Electronic Journal of Statistics
Vol. 13 (2019) 4079–4101
ISSN: 1935-7524
https://doi.org/10.1214/19-EJS1621

Spectral analysis of high-dimensional time series*

Mark Fiecas

School of Public Health, Division of Biostatistics, University of Minnesota
e-mail: mfiecas@umn.edu

Chenlei Leng†

Department of Statistics, University of Warwick
e-mail: C.Leng@warwick.ac.uk

Weidong Liu

School of Mathematical Sciences, Shanghai Jiao Tong University
e-mail: weidongl@sjtu.edu.cn

Yi Yu

Department of Statistics, University of Warwick
e-mail: yi.yu.2@warwick.ac.uk

Abstract: A useful approach for analysing multiple time series is via characterising their spectral density matrix as the frequency domain analog of the covariance matrix. When the dimension of the time series is large compared to their length, regularisation based methods can overcome the curse of dimensionality, but the existing ones lack theoretical justification. This paper develops the first non-asymptotic result for characterising the difference between the sample and population versions of the spectral density matrix, allowing one to justify a range of high-dimensional models for analysing time series. As a concrete example, we apply this result to establish the convergence of the smoothed periodogram estimators and sparse estimators of the inverse of spectral density matrices, namely precision matrices. These results, novel in the frequency domain time series analysis, are corroborated by simulations and an analysis of the Google Flu Trends data.

Keywords and phrases: frequency domain time series, high dimension, functional dependency, smoothed periodogram, sparse precision matrix estimation.

Received November 2018.

*We are very grateful to the Editor and two referees for their constructive comments.
†Leng was supported by a Turing fellowship from the Alan Turing Institute under the EPSRC grant EP/N510129/1

Contents

1. Introduction

Spectral density matrices play a large role in characterising the second order properties of multivariate time series. The spectral density matrix is the frequency domain analog of the covariance matrix, and describes the variance in each dimension or the covariance between dimensions that can be attributed to oscillations in the data within certain frequencies. Just as how partial correlations between the dimensions can be extracted as a function of the inverse of a covariance matrix, conditional relationships attributable to variations in the oscillations of the data can be obtained from the inverse of the spectral density matrix [12]. Thus, it is necessary to obtain a positive-definite estimate of the spectral density matrix, but this can be challenging whenever the dimensionality of the time series is relatively large compared to the length of the time series.

The utility of estimating the inverse of the spectral density matrix is in estimating the frequency domain analog of the partial correlation structure, known as partial coherence, of the multivariate time series. This has been useful in many different contexts, including intensive care monitoring [19], brain connectivity in neuroimaging studies [36], signal processing [27], hydrology [24], and cardiology [34]. High-dimensional time series arise in various scientific contexts, such as in recordings of brain activity from numerous regions of the brain [36] or financial time series obtained from a large number of assets in a portfolio [14]. However, estimating the conditional relationships between the dimensions of the time series using partial coherence can be challenging in finite samples.

There have only been a few papers dedicated to developing rigorous theory in the context of a high-dimensional time series. For instance, [13] and [21] both developed methods to give sparse estimates of the parameters of a vector autoregressive (VAR) model. [2] studied the theoretical properties of regularised estimates of the parameters of a broad class of time series models. [38] considered the estimation of large covariance and precision matrices from high-dimensional

sub-Gaussian or heavier-tailed observations with slowly decaying temporal dependence. [11] studied a Dantzig-selector type regularized estimator for linear functionals of high-dimensional linear processes. These recent works, however, focused primarily on time series models in the time domain.

There have been developments in frequency domain methodologies, though these proposed methodologies lack theoretical justifications. For instance, [17], [18], and [37] developed variations of a shrinkage framework developed by [4] for data-driven ℓ_2-penalised estimation of the spectral density matrix, and illustrated their utility for connectivity analysis in neuroimaging data; [25] developed a graphical lasso approach for estimating a graphical model for high-dimensional time series data in the spectral domain; [1] utilised a dynamic factor model to study the network structure of high-dimensional financial time series. These methodological papers have deep roots in the application areas, where we are aware of the demand of estimating high-dimensional spectral density matrices and its inverse, where the latter is necessary in order to quantify the conditional dependency structure of the data. However, there remains a critical gap in theoretical investigations on frequency domain methodologies for estimating the inverse of high-dimensional spectral density matrices.

The aim of this paper is to study the theoretical behaviours of estimators of the spectral density matrix and its inverse in high dimension. We summarise the main contributions of this paper.

First, it is arguable that the most important ingredient in high-dimensional statistical inference, in contrast with classical ones, is the fixed-sample results. To be specific, in order to allow for high dimensions, a common practice is to exploit concentration inequalities, then to provide fixed-sample results to control the differences between the sample and the population versions, and finally to use union bound arguments to derive desirable results. To the best of our knowledge, this paper is the first to show such fixed-sample results on the error control of the smoothed periodogram matrices in Theorem 3.1. This is a challenging task, and the main difficulty in developing such methods comes from the fact that the text book results on frequency domain time series are limited to asymptotic results only [5, 6].

Second, once the fixed-sample results are established, a wide range of high-dimensional statistics methods are ready to be justified, including estimation, prediction and inference tools. In this paper, we use the sparse precision matrix estimation problem as an example, and demonstrate the theoretical (see Proposition 3.2) and numerical performances of applying the constrained ℓ_1-minimisation for inverse matrix estimation [CLIME, 8] to spectral analysis of time series data. We would like to mention that the possible applications of Theorem 3.1 are way beyond Proposition 3.2, while we use the sparse precision matrix estimation as an example.

The rest of this paper is organised as follows. In Section 2, we explain the methodology used in this paper. The theoretical results are collected in Section 3, including two main theorems. The technical details thereof can be found in the Appendix. In Section 4, we demonstrate the numerical performances of our proposed methods, via simulations and real data analysis.

2. Methodology

2.1. Framework and notation

In order to study the theoretical performances, we adopt the functional dependency framework [45]. Let $\boldsymbol{X}_t = (X_{t,1}, \ldots, X_{t,p})^\top \in \mathbb{R}^p$ be centred random vectors satisfying

$$\boldsymbol{X}_t = G(\ldots, \boldsymbol{e}_{t-1}, \boldsymbol{e}_t) =: G(\mathcal{F}_t), \tag{2.1}$$

where $\boldsymbol{e}_t$ are i.i.d. random vectors, $\mathcal{F}_t = (\ldots, \boldsymbol{e}_{t-1}, \boldsymbol{e}_t)$, and

$$G(\mathcal{F}_t) = (g_1(\mathcal{F}_t), \ldots, g_p(\mathcal{F}_t))^\top.$$

With this notation, we have $X_{t,i} = g_i(\mathcal{F}_t)$ for each $i \in \{1, \ldots, p\}$. Let $\tilde{\boldsymbol{e}}_0, \{\boldsymbol{e}_t, t \in \mathbb{Z}\}$ be i.i.d. random vectors. For $t = 1, 2, \ldots$, define $\tilde{\mathcal{F}}_t = (\ldots, \boldsymbol{e}_{-1}, \tilde{\boldsymbol{e}}_0, \boldsymbol{e}_1, \ldots, \boldsymbol{e}_t)$, i.e. replace $\boldsymbol{e}_0$ with $\tilde{\boldsymbol{e}}_0$ in $\mathcal{F}_t$. Define $X'_{t,i} = g_i(\tilde{\mathcal{F}}_t)$ and

$$\theta_{t,i} = \left(\mathbb{E}|X_{t,i} - X'_{t,i}|^2\right)^{1/2}, \tag{2.2}$$

which is used as a dependency measure. It has been pointed out in [45] that a large family of common time series models, including linear processes, Volterra series, nonlinear transforms of linear processes and some nonlinear time series models, can be characterised by imposing proper conditions on (2.2). As a price we have to pay for this generality, the functional dependency assumption does exclude the long-range memory models, which can be handled by the temporal dependency structure proposed by [38].

For rest of this paper, for any vector $\boldsymbol{v} = (v_1, \ldots, v_m)^\top \in \mathbb{C}^m$, let $\|\boldsymbol{v}\|_q := \left(\sum_{i=1}^m |v_i|^q\right)^{1/q}$ be the ℓ_q-norm of $\boldsymbol{v}$; for any matrix $A = (A_{ij})_{i,j=1}^p \in \mathbb{C}^{p\times p}$, let $\|A\|_w := \sup_{\boldsymbol{v}:\, \|\boldsymbol{v}\|_w \le 1} \|A\boldsymbol{v}\|_w$. We use the sparsity definition in [9] to characterise the sparsity of precision matrices, i.e. let the parameter space $\mathcal{G}_q(c_{n,p}, M_{n,p})$ be denoted by

$$\mathcal{G}_q(c_{n,p}, M_{n,p}) := \left\{ \begin{array}{c} \Theta = (\Theta_{ij})_{i,j=1}^p : \max_{j=1,\ldots,p} \sum_{i=1}^p |\Theta_{ij}|^q \le c_{n,p}, \\ \|\Theta\|_1 \le M_{n,p},\ \lambda_{\max}(\Theta)/\lambda_{\min}(\Theta) \le M_1, \end{array} \right\}, \tag{2.3}$$

where $0 \le q < 1$, $c_{n,p}$ and $M_{n,p}$ are potentially diverging as n and p grow.

2.2. The sparse inverse periodogram estimator

In the following sections, we define the spectral density matrix, introduce the estimators thereof, and propose a method to estimate the inverse of the high-dimensional spectral density matrix for any arbitrary frequency. We convert the time domain time series data into frequency domain using the discrete Fourier transform, which results in the data being a complex-valued vector. Motivated by the properties of the complex-valued normal distribution, we separate the

real and imaginary parts of the transformed data and double the dimension of the vectors. At each frequency point, we adopt a moving window and construct the estimator of the inverse of the periodogram, based on the CLIME estimator proposed in [8]. The detailed algorithm is in Algorithm 1. We will first state our algorithm, and explain the details regarding the smoothed periodogram and its inverse in Sections 2.3 and 2.4, respectively.

Algorithm 1 Sparse Inverse Periodogram Estimation.

procedure SIPE($\{\boldsymbol{X}_i \in \mathbb{R}^p\}_{i=1}^n$, h)
  for $j \in \{-\lfloor (n-1)/2 \rfloor, \ldots, \lfloor n/2 \rfloor\}$ **do**
    $\omega_j \leftarrow j/n$
    $\boldsymbol{d}(\omega_j) \leftarrow \sum_{t=1}^n \boldsymbol{X}_t \exp(-\imath 2\pi\omega_j t)$ ▷ $\boldsymbol{d}(\omega_j) \in \mathbb{C}^p$
  end for
  $D^{\mathrm{C}} \leftarrow \left(\boldsymbol{d}(\omega_{-\lfloor (n-1)/2 \rfloor}), \ldots, \boldsymbol{d}(\omega_{\lfloor n/2 \rfloor})\right)^\top$ ▷ $D^{\mathrm{C}} \in \mathbb{C}^{n\times p}$
  $D \leftarrow (\Re(D^{\mathrm{C}}), \Im(D^{\mathrm{C}}))$ ▷ $D \in \mathbb{R}^{n\times 2p}$
  for $j \in \{-\lfloor (n-1)/2 \rfloor, \ldots, \lfloor n/2 \rfloor\}$ **do**
    $\mathrm{ind} \leftarrow (j-h, j-h+1, \ldots, j+h) \mod (n+1)$
    $\begin{pmatrix} A_1 & B_1 \\ B_2 & A_2 \end{pmatrix} \leftarrow \mathrm{CLIME}(D_{\mathrm{ind}})$ ▷ $D_{\mathrm{ind}} \in \mathbb{R}^{|\mathrm{ind}|\times 2p}$, $A_1, A_2, B_1, B_2 \in \mathbb{R}^{p\times p}$
    $\Theta_j \leftarrow (A_1 + A_2)/2 + \imath(B_1 - B_2)/2$ ▷ $\Theta_j \in \mathbb{C}^{p\times p}$
  end for
  return $\{\Theta_i \in \mathbb{C}^{p\times p}\}_{i=1}^n$.
end procedure

In Algorithm 1, $\Re(\cdot)$ and $\Im(\cdot)$ denote the real and imaginary parts of an object, respectively, and preserve the same format of the object. In our case, the input $D^{\mathrm{C}} \in \mathbb{R}^{n\times p}$, and therefore $\Re(D^{\mathrm{C}}), \Im(D^{\mathrm{C}}) \in \mathbb{R}^{n\times p}$. As for the algorithm CLIME, see Section 2.4 and [8] for details.

2.3. Real-valued smoothed periodogram estimators

Let $\{\boldsymbol{X}_t\}_{t\in\mathbb{Z}}$ be a p-variate mean zero stationary real-valued time series with autocovariance matrix function $\Gamma(h) = \mathrm{Cov}(\boldsymbol{X}_t, \boldsymbol{X}_{t+h}) = \mathbb{E}(\boldsymbol{X}_t \boldsymbol{X}_{t+h}^\top)$, for $h \in \mathbb{Z}$. Under these conditions, $\{\boldsymbol{X}_t\}$ has a continuous spectral density matrix given by

$$f(\omega) = \sum_{h\in\mathbb{Z}} \Gamma(h) \exp(-\imath 2\pi\omega h), \quad \omega \in [-1/2, 1/2).$$

Given an interval of the whole time series, namely $\{\boldsymbol{X}_t\}_{t=1,\ldots,n}$, the periodogram defined at the Fourier frequencies $\{\omega_j = j/n, -\lfloor (n-1)/2 \rfloor \le j \le \lfloor n/2 \rfloor\}$ by $P_n(\omega_j) = n^{-1}\boldsymbol{d}(\omega_j)\boldsymbol{d}^*(\omega_j)$, where $\boldsymbol{d}(\omega_j) = \sum_{t=1}^n \boldsymbol{X}_t \exp(-\imath 2\pi\omega_j t)$, and for any complex-valued vector $\boldsymbol{v}$, $\boldsymbol{v}^*$ denotes $\overline{\boldsymbol{v}}^\top$, i.e. the conjugate transpose of $\boldsymbol{v}$.

When $p = 1$, it is known that $\mathbb{E}(P_n(\omega))$ converges uniformly to $f(\omega)$ on $[-1/2, 1/2]$ [e.g. 6, Proposition 10.3.1], but $P_n(\omega)$ does not converge in probability to $f(\omega)$ as $T \to \infty$ [e.g. 6, Theorem 10.3.2]. A common remedy is to use the smoothed periodogram, given by

$$\widetilde{f}_n(\omega_j) = \frac{1}{2M_n + 1} \sum_{|k|\le M_n} P_n(\omega_{j+k}).$$

When $p = 1$, it can be shown that if $M_n \to \infty$ and $M_n/n \to 0$ as $n \to \infty$, $\widetilde{f}_n(\omega_j)$ is a consistent estimator of $f(\omega_j)$.

When $p \to \infty$ as $T \to \infty$, we are interested in the conditional dependence structures of the pairs of coordinate, namely by defining $\Theta(\omega) = \big(f(\omega)\big)^{-1}$, our goal now is to provide a *sparse* estimator of $\Theta(\omega)$ with desirable large-sample properties. Note that both $f(\omega)$ and $\Theta(\omega)$ are complex-valued matrices. To make the following discussion easier, we first transform them into real-valued matrices.

For any $j = -\lfloor (n-1)/2, \ldots, \lfloor n/2 \rfloor$ and $\omega_j = j/n$, since

$$\begin{aligned}
\widetilde{f}_n(\omega_j) =& \frac{1}{(2M_n+1)n} \sum_{|k| \le -M_n} \boldsymbol{d}(\omega_{j+k}) \big(\boldsymbol{d}(\omega_{j+k})\big)^* \\
=& \frac{1}{(2M_n+1)n} \sum_{|k| \le M_n} \Bigg\{ \Bigg(\sum_{t=1}^{n} \boldsymbol{X}_t \cos(2\pi\omega_{j+k} t) \Bigg) \Bigg(\sum_{t=1}^{n} \boldsymbol{X}_t \cos(2\pi\omega_{j+k} t) \Bigg)^\top \\
&\qquad + \Bigg(\sum_{t=1}^{n} \boldsymbol{X}_t \sin(2\pi\omega_{j+k} t) \Bigg) \Bigg(\sum_{t=1}^{n} \boldsymbol{X}_t \sin(2\pi\omega_{j+k} t) \Bigg)^\top \Bigg\} \\
&+ \imath \frac{1}{(2M_n+1)n} \sum_{|k| \le M_n} \Bigg\{ \Bigg(\sum_{t=1}^{n} \boldsymbol{X}_t \cos(2\pi\omega_{j+k} t) \Bigg) \Bigg(\sum_{t=1}^{n} \boldsymbol{X}_t \sin(2\pi\omega_{j+k} t) \Bigg)^\top \\
&\qquad - \Bigg(\sum_{t=1}^{n} \boldsymbol{X}_t \sin(2\pi\omega_{j+k} t) \Bigg) \Bigg(\sum_{t=1}^{n} \boldsymbol{X}_t \cos(2\pi\omega_{j+k} t) \Bigg)^\top \Bigg\} \\
=:& A_j + \imath B_j,
\end{aligned}$$

it follows from Lemma A.1 in the Appendix, that $\Theta(\omega_j)$ has the form $\widetilde{A}_j + \imath \widetilde{B}_j$, where $\widetilde{A}_j$ and $\widetilde{B}_j$ satisfy

$$\begin{pmatrix} A_j & -B_j \\ B_j & A_j \end{pmatrix} \begin{pmatrix} \widetilde{A}_j & -\widetilde{B}_j \\ \widetilde{B}_j & \widetilde{A}_j \end{pmatrix} = I.$$

Therefore, our problem is transformed to finding the inverse of $\begin{pmatrix} A_j & -B_j \\ B_j & A_j \end{pmatrix}$.

Therefore, for any $j = -\lfloor (n-1)/2, \ldots, \lfloor n/2 \rfloor$ and $\omega_j = j/n$, instead of directly studying $\widetilde{f}_n(\omega_j)$, our targets are now

$$\Sigma_j := \begin{pmatrix} \mathrm{Re} f(\omega_j) & \mathrm{Im} f(\omega_j) \\ -\mathrm{Im} f(\omega_j) & \mathrm{Re} f(\omega_j) \end{pmatrix}$$

and sample version

$$\begin{aligned}
\widehat{\Sigma}_j = \frac{1}{(2M_n+1)n} \sum_{|k| \le M_n} \sum_{s=1}^{n} \sum_{\ell=1}^{n} & \\
\begin{pmatrix} \boldsymbol{X}_s \boldsymbol{X}_\ell^\top \cos(2\pi\omega_{j+k}(s-l)) & \boldsymbol{X}_s \boldsymbol{X}_\ell^\top \sin(2\pi\omega_{j+k}(s-l)) \\ -\boldsymbol{X}_s \boldsymbol{X}_\ell^\top \sin(2\pi\omega_{j+k}(s-l)) & \boldsymbol{X}_s \boldsymbol{X}_\ell^\top \cos(2\pi\omega_{j+k}(s-l)) \end{pmatrix}.
\end{aligned}$$

2.4. Penalised precision matrices at every frequency point

Now we have a sequence of expanded but real-valued smoothed periodogram matrices at every frequency point, i.e. $\{\widehat{\Sigma}_j, j = -\lfloor (n-1)/2 \rfloor, \ldots, \lfloor n/2 \rfloor\}$. As for each one, our goal is to obtain a sparse inverse matrix. In the last decade, a number of statistical methods have been proposed to achieve this goal, including graphical Lasso [e.g. 48], node-wise regression [e.g. 30], constrained ℓ_1-minimisation for inverse matrix estimation [CLIME, 8], adaptive CLIME[9] and the innovated scalable efficient estimation [15], among others.

In this paper, we do not intend to compare different sparse precision matrix estimation methods, but to apply the CLIME method for the sake of simplicity in technical details, and to provide with an example for consistent sparse precision matrix estimation in the high-dimensional frequency domain time series context. For details of the CLIME method, we refer readers to [8], which studies the inverse of the covariance matrices, and in which the sparse precision matrix estimators are obtained based on the sample covariance matrices of i.i.d. random vectors. For completeness, we include the definition of the estimators.

For each $j \in \{-\lfloor (n-1)/2 \rfloor, \ldots, \lfloor n/2 \rfloor\}$, let

$$\widehat{\Theta}_j = (\tilde{\Theta}_{j,kl}) = \underset{\|\widehat{\Sigma}_j \Theta_j - I\|_\infty \le \lambda,\, \Theta_j \in \mathbb{R}^{2p \times 2p}}{\arg\min} \|\Theta_j\|_1. \tag{2.4}$$

In practice, one can also symmetrise the estimator and obtain

$$\widetilde{\Theta}_j = (\tilde{\Theta}_{j,kl}),$$

where

$$\tilde{\Theta}_{j,kl} = \tilde{\Theta}_{j,lk} = \hat{\Theta}_{j,kl} \mathbb{1}\{\hat{\Theta}_{j,kl} \le \hat{\Theta}_{j,lk}\} + \hat{\Theta}_{j,lk} \mathbb{1}\{\hat{\Theta}_{j,lk} \le \tilde{\Theta}_{j,kl}\}.$$

3. Theory

In Theorem 3.1, we will provide fixed-sample results for the spectral density matrix of a high-dimensional time series, in the form of an entry-wise error control between the smoothed periodogram estimator and the spectral density matrix. This is a fundamental step in proving many different types of high-dimensional statistical problems. To theoretically justify the sparse precision matrix estimator we proposed in Section 2, but more importantly, to demonstrate the power of Theorem 3.1, in Proposition 3.2, we show the uniform consistency of the sequence of precision matrices $\{\widehat{\Theta}_j\}$.

As pointed out in Section 2.1, in order to provide the desired results, we are using the functional dependency framework described by (2.1) and (2.2). To further characterise the dependency, we introduce Assumption 1. This is also used in [11], and we refer interested readers there for examples.

Assumption 1. *Assume for some constant* $0 < \rho < 1$,

$$\max_{i=1,\ldots,p} \theta_{t,i} = O(\rho^t),$$

and for some constant $\kappa > 0$ *and* $C_0 > 0$,

$$\max_{1\le i\le p} \mathbb{E}(\exp\{\kappa|X_{0,i}|\}) \le C_0.$$

Note that the fixed-sample result holds for all dimensionality, but in order to achieve desirable consistency results, we need extra conditions on the dimensionality of the data, which is detailed in Assumption 2. Note that we can actually handle a super-polynomial rate of n for p, but in order to be specific, we assume p is of any polynomial rate of n as described in Assumption 2.

Assumption 2. *Assume:*

- *there exists constant* $c > 0$ *such that* $p \le cn^r$ *for some* $r > 0$*;*
- $M_n/T \to 0$*, and there exists a constant* $\delta > 0$ *such that* $M_n^{-1/2}(n/M_n)^\delta \to 0$.

Assumption 3 is only used to achieve the consistency of the sparse precision matrix estimators in Proposition 3.2. Under Assumption 2, Equation (3.1) holds even when the ℓ_1- and ℓ_q-norms of Θ_j, $j = 1, \ldots, n$, diverge, as n grows unbounded. Therefore, Assumption 3 is a reasonably weak condition.

Assumption 3. *Recall the parameter space* $\mathcal{G}_q(c_{n,p}, M_{n,p})$ *defined in* (2.3)*. Assume for* $q \in [0, 1)$ *the following holds:*

$$M_{n,p}^{1-q}\left(\frac{M_{n,p}M_n}{n} + \frac{M_{n,p}n^\delta}{M_n^{1/2+\delta}}\right)^{1-q} c_{n,p} = o(1). \tag{3.1}$$

Theorem 3.1 (Smoothed periodogram)**.** *Under Assumption 1, there exists a constant* $C_1 > 0$ *depending only on* κ *and* C_0 *such that for any* $\delta > 0$ *and* $H > 0$*, the following holds*

$$\mathbb{P}\Bigg\{\sup_{k\in\{-\lfloor (n-1)/2\rfloor,\ldots,\lfloor n/2\rfloor\}} \max_{i,j=1,\ldots,p} |\widetilde{f}_{ij,n}(\omega_k) - f_{ij}(\omega_k)| \\ > C_1 M_n/n + 8(n/M_n)^{1/2+\delta} n^{-1/2}\Bigg\} \le C_2 p^2 n^{-H}, \tag{3.2}$$

where $C_2 > 0$ *is a constant with respect to* n*, depending on* C_1 *and* H*.*

If we further assume Assumption 2, then we have

$$\sup_{k\in\{-\lfloor (n-1)/2\rfloor,\ldots,\lfloor n/2\rfloor\}} \max_{i,j=1,\ldots,p} |\widetilde{f}_{ij,n}(\omega_k) - f_{ij}(\omega_k)| = o_P(1).$$

The fixed-sample result in (3.2) holds for any choices of sample size n, dimensionality p and the smoothing window size M_n. It holds in the functional dependency framework detailed in Assumption 1, and provides an entry-wise error control of the smoothed periodogram and the spectral density matrix. We adopt a union bound argument to handle the dimensionality and to provide a uniform result across the sampled frequency points.

It is worth mentioning that the probability upper bound allows for any $H > 0$, which allows for the dimensionality diverges at any arbitrary polynomial rate as the sample size diverges. This is made explicit in Assumption 2.

The detailed proof of Theorem 3.1 is in the Appendix. Here, we briefly outline the sketch of the proof. We start with a fixed frequency point and a fixed entry in the matrix. In order to bound the errors between the smoothed periodogram matrix $\widetilde{f}$ and the spectral density matrix f, we introduce a series of instrumental quantities, including an m-dependent series using conditional expectations, its truncated version which is truncated in magnitude by $(\log(n))^2$, and a centred version by subtracting the unconditional expectations. The majority of the efforts are therefore dedicated to bound the differences of all these different quantities. Applying triangle inequality yields desirable results for a fixed frequency point and a fixed entry in the matrix. Finally, we apply a union bound argument to obtain (3.2).

Proposition 3.2. *Under Assumptions 1 and the parameter space defined in (2.3), for a constant $\delta > 0$, any $w \in [1, \infty]$ and*

$$\lambda \asymp \frac{M_{n,p} M_n}{n} + \frac{M_{n,p} n^{\delta}}{M_n^{1/2+\delta}},$$

we have for a sufficiently large constant $C_1 > 0$,

$$\mathbb{P}\left(\sup_{j \in \{-\lfloor (n-1)/2 \rfloor, \ldots, \lfloor n/2 \rfloor\}} \|\hat{\Theta}(\omega_j) - \Theta(\omega_j)\|_w \leq C_1 M_{n,p}^{1-q} \lambda^{1-q} c_{n,p} \right) \geq 1 - C_2 p^2 n^{-H}, \tag{3.3}$$

where $C_2 > 0$ is a constant with respect to n, depending on C_1 and H.

If we further assume Assumptions 2 and 3, then we have

$$\sup_{j \in \{-\lfloor (n-1)/2 \rfloor, \ldots, \lfloor n/2 \rfloor\}} \|\hat{\Theta}(\omega_j) - \Theta(\omega_j)\|_w = o_P(1).$$

Proposition 3.2 is an application of Theorem 3.1 on the sparse precision matrix estimation. The proof is in fact straightforward based on (3.2) and the proof techniques developed in [8]. Since it is built upon Theorem 3.1, we allow for the same flexibility that in (3.3), H is allowed to be any positive value, and therefore the dimensionality p is allowed to be of any arbitrary order of the sample size n. In practice, the window width M_n and the penalty level λ are chosen via data-driven methods, which will be elaborated in Section 4.

Remark 1. *For the interest of the periodogram itself, Theorem 3.1 can be interpreted as an error control of the smoothed periodogram estimators in terms of the entry-wise sup-norm. In addition, for general w-norm of matrices, one can obtain an error control of the inverse of $\hat{\Theta})(\omega_j)$ by exploiting Theorem 2.5 in [39]. In particular, when $w = 2$, it has been pointed out in [3] that*

$$\left\| \left\{ \hat{\Theta}(\omega_j) \right\}^{-1} - f(\omega_j) \right\|_2 \asymp \left\| \hat{\Theta}(\omega_j) - \Theta(\omega_j) \right\|_2.$$

4. Numerical results

4.1. Simulations

In this section, we verify our proposed methodology using simulated data. We consider multivariate time series having dimension $p = 10$ or 50 with sample size $n = 200$ or 400. These are challenging scenarios for spectral analysis because the amount of data available to estimate the spectral density matrix and its inverse is related to the smoothing span $2M_n + 1$ used to smooth the periodogram matrix, and not the length of the time series. In our simulations, we choose M_n using the generalised cross-validation (GCV) criterion developed by [33], and the penalty level in CLIME by the Stability Approach to Regularization Selection (StARS) method [28]. We construct the smoothed periodogram matrix $\tilde{f}_n(\omega)$ and calculate its inverse (whenever possible) and use these estimators in order to assess relative performance.

We investigated multiple scenarios in this study: we simulated from (1) a p-variate Gaussian white noise model, (2) a p-variate first-order vector autoregressive (VAR(1)) model, whose parameters we give below, (3) a p-variate VAR(1) model whose conditional dependence structure between the dimensions is driven by a sparse precision matrix of the innovations, and (4) a p-variate vector autoregressive moving average (VARMA) model, whose parameters we give below.

Setting (1) allows us to see how our methodology performs relative to the smoothed periodogram matrix in a very simple scenario where the spectral density matrix and its inverse do not change across frequencies, which allow us to evaluate relative performance only as a function of dimensionality. Setting (2) allows us to see how our methodology performs when the data exhibit some degree of autocorrelation and lagged cross-correlation. To construct the VAR(1) model, we set the $p \times p$ coefficient matrix to be a banded matrix such with diagonal entries set to be 0.5, and for the jth row, $j \in \{1, \ldots, p-2\}$, we set the $(j+1)$th column to be -0.3 and the $(j+2)$th column to be 0.2. We use the identity matrix as the covariance matrix for the innovations in the model. Setting (3) creates heterogeneity in the marginal variances, and hence, in the diagonal elements of the spectral density matrix, but truth has a sparse conditional dependence structure. In particular, we let the VAR(1) coefficient matrix be a diagonal matrix with entries randomly selected from the interval $(0.25, 0.75)$, and a random sign. The precision matrix for the innovations vector is sparse, with off-diagonal elements equal to 0 or 0.5 with probability 0.5. Setting (4) allows us to see how our methodology performs when the data generating mechanism has a relatively more complicated dependency structure. We borrow the parameter settings for the VARMA(2,2) model from another paper [44]. Namely, for the autoregressive part of the model, we use diagonal coefficient matrices such that the diagonal element of the k-th coefficient matrix is equal to $0.4/k$. For the moving average part of the model, the k-th coefficient matrix is block diagonal with 5×5 lower triangular blocks with value $0.3/k$ on the diagonals and $0.3/(2k)$ on the off-diagonal. We use the identity matrix as the covariance matrix for the innovations in the model. Thus, the dependencies between dimensions is null

between the blocks and is non-null within the blocks.

We evaluate performance in the following ways. First, we use the mean integrated squared error (MISE), defined by

$$\mathrm{MISE}\big(\{\widehat{\Theta}(\omega_j)\}_{j=1}^{n}, \{\Theta(\omega_j)\}_{j=1}^{n}\big) = \frac{2}{n}\sum_{j=1}^{n/2}\big\|\widehat{\Theta}(\omega_j) - \Theta(\omega_j)\big\|_*^2,$$

where $\{\omega_j\}_{j=1}^{n/2}$ denote the Fourier frequencies in the interval $(0, 0.5)$, $\|\cdot\|_*$ denotes the Frobenius norm of a matrix but discarding the diagonal entries, i.e. for a matrix $A = (A_{ij}) \in \mathbb{R}^{p\times p}$,

$$\|A\|_* = \sqrt{\sum_{i=1}^{p}\sum_{j\neq i} A_{ij}^2}.$$

The reason we are discarding the diagonal entries is that we are mainly interested in the off-diagonal entries, and the penalisations deployed in obtaining the sparse precision matrix estimators inevitably introduce bias, especially for the diagonal entries. If one would like a better estimator of the diagonal entries, one can adopt an optional second step updating the diagonal entries only by forcing the product of the smoothed periodogram matrix and the sparse precision matrix to be identity. Due to the lack of theoretical guarantees, we omit this optional step in this paper.

We compare our estimator (SIPE) to the naïve inverse of the smoothed periodogram matrix (Naïve), with smoothing span being the modified Daniell kernel with bandwidth picked using the GCV criterion, and the shrinkage estimator (Shrinkage) by [4]. We collect the numerical results averaged over 50 repetitions in each setting in Table 1. Each cell of the table is of the form mean (standard deviation). Since the Naïve estimator and the Shrinkage estimator do not produce sparse estimation, we only report the evaluations on the support recovery for the SIPE. We define the true positive proportion (TPP) and true negative proportion (TNP) as follows.

$$\mathrm{TPP} = \frac{\#\text{non-zero off-diagonal entries in the estimator}}{\#\text{non-zero off-diagonal entries in the truth}},$$
$$\mathrm{TNP} = \frac{\#\text{zero off-diagonal entries in the estimator}}{\#\text{zero off-diagonal entries in the truth}}.$$

The results reported are averaged across all frequencies.

First, looking across all simulation settings, we see that the smoothed periodogram matrix sometimes cannot be inverted, motivating the need for some type of regularisation. The spectral density matrix for the white noise (WN) model is the identity matrix across all frequencies. The Shrinkage is biased towards a scaled identity matrix, hence its superior performance in this setting for all dimensionalities and sample sizes. When the time series data possess autocorrelation, such as in the VAR(1), sparse VAR(1) (sVAR(1)), and VARMA

TABLE 1

Simulation results for estimating the inverse of the spectral density matrix, with mean integrated squared error (MISE), and true positive proportions (TPP) and true negative proportions (TNP). All results entries are in the form of mean (standard deviation). Hyphenated entries (-) denote that the smoothed periodogram matrix could not be inverted. TPP and TNR are reported for SIPE only. MISE entries for all simulation settings except for the VARMA were multiplied by 10^3 for clarity.

		MISE - Precision Matrix			SIPE	
p	T	Naïve	Shrinkage	SIPE	TPP	TNP
WN						
10	200	21.62 (31.58)	0.39 (1.23)	0.17 (0.69)	0.83 (0.05)	0.82 (0.05)
	400	13.17 (21.24)	0.22 (0.62)	0.60 (1.91)	0.74 (0.05)	0.74 (0.05)
50	200	202.44 (258.17)	0.04 (0.01)	0.54 (1.62)	0.71 (0.01)	0.69 (0.01)
	400	13.05 (20.57)	0.02 (0.01)	1.06 (3.11)	0.62 (0.01)	0.62 (0.01)
VAR(1)						
10	200	16.36 (47.10)	3.76 (3.93)	3.60 (0.02)	0.91 (0.02)	0.90 (0.03)
	400	8.94 (18.89)	3.69 (6.13)	3.60 (0.01)	0.89 (0.02)	0.87 (0.02)
50	200	-	3.62 (0.44)	4.18 (3.85)	0.86 (0.01)	0.81 (0.02)
	400	-	3.57 (0.55)	3.71 (1.04)	0.86 (0.01)	0.84 (0.01)
sVAR(1)						
10	200	119.64 (230.51)	12.57 (19.40)	10.18 (20.18)	0.97 (0.02)	0.97 (0.03)
	400	47.04 (99.37)	12.87 (25.12)	9.90 (19.62)	0.94 (0.03)	0.94 (0.03)
50	200	-	17.11 (24.30)	15.97 (23.27)	0.97 (0.01)	0.96 (0.02)
	400	-	14.31 (20.97)	15.96 (23.27)	0.95 (0.02)	0.94 (0.02)
VARMA						
10	200	131.28 (121.97)	1.60 (0.28)	2.06 (0.33)	0.84 (0.03)	1.00 (0.00)
	400	31.03 (11.54)	1.40 (0.17)	2.14 (0.23)	0.80 (0.02)	1.00 (0.00)
50	200	-	7.19 (0.62)	14.02 (0.91)	0.81 (0.01)	0.76 (0.02)
	400	-	6.48 (0.38)	15.39 (0.61)	0.82 (0.01)	0.80 (0.01)

models, SIPE is competitive with the shrinkage estimator with respect to MISE, yet can reasonably estimate the zero and non-zero entries of the precision matrices. In contrast, the shrinkage estimator behaves like a ridge estimator, and hence, by construction cannot obtain sparse estimates of the inverse spectral density matrix. We see that our estimator yields favourable estimates of the spectral precision matrix while giving relatively good estimates on which entries of the spectral precision matrix are truly zero or non-zero.

4.2. Analysis of the Google Flu Trends data

We give an empirical illustration of our proposed methodology by analysing the Google Flu Trends data set. Researchers at Google used select Google search terms to predict influenza activity [20]. The resulting data set consists of weekly predicted numbers of influenza-like-illness related visits out of every 100,000 random outpatient visits within select cities throughout the United States of America. The data set is further aggregated at the state-level and region-level, where the latter comprises of different states. The version of the Google Flu Trends data set we used is the state-level aggregate of log-transformed weekly data from 1 January 2006 to 6 October 2013. The resulting time series thus has $p = 50$ dimensions and length $n = 406$.

The goal of our analysis is to investigate the conditional dependencies of the time series across states. To this end, we need to estimate the partial coherence matrix, which is a function of the inverse of the spectral density matrix. The

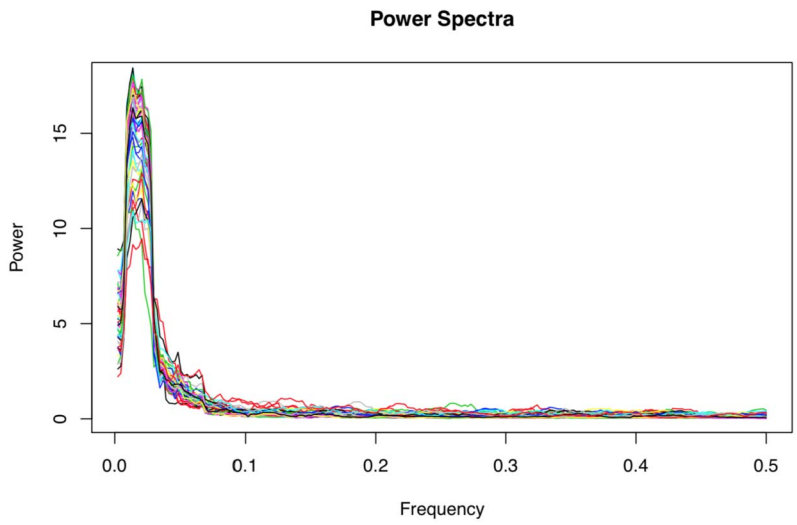

FIG 1. *Power spectra for each state's time series in the Google Flu Trends data. Each colour denotes the power spectrum for one state.*

partial coherence matrix is the frequency domain analog of partial correlation, and can be interpreted as the correlation between two time series that have been bandpass filtered at frequency ω, after removing the linear effects of the other time series. The (j,k)th element of the partial coherence matrix is $\rho_{jk}(\omega) = |\Theta_{jk}(\omega)|^2/[\Theta_{jj}(\omega)\Theta_{kk}(\omega)]$, where $\Theta(\omega)$ is the inverse spectral density matrix. We use our methodology to obtain a sparse estimate of $\Theta(\omega)$, from which we can then obtain estimates of partial coherence. We are only interested in the partial coherence matrix, and so we centre each time series to have mean zero and then we standardised them to have unit variance.

To pick the parameters of our method, we choose M_n using the GCV criterion. Each of the fifty time series were driven by frequencies within the frequency band $(0, 0.10)$, as shown by the diagonal entries of $\widetilde{f}_n$ in Figure 1. Indeed, for each of the fifty time series, the variance attributed to each Fourier frequency outside of this band is less than 5% of the overall variation. Thus, we estimate the partial coherence within this frequency band, and we further summarise our results by taking the median partial coherence within this frequency band. We show our results in Figure 2.

Each of four geographically distinct states (California, New York, Minnesota and Mississippi) yields different conditional independencies as estimated by partial coherence. First, we see a local spatial structure. For instance, we see conditional dependencies between Minnesota and its neighbouring Midwest states, and conditional dependencies between Mississippi and Alabama. Note that the spatial structure of the data was not included in the analysis, yet the association in the time series between neighboring states is still relatively strong even after removing for the effects of the other states. Second, we also conditional

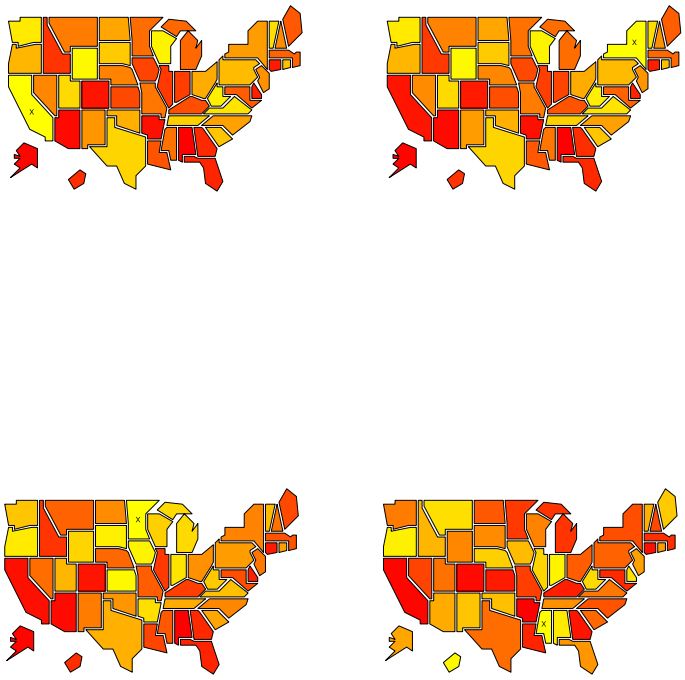

FIG 2. *The conditional dependence structure of the Google Flu Trends data across the fifty states between one state marked by an 'X' and the other forty-nine states. In the top-left, top-right, bottom-left and bottom-right, the 'X'es are California, New York, Minnesota and Mississippi, respectively. Yellow and red colours indicate high and low values of partial coherence, respectively.*

dependencies between geographically distance states, e.g., New York with Washington and Wisconsin. Altogether, the partial coherence gives us information on conditional dependencies that can be attributed to the frequencies driving the variation in the data. Our analysis show both spatially localized and spatially distant conditional dependence structures in the level of influenza activity in the United States, which is consistent with the results by Davis et al. [13].

Appendix

In this section, we collect all the necessary technical details.

Lemma A.1. *Let $Z := A + \imath B \in \mathbb{C}^{p\times p}$, with $A, B \in \mathbb{R}^{p\times p}$. Assume Z is non-singular and the inverse of Z is denoted as Z^{-1}, then $Z^{-1} = \widetilde{A} + \imath\widetilde{B}$, where $\widetilde{A}, \widetilde{B} \in \mathbb{R}^{p\times p}$, satisfying*

$$\begin{pmatrix} A & -B \\ B & A \end{pmatrix} \begin{pmatrix} \widetilde{A} & -\widetilde{B} \\ \widetilde{B} & \widetilde{A} \end{pmatrix} = I_{2\mathrm{p}}.$$

Proof. It follows from the fact $ZZ^{-1} = I_{\mathrm{p}}$ that

$$I_{\mathrm{p}} = (A + \imath B)(\widetilde{A} + \imath\widetilde{B}) = (A\widetilde{A} - B\widetilde{B}) + \imath(B\widetilde{A} + A\widetilde{B}),$$

which is equivalent to

$$\begin{pmatrix} A & -B \\ B & A \end{pmatrix}\begin{pmatrix} \widetilde{A} \\ \widetilde{B} \end{pmatrix} = \begin{pmatrix} I_{\mathrm{p}} \\ 0 \end{pmatrix}, \text{ and } \begin{pmatrix} A & -B \\ B & A \end{pmatrix}\begin{pmatrix} -\widetilde{B} \\ \widetilde{A} \end{pmatrix} = \begin{pmatrix} 0 \\ I_{\mathrm{p}} \end{pmatrix}.$$

Therefore,

$$\begin{pmatrix} A & -B \\ B & A \end{pmatrix}\begin{pmatrix} \widetilde{A} & -\widetilde{B} \\ \widetilde{B} & \widetilde{A} \end{pmatrix} = I_{2\mathrm{p}}. \qquad \square$$

Proof of Theorem 3.1. This proof starts with proving the result for any fixed ω. For any $(i,j) \in \{1,\dots,p\}^{\otimes 2}$, note that the (i,j) entry of the periodogram $P_n(\omega)$ can be written as

$$\begin{aligned} P_{n,ij}(\omega) =& \frac{1}{n}\sum_{t=1}^{n} X_{t,i}\exp(-\imath 2\pi\omega t)\sum_{t=1}^{n} X_{t,j}\exp(\imath 2\pi\omega t) \\ =& \frac{1}{n}\sum_{t=2}^{n}\sum_{l=1}^{t-1} X_{t,i}X_{l,j}\exp(-\imath 2\pi\omega(t-l)) \\ &+ \frac{1}{n}\sum_{l=2}^{n}\sum_{t=1}^{l-1} X_{t,i}X_{l,j}\exp(-\imath 2\pi\omega(t-l)) + \frac{1}{n}\sum_{t=1}^{n} X_{t,i}X_{t,j} \\ =:& P^{(1)}_{n,ij}(\omega) + P^{(2)}_{n,ij}(\omega) + P^{(3)}_{n,ij}. \end{aligned} \tag{A.1}$$

Next, we are to bound the three terms in the right-hand side of (A.1) separately. As for the term (I), we will approximate it by a similar quantities built up by m-dependent random variables. Let

$$\tilde{f}^{(1)}_{n,ij}(\omega) := \frac{1}{2M_n+1}\sum_{s=-M_n}^{M_n} P^{(1)}_{n,ij}(\omega + s/n) = \frac{1}{n}\sum_{t=2}^{n}\sum_{l=1}^{t-1} X_{t,i}X_{l,j}a_{t-l}(\omega),$$

where

$$\begin{aligned} a_k(\omega) &= \frac{1}{2M_n+1}\sum_{s=-M_n}^{M_n}\exp(-\imath 2\pi k(\omega + s/n)) \\ &= \frac{1}{2M_n+1}\exp(-\imath 2\pi k\omega)\sum_{s=-M_n}^{M_n}\exp(-\imath 2\pi ks/n) \\ &= \frac{1}{2M_n+1}\exp(-\imath 2\pi k\omega)\sum_{s=-M_n}^{M_n}\cos(2\pi ks/n) \\ &= \exp(-\imath 2\pi k\omega)\frac{\sin(2\pi(M_n+1/2)k/n)}{(2M_n+1)\sin(\pi k/n)}. \end{aligned} \tag{A.2}$$

The last identity in (A.2) follows from the trigonometric identity that for any α, which is not a multiple of 2π, and any positive integer $m \in \mathbb{N}_+$, we have

$$\sum_{k=0}^{m} \cos(\phi + k\alpha) = \sin((m+1)\alpha/2)\cos(\phi + m\alpha/2)/\sin(\alpha/2).$$

In addition, for any α, which is not a multiple of π, and any $m \in \mathbb{N}_+$, it holds that

$$\left|\frac{\sin(m\alpha)}{\sin(\alpha)}\right| = \left|\frac{\exp(\imath m\alpha)\sin(m\alpha)}{\exp(\imath\alpha)\sin(\alpha)}\right| = \left|\frac{1-\exp(\imath 2m\alpha)}{1-\exp(\imath 2\alpha)}\right| = \left|\sum_{n=0}^{m-1}\exp(\imath 2\alpha)\right| \le m.$$

Then for $k \in \{1, \dots, n-1\}$, we have

$$\frac{\sin^2(2\pi(M_n + 1/2)k/n)}{(2M_n+1)^2\sin^2(\pi k/n)} \le 1. \tag{A.3}$$

For $k \in \{\lfloor n/(M_n+1/2)\rfloor + 1, \dots, \lfloor n - n/(M_n+1/2)\rfloor\}$, we have

$$|\sin(\pi k/n)| \ge \pi/2 \min(k, n-k)/n. \tag{A.4}$$

Combining (A.3) and (A.4) we have

$$\begin{aligned}\sum_{k=1}^{n-1}|a_k(\omega)|^2 &= \sum_{k=1}^{n-1}\frac{\sin^2(2\pi(M_n+1/2)k/n)}{(2M_n+1)^2\sin^2(\pi k/n)} \\ &= \left(\sum_{k=1}^{\lfloor n/(M_n+1/2)\rfloor} + \sum_{k=\lfloor n/(M_n+1/2)\rfloor+1}^{\lfloor n-n/(M_n+1/2)\rfloor} + \sum_{k=\lfloor n-n/(M_n+1/2)\rfloor+1}^{n-1}\right) \\ &\qquad \frac{\sin^2(2\pi(M_n+1/2)k/n)}{(2M_n+1)^2\sin^2(\pi k/n)} \\ &\le \min\left\{2\lceil n/(M_n+1/2)\rceil + \frac{\pi^2}{3}\frac{n^2}{(2M_n+1)^2},\, n-1\right\},\end{aligned}$$

where the last inequality follows from the fact that $\sum_{k=1}^{\infty} = \pi^2/6$. Then,

$$\sum_{k=1}^{n}|a_k(\omega)|^2 \le \min\left\{2\lceil n/(M_n+1/2)\rceil + \frac{\pi^2}{3}\frac{n^2}{(2M_n+1)^2} + 1,\, n\right\} =: A_n;$$

and

$$\max_{k=1,\dots,n}|a_k(\omega)| = 1, \tag{A.5}$$

by noting that $|a_n(\omega)| = 1$.

For $i = 1, \dots, p$, let $\bar{X}_{t,i} = \mathbb{E}(X_{t,i}|\mathcal{F}_{t,m})$, where $\mathcal{F}_{t,m} = (\boldsymbol{e}_{t-m}, \dots, \boldsymbol{e}_t)$ with $m = \lceil(\log(n))^2\rceil$. Note that

$$X_{t,i} - \bar{X}_{t,i} = \sum_{j=m+1}^{\infty}\big(\mathbb{E}(X_{t,i} \mid \mathcal{F}_{t,j}) - \mathbb{E}(X_{t,i} \mid \mathcal{F}_{t,j-1})\big) =: \sum_{j=m+1}^{\infty} d_j,$$

with $\{\mathbb{E}(d_j^2)\}^{1/2} \le \theta_{t,j}$. It follows from Assumption 1 and Theorem 1(ii) in [45], that

$$\{\mathbb{E}((\bar{X}_{t,j} - X_{t,j})^2)\}^{1/2} = O(\rho^m). \tag{A.6}$$

Define

$$\bar{f}_{ij,n}^{(1)}(\omega) = \frac{1}{n}\sum_{t=2}^{n}\sum_{l=1}^{t-1} \bar{X}_{t,i}\bar{X}_{l,j}a_{t-l}(\omega).$$

It follows from Proposition 1 in [29], (A.5) and (A.6) that

$$\mathbb{E}|\bar{f}_{ij,n}^{(1)}(\omega) - \tilde{f}_{ij,n}^{(1)}(\omega)| = O(n\rho^m). \tag{A.7}$$

Now let $M = (\log(n))^2$ and define

$$\breve{X}_{t,i} = \bar{X}_{t,i}\mathbb{1}\{|\bar{X}_{t,i}| \le M\}, \quad \hat{X}_{t,i} = \breve{X}_{t,i} - \mathbb{E}(\breve{X}_{t,i}),$$
$$\breve{f}_{ij,n}^{(1)}(\omega) = \frac{1}{n}\sum_{t=2}^{n}\sum_{l=1}^{t-1} \breve{X}_{t,i}\breve{X}_{l,j}a_{t-l}(\omega), \quad \hat{f}_{ij,T}^{(1)}(\omega) = \frac{1}{n}\sum_{t=2}^{n}\sum_{l=1}^{t-1} \hat{X}_{t,i}\hat{X}_{l,j}a_{t-l}(\omega).$$

Note that for centred random vectors, we have

$$|\mathbb{E}(\bar{X}_{t,i}\mathbb{1}\{|\bar{X}_{t,i}| \le M\})| = |\mathbb{E}(\tilde{X}_{t,i}\mathbb{1}\{|\tilde{X}_{t,i}| > M\})| \le \mathbb{E}\{|\bar{X}_{t,i}|\mathbb{1}\{|\bar{X}_{t,i}| > M\}\}$$
$$\le \int_M^\infty \mathbb{P}\{|\bar{X}_{t,i}| > t\}\,dt \le C_0\int_M^\infty \exp(-\kappa t)\,dt = C_0\kappa^{-1}\exp(-\kappa M), \tag{A.8}$$

where the last inequality follows from Assumption 1 and Markov's inequality.

It also follows from Assumption 1 that

$$\mathbb{P}(\bar{f}_{ij,n}^{(1)}(\omega) \ne \breve{f}_{ij,n}^{(1)}(\omega)) \le \max\Big\{\sum_{t=1}^{n}\mathbb{P}(|X_{t,i}| \ge M), \sum_{t=1}^{n}\mathbb{P}(|X_{t,j}| \ge M)\Big\}$$
$$\le 2C_0 n\exp(-\kappa M). \tag{A.9}$$

To this end, we have for any $\varepsilon > 0$,

$$\begin{aligned}
&\mathbb{P}\{|\tilde{f}_{ij,n}^{(1)}(\omega) - \hat{f}_{ij,n}^{(1)}(\omega)| > \varepsilon\}\\
\le&\mathbb{P}\{|\tilde{f}_{ij,n}^{(1)}(\omega) - \bar{f}_{ij,n}^{(1)}(\omega)| > \varepsilon/3\} + \mathbb{P}\{|\bar{f}_{ij,n}^{(1)}(\omega) - \breve{f}_{ij,n}^{(1)}(\omega)| > \varepsilon/3\}\\
&+ \mathbb{P}\{|\breve{f}_{ij,n}^{(1)}(\omega) - \hat{f}_{ij,n}^{(1)}(\omega)| > \varepsilon/3\}\\
=:&(I) + (II) + (III).
\end{aligned} \tag{A.10}$$

Moreover, it follows from Markov's inequality and (A.7) that

$$(I) \le \frac{\mathbb{E}(|\tilde{f}_{ij,n}^{(1)}(\omega) - \bar{f}_{ij,n}^{(1)}(\omega)|)}{\varepsilon/3} = O(n\rho^m\varepsilon^{-1}). \tag{A.11}$$

It follows from (A.9) that

$$(II) \le \mathbb{P}(\bar{f}_{ij,n}^{(1)}(\omega) \ne \breve{f}_{ij,n}^{(1)}(\omega)) \le 2nC_0\exp\{-\kappa M\}. \tag{A.12}$$

Due to Markov's inequality and (A.8), the following holds

$$(III) = O\big(n\varepsilon^{-1}\exp\big(-2\kappa M\big)\big). \tag{A.13}$$

Finally, combining (A.10), (A.11), (A.12) and (A.13), we obtain that

$$\mathbb{P}\big\{|\tilde{f}^{(1)}_{ij,n}(\omega) - \hat{f}^{(1)}_{ij,n}(\omega)| > \varepsilon\big\} = O\big(n\rho^m\varepsilon^{-1} + n\exp\{-\kappa M\} + n\varepsilon^{-1}\exp\big(-2\kappa M\big)\big). \tag{A.14}$$

Note that $(\hat{X}_{t,i}, \hat{X}_{t,j})$, $1 \le t \le n$ are also m-dependent random vectors with zero means. In addition, we have (A.5),

$$\max_{\substack{t=1,\ldots,n\\ i=1,\ldots,p}} \mathbb{E}(\hat{X}^2_{t,i}) \le K_0, \qquad \max_{\substack{t=1,\ldots,n\\ i=1,\ldots,p}} \mathbb{E}(\hat{X}^4_{t,i}) \le K_0,$$

where K_0 only depends on C_0 following from Assumption 1. Therefore it follows from Proposition 3 in [29] that for any $x \ge 1$, $y \ge 1$ and any constant $Q > 0$ we have,

$$\begin{aligned}
&\mathbb{P}(|\hat{f}^{(1)}_{ij,n}(\omega) - \mathbb{E}\big(\hat{f}^{(1)}_{ij,n}(\omega)\big)| \ge x/n)\\
&\le 2e^{-y/4} + C_1 n^3 M^2\Big(x^{-2}y^2m^3(M^2+n)\sum_{k=1}^{n}|a_k(\omega)|^2\Big)^Q\\
&+ C_1 n^4 M^2 \max\left\{\mathbb{P}\left(|\hat{X}_{0,i}| \ge \frac{C_2 x}{ym^2(M+n^{1/2})}\right), \mathbb{P}\left(|\hat{X}_{0,j}| \ge \frac{C_2 x}{ym^2(M+n^{1/2})}\right)\right\},
\end{aligned}$$

where C_1 and C_2 are positive constants depending only on Q, κ and C_0.

For any $\delta > 0$ and $H > 0$, letting $x = (n/M_n)^{1/2+\delta}n^{1/2}$ and $y = (\log(n))^2$, there exists a constant $C_3 > 0$ only depending on H, κ and C_0, such that

$$\begin{aligned}
&\mathbb{P}(|\hat{f}^{(1)}_{ij,n}(\omega) - \mathbb{E}\big(\hat{f}^{(1)}_{ij,n}(\omega)\big)| \ge (n/M_n)^{1/2+\delta}n^{-1/2})\\
\le& 2n^{-\log(n)/4} + C_3 n^{3-(1+2\delta)Q}A_n^Q M_n^{(1+2\delta)Q}(\log(n))^{4+10Q}\\
\le& C_3 n^{-H}.
\end{aligned} \tag{A.15}$$

Combining (A.14) and (A.15), we have

$$\mathbb{P}(|\tilde{f}^{(1)}_{ij,n}(\omega) - \mathbb{E}\hat{f}^{(1)}_{ij,n}(\omega)| \ge 2(n/M_n)^{1/2+\delta}n^{-1/2}) \le C_3 n^{-H}.$$

We now seek to bound $\mathbb{E}|\hat{f}^{(1)}_{ij,n}(\omega) - \tilde{f}^{(1)}_{ij,n}(\omega)|$. Note that

$$\begin{aligned}
&\mathbb{E}|\hat{f}^{(1)}_{ij,n}(\omega) - \tilde{f}^{(1)}_{ij,n}(\omega)|\\
\le& \mathbb{E}|\hat{f}^{(1)}_{ij,n}(\omega) - \breve{f}^{(1)}_{ij,n}(\omega)| + \mathbb{E}|\breve{f}^{(1)}_{ij,n}(\omega) - \bar{f}^{(1)}_{ij,n}(\omega)| + \mathbb{E}|\bar{f}^{(1)}_{ij,n}(\omega) - \tilde{f}^{(1)}_{ij,n}(\omega)|\\
\le& 2C_0Mn^2\exp(-\kappa M) + O(n\exp\{-\kappa M\}) + O(n\rho^m) \le O(Mn\rho^m),
\end{aligned}$$

which implies

$$\mathbb{P}(|\tilde{f}^{(1)}_{ij,n}(\omega) - \mathbb{E}(\tilde{f}^{(1)}_{ij,n}(\omega))| \ge 3(n/M_n)^{1/2+\delta}n^{-1/2}) \le C_3 n^{-H}. \tag{A.16}$$

Similarly arguments lead to

$$\mathbb{P}(|\tilde{f}^{(2)}_{ij,n}(\omega) - \mathbb{E}(\tilde{f}^{(2)}_{ij,n}(\omega))| \geq 3(n/M_n)^{1/2+\delta} n^{-1/2}) \leq C_3 n^{-H}, \tag{A.17}$$

and

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{t=1}^{n} X_{t,i}X_{t,j} - \mathbb{E}\left\{\frac{1}{n}\sum_{t=1}^{n} X_{t,i}X_{t,j}\right\}\right| \geq n^{-1/2+\delta}\right) \leq C_4 n^{-H}. \tag{A.18}$$

Combining (A.16)-(A.18), we obtain

$$\mathbb{P}(|\tilde{f}_{ij,n}(\omega) - \mathbb{E}(\tilde{f}_{ij,n}(\omega))| \geq 7(n/M_n)^{1/2+\delta} n^{-1/2}) \leq C_5 n^{-H}, \tag{A.19}$$

where $C_4, C_5 > 0$ are constants only depending on δ, H, κ and C_0.

It follows from a slight modification of Theorem 10.4.1 in [6] and Assumption 1 that there exists a constant $C_6 > 0$ only depending on κ and C_0 such that

$$\max_{i,j} |f_{ij}(\omega) - \mathbb{E}(\tilde{f}_{ij,n}(\omega))| \leq C_6 M_n/n. \tag{A.20}$$

Combining (A.19) and (A.20) yields that for any (i,j) we have

$$\mathbb{P}(|\tilde{f}_{ij,n}(\omega) - f_{ij}(\omega)| \geq C_6 M_n/n + 8(n/M_n)^{1/2+\delta} n^{-1/2}) \leq C_5 n^{-H}.$$

Therefore, using the union bound argument we can show that there exists a constant $C > 0$ depending only on κ and C_0 such that for any $\delta > 0$ and $H > 0$ the following holds

$$\mathbb{P}\Bigg\{ \sup_{k\in\{-\lfloor (n-1)/2\rfloor,\ldots,\lfloor n/2\rfloor\}} \max_{i,j=1,\ldots,p} |\widetilde{f}_{ij,n}(\omega_k) - f_{ij}(\omega_k)| \\ > CM_n/n + 8(n/M_n)^{1/2+\delta} n^{-1/2} \Bigg\} \leq p^2 n^{-H}. \qquad \square$$

Proof of Proposition 3.2. It is due to Theorem 7.2 in [9] that for any symmetric matrix A and $w \in [1,\infty]$, the relation $\|A\|_w \leq \|A\|_1$ holds; therefore it is enough to consider only the $w = 1$ case.

Define the event

$$\mathcal{A}_n := \left\{ \sup_{\omega} \max_{k,l=1,\ldots,p} |\widetilde{f}_{ij,n}(\omega) - f_{ij}(\omega)| \leq C\left(\frac{M_n}{n} + \frac{1}{\sqrt{M_n}}\left(\frac{n}{M_n}\right)^{\delta}\right)\right\},$$

where $C > 0$ and $\delta > 0$ are constants. It follows from Therorem 3.1 and Assumption 2 that

$$\mathbb{P}\{\mathcal{A}_n\} \to 1,$$

as $n \to \infty$.

Let $\mathcal{O} := \{-\lfloor (n-1)/2\rfloor, \ldots, \lfloor n/2\rfloor\}$. In the event $\mathcal{A}_n$, we have

$$\sup_{\omega\in\mathcal{O}} \|\widehat{\Sigma}(\omega)\Theta(\omega) - I\|_\infty \leq \sup_{\omega} \|\Theta(\omega)\|_1 \sup_{\omega} \|\widehat{\Sigma}(\omega) - \Sigma(\omega)\|_\infty$$

$$\leq \sup_{\omega} \|\Theta(\omega)\|_1 \sup_{\omega} \|\widetilde{f}_T(\omega) - f(\omega)\|_\infty$$
$$\leq \frac{CM_{n,p}M_n}{n} + \frac{CM_{n,p}n^\delta}{M_n^{1/2+\delta}} \asymp \lambda,$$

where $M_{n,p}$ is defined in (2.3). Then due to the definition of $\widehat{\Theta}(\omega)$, we have for any ω, on the event $\mathcal{A}_n$, $\|\widehat{\Theta}(\omega)\|_1 \leq \|\Theta(\omega)\|_1 \leq M_{n,p}$.

Therefore, in the event $\mathcal{A}_n$,

$$\sup_{\omega\in\mathcal{O}} \|\widehat{\Theta}(\omega) - \Theta(\omega)\|_\infty = \sup_{\omega\in\mathcal{O}} \|(\Theta(\omega)\widehat{\Sigma}(\omega) - I)\widehat{\Theta}(\omega) + \Theta(\omega)(I - \widehat{\Sigma}(\omega)\widehat{\Theta}(\omega))\|_\infty$$
$$\leq \sup_{\omega\in\mathcal{O}} \|\widehat{\Theta}(\omega)\|_1 \sup_{\omega\in\mathcal{O}} \|\Theta(\omega)\widehat{\Sigma}(\omega) - I\|_\infty + \sup_{\omega\in\mathcal{O}} \|\Theta(\omega)\|_1 \sup_{\omega\in\mathcal{O}} \|I - \widehat{\Sigma}(\omega)\widehat{\Theta}(\omega)\|_\infty$$
$$\leq 2M_{n,p}\lambda := t_n.$$

Moreover, we are to bound the ℓ_1 errors. it follows from Lemma 7.1 in [9] that in the event $\mathcal{A}_n$ we have

$$\sup_{\omega\in\mathcal{O}} \|\widehat{\Theta}(\omega) - \Theta(\omega)\|_1 \leq 12c_{n,p}t_n^{1-q},$$

where $c_{n,p}$ is defined in (2.3), and we complete the proof. □

References

[1] Barigozzi, M., & Hallin, M. (2017). A network analysis of the volatility of high dimensional financial series. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **66**, 581–605. MR3632343

[2] Basu, S. & Michailidis, G. (2015). Regularized estimation in sparse high-dimensional time series models. *The Annals of Statistics*, **43**, 1535–1567. MR3357870

[3] Bickel, P. J. & Levina, E. (2008). Covariance regularization by thresholding. *The Annals of Statistics*, **36**, 2577–2604 MR2485008

[4] Böhm, H. & von Sachs, R. (2009). Shrinkage estimation in the frequency domain of multivariate time series. *Journal of Multivariate Analysis.* **100**, 913–935. MR2498723

[5] Brillinger, D. R. (2001). Time Series: Data Analysis and Theory. Vol. 36. SIAM. MR1853554

[6] Brockwell, P. J., and Richard A. D. (2006). Time Series: Theory and Methods. Springer Science & Business Media. MR2839251

[7] Cai, T. T., Zhang, C. H. & Zhou, H. H. (2010). Optimal rates of convergence for covariance matrix estimation. *The Annals of Statistics*, **38**, 2118–2144. MR2676885

[8] Cai, T., Liu, W., & Luo, X. (2011). A constrained ℓ_1 minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association*, **106**, 594–607. MR2847973

[9] Cai, T. T., Liu, W. & Zhou, H. H. (2016). Estimating sparse precision matrix: Optimal rates of convergence and adaptive estimation. *The Annals of Statistics*, **44**, 455–488. MR3476606

[10] Chang, J., Yao, Q., & Zhou, W. (2017). Testing for high-dimensional white noise using maximum cross-correlations. *Biometrika*, **104**, 111–127. MR3626482

[11] Chen, X., Xu, M., & Wu, W. B. (2013). Covariance and precision matrix estimation for high-dimensional time series. *The Annals of Statistics*, **41**, 2994–3021. MR3161455

[12] Dahlhaus, R. (2000). Graphical interaction models for multivariate time series. *Metrika*, **51**, 157–172. MR1790930

[13] Davis, R. A., Zang, P., & Zheng, T. (2016). Sparse vector autoregressive modeling. *Journal of Computational and Graphical Statistics*, **25**, 1077–1096. MR3572029

[14] Fan, J. and Lv, J. and Qi, L. (2011). Sparse high-dimensional models in economics. *Annu. Rev. Econ*, **3**, 291–317.

[15] Fan, Y. & Lv, J. (2016). Innovated scalable efficient estimation in ultra-large Gaussian graphical models. *The Annals of Statistics*, **44**, 2098–2126. MR3546445

[16] Fan, J., & Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, **96**, 1348–1360. MR1946581

[17] Fiecas, M. and Ombao, H. (2011). The generalized shrinkage estimator for the analysis of functional connectivity of brain signals. *The Annals of Applied Statistics*, 1102–1125. MR2840188

[18] Fiecas, M. and von Sachs, R. (2014). Data-driven shrinkage of the spectral density matrix of a high-dimensional time series. *Electronic Journal of Statistics*, **8**, 2975–3003. MR3301298

[19] Gather, U. and Imhoff, M. and Fried, R. (2002). Graphical models for multivariate time series from intensive care monitoring. *Statistics in Medicine*, **21**, 2685–2701.

[20] Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., & Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature*, **457**, 1012–1014.

[21] Guo, S., Wang, Y., & Yao, Q. (2016). High-dimensional and banded vector autoregressions. *Biometrika*, **103**, 889–903 MR3620446

[22] Holbrook, A., Lan, S., Vandenberg-Rodes, A., & Shahbaba, B. (2018). Geodesic Lagrangian Monte Carlo over the space of positive definite matrices: with application to Bayesian spectral density estimation. *Journal of Statistical Computation and Simulation*, **88**, 982–1002. MR3750634

[23] Huang, J., Sun, T., Ying, Z., Yu, Y., & Zhang, C. H. (2013). Oracle inequalities for the lasso in the Cox model. *Annals of Statistics*, **41**, 1142. MR3113806

[24] Jukić, D. and Denić-Jukić, V. (2011). Partial spectral analysis of hydrological time series. *Journal of Hydrology*, **408**, 223–233.

[25] Jung, A., Hannak, G., & Goertz, N. (2015). Graphical lasso based model

selection for time series. *IEEE Signal Processing Letters*, **22**, 1781–1785.
[26] Lam, C., & Yao, Q. (2012). Factor modeling for high-dimensional time series: inference for the number of factors. *The Annals of Statistics*, **40**, 694–726. MR2933663
[27] Leclere, Q. (2009). Multi-channel spectral analysis of multi-pass acquisition measurements. *Mechanical Systems and Signal Processing*, **23**, 1415–1422.
[28] Liu, H., Roeder, K. and Larry, W. (2010) Stability approach to regularization selection (StARS) for high dimensional graphical models. Advances in Neural Information Processing Systems 23, 1432–1440.
[29] Liu, W., & Wu, W. B. (2010). Asymptotics of spectral density estimates. *Econometric Theory*, **26**, 1218–1245. MR2660298
[30] Meinshausen, N., & Bühlmann, P. (2006). High-dimensional graphs and variable selection with the Lasso. *The annals of statistics*, **34**, 1436–1462. MR2278363
[31] Meinshausen, N., & Bühlmann, P. (2010). Stability selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **72**, 417–473. MR2758523
[32] Negahban, S., & Wainwright, M. J. (2011). Estimation of (near) low-rank matrices with noise and high-dimensional scaling. *The Annals of Statistics*, 1069–1097. MR2816348
[33] Ombao, H. C., Raz, J. A., Strawderman, R. L., & von Sachs, R. (2001). A simple generalised crossvalidation method of span selection for periodogram smoothing. *Biometrika*, 88(4), 1186–1192. MR1872229
[34] Orini, M. and Bailón, R. and Laguna, P. and Mainardi, L. T. and Barbieri, R. (2012). A multivariate time-frequency method to characterize the influence of respiration over heart period and arterial pressure. *EURASIP Journal on Advances in Signal Processing*, 2012(1), 214.
[35] Qin, T., & Rohe, K. (2013). Regularized spectral clustering under the degree-corrected stochastic blockmodel. In *Advances in Neural Information Processing Systems*, 3120–3128.
[36] Salvador, R. and Suckling, J. and Schwarzbauer, C. and Bullmore, E. (2005). Undirected graphs of frequency-dependent functional connectivity in whole brain networks. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1457), 937–946.
[37] Schneider-Luftman, D., & Walden, A. T. (2016). Partial coherence estimation via spectral matrix shrinkage under quadratic loss. In *IEEE Transactions on Signal Processing*, **64**, 5767–5777. MR3554582
[38] Shu, H. & Nan, B. (2014). Estimation of large covariance and precision matrices from temporally dependent observations. *arXiv preprint arXiv:1412.5059*. MR3911114
[39] Stewart, G. & Sun, J.-G. (1990). Matrix Perturbation Theory. Computer Science and Scientific Computing. Academic Press. MR1061154
[40] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 267–288. MR1379242
[41] Van de Geer, S. A. (2008). High-dimensional generalized linear models and

the lasso. *The Annals of Statistics*, **36**, 614–645. MR2396809

[42] Wang, D., Yu, Y., & Rinaldo, A. (2017). Optimal covariance change point detection in high dimension. *ArXiv preprint, arXiv:1712.09912*.

[43] Wasserman, L., & Roeder, K. (2009). High dimensional variable selection. *The Annals of Statistics*, **37**, 2178. MR2543689

[44] Wilms, I., Basu, S., Bien, J., & Matteson, D. (2017). Sparse identification and estimation of high-dimensional vector autoregressive moving averages. *ArXiv preprint, arXiv:1707.09208*.

[45] Wu, W. B. (2005). Nonlinear system theory: Another look at dependence. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 14150–14154. MR2172215

[46] Xie, H., & Huang, J. (2009). SCAD-penalized regression in high-dimensional partially linear models. *The Annals of Statistics*, **37**, 673–696. MR2502647

[47] Yu, Y., Bradic, J. & Samworth, R. J. (2018). Confidence intervals for high-dimensional Cox models. *ArXiv preprint arXiv:1803.01150*. MR3650395

[48] Yuan, M. & Lin, Y. (2007). Model selection and estimation in the Gaussian graphical model. *Biometrika*, **94**, 19–35. MR2367824

[49] Zhang, C. H. (2010). Nearly unbiased variable selection under minimax concave penalty. *The Annals of Statistics*, **38**, 894–942. MR2604701

[50] Zhang, C.-H. & Zhang, S. S. (2014) Confidence intervals for low dimensional parameters in high dimensional linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **76**, 217–242. MR3153940

[51] Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **67**, 301–320. MR2137327