ON THE SPARSITY OF SIGNALS IN A RANDOM SAMPLE

JIANG BINYAN

NATIONAL UNIVERSITY OF SINGAPORE

2011

ON THE SPARSITY OF SIGNALS IN A RANDOM SAMPLE

JIANG BINYAN

(B.Sc. University of Science and Technology of China)

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF STATISTICS AND APPLIED PROBABILITY

NATIONAL UNIVERSITY OF SINGAPORE

2011

ACKNOWLEDGEMENTS

I am so grateful that I have Professor Loh Wei-Liem as my supervisor. He is truly a great mentor not only in statistics but also in daily life. I would like to thank him for his guidance, encouragement, time, and endless patience. Next, I would like to thank my senior Li Mengxin and Wang Daqing for discussion on various topics in research. I also thank all my friends who helped me to make life easier as a graduate student. I wish to express my gratitude to the university and the department for supporting me through NUS Graduate Research Scholarship. Finally, I will thank my family for their love and support.

CONTENTS

SUMMARY

The "large p small n" data sets are frequently encountered by various researchers during the past decades. One of the commonly used assumptions for these data sets is that the data set is sparse. Various methods have been developed in dealing with model selection, signal detection or large covariance matrix estimation. However, as far as we know, the problem of estimating the "sparsity" has not been addressed thoroughly yet. Here loosely speaking, sparsity is interpreted as the proportion of parameters taking the value 0.

Our work in this thesis contains two parts. The first part (Chapter 2) deals with estimating the sparsity of a sparse random sequence. An estimator is constructed from a sample analog of certain Hermitian trigonometric matrices. To evaluate our estimator, upper and lower bounds for the minimax convergence rate are derived.

Simulation studies show that our estimator performs well.

The second part (Chapter 3) deals with estimating the sparsity of a large covariance matrix or correlation matrix. This to some degree is related to the problem of finding a universal data-dependent threshold for the elements of a sample correlation matrix. We propose two estimators $\hat{\omega}_1$ and $\hat{\omega}_2$ based on different methods. $\hat{\omega}_1$ is derived assuming that the observations $X_1, ..., X_n$ are n independent random samples from a multivariate normal distribution with mean $\mathbf{0}_p$ and unknown population matrix $\Sigma = (\sigma_{ij})_{p \times p}$. In contrast, $\hat{\omega}_2$ is derived under more general (possibly non-Gaussian) assumptions on the distribution of observations $X_1, ..., X_n$. Consistency of these two estimators are proved under mild conditions. Simulation studies are carried out with a comparison to thresholding estimators derived from cross validation and adaptive cross validation methods.

LIST Of NOTATIONS

$\mathbf{0}_p$	$p \times 1$ vector such that all elements are zero.
$\mathbb{R}^d$	d-dimensional Euclidean space
$\mathbb{C}^d$	d-dimensional complex space
M'	transpose of a matrix M
$a \vee b$	maximum of a and b, where $a, b \in \mathbb{R}$
$a \wedge b$	minimum of a and b, where $a, b \in \mathbb{R}$
$\lfloor \cdot \rfloor$	$\lfloor x \rfloor$ denotes the largest integer less than or equal to $x \in \mathbb{R}$
$I_{\{\cdot\}}$	indicator function
$\mathbf{i}$	$\sqrt{-1}$

List of Tables

CHAPTER 1 Introduction

High dimension, low sample size (HDLSS) data sets are frequently encountered nowadays in many different fields. However it is well known that the statistical analysis of HDLSS data is very challenging and possibly intractable in some instances. Fortunately in many situations, the data can be assumed to have some particular structures. One of the commonly used assumptions of HDLSS data is sparesness, and under this assumption, accurate statistical inference becomes feasible. There are a lot of interesting problems in sparse HDLSS data analysis, and here we mainly focus on two of these problems: (i) sparse signal detection and (ii) sparse covariance selection.

For sparse signal detection problem, the sequence of observations $X_1, \ldots, X_n$ is usually modeled as $X_i = \Theta_i + Z_i$, where $\Theta_1, ..., \Theta_n$ is an unobservable signal sequence and $Z_1, ..., Z_n$ is a sequence of noise. The objective of this problem is to estimate the unobservable sparse signal sequence $\Theta_1, ..., \Theta_n$. For example, Johnstone and Silverman (2004) considered the estimation of sparse sequences observed in Gaussian white noise. More precisely, the Z_i's are $N(0, 1)$ random variables independent of the Θ_i's, and that the Θ_i's are sparse is modeled by using the prior mixture density for Θ_i: $f_{\text{prior}}(\theta) = \omega_0 \delta_0 + (1 - \omega_0)h(\theta)$ where $\omega_0 \in (0, 1]$ is a constant, δ_0 denotes point mass at 0 and h is a density function. Sparsity is now quantified by ω_0, which is the proportion of θ_i's that are zero when $n \to \infty$. Instead of finding estimators for the unobservable signal sequence, in this thesis we are more interested in answering a relatively basic question: "How sparse is the unobservable signal sequence (meaning how many of the θ_i's are 0)?" Or equivalently, we are aiming at estimating ω_0. Johnstone and Silverman (2004) used the posterior median to estimate the signal sequence. Although the signal sequence can be estimated quite well, according to our simulations, the resulting estimator is usually not able to estimate ω_0 well unless ω_0 is close to 1. In fact, the problem of estimating ω_0 has not been addressed a lot in the literature we have covered. In addition, in the literature, $Z_1, ..., Z_n$ are usually assumed to be normally distributed. It would be practically important to study the problem under more general noise distributions.

The second problem is related to sparse covariance matrix estimation. The problem of estimating a large sparse covariance matrix has generated much interest in recent years. Here the literature is huge. This includes El Karoui (2008), Bickel and Levina (2008a, b), Lam and Fan (2009), Cai and Liu (2011) and the references cited therein. In this thesis, we aim at estimating the sparsity of a large population covariance or correlation matrix. As far as we know, this problem has not been studied directly yet. One immediate application of a good sparsity estimator is in choosing the thresholding parameter for thresholding estimators [e.g. Bickel and Levina (2008a, b), Cai and Liu (2011)]. More precisely, an important problem in thresholding methods is to find data-dependent thresholds. However, there are still some problems in the existing methods for finding the thresholds. For example, Bickel and Levina (2008b) used cross validation in finding a data-dependent universal threshold while Cai and Liu (2011) proposed an adaptive thresholding method which adapts heteroscedastic noise. However, cross validation and adaptive cross validation methods are computationally intensive and tend to over-threshold according to our simulations. Another approach in finding thresholds for the elements of a sample covariance matrix where the noise may be heteroscedastic is to find a universal threshold for the sample correlation matrix. However, as far as we know, there is not enough study on this. On the other hand, given a good sparsity estimator, we can find a universal threshold for the elements of a sample correlation matrix such that the sparsity of the resulting thresholded sample correlation matrix equals to the estimated sparsity. In summary, we are aiming at addressing the question "How sparse is a large covariance matrix?". Intuitively, if we can estimate the sparsity well, the corresponding data-dependent thresholds for the covariance matrix could perform well in estimating the true covariance structure.

To conclude this subsection, the problems we study in this thesis are (i) to estimate the sparsity of a sparse random sequence and (ii) to estimate the sparsity of a large sparse covariance matrix. Here, loosely speaking, sparsity is interpreted as the proportion of parameters taking the value 0. In Section 1.1, the literature on estimating a sparse signal sequence will be reviewed. In Section 1.2, some popular methods used in estimating a large sparse covariance matrix will be discussed.

1.1 Signal detection

Signal activity detection is a critical stage in many research fields. The objective of signal detection is to determine the presence or absence of a signal embedded in additive noise. More precisely, we have a sequence of observations $X_1, ..., X_n$, which is usually modeled as $X_i = \theta_i + Z_i, i = 1, ..., n$. Here $\theta_1, .., \theta_n$ is the unobservable signal sequence and $Z_i, ..., Z_n$ is a sequence of noise. The objective is to estimate the positions of those non-zero θ_i's. The unobservable sequence $\theta_1, ..., \theta_n$ is usually assumed to be sparse, in that a number of θ_i's are identically 0.

Next we review some approaches in solving this problem.

• **Multiple hypothesis testing**. This is one of the popular approaches. The problem of determining the presence or absence of a signal is treated as a Hypothesis-Testing problem:

$$H_0 : \theta_i = 0 \quad v.s. \quad H_1 : \theta_i \neq 0 \quad , i = 1, ..., n.$$

Here the literature is huge. This includes Abramovich and Benjamini (1995), Donoho and Jin (2004), Hall and Jin (2010) and the references cited therein.

• **SURE**. Donoho and Johnstone (1995) derived estimators for the sparse signal sequence by minimizing Stein's unbiased risk estimate for the mean squared error of soft thresholding. However, this method is aiming at estimating the signal sequence and the corresponding sparsity of the estimated signal sequence is usually different from the true sparsity.

• **FDR**. Benjamini and Hochberg (1995) proposed the false discovery rate approach which is derived from the principle of controlling the false discovery rate in simultaneous hypothesis testing. This method also led to a spur of further research such as Benjamini and Yekutieli (2001), Storey (2002) and Chung et. al. (2007). However, for different false discovery rate parameter q, the resulting sparsity of the estimated signal sequence varies.

- **Empirical Bayes approach** Johnstone and Silverman (2004) modeled the unobservable signal sequence Θ_i's using the prior mixture density for Θ_i: $f_{\text{prior}}(.) = \omega_0\delta_0 + (1-\omega_0)h(.)$ where $\omega_0 \in (0,1]$ is a constant, δ_0 denotes point mass at 0 and h is a density function. Sparsity is then quantified by ω_0. Notice that the posterior distribution of Θ_i's are also a mixture of point mass at 0 with some continuous distribution function, by using the posterior median as an estimator for each Θ_i, the resulting estimator of the signal sequence will be sparse. However, they assumed that the signal sequence is very sparse in that ω_0 tends to 0 as n tends to infinity.

Above all, the noise $Z_1, ..., Z_n$ are usually assumed to be independently distributed normal random variables [e.g. Johnstone and Silverman (2004), Lee et. al. (2010)] or normal random variables with known covariance matrix or the covariance matrix can be estimated [e.g. Hall and Jin (2010)]. Another commonly used assumption is that the signal sequence $\theta_1, ..., \theta_n$ is very sparse, in that the proportion of zero θ_i's tends to 1 as n tends to infinity [e.g. Donoho and Jin (2004), Hall and Jin (2010)]. In this thesis, we consider the problem of estimating the sparsity of the signal sequence. Consequently, a natural estimator for the set of nonzero θ_i's can be obtained by thresholding the observation sequence based on the estimator of ω_0. In Chapter 2, we propose a more general model as the prior of Θ_i's [see (2.1)], where the sparsity is quantified by ω_0 similar to that in Johnstone and Silverman (2004) and Lee et. al. (2010). Different from the literature, we assume that $0 < \omega_0 \leq 1$ instead of assuming ω_0 tends to 1; and we assume that the noise distribution may be unknown but there is a sequence of pure noise observations $Y_1, ..., Y_m$. Particularly, the Z_i's may not be normally distributed or independent. To evaluate the performance of our estimator, we also derived lower bounds of the minimax risk for estimating ω_0 when the noise is known. Given a good estimator of sparsity it would be interesting to study the problem of estimating the signal sequence, and hopefully, we can obtain good estimators under mild conditions. However, this is beyond the scope of this thesis and will be treated as future work.

1.2 Covariance selection

Let $X_1, \ldots, X_n$ be independent, identically distributed p-dimensional random vectors with mean $\mathbf{0}_p$, covariance matrix $\Sigma = (\sigma_{ij})_{p\times p}$ and correlation matrix $\Gamma = (\rho_{ij})_{p\times p}$. For definiteness, the sample covariance matrix is denoted by $S = (s_{jk})_{p\times p} = (1/n)\sum_{i=1}^{n} X_i X_i'$, and the sample correlation matrix is denoted as $R = (r_{jk})_{p\times p}$ where $r_{jk} = s_{jk}/\sqrt{s_{jj}s_{kk}}$ and $X_i = (X_{1i}, \ldots, X_{pi})'$.

Given observations $X_1, \ldots, X_n$ or S, the problem of estimating the population covariance matrix Σ occurs naturally in many statistical problems that arise in various scientific applications. During the past decades, the "large p small n" data sets are frequently encountered by various researchers and sometimes the estimation problem involves the case where $n < p$. The usual estimator for the covariance matrix Σ is the sample covariance matrix S, where S is distributed according to the Wishart distribution $W_p(\Sigma, n)$. Although S is unbiased, it is known that:

i) The sample eigenvalues of S tend to be more spread out than the population eigenvalues, unless $p/n \to 0$;

ii) S is singular when $n < p$.

Many works have been done to construct better estimators either for the covariance matrix or the concentration matrix. One of the problems people try to solve is i) mentioned above. Stein (1975) proved the "Wishart identity" (also proved independently by Haff (1977)), and proposed a non-asymptotic approach in estimating the covariance matrix, where the eigenvalues of the sample covariance matrix are shrunk. Extension to estimating two covariance matrices based on a similar non-asymptotic approach can be found in Loh (1988) and Loh (1991). A Monte Carlo study of Stein's estimator with comparison to other estimators can be found in Lin and Perlman (1985). Dey and Srinivasan (1986) constructed a class of minimax estimators for Σ, which shrink or expand the sample eigenvalues depending on their magnitudes. However, both Stein's estimator and Dey and Srinivasan's estimator do not preserve the order of eigenvalues and the resulting estimators of the eigenvalues can be negative. Haff (1991) derived an estimator similar to Stein's but was computed under the constraint of maintaining the order of the sample eigenvalues. There are also some authors who estimate covariance matrices from a Bayes perspective. The idea is to specify an appropriate prior for the population covariance matrix and choose a (shrinkage) estimator based on a particular loss function. Yang and Berger (1994) developed the reference non-informative prior for a covariance matrix and obtained expressions for the resulting Bayes estimators, which are comparable to Stein's (1975) and Haff's (1991) estimators. Later, Kass (2001) suggested placing normal prior distributions on the logarithm of the eigenvalues and obtained a shinkage estimator for the covariance matrix.

The other case, which is also the main concern of this thesis, is the case when p and n are both very large, including the case $n < p$. Since the dimension of parameters ($p(p+1)/2$) can be very large relative to the sample size, the problem of estimating a covariance matrix becomes much more difficult. Fortunately, the covariance matrix or concentration matrix is usually believed and assumed to have some structures, such as ordering between variables and sparseness. The shrinkage estimators discussed above are not applicable to the $n < p$ case since the sample covariance matrix is no longer positive definite. Ledoit and Wolf (2004) proposed a well-conditioned shrinkage estimator which is applicable to the case $n < p$. Their estimator is of the form:

$$\Sigma^* = \rho_1 I + \rho_2 S,$$

such that it minimizes the risk with respect to the following loss function:

$$L(\Sigma^*, \Sigma) = tr(\Sigma^* - \Sigma)(\Sigma^* - \Sigma)'/p.$$

However, when the covariance matrix is believed or assumed to be sparse, this estimator does not seem appropriate as the elements of the estimator equal 0 with probability 0. To estimate a large but sparse covariance matrix, we found that there are basically three different approaches in recent literature: penalized likelihood approach, Bayesian approach and thresholding approach.

i) **Penalized likelihood approach**. Estimators are obtained by minimizing the penalized negative normal likelihood for the population covariance matrix or concentration matrix or their corresponding Cholesky factors. Huang et. al. (2006) used LASSO on the off-diagonal elements of the Cholesky factor from the modified Cholesky decomposition. Yuan and Lin (2007) used LASSO for estimating the concentration matrix in the Gaussian graphical model, subjected to the positive definite constraint. Based on the penalized likelihood with L_1 penalty on the off-diagonal elements of the concentration matrix, Friedman et. al. (2008) proposed a simple and fast algorithm for the estimation of a sparse concentration matrix, and Rothman et. al. (2008) obtained the rate of convergence under the Frobenius norm. Lam and Fan (2009) studied not only the LASSO penalty but also other non-convex penalties such as SCAD and hard-thresholding penalty, and obtained explicit rates of convergence.

ii) **Bayesian approach**. As far as we know, there has not been much research done on estimating large sparse covariance matrices using Bayes methods. Wong et. al. (2003) used a prior for the partial correlation matrix that allows elements of the inverse partial correlation matrix to be zero. The computation was carried out using Markov chain Monte Carlo (MCMC). However, their estimator also does not introduce zeros since they used the mean of samples generated from the posterior using MCMC. Also, the computation can be very time consuming when p is large. Smith and Kohn (2002) introduced a prior that introduces zeros in the off-diagonal elements of the Cholesky factor of the concentration matrix. However, the method can only be applied to longitudinal data, which has a relatively simple structure.

iii) **Thresholding approach**. The idea behind this approach is very natural: when we believe that there are many zeros in the covariance matrix, an estimator could possibly be obtained by thresholding some of the off-diagonal elements of the sample covariance matrix or the correlation matrix that have small magnitude to be zero. Bickel and Levina (2008a, b) proposed estimators by tapering or thresholding sample covariance matrices, and showed that the thresholding estimators are consistent over a class of sparse matrices. Rothman, Levina and Zhu (2009) considered thresholding sample covariance matrices with more general thresholding functions possessing a shrinkage property. El Karoui (2008) studied the thresholding estimators under a special notion of sparsity called $\beta - sparsity$, and showed that $\beta - sparse$ matrices, with $\beta < 1/2$, are consistently estimable in the spectral norm. More recently, Cai and Liu (2011) proposed an adaptive thresholding method in thresholding sample covariance matrices which is applicable when the noise is not homoscedastic.

Among the literature mentioned above, although some of the authors were aiming at obtaining sparse estimators for the population covariance matrix, in the other words, they were doing estimation and covariance selection simultaneously, they did not explore the problem of estimating the sparsity of the population covariance matrix directly. In addition, for the thresholding approach, although the idea of thresholding estimator is very natural, it is difficult to answer the question "How to choose a data-dependent threshold?" Methods for finding a data-dependent thresholding parameter in the literature include cross validation (Bickel and Levina (2008 a, b)), and adaptive cross validation (Cai and Liu (2011)). However, cross validation and adaptive cross validation are computationally intensive and tend to over-threshold according to our simulations. Furthermore, these two methods are not designed to address the question "How sparse is the matrix?" directly, therefore the resulting threshold may not perform well in terms of covariance selection.

In Chapter 3 of this thesis, we aim at estimating the sparsity of the population covariance matrix. If the sparsity of the population covariance matrix can be well estimated, we can estimate the covariance structure by thresholding the sample correlation matrix, which is also adaptive to the heteroscedastic case. More specifically, if ω, the sparsity of the population matrix, can be well estimated by $\hat{\omega}$, we can find the corresponding universal threshold in $[0, 1)$ such that the sparsity of the thresholded sample correlation matrix equals to $\hat{\omega}$. This approach to some degree can also be viewed as a method in finding data-dependent thresholds for the sample covariance matrix.

To model the sparsity of the population covariance matrix, motivated by the signal detection problem, we model the population correlation coefficients using a mixture of a point mass at zero and a distribution function G in [-1,1]:

$$(1 - \omega)dG(\rho) + \omega\delta_0(\rho),$$

where $\delta_0(\rho)$ denotes point mass at $\rho = 0$. Then the problem becomes estimating ω as in Chapter 2.

In this thesis, we study only the problem of estimating the sparsity parameter ω. The problem of estimating the population matrix based on a good estimator of ω will be treated as future work, as it is beyond the scope of this thesis.

CHAPTER 2

Signal Detection

2.1 Introduction

Let $X_1, ..., X_n$ be a random sample of observations. Assume that for each $1 \leq i \leq n$, $X_i = \theta_i + Z_i$ where $Z_1, Z_2, ...,$ are stationary, strong mixing random variables with marginal probability density function f_Z. θ_i and Z_i are commonly regarded as the "signal" and "noise" respectively. f_Z and θ_i's are unknown and we assume that there is an independent sample of pure noise observations $Y_1, ..., Y_m$ having the same joint distribution as $Z_1, ..., Z_m$. If $m = \infty$, then f_Z would be known. We assume that the sequence of X_i's may be sparse in that a number of θ_i's are identically 0 and our objective is to estimate the set $\Xi = \{1 \le i \le n : \theta_i = 0\}$.

Our approach is to first estimate the proportion ω_0 of θ_i's such that $\theta_i = 0$. ω_0 is a measure of the sparsity of the signals from the random sample $X_1, .., X_n$. Once an estimate $\hat{\omega}_0$ for ω_0 is obtained, let k be the integer satisfying $(k-1)/n < \hat{\omega}_0 \le k/n$ and $X_{(1)} \le X_{(2)} \le ... \le X_{(n)}$ be the ordered statistics of $|X_1|, ..., |X_n|$. Then a natural estimate for Ξ is as follow:

$$\hat{\Xi} = \{1 \le i \le n : |X_i| \le X_{(k)}\}.$$

More specifically, let $\Theta_1, ..., \Theta_n$ be independent, identically distributed random variables each with cumulative distribution function

$$F_\Theta(\theta) = \sum_{j=0}^{\nu} \omega_j I_{\{\theta \ge \mu_j\}} + (1 - \sum_{j=0}^{\nu} \omega_j) \int_{-\infty}^{\theta} h(y)dy, \quad \forall \theta \in \mathbb{R}, \tag{2.1}$$

where $\mu_0 = 0$, ν is a non-negative integer, $\omega_0 > \omega_1 \ge \cdots \ge \omega_\nu > 0$ are constants satisfying $\sum_{j=0}^{\nu} \omega_i \le 1$, $\mu_1, ..., \mu_\nu$ are non-zero, distinct constants and h is a probability density function. We assume that the Θ_i's are independent of the Z_j's. Consequently, $X_i = \Theta_i + Z_i$ has the mixture probability density function given by:

$$f_X(x) = \sum_{j=0}^{\nu} \omega_j f_Z(x - \mu_j) + (1 - \sum_{j=0}^{\nu} \omega_j) \int_{\mathbb{R}} f_Z(x - y)h(y)dy, \ \forall x \in \mathbb{R}. \tag{2.2}$$

The mixture density given by (2.2) is very general in that the mixing distribution has possibly both discrete and continuous components. We assume that $\nu, \mu_0, ..., \mu_\nu, \omega_0, ..., \omega_\nu$ and h are unknown, and our target is to estimate ω_0, which is the proportion of θ_i's that are zero when $n \to \infty$.

The rest of this chapter is organized as follows. Section 2.2 introduces a number of trigonometric matrices. Proposition 2.1 provides explicit bounds for their largest eigenvalues. Motivated by these bounds, we propose a method-of-moments estimator for ω_0 in Section 2.3 when f_Z is known. Upper bounds for the expected L1 loss of this estimator are also derived. In Section 2.4, we show that the estimator of Lee, et al. (2010) achieves the same upper bounds as our estimator. In Section 2.5 we derive lower bounds for the minimax risk for estimating ω_0 when f_Z is known. Last but not the least, we generalize our estimator in Section 2.6 to the case when f_Z is unknown but there is an independent sample of pure noise observations.

2.2 Trigonometric moment matrices

Following Li and Loh (2011), for any positive integer q, we define a matrix-valued function $T_q : (-1, 1) \to \mathbb{C}^{(q+1)\times(q+1)}$ by:

$$T_q(x) = \begin{pmatrix} 1 & e^{\mathbf{i}x} & e^{\mathbf{i}2x} & \dots & e^{\mathbf{i}qx} \\ e^{-\mathbf{i}x} & 1 & e^{\mathbf{i}x} & \dots & e^{\mathbf{i}(q-1)x} \\ e^{-\mathbf{i}2x} & e^{-\mathbf{i}x} & 1 & \dots & e^{\mathbf{i}(q-2)x} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ e^{-\mathbf{i}qx} & e^{-\mathbf{i}(q-1)x} & e^{-\mathbf{i}(q-2)x} & \dots & 1 \end{pmatrix}$$

$$= \begin{pmatrix} 1 \\ e^{-\mathbf{i}x} \\ \vdots \\ e^{-\mathbf{i}qx} \end{pmatrix} \begin{pmatrix} 1, & e^{\mathbf{i}x}, & \ldots, & e^{\mathbf{i}qx} \end{pmatrix}.$$

where $\mathbf{i} = \sqrt{-1}$. Further define $M_q = ET_q(\Theta) = M_{q,disc} + M_{q,cont}$, where

$$M_{q,disc} = \omega T_q(0) = \begin{pmatrix} \omega & \cdots & \omega \\ \vdots & \ddots & \vdots \\ \omega & \cdots & \omega \end{pmatrix},$$

$$\begin{aligned} M_{q,cont} &= (1-\omega) \int_{\mathbb{R}} T_q(\theta) h(\theta) d\theta \\ &= (1-\omega) \int_0^{2\pi} T_q(\theta) \sum_{j=-\infty}^{\infty} h(\theta + 2\pi j) d\theta. \end{aligned}$$

Notice that $M_q, M_{q,disc}$ and $M_{q,cont}$ are Hermitian matrices. This implies that all their eigenvalues are real-valued. Let $\lambda_i(A)$ denote the ith largest eigenvalues of A where A is an arbitrary $(q+1) \times (q+1)$ Hermitian matrix. Thus $\lambda_1(M_q) \geq \lambda_2(M_q) \geq \cdots \geq \lambda_{q+1}(M_q)$.

For a function $h(x) : x \to \mathbb{R}$, the essential supremum of h is defined by

$$\text{ess sup}_{x \in \mathbb{R}} h = \inf\{a \in \mathbb{R} : \mu(\{x : h(x) > a\}) = 0\},$$

where $\mu(\cdot)$ is the Lebesgue measure. Similarly, the essential infimum of h is defined by

$$\text{ess inf}_{x \in \mathbb{R}} h = \sup\{b \in \mathbb{R} : \mu(\{x : h(x) < b\}) = 0\}.$$

The proof of Proposition 2.1 can be found in Li and Loh (2011).

Proposition 2.1. *Assume that the cumulative distribution function of* Θ *is given by (2.1). With the above notation, suppose* $\omega_0 > \omega_1 \geq \cdots \geq \omega_\nu > 0$ *and* $\mu_1, ..., \mu_\nu$ *are nonzero constants such that* $\frac{\mu_i - \mu_j}{2\pi} \notin \mathbb{Z}$ *for any* $1 \leq i < j \leq \nu$*. Then writing*

$$\Omega = \sqrt{8 \sum_{0 \leq j < k \leq \nu} \frac{\pi_j \pi_k}{|1 - e^{\mathbf{i}\,(\mu_j - \mu_k)}|^2}},$$

we have

$$\begin{aligned} & (q+1)\omega_0 - \Omega + 2\pi(1 - \sum_{k=0}^{\nu} \omega_k) \left\{ \text{ess inf}_{0 \leq \theta < 2\pi} \sum_{j=-\infty}^{\infty} h(\theta + 2\pi j) \right\} \\ \leq \;\; & \lambda_1(M_q) \leq (q+1)\omega_0 + \Omega + 2\pi(1 - \sum_{k=0}^{\nu} \omega_k) \left\{ \text{ess sup}_{0 \leq \theta < 2\pi} \sum_{j=-\infty}^{\infty} h(\theta + 2\pi j) \right\}. \end{aligned}$$

Remark 2.1. We observe from the proof in Li and Loh (2011) that if $\omega_0 > \omega_1 \geq \cdots \geq \omega_\nu > 0$ is not satisfied, the above inequality will become

$$\begin{aligned} & (q+1) \max_{0 \leq k \leq \nu} \omega_k - \Omega + 2\pi(1 - \sum_{k=0}^{\nu} \omega_k) \left\{ \text{ess inf}_{0 \leq \theta < 2\pi} \sum_{j=-\infty}^{\infty} h(\theta + 2\pi j) \right\} \\ & \leq \lambda_1(M_q) \leq \\ & (q+1) \max_{0 \leq k \leq \nu} \omega_k + \Omega + 2\pi(1 - \sum_{k=0}^{\nu} \omega_k) \left\{ \text{ess sup}_{0 \leq \theta < 2\pi} \sum_{j=-\infty}^{\infty} h(\theta + 2\pi j) \right\}, \end{aligned}$$

and our estimator $\hat{\omega}_0$ defined in this chapter will become an estimator for $\max_{0 \leq k \leq \nu} \omega_k$.

The following is an immediate corollary of Proposition 2.1.

Corollary 2.1. *Suppose that* $\text{ess sup}_{0 \leq \theta < 2\pi} \sum_{j=-\infty}^{\infty} h(\theta + 2\pi j) < \infty$*. Then under assumptions of Proposition 2.1,* $\frac{\lambda_1(M_q)}{q} \to \omega_0$ *as* $q \to \infty$*.*

Corollary 2.1 gives, at least in principle, a way for estimating ω_0 by estimating the largest eigenvalue of M_q for a sufficiently large q.

2.3 A method-of-moments estimator when f_Z is known

In this section we assume that the probability density function f_Z of the noise distribution is known. Let $X_1, ..., X_n$ be as in the introduction. Since $X_i = \Theta_i + Z_i$ and that Θ_i and Z_i are independent, we have

$$E(e^{-\mathbf{i}k\Theta_1}) = E(e^{-\mathbf{i}kX_1})[E(e^{-\mathbf{i}kZ_1})]^{-1}, \quad \forall k \in \mathbb{Z},$$

provided the right hand side is well defined. Recall that $Z_1, Z_2, ...,$ are stationary, strongly mixing, mean-zero random variables. For integers $1 \leq a \leq b$, let $\mathscr{F}_a^b = \sigma(Z_i, a \leq i \leq b)$ denote the $\sigma - field$ generated by $\{Z_i, a \leq i \leq b\}$. Define for all $k, l \geq 1$,

$$\alpha(\mathscr{F}_1^{k+1}, \mathscr{F}_{k+l+1}^{\infty}) = \sup_{A \in \mathscr{F}_1^{k+1}, B \in \mathscr{F}_{k+l+1}^{\infty}} |P(A \cap B) - P(A)P(B)|,$$

$$\alpha(l) = \sup_{k \geq 1} \alpha(\mathscr{F}_1^{k+1}, \mathscr{F}_{k+l+1}^{\infty}). \tag{2.3}$$

From the definition of strong mixing, we have $\alpha(l) \to 0$ as $l \to \infty$. Now let q be a positive integer depending only on n. Since $M_q = E[T_q(\theta)]$, we estimate M_q using the sample analog, that is, the $(q+1)\times(q+1)$ matrix $\hat{M}_q$ whose (j,k)th element is given by

$$(\hat{M}_q)_{jk} = \frac{n^{-1}\sum_{i=1}^{n} e^{-\mathbf{i}(j-k)X_i}}{E(e^{-\mathbf{i}(j-k)Z_1})}, \quad \forall 1 \le j,k \le q+1.$$

$\hat{M}_q$ is a Hermitian matrix and hence its eigenvalues are real numbers. We propose as an estimator of ω_0,

$$\hat{\omega}_0 = \begin{cases} 1, & if\ (\lambda_1(\hat{M}_q)-1)/q > 1, \\ (\lambda_1(\hat{M}_q)-1)/q, & if\ 0 \le (\lambda_1(\hat{M}_q)-1)/q \le 1, \\ 0, & if\ (\lambda_1(\hat{M}_q)-1)/q < 0, \end{cases} \tag{2.4}$$

for a sufficiently large integer q. $\hat{\omega}_0$ can be regarded as a bias corrected version of the naive estimator $(\lambda_1(\hat{M}_q)/(q+1)\ \wedge\ 1)$ in the following way. Suppose $\nu = 0$ and h is the density of the uniform distribution on $[a, a+2\pi k)$ for some constant a and integer $k \ne 0$. Then it follows from Proposition 2.1 that $\omega_0 = (\lambda_1(M_q)-1)/q$ and $\hat{\omega}_0$ is an estimate of it.

Following Fan (1991), we call f_Z *supersmooth* of order β if its Fourier transform $\varphi_{f_Z}(t) = \int_{-\infty}^{\infty} e^{\mathbf{i}ts} f_Z(s)ds$ satisfies

$$d_0|t|^{\beta_0}\exp(-|t|^{\beta}/\gamma) \le |\varphi_{f_Z}(t)| \le d_1|t|^{\beta_1}\exp(-|t|^{\beta}/\gamma), \quad as\ |t| \to \infty, \tag{2.5}$$

for some strictly positive constants d_0, d_1, γ, β and constants β_0, β_1. We call f_Z *ordinary smooth* of order β if its Fourier transform $\varphi_{f_Z}(t)$ satisfies

$$d_0|t|^{-\beta} \le |\varphi_{f_Z}(t)| \le d_1|t|^{-\beta}, \quad as\ |t| \to \infty, \tag{2.6}$$

for some strictly positive constants d_0, d_1 and β. Examples for *supersmooth* distributions are normal, Cauchy, and mixture of any *supersmooth* distributions. Examples for *ordinary smooth* distributions include gamma, double exponential and mixture of any *ordinary smooth* distributions.

Definition 2.1. For each constant $C > 0$, define $\mathscr{F}_C$ to be the set of probability density functions f_X given by (2.2) where $\operatorname{ess\,sup}_{0\le\theta<2\pi}\sum_{j=-\infty}^{\infty} h(\theta+2\pi j) \le C$.

Definition 2.2. Let $\mathscr{U}$ denote the set of probability density functions f_X given by (2.2) where $\nu = 0$ and h is the density of a uniform distribution on $[a, a+2\pi k)$ for some constant a and integer $k > 0$.

Notice that, when $C \ge 1/(2\pi)$, we have $\mathscr{U} \subset \mathscr{F}_C$. The following proposition provides upper bounds for the expected L1 loss of $\hat{\omega}_0$.

Proposition 2.2. *Let $\hat{\omega}_0$ be as in (2.4) and α be as in (2.3). Suppose $\varphi_{f_Z}(k) \neq 0$ for all $k \in \mathbb{Z}$. Then*

$$E|\hat{\omega}_0 - \omega_0| \le \frac{\sum_{i=1}^{\nu}\omega_i}{q} + \frac{2}{q}\sum_{k=1}^{q}\frac{1}{\sqrt{n}}\frac{[1+16\sum_{l=1}^{\infty}\alpha(l)]^{\frac{1}{2}}}{|\varphi_{f_Z}(k)|} + \frac{\Omega}{q} + \frac{1-\sum_{k=0}^{\nu}\omega_k}{q}H,$$

where

$$H = \max\{2\pi\operatorname{ess\,sup}_{\theta\in[0,2\pi)}\sum_{j=-\infty}^{\infty} h(\theta+2\pi j) - 1, 1 - 2\pi\operatorname{ess\,inf}_{\theta\in[0,2\pi)}\sum_{j=-\infty}^{\infty} h(\theta+2\pi j)\}.$$

If in addition we have $f_X \in \mathscr{U}$, then

$$E|\hat{\omega}_0 - \omega_0| \le \frac{2}{q}\sum_{k=1}^{q}\frac{1}{\sqrt{n}}\frac{[1+16\sum_{l=1}^{\infty}\alpha(l)]^{\frac{1}{2}}}{|\varphi_{f_Z}(k)|}.$$

Proof. Denote the matrix norm induced by $l_1 - norm$ for vectors as $\| A \|_1$. We have:

$$\| A \|_1 = \max_{1 \le j \le q+1} \sum_{i=1}^{q+1} |a_{ij}|,$$

which is the maximum absolute column sum of the matrix A. Write $\rho(A)$ as the spectral radius of A. We observe from Theorem 5.6.9 of Horn and Johnson (1985) that

$$\begin{aligned}
\lambda_1(\hat{M}_q - M_q) &\le \rho(\hat{M}_q - M_q) \\
&\le \| \hat{M}_q - M_q \|_1 \\
&\le 2 \sum_{k=1}^{q+1} |(\hat{M}_q - M_q)_{q+1,k}|.
\end{aligned}$$

Similarly we have

$$\lambda_{q+1}(\hat{M}_q - M_q) = -\lambda_1(M_q - \hat{M}_q) \ge -2 \sum_{k=1}^{q+1} |(\hat{M}_q - M_q)_{q+1,k}|.$$

By Proposition 2.1, we have

$$\begin{aligned}
\lambda_1(\hat{M}_q) - 1 - q\omega_0 &\ge \lambda_1(M_q) + \lambda_{q+1}(\hat{M}_q - M_q) - 1 - q\omega_0 \\
&\ge 2\pi(1 - \sum_{k=0}^{\nu} \omega_k) \text{ess}\inf_{0 \le \theta < 2\pi} \sum_{j=-\infty}^{\infty} h(\theta + 2\pi j) \\
&\quad + (q+1)\omega_0 - 2 \sum_{k=1}^{q+1} |(\hat{M}_q - M_q)_{q+1,k}| - \Omega - 1 - q\omega_0 \\
&= -(1 - \sum_{k=0}^{\nu} \omega_k)\{1 - 2\pi \text{ess}\inf_{0 \le \theta < 2\pi} \sum_{j=-\infty}^{\infty} h(\theta + 2\pi j)\}
\end{aligned}$$

$$-\sum_{k=1}^{\nu}\omega_k - 2\sum_{k=1}^{q+1}|(\hat{M}_q - M_q)_{q+1,k}| - \Omega,$$

and

$$\begin{aligned}
\lambda_1(\hat{M}_q) - 1 - q\omega_0 &\leq \lambda_1(M_q) + \lambda_1(\hat{M}_q - M_q) - 1 - q\omega_0 \\
&\leq 2\pi(1-\sum_{k=0}^{\nu}\omega_k)\mathrm{ess\,sup}_{0\leq\theta<2\pi}\sum_{j=-\infty}^{\infty}h(\theta+2\pi j) \\
&\quad +(q+1)\omega_0 + 2\sum_{k=1}^{q+1}|(\hat{M}_q - M_q)_{q+1,k}| + \Omega - 1 - q\omega_0 \\
&= -(1-\sum_{k=0}^{\nu}\omega_k)\{1 - 2\pi\mathrm{ess\,sup}_{0\leq\theta<2\pi}\sum_{j=-\infty}^{\infty}h(\theta+2\pi j)\} \\
&\quad -\sum_{k=1}^{\nu}\omega_k + 2\sum_{k=1}^{q+1}|(\hat{M}_q - M_q)_{q+1,k}| + \Omega.
\end{aligned}$$

Since $\mathrm{ess\,sup}_{0\leq\theta<2\pi}\sum_{j=-\infty}^{\infty}h(\theta+2\pi j) < \infty$, and $\int_0^{2\pi}\sum_{j=-\infty}^{\infty}h(\theta+2\pi j) = 1$, we have

$$\mathrm{ess\,inf}_{0\leq\theta<2\pi}\sum_{j=-\infty}^{\infty}h(\theta+2\pi j) \leq \frac{1}{2\pi} \leq \mathrm{ess\,sup}_{0\leq\theta<2\pi}\sum_{j=-\infty}^{\infty}h(\theta+2\pi j),$$

Therefore, we have

$$|\hat{\omega}_0 - \omega_0| \leq \frac{\sum_{i=1}^{\nu}\omega_i}{q} + \frac{2}{q}\sum_{k=1}^{q+1}|(\hat{M}_q - M_q)_{q+1,k}| + \frac{\Omega}{q} + \frac{1-\sum_{k=0}^{\nu}\omega_k}{q}H. \tag{2.7}$$

Also, from Lemma 1 on page 10 of Doukhan (1994) we have

$$\begin{aligned}
&E|\frac{1}{n}\sum_{i=1}^{n}(e^{-\mathbf{i}kX_i} - Ee^{-\mathbf{i}kX_i})|^2 \\
= &\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}E[(e^{-\mathbf{i}kX_i} - Ee^{-\mathbf{i}kX_i})(e^{-\mathbf{i}kX_j} - Ee^{-\mathbf{i}kX_j})] \\
= &\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}[Ee^{\mathbf{i}k(X_j - X_i)} - Ee^{-\mathbf{i}kX_i}Ee^{\mathbf{i}kX_j}]
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{n}(1-|Ee^{\mathbf{i}kX_i}|^2) + \frac{1}{n^2}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}[Ee^{\mathbf{i}k(X_j-X_i)} - Ee^{-\mathbf{i}kX_i}Ee^{\mathbf{i}kX_j}] \\
&\quad + \frac{1}{n^2}\sum_{i=2}^{n}\sum_{j=1}^{i-1}[Ee^{\mathbf{i}k(X_j-X_i)} - Ee^{-\mathbf{i}kX_i}Ee^{\mathbf{i}kX_j}] \\
&= \frac{1}{n}(1-|Ee^{\mathbf{i}kX_i}|^2) + \frac{1}{n^2}\sum_{i=1}^{n-1}\sum_{l=1}^{n-i}[Ee^{\mathbf{i}k(X_{i+l}-X_i)} - Ee^{-\mathbf{i}kX_i}Ee^{\mathbf{i}kX_{i+l}}] \\
&\quad + \frac{1}{n^2}\sum_{i=1}^{n-1}\sum_{l=1}^{n-i}[Ee^{-\mathbf{i}k(X_{i+l}-X_i)} - Ee^{\mathbf{i}kX_i}Ee^{-\mathbf{i}kX_{i+l}}] \\
&= \frac{1}{n}(1-|Ee^{\mathbf{i}kX_i}|^2) + \frac{2}{n^2}\sum_{i=1}^{n-1}\sum_{l=1}^{n-i}|Ee^{-\mathbf{i}k\Theta_i}|^2 Re[Ee^{\mathbf{i}k(Z_{i+l}-Z_i)} - Ee^{-\mathbf{i}kZ_i}Ee^{\mathbf{i}kZ_{i+l}}] \\
&\leq \frac{1}{n} + \frac{16}{n^2}\sum_{i=1}^{n-1}\sum_{l=1}^{n-i}\alpha(l) \\
&\leq \frac{1}{n}[1+16\sum_{l=1}^{\infty}\alpha(l)].
\end{aligned}
$$

Consequently, we have:

$$
\begin{aligned}
\frac{2}{q}\sum_{k=1}^{q+1}E|(\hat{M}_q - M_q)_{q+1,k}| &= \frac{2}{q}\sum_{k=1}^{q}E|\frac{\frac{1}{n}\sum_{i=1}^{n}(e^{-\mathbf{i}kX_i} - Ee^{-\mathbf{i}kX_i})}{\varphi_{f_Z}(k)}| \\
&\leq \frac{2}{q}\sum_{k=1}^{q}\frac{[E|\frac{1}{n}\sum_{i=1}^{n}(e^{-\mathbf{i}kX_i} - Ee^{-\mathbf{i}kX_i})|^2]^{\frac{1}{2}}}{|\varphi_{f_Z}(k)|} \\
&\leq \frac{2}{q}\sum_{k=1}^{q}\frac{1}{\sqrt{n}}\frac{[1+16\sum_{l=1}^{\infty}\alpha(l)]^{\frac{1}{2}}}{|\varphi_{f_Z}(k)|}.
\end{aligned}
$$

Substituting this into (2.7) we have

$$
E|\hat{\omega}_0 - \omega_0| \leq \frac{\sum_{i=1}^{\nu}\omega_i}{q} + \frac{2}{q}\sum_{k=1}^{q}\frac{1}{\sqrt{n}}\frac{[1+16\sum_{l=1}^{\infty}\alpha(l)]^{\frac{1}{2}}}{|\varphi_{f_Z}(k)|} + \frac{\Omega}{q} + \frac{1-\sum_{k=0}^{\nu}\omega_k}{q}H.
$$

This proves the first statement of Proposition 2.2. If $f_X \in \mathscr{U}$, then $\nu = \Omega = 0$ and

$$
2\pi\sum_{j=-\infty}^{\infty}h(\theta+2\pi j) = 1, \quad \forall\ \theta \in \mathbb{R},
$$

which implies that $H = 0$. Consequently,

$$\begin{aligned} E|\hat{\omega}_0 - \omega_0| &\leq \frac{2}{q}\sum_{k=1}^{q+1} E|(\hat{M}_q - M_q)_{q+1,k}| \\ &\leq \frac{2}{q}\sum_{k=1}^{q} \frac{1}{\sqrt{n}} \frac{[1 + 16\sum_{l=1}^{\infty}\alpha(l)]^{\frac{1}{2}}}{|\varphi_{f_Z}(k)|}. \end{aligned}$$

□

The following two theorems are the main results of this section. They establish upper bounds to the minimax convergence rate of $\hat{\omega}_0$ with respect to $f_X \in \mathscr{F}_C$ and f_Z suitably smooth.

Theorem 2.1. *Let $\hat{\omega}_0$ be as in (2.4)and α be as in (2.3) such that $\sum_{l=1}^{\infty}\alpha(l) < \infty$. Suppose f_Z is supersmooth of order β, $\varphi_{f_Z}(k) \neq 0$ for all $k \in \mathbb{Z}$. Then by choosing $q = \lfloor (c\log n)^{\frac{1}{\beta}} \rfloor$ for some constant $0 < c < \gamma/2$, we have, for any $C > 0$,*

$$\sup_{f_X \in \mathscr{F}_C} E_{f_X}|\hat{\omega}_0 - \omega_0| = \mathrm{O}(\frac{1}{\log^{1/\beta} n});$$

By choosing q to be a constant, we have

$$\sup_{f_X \in \mathscr{U}} E_{f_X}|\hat{\omega}_0 - \omega_0| = \mathrm{O}(\frac{1}{\sqrt{n}}).$$

Proof. Since $\varphi_{f_Z}(k) \neq 0$ for all $k \in \mathbb{Z}$, we observe from (2.5) that there exists a constant d such that

$$\frac{1}{\sqrt{n}\min_{k\in\{1,\dots,q\}}|\varphi_{f_Z}(k)|} \leq \frac{dq^{-\beta_0}\exp(q^{\beta/\gamma})}{\sqrt{n}}.$$

Now choosing $q = \lfloor (c \log n)^{\frac{1}{\beta}} \rfloor$ for some constant $0 < c < \gamma/2$, we deduce from Proposition 2.2 that $\sup_{f_X \in \mathscr{F}_C} E_{f_X} |\hat{\omega}_0 - \omega_0| = O(\frac{1}{\log^{1/\beta} n})$. The second statement in this theorem is a straightforward consequence of the second statement of Proposition 2.2. □

The first statement of Theorem 2.1 together with Theorem 2.5 in Section 2.5 show that $\hat{\omega}_0$ achieves the optimal minimax convergence rate with respect to $f_X \in \mathscr{F}_C$ for *supersmooth* f_Z. This includes the case of normal noise since the normal density is *supersmooth.* The second statement of Theorem 2.1 shows that $\hat{\omega}_0$ converges in a $\sqrt{n}$ rate with respect to $f_X \in \mathscr{U}$. The following theorem gives similar results when f_Z is *ordinary smooth.*

Theorem 2.2. *Let $\hat{\omega}_0$ be as in (2.4)and α be as in (2.3) such that $\sum_{l=1}^{\infty} \alpha(l) < \infty$. Suppose f_Z is ordinary smooth of order β, $\varphi_{f_Z}(k) \neq 0$ for all $k \in \mathbb{Z}$. Then by choosing $q = \lfloor cn^{1/(2\beta+2)} \rfloor$ for some constant $c > 0$, we have, for any $C > 0$,*

$$\sup_{f_X \in \mathscr{F}_C} E_{f_X} |\hat{\omega}_0 - \omega_0| = \mathrm{O}(\frac{1}{n^{1/(2\beta+2)}});$$

By choosing q to be a constant, we have

$$\sup_{f_X \in \mathscr{U}} E_{f_X} |\hat{\omega}_0 - \omega_0| = \mathrm{O}(\frac{1}{\sqrt{n}}).$$

Proof. Similar to the proof of Theorem 2.1, since $\varphi_{f_Z}(k) \neq 0$ for all $k \in \mathbb{Z}$, we observe from (2.6) and Proposition 2.2 that by choosing $q = \lfloor cn^{1/(2\beta+2)} \rfloor$ for some constant $c > 0$, we have

$$\sup_{f_X \in \mathscr{F}_C} E_{f_X} |\hat{\omega}_0 - \omega_0| = O(\frac{1}{n^{1/(2\beta+2)}});$$

The second statement of Theorem 2.2 is a straight forward consequence of the second statement of Proposition 2.2. □

2.4 The estimator of Lee, et al. (2010)

When f_Z is symmetric about 0, Lee, *et al.* (2010) proposed an estimator $\tilde{\omega}_0$ for ω_0 where

$$\tilde{\omega}_0 = \frac{1}{2nT} \sum_{j=1}^{n} Re \int_{-T}^{T} \frac{e^{\mathbf{i}tX_j}}{\varphi_{f_Z}(t)} dt.$$

In this section, we show that, $\tilde{\omega}_0$ obtains the same bound as $\hat{\omega}_0$.

Proposition 2.3. *Suppose f_Z is symmetric about 0 such that $\varphi(t) \neq 0$ for any $t \in \mathbb{R}$. Denote the Fourier transform of h as φ_h and assume that $||\varphi_h||_1 < \infty$. Then*

$$E|\tilde{\omega}_0 - \omega_0| \leq \sqrt{\frac{(1-\omega_0)^2 ||\varphi_h||_1^2}{4T^2} + \frac{2}{n}[\frac{1}{T}\int_0^T \frac{1}{\varphi_{f_Z}(t)} dt]^2}.$$

Proof. Notice that

$$E|\tilde{\omega}_0 - \omega_0| \leq \sqrt{E|\tilde{\omega}_0 - \omega_0|^2}$$

$$= \sqrt{E(\tilde{\omega}_0 - E\tilde{\omega}_0)^2 + (E\tilde{\omega}_0 - \omega_0)^2}.$$

By Lemma 1 and (12) of Lee, *et al.* (2010), we immediately have:

$$E|\tilde{\omega}_0 - \omega_0| \leq \sqrt{\frac{(1-\omega_0)^2||\varphi_h||_1^2}{4T^2} + \frac{2}{n}[\frac{1}{T}\int_0^T \frac{1}{\varphi_{f_Z}(t)}dt]^2}.$$

□

Theorem 2.3. *Suppose f_Z is supersmooth of order β and is symmetric about 0 such that $\varphi(t) \neq 0$ for any $t \in \mathbb{R}$. Denote the Fourier transform of h as φ_h and assume that $||\varphi_h||_1 < \infty$. Then by choosing $T = (c\log n)^{1/\beta}$ for some constant $0 < c < \gamma/2$, we have*

$$E|\tilde{\omega}_0 - \omega_0| = \mathrm{O}(\frac{1}{\log^{1/\beta} n}).$$

Proof. When $T = (c\log n)^{1/\beta}$, $\frac{(1-\omega_0)^2||\varphi_h||_1^2}{4T^2} = O(\frac{1}{\log^{2/\beta} n})$, and by (2.5), there exists a constant d such that

$$\frac{1}{\sqrt{n}T}\int_0^T \frac{1}{\varphi_{f_Z}(t)}dt = \frac{d}{\sqrt{n}T}\int_0^T (1+t^{-\beta_0})e^{t^\beta/\gamma}dt = o(\frac{1}{\log^{1/\beta} n}).$$

Therefore, by Proposition 2.3 we have $E|\tilde{\omega}_0 - \omega_0| = O(\frac{1}{\log^{1/\beta} n})$. □

Theorem 2.4. *Suppose f_Z is ordinary smooth of order β and is symmetric about 0 such that $\varphi(t) \neq 0$ for any $t \in \mathbb{R}$. Denote the Fourier transform of h as φ_h and assume that $||\varphi_h||_1 < \infty$. Then by choosing $T = cn^{1/(2\beta+2)}$ for some constant $c > 0$, we have*

$$E|\tilde{\omega}_0 - \omega_0| = \mathrm{O}(\frac{1}{n^{1/(2\beta+2)}}).$$

Proof. When $T = cn^{1/(2\beta+2)}$, $\frac{(1-\omega_0)^2||\varphi_h||_1^2}{4T^2} = O(\frac{1}{n^{1/(\beta+1)}})$, and by (2.5), there exists a constant d such that

$$\frac{1}{\sqrt{n}T}\int_0^T \frac{1}{\varphi_{f_Z}(t)}dt = \frac{d}{\sqrt{n}T}\int_0^T (1+t^\beta)dt = O(\frac{1}{n^{1/(2\beta+2)}}).$$

Therefore, by Proposition 2.3 we have $E|\tilde{\omega}_0 - \omega_0| = O(\frac{1}{n^{1/(2\beta+2)}})$. □

2.5 Lower bounds

In this section, we establish lower bounds to the minimax convergence rate for the problem of estimating ω_0. Assuming that the noise random variables $Z_1, ..., Z_n$ are independent and identically distributed with known marginal density f_Z, and $f_Z \in \mathscr{F}_C$ for some sufficiently large C.

An infinitely differentiable complex-valued function f on $\mathbb{R}$ is called a *Schwartz function* if for all non-negative integers i and j, there exist positive constants $C_{i,j}$ such that

$$|\frac{d^i f(x)}{dx^i}| \leq \frac{C_{i,j}}{(1+|x|^j)}, \quad \forall x \in \mathbb{R}.$$

The set of all Schwartz functions will be denoted by $\mathbf{S}(\mathbb{R})$. Define $\eta : \mathbb{R} \to \mathbb{R}$ by

$$\eta(t) = \begin{cases} [\int_{-1}^1 e^{-1/(1-s^2)}ds]e^{-1/(1-t^2)} & if\ |t| < 1, \\ 0 & if\ |t| \geq 1. \end{cases} \tag{2.8}$$

It is easily seen that $\eta \in \mathbf{S}(\mathbb{R})$. Next let $\alpha > 1$ and $I_{[-b,b]}(t)$ denote the indicator function of the interval $[-b, b]$. Define

$$\psi(t) = I_{[-b,b]} * \eta(t) = \int_{-b}^{b} \eta(t-s)ds, \quad \forall t \in \mathbb{R},$$

where $*$ denotes the convolution operation between two functions. We observe that $\psi \in \mathbf{S}(\mathbb{R})$, $0 \leq \psi(t) \leq 1$ for all $t \in \mathbb{R}$ and

$$\psi(t) = \begin{cases} 1 & if\ |t| \leq b-1, \\ 0 & if\ |t| \geq b+1. \end{cases} \tag{2.9}$$

Next define

$$\check{\psi}(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-\mathbf{i}tx} \psi(t)dt, \quad \forall x \in \mathbb{R}.$$

$\check{\psi}(x) : \mathbb{R} \to \mathbb{R}$ is the inverse Fourier transform of ψ and it follows from Proposition 2.2.11 of Grafakos (2008) that $\check{\psi}(x) \in \mathbf{S}(\mathbb{R})$. In particular, we have $\psi(0) = \int_{-\infty}^{\infty} \check{\psi}(t)dt = 1$ since

$$\psi(t) = \int_{-\infty}^{\infty} e^{\mathbf{i}tx} \check{\psi}(x)dx, \quad \forall t \in \mathbb{R}.$$

Define

$$h_0(x) = \frac{C_r}{(1+x^2)^r}, \quad \forall x \in \mathbb{R},$$

where $r > 1/2$ and C_r are constants such that $\int_{-\infty}^{\infty} h_0(x)dx = 1$. Let $\omega_0 \in (0, 1]$ be a constant. Now choose a_0 and δ_n to be suitably small, strictly positive constants such that $\inf_{x\in\mathbb{R}}\{(1-\omega_0)h_0(x) + a_0\check{\psi}(x/\delta_n)\} \geq 0$. This is indeed possible since

$\check{\psi} \in \mathbf{S}(\mathbb{R})$ and hence decreases to 0 at a super polynomial rate while h_0 decreases to 0 at the exact rate of $|x|^{-2r}$ as $|x| \to \infty$. Define the cumulative distribution functions

$$F(\theta) = \omega_0 I_{\{\theta \geq 0\}} + (1 - \omega_0) \int_{-\infty}^{\theta} h_0(x)dx,$$

$$F^*(\theta) = \omega_0^* I_{\{\theta \geq 0\}} + (1 - \omega_0) \int_{-\infty}^{\theta} h_0(x)dx + a_0 \int_{-\infty}^{\theta} \check{\psi}(x/\delta_n)dx,$$

where $\omega_0 - \omega_0^* = a_0\delta_n$. Let Θ, Θ^* be random variables with distribution functions F, F^* respectively and Z, Z^* be random variables each with density f_Z. We further assume that Θ, Z are independent and Θ^*, Z^* are independent. Define $X = \Theta + Z$ and $X^* = \Theta^* + Z^*$ and let g, g^* denote the density functions of X, X^* respectively. Notice that $g, g^* \in \mathscr{F}_C$ for a sufficiently large C and $\mathscr{F}_C$ is a convex set. In addition, ω_0 as a function of g (see (2.2)) is linear and $\omega_0 - \omega_0^* = a_0\delta_n$. By Theorems 2.1 and 3.1 of Donoho and Liu (1991), lower bounds can be obtained by finding the largest δ_n such that the square of the Hellinger distance between g and g^* is of order $O(\frac{1}{n})$. Since the square of the Hellinger distance is dominated by χ^2 divergence, we want to find the largest δ_n such that the χ^2 divergence

$$\int_{\mathbb{R}} \frac{[g(x) - g^*(x)]^2}{g(x)} dx \leq \frac{c}{n}, \tag{2.10}$$

for some constant $c > 0$. The following two lemmas will be used in proving Theorem 2.5.

Lemma 2.1. *With the above notation, there exists a constant* $C_g > 0$ *such that*

$$g(x) \geq \frac{C_g}{(1+x^2)^r}, \quad \forall x \in \mathbb{R}.$$

Proof. First let $a > 0$ be a constant such that $\int_0^a f_Z(t)dt > 0$. Notice that g is a strictly positive continuous function on $\mathbb{R}$. It suffices to assume that $x \geq a$. By the definition of h_0 we have

$$\begin{aligned} g(x) &= \omega_0 f_Z(x) + (1-\omega_0) h_0 * f_Z(x) \\ &\geq (1-\omega_0) \int_0^a h_0(x-s) f_Z(s) ds \\ &\geq \frac{(1-\omega_0) C_r \int_0^a f_Z(s) ds}{(1+x^2)^r}. \end{aligned}$$

□

Lemma 2.2. *Suppose* $f_z = \mathrm{O}(|x|^{-\kappa})$ *as* $|x| \to \infty$ *for some constant* $\kappa > 1$*. Let* $0 < \delta_n, \kappa_0 < 1$ *be constants such that* $\kappa - \kappa_0 > 1$*. Then there exist constants* M *and* C_M *such that*

$$|\delta_n f_Z(\delta_n x) - \int_{-\infty}^{\infty} \check{\psi}(x-y) \delta_n f_Z(\delta_n y) dy| \leq \frac{C_M}{|\delta_n x|^{\kappa - \kappa_0}}, \quad \forall |\delta_n x| \geq M.$$

Proof. Since $\check{\psi} \in \mathbf{S}(\mathbb{R})$, we have $|\check{\psi}(x)| = O(|x|^{-m_0})$ as $|x| \to \infty$ for some constant m_0 such that $\kappa - \kappa_0 > m_0 \kappa_0$. Consequently,

$$\begin{aligned} & |\int_{-\infty}^{\infty} \check{\psi}(x-y) \delta_n f_Z(\delta_n y) dy| \\ \leq \ & |\int_{|x - y/\delta_n| \leq |x|^{\kappa_0}} \check{\psi}(x - y/\delta_n) f_Z(y) dy| + |\int_{|x-y/\delta_n| > |x|^{\kappa_0}} \check{\psi}(x - y/\delta_n) f_Z(y) dy| \end{aligned}$$

$$
\begin{aligned}
&\leq \; ||\check{\psi}||_\infty \int_{|\delta_n x|-|\delta_n x|^{\kappa_0} \leq y \leq |\delta_n x|+|\delta_n x|^{\kappa_0}} f_Z(y)dy + O(|x|^{-m_0\kappa_0}) \\
&= \; O(|x|^{-(\kappa-\kappa_0)}) \;\; as \; |\delta_n x| \to \infty.
\end{aligned}
$$

□

The following theorem provides lower bounds for the minimax convergence rate when the noise density function is *supersmooth*.

Theorem 2.5. *Let $X_1, ..., X_n$ be as in Section 2.1 with the noise random variables $Z_1, ..., Z_n$ independent and identically distributed. Suppose f_Z is supersmooth of order β and $f_Z(x) = \mathrm{O}(|x|^{-\kappa})$ as $|x| \to \infty$ for some constant $\kappa > 1$. Then for any estimator $\hat{\omega}_0$ based on $X_1, ..., X_n$, we have*

$$
\sup_{f_X \in \mathscr{F}_C} E_{f_X}|\hat{\omega}_0 - \omega_0| > c(\log n)^{-1/\beta},
$$

for some constant $c > 0$ whenever C is sufficiently large.

Proof. We observe that

$$
\begin{aligned}
& \int_{\mathbb{R}} \frac{[g(x) - g^*(x)]^2}{g(x)} dx \\
= & \int_{\mathbb{R}} \frac{[(\omega_0 - \omega_0^*)f_Z(x) - a_0 \int_{-\infty}^{\infty} \check{\psi}((x-y)/\delta_n) f_Z(y)dy]^2}{g(x)} dx \\
= & \int_{\mathbb{R}} \frac{[(\omega_0 - \omega_0^*)f_Z(x) - a_0\delta_n \int_{-\infty}^{\infty} \check{\psi}(x\delta_n^{-1} - y) f_Z(\delta_n y)dy]^2}{g(x)} dx \\
= & \; a_0^2 \delta_n \int_{\mathbb{R}} \frac{[\delta_n f_Z(\delta_n x) - \int_{-\infty}^{\infty} \check{\psi}(x-y)\delta_n f_Z(\delta_n y)dy]^2}{g(\delta_n x)} dx
\end{aligned}
$$

Using Lemmas 2.1 and 2.2 and taking a constant $M_n \geq (M \vee 1)$, the last term of the above equations is less than or equal to

$$\begin{aligned}
&\frac{a_0^2\delta_n}{C_g}\int_{-\infty}^{\infty}(1+|\delta_n x|^2)^r[\delta_n f_Z(\delta_n x) - \int_{-\infty}^{\infty}\check{\psi}(x-y)\delta_n f_Z(\delta_n y)dy]^2 dx \\
\leq\ & \frac{a_0^2\delta_n}{C_g}\int_{|\delta_n x|\geq M_n}\frac{C_M^2(1+|\delta_n x|^2)^r}{|\delta_n x|^{2(\kappa-\kappa_0)}}dx \\
&+\frac{a_0^2\delta_n(1+M_n^2)^r}{C_g}\int_{-\infty}^{\infty}[\delta_n f_Z(\delta_n x) - \int_{-\infty}^{\infty}\check{\psi}(x-y)\delta_n f_Z(\delta_n y)dy]^2 dx \\
\leq\ & \frac{a_0^2\delta_n(1+M_n^2)^r}{C_g}\int_{-\infty}^{\infty}[\delta_n f_Z(\delta_n x) - \int_{-\infty}^{\infty}\check{\psi}(x-y)\delta_n f_Z(\delta_n y)dy]^2 dx \\
&+\frac{2^r a_0^2 C_M^2}{C_g}\int_{|x|\geq M_n}\frac{1}{|x|^{2(\kappa-\kappa_0-r)}}dx \\
\leq\ & \frac{a_0^2\delta_n(1+M_n^2)^r}{C_g}\int_{-\infty}^{\infty}[\delta_n f_Z(\delta_n x) - \int_{-\infty}^{\infty}\check{\psi}(x-y)\delta_n f_Z(\delta_n y)dy]^2 dx \\
&+\frac{2^{r+1}a_0^2C_M^2}{C_g[2(\kappa-\kappa_0-r)-1]M_n^{2(\kappa-\kappa_0-r)-1}} \\
=\ & \frac{a_0^2\delta_n(1+M_n^2)^r}{2\pi C_g}\int_{-\infty}^{\infty}|\varphi_{f_Z}(\frac{t}{\delta_n}) - \psi(t)\varphi_{f_Z}(\frac{t}{\delta_n})|^2 dt \\
&+\frac{2^{r+1}a_0^2C_M^2}{C_g[2(\kappa-\kappa_0-r)-1]M_n^{2(\kappa-\kappa_0-r)-1}} \\
\leq\ & \frac{a_0^2\delta_n(1+M_n^2)^r}{\pi C_g}\int_{b-1}^{\infty}|\varphi_{f_Z}(\frac{t}{\delta_n})|^2 dt + \frac{2^{r+1}a_0^2C_M^2}{C_g[2(\kappa-\kappa_0-r)-1]M_n^{2(\kappa-\kappa_0-r)-1}}.
\end{aligned}$$

where in the last two steps we have used Parseval's identity and (2.9). Choose constants $b > 1, \kappa_0 > 0, r > 1/2$ such that $2(\kappa-\kappa_0-r)-1 > 0$ and $(b-1)^\beta - r > 0$. Now let $M_n = e^{1/(\delta_n^\beta\gamma)}$ and $\delta_n = (c/\log n)^{1/\beta}$ for some constant $0 < c < \gamma^{-1}\min\{2(b-1)^\beta - 2r, 2(\kappa-\kappa_0-r)-1\}$. Then

$$\frac{2^{r+1}a_0^2C_M^2}{C_g[2(\kappa-\kappa_0-r)-1]M_n^{2(\kappa-\kappa_0-r)-1}} = o(1/n), \quad as\ n\to\infty.$$

It follows from (2.5) that

$$\begin{aligned} & \frac{a_0^2\delta_n(1+M_n^2)^r}{\pi C_g}\int_{b-1}^{\infty}|\varphi_{f_Z}(\frac{t}{\delta_n})|^2dt \\ \leq & \frac{2^rM_n^{2r}a_0^2d_1^2\delta_n^{1-2\beta_1}}{\pi C_g}\int_{b-1}^{\infty}t^{2\beta_1}e^{-2t^\beta/(\delta_n^\beta\gamma)}dt \\ \leq & \frac{2^rM_n^{2r}a_0^2d_1^2\delta_n^{1-2\beta_1}e^{-2(b-1)^\beta/(\delta_n^\beta\gamma)}}{\pi C_g}\int_{b-1}^{\infty}t^{2\beta_1}e^{-2[t^\beta-(b-1)^\beta]/(\delta_n^\beta\gamma)}dt \\ = & o(1/n), \end{aligned}$$

as $n\to\infty$. Hence we conclude that (2.10) holds. Notice that $g, g^*\in\mathscr{F}_C$ for a sufficiently large C and $\mathscr{F}_C$ is a convex set. In addition, ω_0 as a function of g (see (2.2)) is linear and $\omega_0-\omega_0^*=a_0\delta_n$. Since the Hellinger distance is dominated by the χ^2 divergence [as in (2.10)], it follows from Theorems 2.1 and 3.1 of Donoho and Liu (1991) that

$$\inf_{\hat{\omega}_0}\sup_{f_X\in\mathscr{F}_C}E_{f_X}|\hat{\omega}_0-\omega_0|>c(\log n)^{-1/\beta},$$

for some constant $c>0$. □

The following theorem provides lower bounds for the minimax convergence rate when the noise density function is *ordinary smooth* with order $\beta>1/2$.

Theorem 2.6. *Let $X_1,...,X_n$ be as in Section 2.1 with the noise random variables $Z_1,...,Z_n$ independent and identically distributed. Suppose f_Z is ordinary smooth of order $\beta>1/2$ and $|d^j\varphi_{f_Z}(t)/dt^j(t)|<c_j|t|^{\beta-j}$ as $|t|\to\infty$ for $j=\{0,1,2\}$ and*

some constants c_0, c_1, c_2. Then for any estimator $\hat{\omega}_0$ based on $X_1, ..., X_n$, we have

$$\sup_{f_X \in \mathscr{F}_C} E_{f_X}|\hat{\omega}_0 - \omega_0| > cn^{-1/(2\beta+1)},$$

for some constant $c > 0$ whenever C is sufficiently large .

Proof. Choose $1/2 < r < 3/2$.

$$\begin{aligned}
&\int_{-\infty}^{+\infty} \frac{[g(x) - g^*(x)]^2}{g(x)} \\
= \quad & a_0^2 \delta_n \int_{-\infty}^{+\infty} \frac{[\delta_n f_Z(x\delta_n) - \int_{-\infty}^{\infty} \breve{\psi}(x-y)\delta_n f_Z(\delta_n y)dy]^2}{g(\delta_n x)} dx \\
= \quad & a_0^2 \delta_n (\int_{|\delta_n x|>1} + \int_{|\delta_n x|\le 1}) \frac{[\delta_n f_Z(x\delta_n) - \int_{-\infty}^{\infty} \breve{\psi}(x-y)\delta_n f_Z(\delta_n y)dy]^2}{g(\delta_n x)} dx.
\end{aligned}$$

From Lemma 2.1 we know that g is a continuous function and does not vanish in $|x| \le 1$. Let $C_1 = \max_{|x|\le 1} g^{-1}(x)$. By Parseval's identity and (2.6), we have

$$\begin{aligned}
I_1 \quad :=& \quad \int_{|\delta_n x|\le 1} \frac{[\delta_n f_Z(x\delta_n) - \int_{-\infty}^{\infty} \breve{\psi}(x-y)\delta_n f_Z(\delta_n y)dy]^2}{g(\delta_n x)} dx \\
\le& \quad \frac{C_1}{2\pi} \int_{-\infty}^{+\infty} |\varphi_{f_Z}(t/\delta_n)(1 - \psi(t))|^2 dt \\
\le& \quad \frac{C_1}{\pi} \int_{b-1}^{+\infty} |\varphi_{f_Z}(t/\delta_n)|^2 dt \\
=& \quad O(\delta_n^{2\beta}).
\end{aligned}$$

Define

$$\phi_{\delta_n}(t) = \frac{d^2(\varphi_{f_Z}(t/\delta_n)(1 - \psi(t)))}{dt^2}.$$

Writing $\varphi_{f_Z}^{(j)}(t) = d^j\varphi_{f_Z}(t)/dt^j$ and $\psi^{(j)}(t) = d^j\psi(t)/dt^j$. From the definition of $\psi(t)$, we know that $\psi^{(j)}(t)$ is bounded for $j = 0, 1, 2$, and notice that $\psi^{(j)}(t) = 0$ for $|t| \geq \alpha + 1, j = 1, 2$, we have, for $t \geq \alpha - 1$ and δ_n small enough, there exists a constant $C_2 > 0$, such that

$$\begin{aligned} |\phi_{\delta_n}(t)| &= |\varphi_{f_Z}^{(2)}(t/\delta_n)\frac{1}{\delta_n^2}(1-\psi(t)) - \varphi_{f_Z}(t/\delta_n)\psi^{(2)}(t) - 2\varphi_f^{(1)}(t/\delta_n)\frac{1}{\delta_n}\psi^{(1)}(t)| \\ &\leq C_2\delta_n^{\beta}t^{-\beta-2}. \end{aligned}$$

By the Fourier inversion formula and (2.6) we have:

$$\begin{aligned} & |\delta_n f_Z(x\delta_n) - \int_{-\infty}^{\infty} \breve{\psi}(x-y)\delta_n f_Z(\delta_n y)dy| \\ = & |\frac{1}{2\pi}\int_{-\infty}^{+\infty} \exp(-\mathbf{i}tx)\varphi_{f_Z}(t/\delta_n)(1-\psi(t))dt| \\ \leq & |\frac{1}{\pi}\int_{b-1}^{+\infty} \exp(-\mathbf{i}tx)\varphi_{f_Z}(t/\delta_n)(1-\psi(t))dt|. \end{aligned}$$

Consequently

$$\begin{aligned} & |\delta_n f_Z(x\delta_n) - \int_{-\infty}^{\infty} \breve{\psi}(x-y)\delta_n f_Z(\delta_n y)dy|^2 \\ \leq & |\frac{1}{\pi}\int_{b-1}^{+\infty} \exp(-\mathbf{i}tx)\varphi_{f_Z}(t/\delta_n)(1-\psi(t))dt|^2 \\ = & |\frac{1}{-x^2\pi}\int_{\alpha-1}^{+\infty} \exp(-\mathbf{i}xt)\phi_{\delta_n}(t)dt|^2 \\ = & (\pi x^2)^{-2}\delta_n^{2\beta}|\int_{b-1}^{+\infty} \exp(-\mathbf{i}xt)\delta_n^{-\beta}\phi_{\delta_n}(t)dt|^2. \end{aligned}$$

By Lemma 2.1, we have:

$$I_2 := \int_{|\delta_n x|>1} \frac{[\delta_n f(x\delta_n) - \int_{-\infty}^{\infty} \breve{\psi}(x-y)\delta_n f(\delta_n y)dy]^2}{g_0(\delta_n x)}dx$$

$$
\begin{aligned}
&\leq \int_{|\delta_n x|>1} \frac{(1+\delta_n^2 x^2)^r}{C_{g_0}} (\pi x^2)^{-2} \delta_n^{2\beta} \left| \int_{\alpha-1}^{+\infty} \exp(-\mathbf{i}xt) \delta_n^{-\beta} \phi_{\delta_n}(t) dt \right|^2 dx \\
&\leq \frac{2^r \delta_n^{2r+2\beta}}{C_{g_0}\pi^2} \int_{|\delta_n x|>1} x^{2r-4} \left| \int_{\alpha-1}^{+\infty} \exp(-\mathbf{i}xt) \delta_n^{-\beta} \phi_{\delta_n}(t) dt \right|^2 dx \\
&\leq \frac{2^r \delta_n^{2r+2\beta}}{C_{g_0}\pi^2} \int_{|\delta_n x|>1} x^{2r-4} \left| \int_{\alpha-1}^{+\infty} C_2 t^{-\beta-2} dt \right|^2 dx \\
&= O(\delta_n^{2\beta+3}).
\end{aligned}
$$

By choosing $\delta_n = dn^{-1/(2\beta+1)}$ for some constant $d > 0$, we have,

$$
\int_{-\infty}^{+\infty} \frac{[g_0(x) - g_n(x)]^2}{g_0(x)} = a_0^2 \delta_n (I_1 + I_2) = O(\delta_n^{2\beta+1}) = O(n^{-1}).
$$

When C is large enough, $g, g^* \in \mathscr{F}_C$, and we conclude from Theorems 2.1 and 3.1 of Donoho and Liu (1991) that

$$
\inf_{\hat{\omega}_0} \sup_{f_X \in \mathscr{F}_C} E_{f_X} |\hat{\omega}_0 - \omega_0| > cn^{-1/(2\beta+1)},
$$

for some constant $c > 0$.

□

2.6 A method-of-moments estimator when f_Z is unknown

Let $X_1, ..., X_n$ be as in Section 2.3. In this section we assume that f_Z is unknown but we assume that there is an independent sample of (pure) noise observations

$Y_1, ..., Y_m$. In Section 2.3, we use $\hat{M}_q$ to estimate M_q. Since now f_Z is unknown, we shall estimate M_q using the $(q+1)\times(q+1)$ matrix $\check{M}_q$ whose (j,k)th element is given by

$$(\check{M}_q)_{jk} = \frac{n^{-1}\sum_{i=1}^{n} e^{-\mathbf{i}(j-k)X_i}}{m^{-1}\sum_{l=1}^{m} e^{-\mathbf{i}(j-k)Y_l}}, \quad \forall 1 \le j,k \le q+1.$$

$\check{M}_q$ is a Hermitian matrix and hence its eigenvalues are real numbers. We propose as an estimator of ω_0,

$$\check{\omega}_0 = \begin{cases} 1, & if\ (\lambda_1(\check{M}_q)-1)/q > 1, \\ (\lambda_1(\check{M}_q)-1)/q, & if\ 0 \le (\lambda_1(\check{M}_q)-1)/q \le 1, \\ 0, & if\ (\lambda_1(\check{M}_q)-1)/q < 0, \end{cases} \tag{2.11}$$

for a sufficiently large integer q.

Proposition 2.4. *Let $\check{\omega}_0$ be as in (2.11) and α be as in (2.3) such that, for a certain constant $c_0 > 0$, $\alpha(l) \le e^{-c_0 l}$ for all integers $l \ge 1$. Then there exists a constant $C_0 > 0$ depending only on c_0 such that for all $m \ge 4$,*

$$\begin{aligned} E|\check{\omega}_0 - \omega_0| \le\ & \frac{\sum_{i=1}^{\nu}\omega_i}{q} + 4\sum_{j=1}^{q}\{\frac{[1+16\sum_{l=1}^{\infty}\alpha(l)]^{\frac{1}{2}}}{q\sqrt{n}|\varphi_{f_Z}(j)|} + \frac{[1+16\sum_{l=1}^{\infty}\alpha(l)]^{\frac{1}{2}}}{q\sqrt{m}|\varphi_{f_Z}(j)|} \\ & + \exp[-\frac{C_0 m|\varphi_{f_Z}(j)|^2}{8(4+\log^2 m)}]\} + \frac{\Omega}{q} + \frac{1-\sum_{k=0}^{\nu}\omega_k}{q}H, \end{aligned}$$

where

$$H = \max\{2\pi \sup_{\theta\in[0,2\pi)} \sum_{j=-\infty}^{\infty} h(\theta+2\pi j) - 1, 1 - 2\pi \inf_{\theta\in[0,2\pi)} \sum_{j=-\infty}^{\infty} h(\theta+2\pi j)\}.$$

If in addition we have $f_X \in \mathscr{U}$, then for all $m \ge 4$,

$$E|\check{\omega}_0 - \omega_0| \le 4\sum_{j=1}^{q}\{\frac{[1+16\sum_{l=1}^{\infty}\alpha(l)]^{\frac{1}{2}}}{q\sqrt{n}|\varphi_{f_Z}(j)|} + \frac{[1+16\sum_{l=1}^{\infty}\alpha(l)]^{\frac{1}{2}}}{q\sqrt{m}|\varphi_{f_Z}(j)|}$$

$$+\exp[-\frac{C_0 m|\varphi_{f_Z}(j)|^2}{8(4+\log^2 m)}]\}.$$

Proof. Similar to (2.7) in the proof of Proposition 2.2, we have

$$|\check{\omega}_0-\omega_0| \le \frac{\sum_{i=1}^{\nu}\omega_i}{q}+\frac{2}{q}\sum_{k=1}^{q+1}|(\check{M}_q-M_q)_{q+1,k}|+\frac{\Omega}{q}+\frac{1-\sum_{k=0}^{\nu}\omega_k}{q}H. \quad (2.12)$$

Also

$$\begin{aligned}
&\sum_{k=1}^{q}|(\check{M}_q-M_q)_{q+1,k}|I_{\{\min_{1\le j\le q}|\frac{m^{-1}\sum_{l=1}^m e^{-\mathbf{i}jY_l}}{Ee^{-\mathbf{i}jY_l}}|>\frac{1}{2}\}}\\
= \;& I_{\{\min_{1\le j\le q}|\frac{m^{-1}\sum_{l=1}^m e^{-\mathbf{i}jY_l}}{Ee^{-\mathbf{i}jY_1}}|>\frac{1}{2}\}}\\
&\times\sum_{j=1}^{q}|\frac{(n^{-1}\sum_{i=1}^n e^{-\mathbf{i}jX_i})(Ee^{-\mathbf{i}jY_1})-(m^{-1}\sum_{l=1}^m e^{-\mathbf{i}jY_l})(Ee^{-\mathbf{i}jX_1})}{(m^{-1}\sum_{l=1}^m e^{-\mathbf{i}jY_l})(Ee^{-\mathbf{i}jY_1})}|\\
\le \;& I_{\{\min_{1\le j\le q}|\frac{m^{-1}\sum_{l=1}^m e^{-\mathbf{i}jY_l}}{Ee^{-\mathbf{i}jY_1}}|>\frac{1}{2}\}}\sum_{j=1}^{q}\big[|\frac{(n^{-1}\sum_{i=1}^n e^{-\mathbf{i}jX_i}-Ee^{-\mathbf{i}jX_1})(Ee^{-\mathbf{i}jY_1}))}{(m^{-1}\sum_{l=1}^m e^{-\mathbf{i}jY_l})(Ee^{-\mathbf{i}jY_1})}|\\
&+|\frac{(m^{-1}\sum_{l=1}^m e^{-\mathbf{i}jY_l}-Ee^{-\mathbf{i}jY_1})(Ee^{-\mathbf{i}jX_1}))}{(m^{-1}\sum_{l=1}^m e^{-\mathbf{i}jY_l})(Ee^{-\mathbf{i}jY_1})}|\big]\\
\le \;& 2\sum_{j=1}^{q}\big[|\frac{\sum_{i=1}^n(e^{-\mathbf{i}jX_i}-Ee^{-\mathbf{i}jX_1})}{nEe^{-\mathbf{i}jY_1}}|+|\frac{\sum_{l=1}^m(e^{-\mathbf{i}jY_l}-Ee^{-\mathbf{i}jY_1})}{mEe^{-\mathbf{i}jY_1}}|\big]. \quad (2.13)
\end{aligned}$$

We recall that in the proof of Proposition 2.2 we have showed that

$$E|\frac{1}{n}\sum_{i=1}^{n}(e^{-\mathbf{i}kX_i}-Ee^{-\mathbf{i}kX_i})|^2 \le \frac{1}{n}[1+16\sum_{l=1}^{\infty}\alpha(l)]. \quad (2.14)$$

Since $Y_1,...,Y_m$ are strongly mixing with $\alpha(l)\le e^{-c_0 l}$ for all integers $l\ge 1$, we observe from Theorem 1 of Merlevede et al. (2009) that there exists a constant $C_0>0$ depending on c_0 only such that for all $m\ge 4$,

$$P(\min_{1\le j\le q}|\frac{m^{-1}\sum_{l=1}^m e^{-\mathbf{i}jY_l}}{Ee^{-\mathbf{i}jY_l}}|\le\frac{1}{2})<4\sum_{j=1}^{q}\exp[-\frac{C_0 m|Ee^{-\mathbf{i}jY_1}|^2}{8(4+\log^2 m)}]. \quad (2.15)$$

It follows from (2.12), (2.13), (2.14) and (2.15) that

$$\begin{aligned}
E|\breve{\omega}_0-\omega_0| &\leq P(\min_{1\leq j\leq q}|\frac{m^{-1}\sum_{l=1}^{m}e^{-\mathbf{i}jY_l}}{Ee^{-\mathbf{i}jY_l}}|\leq\frac{1}{2}) \\
&\quad +E\left\{E|\breve{\omega}_0-\omega_0|I_{\{\min_{1\leq j\leq q}|\frac{m^{-1}\sum_{l=1}^{m}e^{-\mathbf{i}jY_l}}{Ee^{-\mathbf{i}jY_l}}|>\frac{1}{2}\}}\right\} \\
&< 4\sum_{j=1}^{q}\exp[-\frac{C_0m|Ee^{-\mathbf{i}jY_1}|^2}{8(4+\log^2 m)}]+\frac{\sum_{i=1}^{\nu}\omega_i}{q}+\frac{\Omega}{q}+\frac{1-\sum_{k=0}^{\nu}\omega_k}{q}H \\
&\quad +\frac{4}{q}\sum_{j=1}^{q}\left\{|\frac{\sum_{i=1}^{n}(e^{-\mathbf{i}jX_i}-Ee^{-\mathbf{i}jX_1})}{nEe^{-\mathbf{i}jY_1}}|+|\frac{\sum_{l=1}^{m}(e^{-\mathbf{i}jY_l}-Ee^{-\mathbf{i}jY_1})}{mEe^{-\mathbf{i}jY_1}}|\right\} \\
&\leq \frac{\sum_{i=1}^{\nu}\omega_i}{q}+4\sum_{j=1}^{q}\{\frac{[1+16\sum_{l=1}^{\infty}\alpha(l)]^{\frac{1}{2}}}{q\sqrt{n}|\varphi_{f_Z}(j)|}+\frac{[1+16\sum_{l=1}^{\infty}\alpha(l)]^{\frac{1}{2}}}{q\sqrt{m}|\varphi_{f_Z}(j)|} \\
&\quad +\exp[-\frac{C_0m|\varphi_{f_Z}(j)|^2}{8(4+\log^2 m)}]\}+\frac{\Omega}{q}+\frac{1-\sum_{k=0}^{\nu}\omega_k}{q}H.
\end{aligned}$$

This proves the first statement of Proposition 2.4. If $f_X\in\mathscr{U}$, then

$$2\pi\sum_{j=-\infty}^{\infty}h(\theta+2\pi j)=1,\quad\forall\,\theta\in\mathbb{R},$$

which implies from Proposition 2.1 that

$$E|\breve{\omega}_0-\omega_0| \leq \frac{2}{q}\sum_{k=1}^{q+1}E|(\check{M}_q-M_q)_{q+1,k}|,$$

and the rest of the argument is as before. □

The following two theorems establish upper bounds to the minimax convergence rate of $\breve{\omega}_0$ with respect to $f_X\in\mathscr{F}_C$ and f_Z suitably smooth. The proof of Theorem 2.7 is almost the same as the proof of Theorem 2.2 and the proof of Theorem 2.8 is almost the same as the proof of Theorem 2.3.

Theorem 2.7. *Let $\check{\omega}_0$ be as in (2.11)and α be as in (2.3) and there exists a constant $c_0 > 0$ such that $\alpha(l) \leq e^{-c_0 l}$ for all integers $l \geq 1$. Suppose f_Z is supersmooth of order β, $\varphi_{f_Z}(k) \neq 0$ for all $k \in \mathbb{Z}$. Then by choosing $q = \lfloor (c \log n)^{\frac{1}{\beta}} \rfloor$ for some constant $0 < c < \gamma/2$, we have, for any $C > 0$,*

$$\sup_{f_X \in \mathscr{F}_C} E_{f_X} |\check{\omega}_0 - \omega_0| = \mathrm{O}(\frac{1}{\log^{1/\beta}(m \wedge n)}), \quad \text{as } m \wedge n \to \infty.;$$

By choosing q to be a constant, we have

$$\sup_{f_X \in \mathscr{U}} E_{f_X} |\check{\omega}_0 - \omega_0| = \mathrm{O}(\frac{1}{\sqrt{m \wedge n}}), \quad \text{as } m \wedge n \to \infty.$$

Theorem 2.8. *Let $\check{\omega}_0$ be as in (2.11)and α be as in (2.3) and there exists a constant $c_0 > 0$ such that $\alpha(l) \leq e^{-c_0 l}$ for all integers $l \geq 1$. Suppose f_Z is ordinary smooth of order β, $\varphi_{f_Z}(k) \neq 0$ for all $k \in \mathbb{Z}$. Then by choosing $q = \lfloor c n^{1/(2\beta+2)} \rfloor$ for some constant $c > 0$, we have, for any $C > 0$,*

$$\sup_{f_X \in \mathscr{F}_C} E_{f_X} |\check{\omega}_0 - \omega_0| = \mathrm{O}(\frac{1}{(m \wedge n)^{1/(2\beta+2)}});$$

By choosing q to be a constant, we have

$$\sup_{f_X \in \mathscr{U}} E_{f_X} |\check{\omega}_0 - \omega_0| = \mathrm{O}(\frac{1}{\sqrt{m \wedge n}}), \quad \text{as } m \wedge n \to \infty.$$

2.7 Numerical study

Assuming that the noise is unknown and we have two observation sequences: a signal-plus-noise sequence $X_1, ..., X_n$ and a pure noise sequence $Y_1, ..., Y_m$. In this section, we perform a few simulation studies to investigate the performance and properties of our estimator $\breve{\omega}_0$ (see (2.11)). We also compare our estimator to the following empirical Bayes estimators derived from the posterior median estimators of signal sequences in Johnstone and Silverman (2004):

1) ω_{lap}: Proportion of zeros in the empirical Bayes estimator of signals using posterior median when a Laplace prior is used for the non-zero part of the signal and the noise is Gaussian.

2) ω_{cauchy}: Proportion of zeros in the empirical Bayes estimator of signals using posterior median when a Cauchy prior is used for the non-zero part of the signal and the noise is Gaussian.

To compute these two estimators, the R package "EbayesThresh" is used in this simulation. The standard deviation of the noise is estimated from the pure noise data $Y_1, ..., Y_m$.

Recall that in the discussion after we define $\hat{\omega}_0$ in (2.4), a naive estimator inspired by Proposition 2.1 is given by $(\lambda_1(\hat{M}_q)/(q+1) \ \wedge \ 1)$. Similarly, we define $\breve{\omega}_0^* = (\lambda_1(\check{M}_q)/(q+1) \ \wedge \ 1)$. For a given q, we have

(i)if $\lambda_1(\hat{M}_q) \geq q+1$, $\breve{\omega}_0 = \breve{\omega}_0^* = 1$;

(ii)if $\lambda_1(\hat{M}_q) \leq 1$, $\breve{\omega}_0 = 0$, $\breve{\omega}_0^* \leq \frac{1}{q+1}$;

(iii) else $\breve{\omega}_0^* - \breve{\omega}_0 = \frac{1}{q}(1 - \breve{\omega}_0^*)$.

Therefore, Theorems 2.7 and 2.8 are also true for $\breve{\omega}_0^*$. In this simulation, we also compute $\breve{\omega}_0^*$, so that we can check whether there is any significant difference between $\breve{\omega}_0$ and $\breve{\omega}_0^*$ in finite sample simulations.

As in Section 2.1, $X_i = \Theta_i + Z_i, 1 \leq i \leq n$, where n is the sample size. For the true signals $\Theta_1, ..., \Theta_n$, we set $\omega_0 n$ of them to be zero and generate the nonzero Θ_i's using different types of prior distributions given below:

P1. $\Theta = 3$;

P2. $\Theta = 5$;

P3. N(0,10);

P4. $10\exp(1)$;

P5. N(2,1);

P6. exp(0.25);

P7. $U(1, 1 + 2\pi)$.

In models $\mathbf{P1} - \mathbf{P4}$, the signal strength is relatively strong in that most of the nonzero θ_i's are 3 times of the standard deviation of the noise away from 0, while in models **P5 − P7**, the generated nonzero θ_i's are closer to zero.

We generate Z_i using the following different types of distributions:

N1. N(0,1);

N2. $t_5/\sqrt{5/3}$, where dividing by $\sqrt{5/3}$ is to normalize the noise such that $Var(Z_i) = 1$;

N3. $SN(0,1,1)/\sqrt{1-1/\pi}$: skewed normal with location = 0, scale = 1, shape = 1, where deviding by $\sqrt{1-1/\pi}$ is to normalize the noise such that $Var(Z_i) = 1$;

In this simulation, we set the sample size $n = 1000$, and the sample size of the pure noise $m = 2000$. We consider three different types of distributions (**N1** − **N3**) for the noise. For each type of noise, we set the proportion of zeros in the signals (ω_0) to be 0.9, 0.75 and 0.5, and generate the nonzero signals according to the seven types of prior distributions **P1** − **P7**. Once $\breve{\omega}_0$ is evaluated, let k be the integer satisfying $(k-1)/n < \breve{\omega}_0 \leq k/n$. We estimate the set $\Xi = \{1 \leq j \leq n : \theta_j = 0\}$ by $\hat{\Xi} = \{1 \leq j \leq n : |X_j| \leq X_{(k)}\}$, where $X_{(1)} \leq ... \leq X_{(n)}$ are the order statistics of $|X_1|, ..., |X_n|$. The mean and standard deviation of the mean over 100 replications of the following four quantities are computed:

(i) l1-loss: $|\text{estimator} - \omega_0|$;

(ii) Error1: number of times of estimating zero to be nonzero. For our estimator $\breve{\omega}_0$, Error1= $\#\{1 \le j \le n : j \in \Xi, j \notin \hat{\Xi}\}$;

(iii) Error2: number of times of estimating nonzero to be zero. For our estimator $\breve{\omega}_0$, Error2= $\#\{1 \le j \le n : j \notin \Xi, j \in \hat{\Xi}\}$.

(iv) Signal-l1-loss. In addition to comparing the estimate of ω_0, we also compare the L1 loss of the signal sequence defined as: Signal-l1-loss= $\frac{1}{n}\sum_{i=1}^{n} |\hat{\theta}_i - \theta_i|$. For the empirical Bayes method, we use the estimators $\hat{\theta} = (\hat{\theta}_1, ..., \hat{\theta}_n)$ derived from posterior median. In our case, we use a naive estimator by thresholding the empirical Bayes estimator using posterior mean under Laplace prior. More precisely, let $\check{\theta} = (\check{\theta}_1, ..., \check{\theta}_n)$ be the empirical Bayes estimator using posterior mean. Once an estimate $\breve{\omega}_0$ is obtained, let k be the integer satisfying $(k-1)/n < \breve{\omega}_0 \le k/n$ and let $\check{\theta}_{(1)} \le \check{\theta}_{(2)} \le ... \le \check{\theta}_{(n)}$ be the ordered statistics of $|\check{\theta}_1|, ..., |\check{\theta}_n|$. We define the naive estimator $\hat{\theta} = (\hat{\theta}_1, ..., \hat{\theta}_n)$ as:

$$\hat{\theta}_i = \check{\theta}_1 I_{\{|\check{\theta}_i| \ge \check{\theta}_{(k)}\}}, \quad i = 1, ..., n. \tag{2.16}$$

One of the important problem in simulation or in practice would be how to choose a proper q? Write the estimator for a corresponding q as $\breve{\omega}_0^q$. We choose q in the following way:

Step 1. Normalize the data by dividing $X_1, ..., X_n, Y_1, ..., Y_m$ by $10\sigma_Y$, where

σ_Y is the sample standard deviation of $Y_1, ..., Y_m$. We use the same notation for the normalized data.

Step 2. Let $u = 10\sqrt{\log n}$. For each $i = (1 \wedge (u-10)), ..., u+10$, computer $\breve{\omega}_0^{1\wedge(u-10)}, ..., \breve{\omega}_0^{u+10}$ and let k_i be the integer satisfying $(k_i - 1)/n < \breve{\omega}_0^i \leq k_i/n$. Let $X_{(1)} \leq X_{(2)} \leq ... \leq X_{(n)}$ be the ordered statistics of $|X_1|, ..., |X_n|$. Then define:

$$\hat{\Xi}_i = \{1 \leq i \leq n : |X_i| \leq X_{(k_i)}\}.$$

Let F_i be the empirical cumulative distribution function of $\{X_j, j \in \hat{\Xi}_i\}$, for $i = (1 \wedge (u-10)), ..., u+10$ and F_Y be the empirical cumulative distribution function of $Y_1, ..., Y_m$. Define the Kolmogorov-Smirnov distance between $\{X_j, j \in \hat{\Xi}_i\}$ and $Y_1, ..., Y_m$ as:

$$D_i = \sup_{x\in\mathbb{R}} |F_i(x) - F_Y(x)|, \quad i = (1 \wedge (u-10)), ..., u+10.$$

Step 3. Choose q such that D_q is the minimum among $D_{(1\wedge(u-10))}, ..., D_{u+10}$.

The reason for defining u in Step 2 is to make the computation more efficient when n is large. If n is not large, after normalizing the data, we can compute $\breve{\omega}_0^1, ..., \breve{\omega}_0^{\lfloor\sqrt{n}\rfloor}$ in Step 2 instead.

Algorithms in choosing q and computation of quantities such as Error1 and Error2 for $\breve{\omega}_0^*$ are similar as $\breve{\omega}_0$.

Conclusions

- In terms of the l1-loss, $\breve{\omega}_0$ performs well for all the cases simulated while the empirical Bayes estimators based on posterior median only do well for sparse cases (ω_0=0.9 and 0.75 in some cases).

- $\breve{\omega}_0$ and $\breve{\omega}_0^*$ are very close as expected. There are no significant differences between these two estimators.

- Notice that when $\omega_0 = 0.9$, the empirical Bayes estimators perform quite well under normal or t noise. However, in the skewed normal noise case, they do not perform very well. In addition, in the skewed noise case, the l1-loss of the empirical Bayes estimators is larger. On the other hand, $\breve{\omega}_0$ is relatively robust to different types of noise distributions.

- The Signal-l1-loss of the naive thresholding estimator based on $\breve{\omega}_0$ (see (2.16)) is smaller when ω is small ($\omega = 0.75, 0.5$.)

- In terms of Error1 and Error2, when $\omega = 0.75, 0.5$, $|\text{Error1} - \text{Error2}|$ of $\breve{\omega}_0$ and $\breve{\omega}_0^*$ is relatively smaller, implying that $\breve{\omega}_0$ and $\breve{\omega}_0^*$ are relatively less biased than the empirical Bayes estimators. In addition, when $\omega = 0.75, 0.5$, $\text{Error1} + \text{Error2}$ values of $\breve{\omega}_0$ and $\breve{\omega}_0^*$ are also slightly smaller relatively to $\text{Error1} + \text{Error2}$ values of the empirical Bayes estimators.

ω_0		$\breve{\omega}_0$	$\breve{\omega}_0^*$	ω_{lap}	ω_{cauchy}
0.9	l1-loss(sd)	0.021(0.002)	0.020(0.002)	0.019(0.001)	0.029(0.001)
	Error1(sd)	26.3(1.9)	25.5(2.0)	12.0(0.5)	8.3(0.4)
	Error2(sd)	24.1(0.9)	24.7(0.9)	31.5(0.8)	37.0(0.8)
	Signal-l1-loss(sd)	0.176(0.002)	0.176(0.002)	0.170(0.002)	0.181(0.002)
0.75	l1-loss(sd)	0.017(0.001)	0.015(0.001)	0.052(0.002)	0.042(0.002)
	Error1(sd)	38.7(1.3)	37.1(1.5)	72.0(1.3)	64.6(1.2)
	Error2(sd)	33.9(0.8)	34.9(0.8)	20.2(0.4)	23.0(0.5)
	Signal-l1-loss(sd)	0.340(0.002)	0.340(0.002)	0.342(0.002)	0.358(0.002)
0.5	l1-loss(sd)	0.018(0.001)	0.017(0.001)	0.468(0.003)	0.5(0)
	Error1(sd)	46.8(1.3)	39.3(1.0)	468.1(3.1)	500(0)
	Error2(sd)	49.1(1.2)	55.4(0.6)	0.2(<0.1)	0(0)
	Signal-l1-loss(sd)	0.555(0.001)	0.554(0.001)	0.692(0.003)	0.702(0.001)

Table 2.1 Simulation results under the model $\mathbf{P1} + \mathbf{N1}$:Nonzero $\Theta_i = 3$ and $Z_i \sim N(0,1)$.

ω_0		$\breve{\omega}_0$	$\breve{\omega}_0^*$	ω_{lap}	ω_{cauchy}
0.9	l1-loss(sd)	0.022(0.001)	0.019(0.001)	0.027(0.001)	0.024(0.001)
	Error1(sd)	18.1(1.7)	14.9(1.7)	28.1(0.7)	24.0(0.6)
	Error2(sd)	4.9(0.8)	6.3(1.0)	0.2(<0.1)	0.3(<0.1)
	Signal-l1-loss(sd)	0.132(0.002)	0.132(0.002)	0.129(0.001)	0.121(0.001)
0.75	l1-loss(sd)	0.012(0.001)	0.016(0.001)	0.151(0.002)	0.180(0.002)
	Error1(sd)	7.2(0.6)	7.0(0.7)	151.0(1.9)	179.3(2.1)
	Error2(sd)	7.2(0.9)	10.7(1.2)	0(0)	0(0)
	Signal-l1-loss(sd)	0.255(0.002)	0.265(0.002)	0.335(0.002)	0.323(0.001)
0.5	l1-loss(sd)	0.017(0.001)	0.018(0.001)	0.468(0.003)	0.5(0)
	Error1(sd)	46.8(1.4)	39.2(1.1)	468.1(3.2)	500(0)
	Error2(sd)	49.1(1.2)	55.4(0.6)	0.2(<0.1)	0(0)
	Signal-l1-loss(sd)	0.554(0.001)	0.556(0.001)	0.692(0.003)	0.701(0.001)

Table 2.2 Simulation results under the model $\mathbf{P2} + \mathbf{N1}$:Nonzero $\Theta_i = 5$ and $Z_i \sim N(0, 1)$.

ω_0		$\breve{\omega}_0$	$\breve{\omega}_0^*$	ω_{lap}	ω_{cauchy}
	l1-loss(sd)	0.017(0.001)	0.015(0.001)	0.006(0.001)	0.007(0.001)
0.9	Error1(sd)	13.0(1.4)	9.8(0.8)	20.1(0.6)	14.0(0.4)
	Error2(sd)	22.7(0.6)	23.0(0.6)	18.7(0.3)	19.9(0.4)
	Signal-l1-loss(sd)	0.124(0.001)	0.122(0.001)	0.124(0.001)	0.116(0.001)
	l1-loss(sd)	0.016(0.001)	0.022(0.001)	0.064(0.001)	0.068(0.002)
0.75	Error1(sd)	27.7(1.3)	23.3(1.4)	94.7(1.0)	98.5(1.2)
	Error2(sd)	41.7(0.5)	43.1(0.5)	30.7(0.6)	30.5(0.6)
	Signal-l1-loss(sd)	0.282(0.001)	0.278(0.001)	0.311(0.001)	0.288(0.001)
	l1-loss(sd)	0.052(0.002)	0.065(0.002)	0.474(0.003)	0.5(0)
0.5	Error1(sd)	25.7(0.9)	18.9(1.0)	476.7(2.9)	500(0)
	Error2(sd)	77.4(1.0)	84.4(1.4)	3.1(0.5)	0(0)
	Signal-l1-loss(sd	0.505(0.002)	0.503(0.002)	0.690(0.002)	0.619(0.001)

Table 2.3 Simulation results under the model $\mathbf{P3} + \mathbf{N1}$:Nonzero $\Theta_i \sim N(0, 10)$ and $Z_i \sim N(0, 1)$.

ω_0		$\breve{\omega}_0$	$\breve{\omega}_0^*$	ω_{lap}	ω_{cauchy}
0.9	l1-loss(sd)	0.017(0.001)	0.016(0.001)	0.006(0.001)	0.007(0.001)
	Error1(sd)	15.9(1.8)	14.7(1.7)	18.8(0.6)	14.1(0.5)
	Error2(sd)	21.8(0.5)	21.7(0.5)	19.5(0.4)	20.5(0.4)
	Signal-l1-loss(sd)	0.125(0.002)	0.124(0.002)	0.125(0.001)	0.116(0.001)
0.75	l1-loss(sd)	0.023(0.001)	0.025(0.001)	0.056(0.002)	0.057(0.003)
	Error1(sd)	26.2(1.3)	25.0(1.4)	89.6(2.0)	90.9(2.1)
	Error2(sd)	45.7(0.8)	46.3(0.8)	33.6(0.5)	33.7(0.6)
	Signal-l1-loss(sd)	0.277(0.003)	0.277(0.003)	0.306(0.002)	0.285(0.002)
0.5	l1-loss(sd)	0.046(0.002)	0.055(0.002)	0.459(0.003)	0.5(0)
	Error1(sd)	34.5(1.5)	29.8(1.4)	463.7(2.5)	500(0)
	Error2(sd)	80.8(0.8)	84.8(0.9)	4.9(0.4)	0(0)
	Signal-l1-loss(sd)	0.511(0.003)	0.509(0.002)	0.680(0.002)	0.619(0.002)

Table 2.4 Simulation results under the model $\mathbf{P4}+\mathbf{N1}$:Nonzero $\Theta_i \sim 10\exp(1)$ and $Z_i \sim N(0,1)$.

ω_0		$\breve{\omega}_0$	$\breve{\omega}_0^*$	ω_{lap}	ω_{cauchy}
0.9	l1-loss(sd)	0.034(0.002)	0.035(0.002)	0.068(0.001)	0.074(0.001)
	Error1(sd)	24.9(1.5)	24.2(1.5)	4.5(0.3)	2.5(0.3)
	Error2(sd)	58.2(0.9)	58.3(0.9)	72.1(0.6)	76.4(0.5)
	Signal-l1-loss(sd)	0.155(0.001)	0.154(0.001)	0.152(0.001)	0.157(0.001)
0.75	l1-loss(sd)	0.063(0.003)	0.070(0.004)	0.121(0.002)	0.138(0.001)
	Error1(sd)	55.1(2.6)	50.5(2.9)	20.3(0.8)	14.0(0.6)
	Error2(sd)	113.0(1.8)	116.3(1.9)	140.9(1.1)	151.7(1.1)
	Signal-l1-loss(sd)	0.334(0.002)	0.334(0.002)	0.331(0.002)	0.345(0.002)
0.5	l1-loss(sd)	0.109(0.004)	0.116(0.004)	0.079(0.004)	0.086(0.004)
	Error1(sd)	68.1(2.0)	64.9(2.0)	84.8(2.2)	81.6(2.5)
	Error2(sd)	176.9(1.9)	180.3(2.0)	161.4(2.3)	164.7(2.7)
	Signal-l1-loss(sd)	0.567(0.002)	0.557(0.002)	0.559(0.002)	0.580(0.002)

Table 2.5 Simulation results under the model $\mathbf{P5} + \mathbf{N1}$:Nonzero $\Theta_i \sim N(2, 1)$ and $Z_i \sim N(0, 1)$.

ω_0		$\breve{\omega}_0$	$\breve{\omega}_0^*$	ω_{lap}	ω_{cauchy}
0.9	l1-loss(sd)	0.030(0.002)	0.031(0.002)	0.034(0.001)	0.041(0.001)
	Error1(sd)	16.0(1.3)	15.6(1.3)	10.5(0.5)	6.6(0.3)
	Error2(sd)	44.7(0.9)	45.1(0.9)	44.8(0.4)	47.2(0.5)
	Signal-l1-loss(sd)	0.126(0.001)	0.126(0.001)	0.120(0.001)	0.118(0.001)
0.75	l1-loss(sd)	0.060(0.003)	0.062(0.003)	0.048(0.001)	0.057(0.001)
	Error1(sd)	33.7(1.6)	32.4(1.7)	41.5(0.7)	35.1(0.7)
	Error2(sd)	93.5(1.2)	94.6(1.2)	89.9(1.0)	92.6(0.9)
	Signal-l1-loss(sd)	0.279(0.002)	0.278(0.002)	0.280(0.001)	0.273(0.001)
0.5	l1-loss(sd)	0.128(0.003)	0.133(0.003)	0.093(0.004)	0.168(0.005)
	Error1(sd)	39.4(2.0)	37.5(1.8)	181.9(2.7)	239.5(4.0)
	Error2(sd)	167.5(1.6)	170.1(1.7)	90.6(1.2)	71.3(1.5)
	Signal-l1-loss(sd)	0.498(0.002)	0.497(0.003)	0.530(0.002)	0.523(0.002)

Table 2.6 Simulation results under the model $\mathbf{P6} + \mathbf{N1}$:Nonzero $\Theta_i \sim \exp(0.25)$ and $Z_i \sim N(0, 1)$.

ω_0		$\breve{\omega}_0$	$\breve{\omega}_0^*$	ω_{lap}	ω_{cauchy}
0.9	l1-loss(sd)	0.015(0.001)	0.015(0.001)	0.007(<0.001)	0.011(0.001)
	Error1(sd	22.6(1.4)	21.3(1.4)	17.8(0.4)	13.4(0.4)
	Error2(sd)	21.5(0.6)	22.0(0.6)	22.2(0.5)	24.2(0.5)
	Signal-l1-loss(sd)	0.143(0.001)	0.143(0.001)	0.139(0.001)	0.136(0.001)
0.75	l1-loss(sd)	0.021(0.001)	0.026(0.002)	0.052(0.002)	0.051(0.002)
	Error1(sd)	29.5(1.5)	24.6(1.4)	83.7(1.4)	83.1(1.5)
	Error2(sd)	46.3(0.6)	49.0(0.7)	31.8(0.5)	32.0(0.5)
	Signal-l1-loss(sd)	0.309(0.002)	0.306(0.002)	0.326(0.002)	0.320(0.002)
0.5	l1-loss(sd)	0.035(0.002)	0.036(0.002)	0.470(0.003)	0.5(0)
	Error1(sd)	42.8(1.9)	42.4(1.9)	473.0(2.3)	500(0)
	Error2(sd)	71.7(1.2)	72.2(1.2)	2.3(0.2)	0(0)
	Signal-l1-loss(sd)	0.554(0.002)	0.553(0.002)	0.697(0.002)	0.663(0.002)

Table 2.7 Simulation results under the model $\mathbf{P7} + \mathbf{N1}$:Nonzero $\Theta_i \sim U(1, 1 + 2\pi)$ and $Z_i \sim N(0, 1)$.

ω_0		$\breve{\omega}_0$	$\breve{\omega}_0^*$	ω_{lap}	ω_{cauchy}
0.9	l1-loss(sd)	0.018(0.001)	0.017(0.001)	0.009(0.001)	0.013(0.001)
	Error1(sd)	27.8(1.4)	26.3(1.4)	23.8(0.6)	19.0(0.6)
	Error2(sd)	23.2(1.1)	24.2(1.0)	24.4(0.6)	30.3(0.7)
	Signal-l1-loss(sd)	0.196(0.002)	0.196(0.002)	0.191(0.002)	0.202(0.002)
0.75	l1-loss(sd)	0.016(0.001)	0.016(0.001)	0.046(0.002)	0.041(0.002)
	Error1(sd)	40.3(1.2)	38.2(1.2)	64.6(1.6)	60.8(1.7)
	Error2(sd)	31.2(0.9)	33.5(0.9)	18.8(0.6)	20.4(0.7)
	Signal-l1-loss(sd)	0.349(0.002)	0.349(0.002)	0.348(0.002)	0.369(0.002)
0.5	l1-loss(sd)	0.012(0.001)	0.011(0.001)	0.450(0.005)	0.5(0)
	Error1(sd)	39.0(1.0)	22.6(0.7)	448.0(4.6)	500(0)
	Error2(sd)	39.6(0.9)	46.4(0.6)	0.7(0.1)	0(0)
	Signal-l1-loss(sd)	0.524(0.003)	0.525(0.003)	0.648(0.003)	0.669(0.002)

Table 2.8 Simulation results under the model **P1** + **N2**:Nonzero $\Theta_i = 3$ and $Z_i \sim t_5/\sqrt{5/3}$.

ω_0		$\breve{\omega}_0$	$\breve{\omega}_0^*$	ω_{lap}	ω_{cauchy}
0.9	l1-loss(sd)	0.018(0.001)	0.017(0.001)	0.035(0.001)	0.029(0.001)
	Error1(sd)	11.1(1.2)	10.2(0.9)	36.7(1.1)	30.6(0.8)
	Error2(sd)	9.1(0.8)	5.9(0.5)	1.3(0.3)	1.1(0.2)
	Signal-l1-loss(sd)	0.142(0.002)	0.141(0.001)	0.150(0.002)	0.141(0.001)
0.75	l1-loss(sd)	0.015(0.001)	0.016(0.001)	0.114(0.003)	0.125(0.003)
	Error1(sd)	16.4(1.4)	16.0(1.3)	114.7(1.7)	133.5(1.9)
	Error2(sd)	9.5(0.8)	11.1(0.9)	0.4(0.1)	1.5(0.3)
	Signal-l1-loss(sd)	0.269(0.002)	0.270(0.002)	0.330(0.002)	0.319(0.002)
0.5	l1-loss(sd)	0.012(0.001)	0.012(0.001)	0.5(0)	0.5(0)
	Error1(sd)	8.7(0.6)	6.7(0.5)	500(0)	500(0)
	Error2(sd)	10.1(0.7)	11.3(0.8)	0(0)	0(0)
	Signal-l1-loss(sd)	0.474(0.002)	0.475(0.002)	0.679(0.002)	0.618(0.002)

Table 2.9 Simulation results under the model **P2** + **N2**:Nonzero $\Theta_i = 5$ and $Z_i \sim t_5/\sqrt{5/3}$.

ω_0		$\breve{\omega}_0$	$\breve{\omega}_0^*$	ω_{lap}	ω_{cauchy}
0.9	l1-loss(sd)	0.016(0.002)	0.014(0.002)	0.011(0.001)	0.008(0.001)
	Error1(sd)	16.5(1.1)	12.3(1.0)	28.4(0.5)	25.2(0.7)
	Error2(sd)	24.7(1.2)	25.8(1.1)	17.7(0.9)	16.3(0.3)
	Signal-l1-loss(sd)	0.149(0.003)	0.148(0.003)	0.145(0.001)	0.139(0.001)
0.75	l1-loss(sd)	0.024(0.002)	0.026(0.002)	0.035(0.002)	0.036(0.002)
	Error1(sd)	23.5(0.9)	19.9(0.8)	66.6(1.3)	68.1(1.4)
	Error2(sd)	45.3(1.2)	49.8(1.3)	30.2(0.7)	30.0(0.7)
	Signal-l1-loss(sd)	0.286(0.002)	0.287(0.002)	0.300(0.002)	0.290(0.002)
0.5	l1-loss(sd)	0.046(0.002)	0.049(0.002)	0.425(0.005)	0.500(0.001)
	Error1(sd)	31.9(1.1)	28.8(1.0)	465.5(3.9)	499.8(0.3)
	Error2(sd)	75.8(1.0)	79.9(0.9)	7.1(0.4)	0.02(0.01)
	Signal-l1-loss(sd)	0.505(0.002)	0.502(0.002)	0.629(0.003)	0.585(0.002)

Table 2.10 Simulation results under the model $\mathbf{P3} + \mathbf{N2}$:Nonzero $\Theta_i \sim N(0, 10)$ and $Z_i \sim t_5/\sqrt{5/3}$.

ω_0		$\breve{\omega}_0$	$\breve{\omega}_0^*$	ω_{lap}	ω_{cauchy}
0.9	l1-loss(sd)	0.020(0.001)	0.019(0.001)	0.013(0.001)	0.010(0.001)
	Error1(sd)	19.6(1.2)	18.7(1.1)	29.6(1.0)	26.7(0.6)
	Error2(sd)	26.8(0.9)	26.0(0.8)	20.1(0.4)	23.5(0.3)
	Signal-l1-loss(sd)	0.150(0.002)	0.149(0.002)	0.153(0.002)	0.145(0.001)
0.75	l1-loss(sd)	0.023(0.002)	0.024(0.002)	0.030(0.002)	0.032(0.002)
	Error1(sd)	29.4(1.3)	28.7(1.2)	64.7(1.2)	67.2(1.6)
	Error2(sd)	46.5(0.7)	46.6(0.8)	39.1(0.6)	38.3(0.7)
	Signal-l1-loss(sd)	0.290(0.002)	0.291(0.002)	0.303(0.002)	0.290(0.002)
0.5	l1-loss(sd)	0.045(0.002)	0.050(0.002)	0.393(0.005)	0.499(0.001)
	Error1(sd)	36.7(1.3)	32.5(1.2)	401.3(4.6)	499.1(0.9)
	Error2(sd)	79.9(1.2)	84.4(1.2)	11.1(0.5)	0.2(0.1)
	Signal-l1-loss(sd)	0.511(0.002)	0.509(0.002)	0.617(0.003)	0.588(0.002)

Table 2.11 Simulation results under the model **P4**+**N2**:Nonzero $\Theta_i \sim 10\exp(1)$ and $Z_i \sim t_5/\sqrt{5/3}$.

ω_0		$\breve{\omega}_0$	$\breve{\omega}_0^*$	ω_{lap}	ω_{cauchy}
0.9	l1-loss(sd)	0.025(0.002)	0.026(0.002)	0.056(0.001)	0.063(0.001)
	Error1(sd)	35.4(1.7)	36.0(1.7)	13.6(0.6)	10.9(0.5)
	Error2(sd)	53.4(0.9)	53.2(0.9)	69.8(0.7)	74.8(0.7)
	Signal-l1-loss(sd)	0.182(0.001)	0.181(0.001)	0.181(0.001)	0.184(0.001)
0.75	l1-loss(sd)	0.051(0.002)	0.055(0.003)	0.115(0.002)	0.130(0.002)
	Error1(sd)	58.3(1.7)	56.6(1.9)	28.7(0.7)	24.5(0.8)
	Error2(sd)	107.7(1.2)	110.0(1.3)	142.9(1.3)	152.6(1.5)
	Signal-l1-loss(sd)	0.356(0.002)	0.356(0.002)	0.365(0.003)	0.380(0.002)
0.5	l1-loss(sd)	0.095(0.003)	0.099(0.003)	0.112(0.004)	0.122(0.005)
	Error1(sd)	61.4(1.3)	60.2(1.2)	63.4(1.7)	60.7(1.5)
	Error2(sd)	158.3(1.5)	160.6(1.6)	166.9(2.3)	171.5(3.3)
	Signal-l1-loss(sd)	0.556(0.002)	0.555(0.02)	0.568(0.002)	0.601(0.002)

Table 2.12 Simulation results under the model $\mathbf{P5} + \mathbf{N2}$:Nonzero $\Theta_i \sim N(2,1)$ and $Z_i \sim t_5/\sqrt{5/3}$.

ω_0		$\breve{\omega}_0$	$\breve{\omega}_0^*$	ω_{lap}	ω_{cauchy}
0.9	l1-loss(sd)	0.024(0.002)	0.025(0.002)	0.024(0.001)	0.029(0.001)
	Error1(sd)	24.8(1.7)	25.7(1.6)	21.9(0.7)	17.9(0.6)
	Error2(sd)	43.7(0.8)	44.3(0.8)	46.3(0.7)	47.5(0.7)
	Signal-l1-loss(sd)	0.152(0.002)	0.152(0.002)	0.151(0.002)	0.147(0.001)
0.75	l1-loss(sd)	0.049(0.003)	0.051(0.003)	0.048(0.002)	0.053(0.002)
	Error1(sd)	46.2(1.7)	44.4(1.6)	45.2(1.3)	40.3(1.2)
	Error2(sd)	96.7(1.3)	99.3(1.2)	94.3(1.0)	96.5(1.2)
	Signal-l1-loss(sd)	0.301(0.002)	0.301(0.002)	0.301(0.002)	0.299(0.002)
0.5	l1-loss(sd)	0.089(0.003)	0.094(0.003)	0.043(0.003)	0.010(0.005)
	Error1(sd)	51.1(1.3)	49.9(1.4)	124.7(2.4)	168.3(2.9)
	Error2(sd)	152.9(1.4)	150.8(1.5)	102.1(1.8)	85.2(2.1)
	Signal-l1-loss(sd)	0.503(0.002)	0.503(0.002)	0.513(0.002)	0.512(0.002)

Table 2.13 Simulation results under the model **P6**+**N2**:Nonzero $\Theta_i \sim \exp(0.25)$ and $Z_i \sim t_5/\sqrt{5/3}$.

ω_0		$\breve{\omega}_0$	$\breve{\omega}_0^*$	ω_{lap}	ω_{cauchy}
0.9	l1-loss(sd)	0.015(0.001)	0.018(0.001)	0.008(0.001)	0.008(0.001)
	Error1(sd)	19.6(1.2)	17.8(1.2)	26.2(0.7)	21.4(0.5)
	Error2(sd)	26.5(0.7)	28.4(0.8)	22.8(0.6)	24.7(0.6)
	Signal-l1-loss(sd)	0.164(0.002)	0.165(0.002)	0.166(0.002)	0.162(0.002)
0.75	l1-loss(sd)	0.023(0.002)	0.027(0.002)	0.035(0.002)	0.034(0.002)
	Error1(sd)	36.0(1.8)	34.3(1.9)	64.5(1.5)	63.7(1.6)
	Error2(sd)	44.9(1.0)	47.3(1.2)	32.5(0.8)	32.7(0.9)
	Signal-l1-loss(sd)	0.322(0.002)	0.323(0.002)	0.327(0.002)	0.325(0.002)
0.5	l1-loss(sd)	0.027(0.002)	0.024(0.002)	0.413(0.005)	5.0(0)
	Error1(sd)	47.5(1.8)	44.9(1.8)	417.2(4.6)	500(0)
	Error2(sd)	60.6(0.9)	62.5(1.0)	3.9(0.3)	0(0)
	Signal-l1-loss(sd)	0.529(0.002)	0.529(0.002)	0.629(0.003)	0.622(0.002)

Table 2.14 Simulation results under the model $\mathbf{P7}+\mathbf{N2}$:Nonzero $\Theta_i \sim U(1, 1+2\pi)$ and $Z_i \sim t_5/\sqrt{5/3}$.

ω_0		$\breve{\omega}_0$	$\breve{\omega}_0^*$	ω_{lap}	ω_{cauchy}
0.9	l1-loss(sd)	0.015(0.001)	0.014(0.001)	0.095(0.002)	0.077(0.002)
	Error1(sd)	21.5(0.9)	20.0(0.9)	97.7(2.5)	80.8(2.2)
	Error2(sd)	21.3(0.6)	21.5(0.6)	8.0(0.3)	7.3(0.4)
	Signal-l1-loss(sd)	0.170(0.002)	0.170(0.002)	0.227(0.002)	0.210(0.002)
0.75	l1-loss(sd)	0.017(0.001)	0.015(0.001)	0.285(0.004)	0.322(0.005)
	Error1(sd)	26.0(0.8)	25.3(0.9)	288.5(3.8)	324.3(5.5)
	Error2(sd)	34.5(0.8)	33.6(0.8)	1.3(0.1)	1.2(0.2)
	Signal-l1-loss(sd)	0.321(0.002)	0.320(0.002)	0.459(0.003)	0.445(0.003)
0.5	l1-loss(sd)	0.023(0.001)	0.024(0.001)	0.5(0)	0.5(0)
	Error1(sd)	23.9(0.6)	22.4(0.6)	500(0)	500(0)
	Error2(sd)	47.4(0.7)	48.5(0.7)	0(0)	0(0)
	Signal-l1-loss(sd)	0.531(0.002)	0.530(0.002)	0.743(0.002)	0.705(0.002)

Table 2.15 Simulation results under the model **P1** + **N3**:Nonzero $\Theta_i = 3$ and $Z_i \sim SN(0,1,1)/\sqrt{1-1/\pi}$.

ω_0		$\breve{\omega}_0$	$\breve{\omega}_0^*$	ω_{lap}	ω_{cauchy}
0.9	l1-loss(sd)	0.012(0.001)	0.012(0.001)	0.107(0.002)	0.093(0.002)
	Error1(sd)	7.5(0.8)	6.9(0.8)	107.5(2.1)	93.0(1.9)
	Error2(sd)	7.4(0.8)	7.7(0.8)	0.01(0.01)	0.01(0.01)
	Signal-l1-loss(sd)	0.123(0.003)	0.123(0.003)	0.219(0.002)	0.193(0.002)
0.75	l1-loss(sd)	0.012(0.001)	0.013(0.001)	0.326(0.004)	0.403(0.005)
	Error1(sd)	8.7(0.8)	7.9(0.8)	326.3(3.5)	402.5(5.0)
	Error2(sd)	8.9(0.9)	9.3(0.8)	0(0)	0(0)
	Signal-l1-loss(sd)	0.244(0.003)	0.246(0.003)	0.475(0.003)	0.461(0.003)
0.5	l1-loss(sd)	0.011(0.001)	0.013(0.001)	0.5(0)	0.5(0)
	Error1(sd)	4.6(0.6)	3.3(0.6)	500(0)	500(0)
	Error2(sd)	10.1(1.1)	11.2(1.0)	0(0)	0(0)
	Signal-l1-loss(sd)	0.445(0.003)	0.447(0.003)	0.748(0.002)	0.703(0.002)

Table 2.16 Simulation results under the model $\mathbf{P2} + \mathbf{N3}$:Nonzero $\Theta_i = 5$ and $Z_i \sim SN(0,1,1)/\sqrt{1-1/\pi}$.

ω_0		$\breve{\omega}_0$	$\breve{\omega}_0^*$	ω_{lap}	ω_{cauchy}
0.9	l1-loss(sd)	0.021(0.02)	0.023(0.002)	0.065(0.002)	0.055(0.002)
	Error1(sd)	11.1(1.1)	8.8(1.0)	87.5(1.7)	71.2(1.6)
	Error2(sd)	31.9(1.5)	33.1(1.6)	16.8(0.6)	17.9(0.7)
	Signal-l1-loss(sd)	0.183(0.001)	0.181(0.001)	0.219(0.003)	0.201(0.002)
0.75	l1-loss(sd)	0.032(0.02)	0.035(0.002)	0.213(0.004)	0.227(0.004)
	Error1(sd)	24.1(1.5)	20.3(1.4)	242.8(3.3)	275.6(3.9)
	Error2(sd)	60.5(1.2)	66.3(1.1)	28.6(0.7)	28.5(0.6)
	Signal-l1-loss(sd)	0.343(0.003)	0.344(0.003)	0.487(0.003)	0.465(0.003)
0.5	l1-loss(sd)	0.059(0.003)	0.063(0.003)	0.5(0)	0.5(0)
	Error1(sd)	34.5(1.5)	29.6(1.4)	500(0)	500(0)
	Error2(sd)	98.3(1.4)	100.2(1.4)	0(0)	0(0)
	Signal-l1-loss(sd)	0.621(0.002)	0.626(0.002)	0.839(0.002)	0.771(0.002)

Table 2.17 Simulation results under the model $\mathbf{P3}+\mathbf{N3}$:Nonzero $\Theta_i \sim N(0,10)$ and $Z_i \sim SN(0,1,1)/\sqrt{1-1/\pi}$.

ω_0		$\breve{\omega}_0$	$\breve{\omega}_0^*$	ω_{lap}	ω_{cauchy}
0.9	l1-loss(sd)	0.016(0.001)	0.017(0.001)	0.080(0.002)	0.063(0.002)
	Error1(sd)	12.1(1.1)	11.8(1.1)	93.5(1.7)	77.4(1.7)
	Error2(sd)	22.9(0.8)	24.7(0.9)	17.4(0.6)	15.6(0.5)
	Signal-l1-loss(sd)	0.131(0.002)	0.133(0.002)	0.212(0.002)	0.183(0.002)
0.75	l1-loss(sd)	0.026(0.001)	0.027(0.001)	0.227(0.004)	0.253(0.005)
	Error1(sd)	21.5(1.1)	17.9(1.0)	247.9(3.7)	273.3(5.3)
	Error2(sd)	50.2(0.9)	54.3(0.9)	23.4(0.5)	16.6(0.6)
	Signal-l1-loss(sd)	0.267(0.002)	0.267(0.002)	0.422(0.004)	0.409(0.003)
0.5	l1-loss(sd)	0.054(0.001)	0.055(0.001)	0.5(0)	0.5(0)
	Error1(sd)	24.4(0.7)	23.5(0.7)	500(0)	500(0)
	Error2(sd)	79.4(1.1)	77.8(1.0)	0(0)	0(0)
	Signal-l1-loss(sd)	0.479(0.002)	0.483(0.002)	0.732(0.002)	0.693(0.002)

Table 2.18 Simulation results under the model **P4**+**N3**:Nonzero $\Theta_i \sim 10\exp(1)$ and $Z_i \sim SN(0,1,1)/\sqrt{1-1/\pi}$.

ω_0		$\breve{\omega}_0$	$\breve{\omega}_0^*$	ω_{lap}	ω_{cauchy}
0.9	l1-loss(sd)	0.036(0.002)	0.034(0.002)	0.029(0.002)	0.018(0.001)
	Error1(sd)	23.4(1.1)	22.1(1.2)	61.2(1.6)	45.3(1.3)
	Error2(sd)	56.3(0.8)	59.6(0.8)	40.3(0.7)	42.9(0.6)
	Signal-l1-loss(sd)	0.175(0.002)	0.175(0.002)	0.199(0.002)	0.183(0.002)
0.75	l1-loss(sd)	0.069(0.002)	0.070(0.002)	0.101(0.003)	0.083(0.004)
	Error1(sd)	35.6(1.1)	33.8(1.0)	157.5(2.6)	146.5(2.8)
	Error2(sd)	106.3(1.1)	110.6(1.1)	55.3(1.0)	54.2(1.1)
	Signal-l1-loss(sd)	0.349(0.002)	0.348(0.002)	0.390(0.003)	0.375(0.002)
0.5	l1-loss(sd)	0.143(0.002)	0.144(0.002)	0.425(0.006)	0.484(0.004)
	Error1(sd)	41.3(0.9)	37.7(0.8)	425.4(4.5)	484.2(3.7)
	Error2(sd)	183.2(1.3)	188.8(1.3)	12.5(0.8)	1.1(0.7)
	Signal-l1-loss(sd)	0.581(0.002)	0.580(0.002)	0.681(0.004)	0.657(0.003)

Table 2.19 Simulation results under the model $\mathbf{P5}+\mathbf{N3}$:Nonzero $\Theta_i \sim N(2,1)$ and $Z_i \sim SN(0,1,1)/\sqrt{1-1/\pi}$.

ω_0		$\breve{\omega}_0$	$\breve{\omega}_0^*$	ω_{lap}	ω_{cauchy}
0.9	l1-loss(sd)	0.027(0.002)	0.030(0.002)	0.038(0.002)	0.020(0.001)
	Error1(sd)	15.7(1.0)	14.5(1.0)	67.9(1.5)	51.1(1.1)
	Error2(sd)	42.4(1.0)	43.8(1.0)	29.9(0.5)	31.9(0.6)
	Signal-l1-loss(sd)	0.139(0.001)	0.138(0.001)	0.185(0.002)	0.163(0.001)
0.75	l1-loss(sd)	0.061(0.002)	0.066(0.002)	0.151(0.003)	0.152(0.004)
	Error1(sd)	27.6(0.9)	24.9(0.9)	194.1(2.8)	195.8(3.2)
	Error2(sd)	88.4(0.9)	90.7(0.9)	43.1(0.7)	43.1(0.8)
	Signal-l1-loss(sd)	0.291(0.002)	0.289(0.002)	0.396(0.003)	0.369(0.002)
0.5	l1-loss(sd)	0.134(0.002)	0.138(0.002)	0.459(0.004)	0.498(0.001)
	Error1(sd)	27.2(0.8)	28.6(0.8)	446.7(3.6)	498.3(1.1)
	Error2(sd)	160.6(1.4)	163.4(1.3)	8.7(1.0)	0(0)
	Signal-l1-loss(sd)	0.518(0.002)	0.517(0.002)	0.711(0.004)	0.655(0.002)

Table 2.20 Simulation results under the model **P6+N3**:Nonzero $\Theta_i \sim \exp(0.25)$ and $Z_i \sim SN(0,1,1)/\sqrt{1-1/\pi}$.

ω_0		$\breve{\omega}_0$	$\breve{\omega}_0^*$	ω_{lap}	ω_{cauchy}
0.9	l1-loss(sd)	0.013(0.001)	0.013(0.001)	0.089(0.002)	0.070(0.002)
	Error1(sd)	14.3(0.9)	13.2(0.8)	97.5(1.8)	78.9(1.5)
	Error2(sd)	21.5(0.7)	22.5(0.8)	8.1(0.3)	9.0(0.3)
	Signal-l1-loss(sd)	0.147(0.001)	0.147(0.001)	0.220(0.002)	0.193(0.002)
0.75	l1-loss(sd)	0.019(0.001)	0.023(0.001)	0.260(0.004)	0.296(0.005)
	Error1(sd)	23.9(0.7)	20.5(0.5)	266.9(3.2)	302.3(4.3)
	Error2(sd)	41.5(0.9)	43.9(0.9)	7.4(0.4)	6.1(0.4)
	Signal-l1-loss(sd)	0.309(0.001)	0.306(0.001)	0.454(0.002)	0.434(0.002)
0.5	l1-loss(sd)	0.040(0.001)	0.041(0.001)	0.5(0)	0.5(0)
	Error1(sd)	26.8(0.7)	25.9(0.6)	500(0)	500(0)
	Error2(sd)	64.8(0.9)	65.9(0.9)	0(0)	0(0)
	Signal-l1-loss(sd)	0.527(0.002)	0.528(0.002)	0.739(0.002)	0.698(0.002)

Table 2.21 Simulation results under the model $\mathbf{P7}+\mathbf{N3}$:Nonzero $\Theta_i \sim U(1, 1+2\pi)$ and $Z_i \sim SN(0,1,1)/\sqrt{1-1/\pi}$.

CHAPTER 3

Covariance Selection

3.1 Introduction

Recently, there is a surge of interest on the estimation of large dimensional sparse covariance matrices and concentration matrices. Bickel and Levina (2008a, b) proposed estimators by tapering or thresholding sample covariance matrices and showed that they are consistent over a class of sparse matrices. Rothman, Levina and Zhu (2009) considered thresholding sample covariance matrices with more general thresholding functions possessing a shrinkage property. El Karoui (2008) studied the thresholding estimators under a special notion of sparsity called

$\beta - sparsity$ and showed that $\beta - sparse$ matrices, with $\beta < 1/2$ are consistently estimable in spectral norm.

Our objective in this chapter is to estimate the sparsity of the population covariance matrix from a sample correlation matrix. Different from the usual assumption in the literature, we do not assume the population covariance matrix to be very sparse. Our assumption on the population covariance or correlation matrix is that it is believed to have a number of zeros.

One possible application of a good sparsity estimator is in finding a data-dependent threshold for the sample correlation matrix. Bickel and Levina (2008a,b) used cross validation to choose a data-dependent threshold for the sample covariance matrix. However, it is computationally very intensive and tends to over-threshold according to our simulation. Furthermore, when the noise is not homoscedastic, it is more reasonable to find a universal threshold to the sample correlation matrix other than to find a universal threshold to the sample covariance matrix. El Karoui (2008) has established theoretical results for thresholded sample correlation matrices under $\beta - sparsity$, but the methods used for choosing a data-driven threshold is still by resampling. Cai and Liu (2011) proposed an adaptive thresholding method in thresholding sample covariance matrices but they did not deal with sample correlation matrices. However, if the proportion of zeros, say ω, in the population correlation matrix can be well estimated, we can estimate the covariance structure by thresholding the corresponding proportion of smallest (in absolute value) sample correlation coefficients to be zero. This to some degree provides an efficient way of choosing the data-dependent thresholding parameter.

In Section 3.2, we introduce a series of Bernstein-type inequalities and establish a theoretical verification (Theorem 3.1) of our idea of deriving estimators to the covariance structure based on thresholding the sample correlation matrix. In Section 3.3, we propose an empirical Bayes estimator for ω under Gaussian noise. In Section 3.4, we construct a method-of-moments estimator based on trigonometric moment matrices, and derive an upper bound for the expected L1 loss of the estimator. Simulation studies are carried out in Section 3.5 with comparison to estimators derived base on cross-validation methods.

3.2 Sample correlation matrix

In this section, we prove a series of Bernstein-type inequalities. Lemma 1 is an immediate application of the original Bernstein inequality [Bennett (1962)]. Lemma 1 and Lemma 2 are used to show Lemma 3, which establishes an exponential bound for the tail probability of the sample correlation coefficients.

Suppose that $X_1, ..., X_n$ are n i.i.d random observations of Y, which is a p

dimensional random vector with mean $\mathbf{0}$ and covariance matrix $\Sigma_{p\times p} = (\sigma_{ij})_{p\times p}$. We first of all introduce some notations:

$$Y = (Y_1, Y_2, ..., Y_p)';$$

$$X_i = (X_{1i}, X_{2i}, ..., X_{pi})', \quad i = 1, ...n;$$

$$S_j^2 = \sum_{i=1}^{n} X_{ji}^2, \quad j = 1, ...p;$$

$$t_{jk} = n\sigma_{jj}^{1/2}\sigma_{kk}^{1/2}/S_jS_k, \quad 1 \le j, k \le p;$$

$$\rho_{jk} = \sigma_{jk}(\sigma_{jj}\sigma_{kk})^{-1/2}, \quad 1 \le j, k \le p;$$

$$Z_{ji} = X_{ji}/\sigma_{jj}^{1/2}, \quad 1 \le j \le p, 1 \le i \le n;$$

The sample covariance matrix is given by

$$S = (s_{ij})_{p\times p}, \text{ where } s_{ij} = \frac{1}{n}\sum_{k=1}^{n} X_{ik}X_{jk}, \quad 1 \le i, j \le p;$$

The sample correlation coefficients between Y_i and $Y_j, 1 \le i, j \le p$ are given by

$$r_{ij} = \frac{\sum_{k=1}^{n} X_{ik}X_{jk}}{S_iS_j} = \frac{s_{ij}}{\sqrt{s_{ii}s_{jj}}}, \quad 1 \le i, j \le p,$$

and the sample correlation matrix is denoted as $R = (r_{ij})_{p\times p}$. Write the population correlation matrix as $\Gamma = (\rho_{ij})_{p\times p}$. We point out that these notations will be used frequently throughout this chapter.

Assume that, for any $1 \le i \le p$, exists a constant C_0 such that:

$$0 < E|\frac{Y_i^2}{\sigma_{ii}} - 1|^2 \le \sigma^2 < \infty, \quad i = 1, ..., p; \tag{3.1}$$

$$E|\frac{Y_i^2}{\sigma_{ii}} - 1|^r \leq \frac{1}{2}\sigma^2 C_0^{r-2} r!, \qquad r \geq 3. \tag{3.2}$$

It can be easily seen that all results in this section are also true when $Y_i, i = 1, ..., p$ are constants with probability one.

By Bernstein's inequality [Bennett (1962) inequality (7)] we immediately have, for any $x > 0$:

$$P(|\sum_{j=1}^{n} X_{ij}^2/\sigma_{ii} - n| \geq nx) \leq 2\exp\{-\frac{\frac{x^2 n}{E|\frac{Y_i^2}{\sigma_{ii}}-1|^2}}{2 + \frac{2C_0 x}{E|\frac{Y_i^2}{\sigma_{ii}}-1|^2}}\}.$$

Notice that the right hand side of the above inequality is an increasing function of $E|\frac{Y_i^2}{\sigma_{ii}} - 1|^2$, we conclude that:

Lemma 3.1. *For any $1 \leq i \leq p$ and $0 < x \leq K$, there exists a constant $d > 0$, depending on K, C_0 and σ^2 only, such that*

$$P(|\sum_{j=1}^{n} X_{ij}^2/\sigma_{ii} - n| \geq nx) \leq 2\exp\{-dnx^2\}. \tag{3.3}$$

Particularly, when $Y \sim \mathscr{N}(\mathbf{0}, \Sigma_{p\times p})$, we have, for $r \geq 2$:

$$\begin{aligned} E|\frac{Y_i^2}{\sigma_{ii}} - 1|^r &\leq E\max(\frac{Y_i^2}{\sigma_{ii}}, 1)^r \\ &\leq E\frac{Y_i^{2r}}{\sigma_{ii}^r} + 1 \\ &= (2r-1)!! + 1 \\ &\leq 2^r r!. \end{aligned}$$

Therefore, Lemma 1 applies when $X_1, ..., X_n$ $i.i.d$ $\sim \mathscr{N}(\mathbf{0}, \Sigma_{p\times p})$.

Lemma 3.2. *For any $1 \le j,k \le p$ and $0 < x \le K$, there exists a constant $f > 0$, depending on K, C_0 and σ^2 only, such that*

$$P(|\sum_{i=1}^{n}[(Z_{ji}+Z_{ki})^2 - 2(1+\rho_{jk})]| \ge nx) \le 2\exp\{-fnx^2\}; \tag{3.4}$$

$$P(|\sum_{i=1}^{n}[(Z_{ji}-Z_{ki})^2 - 2(1-\rho_{jk})]| \ge nx) \le 2\exp\{-fnx^2\}. \tag{3.5}$$

Proof. To prove (3.4), by Bernstein's inequality, we only need to show that there exist constants $c > 0, w > 0$ depending on K, C_0 and σ^2 only, such that

$$E|(Z_{ji}+Z_{ki})^2 - 2(1+\rho_{jk})|^r \le cw^{r-2}r!, \qquad r \ge 2.$$

Notice that:

$$\begin{aligned}
& E|(Z_{ji}+Z_{ki})^2 - 2(1+\rho_{jk})|^r \\
\le\ & E[|Z_{ji}+Z_{ki}|^2 + 2(1+\rho_{jk})]^r \\
\le\ & E[2|Z_{ji}^2 - 1| + 2|Z_{ki}^2 - 1| + 2(3+\rho_{jk})]^r \\
\le\ & 3^{r-1}E[2^r|Z_{ji}^2 - 1|^r + 2^r|Z_{ki}^2 - 1|^r + 2^r|3+\rho_{jk}|^r],
\end{aligned}$$

where in the last step we use the following inequality:

$$(\frac{a+b+c}{3})^r \le \frac{a^r+b^r+c^r}{3}, \quad for \quad any \quad a,b,c > 0, r \ge 2.$$

By the the definition of $Z_{ji}, 1 \le i,j \le p$ and assumption (3.2) , we have:

$$E|(Z_{ji}+Z_{ki})^2 - 2(1+\rho_{jk})|^r$$

$$
\begin{aligned}
&\leq 3^{r-1}E[2^r|Z_{ji}^2-1|^r+2^r|Z_{ki}^2-1|^r+2^r|3+\rho_{jk}|^r] \\
&\leq 6^r[\sigma^2 C_0^{r-2}r!+4^r] \\
&\leq 36\sigma^2(6C_0+\frac{4}{1\wedge\sigma})^{r-2}r! \\
&= cw^{r-2}r!,
\end{aligned}
$$

where $c = 36\sigma^2$ and $w = 6C_0 + \frac{4}{1\wedge\sigma}$. (3.5) can be proved similarly. □

Next we prove a Bernstein-type inequality for elements of the sample correlation matrix.

Lemma 3.3. *For any $0 < v \leq 2$ and $1 \leq j, k \leq p$, there exist constants $d_1 > 0$ and $d_2 > 0$, depending on C_0 and σ^2 only, such that*

$$
P(|\frac{\sum_{i=1}^n X_{ji}X_{ki}}{S_jS_k}-\rho_{jk}|\geq v)\leq d_1e^{-d_2nv^2}.
$$

Proof. When $\rho_{jk} = \pm 1$, LHS of the inequality equals to zero, and so the inequality holds. Now we consider the case: $-1 < \rho_{jk} < 1$.

$$
\begin{aligned}
P(|\frac{\sum_{i=1}^n X_{ji}X_{ki}}{S_jS_k}-\rho_{jk}|\geq v) &= P(|\frac{\sum_{i=1}^n X_{ji}X_{ki}}{n\sigma_{jj}^{1/2}\sigma_{kk}^{1/2}}\cdot\frac{n\sigma_{jj}^{1/2}\sigma_{kk}^{1/2}}{S_jS_k}-\rho_{jk}|\geq v) \\
&= P(|\frac{1}{n}\sum_{i=1}^n Z_{ji}Z_{ki}\cdot t_{jk}-\rho_{jk}|\geq v) \\
&\leq P(|\frac{1}{n}\sum_{i=1}^n Z_{ji}Z_{ki}\cdot(t_{jk}-1)|\geq\frac{v}{2}) \\
&\quad +P(|\frac{1}{n}\sum_{i=1}^n(Z_{ji}Z_{ki}-\rho_{jk})|\geq\frac{v}{2}).
\end{aligned}
$$

Now,

$$\sum_{i=1}^{n}(Z_{ji}Z_{ki}-\rho_{jk}) = \frac{1}{4}\{\sum_{i=1}^{n}[(Z_{ji}+Z_{ki})^2-2(1+\rho_{jk})] - \sum_{i=1}^{n}[(Z_{ji}-Z_{ki})^2-2(1-\rho_{jk})]\}.$$

By Lemma 3.2, there exists a constant $f_1 > 0$, depending on C_0 and σ^2 only, such that,

$$\begin{aligned}
& P(|\frac{1}{n}\sum_{i=1}^{n}(Z_{ji}Z_{ki}-\rho_{jk})| \geq \frac{v}{2}) \\
= \; & P(|\sum_{i=1}^{n}[(Z_{ji}+Z_{ki})^2-2(1+\rho_{jk})] - \sum_{i=1}^{n}[(Z_{ji}-Z_{ki})^2-2(1-\rho_{jk})]| \geq 2nv) \\
\leq \; & P(|\sum_{i=1}^{n}[(Z_{ji}+Z_{ki})^2-2(1+\rho_{jk})]| \geq nv) \\
& + P(|\sum_{i=1}^{n}[(Z_{ji}-Z_{ki})^2-2(1-\rho_{jk})]| \geq nv) \\
\leq \; & 4e^{-f_1 nv^2}.
\end{aligned}$$

Let $a = \frac{v}{2(|\rho_{jk}|+v)}$, we have:

$$P(|\frac{1}{n}\sum_{i=1}^{n} Z_{ji}Z_{ki}\cdot(t_{jk}-1)| \geq \frac{v}{2}) \leq P(|\sum_{i=1}^{n} Z_{ji}Z_{ki}| \geq \frac{nv}{2a}) + P(|t_{jk}-1| > a).$$

Now,

$$\begin{aligned}
P(|\sum_{i=1}^{n} Z_{ji}Z_{ki}| \geq \frac{nv}{2a}) & \leq P(|\sum_{i=1}^{n} Z_{ji}Z_{ki} - n\rho_{jk}| \geq n(\frac{v}{2a}-|\rho_{jk}|)) \\
& = P(|\sum_{i=1}^{n} Z_{ji}Z_{ki} - n\rho_{jk}| \geq nv).
\end{aligned}$$

As we have shown just now, by replacing $v/2$ to be v, the above inequality can be bounded by $4e^{-f_2 nv^2}$, for some constant $f_2 > 0$, depending on C_0 and σ^2 only.

$$\begin{aligned}
P(|t_{jk}-1|>a) &= P(n\sigma_{jj}^{1/2}\sigma_{kk}^{1/2}/S_jS_k > a+1) + P(n\sigma_{jj}^{1/2}\sigma_{kk}^{1/2}/S_jS_k < 1-a) \\
&= P(\sum_{i=1}^{n} Z_{ji}^2 \sum_{i=1}^{n} Z_{ki}^2 < \frac{n^2}{(1+a)^2}) + P(\sum_{i=1}^{n} Z_{ji}^2 \sum_{i=1}^{n} Z_{ki}^2 > \frac{n^2}{(1-a)^2}) \\
&\leq P(\sum_{i=1}^{n} Z_{ji}^2 < \frac{n}{1+a}) + P(\sum_{i=1}^{n} Z_{ki}^2 < \frac{n}{1+a}) \\
&\quad + P(\sum_{i=1}^{n} Z_{ji}^2 > \frac{n}{1-a}) + P(\sum_{i=1}^{n} Z_{ki}^2 > \frac{n}{1-a}) \\
&= P(\sum_{i=1}^{n}(Z_{ji}^2-1) < -\frac{an}{1+a}) + P(\sum_{i=1}^{n}(Z_{ki}^2-1) < -\frac{an}{1+a}) \\
&\quad + P(\sum_{i=1}^{n}(Z_{ji}^2-1) > \frac{an}{1-a}) + P(\sum_{i=1}^{n}(Z_{ki}^2-1) > \frac{an}{1-a}) \\
&\leq P(|\sum_{i=1}^{n}(Z_{ji}^2-1)| > \frac{an}{1+a}) + P(|\sum_{i=1}^{n}(Z_{ki}^2-1)| > \frac{an}{1+a}).
\end{aligned}$$

By Lemma 3.1, there exists a constant $d > 0$ independent of n and v, such that,

$$\begin{aligned}
P(|t_{jk}-1|>a) &\leq 4e^{-dn(\frac{a}{1+a})^2} \\
&= 4e^{-dn(\frac{v}{3v+2|\rho_{jk}|})^2} \\
&\leq 4e^{-f_3 nv^2},
\end{aligned}$$

where $f_3 = d/(3v+2|\rho_{ik}|)^2 \geq d/64$.

Above all, we have shown that:

$$\begin{aligned}
&P(|\frac{\sum_{i=1}^{n} X_{ji}X_{ki}}{S_jS_k} - \rho_{jk}| \geq v) \\
\leq\ & P(|\frac{1}{n}\sum_{i=1}^{n} Z_{ji}Z_{ki}\cdot(t_{jk}-1)| \geq \frac{v}{2}) + P(|\sum_{i=1}^{n} Z_{ji}Z_{ki}| \geq \frac{nv}{2a}) + P(|t_{jk}-1|>a)
\end{aligned}$$

$$\leq \; 4e^{-f_1 nv^2} + 4e^{-f_2 nv^2} + 4e^{-f_3 nv^2}.$$

The theorem is proved by letting $d_1 = 12$ and $d_2 = min(f_1, f_2, f_3)$. □

The next theorem establishes an upper bound for the probability of covariance selection consistency. First of all, we introduce some notations and assumptions:

1)As defined at the beginning of this section, we write the population correlation matrix as $\Gamma = (\rho_{ij})_{p\times p}$. Now let the set $G = \{(i,j) : \rho_{ij} \neq 0, i < j\}$, with $card(G) = g$, which is the cardinality of G. Assume that:

$|\rho_{ij}| \geq k(n,p)$, if $\rho_{ij} \neq 0$.

2)Defined for the sample correlation matrix $R = (r_{ij})_{p\times p}$, a thresholding operator:

$$T_t(R) = [r_{ij} I_{\{|r_{ij}| \geq t\}}]_{1\leq i,j\leq p},$$

where $t(n,p) < k(n,p)$ is a thresholding parameter.

3)Define $\hat{G} = \{(i,j) : |r_{ij}| > t, i < j\}$.

Under Assumptions 1) and 2), we have:

Theorem 3.1. *There exists $c > 0$, such that,*

$$P(\hat{G} = G) = 1 - O(p^2 exp(-c(t^2 \vee (k-t)^2)n)).$$

Proof.

$$P(\hat{G} \neq G) \leq gP(|r_{ij}| \leq t, (i,j) \in G) + (\frac{p(p-1)}{2} - g)P(|r_{ij}| \geq t, (i,j) \in G^c).$$

Now, by Lemma 3, there exist $d_1 > 0, d_2 > 0$, such that,

$$\begin{aligned} P(|r_{ij}| \leq t, (i,j) \in G) &= P(|\frac{\sum_{k=1}^n X_{ik}X_{jk}}{S_i S_j}| < t,\ \rho_{ij} \geq k) \\ &\leq P(|\frac{\sum_{k=1}^n X_{ik}X_{jk}}{S_i S_j} - \rho_{ij}| \geq k - t) \\ &\leq d_1 \exp\{-d_2(k-t)^2 n\}. \end{aligned}$$

Similarly,

$$P(|r_{ij}| \geq t, (i,j) \in G^c) \leq d_1 \exp\{-d_2 t^2 n\}.$$

Hence:

$$P(\hat{G} \neq G) \leq gd_1 \exp\{-d_2(k-t)^2 n\} + (\frac{p(p-1)}{2} - g)d_1 \exp\{-d_2 t^2 n\}.$$

Therefore

$$P(\hat{G} \neq G) \leq d_1 \frac{p(p-1)}{2} \exp\{-d_2(t^2 \vee (k-t)^2)n\}.$$

□

By setting $k(n,p)$ and $t(n,p)$ properly we have the following corollary:

Corollary 3.1. *Assume that* $k \geq \sqrt{\frac{2\log p}{n} + n^{-\alpha}}$*, for some* $0 < \alpha < 1$*. With* $t = k/2$*, we have:*

$$P(\hat{G} = G) = 1 - O(\exp\{-n^{1-\alpha}\}).$$

3.3 Empirical Bayes estimator under multivariate normal assumption

In this section, we model the prior on ρ_{ij} as a mixture distribution which has a point mass ω at zero. We propose an empirical Bayes estimator $\hat{\omega}_1$ for ω and show that it is consistent. Let the sample correlation coefficients be defined as in Section 3.2.

3.3.1 Assumptions on the prior

Assuming the prior distribution on the correlation coefficients $\rho_{ij}, 1 \leq j < i \leq p$ satisfies:

A1 The marginal distribution of ρ_{ij} is a mixture of a point mass at zero and a distribution function G in [-1,1]:

$$(1-\omega)dG(\rho) + \omega\delta_0(\rho),$$

where $\delta_0(\rho)$ denotes point mass at $\rho = 0$.

A2 Let $\mathscr{F}_{ij} = \sigma(\rho_{ij})$ denote the σ-field generated by ρ_{ij}. Define for all $1 \leq$

$i, j, s, t \leq p$,

$$\alpha(\rho_{ij}, \rho_{st}) = \sup_{A \in \mathscr{F}_{ij}, B \in \mathscr{F}_{st}} |P(A \cap B) - P(A)P(B)|.$$

Assuming that, there exists a constant $0 \leq \nu < 4$, such that

$$\sum_{i,j,s,t:all\ \ distinct} \alpha(\rho_{ij}, \rho_{st}) = O(p^{\nu}).$$

This condition implies that

$$\sum_{i,j,s,t:all\ \ distinct} ReE[(e^{-\mathbf{i}k\rho_{ij}} - Ee^{-\mathbf{i}k\rho_{ij}})(e^{\mathbf{i}k\rho_{st}} - Ee^{\mathbf{i}k\rho_{st}})] = O(p^{\nu}),$$

which is what we really need in Section 3.4.

A3 Let g be the corresponding density function of G. Define

$$g_1(\rho) = \begin{cases} g(\rho), & if\ \ -1 < \rho < 0, \\ 0, & otherwise, \end{cases}$$

and

$$g_2(\rho) = \begin{cases} g(\rho), & if\ 0 < \rho < 1, \\ 0, & otherwise, \end{cases}$$

There exist constants $a \geq 0$ and $0 \leq b < \frac{1}{2}$, such that,

$$If\ \ \int_0^1 g_1(-\rho)d\rho > 0,\ for\ all\ i, n \in \mathbb{N},\ when\ n\ large\ enough,$$
$$\int_0^1 (1-\rho^2)^{\frac{n}{2}} \rho^i g_1(-\rho)d\rho \leq n^b B(\frac{n}{2}+1, \frac{i}{2}+\frac{1}{2});$$
$$for\ all\ i, n \in \mathbb{N}, i \leq n,\ when\ n\ large\ enough,$$

$$\int_0^1 (1-\rho^2)^{\frac{n}{2}}\rho^{2i} g_1(-\rho)d\rho \geq n^{-a}B(\frac{n}{2}+1, i+\frac{1}{2});$$

$$If \quad \int_0^1 g_2(\rho)d\rho > 0, \ for \ all \ i, n \in \mathbb{N}, \ when \ n \ large \ enough,$$

$$\int_0^1 (1-\rho^2)^{\frac{n}{2}}\rho^{i} g_2(\rho)d\rho \leq n^{b}B(\frac{n}{2}+1, \frac{i}{2}+\frac{1}{2});$$

$$for \ all \ i, n \in \mathbb{N}, i \leq n, \ when \ n \ large \ enough,$$

$$\int_0^1 (1-\rho^2)^{\frac{n}{2}}\rho^{2i} g_2(\rho)d\rho \geq n^{-a}B(\frac{n}{2}+1, i+\frac{1}{2});$$

where $B(\cdot,\cdot)$ is the beta function.

This assumption prevents g in degenerating to point mass distributions at 0,-1 and 1. Also, it ensures all the interchange of summation and integration operations throughout this section.

A4 There exists a constant $v \in (0, \frac{1}{2})$, such that ,

$$\int_{-\frac{1}{n^{\frac{1}{2}-v}}}^{\frac{1}{n^{\frac{1}{2}-v}}} dG(\rho) \longrightarrow 0, \quad as \quad n \quad increases.$$

Remark 3.1. Any continuous density function g with support [-1,1] satisfies Assumptions A3 and A4. Also, it is easy to see that for any $a_1, a_2 < \infty$, if $\rho^2 \sim Beta(a_1, a_2)$, A3 and A4 are satisfied. When $0 < \omega < 1$, for any g, such that $\sup_{\rho\in(-1,1)} g(\rho) < \infty$, we can perturb ω a little bit to be $\omega' = \omega - n^{-a}$ for some constant $a > 0$, and replacing g by $g'(\rho) = \frac{1-\omega}{1-\omega'}g(\rho) + \frac{1}{2(1-\omega')}n^{-a}$, then g' will satisfy Assumption A3. This perturbation will only introduce an error which is negligible and we will discuss this in Section 3.5.

3.3.2 Motivation for $\hat{\omega}_1$

The following equation will be used for a few times in this section:

$$\Gamma(z)\Gamma(z+\frac{1}{2}) = 2^{1-2z}\sqrt{\pi}\Gamma(2z).$$

Under normal noise, the density of r_{ij} given ρ_{ij} is [see Anderson (2003), Theorem 4.2.2]:

$$f_{r_{ij}|\rho_{ij}}(r|\rho) = \frac{2^{n-2}(1-\rho^2)^{\frac{n}{2}}(1-r^2)^{\frac{n-3}{2}}}{(n-2)!\pi}\sum_{i=0}^{\infty}\frac{(2\rho r)^i}{i!}\Gamma^2(\frac{n+i}{2}) \quad \forall r \in (-1,1).$$

Denote $\omega' = 1-\omega$. The marginal density of r_{ij} can be simplified:

$$\begin{aligned}
f_{r_{ij}}(r;\omega) &= (1-\omega')f_{r_{ij}|\rho_{ij}}(r|\rho=0) + \omega'\int_{-1}^{1} f_{r_{ij}|\rho_{ij}}(r|\rho)dG(\rho) \\
&= (1-\omega')\frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})\sqrt{\pi}}(1-r^2)^{\frac{n-3}{2}} \\
&\quad +\omega'\int_{-1}^{1}\frac{2^{n-2}(1-\rho^2)^{\frac{n}{2}}(1-r^2)^{\frac{n-3}{2}}}{(n-2)!\pi}\sum_{i=0}^{\infty}\frac{(2\rho r)^{2i}}{(2i)!}\Gamma^2(\frac{n+2i}{2})dG(\rho) \\
&\quad +\omega'\int_{-1}^{1}\frac{2^{n-2}(1-\rho^2)^{\frac{n}{2}}(1-r^2)^{\frac{n-3}{2}}}{(n-2)!\pi}\sum_{i=0}^{\infty}\frac{(2\rho r)^{2i+1}}{(2i+1)!}\Gamma^2(\frac{n+2i+1}{2})dG(\rho) \\
&= \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})\sqrt{\pi}}(1-r^2)^{\frac{n-3}{2}}(1+\omega' a_{ij}),
\end{aligned}$$

where

$$\begin{aligned}
a_{ij} &= -1 + \int_{-1}^{1}\frac{2^{n-2}\Gamma(\frac{n-1}{2})\sqrt{\pi}(1-\rho^2)^{\frac{n}{2}}}{\Gamma(\frac{n}{2})(n-2)!\pi}\sum_{i=0}^{\infty}\frac{(2\rho r)^i}{(i)!}\Gamma^2(\frac{n+i}{2})dG(\rho) \\
&= -1 + \int_{-1}^{1}\frac{2^{n-2}\Gamma(\frac{n-1}{2})\sqrt{\pi}(1-\rho^2)^{\frac{n}{2}}}{\Gamma(\frac{n}{2})(n-2)!\pi}\sum_{i=0}^{\infty}\frac{(2\rho r)^{2i}}{(2i)!}\Gamma^2(\frac{n+2i}{2})dG(\rho)
\end{aligned}$$

$$
\begin{aligned}
&+\int_{-1}^{1}\frac{2^{n-2}\Gamma(\frac{n-1}{2})\sqrt{\pi}(1-\rho^2)^{\frac{n}{2}}}{\Gamma(\frac{n}{2})(n-2)!\pi}\sum_{i=0}^{\infty}\frac{(2\rho r)^{2i+1}}{(2i+1)!}\Gamma^2(\frac{n+2i+1}{2})dG(\rho)\\
=\ &-1+\int_{-1}^{1}\frac{2^{n-2}\Gamma(\frac{n-1}{2})\Gamma(\frac{n}{2})\sqrt{\pi}(1-\rho^2)^{\frac{n}{2}}}{\Gamma^2(\frac{n}{2})(n-2)!\pi}\sum_{i=0}^{\infty}\frac{(2\rho r)^{2i}}{(2i)!}\Gamma^2(\frac{n+2i}{2})dG(\rho)\\
&+\int_{-1}^{1}\frac{2^{n-2}\Gamma(\frac{n-1}{2})\Gamma(\frac{n}{2})\sqrt{\pi}(1-\rho^2)^{\frac{n}{2}}}{\Gamma^2(\frac{n}{2})(n-2)!\pi}\sum_{i=0}^{\infty}\frac{(2\rho r)^{2i+1}}{(2i+1)!}\Gamma^2(\frac{n+2i+1}{2})dG(\rho)\\
=\ &-1+\int_{-1}^{1}\frac{\Gamma(n-1)(1-\rho^2)^{\frac{n}{2}}}{\Gamma^2(\frac{n}{2})(n-2)!}\sum_{i=0}^{\infty}\frac{(2\rho r)^{2i}}{(2i)!}\Gamma^2(\frac{n+2i}{2})dG(\rho)\\
&+\int_{-1}^{1}\frac{\Gamma(n-1)(1-\rho^2)^{\frac{n}{2}}}{\Gamma^2(\frac{n}{2})(n-2)!}\sum_{i=0}^{\infty}\frac{(2\rho r)^{2i+1}}{(2i+1)!}\Gamma^2(\frac{n+2i+1}{2})dG(\rho)\\
=\ &-1+\sum_{i=0}^{\infty}\frac{\Gamma^2(\frac{n+2i}{2})2^{2i}r^{2i}}{\Gamma^2(\frac{n}{2})(2i)!}\int_{-1}^{1}(1-\rho^2)^{\frac{n}{2}}\rho^{2i}dG(\rho)\\
&+\sum_{i=0}^{\infty}\frac{\Gamma^2(\frac{n+2i+1}{2})2^{2i+1}r^{2i+1}}{\Gamma^2(\frac{n}{2})(2i+1)!}\int_{-1}^{1}(1-\rho^2)^{\frac{n}{2}}\rho^{2i+1}dG(\rho).
\end{aligned}
$$

Notice that when $a_{ij} > (<)0$, $f_{r_{ij}}(r;\omega)$ is maximized when $\omega = 0(1)$. In other words, when $a_{ij} > (<)0$, it tends to estimate ρ_{ij} as nonzero(zero). Therefore we propose the following estimator for ω:

$$\hat{\omega}_1 = 1 - \frac{\sum_{1\le i<j\le p} I_{\{a_{ij}>0\}}}{p(p-1)/2}. \tag{3.6}$$

Particularly, when g is an even function in (-1,1), we have:

$$a_{ij} = -1 + 2\sum_{i=0}^{\infty}\frac{\Gamma^2(\frac{n+2i}{2})2^{2i}r^{2i}}{\Gamma^2(\frac{n}{2})(2i)!}\int_0^1(1-\rho^2)^{\frac{n}{2}}\rho^{2i}g(\rho)d\rho,$$

which is an increasing function of r^2, therefore the corresponding threshold is the root of the following equation:

$$a_{ij}(r) = 0.$$

3.3.3 Properties of $\hat{\omega}_1$

Lemma 3.4 and Lemma 3.5 to some degree provide a lower bound and an upper bound for the thresholding parameter corresponding to our estimator $\hat{\omega}_1$. Theorem 3.2 shows that $\hat{\omega}_1$ is consistent in estimating ω.

We first of all introduce an inequality that will be used quite often in this subsection:

Lemma 3.4. *for any $n, k \in Z^+$, we have:*

$$\frac{1}{\sqrt{\frac{n}{2}+k}} < \frac{\Gamma(\frac{n}{2}+k)}{\Gamma(\frac{n}{2}+k+\frac{1}{2})} < \frac{1}{\sqrt{\frac{n}{2}+k-\frac{1}{4}}}. \tag{3.7}$$

Proof. This is a direct result of (4.2) of Bustoz and Ismail (1986). □

The following lemma provides an asymptotic lower bound for the thresholding parameter.

Lemma 3.5. *Under Assumption A3, for $\forall n$ large enough, and $0 < c < \frac{1}{2} - b$,*

$$a_{ij} > 0 \Rightarrow r_{ij}^2 > \frac{(\frac{1}{2}-b-c)\log n}{n}.$$

Proof. It is enough to show that for $\forall n$ large enough and $0 < c < \frac{1}{2} - b$, $r^2 \leq \frac{(\frac{1}{2}-b-c)\log n}{n}$ implies $a_{ij}(r) < 0$.

Write:

$$
\begin{aligned}
& \sum_{i=0}^{\infty} \frac{\Gamma^2(\frac{n+2i}{2}) 2^{2i} r^{2i}}{\Gamma^2(\frac{n}{2})(2i)!} \int_0^1 (1-\rho^2)^{\frac{n}{2}} \rho^{2i} (g_1(-\rho) + g_2(\rho)) d\rho \\
= \; & \int_0^1 (1-\rho^2)^{\frac{n}{2}} (g_1(-\rho) + g_2(\rho)) d\rho \\
& + \sum_{i=1}^{[\frac{n}{2}]} \frac{\Gamma^2(\frac{n+2i}{2}) 2^{2i} r^{2i}}{\Gamma^2(\frac{n}{2})(2i)!} \int_0^1 (1-\rho^2)^{\frac{n}{2}} \rho^{2i} (g_1(-\rho) + g_2(\rho)) d\rho \\
& + \sum_{i=[\frac{n}{2}]+1}^{\infty} \frac{\Gamma^2(\frac{n+2i}{2}) 2^{2i} r^{2i}}{\Gamma^2(\frac{n}{2})(2i)!} \int_0^1 (1-\rho^2)^{\frac{n}{2}} \rho^{2i} (g_1(-\rho) + g_2(\rho)) d\rho \\
=: \; & I + II + III;
\end{aligned}
$$

$$
\begin{aligned}
& \sum_{i=0}^{\infty} \frac{\Gamma^2(\frac{n+2i+1}{2}) 2^{2i+1} r^{2i+1}}{\Gamma^2(\frac{n}{2})(2i+1)!} \int_{-1}^1 (1-\rho^2)^{\frac{n}{2}} \rho^{2i+1} (g_1(-\rho) + g_2(\rho)) d\rho \\
= \; & \sum_{i=0}^{[\frac{n}{2}]} \frac{\Gamma^2(\frac{n+2i+1}{2}) 2^{2i+1} r^{2i+1}}{\Gamma^2(\frac{n}{2})(2i+1)!} \int_0^1 (1-\rho^2)^{\frac{n}{2}} \rho^{2i+1} (g_1(-\rho) + g_2(\rho)) d\rho \\
& + \sum_{i=[\frac{n}{2}]+1}^{\infty} \frac{\Gamma^2(\frac{n+2i+1}{2}) 2^{2i+1} r^{2i+1}}{\Gamma^2(\frac{n}{2})(2i+1)!} \int_0^1 (1-\rho^2)^{\frac{n}{2}} \rho^{2i+1} (g_1(-\rho) + g_2(\rho)) d\rho \\
=: \; & II' + III'.
\end{aligned}
$$

From Assumption A3 and Lemma 3.4, we have:

$$
I \le 2n^b \frac{\Gamma(\frac{n}{2}+1)\Gamma(\frac{1}{2})}{\Gamma(\frac{n}{2}+\frac{3}{2})} \le \frac{2n^b\sqrt{\pi}}{\sqrt{\frac{n}{2}+\frac{3}{4}}} \quad .
$$

For III, we have:

$$
III
$$

$$
\begin{aligned}
&\leq 2n^b \sum_{i>[\frac{n}{2}]} \frac{\Gamma^2(\frac{n+2i}{2})2^{2i}r^{2i}}{\Gamma^2(\frac{n}{2})(2i)!} \cdot \frac{\Gamma(\frac{n}{2}+1)\Gamma(i+\frac{1}{2})}{\Gamma(\frac{n}{2}+i+\frac{3}{2})} \\
&= 2n^b \sum_{i>[\frac{n}{2}]} (2r)^{2i} \cdot \frac{\Gamma(\frac{n}{2}+1)}{\Gamma(\frac{n}{2})} \cdot \frac{\Gamma(\frac{n}{2}+i)}{(\frac{n}{2}+i+\frac{1}{2})\Gamma(\frac{n}{2}+i+\frac{1}{2})} \cdot \frac{\Gamma(\frac{n}{2}+i)\Gamma(i+\frac{1}{2})}{\Gamma(\frac{n}{2})(2i)!} \\
&= 2n^b \sum_{i>[\frac{n}{2}]} (2r)^{2i} \cdot \frac{\frac{n}{2}}{\frac{n}{2}+i+\frac{1}{2}} \cdot \frac{\Gamma(\frac{n}{2}+i)}{\Gamma(\frac{n}{2}+i+\frac{1}{2})} \cdot \frac{\Gamma(i+\frac{1}{2})}{\Gamma(i+1)} \cdot \frac{(\frac{n}{2}+i-1)\cdots\frac{n}{2}}{2i\cdots(i+1)} \\
&\leq 2n^b \sum_{i>[\frac{n}{2}]} (2r)^{2i} \cdot \frac{1}{2} \cdot \frac{1}{\sqrt{\frac{n}{2}+i-\frac{1}{4}}} \cdot \frac{1}{\sqrt{i+\frac{1}{4}}} \\
&\leq n^b \sum_{i>[\frac{n}{2}]} \frac{\sqrt{2}}{n}(2r)^{2i} \quad .
\end{aligned}
$$

Similarly, for III', we have:

$$
\begin{aligned}
&|III'| \\
&\leq 2n^b \sum_{i>[\frac{n}{2}]} \frac{\Gamma^2(\frac{n+2i+1}{2})2^{2i+1}|r|^{2i+1}}{\Gamma^2(\frac{n}{2})(2i+1)!} \cdot \frac{\Gamma(\frac{n}{2}+1)\Gamma(i+1)}{\Gamma(\frac{n}{2}+i+2)} \\
&= 2n^b \sum_{i>[\frac{n}{2}]} |2r|^{2i+1} \cdot \frac{\Gamma(\frac{n}{2}+1)}{\Gamma(\frac{n}{2})} \cdot \frac{\Gamma(\frac{n}{2}+i+\frac{1}{2})}{(\frac{n}{2}+i+1)\Gamma(\frac{n}{2}+i+1)} \cdot \frac{\Gamma(\frac{n}{2}+i+\frac{1}{2})\Gamma(i+1)}{\Gamma(\frac{n}{2})(2i+1)!} \\
&= 2n^b \sum_{i>[\frac{n}{2}]} \frac{|2r|^{2i+1}\frac{n}{2}}{\frac{n}{2}+i+1} \cdot \frac{\Gamma(\frac{n}{2}+i+\frac{1}{2})}{\Gamma(\frac{n}{2}+i+1)} \cdot \frac{\Gamma(\frac{n}{2}+i+\frac{1}{2})}{\Gamma(\frac{n}{2}+i+1)} \cdot \frac{\Gamma(\frac{n}{2}+i+1)}{\Gamma(\frac{n}{2})} \cdot \frac{\Gamma(i+1)}{(2i+1)!} \\
&\leq 2n^b \sum_{i>[\frac{n}{2}]} |2r|^{2i+1} \cdot \frac{1}{2} \cdot \frac{1}{\frac{n}{2}+i+\frac{1}{4}} \cdot \frac{(\frac{n}{2}+i)(\frac{n}{2}+i-1)\cdots\frac{n}{2}}{(2i+1)2i\cdots(i+1)} \\
&\leq n^b \sum_{i>[\frac{n}{2}]} \frac{1}{n}|2r|^{2i+1} \quad .
\end{aligned}
$$

For II and II', we have:

$$
\begin{aligned}
II &\leq 2n^b \sum_{i=1}^{[\frac{n}{2}]} \frac{\Gamma^2(\frac{n+2i}{2})2^{2i}r^{2i}}{\Gamma^2(\frac{n}{2})(2i)!} \cdot \frac{\Gamma(\frac{n}{2}+1)\Gamma(i+\frac{1}{2})}{\Gamma(\frac{n}{2}+i+\frac{3}{2})} \\
&= 2n^b \sum_{i=1}^{[\frac{n}{2}]} (2r)^{2i} \cdot \frac{\Gamma(i+\frac{1}{2})}{\Gamma(2i+1)} \cdot \frac{\Gamma(\frac{n}{2}+i)}{\Gamma(\frac{n}{2})} \cdot \frac{\Gamma(\frac{n}{2}+1)}{\Gamma(\frac{n}{2})} \cdot \frac{\Gamma(\frac{n}{2}+i)}{(\frac{n}{2}+i+\frac{1}{2})\Gamma(\frac{n}{2}+i+\frac{1}{2})}
\end{aligned}
$$

$$
\begin{aligned}
&= 2n^b \sum_{i=1}^{[\frac{n}{2}]} (2r)^{2i} \cdot \frac{\sqrt{\pi}2^{-2i}}{\Gamma(i+1)} \cdot (\frac{n}{2}+i-1)\cdots\frac{n}{2} \cdot \frac{\Gamma(\frac{n}{2}+i)}{\Gamma(\frac{n}{2}+i+\frac{1}{2})} \cdot \frac{\frac{n}{2}}{\frac{n}{2}+i+\frac{1}{2}} \\
&\leq 2n^b\sqrt{\pi} \sum_{i=1}^{[\frac{n}{2}]} \frac{(nr^2)^i}{i!} \cdot \frac{1}{\sqrt{\frac{n}{2}+i-\frac{1}{4}}} \\
&\leq 2n^b\sqrt{2\pi}e^{nr^2}\frac{1}{\sqrt{n}} \quad .
\end{aligned}
$$

$$
\begin{aligned}
&|II'| \\
\leq\; & 2n^b \sum_{i=1}^{[\frac{n}{2}]} \frac{\Gamma^2(\frac{n+2i+1}{2})2^{2i+1}|r|^{2i+1}}{\Gamma^2(\frac{n}{2})(2i+1)!} \cdot \frac{\Gamma(\frac{n}{2}+1)\Gamma(i+1)}{\Gamma(\frac{n}{2}+i+2)} \\
=\; & 4|r|n^b \sum_{i=1}^{[\frac{n}{2}]} \frac{(2r)^{2i}\Gamma(i+1)}{\Gamma(2i+2)} \cdot \frac{\Gamma(\frac{n}{2}+i+\frac{1}{2})}{\Gamma(\frac{n}{2})} \cdot \frac{\Gamma(\frac{n}{2}+1)}{\Gamma(\frac{n}{2})} \cdot \frac{\Gamma(\frac{n}{2}+i+\frac{1}{2})}{(\frac{n}{2}+i+1)\Gamma(\frac{n}{2}+i+1)} \\
=\; & 4n^b|r| \sum_{i=1}^{[\frac{n}{2}]} (2r)^{2i} \cdot \frac{\sqrt{\pi}2^{-2i-1}\Gamma(i+1)}{\Gamma(i+\frac{3}{2})\Gamma(i+1)} \cdot \frac{\Gamma(\frac{n}{2}+i)}{\Gamma(\frac{n}{2})} \cdot \frac{\Gamma(\frac{n}{2}+i+\frac{1}{2})}{\Gamma(\frac{n}{2}+i)} \\
& \cdot \frac{n}{n+2i+2} \cdot \frac{\Gamma(\frac{n}{2}+i+\frac{1}{2})}{\Gamma(\frac{n}{2}+i+1)} \\
\leq\; & 2n^b\sqrt{\pi}|r| \sum_{i=1}^{[\frac{n}{2}]} \frac{(r^2)^i}{i!} \cdot \frac{\Gamma(i+1)}{\Gamma(i+\frac{3}{2})} \cdot (\frac{n}{2}+i-1)\cdots\frac{n}{2} \cdot \frac{\Gamma(\frac{n}{2}+i+\frac{1}{2})}{\Gamma(\frac{n}{2}+i)} \cdot \frac{\Gamma(\frac{n}{2}+i+\frac{1}{2})}{\Gamma(\frac{n}{2}+i+1)} \\
\leq\; & 2n^b\sqrt{\pi}|r| \sum_{i=1}^{[\frac{n}{2}]} \frac{(nr^2)^i}{i!} \cdot \frac{1}{\sqrt{i+\frac{3}{4}}} \cdot \sqrt{\frac{n}{2}+i} \cdot \frac{1}{\sqrt{\frac{n}{2}+i+\frac{1}{4}}} \\
\leq\; & 2n^b|r|\sqrt{\pi}e^{nr^2} \quad .
\end{aligned}
$$

Therefore, we have, when $r^2 \leq \frac{(\frac{1}{2}-b-c)\log n}{n}$:

$$
\begin{aligned}
a_{ij} &\leq -1 + I + III + |III'| + II + |II'| \\
&\leq -1 + \frac{2n^b\sqrt{\pi}}{\sqrt{\frac{n}{2}+\frac{3}{4}}} + n^b \sum_{i>[\frac{n}{2}]} \frac{\sqrt{2}}{n}(2r)^{2i} +
\end{aligned}
$$

$$n^b \sum_{i>[\frac{n}{2}]} \frac{1}{n}|2r|^{2i+1} + 2n^b\sqrt{2\pi}e^{nr^2}\frac{1}{\sqrt{n}} + 2n^b|r|\sqrt{\pi}e^{nr^2}$$
$$= \quad -1 + O(n^{-(1/2-b)}) + O(\frac{\sqrt{\log n}}{n^c}),$$

which will tend to -1 when n tends to infinity.

□

Lemma 3.6. *Under Assumption A3, for* $\forall n$ *large enough,*

1) If $\int_0^1 g_2(\rho)d\rho = 0$,

$$a_{ij} \leq 0 \Rightarrow r_{ij} > -\sqrt{\frac{2(a+1)\log n}{n}} \tag{3.8}$$

2) If $\int_0^1 g_2(\rho)d\rho = 1$,

$$a_{ij} \leq 0 \Rightarrow r_{ij} < \sqrt{\frac{2(a+1)\log n}{n}} \tag{3.9}$$

3) If $0 < \int_0^1 g_2(\rho)d\rho < 1$,

$$a_{ij} \leq 0 \Rightarrow r_{ij}^2 < \frac{2(a+1)\log n}{n}. \tag{3.10}$$

Proof. To prove (3.8), we show that when $r \leq -\sqrt{\frac{2(a+1)\log n}{n}}$, $a_{ij}(r) > 0$ for any n large enough. Notice that when $\int_0^1 g_2(\rho)d\rho = 0$, for any $-1 < r \leq -\sqrt{\frac{2(a+1)\log n}{n}}$, we have:

$$
\begin{aligned}
a_{ij} &= -1+\sum_{i=0}^{\infty}\frac{\Gamma^2(\frac{n+2i}{2})2^{2i}r^{2i}}{\Gamma^2(\frac{n}{2})(2i)!}\int_{-1}^{0}(1-\rho^2)^{\frac{n}{2}}\rho^{2i}g_1(\rho)d\rho \\
&\quad +\sum_{i=0}^{\infty}\frac{\Gamma^2(\frac{n+2i+1}{2})2^{2i+1}r^{2i+1}}{\Gamma^2(\frac{n}{2})(2i+1)!}\int_{-1}^{0}(1-\rho^2)^{\frac{n}{2}}\rho^{2i+1}g_1(\rho)d\rho \\
&> -1+\sum_{i=0}^{\infty}\frac{\Gamma^2(\frac{n+2i}{2})2^{2i}r^{2i}}{\Gamma^2(\frac{n}{2})(2i)!}\int_{0}^{1}(1-\rho^2)^{\frac{n}{2}}\rho^{2i}g_1(-\rho)d\rho
\end{aligned}
$$

By Assumption A3,

$$
\begin{aligned}
&\sum_{i=0}^{\infty}\frac{\Gamma^2(\frac{n+2i}{2})2^{2i}r^{2i}}{\Gamma^2(\frac{n}{2})(2i)!}\int_{0}^{1}(1-\rho^2)^{\frac{n}{2}}\rho^{2i}g_1(-\rho)d\rho \\
\geq\ & n^{-a}\sum_{i=1}^{[\frac{n}{2}]}\frac{\Gamma^2(\frac{n+2i}{2})2^{2i}r^{2i}}{\Gamma^2(\frac{n}{2})(2i)!}\cdot\frac{\Gamma(\frac{n}{2}+1)\Gamma(i+\frac{1}{2})}{\Gamma(\frac{n}{2}+i+\frac{3}{2})} \\
=\ & n^{-a}\sum_{i=1}^{[\frac{n}{2}]}(2r)^{2i}\frac{\Gamma(i+\frac{1}{2})}{\Gamma(2i+1)}\cdot\frac{\Gamma(\frac{n+2i}{2})}{\Gamma(\frac{n}{2})}\cdot\frac{\Gamma(\frac{n}{2}+i)}{\Gamma(\frac{n}{2}+i+\frac{3}{2})}\cdot\frac{\Gamma(\frac{n}{2}+1)}{\Gamma(\frac{n}{2})} \\
=\ & n^{-a}\sum_{i=1}^{[\frac{n}{2}]}(2r)^{2i}\cdot\frac{\sqrt{\pi}2^{-2i}}{\Gamma(i+1)}\cdot(\frac{n}{2}+i-1)\cdots\frac{n}{2}\cdot\frac{\Gamma(\frac{n}{2}+i)}{\Gamma(\frac{n}{2}+i+\frac{1}{2})}\cdot\frac{\frac{n}{2}}{\frac{n}{2}+i+\frac{1}{2}} \\
\geq\ & n^{-a}\sqrt{\pi}\sum_{i=1}^{[\frac{n}{2}]}\frac{r^{2i}}{i!}\cdot(\frac{n}{2})^{i}\cdot\frac{1}{\sqrt{\frac{n}{2}+i}}\cdot\frac{1}{3} \\
\geq\ & \frac{n^{-a}\sqrt{\pi}}{3}\sum_{i=1}^{[\frac{n}{2}]}\frac{(nr^2/2)^i}{i!}\cdot\frac{1}{\sqrt{n}} \\
=\ & \frac{n^{-a}\sqrt{\pi}}{3}\left(\frac{e^{nr^2/2}}{\sqrt{n}}-\frac{1}{\sqrt{n}}-\sum_{i>[\frac{n}{2}]}\frac{(nr^2/2)^i}{i!}\cdot\frac{1}{\sqrt{n}}\right).
\end{aligned}
$$

Notice that $\sum_{i=0}^{\infty}\frac{\Gamma^2(\frac{n+2i}{2})2^{2i}r^{2i}}{\Gamma^2(\frac{n}{2})(2i)!}\int_0^1(1-\rho^2)^{\frac{n}{2}}\rho^{2i}g_1(-\rho)d\rho$ is an decreasing function of $-1<r\leq-\sqrt{\frac{2(a+1)\log n}{n}}$, when $r=-\sqrt{\frac{2(a+1)\log n}{n}}$, $\frac{n^{-a}e^{nr^2/2}}{\sqrt{n}}=\sqrt{n}$ and, by

Stirling's formula, we have that,

$$\sum_{i>[\frac{n}{2}]} \frac{(nr^2/2)^i}{i!} \cdot \frac{1}{\sqrt{n}} \le \sum_{i>[\frac{n}{2}]} \frac{(nr^2/2)^i}{i(\frac{i}{e})^i} \cdot \frac{1}{\sqrt{n}} \le 2 \sum_{i>[\frac{n}{2}]} (er^2)^i \cdot n^{-\frac{3}{2}},$$

which will tend to zero for $r = -\sqrt{\frac{2(a+1)\log n}{n}}$ and when n tends to infinity. Therefore, we conclude that: for any $-1 < r \le -\sqrt{\frac{2(a+1)\log n}{n}}$, $a_{ij}(r) > 0$ for any n large enough.

(3.9) can be proved similarly.

To prove (3.10), first of all, by looking at the density function $f_{r_{ij}|\rho_{ij}}(r|\rho)$ we have, for any $-1 < r, \rho < 1$:

$$\sum_{i=0}^{\infty} \frac{(2\rho r)^i}{i!} \Gamma^2(\frac{n+i}{2}) > 0.$$

Therefore, for any $-1 < r \le -\sqrt{\frac{2(a+1)\log n}{n}}$,

$$\begin{aligned} a_{ij} &> -1 + \sum_{i=0}^{\infty} \frac{\Gamma^2(\frac{n+i}{2}) 2^i r^i}{\Gamma^2(\frac{n}{2})(i)!} \int_{-1}^{0} (1-\rho^2)^{\frac{n}{2}} \rho^i g_1(\rho) d\rho \\ &> -1 + \sum_{i=0}^{\infty} \frac{\Gamma^2(\frac{n+2i}{2}) 2^{2i} r^{2i}}{\Gamma^2(\frac{n}{2})(2i)!} \int_{0}^{1} (1-\rho^2)^{\frac{n}{2}} \rho^{2i} g_1(-\rho) d\rho, \end{aligned}$$

which will be positive for any n large enough as shown before. Similarly, for any $\sqrt{\frac{2(a+1)\log n}{n}} \le r < 1$, we have,

$$\begin{aligned} a_{ij} &> -1 + \sum_{i=0}^{\infty} \frac{\Gamma^2(\frac{n+i}{2}) 2^i r^i}{\Gamma^2(\frac{n}{2})(i)!} \int_{0}^{1} (1-\rho^2)^{\frac{n}{2}} \rho^i g_2(\rho) d\rho \\ &> -1 + \sum_{i=0}^{\infty} \frac{\Gamma^2(\frac{n+2i}{2}) 2^{2i} r^{2i}}{\Gamma^2(\frac{n}{2})(2i)!} \int_{0}^{1} (1-\rho^2)^{\frac{n}{2}} \rho^{2i} g_2(\rho) d\rho, \end{aligned}$$

which will tend to be positive for any n large enough.

□

Particularly, if g is a bounded function (both from below and above for some positive constants) with support [-1,1], we have $a = b = 0$ and the two lemmas above implies that, for any n large enough, the threshold is within the interval: $(\frac{\log n}{2n}, \frac{2\log n}{n})$.

The following observations will be used in the proof of consistency of $\hat{\omega}_1$:

For any $1 \le i < j \le p, 1 \le s < t \le p$,

$$\begin{aligned}
&No.\ of\ pairs\ \{(i,j),(s,t)\} = \frac{p^2(p-1)^2}{4};\\
&No.\ of\ pairs\ \{(i,j),(s,t):\ i,j,s,t\ all\ distinct\} = \frac{p(p-1)(p-2)(p-3)}{4};\\
&No.\ of\ pairs\{(i,j),(s,t):(i,j)\cap(s,t) = i\ or\ j\} = p(p-1)(p-2);\\
&No.\ of\ pairs\{(i,j),(s,t): i = s, j = t\} = \frac{p(p-1)}{2}.
\end{aligned}$$

Theorem 3.2. *Under Assumptions A1-A4,*

$$\hat{\omega}_1 \to \omega,\quad in\ \ probability\ \ as\ \ n\ \ and\ \ p\ \ \to\ \infty.$$

Proof. First of all, by Markov's inequality and Assumption A2, for any $\varepsilon > 0$,

$$P(|\frac{2}{p(p-1)}\sum_{1\le i<j\le p} I_{\{\rho_{ij}\neq 0\}} - (1-\omega)| > \varepsilon)$$

$$
\begin{aligned}
&= P(|\frac{2}{p(p-1)}\sum_{1\le i<j\le p} I_{\{\rho_{ij}\neq 0\}} - E\frac{2}{p(p-1)}\sum_{1\le i<j\le p} I_{\{\rho_{ij}\neq 0\}}| > \varepsilon) \\
&\le \varepsilon^{-2} Var\left[\frac{2}{p(p-1)}\sum_{1\le i<j\le p} I_{\{\rho_{ij}\neq 0\}}\right] \\
&= \frac{4\varepsilon^{-2}}{p^2(p-1)^2}[\frac{p(p-1)}{2}Var(I_{\{\rho_{12}\neq 0\}}) + \\
&\quad p(p-1)(p-2)Cov(I_{\{\rho_{12}\neq 0\}}, I_{\{\rho_{23}\neq 0\}}) + O(p^{\nu})] \\
&\le \frac{4\varepsilon^{-2}}{p^2(p-1)^2}\left[\frac{p(p-1)}{2}\omega(1-\omega) + p(p-1)(p-2)(1-\omega)\omega + O(p^{\nu})\right],
\end{aligned}
$$

where in the last step we have use Assumption A2 and the fact that

$$
\begin{aligned}
& Var(I_{\{\rho_{12}\neq 0\}}) = \omega(1-\omega); \\
& E(I_{\{\rho_{12}\neq 0\}} - EI_{\{\rho_{12}\neq 0\}})(I_{\{\rho_{23}\neq 0\}} - EI_{\{\rho_{23}\neq 0\}}) \\
& \le [E(I_{\{\rho_{12}\neq 0\}} - EI_{\{\rho_{12}\neq 0\}})^2]^{\frac{1}{2}}[E(I_{\{\rho_{23}\neq 0\}} - EI_{\{\rho_{23}\neq 0\}})^2]^{\frac{1}{2}} \\
& = (1-\omega)\omega.
\end{aligned}
$$

Therefore, $\frac{2}{p(p-1)}\sum_{1\le i<j\le p} I_{\{\rho_{ij}\neq 0\}}$ converges in probability to $1-\omega$ as p tends to infinity.

To prove that $\hat{\omega}_1 \to \omega$ in probability, it suffices to show that $1-\hat{\omega}_1$ converge to $\frac{2}{p(p-1)}\sum_{1\le i<j\le p} I_{\{\rho_{ij}\neq 0\}}$ in probability. Now by Markov's inequality we have, for any constant $\varepsilon > 0$:

$$
\begin{aligned}
& P(|1-\hat{\omega}_1 - \frac{2}{p(p-1)}\sum_{1\le i<j\le p} I_{\{\rho_{ij}\neq 0\}}| > \varepsilon) \\
= \; & P(\frac{2}{p(p-1)}|\sum_{1\le i<j\le p} [I_{\{a_{ij}>0\}} - I_{\{\rho_{ij}\neq 0\}}]| > \varepsilon)
\end{aligned}
$$

$$\leq \frac{4}{\varepsilon^2 p^2 (p-1)^2} E[\sum_{1\leq i<j\leq p} (I_{\{a_{ij}>0\}} - I_{\{\rho_{ij}\neq 0\}})]^2.$$

Write

$$Y_{ij} = I_{\{a_{ij}>0\}} - I_{\{\rho_{ij}\neq 0\}}, \quad \forall\ 1 \leq i < j \leq p.$$

Notice that,

$$E(\sum_{1\leq i<j\leq p} Y_{ij})^2 \leq \frac{p(p-1)}{2} \sum_{1\leq i<j\leq p} EY_{ij}^2 = \frac{p^2(p-1)^2}{4} EY_{ij}^2.$$

By Lemma 3.5, for any n large enough, we have:

$$\begin{aligned} EY_{ij}^2 &= P(a_{ij} > 0, \rho_{ij} = 0) + P(a_{ij} \leq 0, \rho_{ij} \neq 0) \\ &\leq P(r_{ij}^2 > \frac{(1/2 - b - c)\log n}{n} | \rho_{ij} = 0)\omega + P(0 < |\rho_{ij}| \leq n^{-\frac{1}{2}+v}) \\ &\quad + P(a_{ij} \leq 0, |\rho_{ij}| > n^{-\frac{1}{2}+v}). \end{aligned}$$

Now, by Lemma 3.3, there exist constants $d, f > 0$, such that

$$P(r_{ij}^2 > \frac{(1/2 - b - c)\log n}{n} | \rho_{ij} = 0) \leq de^{-f\log n} = dn^{-f} \to 0.$$

By Assumption A4,

$$P(0 < |\rho_{ij}| \leq n^{-\frac{1}{2}+v}) \to 0.$$

If $0 < \int_0^1 g_2(\rho)d\rho < 1$, by Lemma 3.3 and Lemma 3.6, for n large enough, there exist $l, m > 0$, such that:

$$P(a_{ij} \leq 0, |\rho_{ij}| > n^{-\frac{1}{2}+v})$$

$$
\begin{aligned}
&= (1-\omega)\left[\int_{n^{-\frac{1}{2}+v}}^{1} P(a_{ij}\le 0|\rho_{ij}=\rho)dG(\rho) + \int_{-1}^{-n^{-\frac{1}{2}+v}} P(a_{ij}\le 0|\rho_{ij}=\rho)dG(\rho)\right] \\
&\le (1-\omega)\big[\int_{n^{-\frac{1}{2}+v}}^{1} P(r_{ij}^2 \le \frac{2(a+1)\log n}{n}|\rho_{ij}=\rho)dG(\rho) \\
&\quad + \int_{-1}^{-n^{-\frac{1}{2}+v}} P(r_{ij}^2 \le \frac{2(a+1)\log n}{n}|\rho_{ij}=\rho)dG(\rho)\big] \\
&\le (1-\omega)\big[\int_{n^{-\frac{1}{2}+v}}^{1} P(|r_{ij}-\rho_{ij}| > \frac{1}{2}n^{-\frac{1}{2}+v}|\rho_{ij}=\rho)dG(\rho) \\
&\quad + \int_{-1}^{-n^{-\frac{1}{2}+v}} P(|r_{ij}-\rho_{ij}| > \frac{1}{2}n^{-\frac{1}{2}+v}|\rho_{ij}=\rho)dG(\rho)\big] \\
&\le 2(1-\omega)le^{-\frac{m}{2}n^{2v}},
\end{aligned}
$$

which will converge to zero when n increases. Similarly, if $\int_0^1 g_2(\rho)d\rho = 0 \textit{ or } 1$, we can show that $P(a_{ij} \le 0, |\rho_{ij}| > n^{-\frac{1}{2}+v})$ converges to zero as n tends to infinity.

Hence $EY_{ij}^2 \to 0$ and consequently,

$$P(|1-\hat{\omega}_1 - \frac{2}{p(p-1)}\sum_{1\le i<j\le p} I_{\{\rho_{ij}\neq 0\}}| > \varepsilon) \le EY_{ij}^2/\varepsilon^2 \to 0.$$

□

From the proof of Theorem 3.2, we know that for any constant $c > 0$, if we choose the thresholding parameter t to be $c\sqrt{\log n/n}$, we can estimate ω consistently as n tends to infinity as long as Assumption A4 is satisfied. Notice that a_{ij} can be written as:

$$a_{ij} = -1 + \frac{\int_{-1}^{1} f_{r_{ij}|\rho_{ij}}(r|\rho)dG(\rho)}{f_{r_{ij}|\rho_{ij}=0}(r|\rho=0)}. \tag{3.11}$$

Therefore our estimator $\hat{\omega}_1$ is aiming to choose c based on comparing whether the ratio between the marginal likelihood given $\rho \neq 0$ and the likelihood given $\rho = 0$ is greater than 1 or not. Practically, $G(\rho)$ is usually unknown. However, all the lemmas and Theorem 3.2 are valid for any prior such that Assumptions A1-A4 are satisfied. Practically, we propose to threshold r_{ij} to be $r_{ij}I_{\{|r_{ij}|\leq\sqrt{2\log n/n}\}}, 1 \leq i < j \leq p$ and treat those nonzero $r_{ij}I_{\{|r_{ij}|\leq\sqrt{2\log n/n}\}}$ as random samples from $G(\rho)$. Therefore we can achieve an estimator for g, say, $\hat{g}$, and construct an estimator: $\hat{\omega}_1(\hat{g})$. As long as $\hat{g}(\rho)$ is a bounded function in [-1,1], $\hat{\omega}_1(\hat{g})$ will be consistent in estimating ω. More discussions will be given in Section 3.5.

3.4 Method-of-moments estimator

In this section, we assume that $X_1, ..., X_n$ satisfy the moment conditions (3.1) and (3.2) introduced in Section 3.2. Similar to the last section, we first of all propose some assumptions on the marginal prior density for the $p(p-1)/2$ correlation coefficients:

A5:

$$\rho_{ij} \sim \omega\delta_0 + (1-\omega)h_{ij}(\rho), \quad 1 \leq j < i \leq p,$$

such that $\sup_{\theta\in(-1,1)} h(\rho) < \infty$;

A6: for any $1 \le j < i \le p, 1 \le t < s \le p$, if $\{i,j\} \cap \{s,t\} = \emptyset$, ρ_{ij} is independent of $\rho_{s,t}$.

A2* Since we are assuming the marginal prior densities of the ρ_{ij}'s are the same, a more natural assumption than $A2$ would be that, we assume the prior distribution of $\rho_{ij}, 1 \le i < j \le p$ is invariant with respect to the subscripts. Consequently, Assumption A2 becomes

$$ReE[(e^{-\mathbf{i}k\rho_{ij}} - Ee^{-\mathbf{i}k\rho_{ij}})(e^{\mathbf{i}k\rho_{st}} - Ee^{\mathbf{i}k\rho_{st}})] = o(1) \;\; as \;\; p \to \infty, \forall\, i,j,s,t: \; distinct.$$

Remark 3.2. Let $\Sigma = R'R$ be the Cholesky decomposition of Σ with the matrix R upper triangular. For any distribution of R such that the rows of R are independent of each other, the corresponding distribution of Σ will satisfy Assumption A6. Examples can be seen in Model 2 and Model 3 in the simulation study.

It is easy to see that Assumption A5 is slightly stronger than Assumptions A1, A3 and A4. Assumption A6 is stronger than A2. We first look at the problem under Assumptions A5 and A6. Later we will relax Assumption A6 to Assumption A2.

For any positive integer q, we define a matrix-valued function $T_q : (-1, 1) \to$

$\mathbb{C}^{(q+1)\times(q+1)}$ by:

$$T_q(\rho) = \begin{pmatrix} 1 & e^{\mathbf{i}\rho} & e^{\mathbf{i}2\rho} & \dots & e^{\mathbf{i}q\rho} \\ e^{-\mathbf{i}\rho} & 1 & e^{\mathbf{i}\rho} & \dots & e^{\mathbf{i}(q-1)\rho} \\ e^{-\mathbf{i}2\rho} & e^{-\mathbf{i}\rho} & 1 & \dots & e^{\mathbf{i}(q-2)\rho} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ e^{-\mathbf{i}q\rho} & e^{-\mathbf{i}(q-1)\rho} & e^{-\mathbf{i}(q-2)\rho} & \dots & 1 \end{pmatrix} = \begin{pmatrix} 1 \\ e^{-\mathbf{i}\rho} \\ \vdots \\ e^{-\mathbf{i}q\rho} \end{pmatrix} \begin{pmatrix} 1, & e^{\mathbf{i}\rho}, & \dots, & e^{\mathbf{i}q\rho} \end{pmatrix}.$$

where $\mathbf{i} = \sqrt{-1}$. Further define $M_q = ET_q(\rho) = M_{q,disc} + M_{q,cont}$, where

$$M_{q,disc} = \omega T_q(0) = \begin{pmatrix} \omega & \cdots & \omega \\ \vdots & \ddots & \vdots \\ \omega & \cdots & \omega \end{pmatrix},$$

$$M_{q,cont} = (1-\omega)\int_{-1}^{1} T_q(\rho)h(\rho)d\rho.$$

Let $\lambda_i(A)$ denote the ith largest eigenvalues of A where A is an arbitrary $(q+1)\times(q+1)$ Hermitian matrix.

Lemma 3.7. *With the notation and assumptions of ρ given above, we have: (i)$\lambda_1(M_{q,disc}) = (q+1)\omega$ and $\lambda_i(M_{q,disc}) = 0$ for all $i = 2, ..., q+1$, and (ii):*

$$0 \leq \lambda_{q+1}(M_{q,cont}) \leq \lambda_1(M_{q,cont}) \leq 2\pi(1-\omega) \sup_{-1<\rho<1} h(\rho).$$

Proof. (i) is straightforward. For (ii), let $a = (a_1, ..., a_{q+1})' \in \mathbb{C}^{(q+1)}$ and $\bar{a} = (\bar{a}_1, ..., \bar{a}_{q+1})'$ is the complex conjugate of a. Then

$$\begin{aligned}
\frac{\bar{a}' M_{q,cont} a}{\bar{a}' a} &= \frac{(1-\omega) \sum_{k=1}^{q+1} \sum_{j=1}^{q+1} \int_{-1}^{1} a_k \bar{a}_j e^{\mathbf{i}(k-j)\rho} h(\rho) d\rho}{\bar{a}' a} \\
&= \frac{(1-\omega) \int_{-1}^{1} |\sum_{k=1}^{q+1} a_k e^{\mathbf{i}k\rho}|^2 h(\rho) d\rho}{\sum_{k=1}^{q+1} |a_k|^2} \\
&= \frac{2\pi(1-\omega) \int_{-1}^{1} |\sum_{k=1}^{q+1} a_k e^{\mathbf{i}k\rho}|^2 h(\rho) d\rho}{\int_{-\pi}^{\pi} |\sum_{k=1}^{q+1} a_k e^{\mathbf{i}k\rho}|^2 d\rho}.
\end{aligned}$$

Thus for arbitrary $a \in \mathbb{C}^{(q+1)}$ such that $|a| = 1$,

$$0 \leq \lambda_{q+1}(M_{q,cont}) \leq \lambda_1(M_{q,cont}) \leq 2\pi(1-\omega) \sup_{-1<\rho<1} h(\rho).$$

□

Theorem 3.3. *With the above notation, we have:*

$$0 \leq \frac{\lambda_1(M_q)}{q+1} - \omega \leq \frac{2\pi(1-\omega)\sup_{-1<\rho<1} h(\rho)}{q+1}.$$

Also, for $i = 2, ..., q+1$,

$$0 \leq \frac{\lambda_i(M_q)}{q+1} \leq \frac{2\pi(1-\omega)\sup_{-1<\rho<1} h(\rho)}{q+1}.$$

Proof. Since $M_q = M_{q,disc} + M_{q,cont}$, by Lemma 3.7 we have

$$\begin{aligned}
\lambda_1(M_q) &\geq \lambda_1(M_{q,disc}) + \lambda_{q+1}(M_{q,cont}) \\
&\geq (q+1)\omega, \\
\lambda_1(M_q) &\leq \lambda_1(M_{q,disc}) + \lambda_1(M_{q,cont})
\end{aligned}$$

$$\leq \quad (q+1)\omega + 2\pi(1-\omega) \sup_{-1<\rho<1} h(\rho).$$

For $i = 2, ..., q+1$, by Lemma 3.7, M_q is nonnegative definite therefore $0 \leq \frac{\lambda_i(M_q)}{q+1}$. Also, by Theorem A.8. of Bai and Silverstein, we have,

$$\begin{aligned} \lambda_i(M_q) &\leq \lambda_i(M_{q,disc}) + \lambda_1(M_{q,cont}) \\ &\leq 2\pi(1-\omega) \sup_{-1<\rho<1} h(\rho). \end{aligned}$$

□

The following is an immediate corollary of Theorem 3.3.

Corollary 3.2. *Suppose that* $\sup_{-1<\rho<1} h(\rho) < \infty$. *Then* $\frac{\lambda_1(M_q)}{q+1} \to \omega$ *as* $q \to \infty$.

This corollary gives, at least in principle, a way for estimating ω by estimating the largest eigenvalue of M_q for a sufficiently large q.

We estimate M_q by the $(q+1) \times (q+1)$ matrix $\hat{M}_q$ whose (k,l)th element is given by:

$$(\hat{M}_q)_{k,l} = \frac{2}{p(p-1)} \sum_{1 \leq j < i \leq p} e^{-\mathbf{i}(k-l)r_{ij}}, \quad \forall 1 \leq k, l \leq q+1.$$

Define

$$\hat{\omega}_2 = \frac{\lambda_1(\hat{M}_q)}{q+1}.$$

Then we have:

Proposition 3.1.

$$E\,|\hat{\omega}_2 - \omega| \leq \frac{2}{q+1}\sum_{k=1}^{q} E|(\hat{M}_q - M_q)_{q+1,k}| + \frac{2\pi(1-\omega)}{q+1}\sup_{\theta\in(-1,1)} h(\theta).$$

Proof. Denote the matrix norm induced by $l_1 - norm$ for vectors as $\parallel A \parallel_1$. We have:

$$\parallel A \parallel_1 = \max_{1\leq j\leq q+1}\sum_{i=1}^{q+1}|a_{ij}|,$$

which is the maximum absolute column sum of the matrix A. Write $\rho(A)$ as the spectral radius of A. We observe from Theorem 5.6.9 of Horn and Johnson (1985) that

$$\begin{aligned}
\lambda_1(\hat{M}_q - M_q) &\leq \rho(\hat{M}_q - M_q) \\
&\leq \parallel \hat{M}_q - M_q \parallel_1 \\
&\leq 2\sum_{k=1}^{q+1}|(\hat{M}_q - M_q)_{q+1,k}|.
\end{aligned}$$

Let $\tilde{e} = (1/\sqrt{q+1}, ..., 1/\sqrt{q+1})$. Then it follows from Theorem 3.3 that

$$\begin{aligned}
\lambda_1(\hat{M}_q) &\geq \tilde{e}' M_{q,disc}\tilde{e} + \tilde{e}' M_{q,cont}\tilde{e} + \tilde{e}'(\hat{M}_q - M_q)\tilde{e} \\
&\geq (q+1)\omega - \frac{1}{q+1}\sum_{j=1}^{q+1}\sum_{k=1}^{q+1}|(\hat{M}_q - M_q)_{j,k}| \\
&\geq (q+1)\omega - 2\sum_{k=1}^{q+1}|(\hat{M}_q - M_q)_{q+1,k}|,
\end{aligned}$$

and

$$\lambda_1(\hat{M}_q) \leq \lambda_1(M_q) + \lambda_1(\hat{M}_q - M_q)$$

$$\le \quad (q+1)\omega + 2\pi(1-\omega) \sup_{-1<\rho<1} h(\rho) + 2\sum_{k=1}^{q+1} |(\hat{M}_q - M_q)_{q+1,k}|.$$

Thus we have:

$$|\lambda_1(\hat{M}_q) - (q+1)\omega| \le 2\pi(1-\omega) \sup_{-1<\rho<1} h(\rho) + 2\sum_{k=1}^{q+1} |(\hat{M}_q - M_q)_{q+1,k}|.$$

The theorem is proved by dividing $q+1$ on both sides of the above inequality. □

Let the sample correlation matrix $R = (r_{ij})_{p\times p}$ and the population correlation matrix $\Gamma = (\rho_{ij})_{p\times p}$ be define as in Section 3.2. The following lemma will be used in the proof of the main theorems of this section.

Lemma 3.8. *With the notation of Section 3.2, we have*

$$E(r_{ij} - \rho_{ij})^2 \le \frac{8.5}{n} E\big(\frac{X_{i1}^4}{\sigma_{ii}^2} + \frac{X_{j1}^4}{\sigma_{jj}^2}\big), \quad \forall 1 \le i,j \le p.$$

Proof. Denote E_Σ as the conditional expectation given Σ. We observe that for $1 \le i,j \le$ and constant $a > 0$,

$$\begin{aligned} E_\Sigma(r_{ij} - \rho_{ij})^2 \quad &\le \quad E_\Sigma\left[(r_{ij} - \rho_{ij})^2 I_{\{|s_{ii}-\sigma_{ii}| \le \sigma_{ii}/a\}} I_{\{|s_{jj}-\sigma_{jj}| \le \sigma_{jj}/a\}}\right] \\ &\quad +4\sum_{l=i,j} P(|s_{ll} - \sigma_{ll}| > \sigma_{ll}/a|\Sigma). \end{aligned}$$

Notice that

$$\begin{aligned} P(|s_{ll} - \sigma_{ll}| > \sigma_{ll}/a|\Sigma) \quad &\le \quad \frac{a^2}{\sigma_{ll}^2} E_\Sigma[\frac{1}{n}\sum_{k=1}^{n}(X_{lk}^2 - \sigma_{ll})]^2 \\ &= \quad \frac{a^2}{n\sigma_{ll}^2} E_\Sigma(X_{l1}^2 - \sigma_{ll})^2 \end{aligned}$$

$$\leq \frac{a^2}{n} E_\Sigma \frac{X_{l1}^4}{\sigma_{ll}^2}, \qquad l = i, j,$$

and

$$\begin{aligned}
& E_\Sigma \left[(r_{ij} - \rho_{ij})^2 I_{\{|s_{ii} - \sigma_{ii}| \leq \sigma_{ii}/a\}} I_{\{|s_{jj} - \sigma_{jj}| \leq \sigma_{jj}/a\}} \right] \\
\leq\ & E_\Sigma \Big(\frac{s_{ij}}{\sqrt{s_{ii} s_{jj}}} - \frac{s_{ij}}{\sqrt{\sigma_{ii} s_{jj}}} + \frac{s_{ij}}{\sqrt{\sigma_{ii} s_{jj}}} - \frac{s_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}} + \frac{s_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}} - \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}} \Big)^2 \\
& \times I_{\{|s_{ii} - \sigma_{ii}| \leq \sigma_{ii}/a\}} I_{\{|s_{jj} - \sigma_{jj}| \leq \sigma_{jj}/a\}} \\
=\ & E_\Sigma \Big\{ \Big[\frac{s_{ij}(\sigma_{ii} - s_{ii})}{(\sqrt{\sigma_{ii}} + \sqrt{s_{ii}})\sqrt{s_{ii}\sigma_{ii}s_{jj}}} + \frac{s_{ij}(\sigma_{jj} - s_{jj})}{(\sqrt{\sigma_{jj}} + \sqrt{s_{jj}})\sqrt{\sigma_{ii}\sigma_{jj}s_{jj}}} + \frac{s_{ij} - \sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}} \Big]^2 \\
& \times I_{\{|s_{ii} - \sigma_{ii}| \leq \sigma_{ii}/a\}} I_{\{|s_{jj} - \sigma_{jj}| \leq \sigma_{jj}/a\}} \Big\} \\
\leq\ & E_\Sigma \Big[\frac{|\sigma_{ii} - s_{ii}|}{\sigma_{ii}} + \frac{\sqrt{1 + 1/a}|\sigma_{jj} - s_{jj}|}{\sigma_{jj}} + \frac{|s_{ij} - \sigma_{ij}|}{\sqrt{\sigma_{ii}\sigma_{jj}}} \Big]^2 \\
\leq\ & 3E_\Sigma \Big[\frac{(\sigma_{ii} - s_{ii})^2}{\sigma_{ii}^2} + \frac{(1 + 1/a)(\sigma_{jj} - s_{jj})^2}{\sigma_{jj}^2} + \frac{(s_{ij} - \sigma_{ij})^2}{\sigma_{ii}\sigma_{jj}} \Big] \\
=\ & \frac{3}{n} \Big[\frac{E_\Sigma (X_{i1}^2 - \sigma_{ii})^2}{\sigma_{ii}^2} + \frac{(1 + 1/a) E_\Sigma (X_{j1}^2 - \sigma_{jj})^2}{\sigma_{jj}^2} + \frac{E_\Sigma (X_{i1} X_{j1} - \sigma_{ij})^2}{\sigma_{ii}\sigma_{jj}} \Big] \\
\leq\ & \frac{1}{n} E_\Sigma \Big[\frac{9}{2} \frac{X_{i1}^4}{\sigma_{ii}^2} + (\frac{9}{2} + 3/a) \frac{X_{j1}^4}{\sigma_{jj}^2} \Big],
\end{aligned}$$

where in the last step we have used the fact that

$$\frac{E_\Sigma (X_{i1} X_{j1} - \sigma_{ij})^2}{\sigma_{ii}\sigma_{jj}} \leq \frac{E_\Sigma X_{i1}^2 X_{j1}^2}{\sigma_{ii}\sigma_{jj}} \leq \frac{1}{2} E_\Sigma \Big(\frac{X_{i1}^4}{\sigma_{ii}^2} + \frac{X_{j1}^4}{\sigma_{jj}^2} \Big).$$

Consequently we conclude that

$$E_\Sigma (r_{ij} - \rho_{ij})^2 \leq \frac{1}{n} E_\Sigma \big[(4a^2 + \frac{9}{2}) \frac{X_{i1}^4}{\sigma_{ii}^2} + (4a^2 + \frac{9}{2} + \frac{3}{a}) \frac{X_{j1}^4}{\sigma_{jj}^2} \big].$$

By changing the indices i with j we have:

$$E_\Sigma (r_{ij} - \rho_{ij})^2 \leq \frac{1}{n} E_\Sigma \big[(4a^2 + \frac{9}{2}) \frac{X_{j1}^4}{\sigma_{jj}^2} + (4a^2 + \frac{9}{2} + \frac{3}{a}) \frac{X_{i1}^4}{\sigma_{ii}^2} \big].$$

Therefore we conclude that

$$E_\Sigma(r_{ij}-\rho_{ij})^2 \le \frac{8a^2+9+\frac{3}{a}}{2n} E_\Sigma\Big[\frac{X_{i1}^4}{\sigma_{ii}^2}+\frac{X_{j1}^4}{\sigma_{jj}^2}\Big], \quad \forall 1 \le i,j \le p.$$

By letting $a=(3/16)^{1/3}$, the right hand side of the above equation is minimized:

$$E_\Sigma(r_{ij}-\rho_{ij})^2 \le \frac{8.5}{n} E_\Sigma\Big[\frac{X_{i1}^4}{\sigma_{ii}^2}+\frac{X_{j1}^4}{\sigma_{jj}^2}\Big], \quad \forall 1 \le i,j \le p.$$

Lemma 3.8 is proved by taking expectation with respect to Σ in the above inequality. $\square$

Theorems 3.4, 3,5 and 3.6 are the main results of this section. They provide upper bounds to the convergence rate of $\hat{\omega}_2$.

Theorem 3.4. *Suppose $X_1,...,X_n$ are i.i.d. mean zero random vectors such that $0<\sup_{i\in\{1,2,..,\}} EX_{i1}^4/\sigma_{ii}^2<\infty$, where X_{i1} is the ith element of $X_1=(X_{11},...,X_{p1})'$. Under Assumptions A5, A6 and with $\hat{\omega}_2$ defined above, there exists a constant $c>0$ large enough, such that*

$$E\left|\hat{\omega}_2-\omega\right| \le \frac{cq}{\sqrt{n}}+\frac{8}{\sqrt{p}}+\frac{2\pi(1-\omega)}{q+1}\sup_{\rho\in(-1,1)} h(\rho).$$

Particularly, when $X_1,...,X_n$ are i.i.d. mean zero multivariate normal random vectors, we have,

$$E\left|\hat{\omega}_2-\omega\right| \le \frac{q}{\sqrt{n}}+\frac{8}{\sqrt{p}}+\frac{2\pi(1-\omega)}{q+1}\sup_{\rho\in(-1,1)} h(\rho).$$

Proof. By Proposition 3.1, we only need to bound $\sum_{k=1}^{q} E|(\hat{M}_q - M_q)_{q+1,k}|$. Write $N = p(p-1)/2$, we have:

$$
\begin{aligned}
& E|N^{-1} \sum_{1 \le j < i \le p} (e^{-\mathbf{i}kr_{ij}} - Ee^{-\mathbf{i}k\rho_{ij}})|^2 \\
= \; & N^{-2} \sum_{1 \le j < i \le p} \sum_{1 \le t < s \le p} E(e^{-\mathbf{i}kr_{ij}} - Ee^{-\mathbf{i}k\rho_{ij}})(e^{\mathbf{i}kr_{st}} - Ee^{\mathbf{i}k\rho_{st}}) \\
= \; & N^{-2} \sum_{1 \le j < i \le p} \sum_{1 \le t < s \le p} E[(e^{-\mathbf{i}kr_{ij}} - e^{-\mathbf{i}k\rho_{ij}})(e^{\mathbf{i}kr_{st}} - e^{\mathbf{i}k\rho_{st}}) \\
& + (e^{-\mathbf{i}kr_{ij}} - e^{-\mathbf{i}k\rho_{ij}})(e^{\mathbf{i}k\rho_{st}} - Ee^{\mathbf{i}k\rho_{st}}) + (e^{-\mathbf{i}k\rho_{ij}} - Ee^{-\mathbf{i}k\rho_{ij}})(e^{\mathbf{i}kr_{st}} - e^{\mathbf{i}k\rho_{st}}) \\
& + (e^{-\mathbf{i}k\rho_{ij}} - Ee^{-\mathbf{i}k\rho_{ij}})(e^{\mathbf{i}k\rho_{st}} - Ee^{\mathbf{i}k\rho_{st}})]
\end{aligned}
$$

When $\{i,j\} \cap \{s,t\} = \emptyset$, the last three terms in the above equation equal to zero and

$$
\begin{aligned}
& E(e^{-\mathbf{i}kr_{ij}} - e^{-\mathbf{i}k\rho_{ij}})(e^{\mathbf{i}kr_{st}} - e^{\mathbf{i}k\rho_{st}}) \\
\le \; & \sqrt{E|(e^{-\mathbf{i}kr_{ij}} - e^{-\mathbf{i}k\rho_{ij}})|^2 E|(e^{-\mathbf{i}kr_{st}} - e^{-\mathbf{i}k\rho_{st}})|^2} \\
\le \; & \sqrt{E|e^{-\mathbf{i}k\rho_{ij}}(e^{-\mathbf{i}k(r_{ij}-\rho_{ij})} - 1)|^2 E|e^{-\mathbf{i}k\rho_{st}}(e^{-\mathbf{i}k(r_{st}-\rho_{st})} - 1)|^2} \\
= \; & \sqrt{E|e^{-\mathbf{i}k(r_{ij}-\rho_{ij})} - 1|^2 E|e^{-\mathbf{i}k(r_{st}-\rho_{st})} - 1|^2} \\
= \; & 2\sqrt{E(1 - \cos k(r_{ij} - \rho_{ij}))E(1 - \cos k(r_{st} - \rho_{st}))} \\
= \; & 4\sqrt{E\sin^2(k(r_{ij} - \rho_{ij})/2)E\sin^2(k(r_{st} - \rho_{st})/2)} \\
\le \; & k^2\sqrt{E(r_{ij} - \rho_{ij})^2 E(r_{st} - \rho_{st})^2}.
\end{aligned}
$$

By Lemma 3.8, there exists a constant $c > 0$, depending on $\sup_{i \in \{1,2,..,\}} EX_{i1}^4$ only such that for any $1 \le i, j \le p$

$$
E(r_{ij} - \rho_{ij})^2 \le \frac{c^2}{n}.
$$

When $\{i,j\} \cap \{s,t\} \neq \emptyset$

$$\begin{aligned} & Re(E(e^{-\mathbf{i}kr_{ij}} - Ee^{-\mathbf{i}k\rho_{ij}})(e^{\mathbf{i}kr_{st}} - Ee^{\mathbf{i}k\rho_{st}})) \\ \leq\ & E|(e^{-\mathbf{i}kr_{ij}} - Ee^{-\mathbf{i}k\rho_{ij}})(e^{\mathbf{i}kr_{st}} - Ee^{\mathbf{i}k\rho_{st}})| \\ \leq\ & 4. \end{aligned}$$

Therefore,

$$\begin{aligned} E|N^{-1} \sum_{1 \leq j < i \leq p} (e^{-\mathbf{i}kr_{ij}} - Ee^{-\mathbf{i}k\rho_{ij}})|^2 &\leq \frac{c^2k^2}{n} + \frac{4p(p-1)(p-\frac{3}{2})}{N^2} \\ &< \frac{c^2k^2}{n} + \frac{16}{p}. \end{aligned}$$

By Holder's inequality we have:

$$\begin{aligned} \sum_{k=1}^{q} E|(\hat{M}_q - M_q)_{q+1,k}| &\leq \sum_{k=1}^{q} (E|N^{-1} \sum_{1 \leq j < i \leq p} (e^{-\mathbf{i}kr_{ij}} - Ee^{-\mathbf{i}k\rho_{ij}})|^2)^{\frac{1}{2}} \\ &< \sum_{k=1}^{q} (\frac{ck}{\sqrt{n}} + \frac{4}{\sqrt{p}}) \\ &= \frac{cq(q+1)}{2\sqrt{n}} + \frac{4q}{\sqrt{p}}. \end{aligned}$$

Consequently, by Proposition 3.1, we have

$$E\,|\hat{\omega}_2 - \omega| \leq \frac{cq}{\sqrt{n}} + \frac{8}{\sqrt{p}} + \frac{2\pi(1-\omega)}{q+1} \sup_{\rho \in (-1,1)} h(\rho).$$

Particularly, when $X_1, ..., X_n$ are i.i.d. mean zero multivariate normal random vectors, we observe from Kendall (1960) that given ρ

$$E(r-\rho)^2 = (1-\rho^2)^2[\frac{1}{n} + \frac{23\rho^2}{4n^2} + O(n^{-3})].$$

Therefore, when $\{i,j\}\cap\{s,t\}=\emptyset$,

$$\begin{aligned} E(e^{-\mathbf{i}kr_{ij}}-e^{-\mathbf{i}k\rho_{ij}})(e^{\mathbf{i}kr_{st}}-e^{\mathbf{i}k\rho_{st}}) &\le Ek^2(r_{ij}-\rho_{ij})^2 \\ &\le \frac{k^2}{n}+O(k^2n^{-2}). \end{aligned}$$

Therefore,

$$\begin{aligned} & E\left|\hat{\omega}_2-\omega\right| \\ \le\ & \frac{2}{q+1}\sum_{k=1}^{q}E|(\hat{M}_q-M_q)_{q+1,k}|+\frac{2\pi(1-\omega)}{q+1}\sup_{\rho\in(-1,1)}h(\rho) \\ \le\ & \frac{2}{q+1}\sum_{k=1}^{q}(E|N^{-1}\sum_{1\le j<i\le p}(e^{-\mathbf{i}kr_{ij}}-Ee^{-\mathbf{i}k\rho_{ij}})|^2)^{\frac{1}{2}}+\frac{2\pi(1-\omega)}{q+1}\sup_{\rho\in(-1,1)}h(\rho) \\ \le\ & \frac{2}{q+1}\sum_{k=1}^{q}(\frac{k}{\sqrt{n}}+\frac{4}{\sqrt{p}}+O(kn^{-1}))+\frac{2\pi(1-\omega)}{q+1}\sup_{\rho\in(-1,1)}h(\rho) \\ \le\ & \frac{q}{\sqrt{n}}+\frac{8}{\sqrt{p}}+O(qn^{-1})+\frac{2\pi(1-\omega)}{q+1}\sup_{\theta\in(-1,1)}h(\theta). \end{aligned}$$

□

By letting $q=O(n^{1/4})$ in Theorem 3.4 we immediately have:

Corollary 3.3. *Under the assumptions of Theorem 3.4, with $\hat{\omega}_2$ defined above, there exists a constant $c>0$ large enough, such that*

$$E\left|\hat{\omega}_2-\omega\right|=\mathrm{O}(\frac{1}{n^{1/4}}\vee\frac{1}{p^{1/2}}).$$

Next we relax Assumption A6 to Assumption A2 and prove the theorem.

Theorem 3.5. *Suppose* $X_1, ..., X_n$ *are i.i.d. mean zero random vectors such that* $0 < \sup_{i\in\{1,2,...\}} EX_{i1}^4/\sigma_{ii}^2 < \infty$, *where* X_{i1} *is the* i*th element of* $X_1 = (X_{11}, ..., X_{p1})'$. *Under Assumptions A2 and A5, with* $\hat{\omega}_2$ *defined above, there exists a constant* $c > 0$ *large enough, such that*

$$\begin{aligned} & E\left|\hat{\omega}_2 - \omega\right| \\ \leq \; & \frac{cq}{\sqrt{n}} + \frac{4\sqrt{2c(q+1)}}{3n^{1/4}} + \frac{8}{\sqrt{p}} + O(\frac{1}{p^{2-\nu/2}}) + \frac{2\pi(1-\omega)}{q+1} \sup_{\rho\in(-1,1)} h(\rho). \end{aligned}$$

Particularly, when $X_1, ..., X_n$ *are i.i.d. mean zero multivariate normal random vectors, we have,*

$$E\left|\hat{\omega}_2 - \omega\right| \leq \frac{q}{\sqrt{n}} + \frac{4\sqrt{2(q+1)}}{3n^{1/4}} + \frac{8}{\sqrt{p}} + O(\frac{1}{p^{2-\nu/2}}) + \frac{2\pi(1-\omega)}{q+1} \sup_{\rho\in(-1,1)} h(\rho).$$

Proof. Similar to the proof of Theorem 3.4, write $N = p(p-1)/2$, we have:

$$\begin{aligned} & E|N^{-1} \sum_{1\leq j<i\leq p} (e^{-\mathbf{i}kr_{ij}} - Ee^{-\mathbf{i}k\rho_{ij}})|^2 \\ = \; & N^{-2} \sum_{1\leq j<i\leq p} \sum_{1\leq t<s\leq p} E(e^{-\mathbf{i}kr_{ij}} - Ee^{-\mathbf{i}k\rho_{ij}})(e^{\mathbf{i}kr_{st}} - Ee^{\mathbf{i}k\rho_{st}}) \\ = \; & N^{-2} \sum_{1\leq j<i\leq p} \sum_{1\leq t<s\leq p} E[(e^{-\mathbf{i}kr_{ij}} - e^{-\mathbf{i}k\rho_{ij}})(e^{\mathbf{i}kr_{st}} - e^{\mathbf{i}k\rho_{st}}) \\ & +(e^{-\mathbf{i}kr_{ij}} - e^{-\mathbf{i}k\rho_{ij}})(e^{\mathbf{i}k\rho_{st}} - Ee^{\mathbf{i}k\rho_{st}}) + (e^{-\mathbf{i}k\rho_{ij}} - Ee^{-\mathbf{i}k\rho_{ij}})(e^{\mathbf{i}kr_{st}} - e^{\mathbf{i}k\rho_{st}}) \\ & +(e^{-\mathbf{i}k\rho_{ij}} - Ee^{-\mathbf{i}k\rho_{ij}})(e^{\mathbf{i}k\rho_{st}} - Ee^{\mathbf{i}k\rho_{st}})]. \end{aligned}$$

By Lemma 3.8, there exists a constant $c > 0$ depending on $\sup_{i\in\{1,2,..,\}} EX_{i1}^4$ only such that for any $1 \leq i, j \leq p$

$$E(r_{ij} - \rho_{ij})^2 \leq \frac{c^2}{n}.$$

Therefore, when $\{i,j\} \cap \{s,t\} = \emptyset$,

$$\begin{aligned}
& Re[E(e^{-\mathbf{i}kr_{ij}} - e^{-\mathbf{i}k\rho_{ij}})(e^{\mathbf{i}kr_{st}} - e^{\mathbf{i}k\rho_{st}})] \\
\leq & \sqrt{E|(e^{-\mathbf{i}kr_{ij}} - e^{-\mathbf{i}k\rho_{ij}})|^2 E|(e^{-\mathbf{i}kr_{st}} - e^{-\mathbf{i}k\rho_{st}})|^2} \\
\leq & \sqrt{E|e^{-\mathbf{i}k\rho_{ij}}(e^{-\mathbf{i}k(r_{ij}-\rho_{ij})} - 1)|^2 E|e^{-\mathbf{i}k\rho_{st}}(e^{-\mathbf{i}k(r_{st}-\rho_{st})} - 1)|^2} \\
= & \sqrt{E|e^{-\mathbf{i}k(r_{ij}-\rho_{ij})} - 1|^2 E|e^{-\mathbf{i}k(r_{st}-\rho_{st})} - 1|^2} \\
= & 2\sqrt{E(1-\cos k(r_{ij}-\rho_{ij}))E(1-\cos k(r_{st}-\rho_{st}))} \\
= & 4\sqrt{E\sin^2(k(r_{ij}-\rho_{ij})/2)E\sin^2(k(r_{st}-\rho_{st})/2)} \\
\leq & k^2\sqrt{E(r_{ij}-\rho_{ij})^2 E(r_{st}-\rho_{st})^2} \\
\leq & \frac{c^2k^2}{n}.
\end{aligned}$$

$$\begin{aligned}
& Re[E(e^{-\mathbf{i}kr_{ij}} - e^{-\mathbf{i}k\rho_{ij}})(e^{\mathbf{i}k\rho_{st}} - Ee^{\mathbf{i}k\rho_{st}})] \\
\leq & \left\{E|e^{-\mathbf{i}kr_{ij}} - e^{-\mathbf{i}k\rho_{ij}}|^2 E|e^{\mathbf{i}k\rho_{st}} - Ee^{\mathbf{i}k\rho_{st}}|^2\right\}^{\frac{1}{2}} \\
\leq & (E|e^{-\mathbf{i}kr_{ij}} - e^{-\mathbf{i}k\rho_{ij}}|^2)^{\frac{1}{2}} \\
\leq & (Ek^2(r_{ij}-\rho_{ij})^2)^{\frac{1}{2}} \\
= & \frac{ck}{\sqrt{n}}.
\end{aligned}$$

Similarly,

$$Re[E(e^{-\mathbf{i}k\rho_{ij}} - Ee^{-\mathbf{i}k\rho_{ij}})(e^{\mathbf{i}kr_{st}} - e^{\mathbf{i}k\rho_{st}})] \leq \frac{ck}{\sqrt{n}}.$$

Moreover,

$$Re[E(e^{-\mathbf{i}k\rho_{ij}} - Ee^{-\mathbf{i}k\rho_{ij}})(e^{\mathbf{i}k\rho_{st}} - Ee^{\mathbf{i}k\rho_{st}})] \leq 4\alpha(\rho_{ij}, \rho_{st}).$$

Therefore, by Assumption A2, we have,

$$\begin{aligned} & N^{-2} \sum_{1\le j<i\le p, 1\le t<s\le p, i,j,s,t: all\ distinct} E(e^{-\mathbf{i}kr_{ij}} - Ee^{-\mathbf{i}k\rho_{ij}})(e^{\mathbf{i}kr_{st}} - Ee^{\mathbf{i}k\rho_{st}}) \\ \le\ & \frac{k^2c^2}{n} + \frac{2ck}{\sqrt{n}} + O(\frac{1}{p^{4-\nu}}). \end{aligned}$$

When $\{i,j\} \cap \{s,t\} \neq \emptyset$

$$\begin{aligned} & Re(E(e^{-\mathbf{i}kr_{ij}} - Ee^{-\mathbf{i}k\rho_{ij}})(e^{\mathbf{i}kr_{st}} - Ee^{\mathbf{i}k\rho_{st}})) \\ \le\ & E|(e^{-\mathbf{i}kr_{ij}} - Ee^{-\mathbf{i}k\rho_{ij}})(e^{\mathbf{i}kr_{st}} - Ee^{\mathbf{i}k\rho_{st}})| \\ \le\ & 4. \end{aligned}$$

Therefore,

$$\begin{aligned} & E|N^{-1} \sum_{1\le j<i\le p} (e^{-\mathbf{i}kr_{ij}} - Ee^{-\mathbf{i}k\rho_{ij}})|^2 \\ \le\ & \frac{k^2c^2}{n} + \frac{2ck}{\sqrt{n}} + O(\frac{1}{p^{4-\nu}}) + \frac{4p(p-1)(p-\frac{3}{2})}{N^2} \\ <\ & \frac{k^2c^2}{n} + \frac{2ck}{\sqrt{n}} + O(\frac{1}{p^{4-\nu}}) + \frac{16}{p}. \end{aligned}$$

By Holder's inequality we have:

$$\begin{aligned} & \sum_{k=1}^{q} E|(\hat{M}_q - M_q)_{q+1,k}| \\ \le\ & \sum_{k=1}^{q} (E|N^{-1} \sum_{1\le j<i\le p} (e^{-\mathbf{i}kr_{ij}} - Ee^{-\mathbf{i}k\rho_{ij}})|^2)^{\frac{1}{2}} \\ <\ & \sum_{k=1}^{q} \left[\frac{kc}{\sqrt{n}} + \frac{\sqrt{2ck}}{n^{1/4}} + \frac{4}{\sqrt{p}} + O(\frac{1}{p^{2-\nu/2}}) \right] \end{aligned}$$

$$\leq \quad \frac{cq(q+1)}{2\sqrt{n}} + \frac{2\sqrt{2c}(q+1)^{3/2}}{3n^{1/4}} + \frac{4q}{\sqrt{p}} + O(\frac{q}{p^{2-\nu/2}}).$$

Consequently, by Proposition 3.1, we have

$$\begin{aligned} & E\,|\hat{\omega}_2 - \omega| \\ \leq \quad & \frac{cq}{\sqrt{n}} + \frac{4\sqrt{2c(q+1)}}{3n^{1/4}} + \frac{8}{\sqrt{p}} + O(\frac{1}{p^{2-\nu/2}}) + \frac{2\pi(1-\omega)}{q+1} \sup_{\rho\in(-1,1)} h(\rho). \end{aligned}$$

Particularly, when $X_1, ..., X_n$ are i.i.d. mean zero multivariate normal random vectors, same as the proof of Theorem 3.4, by using the following inequality,

$$E(r-\rho)^2 \leq \frac{1}{n} + O(\frac{1}{n^2}),$$

we have:

$$\begin{aligned} & E\,|\hat{\omega}_2 - \omega| \\ \leq \quad & \frac{2}{q+1}\sum_{k=1}^{q} E|(\hat{M}_q - M_q)_{q+1,k}| + \frac{2\pi(1-\omega)}{q+1} \sup_{\rho\in(-1,1)} h(\rho) \\ \leq \quad & \frac{2}{q+1}\sum_{k=1}^{q} (E|N^{-1}\sum_{1\leq j<i\leq p}(e^{-\mathbf{i}kr_{ij}} - Ee^{-\mathbf{i}k\rho_{ij}})|^2)^{\frac{1}{2}} + \frac{2\pi(1-\omega)}{q+1} \sup_{\rho\in(-1,1)} h(\rho) \\ \leq \quad & \frac{2}{q+1}\sum_{k=1}^{q} \big[\frac{k}{\sqrt{n}} + \frac{\sqrt{2k}}{n^{1/4}} + \frac{4}{\sqrt{p}} + O(\frac{1}{p^{2-\nu/2}}) + O(kn^{-1} + k^{\frac{1}{2}}n^{-\frac{1}{2}})\big] \\ & + \frac{2\pi(1-\omega)}{q+1} \sup_{\rho\in(-1,1)} h(\rho) \\ \leq \quad & \frac{q}{\sqrt{n}} + \frac{4\sqrt{2(q+1)}}{3n^{1/4}} + \frac{8}{\sqrt{p}} + O(\frac{1}{p^{2-\nu/2}}) + O(qn^{-1} + q^{\frac{1}{2}}n^{-\frac{1}{2}}) \\ & + \frac{2\pi(1-\omega)}{q+1} \sup_{\theta\in(-1,1)} h(\theta). \end{aligned}$$

□

By letting $q = O(n^{1/6})$ in Theorem 3.5 we immediately have:

Corollary 3.4. *Under assumptions of Theorem 3.5, there exists a constant $c > 0$ large enough, such that*

$$E\left|\hat{\omega}_2 - \omega\right| = \mathrm{O}\left(\frac{1}{n^{1/6}} \vee \frac{1}{p^{1/2\wedge(2-\nu/2)}}\right).$$

Similar to Theorem 3.5 we have,

Theorem 3.6. *Suppose $X_1, ..., X_n$ are i.i.d. mean zero random vectors such that $0 < \sup_{i=1,2,...} EX_{i1}^4/\sigma_{ii}^2 < \infty$. Under Assumptions $A2^*, A5$ and with $\hat{\omega}_2$ defined above, there exists a constant $c > 0$ depending on EX_{11}^4/σ_{ii}^2 only, such that*

$$E\left|\hat{\omega}_2 - \omega\right| \to 0 \text{ as } p \wedge n \to \infty.$$

3.5 Numerical study

In practice, we need to determine the prior density $g(\rho)$ to derive $\hat{\omega}_1(g)$, and we need to choose a proper q for $\hat{\omega}_2$.

Assuming that $0 < \omega < 1$. Instead of estimating ω, we look at the problem of estimate $\omega' = \omega - n^{-a}$ for some positive constant a, and define $g'(\rho) = \frac{1-\omega}{1-\omega'}g(\rho) + \frac{1}{2(1-\omega')}n^{-a}$. We can see that g' satisfies Assumptions A3 and A4, and therefore, (3.12), (3.13), (3.14) and (3.15) are true for $a_{ij}(g')$. In addition, if

Assumptions A1 and A2 are satisfied, $\hat{\omega}_1(g')$ is consistent in estimating ω', or ω. Therefore, to some degree, Assumptions A3 and A4 can be relaxed to be $\sup_{\rho\in(-1,1)} g(\rho) < \infty$ if we replace our estimator by $\hat{\omega}_1(g')$. To compute $\hat{\omega}_1(g')$ numerically, we threshold $|r_{ij}|, 1 \leq i < j \leq p$ by $2\sqrt{\log n/n}$ and use the empirical cumulative distribution function(CDF), say $\hat{G}$, of those nonzero r_{ij} as the CDF of the prior of the continuous part of ρ, and construct an estimator $\hat{\omega}_1(d\hat{G})$. Notice that $\hat{G}$ will satisfy Assumptions A1-A4 with probability tending to one as long as $\sup_{\rho\in(-1,1)} g(\rho) < \infty$. Therefore, $\hat{\omega}_1(d\hat{G})$ is consistent in estimating ω. In the simulation, if $\frac{2}{p(p-1)} \sum_{1\leq i<j\leq p} I_{\{|r_{ij}|>2\sqrt{\log n/n}\}}$ is very small, we set $\hat{\omega}_1 = 1$.

For $\hat{\omega}_2$, we choose q in the following way:

Step 1. Compute estimators for $q = 1, ..., [2\sqrt{n}]$, write them as $\hat{\omega}_2(1)$,..., $\hat{\omega}_2([2\sqrt{n}])$.

Step 2. For each $\hat{\omega}_2(q)$, denote $N_q = [\hat{\omega}_2(q)p(p-1)/2]$, $q = 1, ..., [2\sqrt{n}]$. Let $r_{(1)} \leq ... \leq r_{(p(p-1)/2)}$ be the order statistics of $|r_{ij}|, 1 \leq i < j \leq p$ and define $\hat{\Lambda}_q = \{(i,j) : |r_{ij}| \leq r_{(N_q)}, 1 \leq i < j \leq p\}$. Let F_q be the empirical CDF of $\{r_{ij} : (i,j) \in \hat{\Lambda}_q, 1 \leq i < j \leq p\}$, $q = 1, ..., [2\sqrt{n}]$. Define the Kolmogorov-Smirnov distance between F_q and $F(x)$ as:

$$D_q = \sup_{0<x<1} |F_q(x) - F(x)|,$$

where $F(x)$ is the CDF of r_{ij} given $\rho = 0$.

Step 3. Choose q such that D_q is the minimum among $D_1, ..., D_{[2\sqrt{n}]}$.

For the multivariate normal case, F is given by:

$$F(x) = \int_{-1}^{x} \frac{\Gamma[\frac{1}{2}(n-1)]}{\Gamma[\frac{1}{2}(n-2)]\sqrt{\pi}}(1-x^2)^{\frac{1}{2}(n-4)}.$$

In this simulation, we consider $n = 100$ and $p = 50, 100, 200$ for different types of covariance matrices for the multivariate normal case over 100 replications. For a given estimate $\hat{\omega}$ of ω, we threshold the $[p(p-1)/2\hat{\omega}]$ smallest (in term of absolute value) sample correlation coefficients among all the $p(p-1)/2$ different off-diagonal elements of the sample correlation matrix to be zero, and denote this estimator to be $T(R) = (t_{ij})_{p\times p}$. We compare our estimators to $\hat{\omega}_{cv}$ and $\hat{\omega}_{acv}$ representing the estimator computed base on cross validation and adaptive cross validation under Frobenius norm correspondingly. Similar to what we did in Section 2.7, we also compute the following estimator in this simulation:

$$\hat{\omega}_3 = \begin{cases} 1, & \text{if } \{\lambda_1(\hat{M}_q) - 1\}/q > 1, \\ \{\lambda_1(\hat{M}_q) - 1\}/q, & \text{if } 0 \leq \{\lambda_1(\hat{M}_q) - 1\}/q \leq 1, \\ 0, & \text{if } \{\lambda_1(\hat{M}_q) - 1\}/q < 0, \end{cases} \tag{3.12}$$

Same as what we have discussed in Section 2.7, $|\hat{\omega}_2(q) - \hat{\omega}_3(q)| \leq \frac{1}{q}$, therefore similar results in Section 3.4 can be obtained for $\hat{\omega}_3$.

Mean and its standard deviation of the following quantities over 100 replications are computed and compared:

1)Error1=$\sum_{1\le i<j\le p} I_{\{\rho_{ij}=0,t_{ij}\neq 0\}}$, i.e., it counts the number of times of estimating zero to be nonzero.

2)Error2=$\sum_{1\le i<j\le p} I_{\{\rho_{ij}\neq 0,t_{ij}=0\}}$, i.e., it counts the number of times of estimating nonzero to be zero.

3)l1-loss: $|\text{estimator} - \omega|$

Let Σ be the population covariance matrix.

Model 1 $\Sigma = (\sigma_{ij})_{1\le i,j\le p}$, where $\sigma_{ij} = \sigma$ for any $1 \le i,j \le p/2, i \neq j$, $\sigma_{ii} = 1$, $i = 1,...p$ and $\sigma_{ij} = 0$ otherwise. We set $\sigma = 0.2, 0.5$.

Model 2 $\Sigma = TT'$ where $T = (t_{ij})_{p\times p}$ is a lower triangular matrix with $t_{ii} = 0.01, i = 1,...,p$ and $t_{ij} = U(0,1) \times Ber(0.05), 1 \le j < i \le p$. Here $U(0,1)$ representing a random variable uniformly distributed in $(0,1)$ and $Ber(0.05)$ is a Bernoulli random variable which takes value 1 with probability 0.05 and 0 with probability 0.95. This way of generating the population covariance matrix ensures positive definiteness and introduces zeros in Σ. Furthermore, the nonzero elements in the off-diagonal of the population correlation matrix will be able to cover values from 0 to 1.

Model 3 We generate Σ based on Lemma 3 of Wong, Carter and Kohn (2003). They derived the marginal distribution for the correlation coefficients under their model (adopting their notation):

$$p(dC_{ij}|C_{\{-ij\}}) = I_{\{|C_{ij}-a|<b\sqrt{c}\}} \frac{I_{\{C_{ij}=0\}} + (dC_{ij})h(J_{\{-ij\}})}{I_{\{|a|<b\sqrt{c}\}} + 2b\sqrt{c}h(J_{\{-ij\}})}.$$

where $C = (C_{ij})_{p\times p}$ is a correlation matrix and $C_{-\{ij\}} = \{C_{st}, 1 \leq t < s \leq p, (s,t) \neq (i,j)\}$.

When $i = p$ and $j = p-1$, let $C = R'R$ be the Cholesky decomposition of C with the matrix R upper triangular. a, b, c are defined as:

$$a = \sum_{j=1}^{p-2} R_{j,p-1}R_{j,p}, \quad b = R_{p-1,p-1}, \quad c = C_{pp} - \sum_{j=1}^{p-2} R_{j,p}^2;$$

For other values of i, j, we can always permute the indices i, j with $p, p-1$. From Lemma 1 of Wong, Carter and Kohn (2003), we observe that to ensure the positive definiteness of matrix C, C_{ij} can only be chosen within the interval: $(-b\sqrt{c} + a, b\sqrt{c} + a)$. Thus Lemma 3 in their paper to some degree provides a way of adapting C_{ij} to be 0 with some positive probability if $0 \in (-b\sqrt{c}+a, b\sqrt{c}+a)$ and $C_{ij} \sim U(-b\sqrt{c} + a, b\sqrt{c} + a)$ otherwise. We modify the marginal distribution of C_{ij} to be:

$$p(dC_{ij}|C_{\{-ij\}}) = I_{\{|C_{ij}-a|<b\sqrt{c}\}} \frac{I_{\{C_{ij}=0\}} + (dC_{ij})H}{I_{\{|a|<b\sqrt{c}\}} + 2b\sqrt{c}H}.$$

where H is a constant. We generate C using Gibbs sampling for:

Model 3.1$H = 0.8$ and $p = 50$. We run the chain for 10000 times with initial value: $C^{(0)} = I_p$. Write the resulting samples of C as: $C^{(1)}, ..., C^{(10000)}$. Let

$\Sigma = C^{(10000)}$. The corresponding proportion of zeros in Σ in our simulation is: $\omega = 0.7428571$.

Model 3.2$H = 0.8$ and $p = 100$. We run the chain for 10000 times with initial value: $C^{(0)} = (C_{ij})_{p\times p}$, where $C_{ij} = 0.9^{|i-j|}, 1 \leq i, j \leq p$. Let $\Sigma = C^{(10000)}$. The corresponding proportion of zeros in Σ in our simulation is: $\omega = 0.77111111$.

Model 3.3$H = 0.8$ and $p = 200$. We run the chain for 10000 times with initial value: $C^{(0)} = (C_{ij})_{p\times p}$, where $C_{ij} = 0.9^{|i-j|}, 1 \leq i, j \leq p$. Let $\Sigma = C^{(10000)}$. The corresponding proportion of zeros in Σ in our simulation is: $\omega = 0.8201005$.

For the l1-loss, under Model 1 $\sigma = 0.2$ case and Model 2, both $\hat{\omega}_1$ and $\hat{\omega}_2$ outperform $\hat{\omega}_{cv}$ and $\hat{\omega}_{acv}$. In fact, from Tables 3.1-3.3, we can see that cross validation methods tends to over threshold when $\sigma = 0.2$, which is relatively small. Under Model 1, we can also see that the l1-loss of $\hat{\omega}_2$ is slightly smaller when p is larger. In all cases, $\hat{\omega}_1$ has smaller l1-loss than $\hat{\omega}_{cv}$ and $\hat{\omega}_{acv}$. $\hat{\omega}_2$ and $\hat{\omega}_3$ are very close as expected. There are no significant differences between these two estimators. When the nonzero σ_{ij}'s are relatively small, $\hat{\omega}_2$ outperforms $\hat{\omega}_1$ $\hat{\omega}_{cv}$ and $\hat{\omega}_{acv}$, while when the nonzero σ's are relatively far away from 0, $\hat{\omega}_2$ has a larger l1-loss due to bias.

$p=50$	$\hat{\omega}_1$	$\hat{\omega}_2$	$\hat{\omega}_3$	$\hat{\omega}_{cv}$	$\hat{\omega}_{acv}$
l1-loss(sd)	0.062(0.003)	0.033(0.003)	0.033(0.003)	0.158(0.006)	0.217(0.007)
Error1(sd)	37.5(0.9)	90.3(2.7)	89.1(2.8)	21.9(3.5)	16.3(6.1)
Error2(sd)	114.1(3.1)	104.5(3.4)	105.0(3.4)	203.3(6.9)	262.8(8.1)

Table 3.1 Summary of simulation results over 100 replications under Model 1 when $p=50$, ω=0.755102 and $\sigma=0.2$.

$p=100$	$\hat{\omega}_1$	$\hat{\omega}_2$	$\hat{\omega}_3$	$\hat{\omega}_{cv}$	$\hat{\omega}_{acv}$
l1-loss(sd)	0.068(0.002)	0.025(0.002)	0.025(0.002)	0.163(0.006)	0.218(0.008)
Error1(sd)	147.8(2.5)	352.5(7.7)	354.2(7.7)	66.3(8.8)	84.6(30.1)
Error2(sd)	483.2(11.3)	443.3(9.9)	443.4(10.4)	862.7(25.4)	1058.8(35.3)

Table 3.2 Summary of simulation results over 100 replications under Model 1 when $p=100$, ω=0.7525253 and $\sigma=0.2$.

$p=200$	$\hat{\omega}_1$	$\hat{\omega}_2$	$\hat{\omega}_3$	$\hat{\omega}_{cv}$	$\hat{\omega}_{acv}$
l1-loss(sd)	0.073(0.002)	0.025(0.002)	0.023(0.002)	0.174(0.006)	0.219(0.008)
Error1(sd)	595.4(4.5)	1415.8(26.3)	1452.9(28.1)	217.6(30.9)	125.9(34.0)
Error2(sd)	2049.4(48.0)	1886.2(40.7)	1862.2(39.9)	3634.1(99.0)	4477.2(118.2)

Table 3.3 Summary of simulation results over 100 replications under Model 1 when $p=200$, ω=0.7512563 and $\sigma=0.2$.

$p = 50$	$\hat{\omega}_1$	$\hat{\omega}_2$	$\hat{\omega}_3$	$\hat{\omega}_{cv}$	$\hat{\omega}_{acv}$
l1-loss(sd)	0.006(0.001)	0.025(0.002)	0.026(0.002)	0.010(0.001)	0.015(0.002)
Error1(sd)	8.0(0.6)	21.4(3.1)	21.9(3.1)	13.2(1.3)	19.1(2.1)
Error2(sd)	0.3(0.1)	10.8(1.5)	10.3(1.7)	1.7(0.4)	1.0(0.2)

Table 3.4 Summary of simulation results over 100 replications under Model 1 when $p = 50$, ω=0.755102 and $\sigma = 0.5$.

$p = 100$	$\hat{\omega}_1$	$\hat{\omega}_2$	$\hat{\omega}_3$	$\hat{\omega}_{cv}$	$\hat{\omega}_{acv}$
l1-loss(sd)	0.006(<0.001)	0.019(0.001)	0.020(0.002)	0.008(0.001)	0.015(0.002)
Error1(sd)	32.6(2.0)	64.4(8.0)	72.0(9.1)	47.2(3.9)	79.6(7.8)
Error2(sd)	0.8(0.2)	30.8(4.8)	28.2(4.8)	8.5(1.5)	4.0(0.7)

Table 3.5 Summary of simulation results over 100 replications under Model 1 when $p = 100$, ω=0.7525253 and $\sigma = 0.5$.

$p = 200$	$\hat{\omega}_1$	$\hat{\omega}_2$	$\hat{\omega}_3$	$\hat{\omega}_{cv}$	$\hat{\omega}_{acv}$
l1-loss(sd)	0.006(<0.001)	0.011(0.001)	0.012(0.001)	0.008(0.001)	0.009(0.001)
Error1(sd)	122.0(6.5)	160.7(17.7)	194.2(24.0)	182.0(15.5)	201.6(15.7)
Error2(sd)	2.3(0.4)	68.4(12.9)	57.6(9.6)	26.2(4.3)	14.4(2.1)

Table 3.6 Summary of simulation results over 100 replications under Model 1 when $p = 200$, ω=0.7512563 and $\sigma = 0.5$.

$p = 50$	$\hat{\omega}_1$	$\hat{\omega}_2$	$\hat{\omega}_3$	$\hat{\omega}_{cv}$	$\hat{\omega}_{acv}$
l1-loss(sd)	0.020(0.001)	0.028(0.002)	0.028(0.002)	0.054(0.003)	0.028(0.002)
Error1(sd)	12.5(0.8)	26.4(3.2)	25.6(3.1)	1.4(0.2)	14.9(3.5)
Error2(sd)	35.8(0.3)	41.1(1.2)	40.8(1.1)	67.5(0.3)	39.3(0.5)

Table 3.7 Summary of simulation results over 100 replications under Model 2 when $p = 50$, ω=0.9306122.

$p = 100$	$\hat{\omega}_1$	$\hat{\omega}_2$	$\hat{\omega}_3$	$\hat{\omega}_{cv}$	$\hat{\omega}_{acv}$
l1-loss(sd)	0.048(0.001)	0.047(0.002)	0.048(0.002)	0.073(0.001)	0.061(0.001)
Error1(sd)	68.9(2.0)	88.3(7.0)	87.3(7.3)	46.9(1.8)	38.6(3.5)
Error2(sd)	304.1(0.9)	330.6(3.2)	321.8(3.3)	406.8(1.2)	341.4(2.2)

Table 3.8 Summary of simulation results over 100 replications under Model 2 when $p = 100$, ω=0.8911111.

$p = 200$	$\hat{\omega}_1$	$\hat{\omega}_2$	$\hat{\omega}_3$	$\hat{\omega}_{cv}$	$\hat{\omega}_{acv}$
l1-loss(sd)	0.109(<0.001)	0.099(0.002)	0.099(0.002)	0.166(0.001)	0.158(0.001)
Error1(sd)	382.8(5.6)	619.9(25.7)	625.0(26.9)	50.9(8.8)	35.9(4.3)
Error2(sd)	2548.5(4.4)	2592.4(13.2)	2593(13.6)	3341.8(34.6)	3181.1(14.9)

Table 3.9 Summary of simulation results over 100 replications under Model 2 when $p = 200$, ω=0.8125628.

$p = 50$	$\hat{\omega}_1$	$\hat{\omega}_2$	$\hat{\omega}_3$	$\hat{\omega}_{cv}$	$\hat{\omega}_{acv}$
l1-loss(sd)	0.175(0.001)	0.138(0.002)	0.138(0.002)	0.195(0.002)	0.203(0.002)
Error1(sd)	27.1(0.7)	36.1(1.4)	35.6(1.4)	7.8(0.7)	7.6(1.1)
Error2(sd)	241.4(0.5)	204.7(1.2)	25.0(1.2)	246.9(1.5)	256.4(2.0)

Table 3.10 Summary of simulation results over 100 replications under Model 3.1, where $\omega = 0.7428571$.

$p = 100$	$\hat{\omega}_1$	$\hat{\omega}_2$	$\hat{\omega}_3$	$\hat{\omega}_{cv}$	$\hat{\omega}_{acv}$
l1-loss(sd)	0.166(<0.001)	0.161(0.001)	0.162(0.001)	0.205(<0.001)	0.227(<0.001)
Error1(sd)	141.7(1.4)	109.2(2.8)	106.5(2.8)	16.5(0.6)	0.6(0.2)
Error2(sd)	962.2(1.0)	906.9(2.3)	908.6(2.2)	1030.4(1.7)	1125.9(1.7)

Table 3.11 Summary of simulation results over 100 replications under Model 3.2, where $\omega = 0.77111111$.

$p = 200$	$\hat{\omega}_1$	$\hat{\omega}_2$	$\hat{\omega}_3$	$\hat{\omega}_{cv}$	$\hat{\omega}_{acv}$
l1-loss(sd)	0.126(<0.001)	0.146(<0.001)	0.146(0.001)	0.168<0.001)	0.180(<0.001)
Error1(sd)	652.6(4.3)	289.3(4.9)	285.9(2.8)	61.0(1.8)	0.3(0.1)
Error2(sd)	3150.8(1.8)	3188.9(3.2)	3193.6(2.2)	3408.1(2.7)	3577.8(0.8)

Table 3.12 Summary of simulation results over 100 replications under Model 3.3, where $\omega = 0.8201005$.

CHAPTER 4

Conclusion

This study established consistent parameter estimation for the sparsity of a sparse signal sequence and the sparsity of a sparse covariance matrix.

In Chapter 2, we modeled the sparse signal sequence by (2.1) and (2.2) and proposed a method-of-moments estimator for the sparsity parameter ω_0. Particularly, different from most of the literature, ω_0 is assumed to be any number in $(0, 1]$ and the noise is assumed to be strong mixing and possibly non-Gaussian. To evaluate our estimator, upper bounds of the expected L1 loss of our estimator were derived. In addition, when the noise is assumed to be known, we derived lower bounds for the minimax risk of estimating ω_0. By comparing the upper and lower bounds, we concluded that our estimator achieves the optimal minimax convergence rate when the density of the noise is *supersmooth*. Simulation studies showed that our estimator performs well for different values of ω_0 and different types of noise distributions. In finite sample simulations, when the true ω_0 is large, meaning the signal sequence is very sparse, our estimator loses a bit to the empirical Bayes estimators. This might be due to the fact that the bias of our estimator is of order q^{-1}, which can be seen from Proposition 2.1.

In Chapter 3, we estimated the sparsity of a sparse covariance matrix or correlation matrix. We proposed two estimators: (1) $\hat{\omega}_1$, an empirical Bayes estimator; (2) $\hat{\omega}_2$, a method-of-moments estimator. $\hat{\omega}_1$ is derived under Gaussian assumption while $\hat{\omega}_2$ is more general. Consistency of these two estimators was proved. Simulation studies were carried out with a comparison to $\hat{\omega}_{cv}$ and $\hat{\omega}_{acv}$, the thresholding estimators derived from cross validation and adaptive cross validation methods respectively. Our estimators performed well in the simulations we conducted. More specifically, under the models we studied in the simulations, when the non-zero elements in the population correlation matrix were small, $\hat{\omega}_2$ outperformed $\hat{\omega}_1, \hat{\omega}_{cv}$ and $\hat{\omega}_{acv}$. When most of the non-zero elements of the population correlation matrix were large, $\hat{\omega}_1, \hat{\omega}_{cv}$ and $\hat{\omega}_{acv}$ performed well since the non-zero elements were relatively far away from 0. As for $\hat{\omega}_2$, it was still able to estimate ω well but lost a bit to other estimators. This might be due to the fact that the bias of $\hat{\omega}_2$ is of order

q^{-1}, which can be seen from Theorem 3.3. In addition, $\hat{\omega}_1$ outperformed the cross validation and adaptive cross validation estimators in all the cases we simulated.

The following are two open problems for future work:

1. How to estimate the signal sequence when the sparsity of the sequence is known or can be well estimated?

2. How to estimate the population covariance matrix or population correlation matrix when the sparsity of the matrix is known or can be well estimated?

Bibliography

[1] Abramovich, F. and Benjamini, Y. (1995). Thresholding of wavelet coefficients as multiple hypotheses testing procedure. *In Wavelets and Statistics, Lecture Notes in Statistics* **103**, 5-14. Springer, New York.

[2] Anderson, T.W. (2003). *An Introduction to Multivariate Statistical Analysis,* 3rd ed. Wiley Series in Probability and Statistics. Wiley, New York.

[3] Bai, Z.D. and Silverman, J.W. (2010). *Spectral Analysis of Large Dimensional Random Matrices,* 2nd ed. Springer Series in Statistics. Springer, New York.

[4] Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B* **57**, 289-300.

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

[6] Bennett, G. (1962). Inequalities for the sum of independent random variables. *J. Am. Statist. Assoc.* **48**, 33-45.

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

[8] Bickel, P.J. and Levina, E. (2008b). Covariance regularization by thresholding. *Ann. Statist.* **48**, 2577-2604.

[9] Bustoz, J. and Ismail, M. (1986). On Gamma function inequalities. *Math. Comput.* **47**, 659-667.

[10] Cai, T. and Liu, W. (2011). Adaptive thresholding for sparse covariance matrix estimation. *J. Am. Statist. Assoc.* **106**, 672-684.

[11] Chung, P.J., Bohme, J.F., Mecklenbrauker, C.F. and Hero, A.O. (2007). Detection of the number of signals using the Benjamini-Hochberg procedure. *IEEE Trans. Signal Process.* **55**, 2497-2508.

[12] Dey, D.K. and Srinivasan, C. (1986). Estimation of a covariance matrix under Stein's loss. *Ann. Statist.* **13**, 1581-1591.

[13] Donoho, D.L. and Jin, J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *Ann. Statist.* **32**, 962-994.

[14] Donoho, D.L. and Johnstone, I.M. (1995). Adapting to unknown smoothness via wavelet shrinkage. *J. Am. Statist. Assoc.* **90**, 1200-1224.

[15] Donoho, D.L. and Liu, R.C. (1991). Geometrizing rates of convergence, II. *Ann. Statist.* **19**, 633-667.

[16] Doukhan, P. (1994). *Mixing: Properties and Examples, Lecture Notes in Statistics* **85**. Springer, New York.

[17] El Karoui, N. (2008). Operator norm consistent estimation of large-dimensional sparse covariance matrices. *Ann. Statist.* **48**, 2717-2756.

[18] Fan, J. (1991). On the optimal rates of convergence of nonparametric deconvolution problems. *Ann. Statist.* **19**, 1257-1272.

[19] Friedman, J., Hastie, T. and Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical LASSO. *Biostatistics* **9**, 432-441.

[20] Grafakos, L. (2008). *Classical Fourier Analysis,* 2nd ed. Springer, New York.

[21] Haff, L.R. (1977). Minimax estimators for a multinormal precision matrix. *J. Multivariate Anal.* **7**, 374-385.

[22] Haff, L.R. (1991). The variational form of certain Bayes estimators. *Ann. Statist.* **19**, 1163-1190.

[23] Hall, P. and Jin, J. (2010). Innovated higher criticism for detecting sparse signals in correlated noise. *Ann. Statist.* **38**, 1686-1732.

[24] Horn, R.A. and Johnson, C.R. (1985). *Matrix Analysis.* Cambridge Univ. Press, Cambridge.

[25] Huang, J., Liu, N., Pourahmadi, M. and Liu, L. (2006). Covariance matrix selection and estimation via penalized normal likelihood. *Biometrika* **93**, 85-98.

[26] Johnstone, I.M. and Silverman, B.W. (2004). Needles and straw in haystacks: Empirical Bayes estimates of possibly sparse sequences. *Ann. Statist.* **32**, 1594-1649.

[27] Kass, R. (2001). Shrinkage estimators for covariance matrices. *Biometrics* **57**, 1173-1184.

[28] Kendall, M.G. (1960). The evergreen correlation coefficient. *Contributions to Probability and Statistics, Essays in Honor of Harold Hotelling*, 274-277. Stanford Univ. Press, Stanford.

[29] Lam, C. and Fan, J. (2009). Sparsistency and rates of convergence in large covariance matrix estimation. *Ann. Statist.* **37**, 4254-4278.

[30] Ledoit, O. and Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *J. Multivariate Anal.* **88**, 365-411.

[31] Lee, M., Shen, H., Burch, C. and Marron, J.S. (2010). Direct deconvolution density estimation of a mixture distribution motivated by mutation effects distribution. *J. Nonparametr. Statist.* **22**, 1-22.

[32] Li, M. and Loh, W.L. (2011). Estimating the number of neurons in multi-neuronal spike trains. *Ann. Appl. Statist.* **5**, 176-200.

[33] Lin, S.P. and Perlman, M. (1985). A Monte Carlo comparison of four estimators of a covariance matrix. *Multivariate Analysis-VI.* 411-426.

[34] Loh, W.L. (1988). *Estimating Covariance Matrices.* Stanford University Ph.D. thesis.

[35] Loh, W.L. (1991) Estimating covariance matrices. *Ann. Statist.* **19**, 283-296.

[36] Merlevede, F., Peligrad, M. and Rios, E. (2009). Bernstein inequality and moderate deviations under strong mixing conditions. In *High Dimensional Probability V: the Luminy vol.*, 273-292, *IMS Collections.* Inst. Math. Statist.

[37] Raykar, V. and Zhao, H. (2011). Empirical Bayesian thresholding for sparse signal using mixture loss functions. *Statist. Sinica* **21**, 449-474.

[38] Rothman, A.H., Bickel, P.J., Levina, E. and Zhu, J. (2008). Sparse permutation invariant covariance estimation. *Electron. J. Statist.* **2**, 494-515.

[39] Smith, M. and Kohn, R. (2002). Parsimonious covariance matrix estimation for longitudinal data. *J. Am. Statist. Assoc.* **97**, 1141-1153.

[40] Stein, C. (1975). *Rietz Lecture.* 39th annual meeting IMS. Atlanta, Georgia.

[41] Storey, J.D. (2002). A direct approach to false discovery rates. *J. Roy. Statist. Soc. Ser. B* **64**, 479-498.

[42] Wong, F., Carter, C.K. and Kohn, R. (2003). Efficient estimation of covariance selection models. *Biometrika* **90**, 809-830.

[43] Yang, R. and Berger, J.O. (1994). Estimation of a covariance matrix using the reference prior. *Ann. Statist.* **22**, 1195-1211.

[44] Yuan, M. and Lin, Y. (2007). Model selection and estimation in the Gaussian graphical model. *Biometrika* **94**, 19-35.