0 < \alpha \leq 3\) and \(\|\Theta^{ij}\|_{\text{max}} \asymp |i - j|^{-\alpha}\) for any \(i \neq j\), \(\Gamma \notin A_n(a_n)\).

Chen, Xu and Wu (2013) considered hard thresholding estimators of the covariance matrix using the functional dependence measure of Wu (2005). Assume \(\{X_{1t}\}\), the first row of \(X_{p \times n}\), is a stationary process with autocovariance \(\gamma_1(t)\), and suppose \(EX_{1t} = 0\) following their consideration, then \(\gamma_1(t) = EX_{1,0}X_{1t}\). By the argument in the proof of Theorem 1 in Wu and Pourahmadi (2009) together with Lyapunov’s inequality [Karr (1993)] and Theorem 1 of Wu (2005), one can obtain that their model requires \(\sum_{t=0}^{\infty} \gamma_1(t) < \infty\). However, when \(0 < \alpha \leq 1\) and \(|\gamma_1(t)| \sim C t^{-\alpha}\) for a constant \(C > 0\), \(\sum_{t=0}^{\infty} |\gamma_1(t)| = \infty\). Hence the model considered by Chen, Xu and Wu (2013) does not apply to \(B(C, \alpha)\) when \(0 < \alpha \leq 1\).

Zhou (2014) was interested in estimating a separable \(\Delta = \Phi \otimes \Sigma\), where \(\Phi = (\phi_{ij})_{n \times n}\) is an \(n \times n\) matrix. It is easy to see that \(\theta_{kl}^{ij} = \phi_{ij}\), indicating a restrictive model with homogeneous decaying rate among all \(p\) time series. Besides, her method can only estimate \(\Sigma\) up to a scaling factor because \(\eta \Phi \otimes \frac{1}{\eta} \Sigma = \Phi \otimes \Sigma\) for any constant \(\eta > 0\).

Now consider an example of the ICA-FIX denoised rfMRI data of a single subject obtained from the WU-Minn Human Connectome Project database (www.humanconnectome.org). The rfMRI data has been preprocessed [Glasser et al. (2013); Smith et al. (2013)], resulting in corrected and cleaned images. The data consist of 1,200 timepoint images and 229,404 brain voxels of voxel-size 2 \(\times\) 2 \(\times\) 2 mm\(^3\). We discard the first 10 images because of early non-steady magnetization. The functional brain nodes are defined using the grid-based method [Sripada et al. (2014a,b); Watanabe et al. (2014)], with 907 nodes placed in a regular three-dimensional grid spaced at 12 mm intervals throughout the brain. Each node consists of a 3-mm voxel-center-to-voxel-center radius pseudosphere, which encompasses 19 voxels. The time series for each of the nodes are the spatially averaged time series of the 19 voxels within the node. All node time series pass the Priestley-Subba Rao test for stationarity [Priestley and Subba Rao (1969)] with significance level 0.05 for \(p\)-values adjusted by the false discovery rate (FDR) controlling procedure of Benjamini and Yekutieli (2001). Hence the within-rates or autocorrelations \(\theta_{kk}^{ij}\) can be approximated by sample autocorrelations \(\hat{\rho}_k(t)\). For computational simplicity, we only check the autocorrelations. One may make a mild assumption that the cross-rates are dominated...
Fig 1. Sample autocorrelations of brain nodes
by the within-rates in the sense that $|\theta_{ij}^{kl}| \leq C|\theta_{kk}^{ij}|$ for a constant $C > 0$, thus only need to check the within-rates in practice.

Figure 1(a) shows that $\max_{1 \leq i \leq p} |\hat{\rho}_i(t)|$ can be bounded by $10^8 t^{-3}$, but not by $10^7 t^{-3}$. Thus the weak temporal dependence of Bhattacharjee and Bose (2014) do not seem to fit the data well. For a randomly selected brain node, the least squares fitting for a log-linear model yields $|\hat{\rho}_1(t)| = 0.26 t^{-0.50}$, thus the applicability of Chen, Xu and Wu (2013) is in question. Figure 1(b) illustrates the estimated autocorrelations for two randomly selected brain nodes, which clearly have different patterns, indicating that the assumption of homogeneous decaying-rates for all time series in Zhou (2014) does not hold. Under the mild assumption that the cross-rates are dominated by the within-rates, we obtain that the fMRI data have the long memory temporal dependence with estimated parameter $\hat{\alpha} \leq 0.25$ since $\max_{1 \leq i \leq p} |\hat{\rho}_i(t)| \leq t^{-0.25}$ (see Figure 1(a)).

3. Main Results. We consider the generalized thresholding and banding estimators based on the sample covariance matrix defined by (1), i.e.,

$$
\hat{\Sigma} = (\hat{\sigma}_{ij})_{p \times p} = \left( \frac{1}{n} \sum_{k=1}^{n} X_{ik}X_{jk} - \bar{X}_i\bar{X}_j \right)_{p \times p}
$$

with $\bar{X}_i = \frac{1}{n} \sum_{k=1}^{n} X_{ik}$, and the corresponding sample correlation matrix by

$$
\hat{R} = (\hat{\rho}_{ij})_{p \times p} = \left( \frac{\hat{\sigma}_{ij}}{\sqrt{\hat{\sigma}_{ii}\hat{\sigma}_{jj}}} \right)_{p \times p}.
$$

Since $\mathcal{B}(C, \alpha) \subset \mathcal{A}(f,g)$ for some $f$ and $g$, we first provide the theoretical results for generalized thresholding and banding estimators under the general temporal dependence $\mathcal{A}(f,g)$ in Subsections 3.1 and 3.2, respectively, then show the detailed convergence rates for these estimators under the long memory temporal dependence $\mathcal{B}(C, \alpha)$ by specifying its associated $f$ and $g$ in Subsection 3.3.

3.1. Generalized thresholding for general dependence. Consider the class of “approximately sparse” covariance matrices [Bickel and Levina (2008b); Rothman, Levina and Zhu (2009)]

$$
\mathcal{U}(q, c_0(p), M_0) = \left\{ \Sigma : \sigma_{ii} \leq M_0, \sum_{j=1}^{p} |\sigma_{ij}|^q \leq c_0(p) \text{ for all } i \right\}.
$$
and the corresponding class of correlation matrices

\[ R(q, c_0(p)) = \left\{ R : \sum_{j=1}^{p} |\rho_{ij}|^q \leq c_0(p) \text{ for all } i \right\}, \]

where \( 0 \leq q < 1 \). For any thresholding parameter \( \tau \geq 0 \), define a generalized thresholding function [Rothman, Levina and Zhu (2009)] by \( s_\tau : \mathbb{R} \to \mathbb{R} \) satisfying the following conditions for all \( z \in \mathbb{R} \):

(i) \( |s_\tau(z)| \leq |z| \); 
(ii) \( s_\tau(z) = 0 \) for \( |z| \leq \tau \); 
(iii) \( |s_\tau(z) - z| \leq \tau \).

The generalized thresholding satisfying conditions (i) through (iii) covers many popular thresholding functions, including hard thresholding

\[ s_\tau^H(z) = z \mathbb{1}(|z| > \tau), \]

soft thresholding

\[ s_\tau^S(z) = \text{sign}(z)(|z| - \tau)_+, \]

smoothly clipped absolute deviation [SCAD; Fan and Li (2001)] and adaptive lasso [Zou (2006)] thresholdings. See details about these examples in Rothman, Levina and Zhu (2009).

We define the generalized thresholding estimators of \( \Sigma \) and \( R \) by

\[ S_\tau(\hat{\Sigma}) = (s_\tau(\hat{\sigma}_{ij}))_{p \times p} \]

and

\[ S_\tau(\hat{R}) = \left( s_\tau(\hat{\rho}_{ij}) \mathbb{1}(i \neq j) + 1(i = j) \right)_{p \times p}, \]

respectively. Then we have the following theoretical results.

**Theorem 1 (Generalized thresholding).** Suppose \( \text{vec}(X_{p \times n}) \sim \mathcal{N}_{np}(I_n \otimes \mu_p, \Delta) \). Uniformly on \( U(q, c_0(p), M_0) \) and \( A(f(n, p), g(n, p)) \), for a sufficiently large constant \( M > 0 \), if \( f = O(n g^2 \log p_n) \) and \( \tau = M \tau' \) with \( \tau' = g \log p_n = o(1) \), then

\[ \|S_\tau(\hat{\Sigma}) - \Sigma\|_2 = O_p(c_0(p)\tau'^{1-q}), \]

(6)

and

\[ \frac{1}{p}\|S_\tau(\hat{\Sigma}) - \Sigma\|_F^2 = O_p(c_0(p)\tau'^{2-q}). \]

(7)
Additionally,

\begin{equation}
\left( E\|S_\tau(\hat{\Sigma}) - \Sigma\|^2 \right)^{\frac{1}{2}} = O\left( c_0(p)\tau^{n-\eta} \right),
\end{equation}

\begin{equation}
\frac{1}{p} E\|S_\tau(\hat{\Sigma}) - \Sigma\|^2_F = O\left( c_0(p)\tau^{2-\eta} \right).
\end{equation}

**Corollary 1.** Theorem 1 holds if \( \hat{\Sigma}, \Sigma \) and \( U(q, c_0(p), M_0) \) are replaced by \( \hat{R}, R \) and \( R(q, c_0(p)) \), respectively.

**Theorem 2 (Sparsistency and sign-consistency).** Suppose \( \text{vec}(X_{p \times n}) \sim N_{np}(I_\tau \otimes \mu_p, \Delta) \), \( \Gamma \in A(f(n,p), g(n,p)) \) and \( \sigma_{ii} \leq M_0 \) for all \( i \). For a sufficiently large constant \( M_1 > 0 \), if \( f = O(n g^2 \log p_n) \) and \( \tau = M_1 \tau' \) with \( \tau' = g \log p_n = o(1) \), then

\[ s_\tau(\hat{\sigma}_{ij}) = 0 \text{ for all } (i, j) \text{ such that } \sigma_{ij} = 0 \]

with probability tending to 1. If additionally assuming that all nonzero elements of \( \Sigma \) satisfy \( |\sigma_{ij}| \geq \xi \) where \( \xi - \tau \geq M_2 \tau' \) for a sufficiently large constant \( M_2 > 0 \) independent of \( M_1 \), we also have, with probability tending to 1,

\[ \text{sign}(s_\tau(\hat{\sigma}_{ij}) \cdot \sigma_{ij}) = 1 \text{ for all } (i, j) \text{ such that } \sigma_{ij} \neq 0. \]

**Corollary 2.** Theorem 2 holds if \( \hat{\sigma}_{ij} \) and \( \sigma_{ij} \) are replaced by \( \hat{\rho}_{ij} \) and \( \rho_{ij} \), respectively, without imposing the bound for \( \sigma_{ii} \).

### 3.2. Banding for general dependence.

Following Bickel and Levina (2008a), define the class of bandable covariance matrices

\begin{equation}
\hat{U}(M_0, \beta, C_0) = \left\{ \Sigma : \sigma_{ii} \leq M_0, \max_j \sum_{i : |i-j| > k} |\sigma_{ij}| \leq C_0 k^{-\beta}, \beta > 0 \text{ for all } k > 0 \right\},
\end{equation}

and the corresponding class of correlation matrices

\[ \hat{R}(\beta, C_0) = \left\{ R : \max_j \sum_{i : |i-j| > k} |\rho_{ij}| \leq C_0 k^{-\beta}, \beta > 0 \text{ for all } k > 0 \right\}. \]

The banding operator for any square matrix \( M = (M_{ij})_{p \times p} \) is defined as

\[ B_k(M) = (M_{ij} 1(|i-j| \leq k))_{p \times p}, \text{ } 0 \leq k < p. \]

Then we have the following consistency results for banding estimators.
Theorem 3 (Banding). Suppose \( \text{vec}(X_{p \times n}) \sim \mathcal{N}_{np}(I_n \otimes \mu, \Delta) \). Uniformly on \( \tilde{U}(M_0, \beta, C_0) \) and \( A(f(n,p), g(n,p)) \), if \( f = O(n g^2 \log p_n) \) and \( k \sim \tau^{-\frac{1}{\beta+1}} \) with \( \tau' = g \log p_n = o(1) \), then

\[
\left\| B_k(\hat{\Sigma}) - \Sigma \right\|_2^2 = O_p\left( \tau'^{-\frac{2\beta+1}{\beta+1}} \right) = \frac{1}{p} \left\| B_k(\hat{\Sigma}) - \Sigma \right\|_F^2,
\]

\[
E\left\| B_k(\hat{\Sigma}) - \Sigma \right\|_2^2 = O\left( \tau'^{-\frac{2\beta+1}{\beta+1}} \right) = \frac{1}{p} E\left\| B_k(\hat{\Sigma}) - \Sigma \right\|_F^2.
\]

If additionally assume \( |\sigma_{ij}| \leq C|i - j|^{-(\beta+1)} \) for all \( i, j \) such that \( |i - j| \geq 1 \), then we have smaller bounds for Frobenius norms as follows:

\[
\frac{1}{p} \left\| B_k(\hat{\Sigma}) - \Sigma \right\|_F^2 = O_p\left( \tau'^{-\frac{2\beta+1}{\beta+1}} \right),
\]

\[
\frac{1}{p} E\left\| B_k(\hat{\Sigma}) - \Sigma \right\|_F^2 = O\left( \tau'^{-\frac{2\beta+1}{\beta+1}} \right).
\]

Corollary 3. Theorem 3 holds if \( \hat{\Sigma}, \Sigma, \sigma_{ij} \) and \( \tilde{U}(M_0, \beta, C_0) \) are replaced by \( \hat{R}, R, \rho_{ij} \) and \( \tilde{R}(\beta, C_0) \), respectively.

Remark 1. Define

\[
\mathcal{V}(M_0, \beta, C) = \left\{ \Sigma : \sigma_{ii} \leq M_0, \ |\sigma_{ij}| \leq C|i - j|^{-(\beta+1)}, \beta > 0 \text{ for all } i, j : |i - j| \geq 1 \right\}
\]

and

\[
\mathcal{W}(M_0, \beta, C) = \left\{ R : |\rho_{ij}| \leq C|i - j|^{-(\beta+1)}, \beta > 0 \text{ for all } i, j : |i - j| \geq 1 \right\}.
\]

It can be shown that \( \mathcal{V}(M_0, \beta, C) \subset \tilde{U}(M_0, \beta, C_0) \) and \( \mathcal{W}(\beta, C) \subset \tilde{R}(\beta, C_0) \) for a suitable choice of \( C_0 \). If \( q > 1/(1 + \beta) \), there exists a constant \( c_0 > 0 \) such that \( \mathcal{V}(M_0, \beta, C) \subset \mathcal{U}(q, c_0, M_0) \) and \( \mathcal{W}(\beta, C) \subset \mathcal{R}(q, c_0) \) [See Bickel and Levina (2008b)]. By Theorems 1 and 3 and Corollaries 1 and 3, the banding estimators should have no slower convergence rates than the thresholding estimators respectively on \( \mathcal{V}(M_0, \beta, C) \) and \( \mathcal{W}(\beta, C) \) for \( q > 1/(1 + \beta) \). Thus, one would expect better estimators using banding if the banding order can be found by using methods like Isomap [Wagaman and Levina (2009)].
3.3. Main results for long memory dependence. As shown in Subsection 2.2, \( \mathcal{B}(C, \alpha) \subseteq \mathcal{A}(f, g) \) for some suitable choice of \( f \) and \( g \). In this subsection, we specify a choice of \( f \) and \( g \) that makes the main results for general temporal dependence applicable to the long memory temporal dependence. Our choice of \( f \) and \( g \) may not be optimal but is easy to obtain under assumed conditions.

**Theorem 4.** All the theorems and corollaries in Subsections 3.1 and 3.2 hold when \( \mathcal{A}(f, g) \) is replaced by \( \mathcal{B}(C, \alpha) \) and \( \tau' = g \log p_n = o(1) \) with \( g \) satisfying (3) and

\[
  f \sim \begin{cases} 
  n^{1-2\alpha/3}(\log p_n)^{1/3}, & 0 < \alpha < 1, \\
  (n \log p_n)^{\frac{1}{\alpha+1}}, & \alpha \geq 1.
  \end{cases}
\]

**Remark 2.** From the consistency theorems and corollaries in Subsections 3.1 and 3.2, we can see that the convergence rates are determined by \( \tau' \). Define

\[
  f_0 = \begin{cases} 
  n^{1-2\alpha/3}(\log p_n)^{1/3}, & 0 < \alpha < 1, \\
  (n \log p_n)^{\frac{1}{\alpha+1}}, & \alpha \geq 1,
  \end{cases}
\]

and

\[
  g_0 = 2C \times \begin{cases} 
  \frac{(n/f_0)^{1-\alpha}-\alpha}{(1-\alpha)f_0^{\alpha}}, & \alpha \neq 1, \\
  f_0^{-\frac{1}{3}}(1 + \log(n/f_0)), & \alpha = 1.
  \end{cases}
\]

To give a simplified expression for the magnitude of \( \tau' \), we treat \( \alpha \) as a fixed value and drop smaller magnitude items in \( g_0 \log p_n \). By Lemma 2 given in Section 6, \( f_0 \sim f = o(n/\log p_n) = o(n) \), thus \( n/f_0 \to \infty \) as \( n \to \infty \). Hence, for any fixed \( \alpha \),

(13) \[
  \tau' \sim g_0 \log p_n
\]

\[
  \propto \begin{cases} 
  n^{1-\alpha}f_0^{-1} \log p_n, & 0 < \alpha < 1, \\
  f_0^{-1}(\log p_n) \log(n/f_0), & \alpha = 1, \\
  f_0^{-\alpha} \log p_n, & \alpha > 1,
  \end{cases}
\]

\[
  = \begin{cases} 
  n^{-\alpha/3}(\log p_n)^{2/3}, & 0 < \alpha < 1, \\
  n^{-1/3}(\log p_n)^{2/3} \log(n^{2/3}(\log p_n)^{-1/3}), & \alpha = 1, \\
  n^{-\alpha} \log p_n^{\frac{\alpha+1}{\alpha+1}}, & \alpha > 1.
  \end{cases}
\]

**Remark 3.** It is noticeable from (13) that when \( \alpha = \infty \), the theoretical results for long memory temporal dependence reduce to those for i.i.d. observations given in Bickel and Levina (2008a,b) and Rothman, Levina and Zhu (2009) when \( p \geq n \) or \( p < n = O(p^c) \) with a constant \( c > 0 \).
Remark 4. The discontinuity of the simplified magnitude of $\tau'$ in (13) at $\alpha = 1$ is due to dropping smaller magnitude items in $g_0 \log p_n$ by treating $\alpha$ as a fixed value. However, these items are not ignorable when $\alpha \to 1$. In fact, one can show that $g_0 \log p_n$ is continuous in $\alpha$, and further, the simplified magnitude of $\tau'$ is monotone in $\alpha$ for large enough $n$ that may depend on $\alpha$.

4. Cross-validation. We propose a cross-validation (CV) method that includes the following three steps:

1. Split the data $X_{p \times n}$ into $H_1 \geq 4$ (almost) equal-sized nonoverlapping blocks $X_i^*$, $i = 1, \ldots, H_1$, such that $X_{p \times n} = (X_1^*, X_2^*, \ldots, X_{H_1}^*)$. For each $i \in \{1, \ldots, H_1\}$, set aside block $X_i^*$ that will be used as the validation data, and use the remaining data after further dropping the neighboring block at either side of $X_i^*$ as the training data that are denoted by $X_i^{**}$.

2. Randomly sample $H_2$ blocks $X_{H_1+1}^*, \ldots, X_{H_1+H_2}^*$ from $X_{p \times n}$, where $X_{H_1+j}^*$ consists of $\lceil n/H_1 \rceil$ consecutive columns of $X_{p \times n}$ for each $j = 1, \ldots, H_2$. Note that these sampled blocks can overlap. For each $i \in \{H_1+1, \ldots, H_1+H_2\}$, set aside block $X_i^*$ as the validation data, and use the remaining data by further excluding the $\lceil n/H_1 \rceil$ columns at either side of $X_i^*$ from $X_{p \times n}$ as the training data that are denoted by $X_i^{**}$.

3. Let $H = H_1 + H_2$. Select the optimal thresholding parameter $\tau$ among a prespecified set of candidates $\{\tau_j\}_{j=1}^J$ and denote it by

$$
\tau_s^\Sigma = \arg \min_{1 \leq j \leq J} \frac{1}{H} \sum_{i=1}^{H} \| S_{\tau_j}(\hat{\Sigma}_i^{**}) - \hat{\Sigma}_i^* \|_F^2
$$

or

$$
\tau_s^R = \arg \min_{1 \leq j \leq J} \frac{1}{H} \sum_{i=1}^{H} \| S_{\tau_j}(\hat{R}_i^{**}) - \hat{R}_i^* \|_F^2,
$$

where $\hat{\Sigma}_i^*$ or $\hat{R}_i^*$, and $\hat{\Sigma}_i^{**}$ or $\hat{R}_i^{**}$ are the corresponding sample covariance or correlation matrices based on $X_i^*$ and $X_i^{**}$, respectively.

Since the temporal dependence between observations $X_i$ decays at least with a power rate, the training data and the validation data are nearly independent if the temporal distance is sufficiently large. We use “gap” blocks, each of size $\approx \lceil n/H_1 \rceil$, to separate training and validation datasets in our CV method. The idea of using “gap” blocks has been employed by the hv-block CV of Racine (2000) for linear models with dependent data. Similar to the $K$-fold CV for i.i.d. observations, Step 1 guarantees all observations are used for both training and validation. To reduce the variability of Step 1 due
to the restriction of keeping the temporal ordering of the observations, Step 2 allows more data splits. This is particularly useful when Step 1 only allows a small number of data splits due to large-size of the “gap” block and/or limited sample size $n$. Step 2 is inspired by the commonly used repeated random subsampling CV for i.i.d. observations [Syed, Principe and Pardalos (2012)]. As many others have shown [Bickel and Levina (2008b); Rothman, Levina and Zhu (2009); Cai and Liu (2011)], using Frobenius norm for tuning parameter selection is adequate. It is easy to show the validity of the proposed CV for i.i.d. observations following Bickel and Levina (2008b), but the theoretical justification for temporally dependent data remains open. However, our simulation studies show that the method performs well for data with long memory temporal dependence.

5. Simulation Studies. In this section, we evaluate the numerical performance of the hard and soft thresholding estimators for correlation matrix, which has practical significance for the brain functional connectivity research. Data are generated from multivariate Gaussian distribution with $\mu_p = 0$, rate matrix $\Gamma$ of the form

\[
(14) \quad \Theta^{ij} = \begin{cases} 
J_{p \times p}, & 1 \leq i = j \leq n, \\
\theta|i - j|^{-\alpha}J_{p \times p}, & 1 \leq i \neq j \leq n,
\end{cases}
\]

and the covariance matrix $\Sigma$ from one of the following four models:

- Model 1: $\sigma_{ij} = \rho^{|i-j|};$
- Model 2: $\sigma_{ij} = (1 - \frac{|i-j|}{10})_+;$
- Model 3: $\sigma_{ij} = (1 - \frac{|i-j|}{0.1p})_+;$
- Model 4: $\sigma_{ij} = (1 - \frac{|i-j|}{0.5p})_+.$

In particular, let $X^{(1)}, \ldots, X^{(p)}$ be the rows of data matrix $X_{p \times n}$, and define $\Sigma = (\tilde{\sigma}_{ij})_{n \times n}$ with

\[
\tilde{\sigma}_{ij} = \begin{cases} 
1, & 1 \leq i = j \leq n, \\
\theta|i - j|^{-\alpha}, & 1 \leq i \neq j \leq n,
\end{cases}
\]

then proceed in the following:

- Model 1: generate $X^{(s)} = \rho X^{(s-1)} + e_s$, where $X^{(1)} \sim N_n \left(0, \tilde{\Sigma}\right)$, and $e_s \overset{i.i.d.}{\sim} N_n \left(0, (1 - \rho^2) \tilde{\Sigma}\right), s = 2, \ldots, p.$
- Model 2, 3 and 4: generate $X^{(s)} = \sum_{k=0}^{m-1} e_{s-k}$, where $e_{s-k} \overset{i.i.d.}{\sim} N_n \left(0, \frac{1}{m} \tilde{\Sigma}\right)$, $s = 1, \ldots, p$, and $m = 10, 0.1p$ and $0.5p$ for Model 2, 3 and 4, respectively.
The rate matrix in (14) belongs to $\mathcal{B}(\theta, \alpha)$. For the covariance matrix $\Sigma$ (in fact the correlation matrix), Model 1 is a standard test case in the literature [Bickel and Levina (2008b); Rothman, Levina and Zhu (2009)], Model 2 is the most sparse case, and Model 4 is the least sparse case. Models 3 and 4 were considered by Rothman, Levina and Zhu (2009) and Wagaman and Levina (2009).

We measure the estimation performance by the spectral norm loss

$$L(S_r(\hat{R}), R) = \|S_r(\hat{R}) - R\|_2.$$ 

To evaluate the ability of generalized thresholding in recovering sparsity, we use the true-positive rate (TPR) and the false-positive rate (FPR) [Rothman, Levina and Zhu (2009)] defined as

$$TPR = \frac{\#\{(i, j) : s_r(\hat{\rho}_{ij}) \neq 0 \text{ and } \rho_{ij} \neq 0, i \neq j\}}{\#\{(i, j) : \rho_{ij} \neq 0, i \neq j\}},$$

$$FPR = \frac{\#\{(i, j) : s_r(\hat{\rho}_{ij}) \neq 0 \text{ and } \rho_{ij} = 0, i \neq j\}}{\#\{(i, j) : \rho_{ij} = 0, i \neq j\}}.$$

Simulations are conducted with a fixed sample size $n = 500$, variable dimension $p$ ranging from 100 to 2,000, and 100 replications for each setting, for which $\theta$ is fixed at 0.6, $\alpha$ varies from 0.1 to 2, and $\rho$ is fixed at 0.7. The CV is implemented with $H_1 = H_2 = 10$. The case with i.i.d. is also considered. In the following tables, “i.i.d.” denotes the proposed intuitive CV for thresholding parameter selection, and “i.i.d.*” denotes the ordinary 10-fold CV [Fang, Wang and Feng (2013)]. The candidate thresholding values $\{\tau_j\}_{j=1}^J$ are ranging from 0.01 to 0.99 with increments of size 0.01.

From Tables 1–4 we can see that the performance of both hard and soft thresholding estimators depends on the sparsity level of the correlation matrix, where the most sparse Models 1 and 2 perform the best. The performance improves when the temporal dependence decreases. Overall the hard thresholding performs better than the soft thresholding. The intuitive CV performs similarly as the ordinary CV for i.i.d. case.

6. Proofs. We start with providing three technical lemmas that are helpful to the proofs of the theoretical results.

**Lemma 1.** Let $X = (X_1, X_2, \ldots, X_n)^T \sim \mathcal{N}_n(0, \Sigma)$, then

$$P[|\bar{X}|^2 \geq u] \leq \exp\left\{-\frac{n^2u}{8}\right\} + n \exp\left\{-\frac{u}{8\|\Sigma - I_{n \times n}\|_1}\right\},$$

If $\Sigma = I_{n \times n}$, then the second term on the right hand side of the inequality is set to be 0.
### Table 1

**Average (standard error) spectral norm loss for Model 1**

| $\alpha$ | $p$  | Sample  | Hard  | Soft  |
|----------|------|---------|-------|-------|
|          | 100  | 4.59(0.83) | 1.55(0.18) | 1.89(0.16) |
|          | 500  | 23.08(1.81) | 2.08(0.25) | 2.72(0.29) |
|          | 1000 | 46.03(2.91) | 2.27(0.30) | 3.40(0.52) |
|          | 2000 | 91.92(3.85) | 2.67(0.53) | 4.50(0.62) |
| 0.5      | 100  | 3.69(0.50)  | 1.53(0.14) | 1.92(0.14) |
|          | 500  | 16.33(1.24) | 1.67(0.09) | 2.26(0.06) |
|          | 1000 | 31.99(1.93) | 1.78(0.06) | 2.45(0.04) |
|          | 2000 | 63.09(2.93) | 1.89(0.06) | 2.91(0.04) |
| 1        | 100  | 2.52(0.31)  | 1.30(0.13) | 1.69(0.13) |
|          | 500  | 8.61(0.52)  | 1.67(0.09) | 2.26(0.06) |
|          | 1000 | 15.37(0.69) | 1.78(0.06) | 2.45(0.04) |
|          | 2000 | 28.51(1.04) | 1.89(0.06) | 2.65(0.04) |
| 1.5      | 100  | 2.25(0.27)  | 1.23(0.13) | 1.62(0.12) |
|          | 500  | 7.14(0.30)  | 1.58(0.08) | 2.19(0.07) |
|          | 1000 | 11.81(0.35) | 1.69(0.06) | 2.36(0.04) |
|          | 2000 | 20.33(0.47) | 1.82(0.06) | 2.55(0.04) |
| 2        | 100  | 2.16(0.25)  | 1.21(0.13) | 1.60(0.13) |
|          | 500  | 6.75(0.26)  | 1.55(0.08) | 2.14(0.05) |
|          | 1000 | 10.97(0.28) | 1.66(0.06) | 2.35(0.04) |
|          | 2000 | 18.38(0.39) | 1.77(0.06) | 2.51(0.05) |
| i.i.d.   | 100  | 1.59(0.16)  | 0.91(0.10) | 1.27(0.12) |
|          | 500  | 4.47(0.17)  | 1.21(0.07) | 1.78(0.05) |
|          | 1000 | 7.00(0.17)  | 1.29(0.06) | 1.91(0.04) |
|          | 2000 | 11.10(0.20) | 1.38(0.05) | 2.04(0.04) |
| i.i.d.*  | 100  | 1.59(0.16)  | 0.86(0.10) | 1.21(0.10) |
|          | 500  | 4.47(0.17)  | 1.12(0.07) | 1.70(0.05) |
|          | 1000 | 7.00(0.17)  | 1.20(0.06) | 1.83(0.04) |
|          | 2000 | 11.10(0.20) | 1.28(0.05) | 1.96(0.03) |

**Proof.** Let $Y \sim N_n(0,I_{n\times n})$ and $W = \tilde{\Sigma}^{1/2}Y$, where $\tilde{\Sigma} = \tilde{\Sigma}^{1/2}(\tilde{\Sigma}^{1/2})^T$. Thus $X \overset{d}{=} W$. Denote $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i = \frac{1}{n} I_n^T X$. Then

$$P \left[ |X|^2 \geq u \right] = P \left[ |W|^2 \geq u \right]$$

(16)

$$\leq P \left[ |\bar{Y}|^2 \geq \frac{u}{4} \right] + P \left[ \frac{1}{n} I_n^T (\tilde{\Sigma}^{1/2} - I_{n\times n}) Y \geq \frac{u}{4} \right].$$

We have

(17)

$$P \left[ |\bar{Y}|^2 \geq \frac{u}{4} \right] \leq \exp \left\{ -\frac{n^2 u}{8} \right\}.$$
Table 2
Average (standard error) spectral norm loss, and TPRs and FPRs for Model 2

| α   | p   | Sample Hard Soft | TPR/FPR | Spectral norm loss | TPR/FPR |
|-----|-----|------------------|---------|--------------------|---------|
| 0.1 | 100 | 4.78(1.15)       | 2.26(0.29) | 0.80/0.00         | 0.95/0.27 |
|     | 500 | 22.89(2.96)      | 3.44(0.50) | 0.72/0.00         | 0.90/0.08 |
|     | 1000| 46.37(4.16)      | 4.46(0.86) | 0.69/0.00         | 0.87/0.04 |
|     | 2000| 92.09(5.79)      | 6.00(1.13) | 0.67/0.00         | 0.85/0.02 |
| 0.5 | 100 | 4.10(0.82)       | 2.28(0.27) | 0.80/0.00         | 0.95/0.25 |
|     | 500 | 16.94(1.81)      | 3.15(0.11) | 0.78/0.00         | 0.92/0.06 |
|     | 1000| 33.03(2.72)      | 3.46(0.11) | 0.76/0.00         | 0.88/0.04 |
|     | 2000| 64.14(3.97)      | 3.75(0.10) | 0.75/0.00         | 0.85/0.02 |
| 1   | 100 | 3.04(0.43)       | 1.96(0.24) | 0.83/0.00         | 0.96/0.25 |
|     | 500 | 10.35(0.65)      | 2.58(0.11) | 0.78/0.00         | 0.92/0.06 |
|     | 1000| 17.65(1.09)      | 3.08(0.08) | 0.76/0.00         | 0.88/0.02 |
|     | 2000| 31.15(1.43)      | 3.21(0.08) | 0.75/0.00         | 0.85/0.02 |
| 1.5 | 100 | 2.79(0.36)       | 1.86(0.23) | 0.84/0.00         | 0.97/0.25 |
|     | 500 | 8.99(0.48)       | 2.54(0.11) | 0.80/0.00         | 0.93/0.07 |
|     | 1000| 14.56(0.73)      | 2.79(0.09) | 0.78/0.00         | 0.91/0.03 |
|     | 2000| 24.10(1.11)      | 2.99(0.07) | 0.77/0.00         | 0.90/0.02 |
| i.i.d| 100| 2.03(0.25)       | 1.41(0.19) | 0.89/0.00         | 0.99/0.26 |
|     | 500 | 5.96(0.31)       | 1.95(0.08) | 0.86/0.00         | 0.96/0.07 |
|     | 1000| 9.16(0.35)       | 2.16(0.09) | 0.85/0.00         | 0.95/0.04 |
|     | 2000| 14.25(0.38)      | 2.36(0.07) | 0.83/0.00         | 0.93/0.01 |
| i.i.d*| 100| 2.03(0.25)       | 1.39(0.20) | 0.91/0.00         | 0.99/0.31 |
|     | 500 | 5.96(0.31)       | 1.96(0.08) | 0.88/0.00         | 0.97/0.12 |
|     | 1000| 9.16(0.35)       | 2.14(0.08) | 0.87/0.00         | 0.96/0.06 |
|     | 2000| 14.25(0.38)      | 2.29(0.06) | 0.86/0.00         | 0.95/0.03 |

On the other hand, if \( \tilde{\Sigma} \neq I_{n \times n} \), let \( Z = AY \) with \( A = \tilde{\Sigma}^{-1} - I_{n \times n} \), then

\[
P \left[ \frac{1}{n} \| \tilde{\Sigma}^{-1} Y \|_2^2 \geq \frac{u}{4} \right] = P \left[ \| \tilde{\Sigma}^{-1/2} Y \|_2^2 \geq \frac{u}{4} \right]
\]

\[
\leq P \left[ \frac{Y^T A^T A Y}{Y^T Y} \geq \frac{nu}{4} \right] \leq P \left[ \sup_{U \neq 0} \frac{U^T A^T A U}{U^T U} Y^T Y \geq \frac{nu}{4} \right]
\]

\[
= P \left[ \| A^T A \|_2^2 Y^T Y \geq \frac{nu}{4} \right] \leq n P \left[ \| A^T A \|_2^2 \geq \frac{u}{4} \right]
\]

\[
\leq n \exp \left\{ -\frac{u}{8\| A^T A \|_2^2} \right\}.
\]
Table 3

Average (standard error) spectral norm loss, and TPRs and FPRs for Model 3

| α  | p   | Sample Hard | Soft Hard | Soft | TPR/FPR |
|----|-----|------------|----------|------|---------|
|    |     | Spectral norm loss | 1.36(0.43) | 2.26(0.29) | 0.80/0.00 | 0.95/0.27 |
| 0.1| 100 | 4.78(1.15)  | 1.36(0.43) | 2.26(0.29) | 0.80/0.00 | 0.95/0.27 |
|    | 500 | 24.46(6.64) | 6.70(1.82) | 11.50(1.48) | 0.77/0.00 | 0.93/0.28 |
|    | 1000| 48.35(11.46) | 13.62(3.50) | 23.25(2.90) | 0.77/0.00 | 0.92/0.28 |
|    | 2000| 98.93(19.49) | 27.85(7.76) | 46.52(5.84) | 0.77/0.00 | 0.92/0.27 |
| 0.5| 100 | 4.10(0.82)  | 1.27(0.31) | 2.28(0.27) | 0.80/0.00 | 0.95/0.25 |
|    | 500 | 20.75(4.10) | 6.13(1.19) | 11.51(1.31) | 0.78/0.00 | 0.93/0.26 |
|    | 1000| 41.61(7.20) | 12.29(2.02) | 23.38(2.67) | 0.77/0.00 | 0.92/0.27 |
|    | 2000| 83.17(13.78) | 24.84(4.33) | 46.42(5.33) | 0.77/0.00 | 0.92/0.26 |
| 1  | 100 | 3.04(0.43)  | 1.02(0.20) | 1.96(0.24) | 0.83/0.00 | 0.96/0.25 |
|    | 500 | 15.39(1.97) | 4.83(0.87) | 9.88(1.07)  | 0.81/0.00 | 0.94/0.25 |
|    | 1000| 30.75(4.31) | 9.78(1.67) | 20.17(2.26) | 0.80/0.00 | 0.93/0.26 |
|    | 2000| 61.36(9.16) | 19.86(3.83) | 39.8(4.81)  | 0.80/0.00 | 0.93/0.25 |
| 1.5| 100 | 2.79(0.36)  | 0.94(0.18) | 1.86(0.23) | 0.84/0.00 | 0.97/0.25 |
|    | 500 | 14.21(1.71) | 4.51(0.78) | 9.36(1.06)  | 0.82/0.00 | 0.94/0.25 |
|    | 1000| 28.41(3.95) | 9.09(1.57) | 19.18(2.12) | 0.81/0.00 | 0.94/0.25 |
|    | 2000| 56.20(7.43) | 18.53(3.61) | 37.53(4.81) | 0.81/0.00 | 0.94/0.25 |
| 2  | 100 | 2.71(0.35)  | 0.93(0.19) | 1.81(0.23) | 0.85/0.00 | 0.97/0.25 |
|    | 500 | 13.83(1.63) | 4.38(0.78) | 9.19(1.06)  | 0.82/0.00 | 0.95/0.25 |
|    | 1000| 27.77(3.81) | 8.81(1.50) | 18.86(2.03) | 0.81/0.00 | 0.94/0.27 |
|    | 2000| 54.42(6.86) | 18.01(3.51) | 36.76(4.83) | 0.82/0.00 | 0.94/0.26 |
| i.i.d| 100 | 2.03(0.25)  | 0.67(0.17) | 1.41(0.19) | 0.89/0.00 | 0.99/0.26 |
|    | 500 | 10.23(1.31) | 3.15(0.51) | 7.07(0.86)  | 0.86/0.00 | 0.96/0.25 |
|    | 1000| 20.37(2.61) | 6.49(1.12) | 14.55(1.61) | 0.86/0.00 | 0.96/0.26 |
|    | 2000| 39.86(4.71) | 13.14(3.04) | 28.29(3.76) | 0.86/0.00 | 0.96/0.26 |
| i.i.d*| 100 | 2.03(0.25)  | 0.69(0.21) | 1.39(0.20) | 0.91/0.00 | 0.99/0.31 |
|    | 500 | 10.23(1.31) | 3.25(0.67) | 7.02(0.91)  | 0.88/0.00 | 0.97/0.32 |
|    | 1000| 20.37(2.61) | 6.72(1.54) | 14.46(1.68) | 0.87/0.00 | 0.96/0.32 |
|    | 2000| 39.86(4.71) | 13.36(3.41) | 28.02(3.78) | 0.87/0.00 | 0.96/0.31 |

Since

\[
\|A^T A\|_2 \leq (\|A\|_2)^2 = \max_i |\lambda_i^2 - 1|^2 \leq \max_i |(\lambda_i^2 - 1)(\lambda_i^2 + 1)|
\]

\[
\leq \max_i |\lambda_i - 1| = \|\hat{\Sigma} - I_{n \times n}\|_2 \leq \|\hat{\Sigma} - I_{n \times n}\|_1,
\]

where \(\lambda_i \geq 0\), \(i = 1, \ldots, n\), are the eigenvalues of \(\Sigma\), we obtain

\[
(18) \quad P \left[ \frac{1}{u n} I_n^T (\hat{\Sigma}^\frac{1}{2} - I_{n \times n}) Y \right]^2 \geq \frac{u}{4} \leq n \exp \left\{ -\frac{u}{8\|\Sigma - I_{n \times n}\|_1} \right\}.
\]

Therefore from (16), (17) and (18), we obtain (15).

\[\square\]
Table 4

| α   | p  | Sample Hard | Soft Hard | Spectral norm loss | TPR/FPR |
|-----|-----|-------------|-----------|--------------------|---------|
|     | 0.1 |             |           |                    |         |
|     | 100 | 4.23(1.56)  | 3.53(1.28)| 4.07(1.26)         | 0.92/0.23 0.99/0.75 |
|     | 500 | 22.35(9.77)| 18.78(9.26)| 21.12(7.73)         | 0.92/0.27 0.99/0.77 |
|     | 1000| 43.75(18.71)| 35.74(14.00)| 41.65(14.83)         | 0.92/0.25 0.99/0.80 |
|     | 2000| 88.19(36.88)| 72.48(33.28)| 81.63(25.81)         | 0.91/0.18 0.99/0.74 |
|     | 0.5 |             |           |                    |         |
|     | 100 | 4.03(1.24)  | 3.52(1.31)| 4.08(1.16)         | 0.92/0.26 0.98/0.73 |
|     | 500 | 21.25(7.79)| 18.60(8.01)| 21.05(6.97)         | 0.92/0.31 0.99/0.74 |
|     | 1000| 42.62(14.26)| 36.43(13.50)| 42.64(13.59)         | 0.92/0.31 0.99/0.77 |
|     | 2000| 84.34(35.89)| 71.81(34.32)| 83.71(30.39)         | 0.91/0.22 0.98/0.71 |
|     | 1   |             |           |                    |         |
|     | 100 | 3.39(1.03)  | 3.00(1.13)| 3.54(1.02)         | 0.93/0.29 0.99/0.73 |
|     | 500 | 16.98(6.01)| 15.24(6.50)| 17.74(5.61)         | 0.93/0.31 0.99/0.71 |
|     | 1000| 35.18(11.16)| 31.08(11.54)| 36.19(10.93)         | 0.93/0.33 0.99/0.76 |
|     | 2000| 69.59(28.02)| 60.97(28.57)| 73.31(26.69)         | 0.93/0.24 0.99/0.70 |
|     | 1.5 |             |           |                    |         |
|     | 100 | 3.21(1.03)  | 2.86(1.13)| 3.38(1.01)         | 0.94/0.30 0.99/0.72 |
|     | 500 | 15.75(5.87)| 14.44(6.28)| 16.68(5.26)         | 0.94/0.33 0.99/0.70 |
|     | 1000| 33.02(10.39)| 29.02(11.10)| 33.70(10.12)         | 0.93/0.31 0.99/0.77 |
|     | 2000| 65.46(25.64)| 58.21(27.15)| 69.85(24.88)         | 0.93/0.26 0.99/0.71 |
|     | 2   |             |           |                    |         |
|     | 100 | 3.15(1.04)  | 2.81(1.15)| 3.32(1.04)         | 0.94/0.29 0.99/0.72 |
|     | 500 | 15.39(5.89)| 13.97(6.23)| 16.37(5.15)         | 0.94/0.29 0.99/0.70 |
|     | 1000| 32.32(10.13)| 28.27(10.97)| 32.85(9.86)         | 0.93/0.29 0.99/0.77 |
|     | 2000| 64.20(24.78)| 57.17(26.49)| 68.54(23.89)         | 0.93/0.28 0.99/0.71 |
|     | i.i.d|           |           |                    |         |
|     | 100 | 2.37(0.80)  | 2.06(0.87)| 2.52(0.83)         | 0.95/0.23 0.99/0.71 |
|     | 500 | 11.87(4.14)| 10.35(4.42)| 12.58(3.82)         | 0.94/0.22 0.99/0.70 |
|     | 1000| 24.17(7.36)| 20.78(8.67)| 24.31(7.67)         | 0.95/0.27 0.99/0.75 |
|     | 2000| 43.99(13.54)| 37.84(14.43)| 48.02(14.50)         | 0.94/0.19 0.99/0.68 |
|     | i.i.d*|         |           |                    |         |
|     | 100 | 2.37(0.80)  | 2.06(0.90)| 2.43(0.73)         | 0.95/0.20 0.99/0.74 |
|     | 500 | 11.87(4.14)| 10.15(4.57)| 12.34(3.74)         | 0.95/0.19 0.99/0.71 |
|     | 1000| 24.17(7.36)| 20.60(8.74)| 24.36(7.20)         | 0.95/0.25 0.99/0.77 |
|     | 2000| 43.99(13.54)| 37.35(14.84)| 46.44(13.67)         | 0.95/0.16 0.99/0.73 |

**Lemma 2.** Let \( 1 \leq f \leq n \). If \( f = O(n g^2 \log p_n) \) and \( \tau' = g \log p_n = o(1) \), then \( f = o(n/\log p_n) \) and \( n^{-1} \log p_n = o(1) \). Moreover, for any positive constants \( M', c_1 \) and \( c_2 \), there exists a constant \( M > 0 \) such that for sufficiently large \( n \),

\[
p_n^c_1 \exp\{-c_2 n^2 u/f^2\} < n_p^c_1 \exp\{-c_2 nu^2/f\} < n_p^c_1 \exp\{-c_2 u/g\} \leq n_p^{-M'},
\]

where \( u = M \tau' \).

**Proof.** Because \( \tau' = g \log p_n = o(1) \), i.e., \( g^2 \log p_n = o(1/\log p_n) \), we have \( 1 \leq f = O(n g^2 \log p_n) = o(n/\log p_n) \), thus \( n^{-1} \log p_n = o(1) \).

Since \( f = O(n g^2 \log p_n) \), there exist positive constants \( c_3 \) and \( N_1(c_3) \) such that \( f < c_3 n g^2 \log p_n \) when \( n > N_1(c_3) \). Then let \( u = M \tau' = M g \log p_n \), we
have

\[ nu^2/f > n(Mg \log p_n)^2/(c_3n g^2 \log p_n) \geq u/g \]

if \( M \geq c_3 \) and \( n > N_1(c_3) \). By \( \tau' = o(1) \), we have for any constant \( M > 0 \), there exists a constant \( N_2(M) > 0 \) such that \( \tau' < 1/M \) when \( n > N_2(M) \). Thus, when \( n > N_2(M) \),

\[ nu^2/f < n^2u/f^2, \]

which follows from \( u = M\tau' < 1 \) and \( n/f \geq 1 \). Thus by (19) and (20), we have

\[ p_n \exp\left\{-c_2n^2u/f^2\right\} < p_n \exp\left\{-c_2nu/f\right\} < p_n \exp\{-c_2u/g\} \]

when \( n > \max\{N_1(c_3), N_2(M)\} \) and \( M \geq c_3 \). Let \( M \geq \max\{c_3, (c_1 + M')/c_2\} \), then

\[ p_n \exp\{-(c_1 - c_2M) \log p_n\} = p_n^{c_1-c_2 M} \leq p_n^{-M'}. \]

This completes the proof.

**Lemma 3.** For any \( 0 \leq u \leq 4n \),

\[ P \left[ \left| \chi_n^2 - n \right| \geq u \right] \leq 2 \exp\left\{-\frac{u^2}{16n}\right\}. \]

**Proof.** The inequality is obvious for the case \( u = 0 \), so we consider the case \( 0 < u \leq 4n \). From Lemma 1 in Laurent and Massart (2000), we have for any \( x > 0 \),

\[ P \left[ \chi_n^2 - n \geq 2\sqrt{nx} + 2x \right] \leq \exp(-x), \]

and

\[ P \left[ \chi_n^2 - n \leq -2\sqrt{nx} \right] \leq \exp(-x). \]

Let \( x = \frac{u^2}{16n} \), then \( 2\sqrt{nx} + 2x = \frac{u}{2} + \frac{u^2}{8n} \leq u \) since \( 0 < u \leq 4n \). From (22),

\[ P \left[ \chi_n^2 - n \geq u \right] \leq P \left[ \chi_n^2 - n \geq \frac{u}{2} + \frac{u^2}{8n} \right] \leq \exp\left\{-\frac{u^2}{16n}\right\}. \]

Plugging \( x = \frac{u^2}{16n} \) into (23) yields

\[ P \left[ \chi_n^2 - n \leq -u \right] \leq \exp\left\{-\frac{u^2}{4n}\right\} \leq \exp\left\{-\frac{u^2}{16n}\right\}. \]

Combining (24) and (25), we obtain (21).
Proof of Theorem 1. Without loss of generality, we assume the unknown true $\mu_p = 0$. Note that

$$
\max_{1 \leq i, j \leq p} |\hat{\sigma}_{ij} - \sigma_{ij}| \leq \max_{1 \leq i, j \leq p} |\bar{X}_i \bar{X}_j| + \max_{1 \leq i, j \leq p} \left| \frac{1}{n} \sum_{k=1}^{n} X_{ik} X_{jk} - \sigma_{ij} \right|
$$

$$
\leq \max_{1 \leq i \leq p} |\bar{X}_i|^2 + \max_{1 \leq i, j \leq p} \left| \frac{1}{n} \sum_{k=1}^{n} X_{ik} X_{jk} - \sigma_{ij} \right|
$$

(26)

Let $Z_{ij} = \frac{X_{ij}}{\sqrt{\sigma_{ii}}}$, $A_{r,f} = \{ k \in \mathbb{Z}^+ \cup \{0\} : kf + r \leq n \}$, $r \in \{1, \ldots, f\}$, and $C_{r,f}$ be the cardinality of $A_{r,f}$. For a fixed integer $f$ and any integer $1 \leq j \leq n$, we have $j = kf + r$, where $k = \lfloor \frac{j}{f} \rfloor$ if $\frac{j}{f}$ is not an integer, and otherwise $k = \frac{j}{f} - 1$. Hence,

$$
\sum_{j=1}^{n} X_{ij} = \sum_{r=1}^{f} \sum_{k \in A_{r,f}} X_{i,kf+r}
$$

and

(27)

$$
n = \sum_{r=1}^{f} C_{r,f}.
$$

Moreover, for any $r \in \{1, \ldots, f\}$,

$$
n/f - 2 \leq \lfloor n/f \rfloor - 1 \leq C_{r,f} - 1 \leq \lceil n/f \rceil \leq n/f,
$$

thus $n - f \leq fC_{r,f} \leq 2n$. By $f = O(n^g \log p_n)$ and $g \log p_n = o(1)$, we have $f = o(n \log p_n)$. Hence, $n \asymp n - f \leq fC_{r,f} \leq 2n$. 


Now, for any $u > 0$,

\[
P \left[ \max_{1 \leq i \leq p} |\bar{X}_i|^2 \geq u \right] \leq \sum_{i=1}^{p} P \left[ |\bar{X}_i| \geq u^{1/2} \right] = \sum_{i=1}^{p} \sum_{j=1}^{n} P \left[ |X_{ij}| \geq nu^{1/2} \right]
\]

(28)

Let $\Delta^{ifr}$ be the covariance matrix of $\text{vec}\{Z_{i,kf+r} : k \in A_{r,f}\}$, i.e.,

\[
\Delta^{ifr} = \left( \text{cov}(Z_{i,kf+r}, Z_{i,lf+r}) \right)_{C_r,f \times C_r,f} = \left( \theta_{ii}^{kf+r,lf+r} 1(k \neq l) + 1(k = l) \right)_{C_r,f \times C_r,f}.
\]

Then

\[
\|\Delta^{ifr} - I_{C_r,f \times C_r,f}\|_1 = \max_{l \neq \ell} \sum_{k} |\theta_{ii}^{kf+r,lf+r}|
\]

(29)

by assumption (2). From Lemma 1 and (29),

\[
P \left[ \frac{1}{C_{r,f}} \sum_{k \in A_{r,f}} Z_{i,kf+r} \geq \frac{n}{f^{C_{r,f}}} \sqrt{\frac{u}{\sigma_{ii}}} \right] \leq \exp \left\{ -\frac{n^2u}{8f^2\sigma_{ii}} \right\} + C_{r,f} \exp \left\{ -\frac{n^2u}{8C_{r,f}^2f^2g\sigma_{ii}} \right\}.
\]

(30)
By (28), (30), $\sigma_{ii} \leq M_0$ in (4) and $fC_{r,f} \leq 2n$ prior to (28), we have

$$P \left[ \max_{1 \leq i \leq p} |\overline{X}_i|^2 \geq u \right] \leq \rho \exp \left\{ -\frac{n^2u}{8f^2M_0} \right\} + \rho \sum_{r=1}^{f} C_{r,f} \exp \left\{ -\frac{n^2u}{8C_{r,f}^2f^2gM_0} \right\}$$

$$\leq \rho^2 \exp \left\{ -\frac{n^2u}{8f^2M_0} \right\} + \rho^2 \exp \left\{ -\frac{u}{32gM_0} \right\},$$

(31) following from $f \leq n \leq p_n$ and (27).

On the other hand, we have

$$P \left[ \max_{1 \leq i,j \leq p} \left| \frac{1}{n} \sum_{k=1}^{n} X_{ik}X_{jk} - \sigma_{ij} \right| \geq u \right]$$

$$\leq \rho^2 \left[ \frac{1}{n} \sum_{k=1}^{n} X_{ik}X_{jk} - \sigma_{ij} \right] \geq u \right]$$

$$= \rho^2 \left[ \sum_{k=1}^{n} Z_{ik}Z_{jk} - n\rho_{ij} \geq \frac{nu}{\sqrt{\sigma_{ii}\sigma_{jj}}} \right]$$

$$\leq \rho^2 \left[ \sum_{k=1}^{n} (Z_{ik} + Z_{jk})^2 - 2n(1 + \rho_{ij}) \geq \frac{2nu}{\sqrt{\sigma_{ii}\sigma_{jj}}} \right]$$

$$+ \rho^2 \left[ \sum_{k=1}^{n} (Z_{ik} - Z_{jk})^2 - 2n(1 - \rho_{ij}) \geq \frac{2nu}{\sqrt{\sigma_{ii}\sigma_{jj}}} \right]$$

(32)

Without loss of generality, we only consider

$$P \left[ \sum_{k=1}^{n} (Z_{ik} - Z_{jk})^2 - 2n(1 - \rho_{ij}) \geq \frac{2nu}{\sqrt{\sigma_{ii}\sigma_{jj}}} \right].$$

(I) If $\rho_{ij} = 1$, then the equality in the Cauchy-Schwarz inequality holds, thus $P(Z_{ik} = Z_{jk}) = 1$. Hence,

$$P \left[ \sum_{k=1}^{n} (Z_{ik} - Z_{jk})^2 - 2n(1 - \rho_{ij}) \geq \frac{2nu}{\sqrt{\sigma_{ii}\sigma_{jj}}} \right] = 0.$$
(II) If \( \rho_{ij} \neq 1 \), we consider

\[
P \left[ \sum_{k=1}^{n} (Z_{ik} - Z_{jk})^2 - 2n(1 - \rho_{ij}) \right] \geq \frac{2nu}{\sqrt{\sigma_{ii}\sigma_{jj}}} \]

\[
= P \left[ \sum_{r=1}^{f} \sum_{k \in A_{r,f}} (Z_{ikf+r} - Z_{jkf+r})^2 - 2 \sum_{r=1}^{f} C_{r,f}(1 - \rho_{ij}) \right] \geq \frac{2nu}{\sqrt{\sigma_{ii}\sigma_{jj}}} \]

\[
\leq \sum_{r=1}^{f} P \left[ \sum_{k \in A_{r,f}} \left( \frac{Z_{ikf+r} - Z_{jkf+r}}{\sqrt{2(1 - \rho_{ij})}} \right)^2 - C_{r,f} \right] \geq \frac{nu}{(1 - \rho_{ij})\sqrt{\sigma_{ii}\sigma_{jj}}} \]

\[
\leq \sum_{r=1}^{f} P \left[ Y^T (\Gamma_{ijrf} - I_{C_{r,f} \times C_{r,f}}) Y \right] \geq \frac{nu}{2(1 - \rho_{ij})f\sqrt{\sigma_{ii}\sigma_{jj}}} \]

(33)

\[
\leq \sum_{r=1}^{f} P \left| \chi^2_{r,f} - C_{r,f} \geq \frac{nu}{4fM_0} \right|
\]

where

\[
Y \sim N_{C_{r,f}}(0, I_{C_{r,f} \times C_{r,f}}),
\]

\[
vec \left\{ \frac{Z_{ikf+r} - Z_{jkf+r}}{\sqrt{2(1 - \rho_{ij})}} : k \in A_{r,f} \right\} \sim N_{C_{r,f}}(0, \Gamma_{ijrf}),
\]

and \( \Gamma_{ijrf} = \left( \gamma_{ijkl}^{ijrf} \right)_{C_{r,f} \times C_{r,f}} \) with

\[
\gamma_{ijkl}^{ijrf} = \begin{cases} 
\theta_{ii}^{k+f+r,lf+r} - \left( \theta_{ij}^{k+f+r,lf+r} + \theta_{ji}^{k+f+r,lf+r} \right)\rho_{ij} & \text{for } k \neq l \\
1, & \text{for } k = l
\end{cases}
\]

and

\[
\gamma_{kkl}^{ijrf} = \begin{cases} 
\theta_{ii}^{k+f+r,lf+r} - \left( \theta_{ij}^{k+f+r,lf+r} + \theta_{ji}^{k+f+r,lf+r} \right)\rho_{ij} & \text{for } k \neq l \\
1, & \text{for } k = l
\end{cases}
\]
and $k, l \in A_{r,f}$. We consider the second term in (33) first. Similar to (29), we have

\begin{align}
\| \Gamma_{ijrf} - I_{C_{r,f} \times C_{r,f}} \|_2 & \leq \| \Gamma_{ijrf} - I_{C_{r,f} \times C_{r,f}} \|_1 \leq 2(1 - \rho_{ij})^{-1} g.
\end{align}

Similar to (18), by (34), we have

\begin{align*}
P \left[ | \chi_{C_{r,f}}^{2} - C_{r,f} | \geq \frac{nu}{4fM_0} \right] & \leq 2 \exp \left\{ - \frac{n^2u^2}{512fM_0^2} \right\}.
\end{align*}

Plugging (35) and (36) into (33) yields

\begin{align*}
P \left[ \sum_{k=1}^{n} (Z_{ik} - Z_{jk})^2 - 2n(1 - \rho_{ij}) \right] & \geq \frac{2nu}{\sqrt{\sigma_{ii}\sigma_{jj}}} \\
& \leq 2f \exp \left\{ - \frac{n^2u^2}{512fM_0^2} \right\} + \sum_{r=1}^{f} C_{r,f} \exp \left\{ - \frac{c_5u}{M_0g} \right\} \\
& = 2f \exp \left\{ - \frac{n^2u^2}{512fM_0^2} \right\} + n \exp \left\{ - \frac{c_5u}{M_0g} \right\}.
\end{align*}
From (I) and (II), for any $\rho_{ij}$, when $n > N(M_1)$,

$$p^2 P \left[ \sum_{k=1}^{n} (Z_{ik} - Z_{jk})^2 - 2n(1 - \rho_{ij}) \geq \frac{2nu}{\sqrt{\sigma_{ii}\sigma_{jj}}} \right]$$

$$\leq 2p_n^3 \exp \left\{ -\frac{nu^2}{512fM_0^2} \right\} + P_n^3 \exp \left\{ -\frac{c_5u}{M_0g} \right\}.$$  

Similarly, when $n > N(M_1)$,

$$p^2 P \left[ \sum_{k=1}^{n} (Z_{ik} + Z_{jk})^2 - 2n(1 + \rho_{ij}) \geq \frac{2nu}{\sqrt{\sigma_{ii}\sigma_{jj}}} \right]$$

$$\leq 2p_n^3 \exp \left\{ -\frac{nu^2}{512fM_0^2} \right\} + P_n^3 \exp \left\{ -\frac{c_5u}{M_0g} \right\}.$$  

Therefore by (32), when $n > N(M_1)$,

$$P \left[ \max_{1 \leq i,j \leq p} \left| \frac{1}{n} \sum_{k=1}^{n} X_{ik}X_{jk} - \sigma_{ij} \right| \geq u \right]$$

$$\leq 4p_n^3 \exp \left\{ -\frac{nu^2}{512fM_0^2} \right\} + 2p_n^3 \exp \left\{ -\frac{c_5u}{M_0g} \right\}. \quad (37)$$

From (26), (31) and (37), when $n > N(M_1)$, we obtain

$$P \left[ \max_{1 \leq i,j \leq p} |\hat{\sigma}_{ij} - \sigma_{ij}| \geq 2u \right]$$

$$\leq p_n^2 \exp \left\{ -\frac{n^2u}{8f^2M_0} \right\} + p_n^2 \exp \left\{ -\frac{u}{32gM_0} \right\}$$

$$+ 4p_n^3 \exp \left\{ -\frac{nu^2}{512fM_0^2} \right\} + 2p_n^3 \exp \left\{ -\frac{c_5u}{M_0g} \right\}.$$  

Since $f = O(ng^2\log p_n)$ and $\tau' = g\log p_n = o(1)$, by Lemma 2, for any constant $M' > 0$, there exists a constant $M'_1 > 0$ such that

$$P \left[ \max_{1 \leq i,j \leq p} |\hat{\sigma}_{ij} - \sigma_{ij}| \geq 2u \right] = O(p_n^{-M'})$$

where $u = M'_1\tau'$. For any constant $M \geq 2M'_1$,

$$P \left[ \max_{1 \leq i,j \leq p} |\hat{\sigma}_{ij} - \sigma_{ij}| \geq M\tau' \right] \leq P \left[ \max_{1 \leq i,j \leq p} |\hat{\sigma}_{ij} - \sigma_{ij}| \geq 2M_1\tau' \right].$$
thus,

\[ P \left[ \max_{1 \leq i, j \leq p} |\hat{\sigma}_{ij} - \sigma_{ij}| \geq \tau \right] = O \left( n^{-M'} \right) \]  

with \( \tau = M\tau' \).

Then following the similar lines of the proof of Theorem 1 after equation (12) in Bickel and Levina (2008b) and the proof of Theorem 1 in Rothman, Levina and Zhu (2009), we obtain that for any constant \( M' > 0 \), there exists a constant \( M_2 \geq 2M_1' \) such that

\[ P \left[ \|S_{\tau}(\hat{\Sigma}) - \Sigma\|_2 \geq C_1c_0(p)\tau'^{1-q} \right] \leq P \left[ \|S_{\tau}(\hat{\Sigma}) - \Sigma\|_1 \geq C_1c_0(p)\tau'^{1-q} \right] = O \left( n^{-M'} \right), \]  

where \( \tau = M\tau' \) with any constant \( M \geq M_2 \) and some constant \( C_1 > 0 \) dependent on \( M \). Thus, we obtain (6).

By condition (iii) of the generalized thresholding function and (38), we have

\[
P \left[ \|S_{\tau}(\hat{\Sigma}) - \Sigma\|_\infty \geq 2\tau \right] 
= P \left[ \max_{1 \leq i, j \leq p} |s_{\tau}(\hat{\sigma}_{ij}) - \sigma_{ij}| \geq 2\tau \right] 
\leq P \left[ \max_{1 \leq i, j \leq p} |s_{\tau}(\hat{\sigma}_{ij}) - \hat{\sigma}_{ij}| + \max_{1 \leq i, j \leq p} |\hat{\sigma}_{ij} - \sigma_{ij}| \geq 2\tau \right] 
\leq P \left[ \tau + \max_{1 \leq i, j \leq p} |\hat{\sigma}_{ij} - \sigma_{ij}| \geq 2\tau \right] = O \left( n^{-M'} \right). \]

By (39), (40) and the inequality \( \|M\|_F^2 \leq p\|M\|_1\|M\|_\infty \) for any \( p \times p \) matrix, we have

\[
P \left[ \frac{1}{p} \|S_{\tau}(\hat{\Sigma}) - \Sigma\|_F^2 \geq 2\tau C_1c_0(p)\tau'^{1-q} \right] 
\leq P \left[ \|S_{\tau}(\hat{\Sigma}) - \Sigma\|_1 \|S_{\tau}(\hat{\Sigma}) - \Sigma\|_\infty \geq 2\tau C_1c_0(p)\tau'^{1-q} \right] 
\leq P \left[ \|S_{\tau}(\hat{\Sigma}) - \Sigma\|_\infty \geq 2\tau \right] + P \left[ \|S_{\tau}(\hat{\Sigma}) - \Sigma\|_2 \geq C_1c_0(p)\tau'^{1-q} \right] = O \left( n^{-M'} \right). \]

Thus, (7) holds.
For the convergence in mean square, we consider
\[
E\|S_\tau(\hat{\Sigma}) - \Sigma\|_F^2
= E\left[\|S_\tau(\hat{\Sigma}) - \hat{\Sigma}\|_F^2 \mathbf{1}\left(\|S_\tau(\hat{\Sigma}) - \Sigma\|_F \geq C_1 c_0(p)\tau^{n-1-q}\right)\right]
+ E\left[\|S_\tau(\hat{\Sigma}) - \Sigma\|_F^2 \mathbf{1}\left(\|S_\tau(\hat{\Sigma}) - \Sigma\|_F < C_1 c_0(p)\tau^{n-1-q}\right)\right]
\leq \left(E\|S_\tau(\hat{\Sigma}) - \Sigma\|_F^2\right)^{\frac{1}{2}} \frac{1}{2} \left(P\left(\|S_\tau(\hat{\Sigma}) - \Sigma\|_F \geq C_1 c_0(p)\tau^{n-1-q}\right)\right)^{\frac{1}{2}}
+ (C_1 c_0(p)\tau^{n-1-q})^2
\]
(41)
\leq \left(E\|S_\tau(\hat{\Sigma}) - \Sigma\|_F^2\right)^{\frac{1}{2}} O\left(p_n^{-\frac{2}{3}}\right) + (C_1 c_0(p)\tau^{n-1-q})^2.

By condition (iii) in the definition of the generalized thresholding function, we have
\[
\|S_\tau(\hat{\Sigma}) - \Sigma\|_F \leq p\tau,
\]
then,
\[
E\|S_\tau(\hat{\Sigma}) - \Sigma\|_F^4 \leq E\left(\|S_\tau(\hat{\Sigma}) - \Sigma\|_F^2 + \|\hat{\Sigma} - \Sigma\|_F^2\right)^4
= E\left[\|S_\tau(\hat{\Sigma}) - \hat{\Sigma}\|_F^4 + \|\hat{\Sigma} - \Sigma\|_F^4 + 4\|S_\tau(\hat{\Sigma}) - \hat{\Sigma}\|_F^2 \|\hat{\Sigma} - \Sigma\|_F^2 + 6\|S_\tau(\hat{\Sigma}) - \hat{\Sigma}\|_F^2 \|\hat{\Sigma} - \Sigma\|_F^2\right]
\leq p^4 \tau^4 + E\|\hat{\Sigma} - \Sigma\|_F^4 + 4p^3 \tau^3 E\|\hat{\Sigma} - \Sigma\|_F^2
+ 4p \tau E\|\hat{\Sigma} - \Sigma\|_F^3 + 6p^2 \tau^2 E\|\hat{\Sigma} - \Sigma\|_F^2
\leq p^4 \tau^4 + E\|\hat{\Sigma} - \Sigma\|_F^4 + 4p^3 \tau^3 \left(E\|\hat{\Sigma} - \Sigma\|_F^2\right)^{\frac{1}{2}}
+ 4p \tau \left(E\|\hat{\Sigma} - \Sigma\|_F^6\right)^{\frac{1}{2}} + 6p^2 \tau^2 E\|\hat{\Sigma} - \Sigma\|_F^2.
\]
(42)
It is easy to see that for \(m = 1, 2, 3\),
\[
\|\hat{\Sigma} - \Sigma\|_F^{2m} = \left\{ \sum_{1 \leq i, j \leq p} \left(\frac{1}{n} \sum_{k=1}^{n} X_{ik}X_{jk} - \bar{X}_i\bar{X}_j - \sigma_{ij}\right)\right\}^m
\]
(43)
is a polynomial, in variables \(X_{ij}, 1 \leq i \leq p, 1 \leq j \leq n\), of degree \(4m\), and the number of its terms is bounded by \(p_n^{C(m)}\) with a positive constant \(C(m)\). Since \(\sigma_{ij} \leq M_0\) for \(1 \leq i \leq p\), \(|EX_{ij}X_{kl}| = |Cov(X_{ij}, X_{kl})| \leq M_0\) for \(1 \leq i, k \leq p, 1 \leq j, l \leq n\). Thus by Isserlis’ theorem [Isserlis (1918)],
there exists a positive constant that bounds all terms in the polynomial. Therefore,

\begin{equation}
E \parallel \Sigma - \Sigma \parallel^2_F = O(p_n^{C(m)}).
\end{equation}

Combining (41), (42) and (44), there exist constants \( c_6, c_7 > 0 \) such that

\begin{equation}
E \| S_r(\hat{\Sigma}) - \Sigma \|_2^2 \leq c_6 p_n^{c_7 \tau' - \frac{M'}{2}} + (C_1 c_0(p) \tau'^{1-q})^2.
\end{equation}

By \( 1 \leq f = O(n g^2 \log p_n) \), we have \( O(\tau') = O(g \log p_n) = \sqrt{\log p_n} \). So we can let \( M' \) be sufficiently large such that \( p_n^{c_7 \tau' - \frac{M'}{2}} = O((c_0(p) \tau'^{1-q})^2) \), then (8) holds. Similarly, we can obtain (9).

**Proof of Corollary 1.** The key of the proof is to show that for any constant \( M' > 0 \), there exists a constant \( C' > 0 \) such that

\begin{equation}
P \left[ \max_{1 \leq i,j \leq p} |\hat{\sigma}_{ij} - \sigma_{ij}| \geq C' \tau' \right] = O(p_n^{-M'}).\end{equation}

Similarly to (26),

\begin{equation}
\max_{1 \leq i,j \leq p} \left| \frac{\hat{\sigma}_{ij} - \sigma_{ij}}{\sqrt{\sigma_{ii} \sigma_{jj}}} \right| \leq \max_{1 \leq i,j \leq p} \left| \frac{\hat{X}_i \hat{X}_j}{\sqrt{\sigma_{ii} \sigma_{jj}}} \right| + \max_{1 \leq i,j \leq p} \left| \frac{1}{n \sqrt{\sigma_{ii} \sigma_{jj}}} \sum_{k=1}^{n} X_{ik} X_{jk} - \rho_{ij} \right| \leq \max_{1 \leq i,j \leq p} \left| \frac{\hat{X}_i}{\sqrt{\sigma_{ii}}} \right|^2 + \max_{1 \leq i,j \leq p} \left| \frac{1}{n \sqrt{\sigma_{ii} \sigma_{jj}}} \sum_{k=1}^{n} X_{ik} X_{jk} - \rho_{ij} \right|
\end{equation}

\begin{equation}
= \max_{1 \leq i,j \leq p} \left| \frac{Z_i}{\sqrt{\sigma_{ii}}} \right|^2 + \max_{1 \leq i,j \leq p} \left| \frac{1}{n} \sum_{k=1}^{n} Z_{ik} Z_{jk} - \rho_{ij} \right|,
\end{equation}

where \( Z_{ik} = \frac{X_{ik}}{\sqrt{\sigma_{ii}}} \). In (46), since \( \max_{1 \leq i,j \leq p} |\rho_{ij}| \leq 1 \), we do not need to assume \( \max_{1 \leq i \leq p} |\sigma_{ii}| \leq M_0 \) any more and we impose the “approximately sparse” assumption (5) on \( R \) instead of (4) on \( \Sigma \). Then following the similar lines of the proof of Theorem 1 up to equation (38), we can obtain that for any constant \( M_1 > 0 \), there exists a constant \( C_1 > 0 \) such that

\begin{equation}
O(p_n^{-M_1}) = P \left[ \max_{1 \leq i,j \leq p} \left| \frac{\hat{\sigma}_{ij} - \sigma_{ij}}{\sqrt{\sigma_{ii} \sigma_{jj}}} \right| \geq C_1 \tau' \right] \geq P \left[ \frac{\hat{\sigma}_{ij} - \sigma_{ij}}{\sqrt{\sigma_{ii} \sigma_{jj}}} \geq C_1 \tau' \right]
\end{equation}

for any \( 1 \leq i, j \leq p \). Thus let \( i = j \), then

\begin{equation}
O(p_n^{-M_1}) = P \left[ \frac{\hat{\sigma}_{ii}}{\sigma_{ii}} - 1 \right] \geq C_1 \tau' \geq P \left[ \left| \frac{\hat{\sigma}_{ii}}{\sigma_{ii}} - 1 \right| \geq C_1 \tau' \right],
\end{equation}
and

\[
O(p_n^{-M_1}) = P \left[ \left\| \hat{\sigma}_{ii} \hat{\sigma}_{jj} \right\| - 1 \geq C_1 \tau' \right] + P \left[ \left\| \hat{\sigma}_{jj} \right\| - 1 \geq C_1 \tau' \right]
\]

\[
\geq P \left[ \left\| \hat{\sigma}_{ii} \right\| - 1 \left\| \hat{\sigma}_{jj} \right\| - 1 \geq C_2 \tau'^2 \right]
\]

\[
= P \left[ \left\| \frac{\hat{\sigma}_{ii} \hat{\sigma}_{jj}}{\hat{\sigma}_{ii} \hat{\sigma}_{jj}} - 1 \right\| \right. \left\| \frac{\hat{\sigma}_{ii}}{\hat{\sigma}_{ii}} - 1 \right\| + 1 \left\| \frac{\hat{\sigma}_{jj}}{\hat{\sigma}_{jj}} - 1 \right\| \geq C_2 \tau'^2 \right]
\]

\[
\geq P \left[ \left\| \frac{\hat{\sigma}_{ii} \hat{\sigma}_{jj}}{\hat{\sigma}_{ii} \hat{\sigma}_{jj}} - 1 \right\| \right. \left\| \frac{\hat{\sigma}_{ii}}{\hat{\sigma}_{ii}} - 1 \right\| + \left\| \frac{\hat{\sigma}_{jj}}{\hat{\sigma}_{jj}} - 1 \right\| \right. \left\| \frac{\hat{\sigma}_{ii}}{\hat{\sigma}_{ii}} - 1 \right\| \left. \left\| \frac{\hat{\sigma}_{jj}}{\hat{\sigma}_{jj}} - 1 \right\| \right. \left. \geq C_2 \tau'^2 \right]
\]

By (48), (49) and \( \tau' = o(1) \),

\[
P \left[ \left\| \frac{\hat{\sigma}_{ii} \hat{\sigma}_{jj}}{\hat{\sigma}_{ii} \hat{\sigma}_{jj}} - 1 \right\| \geq 3C_1 \tau' \right]
\]

\[
\leq P \left[ \left\| \frac{\hat{\sigma}_{ii} \hat{\sigma}_{jj}}{\hat{\sigma}_{ii} \hat{\sigma}_{jj}} - 1 \right\| \geq 3C_1 \tau', \left\| \frac{\hat{\sigma}_{ii}}{\hat{\sigma}_{ii}} - 1 \right\| \leq C_1 \tau', \left\| \frac{\hat{\sigma}_{jj}}{\hat{\sigma}_{jj}} - 1 \right\| \leq C_1 \tau' \right]
\]

\[
+ P \left[ \left\| \frac{\hat{\sigma}_{ii}}{\hat{\sigma}_{ii}} - 1 \right\| \geq C_1 \tau' \text{ or } \left\| \frac{\hat{\sigma}_{jj}}{\hat{\sigma}_{jj}} - 1 \right\| \geq C_1 \tau' \right]
\]

\[
\leq P \left[ \left\| \frac{\hat{\sigma}_{ii} \hat{\sigma}_{jj}}{\hat{\sigma}_{ii} \hat{\sigma}_{jj}} - 1 \right\| \geq C_1 \tau' + \left\| \frac{\hat{\sigma}_{ii}}{\hat{\sigma}_{ii}} - 1 \right\| + \left\| \frac{\hat{\sigma}_{jj}}{\hat{\sigma}_{jj}} - 1 \right\| \right] + O(p_n^{-M_1})
\]

\[
\leq P \left[ \left\| \frac{\hat{\sigma}_{ii} \hat{\sigma}_{jj}}{\hat{\sigma}_{ii} \hat{\sigma}_{jj}} - 1 \right\| \geq C_1 \tau'^2 + \left\| \frac{\hat{\sigma}_{ii}}{\hat{\sigma}_{ii}} - 1 \right\| + \left\| \frac{\hat{\sigma}_{jj}}{\hat{\sigma}_{jj}} - 1 \right\| \right] + O(p_n^{-M_1})
\]

(50)

\[= O(p_n^{-M_1}).\]
Then,
\[
P \left[ \max_{1 \leq i,j \leq p} |\hat{\rho}_{ij} - \rho_{ij}| \geq 4C_1 \tau' \right] \\
\leq P \left[ \max_{1 \leq i,j \leq p} \left| \frac{\hat{\sigma}_{ij}}{\sqrt{\hat{\sigma}_{ii}\hat{\sigma}_{jj}}} - \frac{\hat{\sigma}_{ij}}{\sqrt{\hat{\sigma}_{ii}\hat{\sigma}_{jj}}} \right| + \max_{1 \leq i,j \leq p} \left| \frac{\hat{\sigma}_{ij} - \sigma_{ij}}{\sqrt{\hat{\sigma}_{ii}\hat{\sigma}_{jj}}} \right| \geq 4C_1 \tau' \right] \\
\leq P \left[ \max_{1 \leq i,j \leq p} \left( \frac{\hat{\sigma}_{ij}}{\sqrt{\hat{\sigma}_{ii}\hat{\sigma}_{jj}}} - 1 \right) \geq 3C_1 \tau' \right] \\
+ P \left[ \max_{1 \leq i,j \leq p} \left| \frac{\hat{\sigma}_{ij} - \sigma_{ij}}{\sqrt{\hat{\sigma}_{ii}\hat{\sigma}_{jj}}} \right| \geq C_1 \tau' \right] \\
\leq P \left[ \max_{1 \leq i,j \leq p} \left| \frac{\hat{\sigma}_{ij}}{\sqrt{\hat{\sigma}_{ii}\hat{\sigma}_{jj}}} - 1 \right| \geq 3C_1 \tau' \right] + P \left[ \max_{1 \leq i,j \leq p} \left| \frac{\hat{\sigma}_{ij} - \sigma_{ij}}{\sqrt{\hat{\sigma}_{ii}\hat{\sigma}_{jj}}} \right| \geq C_1 \tau' \right] \\
= O(p_n^{1-M_1+2}),
\]
following from (47) and (50). Equation (45) holds by Letting $C' = 4C_1$ and $M' = M_1 - 2 > 0$. Then following similar lines of the proof of Theorem 1 after equation (38), it is easy to see that $\|S_\tau(\hat{R}) - R\|_F^4$, corresponding to (41) for $R$, is bounded by $16p^4$, which completes the proof. \hfill \square

**Proofs of Theorem 2 and Corollary 2.** The proofs follow the same lines of the proof of Theorem 2 in Rothman, Levina and Zhu (2009) by replacing their equation (A.4) with equations (38) and (45), respectively. Details are hence omitted. \hfill \square

**Proof of Theorem 3.** We first establish the following inequalities:
\[
\|B_k(\hat{\Sigma}) - \Sigma\|_2 \leq \|B_k(\hat{\Sigma}) - \Sigma\|_1 \\
\leq \|B_k(\Sigma) - \Sigma\|_1 + \|B_k(\hat{\Sigma}) - B_k(\Sigma)\|_1 \\
\leq \max_j \sum_{i:|i-j|>k} |\sigma_{ij}| + (2k + 1) \max_{i,j} |\hat{\sigma}_{ij} - \sigma_{ij}|,
\]
and

\[
\frac{1}{p} \| B_k(\hat{\Sigma}) - \Sigma \|_F^2 \\
\leq \max_j \sum_i (\hat{\sigma}_{ij} \mathbf{1}(|i - j| \leq k) - \sigma_{ij})^2 \\
\leq \max_j \sum_i (|\hat{\sigma}_{ij} \mathbf{1}(|i - j| \leq k) - \sigma_{ij} \mathbf{1}(|i - j| \leq k)| + |\sigma_{ij} \mathbf{1}(|i - j| \leq k) - \sigma_{ij}|)^2 \\
\leq (2k + 1) \max_{i,j} |\hat{\sigma}_{ij} - \sigma_{ij}|^2 + 2 \left( \max_{i,j} |\hat{\sigma}_{ij} - \sigma_{ij}| \right) \left( \max_j \sum_{i:|i-j|>k} |\sigma_{ij}| \right) \\
+ \max_j \sum_{i:|i-j|>k} |\sigma_{ij}|^2.
\]

Since

\[
\max_j \sum_{i:|i-j|>k} |\sigma_{ij}| \leq C_0 k^{-\beta},
\]

by (38) we have for any constant $M_1 > 0$ there exists a constant $C_1 > 0$ such that

\[
P \left[ \| B_k(\hat{\Sigma}) - \Sigma \|_2 \geq C_1 (k^{-\beta} + k\tau') \right] = O(p_n^{-M_1}),
\]

\[
P \left[ \frac{1}{p} \| B_k(\hat{\Sigma}) - \Sigma \|_F^2 \geq C_1 (k\tau'^2 + k^{-\beta} \tau' + k^{-2\beta}) \right] = O(p_n^{-M_1}),
\]

and under the additional assumption that $|\sigma_{ij}| \leq C|i - j|^{-(\beta+1)}$ for all $i,j : |i - j| \geq 1$,

\[
P \left[ \frac{1}{p} \| B_k(\hat{\Sigma}) - \Sigma \|_F^2 \geq C_1 (k\tau'^2 + k^{-\beta} \tau' + k^{-2\beta-1}) \right] = O(p_n^{-M_1}).
\]

Setting $k \asymp \tau'^{-1/(\beta+1)}$, we obtain (11) and (12).

For the convergence in mean square, we only consider the case under spectral norm. The proof for Frobenius norm follows similarly. Similar to (41), for any constant $M_1 > 0$ there exists a constant $c_1 > 0$ such that

\[
E \| B_k(\hat{\Sigma}) - \Sigma \|_2^2 \leq c_1 \tau'^{(2\beta/\beta+1)} + (E \| B_k(\hat{\Sigma}) - \Sigma \|_F^4)^{1/2} \quad O(p_n^{-M_1/2}).
\]

Similarly to the argument for (44), by Isserlis’ theorem [Isserlis (1918)], there exists a constant $c_2 > 0$ independent of $M_1$ such that

\[
E \| B_k(\hat{\Sigma}) - \Sigma \|_F^4 = O(p_n^{c_2}).
\]
Plugging (52) into (51) yields
\[
E \| B_k(\hat{\Sigma}) - \Sigma \|^2_2 \leq c_1 \tau'^{(2\beta/(\beta + 1))} + O(p_n^{(c_2-M_1)/2}).
\]
By \(1 \leq f = O(n g^2 \log p_n)\), we have \(O(\tau') = O(g \log p_n) = \sqrt{\log p_n / n}\). We can let \(M_1\) be sufficiently large so that \(p_n^{(c_2-M_1)/2} = O(\tau'^{(2\beta/(\beta + 1))})\), we then obtain \(E \| B_k(\hat{\Sigma}) - \Sigma \|^2_2 = O(\tau'^{(2\beta/(\beta + 1))})\).

**Proof of Corollary 3.** The proof follows the same lines of the proof of Theorem 3 using equation (45) and the inequality \(E \| B_k(\hat{R}) - R \|^4_F \leq 16p^4\).

**Proof of Theorem 4.** We only need to find appropriate \(f\) and \(g\). In all theorems and corollaries for the general temporal dependence, we assume \(f = O(n g^2 \log p_n)\). To obtain an easy solution of \(f\) satisfying this assumption, we set \(f \sim C_{f,n} n g^2 \log p_n\) with
\[
C_{f,n} = \frac{1}{4C^2} \times \begin{cases} 
\left[ \frac{1-\alpha (f/n)^{1-\alpha}}{1-\alpha} \right]^{-2}, & 0 < \alpha < 1, \\
\left[ 1 + \log(n/f) \right]^{-2}, & \alpha = 1, \\
\left[ \frac{\alpha - (f/n)^{\alpha - 1}}{\alpha - 1} \right]^{-2}, & \alpha > 1.
\end{cases}
\]
It is easy to see that \(C_{f,n} \leq \frac{1}{4C^2}\). Next, we set \(g\) to satisfy (3). Then if \(0 < \alpha < 1\),
\[
f \sim \frac{1}{4C^2} \left[ \frac{1-\alpha (f/n)^{1-\alpha}}{1-\alpha} \right]^{-2} n g^2 \log p_n \\
= (n \log p_n) \left[ \frac{(n/f)^{1-\alpha} - \alpha}{(1-\alpha) f^\alpha} \right]^2 \left[ \frac{1-\alpha (f/n)^{1-\alpha}}{1-\alpha} \right]^{-2} \\
= (n \log p_n) \left[ \frac{(n/f)^{1-\alpha} - \alpha}{f^\alpha} \frac{1-\alpha (f/n)^{1-\alpha}}{1-\alpha} \right]^2 \left[ \frac{1-\alpha (f/n)^{1-\alpha}}{1-\alpha} \right]^{-2} \\
= f^{-2} n^{3-2\alpha} \log p_n.
\]
If \(\alpha > 1\),
\[
f \sim \frac{1}{4C^2} \left[ \frac{\alpha - (f/n)^{\alpha-1}}{\alpha - 1} \right]^{-2} n g^2 \log p_n \\
= (n \log p_n) \left[ f^{-\alpha} \frac{(n/f)^{1-\alpha} - \alpha}{1-\alpha} \right]^2 \left[ \frac{\alpha - (f/n)^{\alpha-1}}{\alpha - 1} \right]^{-2} \\
= f^{-2\alpha} n \log p_n.
\]
If $\alpha = 1$,

$$f \sim (n \log p_n) f^{-2} (1 + \log(n/f))^2 [1 + \log(n/f)]^{-2} = (n \log p_n) f^{-2}. $$

Thus,

$$f \sim \begin{cases} 
  n^{1-2\alpha/3} (\log p_n)^{1/3}, & 0 < \alpha < 1, \\
  (n \log p_n)^{1/(\alpha+1)}, & \alpha \geq 1. 
\end{cases}$$

Acknowledgements. The rfMRI data were provided by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657), which is funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research, and by the McDonnell Center for Systems Neuroscience at Washington University.

References.

Bai, Z. D. and Yin, Y. Q. (1993). Limit of the smallest eigenvalue of a large-dimensional sample covariance matrix. *Ann. Probab.* 21 1275–1294. MR1235416

Benjamini, Y. and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Ann. Statist.* 29 1165–1188. MR1869245

Bhattacharjee, M. and Bose, A. (2014). Consistency of large dimensional sample covariance matrix under weak dependence. *Stat. Methodol.* 20 11–26. MR3205718

Bickel, P. J. and Levina, E. (2008a). Regularized estimation of large covariance matrices. *Ann. Statist.* 36 199–227. MR2387969

Bickel, P. J. and Levina, E. (2008b). Covariance regularization by thresholding. *Ann. Statist.* 36 2577–2604. MR2485008

Bradley, R. C. (2005). Basic properties of strong mixing conditions. A survey and some open questions. Update of, and a supplement to, the 1986 original. *Probab. Surv.* 2 107–144. MR2178042

Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods*, 2nd ed. Springer-Verlag, New York. MR1093459

Cai, T. T. and Liu, W. D. (2011). Adaptive thresholding for sparse covariance matrix estimation. *J. Amer. Statist. Assoc.* 106 672–684. MR2847949

Cai, T. T. and Yuan, M. (2012). Adaptive covariance matrix estimation through block thresholding. *Ann. Statist.* 40 2014–2042. MR3059075

Cai, T. T., Zhang, C. H. and Zhou, H. H. (2010). Optimal rates of convergence for covariance matrix estimation. *Ann. Statist.* 38 2118–2144. MR2676885

Cao, G., Bachega, L. R. and Bouman, C. A. (2011). The sparse matrix transform for covariance estimation and analysis of high dimensional signals. *IEEE Trans. Image Process.* 20 625–640. MR2799176

Chen, X., Xu, M. and Wu, W. B. (2013). Covariance and precision matrix estimation for high-dimensional time series. *Ann. Statist.* 41 2994–3021. MR3161455

El Karoui, N. (2008). Operator norm consistent estimation of large-dimensional sparse covariance matrices. *Ann. Statist.* 36 2717–2756. MR2485011
FAN, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc.* 96 1348–1360. MR1946581

FANG, Y. X., WANG, B. H. and FENG, Y. (2013). Tuning parameter selection in regularized estimations of large covariance matrices. Preprint. Available at arXiv: 1308.3416.

GLASSER, M. F., SOTIROPOULOS, S. N., WILSON, J. A., COALSON, T. S., FISCHL, B., ANDERSSON, J. L., XU, J., JBAADI, S., WEBSTER, M., POLIMENI, J. R., VAN ESSEN, D. C. and JENKINSON, M. (2013). The minimal preprocessing pipelines for the Human Connectome Project. *NeuroImage* 80 105–124.

GOLUB, G. H. and VAN LOAN, C. F. (1996). *Matrix Computations*, 3rd ed. Johns Hopkins University Press, Baltimore, MD. MR1417720

HAMILTON, J. D. (1994). *Time Series Analysis*. Princeton University Press, Princeton, NJ. MR1278033

HOSKING, J. R. M. (2008). Fractional differencing. *Biometrika* 88 165–176. MR0614953

HU, T. C., ROSALSKY, A. and VOLODIN, A. (2008). On convergence properties of sums of dependent random variables under second moment and covariance restrictions. *Statist. Probab. Lett.* 78 1999–2005. MR2458009

HUANG, J. Z., LIU, N., POURAHMADI, M. and LIU, L. (2006). Covariance matrix selection and estimation via penalised normal likelihood. *Biometrika* 93 85–98. MR2277742

ISSERLIS, L. (1918). On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables. *Biometrika* 12 134–139.

JIANG, T., (2003). The limiting distributions of eigenvalues of sample correlation matrices. *Sankhyā* A 66 35–48.

KARR, A. F. (1993). *Probability*. Springer-Verlag, New York. MR1231974

LAM, C. and FAN, J. (2009). Sparsistency and rates of convergence in large covariance matrix estimation. *Ann. Statist.* 37 4254–4278. MR2572459

LAURENT, B. and MASSART, P. (2000). Adaptive estimation of a quadratic functional by model selection. *Ann. Statist.* 28 1302–1338. MR1805785

POWER, J. D., COHEN, A. L., NELSON, S. M., WIG, G. S., BARNES, K. A., CHURCH, J. A., VOGEL, A. C., LAUMANN, T. O., MIEZIN, F. M., SCHLAGGAR, B. L. and PETERSEN, S. E. (2011). Functional network organization of the human brain. *Neuron* 72 665–678.

PRIESTLEY, M. B., and SUBBA RAO, T., (1969). A test for non-stationarity of time-series. *J. Roy. Statist. Soc. Ser. B* 31 140–149. MR0269062

RACINE, J. (2000). Consistent cross-validation model-selection for dependent data: h-block cross-validation. *Journal of Econometrics* 99 39–61.

ROTHMAN, A. J., LEVINA, E. and ZHU, J. (2009). Generalized thresholding of large covariance matrices. *J. Amer. Statist. Assoc.* 104 177–186. MR2504372

ROTHMAN, A. J., LEVINA, E. and ZHU, J. (2010). A new approach to Cholesky-based covariance regularization in high dimensions. *Biometrika* 97 539–550. MR2672482

SMITH, S. M., BECHMANN, C. F., ANDERSSON, J., AUERBACH, E. J., BLIJBERGSTOCH, J., DOAUD, G., DUFF, E., FEINBERG, D. A., GRIFFANTI, L., HARMS, M. P., KELLY, M., LAUMANN, T., MILLER, K. L., MOELLER, S., PETERSEN, S., POWER, J., SALIMI-KHORSHIDI, G., SNYDER, A. Z., VU, A. T., WOOLRICH, M. W., XU, J., YACOUB, E., UGBRIL, K., ESSEN, D. C. V. and GLASSER, M. F. (2013). Resting-state fMRI in the Human Connectome Project. *NeuroImage* 80 144–168.

SRIPADA, C., ANGSTSTADT, M., KESSLER, D., PHAN, K. L., LIBERZON, I., EVANS, G. W., WELSH, R. C., KIM, P. and SWAIN, J. E. (2014a). Volitional regulation of emotions produces distributed alterations in connectivity between visual, attention control, and default networks. *NeuroImage* 89 110–121.

SRIPADA, C., KESSLER, D., FANG, Y., WELSH, R. C., KUMAR, K. P., and
Angstadt, M. (2014b). Disrupted network architecture of the resting brain in attention-deficit/hyperactivity disorder. *Human Brain Mapping* **35** 4693–4705.

Syed, M. N., Principe, J. C. and Pardalos, P. M. (2012). Correntropy in data classification. *Dynamics of information systems: mathematical foundations, Springer Proc. Math. Stat.* **20** 81–117. MR3067311

Wagaman, A. S. and Levina, E. (2009). Discovering sparse covariance structures with the isomap. *J. Comput. Graph. Statist.* **18** 551–572. MR2751641

Watanabe, T., Kessler, D., Scott, C., Angstadt, M. and Sripada, C. (2014). Disease prediction based on functional connectomes using a scalable and spatially-informed support vector machine. *NeuroImage* **96** 183–202.

Wu, W. B. (2005). Nonlinear system theory: another look at dependence. *Proc. Natl. Acad. Sci. USA* **102** 14150–14154. MR2172215

Wu, W. B. and Pourahmadi, M. (2009). Banding sample autocovariance matrices of stationary processes. *Statist. Sinica* **19** 1755–1768. MR2589209

Zhou, S. (2014). Gemini: graph estimation with matrix variate normal instances. *Ann. Statist.* **42** 532–562, MR3210978

Zou, H. (2006). The adaptive lasso and its oracle properties. *J. Amer. Statist. Assoc.* **101** 1418–1429. MR2279460

Department of Biostatistics
University of Michigan
1415 Washington Heights
Ann Arbor, Michigan 48109-2029
USA
E-mail: haishu@umich.edu
bnan@umich.edu