Qiwei Yao, Jinyuan Chang, Yumou Qiu and Tao Zou

Confidence regions for entries of a large precision matrix

Article (Published version) (Refereed)

Original citation:
Yao, Qiwei and Chang, Jinyuan and Qiu, Yumou and Zou, Tao (2018) *Confidence regions for entries of a large precision matrix*. Journal of Econometrics. ISSN 0304-4076
DOI: 10.1016/j.jeconom.2018.03.020

ARTICLE IN PRESS
Journal of Econometrics ▮ (▮▮▮▮) ▮▮▮–▮▮▮
Contents lists available at ScienceDirect
Journal of Econometrics
journal homepage: www.elsevier.com/locate/jeconom

Confidence regions for entries of a large precision matrix

Jinyuan Chang [a], Yumou Qiu [b], Qiwei Yao [c,*], Tao Zou [d]

[a] School of Statistics, Southwestern University of Finance and Economics, Chengdu, Sichuan 611130, China
[b] Department of Statistics, University of Nebraska-Lincoln, Lincoln, NE 68583-0963, USA
[c] Department of Statistics, London School of Economics, London, WC2A 2AE, UK
[d] College of Business and Economics, The Australian National University, Acton, ACT 2601, Australia

ARTICLE INFO

Article history:
Received 21 April 2017
Received in revised form 24 October 2017
Accepted 29 March 2018
Available online xxxx

JEL classification:
C12
C13
C15

Keywords:
Bias correction
Dependent data
High dimensionality
Kernel estimation
Parametric bootstrap
Precision matrix

ABSTRACT

We consider the statistical inference for high-dimensional precision matrices. Specifically, we propose a data-driven procedure for constructing a class of simultaneous confidence regions for a subset of the entries of a large precision matrix. The confidence regions can be applied to test for specific structures of a precision matrix, and to recover its nonzero components. We first construct an estimator for the precision matrix via penalized node-wise regression. We then develop the Gaussian approximation to approximate the distribution of the maximum difference between the estimated and the true precision coefficients. A computationally feasible parametric bootstrap algorithm is developed to implement the proposed procedure. The theoretical justification is established under the setting which allows temporal dependence among observations. Therefore the proposed procedure is applicable to both independent and identically distributed data and time series data. Numerical results with both simulated and real data confirm the good performance of the proposed method.

1. Introduction

With an ever-increasing capacity of collecting and storing data, industry, business and government offices all encounter the task of analyzing the data of unprecedented size arisen from various practical fields such as panel studies of economic, social and natural (such as weather) phenomena, financial market analysis, genetic studies and communications engineering. A significant feature of these data is that the number of variables recorded on each individual is large or extremely large. Meanwhile, in many empirical studies, observations taken over different times are dependent with each other. Therefore, many well-developed statistical inference methods for independent and identically distributed (i.i.d.) data may no longer be applicable. Those features of modern data bring both opportunities and challenges to statisticians and econometricians.

The entries of covariance matrix measure the marginal linear dependence of two components of a random vector. There is a large body of literature on estimation and hypothesis testing of high-dimensional covariance matrices with i.i.d. data, including Bickel and Levina (2008a, b), Qiu and Chen (2012), Cai et al. (2013), Chang et al. (2017b) and references within. In order to capture the conditional dependence of two components of a random vector conditionally on all the others, the Gaussian graphical model (GGM) has been widely used. Under GGM, conditional independence of two components is equivalent to the fact that the correspondent entry of the precision matrix (i.e. the inverse of the covariance matrix) is zero. Therefore, the conditional dependence among components of a random vector can be well understood by investigating

* Corresponding author.
E-mail addresses: changjinyuan@swufe.edu.cn (J. Chang), yumouqiu@unl.edu (Y. Qiu), q.yao@lse.ac.uk (Q. Yao), tao.zou@anu.edu.au (T. Zou).

https://doi.org/10.1016/j.jeconom.2018.03.020

the structure of its precision matrix. Beyond GGM, the bijection relationship between the conditional dependence and the precision matrix may not hold. Nevertheless, the precision matrix still plays an important role in, among others, linear regression (van de Geer et al., 2014), linear prediction and kriging, and partial correlation graphs (Huang et al., 2010). See also Examples 1–3 in Section 2.

Let $\boldsymbol{\Omega}$ denote a $p \times p$ precision matrix and p be large. With i.i.d. observations, Yuan and Lin (2007) and Friedman et al. (2008) adopted graphical Lasso to estimate $\boldsymbol{\Omega}$ by maximizing the likelihood with an L_1 penalty. Meinshausen and Bühlmann (2006) introduced a neighborhood selection procedure which estimates $\boldsymbol{\Omega}$ by finding the nonzero regression coefficients of each component on all the other components using Lasso (Tibshirani, 1996) or Dantzig method (Candes and Tao, 2007). Also see Cai et al. (2011), Xue and Zou (2012) and Sun and Zhang (2013) for other penalized estimation methods. Chen et al. (2013) investigated the theoretical properties of the graphical Lasso estimator for $\boldsymbol{\Omega}$ with dependent observations. Though these methods provide consistent estimators for $\boldsymbol{\Omega}$ under some structural assumptions (for example, sparsity) imposed on $\boldsymbol{\Omega}$, they cannot be used for statistical inference directly due to the non-negligible estimation biases, caused by the penalization, which are of order slower than $n^{-1/2}$.

The bias issue has been successfully overcome with i.i.d. Gaussian observations by, for example, Liu (2013) based on p node-wise regressions method. Furthermore, Ren et al. (2015) proposed a novel estimator for each entry of $\boldsymbol{\Omega}$ based on pairwise L_1 penalized regression, and showed that their estimators achieved the minimax optimal rate with no bias terms. In spite of $\frac{p(p-1)}{2}$ pairs among p components, their method in practice only requires at most $p(1+\bar{s})$ pairwise L_1 penalized regressions, where $\bar{s}$ is the average size of the selected node-wise regression models.

The major contribution of this paper is to construct the confidence regions for subsets of the entries of $\boldsymbol{\Omega}$. To our best knowledge, this is the first attempt of this kind. Furthermore we provide the asymptotic justification under the setting which allows dependent observations, and, hence, includes i.i.d. data as a special case. See also Remark 2 in Section 3.2. More precisely, let $\mathcal{S} \subset \{1,\ldots,p\}^2$ be a given index set of interest, whose cardinality $|\mathcal{S}|$ can be finite or grow with p. Let $\boldsymbol{\Omega}_{\mathcal{S}}$ be the vector consisting of the entries of $\boldsymbol{\Omega}$ with their indices in $\mathcal{S}$. We propose a class of data-driven confidence regions $\{\mathcal{C}_{\mathcal{S},\alpha}\}_{0<\alpha<1}$ for $\boldsymbol{\Omega}_{\mathcal{S}}$ such that $\sup_{0<\alpha<1}|\mathbb{P}(\boldsymbol{\Omega}_{\mathcal{S}} \in \mathcal{C}_{\mathcal{S},\alpha}) - \alpha| \to 0$ when both $n, p \to \infty$, where n denotes the sample size. The potential application of $\mathcal{C}_{\mathcal{S},\alpha}$ is wide, including, for example, testing for some specific structures of $\boldsymbol{\Omega}$, and detecting and recovering nonzero entries of $\boldsymbol{\Omega}$ consistently.

For any matrix $\mathbf{A} = (a_{ij})$, let $|\mathbf{A}|_\infty = \max_{i,j}|a_{ij}|$ be its element-wise L_∞-norm. We proceed as follows. First we propose a bias corrected estimator $\widehat{\boldsymbol{\Omega}}_{\mathcal{S}}$ for $\boldsymbol{\Omega}_{\mathcal{S}}$ via penalized node-wise regressions, and develop an asymptotic expansion for $n^{1/2}(\widehat{\boldsymbol{\Omega}}_{\mathcal{S}} - \boldsymbol{\Omega}_{\mathcal{S}})$ without assuming Gaussianity. As the leading term in the asymptotic expansion is a partial sum, we approximate the distribution of $n^{1/2}|\widehat{\boldsymbol{\Omega}}_{\mathcal{S}} - \boldsymbol{\Omega}_{\mathcal{S}}|_\infty$ by that of the L_∞-norm of a high-dimensional normal distributed random vector with mean zero and covariance being an estimated long-run covariance matrix of an unobservable process. This normal approximation, inspired by Chernozhukov et al. (2013, 2014), paves the way for evaluating the probabilistic behavior of $n^{1/2}|\widehat{\boldsymbol{\Omega}}_{\mathcal{S}} - \boldsymbol{\Omega}_{\mathcal{S}}|_\infty$ by parametric bootstrap.

It is worth pointing out that the kernel estimator for long-run covariances, initially proposed by Andrews (1991) for the problem with fixed dimension (i.e. p fixed), also works under our setting with $p \to \infty$ without requiring any structural assumptions on the underlying long-run covariance matrix. Owning to the form of this kernel estimator, the parametric bootstrap sampling can be implemented in an efficient manner in terms of both computational complexity and the required storage space; see Remark 4 in Section 3.2.

The rest of the paper is organized as follows. Section 2 introduces the problem to be solved and its background. The proposed procedure and its theoretical properties are presented in Section 3. Section 4 discusses the applications of our results. Simulation studies and a real data analysis are reported in Sections 5 and 6, respectively. All the technical proofs are relegated to the Appendix. We conclude this section by introducing some notation that is used throughout the paper. We write $a_n \asymp b_n$ to mean $0 < \liminf_{n\to\infty}|a_n/b_n| \le \limsup_{n\to\infty}|a_n/b_n| < \infty$. We say $x_{n,j} = o_p(a_n)$ uniformly over $j \in \mathcal{J}$ if $\max_{j\in\mathcal{J}}|x_{n,j}/a_n| \xrightarrow{p} 0$ as $n \to \infty$. Let $|\cdot|_1$ and $|\cdot|_0$ denote, respectively, the L_1- and L_0-norm of a vector.

2. Preliminaries

Let $\mathbf{y}_1,\ldots,\mathbf{y}_n$ be n observations from an $\mathbb{R}^p$-valued time series, where $\mathbf{y}_t = (y_{1,t},\ldots,y_{p,t})^{\mathrm{T}}$ and each $\mathbf{y}_t$ has the constant first two moments, i.e. $\mathbb{E}(\mathbf{y}_t) = \boldsymbol{\mu}$ and $\mathrm{Cov}(\mathbf{y}_t) = \boldsymbol{\Sigma}$ for each t. Let $\boldsymbol{\Omega} = \boldsymbol{\Sigma}^{-1}$ be the precision matrix. We assume that $\{\mathbf{y}_t\}$ is β-mixing in the sense that $\beta_k \to 0$ as $k \to \infty$, where

$$\beta_k = \sup_t \mathbb{E}\left\{ \sup_{B\in\mathscr{F}_{t+k}^{\infty}} \left|\mathbb{P}(B|\mathscr{F}_{-\infty}^{t}) - \mathbb{P}(B)\right| \right\}.$$

Here $\mathscr{F}_{-\infty}^{t}$ and $\mathscr{F}_{t+k}^{\infty}$ are the σ-fields generated respectively by $\{\mathbf{y}_u\}_{u\le t}$ and $\{\mathbf{y}_u\}_{u\ge t+k}$. β-mixing is a mild condition for time series. It is known that causal ARMA processes with continuous innovation distributions, stationary Markov chains under some mild conditions and stationary GARCH models with finite second moments and continuous innovation distributions are all β-mixing. We refer to Section 2.6 of Fan and Yao (2003) for the further details on β-mixing condition.

For a given index set $\mathcal{S} \subset \{1,\ldots,p\}^2$, recall $\boldsymbol{\Omega}_{\mathcal{S}}$ denotes the vector consisting of the entries of $\boldsymbol{\Omega}$ with their indices in $\mathcal{S}$. We are interested in constructing a class of confidence regions $\{\mathcal{C}_{\mathcal{S},\alpha}\}_{0<\alpha<1}$ for $\boldsymbol{\Omega}_{\mathcal{S}}$ such that

$$\sup_{0<\alpha<1}\left|\mathbb{P}(\boldsymbol{\Omega}_{\mathcal{S}} \in \mathcal{C}_{\mathcal{S},\alpha}) - \alpha\right| \to 0 \quad \text{as } n, p \to \infty. \tag{1}$$

Please cite this article in press as: Chang J., et al., Confidence regions for entries of a large precision matrix. Journal of Econometrics (2018), https://doi.org/10.1016/j.jeconom.2018.03.020.

We also allow $r \equiv |\mathcal{S}|$, the length of vector $\boldsymbol{\Omega}_{\mathcal{S}}$, either to be fixed or to go to infinity together with p. The largest r can be p^2. We first give several motivating examples.

Example 1 (*High-dimensional Linear Regression*)**.** Consider linear regression $z_t = \mathbf{x}_t^{\mathrm{T}}\boldsymbol{\gamma}+\varepsilon_t$ with $\mathbb{E}(\mathbf{x}_t\varepsilon_t) = \mathbf{0}$, where $\mathbf{x}_t$ consists of m explanatory variables and m is large, and $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_m)^{\mathrm{T}} = \{\mathbb{E}(\mathbf{x}_t\mathbf{x}_t^{\mathrm{T}})\}^{-1}\mathbb{E}(\mathbf{x}_t z_t)$ are true regression coefficients. In order to identify non-zero regression coefficients, we test the hypotheses

$$H_0 : \gamma_l = 0 \text{ for all } l \in \mathcal{A} \quad \text{vs.} \quad H_1 : \gamma_l \neq 0 \text{ for some } l \in \mathcal{A}, \tag{2}$$

where $\mathcal{A} \subset \{1, \ldots, m\}$ is a given index set of interest. Let $\mathbf{y}_t = (z_t, \mathbf{x}_t^{\mathrm{T}})^{\mathrm{T}}$, and $\boldsymbol{\Omega} = (\omega_{j_1,j_2})_{p\times p}$ be the precision matrix of $\mathbf{y}_t$. It can be shown that $(\omega_{1,2}, \ldots, \omega_{1,p})^{\mathrm{T}} = -c\boldsymbol{\gamma}$, where $c = [\mathrm{Var}(z_t) - \mathbb{E}(\mathbf{x}_t^{\mathrm{T}}z_t)\{\mathbb{E}(\mathbf{x}_t\mathbf{x}_t^{\mathrm{T}})\}^{-1}\mathbb{E}(\mathbf{x}_t z_t)]^{-1} > 0$. Thus, (2) can be equivalently expressed as

$$H_0 : \omega_{j_1,j_2} = 0 \text{ for any } (j_1, j_2) \in \mathcal{S} \quad \text{vs.} \quad H_1 : \omega_{j_1,j_2} \neq 0 \text{ for some } (j_1, j_2) \in \mathcal{S}, \tag{3}$$

where $\mathcal{S} = \{(1, l) : l - 1 \in \mathcal{A}\}$. We reject H_0 at the significance level α if $\mathcal{C}_{\mathcal{S},\alpha}$ does not contain the origin of $\mathbb{R}^r$ with $r = |\mathcal{A}|$.

Example 2 (*Linear Prediction and Kriging*)**.** In the context of predicting a random variable z_t based on an observed p-dimensional vector $\mathbf{x}_t$, the best linear predictor in the sense of minimizing the mean squared predictive error is $\mathrm{Cov}(z_t, \mathbf{x}_t)\boldsymbol{\Omega}\mathbf{x}_t$, where $\boldsymbol{\Omega}$ is the precision matrix of $\mathbf{x}_t$. Here we assume the means of both z_t and $\mathbf{x}_t$ are zero, to simplify the notation. We also assume that any redundant components of $\mathbf{x}_t$ have been removed by applying the techniques described in Example 1.

To obtain a consistent estimate for $\boldsymbol{\Omega}$ when p is large, it is necessary to impose some structural assumptions on $\boldsymbol{\Omega}$. In the context of kriging (i.e. linear prediction in the context of spatial or spatial–temporal statistics), some lower-dimensional factor structures have been explored. See Huang et al. (2017) and the references within. Bandness/bandableness is another popular structural assumption often used in estimating large covariance or precision matrices (Bickel and Levina, 2008a). To investigate a banded structure for $\boldsymbol{\Omega}$, one may test the hypotheses

$$H_0 : \omega_{j_1,j_2} = 0 \text{ for any } |j_1 - j_2| > k \quad \text{vs.} \quad H_1 : \omega_{j_1,j_2} \neq 0 \text{ for some } |j_1 - j_2| > k, \tag{4}$$

where $1 \le k < p$ is a prespecified integer. We reject H_0 if confidence region $\mathcal{C}_{\mathcal{S},\alpha}$ does not contain the origin $\mathbb{R}^r$, where $\mathcal{S} = \{(j_1, j_2) : 1 \le j_1, j_2 \le p,\ j_2 - j_1 > k\}$ and $r = (p - k)(p - k - 1)/2$.

Example 3 (*Partial Correlation Network*)**.** Given a precision matrix $\boldsymbol{\Omega} = (\omega_{j_1,j_2})_{p\times p}$, we can define an undirected network $G = (V, E)$ where the vertex set $V = \{1, \ldots, p\}$ represents the p components of $\mathbf{y}$ and the edge set $E = \{(j_1, j_2) \in V \times V : \omega_{j_1,j_2} \neq 0,\ j_1 < j_2\}$ are the pairs of variables with non-zero precision coefficients. Let $\rho_{j_1,j_2} = \mathrm{Corr}(\varepsilon_{j_1}, \varepsilon_{j_2})$ be the partial correlation between the j_1th and the j_2th components of $\mathbf{y}$ for any $j_1 \neq j_2$, where ε_{j_1} and ε_{j_2} are the errors of the best linear predictors of y_{j_1} and y_{j_2} given $\mathbf{y}_{-(j_1,j_2)} = \{y_k : k \neq j_1, j_2\}$, respectively. From Lemma 1 of Peng et al. (2009), it is known that $\rho_{j_1,j_2} = -\frac{\omega_{j_1,j_2}}{\sqrt{\omega_{j_1,j_1}\omega_{j_2,j_2}}}$. Therefore, the network $G = (V, E)$ also represents the partial correlation graph of $\mathbf{y}$. The vertices $(j_1, j_2) \notin E$ if and only if y_{j_1} and y_{j_2} are partially uncorrelated. The GGM assumes in addition that $\mathbf{y}$ is multivariate normal. Then $\boldsymbol{\Omega}$ depicts the conditional dependence among the p vertices of the network, i.e. ω_{j_1,j_2} is the conditional correlation between the j_1th and j_2th vertices given all the others.

Neighborhood and community are two basic features in a network. The neighborhood of the jth vertex, denoted by $\mathcal{N}_j$, is the set of all the vertices directly connected to it. For most of the spatial data, it is believed that the partial correlation neighborhood is related to the spatial neighborhood. Let $\mathcal{N}_j(k)$ be the set including the first k closest vertices to the jth vertex in the spatial domain. It is of great interest to test $H_0 : \mathcal{N}_j = \mathcal{N}_j(k)$ vs. $H_1 : \mathcal{N}_j \neq \mathcal{N}_j(k)$ for some pre-specified positive constant k. A community in a network is a group of vertices that have heavier connectivity within the group than outside the group. For graph estimation, we want to maximize the within-community connectivity and reduce the between-community connectivity. Therefore, it is of practical importance to explore the connectivity between different communities. Assume the p components of $\mathbf{y}$ are decomposed into K disjoint communities $V_1, \ldots, V_K$. We are interested in recovering $\mathcal{D} = \{(k_1, k_2) : \omega_{j_1,j_2} \neq 0 \text{ for some } j_1 \in V_{k_1} \text{ and } j_2 \in V_{k_2}\}$.

3. Main results

3.1. Estimation of $\boldsymbol{\Omega}$

We first recall the relationship between a precision matrix and node-wise regressions. For a random vector $\mathbf{y} = (y_1, \ldots, y_p)^{\mathrm{T}}$ with mean $\boldsymbol{\mu} = \mathbf{0}$ and covariance $\boldsymbol{\Sigma}$, we consider p node-wise regressions

$$y_{j_1} = \sum_{j_2 \neq j_1} \alpha_{j_1,j_2} y_{j_2} + \epsilon_{j_1} \quad (j_1 = 1, \ldots, p). \tag{5}$$

Let $\mathbf{y}_{-j_1} = \{y_{j_2} : j_2 \neq j_1\}$. The regression error ϵ_{j_1} is uncorrelated with $\mathbf{y}_{-j_1}$ if and only if $\alpha_{j_1,j_2} = -\frac{\omega_{j_1,j_2}}{\omega_{j_1,j_1}}$ for any $j_2 \neq j_1$. Under this condition, $\mathrm{Cov}(\epsilon_{j_1}, \epsilon_{j_2}) = \frac{\omega_{j_1,j_2}}{\omega_{j_1,j_1}\omega_{j_2,j_2}}$ for any j_1 and j_2. Let $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_p)^{\mathrm{T}}$ and $\mathbf{V} = \mathrm{Cov}(\boldsymbol{\epsilon}) = (v_{j_1,j_2})_{p\times p}$. Then

$\boldsymbol{\Omega} = \{\mathrm{diag}(\mathbf{V})\}^{-1}\mathbf{V}\{\mathrm{diag}(\mathbf{V})\}^{-1}$; see Lemma 1 of Peng et al. (2009). This relationship between $\boldsymbol{\Omega}$ and $\mathbf{V}$ provides a way to learn $\boldsymbol{\Omega}$ by the regression errors in (5).

Since the error vector $\boldsymbol{\epsilon}$ in (5) is unobservable in practice, its "proxy" – the residuals of the node-wise regressions – can be used to estimate $\mathbf{V}$. Let $\boldsymbol{\alpha}_j = (\alpha_{j,1}, \ldots, \alpha_{j,j-1}, -1, \alpha_{j,j+1}, \ldots, \alpha_{j,p})^{\mathrm{T}}$. For each $j = 1, \ldots, p$, we may fit the high-dimensional linear regression

$$y_{j,t} = \sum_{k \neq j} \alpha_{j,k} y_{k,t} + \epsilon_{j,t} \quad (t = 1, \ldots, n) \tag{6}$$

by Lasso (Tibshirani, 1996), Dantzig estimation (Candes and Tao, 2007) or scaled Lasso (Sun and Zhang, 2012). For the case $\boldsymbol{\mu} \neq \mathbf{0}$, the regression (6) will be conducted on the centered data $\mathbf{y}_t - \bar{\mathbf{y}}$, where $\bar{\mathbf{y}} = n^{-1}\sum_{t=1}^{n}\mathbf{y}_t$ is the sample mean. For simplicity, we adopt Lasso estimation. Let $\widehat{\boldsymbol{\alpha}}_j$ be the Lasso estimator of $\boldsymbol{\alpha}_j$ defined as follows:

$$\widehat{\boldsymbol{\alpha}}_j = \arg\min_{\boldsymbol{\gamma} \in \Theta_j} \left[\frac{1}{n} \sum_{t=1}^{n} (\boldsymbol{\gamma}^{\mathrm{T}} \mathbf{y}_t)^2 + 2\lambda_j |\boldsymbol{\gamma}|_1 \right], \tag{7}$$

where $\Theta_j = \{\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_p)^{\mathrm{T}} \in \mathbb{R}^p : \gamma_j = -1\}$ and λ_j is the tuning parameter. For each t, the residual

$$\widehat{\epsilon}_{j,t} = -\widehat{\boldsymbol{\alpha}}_j^{\mathrm{T}} \mathbf{y}_t \tag{8}$$

provides an estimate of $\epsilon_{j,t}$. Write $\widehat{\boldsymbol{\epsilon}}_t = (\widehat{\epsilon}_{1,t}, \ldots, \widehat{\epsilon}_{p,t})^{\mathrm{T}}$ and let $\widetilde{\mathbf{V}} = (\widetilde{v}_{j_1,j_2})_{p \times p}$ be the sample covariance of $\{\widehat{\boldsymbol{\epsilon}}_t\}_{t=1}^n$, where $\widetilde{v}_{j_1,j_2} = n^{-1}\sum_{t=1}^{n}\widehat{\epsilon}_{j_1,t}\widehat{\epsilon}_{j_2,t}$. It is well known that $n^{-1}\sum_{t=1}^{n}\epsilon_{j_1,t}\epsilon_{j_2,t}$ is an unbiased estimator of v_{j_1,j_2}, however, replacing $\epsilon_{j_1,t}$ by $\widehat{\epsilon}_{j_1,t}$ will incur a bias term. Specifically, as shown in Lemma 3 in the Appendix, under Conditions 1–3 and some mild restrictions on the sparsity of $\boldsymbol{\Omega}$ and the growth rate of p with respect to n, it holds that

$$\begin{aligned} \widetilde{v}_{j_1,j_2} - \frac{1}{n}\sum_{t=1}^{n}\epsilon_{j_1,t}\epsilon_{j_2,t} = &-(\widehat{\alpha}_{j_1,j_2} - \alpha_{j_1,j_2})\left(\frac{1}{n}\sum_{t=1}^{n}\epsilon_{j_2,t}^2\right)\mathbb{I}(j_1 \neq j_2) \\ &- (\widehat{\alpha}_{j_2,j_1} - \alpha_{j_2,j_1})\left(\frac{1}{n}\sum_{t=1}^{n}\epsilon_{j_1,t}^2\right)\mathbb{I}(j_1 \neq j_2) + o_p\{(n\log p)^{-1/2}\}. \end{aligned} \tag{9}$$

Here the higher order term $o_p\{(n\log p)^{-1/2}\}$ is uniform over all j_1 and j_2. Since $n^{-1}\sum_{t=1}^{n}\epsilon_{j,t}^2$ is $n^{1/2}$-consistent for $v_{j,j}$, (9) implies that $\widetilde{v}_{j,j}$ is also $n^{1/2}$-consistent for $v_{j,j}$. However, for any $j_1 \neq j_2$, due to the slow convergence rates of the Lasso estimators $\widehat{\alpha}_{j_1,j_2}$ and $\widehat{\alpha}_{j_2,j_1}$, $\widetilde{v}_{j_1,j_2}$ is no longer $n^{1/2}$-consistent for v_{j_1,j_2}. To eliminate the bias, we employ an estimator for v_{j_1,j_2}:

$$\widehat{v}_{j_1,j_2} = \begin{cases} -\dfrac{1}{n}\displaystyle\sum_{t=1}^{n}(\widehat{\epsilon}_{j_1,t}\widehat{\epsilon}_{j_2,t} + \widehat{\alpha}_{j_1,j_2}\widehat{\epsilon}_{j_2,t}^2 + \widehat{\alpha}_{j_2,j_1}\widehat{\epsilon}_{j_1,t}^2), & j_1 \neq j_2; \\ \dfrac{1}{n}\displaystyle\sum_{t=1}^{n}\widehat{\epsilon}_{j_1,t}\widehat{\epsilon}_{j_2,t}, & j_1 = j_2. \end{cases} \tag{10}$$

By noticing that $\boldsymbol{\Omega} = \{\mathrm{diag}(\mathbf{V})\}^{-1}\mathbf{V}\{\mathrm{diag}(\mathbf{V})\}^{-1}$, we estimate ω_{j_1,j_2} by

$$\widehat{\omega}_{j_1,j_2} = \frac{\widehat{v}_{j_1,j_2}}{\widehat{v}_{j_1,j_1}\widehat{v}_{j_2,j_2}} \tag{11}$$

for any j_1 and j_2. We need to point out that the asymptotic expansion (9) is still valid for other penalized methods such as Dantzig estimation (Candes and Tao, 2007) and scaled Lasso (Sun and Zhang, 2012). Hence, we can also estimate v_{j_1,j_2} and ω_{j_1,j_2} as (10) and (11), respectively, based on the residuals $\{\widehat{\boldsymbol{\epsilon}}_t\}_{t=1}^n$ obtained by other penalized methods. To study the theoretical properties of this estimator $\widehat{\omega}_{j_1,j_2}$, we need the following regularity conditions.

Condition 1. *There exist constants $K_1 > 0$, $K_2 > 1$, $0 < \gamma_1 \le 2$ and $0 < \gamma_2 \le 2$ independent of p and n such that for each $t = 1, \ldots, n$,*

$$\max_{1 \le j \le p} \mathbb{E}\{\exp(K_1|y_{j,t}|^{\gamma_1})\} \le K_2 \quad \textit{and} \quad \max_{1 \le j \le p} \mathbb{E}\{\exp(K_1|\epsilon_{j,t}|^{\gamma_2})\} \le K_2.$$

Condition 2. *The eigenvalues of $\boldsymbol{\Sigma}$ are uniformly bounded away from zero and infinity.*

Condition 3. *There exist constants $K_3 > 0$ and $\gamma_3 > 0$ independent of p and n such that $\beta_k \le \exp(-K_3 k^{\gamma_3})$ for any positive k.*

Condition 1 implies $\max_{1 \le j \le p}\mathbb{P}(|y_{j,t}| \ge x) \le K_2\exp(-K_1x^{\gamma_1})$ and $\max_{1 \le j \le p}\mathbb{P}(|\epsilon_{j,t}| \ge x) \le K_2\exp(-K_1x^{\gamma_2})$ for any $x > 0$ and $t = 1, \ldots, n$. It ensures the exponential upper bounds for the tail probabilities of the statistics concerned (see for example Lemma 1 in the Appendix), which makes our procedure work for p diverging at some exponential rate of n.

Condition 2 implies the bounded eigenvalues of $\boldsymbol{\Sigma}$ and $\boldsymbol{\Omega}$, which is commonly assumed in the literatures of high-dimensional data analysis. Condition 3 for the β-mixing coefficients of $\{\mathbf{y}_t\}$ is mild. Causal ARMA processes with continuous innovation distributions are β-mixing with exponentially decaying β_k. So are stationary Markov chains satisfying certain conditions. See Section 2.6.1 of Fan and Yao (2003) and the references therein. In fact, stationary GARCH models with finite second moments and continuous innovation distributions are also β-mixing with exponentially decaying β_k; see Proposition 12 of Carrasco and Chen (2002). If we only require $\sup_t \max_{1\le j\le p}\mathbb{P}(|y_{j,t}|>x)=O\{x^{-2(\nu+\iota)}\}$ and $\sup_t \max_{1\le j\le p}\mathbb{P}(|\epsilon_{j,t}|>x)=O\{x^{-2(\nu+\iota)}\}$ for any $x>0$ in Condition 1 and $\beta_k=O\{k^{-\nu(\nu+\iota)/(2\iota)}\}$ in Condition 3 for some $\nu>2$ and $\iota>0$, we can apply the Fuk–Nagaev-type inequalities to construct the upper bounds for the tail probabilities of the statistics if p diverges at some polynomial rate of n. We refer to Section 3.2 of Chang et al. (2018) for the implementation of the Fuk–Nagaev-type inequalities in such a scenario. The β-mixing condition can be replaced by the α-mixing condition, under which we can justify the proposed method for p diverging at some polynomial rate of n by using the Fuk–Nagaev-type inequalities. However, it remains an open problem to establish the relevant properties under α-mixing for p diverging at some exponential rate of n.

Proposition 1. *Let $s=\max_{1\le j\le p}|\boldsymbol{\alpha}_j|_0$ and select the tuning parameter λ_j in (7) satisfying $\lambda_j \asymp (n^{-1}\log p)^{1/2}$ for each $j=1,\ldots,p$. Under Conditions 1–3, if $s^2(\log p)^3 n^{-1}=o(1)$ and $\log p=o(n^{\varrho_1})$ for a positive constant ϱ_1 specified in the proof of this proposition in the Appendix, it holds that*

$$\widehat{\omega}_{j_1,j_2}-\omega_{j_1,j_2}=-\frac{\delta_{j_1,j_2}}{v_{j_1,j_1}v_{j_2,j_2}}+o_p\{(n\log p)^{-1/2}\},$$

where $\delta_{j_1,j_2}=n^{-1}\sum_{t=1}^n(\epsilon_{j_1,t}\epsilon_{j_2,t}-v_{j_1,j_2})$ for any j_1 and j_2, and $o_p\{(n\log p)^{-1/2}\}$ is a uniform higher order term.

We see from Proposition 1 that $\widehat{\omega}_{j_1,j_2}$ is centered at the true parameter ω_{j_1,j_2} with a standard deviation at the order $n^{-1/2}$. Since α_{j_1,j_2} is proportional to ω_{j_1,j_2}, it follows from $s^2(\log p)^3n^{-1}=o(1)$ that $\boldsymbol{\Omega}$ is sparse. When the maximum number of nonzero elements in each row of $\boldsymbol{\Omega}$ is of the order smaller than $n^{1/2}(\log p)^{-3/2}$, Proposition 1 holds even when p is of an exponential rate of n. Similar to the asymptotic expansion for $\widehat{\omega}_{j_1,j_2}$ in Proposition 1, Liu (2013) gave an asymptotic expansion for $-\widehat{v}_{j_1,j_2}$ with $j_1\neq j_2$. More specifically, with i.i.d. data, he showed that $-\widehat{v}_{j_1,j_2}=-\frac{b_{j_1,j_2}\omega_{j_1,j_2}}{\omega_{j_1,j_1}\omega_{j_2,j_2}}+\delta_{j_1,j_2}+R$ for δ_{j_1,j_2} specified in Proposition 1 and $b_{j_1,j_2}=\omega_{j_1,j_1}\widehat{v}_{j_1,j_1}+\omega_{j_2,j_2}\widehat{v}_{j_2,j_2}-1$, where R is a remainder term with the convergence rate faster than $n^{-1/2}$. It follows from the central limit theorem that $-n^{1/2}c_{j_1,j_2}(\widehat{v}_{j_1,j_2}-\frac{b_{j_1,j_2}\omega_{j_1,j_2}}{\omega_{j_1,j_1}\omega_{j_2,j_2}})$ converges to standard normal distribution with some suitable scale c_{j_1,j_2}, which indicates that $-n^{1/2}c_{j_1,j_2}\widehat{v}_{j_1,j_2}$ can be used as the testing statistic to test $\omega_{j_1,j_2}=0$ or not. Notice that $\widehat{v}_{j,j}=\omega_{j,j}^{-1}+O_p(n^{-1/2})$ which implies $b_{j_1,j_2}=1+O_p(n^{-1/2})$. Hence, the magnitude of $-n^{1/2}c_{j_1,j_2}\widehat{v}_{j_1,j_2}$ will be large if $\omega_{j_1,j_2}\neq 0$. This indicates that the asymptotic expansion given in Liu (2013) is enough for identifying non-zero entries of $\boldsymbol{\Omega}$. However, it is not enough for constructing the confidence interval for ω_{j_1,j_2} due to the fact that it does not contain the asymptotic expansion of $\widehat{\omega}_{j_1,j_2}$.

3.2. Confidence regions

Let $\boldsymbol{\Delta}=-n^{-1}\sum_{t=1}^n(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^{\mathrm{T}}-\mathbf{V})$. It follows from Proposition 1 that

$$\widehat{\boldsymbol{\Omega}}-\boldsymbol{\Omega}=\boldsymbol{\Pi}+\boldsymbol{\Upsilon}\ \text{for}\ \boldsymbol{\Pi}=\{\mathrm{diag}(\mathbf{V})\}^{-1}\boldsymbol{\Delta}\{\mathrm{diag}(\mathbf{V})\}^{-1},$$

where $|\boldsymbol{\Upsilon}|_\infty=o_p\{(n\log p)^{-1/2}\}$. Restricted on a given index set $\mathcal{S}$ with $r=|\mathcal{S}|$, we have

$$\widehat{\boldsymbol{\Omega}}_{\mathcal{S}}-\boldsymbol{\Omega}_{\mathcal{S}}=\boldsymbol{\Pi}_{\mathcal{S}}+\boldsymbol{\Upsilon}_{\mathcal{S}}. \tag{12}$$

Based on (12), we consider two kinds of confidence regions:

$$\begin{aligned}\mathcal{C}_{\mathcal{S},\alpha,1}&=\{\mathbf{a}\in\mathbb{R}^r: n^{1/2}|\widehat{\boldsymbol{\Omega}}_{\mathcal{S}}-\mathbf{a}|_\infty\le q_{\mathcal{S},\alpha,1}\},\\ \mathcal{C}_{\mathcal{S},\alpha,2}&=\{\mathbf{a}\in\mathbb{R}^r: n^{1/2}|\widehat{\mathbf{D}}^{-1}(\widehat{\boldsymbol{\Omega}}_{\mathcal{S}}-\mathbf{a})|_\infty\le q_{\mathcal{S},\alpha,2}\},\end{aligned} \tag{13}$$

where $\widehat{\mathbf{D}}$ is an $r\times r$ diagonal matrix, specified in Remark 5, of which the elements are the estimated standard deviations of the r components in $n^{1/2}(\widehat{\boldsymbol{\Omega}}_{\mathcal{S}}-\boldsymbol{\Omega}_{\mathcal{S}})$. Here $q_{\mathcal{S},\alpha,1}$ and $q_{\mathcal{S},\alpha,2}$ are two critical values to be determined. $\mathcal{C}_{\mathcal{S},\alpha,1}$ and $\mathcal{C}_{\mathcal{S},\alpha,2}$ represent the so-called "non-Studentized-type" and "Studentized-type" confidence regions for $\boldsymbol{\Omega}_{\mathcal{S}}$, respectively. The Studentized-type confidence regions perform better than the non-Studentized-type ones when the heteroscedasticity exists, however, the performance of the non-Studentized-type confidence regions is more stable when the sample size n is fairly small. See Chang et al. (2017a).

In the sequel, we mainly focus on estimating the critical value $q_{\mathcal{S},\alpha,1}$ in (13), as $q_{\mathcal{S},\alpha,2}$ can be estimated in the similar manner; see Remark 5. To determine $q_{\mathcal{S},\alpha,1}$, we need to first characterize the probabilistic behavior of $n^{1/2}|\widehat{\boldsymbol{\Omega}}_{\mathcal{S}}-\boldsymbol{\Omega}_{\mathcal{S}}|_\infty$. Since $\boldsymbol{\Upsilon}_{\mathcal{S}}$ is a higher order term, $n^{1/2}|\widehat{\boldsymbol{\Omega}}_{\mathcal{S}}-\boldsymbol{\Omega}_{\mathcal{S}}|_\infty$ will behave similarly as $n^{1/2}|\boldsymbol{\Pi}_{\mathcal{S}}|_\infty$ when n is large. Notice that each element of $n^{1/2}\boldsymbol{\Pi}_{\mathcal{S}}$ is asymptotically normal distributed. Following the idea of Chernozhukov et al. (2013), it can be proved that the limiting behavior of $n^{1/2}|\boldsymbol{\Pi}_{\mathcal{S}}|_\infty$ can be approximated by that of the L_∞-norm of a certain multivariate normal vector.

Please cite this article in press as: Chang J., et al., Confidence regions for entries of a large precision matrix. Journal of Econometrics (2018), https://doi.org/10.1016/j.jeconom.2018.03.020.

See Theorem 1. More specifically, for each t, let $\boldsymbol{\varsigma}_t$ be an r-dimensional vector whose jth element is $\frac{\epsilon_{\chi_1(j),t}\epsilon_{\chi_2(j),t}-v_{\chi(j)}}{v_{\chi_1(j),\chi_1(j)}v_{\chi_2(j),\chi_2(j)}}$ where $\boldsymbol{\chi}(\cdot)=\{\chi_1(\cdot),\chi_2(\cdot)\}$ is a bijective mapping from $\{1,\ldots,r\}$ to $\mathcal{S}$ such that $\boldsymbol{\Omega}_{\mathcal{S}}=\{\omega_{\boldsymbol{\chi}(1)},\ldots,\omega_{\boldsymbol{\chi}(r)}\}^{\mathrm{T}}$. Then, we have

$$\boldsymbol{\Pi}_{\mathcal{S}}=-\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{\varsigma}_t.$$

Denote by $\mathbf{W}$ the long-run covariance of $\{\boldsymbol{\varsigma}_t\}_{t=1}^n$, namely,

$$\mathbf{W}=\mathbb{E}\bigg\{\bigg(\frac{1}{n^{1/2}}\sum_{t=1}^{n}\boldsymbol{\varsigma}_t\bigg)\bigg(\frac{1}{n^{1/2}}\sum_{t=1}^{n}\boldsymbol{\varsigma}_t\bigg)^{\mathrm{T}}\bigg\}. \quad (14)$$

Let $\boldsymbol{\eta}_t=(\eta_{1,t},\ldots,\eta_{r,t})^{\mathrm{T}}$ where $\eta_{j,t}=\epsilon_{\chi_1(j),t}\epsilon_{\chi_2(j),t}-v_{\boldsymbol{\chi}(j)}$. Then $\mathbf{W}$ specified in (14) can be written as

$$\mathbf{W}=\mathbf{H}\mathbb{E}\bigg\{\bigg(\frac{1}{n^{1/2}}\sum_{t=1}^{n}\boldsymbol{\eta}_t\bigg)\bigg(\frac{1}{n^{1/2}}\sum_{t=1}^{n}\boldsymbol{\eta}_t\bigg)^{\mathrm{T}}\bigg\}\mathbf{H} \quad (15)$$

where $\mathbf{H}=\mathrm{diag}\{v^{-1}_{\chi_1(1),\chi_1(1)}v^{-1}_{\chi_2(1),\chi_2(1)},\ldots,v^{-1}_{\chi_1(r),\chi_1(r)}v^{-1}_{\chi_2(r),\chi_2(r)}\}$. To study the asymptotical distribution of the average of the temporally dependent sequence $\{\boldsymbol{\varsigma}_t\}_{t=1}^n$ and its long-run covariance $\mathbf{W}$, we introduce the following condition on $\{\boldsymbol{\eta}_t\}_{t=1}^n$.

Condition 4. *There exists constant $K_4>0$ such that*

$$\liminf_{b\to\infty}\inf_{1\le\ell\le n+1-b}\mathbb{E}\bigg(\bigg|\frac{1}{b^{1/2}}\sum_{t=\ell}^{\ell+b-1}\eta_{j,t}\bigg|^2\bigg)>K_4$$

for each $j=1,\ldots,r$.

Condition 4 is for the validity of the Gaussian approximation for dependent data. Under Conditions 1 and 3, Davydov inequality (Davydov, 1968) entails $\limsup_{b\to\infty}\sup_{1\le\ell\le n+1-b}\mathbb{E}(|b^{-1/2}\sum_{t=\ell}^{\ell+b-1}\eta_{j,t}|^2)<K_5$ for some universal constant $K_5>0$. Together with Condition 4, they match the requirements of Gaussian approximation imposed on the long-run covariance of $\{\eta_{j,t}\}_{t=\ell}^{\ell+b-1}$ for $j=1,\ldots,r$ and $\ell=1,\ldots,n+1-b$. See Theorem B.1 of Chernozhukov et al. (2014). If $\{\eta_{j,t}\}$ is stationary, $\mathbb{E}(|b^{-1/2}\sum_{t=\ell}^{\ell+b-1}\eta_{j,t}|^2)=\mathbb{E}(\eta_{j,1}^2)+\sum_{k=1}^{b-1}(1-kb^{-1})\mathrm{Cov}(\eta_{j,1},\eta_{j,1+k})$. Under the stationarity assumption on each sequence $\{\eta_{j,t}\}$, Condition 4 is equivalent to $\sum_{k=0}^{\infty}\mathrm{Cov}(\eta_{j,1},\eta_{j,1+k})>K_4$ for any $j=1,\ldots,r$. Now we are ready to state our main result.

Theorem 1. *Let $\boldsymbol{\xi}\sim N(\mathbf{0},\mathbf{W})$ for $\mathbf{W}$ specified in (14). Under the conditions of Proposition 1 and Condition 4, we have*

$$\sup_{x>0}\big|\mathbb{P}\big(n^{1/2}|\widehat{\boldsymbol{\Omega}}_{\mathcal{S}}-\boldsymbol{\Omega}_{\mathcal{S}}|_\infty>x\big)-\mathbb{P}(|\boldsymbol{\xi}|_\infty>x)\big|\to 0$$

as $n\to\infty$, provided that $s^2(\log p)^3n^{-1}=o(1)$ and $\log p=o(n^{\varrho_2})$ where $s=\max_{1\le j\le p}|\boldsymbol{\alpha}_j|_0$ and ϱ_2 is a positive constant specified in the proof of this theorem in the Appendix.

Remark 1. Theorem 1 shows that the Kolmogorov distance between the distributions of $n^{1/2}|\widehat{\boldsymbol{\Omega}}_{\mathcal{S}}-\boldsymbol{\Omega}_{\mathcal{S}}|_\infty$ and $|\boldsymbol{\xi}|_\infty$ converges to zero. More specifically, as shown in the proof of Theorem 1 in the Appendix, this convergence rate is $O(n^{-C})$ for some constant $C>0$ without requiring any structural assumption on the underlying covariance $\mathbf{W}$. Note that $n^{1/2}|\widehat{\boldsymbol{\Omega}}_{\mathcal{S}}-\boldsymbol{\Omega}_{\mathcal{S}}|_\infty$ may converge weakly to an extreme value distribution, which however requires some more stringent assumptions on the structure of $\mathbf{W}$. Furthermore the slow convergence to the extreme value distribution, i.e. typically slower than $O(n^{-C})$, entails an less accurate approximation than that implied by Theorem 1. We need to point out that there is also a requirement imposed on the diverging rate of $r=|\mathcal{S}|$ such as $\log r=o(n^C)$ for some constant $C>0$ in the proof of Theorem 1. Since $r\le p^2$, such requirement is satisfied automatically when the requirements on p in Theorem 1 are required.

Theorem 1 provides a guideline to approximate the distribution of $n^{1/2}|\widehat{\boldsymbol{\Omega}}_{\mathcal{S}}-\boldsymbol{\Omega}_{\mathcal{S}}|_\infty$. To implement it in practice, we need to propose an estimator for $\mathbf{W}$. Denote by $\boldsymbol{\Xi}$ the matrix sandwiched by $\mathbf{H}$'s on the right-hand side of (15), which is the long-run covariance of $\{\boldsymbol{\eta}_t\}_{t=1}^n$. Notice that $\widehat{v}_{j,j}$ defined in (10) is $n^{1/2}$-consistent to $v_{j,j}$, we can estimate $\mathbf{H}$ by

$$\widehat{\mathbf{H}}=\mathrm{diag}\big\{\widehat{v}^{-1}_{\chi_1(1),\chi_1(1)}\widehat{v}^{-1}_{\chi_2(1),\chi_2(1)},\ldots,\widehat{v}^{-1}_{\chi_1(r),\chi_1(r)}\widehat{v}^{-1}_{\chi_2(r),\chi_2(r)}\big\}. \quad (16)$$

Let $\widehat{\boldsymbol{\eta}}_t=(\widehat{\eta}_{1,t},\ldots,\widehat{\eta}_{r,t})^{\mathrm{T}}$ for $\widehat{\eta}_{j,t}=\widehat{\epsilon}_{\chi_1(j),t}\widehat{\epsilon}_{\chi_2(j),t}-\widehat{v}_{\boldsymbol{\chi}(j)}$, and define

$$\widehat{\boldsymbol{\Gamma}}_k=\begin{cases}\dfrac{1}{n}\displaystyle\sum_{t=k+1}^{n}\widehat{\boldsymbol{\eta}}_t\widehat{\boldsymbol{\eta}}_{t-k}^{\mathrm{T}}, & k\ge 0;\\[2ex] \dfrac{1}{n}\displaystyle\sum_{t=-k+1}^{n}\widehat{\boldsymbol{\eta}}_{t+k}\widehat{\boldsymbol{\eta}}_t^{\mathrm{T}}, & k<0.\end{cases}$$

Based on the $\widehat{\boldsymbol{\Gamma}}_k$'s, we propose a kernel estimator suggested by Andrews (1991) for $\boldsymbol{\Xi}$ as

$$\widehat{\boldsymbol{\Xi}} = \sum_{k=-n+1}^{n-1} \mathcal{K}\left(\frac{k}{S_n}\right)\widehat{\boldsymbol{\Gamma}}_k \tag{17}$$

where S_n is the bandwidth, $\mathcal{K}(\cdot)$ is a symmetric kernel function that is continuous at 0 and satisfying $\mathcal{K}(0) = 1$, $|\mathcal{K}(u)| \le 1$ for any $u \in \mathbb{R}$, and $\int_{-\infty}^{\infty}\mathcal{K}^2(u)du < \infty$. Given $\widehat{\mathbf{H}}$ and $\widehat{\boldsymbol{\Xi}}$ defined respectively in (16) and (17), an estimator for $\mathbf{W}$ is given by

$$\widehat{\mathbf{W}} = \widehat{\mathbf{H}}\widehat{\boldsymbol{\Xi}}\widehat{\mathbf{H}}. \tag{18}$$

Theorem 2 shows that we can approximate the distribution of $n^{1/2}|\widehat{\boldsymbol{\Omega}}_{\mathcal{S}} - \boldsymbol{\Omega}_{\mathcal{S}}|_\infty$ by that of $|\widehat{\boldsymbol{\xi}}|_\infty$ for $\widehat{\boldsymbol{\xi}} \sim N(\mathbf{0}, \widehat{\mathbf{W}})$.

Remark 2. Andrews (1991) systematically investigated the theoretical properties for the kernel estimator for the long-run covariance matrix when p is fixed. It shows that the Quadratic Spectral kernel

$$\mathcal{K}_{QS}(u) = \frac{25}{12\pi^2u^2}\left\{\frac{\sin(6\pi u/5)}{6\pi u/5} - \cos(6\pi u/5)\right\}$$

is optimal in the sense of minimizing the asymptotic truncated mean square error. In our numerical work, we adopt this quadratic spectral kernel with the data-driven selected bandwidth proposed in Section 6 of Andrews (1991), though our theoretical analysis applies to general kernel functions. Both our theoretical and simulation results show that this kernel estimator $\widehat{\boldsymbol{\Xi}}$ still works when p is large in relation to n. There also exist other estimation methods for long-run covariances, including the estimation utilizing moving block bootstrap (Lahiri, 2003; Nordman and Lahiri, 2005). Also see den Haan and Levin (1997) and Kiefer et al. (2000). Compared to those methods, an added advantage of using the kernel estimator is the computational efficiency in terms of both speed and storage space especially when p is large; see Remark 4. When the observations are i.i.d., a special case of our setting, $\mathbf{W}$ as in (14) is degenerated to $\mathbb{E}(\boldsymbol{\varsigma}_t\boldsymbol{\varsigma}_t^{\mathrm{T}})$, the marginal covariance of $\boldsymbol{\varsigma}_t$. We can apply $n^{-1}\sum_{t=1}^n\widehat{\boldsymbol{\eta}}_t\widehat{\boldsymbol{\eta}}_t^{\mathrm{T}}$ to estimate $\boldsymbol{\Xi}$, and then use $\widehat{\mathbf{H}}(n^{-1}\sum_{t=1}^n\widehat{\boldsymbol{\eta}}_t\widehat{\boldsymbol{\eta}}_t^{\mathrm{T}})\widehat{\mathbf{H}}$ to estimate $\mathbf{W}$ with $\widehat{\mathbf{H}}$ as in (16).

Theorem 2. *Let $\widehat{\boldsymbol{\xi}} \sim N(\mathbf{0}, \widehat{\mathbf{W}})$ for $\widehat{\mathbf{W}}$ specified in (18). Assume the kernel function $\mathcal{K}(\cdot)$ satisfy $|\mathcal{K}(x)| \asymp |x|^{-\tau}$ as $x \to \infty$ for some $\tau > 1$, and the bandwidth $S_n \asymp n^\rho$ for some $0 < \rho < \min\{\frac{\tau-1}{3\tau}, \frac{\gamma_3}{2\gamma_3+1}\}$ and γ_3 in Condition 3. Under the conditions of Theorem 1, it holds that*

$$\sup_{x>0}\left|\mathbb{P}\left(n^{1/2}|\widehat{\boldsymbol{\Omega}}_{\mathcal{S}} - \boldsymbol{\Omega}_{\mathcal{S}}|_\infty > x\right) - \mathbb{P}\left(|\widehat{\boldsymbol{\xi}}|_\infty > x|\mathcal{Y}_n\right)\right| \xrightarrow{p} 0$$

as $n \to \infty$, provided that $s^2(\log p)n^{-1}\max\{S_n^2, (\log p)^2\} = o(1)$ and $\log p = o(n^{\varrho_3})$ where $s = \max_{1\le j\le p}|\boldsymbol{\alpha}_j|_0$, ϱ_3 is a positive constant specified in the proof of this theorem in the Appendix, *and $\mathcal{Y}_n = \{\mathbf{y}_1, \ldots, \mathbf{y}_n\}$.*

Remark 3. Theorem 2 is valid for any $\widehat{\mathbf{W}}$ satisfying $|\widehat{\mathbf{W}} - \mathbf{W}|_\infty = o_p(1)$; see Chernozhukov et al. (2013). Different from the common practice in estimating large covariance matrices, we construct $\widehat{\mathbf{W}}$ in (18) without imposing any structural assumptions on $\mathbf{W}$.

In practice, we approximate the distribution of $|\widehat{\boldsymbol{\xi}}|_\infty$ by Monto Carlo simulation. Specifically, let $\widehat{\boldsymbol{\xi}}_1, \ldots, \widehat{\boldsymbol{\xi}}_M$ be i.i.d. r-dimensional random vectors drawn from $N(\mathbf{0}, \widehat{\mathbf{W}})$. Then the conditional distribution of $|\widehat{\boldsymbol{\xi}}|_\infty$ given $\mathcal{Y}_n$ can be approximated by the empirical distribution of $\{|\widehat{\boldsymbol{\xi}}_1|_\infty, \ldots, |\widehat{\boldsymbol{\xi}}_M|_\infty\}$, namely,

$$\widehat{F}_M(x) = \frac{1}{M}\sum_{m=1}^M \mathbb{I}\left\{|\widehat{\boldsymbol{\xi}}_m|_\infty \le x\right\}.$$

Then, $q_{\mathcal{S},\alpha,1}$ specified in (13) can be estimated by

$$\widehat{q}_{\mathcal{S},\alpha,1} = \inf\{x \in \mathbb{R} : \widehat{F}_M(x) \ge 1 - \alpha\}. \tag{19}$$

To improve computational efficiency, we propose the following Kernel based Multiplier Bootstrap (KMB) procedure to generate $\widehat{\boldsymbol{\xi}} \sim N(\mathbf{0}, \widehat{\mathbf{W}})$, which is much more efficient when r is large.

Step 1. Generate $\mathbf{g} = (g_1, \ldots, g_n)^{\mathrm{T}}$ from $N(\mathbf{0}, \mathbf{A})$, where $\mathbf{A}$ is the $n \times n$ matrix with $\mathcal{K}(|i - j|/S_n)$ as its (i, j)th element.
Step 2. Let $\widehat{\boldsymbol{\xi}} = n^{-1/2}\widehat{\mathbf{H}}(\sum_{t=1}^n g_t\widehat{\boldsymbol{\eta}}_t)$, where $\widehat{\mathbf{H}}$ is defined in (16).

Remark 4. The standard approach to draw a random vector $\widehat{\boldsymbol{\xi}} \sim N(\mathbf{0}, \widehat{\mathbf{W}})$ consists of three steps: (i) perform the Cholesky decomposition on the $r \times r$ matrix $\widehat{\mathbf{W}} = \mathbf{L}^{\mathrm{T}}\mathbf{L}$, (ii) generate r independent standard normal random variables $\mathbf{z} = (z_1, \ldots, z_r)^{\mathrm{T}}$, (iii) perform transformation $\widehat{\boldsymbol{\xi}} = \mathbf{L}^{\mathrm{T}}\mathbf{z}$. Thus, it requires to store matrix $\widehat{\mathbf{W}}$ and $\{\widehat{\boldsymbol{\eta}}_t\}_{t=1}^n$, which amounts to the storage costs $O(r^2)$ and $O(rn)$, respectively. The computational complexity is $O(r^2n + r^3)$, mainly due to computing $\widehat{\mathbf{W}}$ and the Cholesky decomposition. Note that r could be in the order of $O(p^2)$. In contrast the KMB scheme described above only needs to store $\{\widehat{\boldsymbol{\eta}}_t\}_{t=1}^n$ and $\mathbf{A}$, and draw an n-dimensional random vector $\mathbf{g} \sim N(\mathbf{0}, \mathbf{A})$ in each parametric bootstrap sample. This amounts to total storage cost $O(rn + n^2)$. More significantly, the computational complexity is only $O(n^3)$ which is independent of r and p.

Remark 5. For the Studentized-type confidence regions $\mathcal{C}_{\mathcal{S},\alpha,2}$ defined in (13), we can choose the diagonal matrix $\widehat{\mathbf{D}} = \{\mathrm{diag}(\widehat{\mathbf{W}})\}^{1/2}$ for $\widehat{\mathbf{W}}$ specified in (18). Correspondingly, for $\widehat{\boldsymbol{\xi}} \sim N(\mathbf{0}, \widehat{\mathbf{D}}^{-1}\widehat{\mathbf{W}}\widehat{\mathbf{D}}^{-1})$, it can be proved, in the similar manner as that for Theorem 2, that

$$\sup_{x>0}\left|\mathbb{P}\left\{n^{1/2}|\widehat{\mathbf{D}}^{-1}(\widehat{\boldsymbol{\Omega}}_{\mathcal{S}} - \boldsymbol{\Omega}_{\mathcal{S}})|_\infty > x\right\} - \mathbb{P}(|\widehat{\boldsymbol{\xi}}|_\infty > x|\mathcal{Y}_n)\right| \xrightarrow{p} 0 \ \text{as} \ n \to \infty.$$

To approximate the distribution of $n^{1/2}|\widehat{\mathbf{D}}^{-1}(\widehat{\boldsymbol{\Omega}}_{\mathcal{S}} - \boldsymbol{\Omega}_{\mathcal{S}})|_\infty$, we only need to replace the Step 2 in the KMB procedure by

Step 2′. Let $\widehat{\boldsymbol{\xi}} = n^{-1/2}\widehat{\mathbf{D}}^{-1}\widehat{\mathbf{H}}(\sum_{t=1}^n g_t\widehat{\boldsymbol{\eta}}_t)$ where $\widehat{\mathbf{H}}$ is defined in (16).

Based on the i.i.d. random vectors $\widehat{\boldsymbol{\xi}}_1, \ldots, \widehat{\boldsymbol{\xi}}_M$ generated by Steps 1 and 2′, we can estimate $q_{\mathcal{S},\alpha,2}$ via $\widehat{q}_{\mathcal{S},\alpha,2}$, which is calculated the same as $\widehat{q}_{\mathcal{S},\alpha,1}$ in (19). We call the procedure combining Steps 1 and 2′ as Studentized Kernel based Multiplier Bootstrap (SKMB).

4. Applications

4.1. Testing structures of $\boldsymbol{\Omega}$

Many statistical applications require to explore or to detect some specific structures of the precision matrix $\boldsymbol{\Omega} = (\omega_{j_1,j_2})_{p\times p}$. Given an index set $\mathcal{S}$ of interest and a set of pre-specified constants $\{c_{j_1,j_2}\}$, we test the hypotheses

$$H_0: \omega_{j_1,j_2} = c_{j_1,j_2} \ \text{for any } (j_1,j_2) \in \mathcal{S} \quad \text{vs.} \quad H_1: \omega_{j_1,j_2} \neq c_{j_1,j_2} \ \text{for some } (j_1,j_2) \in \mathcal{S}.$$

Recall that $\boldsymbol{\chi}(\cdot) = \{\chi_1(\cdot), \chi_2(\cdot)\}$ is a bijective mapping from $\{1, \ldots, r\}$ to $\mathcal{S}$ such that $\boldsymbol{\Omega}_{\mathcal{S}} = \{\omega_{\boldsymbol{\chi}(1)}, \ldots, \omega_{\boldsymbol{\chi}(r)}\}^{\mathrm{T}}$. Let $r = |\mathcal{S}|$ and $\mathbf{c} = \{c_{\boldsymbol{\chi}(1)}, \ldots, c_{\boldsymbol{\chi}(r)}\}^{\mathrm{T}}$. A usual choice of $\mathbf{c}$ is the zero vector, corresponding to the test for non-zero structures of $\boldsymbol{\Omega}$. Given a prescribed level $\alpha \in (0, 1)$, define $\Psi_\alpha = \mathbb{I}\{\mathbf{c} \notin \mathcal{C}_{\mathcal{S},1-\alpha,1}\}$ for $\mathcal{C}_{\mathcal{S},1-\alpha,1}$ specified in (13). Then, we reject the null hypothesis H_0 at level α if $\Psi_\alpha = 1$. This procedure is equivalent to the test based on the L_∞-type statistic $n^{1/2}|\widehat{\boldsymbol{\Omega}}_{\mathcal{S}} - \mathbf{c}|_\infty$ that rejects H_0 if $n^{1/2}|\widehat{\boldsymbol{\Omega}}_{\mathcal{S}} - \mathbf{c}|_\infty > \widehat{q}_{\mathcal{S},1-\alpha,1}$. The L_∞-type statistics are widely used in testing high-dimensional means and covariances. See, for example, Cai et al. (2013) and Chang et al. (2017a, b). The following corollary gives the empirical size and power of the proposed testing procedure Ψ_α.

Corollary 1. *Assume conditions of Theorem 2 hold. It holds that:* (i) $\mathbb{P}_{H_0}(\Psi_\alpha = 1) \to \alpha$ *as* $n \to \infty$; (ii) *if* $\max_{(j_1,j_2)\in\mathcal{S}}|\omega_{j_1,j_2} - c_{j_1,j_2}| \geq C(n^{-1}\log p)^{1/2}\max_{1\leq j\leq r} w_{j,j}^{1/2}$ *where* $w_{j,j}$ *is the jth component in the diagonal of* $\mathbf{W}$ *defined in* (14), *and* C *is a constant larger than* $\sqrt{2}$, *then* $\mathbb{P}_{H_1}(\Psi_\alpha = 1) \to 1$ *as* $n \to \infty$.

Corollary 1 implies that the empirical size of the proposed testing procedure Ψ_α will converge to the nominal level α under H_0. The condition $\max_{(j_1,j_2)\in\mathcal{S}}|\omega_{j_1,j_2} - c_{j_1,j_2}| \geq C(n^{-1}\log p)^{1/2}\max_{1\leq j\leq r} w_{j,j}^{1/2}$ specifies the maximal deviation of the precision matrix from the null hypothesis $H_0 : \omega_{j_1,j_2} = c_{j_1,j_2}$ for any $(j_1, j_2) \in \mathcal{S}$, which is a commonly used condition for studying the power of the L_∞-type test. See Cai et al. (2013) and Chang et al. (2017a, b). Corollary 1 shows that the power of the proposed test Ψ_α will approach 1 if such condition holds for some constant $C > \sqrt{2}$. A "Studentized-type" test can be similarly constructed via replacing $n^{1/2}|\widehat{\boldsymbol{\Omega}}_{\mathcal{S}} - \mathbf{c}|_\infty$ and $\widehat{q}_{\mathcal{S},1-\alpha,1}$ by $n^{1/2}|\widehat{\mathbf{D}}^{-1}(\widehat{\boldsymbol{\Omega}}_{\mathcal{S}} - \mathbf{c})|_\infty$ and $\widehat{q}_{\mathcal{S},1-\alpha,2}$ in (13), respectively.

4.2. Support recovering of $\boldsymbol{\Omega}$

In studying partial correlation networks or GGM, we are interested in identifying the edges between nodes. This is equivalent to recover the non-zero components in the associated precision matrix. Let $\mathcal{M}_0 = \{(j_1, j_2) : \omega_{j_1,j_2} \neq 0\}$ be the set of indices with non-zero precision coefficients. Choose $\mathcal{S} = \{1, \ldots, p\}^2$. Note that $\mathcal{C}_{\mathcal{S},\alpha,1}$ provides simultaneous confidence regions for all the entries of $\boldsymbol{\Omega}$. To recover the set $\mathcal{M}_0$ consistently, we choose those precision coefficients whose confidence intervals do not include zero. For any m-dimensional vector $\mathbf{u} = (u_1, \ldots, u_m)^{\mathrm{T}}$, let $\mathrm{supp}(\mathbf{u}) = \{j : u_j \neq 0\}$ be the support set of $\mathbf{u}$. Recall $\boldsymbol{\chi}(\cdot) = \{\chi_1(\cdot), \chi_2(\cdot)\}$ is a bijective mapping from $\{1, \ldots, r\}$ to $\mathcal{S}$ such that $\boldsymbol{\Omega}_{\mathcal{S}} = \{\omega_{\boldsymbol{\chi}(1)}, \ldots, \omega_{\boldsymbol{\chi}(r)}\}^{\mathrm{T}}$. For any $\alpha \in (0, 1)$, let

$$\widehat{\mathcal{M}}_{n,\alpha} = \left\{\boldsymbol{\chi}^{-1}(l) : l \in \bigcap_{\mathbf{u}\in\mathcal{C}_{\mathcal{S},1-\alpha,1}} \mathrm{supp}(\mathbf{u})\right\}$$

be the estimate of $\mathcal{M}_0$.

In our context, note that the false positive means estimating the zero ω_{j_1,j_2} as non-zero. Let FP be the number of false positive errors conducted by the estimated signal set $\widehat{\mathcal{M}}_{n,\alpha}$. Let the family wise error rate (FWER) be the probability of conducting any false positive errors, namely, FWER $= \mathbb{P}(\mathrm{FP} > 0)$. See Hochberg and Tamhane (2009) for various types of error rates in multiple testing procedures. Notice that $\mathbb{P}(\mathrm{FP} > 0) \leq \mathbb{P}(\boldsymbol{\Omega}_{\mathcal{S}} \notin \mathcal{C}_{\mathcal{S},1-\alpha,1}) = \alpha\{1 + o(1)\}$. This shows that the proposed method is able to control family wise error rate at level α for any $\alpha \in (0, 1)$. The following corollary further shows the consistency of $\widehat{\mathcal{M}}_{n,\alpha}$.

Please cite this article in press as: Chang J., et al., Confidence regions for entries of a large precision matrix. Journal of Econometrics (2018), https://doi.org/10.1016/j.jeconom.2018.03.020.

Corollary 2. *Assume conditions of Theorem 2 hold, and the signals satisfy* $\min_{(j_1,j_2)\in\mathcal{M}_0}|\omega_{j_1,j_2}| \geq C(n^{-1}\log p)^{1/2}\max_{1\leq j\leq r} w_{j,j}^{1/2}$ *where* $w_{j,j}$ *is the jth component in the diagonal of* $\mathbf{W}$ *defined in (14), and C is a constant larger than* $\sqrt{2}$*. Selecting* $\alpha \to 0$ *such that* $1/\alpha = o(p)$*, it holds that* $\mathbb{P}(\widehat{\mathcal{M}}_{n,\alpha} = \mathcal{M}_0) \to 1$ *as* $n \to \infty$*.*

From Corollary 2, we see that the selected set $\widehat{\mathcal{M}}_{n,\alpha}$ can identify the true set $\mathcal{M}_0$ consistently if the minimum signal strength satisfies $\min_{(j_1,j_2)\in\mathcal{M}_0}|\omega_{j_1,j_2}| \geq C(n^{-1}\log p)^{1/2}\max_{1\leq j\leq r} w_{j,j}^{1/2}$ for some constant $C > \sqrt{2}$. Notice from Corollary 1 that only the maximum signal is required in the power analysis of the proposed testing procedure. Compared to signal detection, signal recovery is a more challenging problem. The full support recovery of $\boldsymbol{\Omega}$ requires all non-zero $|\omega_{l_1,l_2}|$ larger than a specific level. Similarly, we can also define $\widehat{\mathcal{M}}_{n,\alpha}$ via replacing $\mathcal{C}_{\mathcal{S},1-\alpha,1}$ by its "Studentized-type" analogue $\mathcal{C}_{\mathcal{S},1-\alpha,2}$ in (13).

5. Numerical study

In this section, we evaluate the performance of the proposed KMB and SKMB procedures in finite samples. Let $\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_n$ be i.i.d. p-dimensional samples from $N(\mathbf{0}, \boldsymbol{\Sigma})$. The observed data were generated from the model $\mathbf{y}_1 = \boldsymbol{\varepsilon}_1$ and $\mathbf{y}_t = \rho\mathbf{y}_{t-1} + (1-\rho^2)^{1/2}\boldsymbol{\varepsilon}_t$ for $t \geq 2$. The parameter ρ was set to be 0 and 0.3, which captures the temporal dependence among observations. We chose the sample size $n = 150$ and 300, and the dimension $p = 100$, 500 and 1500 in the simulation. Let $\boldsymbol{\Sigma} = \{\mathrm{diag}(\boldsymbol{\Sigma}_*^{-1})\}^{1/2}\boldsymbol{\Sigma}_*\{\mathrm{diag}(\boldsymbol{\Sigma}_*^{-1})\}^{1/2}$ based on a positive definite matrix $\boldsymbol{\Sigma}_*$. The following two settings were considered for $\boldsymbol{\Sigma}_* = (\sigma^*_{j_1,j_2})_{1\leq j_1,j_2\leq p}$.

A. Let $\sigma^*_{j_1,j_2} = 0.5^{|j_1-j_2|}$ for any $1 \leq j_1, j_2 \leq p$.
B. Let $\sigma^*_{j,j} = 1$ for any $j = 1, \ldots, p$, $\sigma^*_{j_1,j_2} = 0.5$ for $5(h-1)+1 \leq j_1 \neq j_2 \leq 5h$, where $h = 1, \ldots, p/5$, and $\sigma^*_{j_1,j_2} = 0$ otherwise.

Structures A and B lead to, respectively, the banded and block diagonal structures for the precision matrix $\boldsymbol{\Omega} = \boldsymbol{\Sigma}^{-1}$. Note that, based on such defined covariance $\boldsymbol{\Sigma}$, the diagonal elements of the precision matrix are unit. For each of the precision matrices, we considered two choices for the index set $\mathcal{S}$: (i) all zero components of $\boldsymbol{\Omega}$, i.e. $\mathcal{S} = \{(j_1,j_2) : \omega_{j_1,j_2} = 0\}$, and (ii) all the components excluded the ones on the main diagonal, i.e. $\mathcal{S} = \{(j_1,j_2) : j_1 \neq j_2\}$. Notice that the sets of all zero components in $\boldsymbol{\Omega}$ for structures A and B are $\{(j_1,j_2) : |j_1 - j_2| > 1\}$ and $\cap_{h=1}^{p/5}\{(j_1,j_2) : 5(h-1)+1 \leq j_1, j_2 \leq 5h\}^c$, respectively. As we illustrate in the footnote,[1] the index sets $\mathcal{S}$ in the setting (i) and (ii) mimic, respectively, the homogeneous and heteroscedastic cases for the variances of $n^{1/2}(\widehat{\omega}_{j_1,j_2} - \omega_{j_1,j_2})$ among $(j_1,j_2) \in \mathcal{S}$.

For each of the cases above, we examined the accuracy of the proposed KMB and SKMB approximations to the distributions of the non-Studentized-type statistic $n^{1/2}|\widehat{\boldsymbol{\Omega}}_{\mathcal{S}} - \boldsymbol{\Omega}_{\mathcal{S}}|_\infty$ and the Studentized-type statistic $n^{1/2}|\widehat{\mathbf{D}}^{-1}(\widehat{\boldsymbol{\Omega}}_{\mathcal{S}} - \boldsymbol{\Omega}_{\mathcal{S}})|_\infty$, respectively. Denote by $F_{1n}(\cdot)$ and $F_{2n}(\cdot)$ the distribution functions of $n^{1/2}|\widehat{\boldsymbol{\Omega}}_{\mathcal{S}} - \boldsymbol{\Omega}_{\mathcal{S}}|_\infty$ and $n^{1/2}|\widehat{\mathbf{D}}^{-1}(\widehat{\boldsymbol{\Omega}}_{\mathcal{S}} - \boldsymbol{\Omega}_{\mathcal{S}})|_\infty$, respectively. In each of the 1000 independent repetitions, we first draw a sample with size n following the above discussed data generating mechanism, and then computed the associated values of $n^{1/2}|\widehat{\boldsymbol{\Omega}}_{\mathcal{S}} - \boldsymbol{\Omega}_{\mathcal{S}}|_\infty$ and $n^{1/2}|\widehat{\mathbf{D}}^{-1}(\widehat{\boldsymbol{\Omega}}_{\mathcal{S}} - \boldsymbol{\Omega}_{\mathcal{S}})|_\infty$ in this sample. Since $F_{1n}(\cdot)$ and $F_{2n}(\cdot)$ are unknown, we used the empirical distributions of $n^{1/2}|\widehat{\boldsymbol{\Omega}}_{\mathcal{S}} - \boldsymbol{\Omega}_{\mathcal{S}}|_\infty$ and $n^{1/2}|\widehat{\mathbf{D}}^{-1}(\widehat{\boldsymbol{\Omega}}_{\mathcal{S}} - \boldsymbol{\Omega}_{\mathcal{S}})|_\infty$ over 1000 repetitions, denoted as $F^*_{1n}(\cdot)$ and $F^*_{2n}(\cdot)$, to approximate them, respectively. For each repetition i, we applied the KMB and SKMB procedures to estimate the $100(1-\alpha)\%$ quantiles of $n^{1/2}|\widehat{\boldsymbol{\Omega}}_{\mathcal{S}} - \boldsymbol{\Omega}_{\mathcal{S}}|_\infty$ and $n^{1/2}|\widehat{\mathbf{D}}^{-1}(\widehat{\boldsymbol{\Omega}}_{\mathcal{S}} - \boldsymbol{\Omega}_{\mathcal{S}})|_\infty$, denoted as $\widehat{q}^{(i)}_{\mathcal{S},\alpha,1}$ and $\widehat{q}^{(i)}_{\mathcal{S},\alpha,2}$, respectively, with $M = 3000$, and then computed their associated empirical coverages $F^*_{1n}(\widehat{q}^{(i)}_{\mathcal{S},\alpha,1})$ and $F^*_{2n}(\widehat{q}^{(i)}_{\mathcal{S},\alpha,2})$. We considered $\alpha = 0.075$, 0.050 and 0.025 in the simulation. We report the averages and standard deviations of $\{F^*_{1n}(\widehat{q}^{(i)}_{\mathcal{S},\alpha,1})\}_{i=1}^{1000}$ and $\{F^*_{2n}(\widehat{q}^{(i)}_{\mathcal{S},\alpha,2})\}_{i=1}^{1000}$ in Tables 1–3. Due to the selection of the tuning parameter λ_j in (7) depends on the standard deviation of the error term $\epsilon_{j,t}$, we adopted the scaled Lasso (Sun and Zhang, 2012) in the simulation which can estimate the regression coefficients and the variance of the error simultaneously. The tuning parameters in scale Lasso were selected according to Ren et al. (2015).

It is worth noting that in order to accomplish the statistical computing for large p under the R environment in high speed, we programmed the generation of random numbers and most loops into C functions such that we utilized ".C()" routine to call those C functions from R. However, the computation of the two types of statistics involves the fitting of the p node-wise regressions. As a consequence, the simulation for large p still requires a large amount of computation time. In order to overcome this time-consuming issue, the computation in this numerical study was undertaken with the assistance of

[1] It follows from Proposition 1 that $\mathrm{Var}\{n^{1/2}(\widehat{\omega}_{j_1,j_2} - \omega_{j_1,j_2})\} = v_{j_1,j_1}^{-2}v_{j_2,j_2}^{-2}\mathrm{Var}\{n^{-1/2}\sum_{t=1}^n(\epsilon_{j_1,t}\epsilon_{j_2,t} - v_{j_1,j_2})\}\{1+o(1)\}$, where the term $o(1)$ holds uniformly over (j_1,j_2). Recall $\epsilon_{j,t} = -\boldsymbol{\alpha}_j^{\mathrm{T}}\mathbf{y}_t$ and $\mathbf{y}_t = (1-\rho^2)^{1/2}\sum_{k=0}^\infty\rho^k\boldsymbol{\varepsilon}_{t-k}$, if $\omega_{j_1,j_2} = 0$ which is equivalent to $v_{j_1,j_2} = 0$, then it holds that $\mathrm{Var}(n^{-1/2}\sum_{t=1}^n\epsilon_{j_1,t}\epsilon_{j_2,t}) = n^{-1}(1-\rho^2)^2\sum_{t_1,t_2=1}^n\mathbb{E}\{(\sum_{k=0}^\infty\rho^k\boldsymbol{\alpha}_{j_1}^{\mathrm{T}}\boldsymbol{\varepsilon}_{t_1-k})(\sum_{k=0}^\infty\rho^k\boldsymbol{\alpha}_{j_2}^{\mathrm{T}}\boldsymbol{\varepsilon}_{t_1-k})(\sum_{k=0}^\infty\rho^k\boldsymbol{\alpha}_{j_1}^{\mathrm{T}}\boldsymbol{\varepsilon}_{t_2-k})(\sum_{k=0}^\infty\rho^k\boldsymbol{\alpha}_{j_2}^{\mathrm{T}}\boldsymbol{\varepsilon}_{t_2-k})\}$. Since $\boldsymbol{\varepsilon}_t$'s are i.i.d., together with $v_{j_1,j_2} = 0$, we have $\mathbb{E}\{(\sum_{k=0}^\infty\rho^k\boldsymbol{\alpha}_{j_1}^{\mathrm{T}}\boldsymbol{\varepsilon}_{t_1-k})(\sum_{k=0}^\infty\rho^k\boldsymbol{\alpha}_{j_2}^{\mathrm{T}}\boldsymbol{\varepsilon}_{t_1-k})(\sum_{k=0}^\infty\rho^k\boldsymbol{\alpha}_{j_1}^{\mathrm{T}}\boldsymbol{\varepsilon}_{t_2-k})(\sum_{k=0}^\infty\rho^k\boldsymbol{\alpha}_{j_2}^{\mathrm{T}}\boldsymbol{\varepsilon}_{t_2-k})\} = \rho^{2t_2-2t_1}(1-\rho^2)^{-2}\mathbb{E}(\epsilon_{j_1,t}^2\epsilon_{j_2,t}^2)$ for any $t_2 \geq t_1$, which implies $\mathrm{Var}(n^{-1/2}\sum_{t=1}^n\epsilon_{j_1,t}\epsilon_{j_2,t}) = [1+2(1-\rho^2)^{-2}n^{-1}\{(n-1)\rho^{2n}-(n-2)\rho^{2n+2}-\rho^4\}]\mathbb{E}(\epsilon_{j_1,t}^2\epsilon_{j_2,t}^2)$ for any (j_1,j_2) such that $\omega_{j_1,j_2} = 0$. On the other hand, based on the Gaussian assumption, since $v_{j_1,j_2} = \mathbb{E}(\epsilon_{j_1,t}\epsilon_{j_2,t}) = 0$, we know the two normal distributed random variables $\epsilon_{j_1,t}$ and $\epsilon_{j_2,t}$ are independent, which leads to $\mathbb{E}(\epsilon_{j_1,t}^2\epsilon_{j_2,t}^2) = \mathbb{E}(\epsilon_{j_1,t}^2)\mathbb{E}(\epsilon_{j_2,t}^2) = v_{j_1,j_1}v_{j_2,j_2}$. Therefore, $\mathrm{Var}\{n^{1/2}(\widehat{\omega}_{j_1,j_2}-\omega_{j_1,j_2})\} = v_{j_1,j_1}^{-1}v_{j_2,j_2}^{-1}[1+2(1-\rho^2)^{-2}n^{-1}\{(n-1)\rho^{2n}-(n-2)\rho^{2n+2}-\rho^4\}]\{1+o(1)\}$ for any (j_1,j_2) such that $\omega_{j_1,j_2} = 0$. Notice that $\omega_{j,j} = 1$ in our setting for any j, then $v_{j,j} = \omega_{j,j}^{-1} = 1$. Hence, the variances of $n^{1/2}(\widehat{\omega}_{j_1,j_2} - \omega_{j_1,j_2})$ for any (j_1,j_2) such that $\omega_{j_1,j_2} = 0$ are almost identical.

Please cite this article in press as: Chang J., et al., Confidence regions for entries of a large precision matrix. Journal of Econometrics (2018), https://doi.org/10.1016/j.jeconom.2018.03.020.

Table 1
Averages of empirical coverages and their standard deviations (in parentheses) for $p = 100$.

Covariance structure	ρ	$1-\alpha$	$n = 150$				$n = 300$			
			$\mathcal{S} = \{(j_1,j_2) : \omega_{j_1,j_2} = 0\}$		$\mathcal{S} = \{(j_1,j_2) : j_1 \neq j_2\}$		$\mathcal{S} = \{(j_1,j_2) : \omega_{j_1,j_2} = 0\}$		$\mathcal{S} = \{(j_1,j_2) : j_1 \neq j_2\}$	
			KMB	SKMB	KMB	SKMB	KMB	SKMB	KMB	SKMB
A	0	0.925	0.963(0.013)	0.919(0.005)	0.885(0.022)	0.906(0.007)	0.954(0.011)	0.939(0.004)	0.915(0.014)	0.937(0.004)
		0.950	0.978(0.008)	0.949(0.007)	0.913(0.016)	0.941(0.007)	0.972(0.007)	0.956(0.002)	0.941(0.011)	0.954(0.002)
		0.975	0.991(0.004)	0.978(0.003)	0.950(0.014)	0.976(0.003)	0.985(0.003)	0.982(0.002)	0.963(0.006)	0.981(0.002)
	0.3	0.925	0.950(0.014)	0.888(0.014)	0.835(0.029)	0.875(0.014)	0.955(0.012)	0.920(0.009)	0.890(0.019)	0.916(0.010)
		0.950	0.967(0.010)	0.930(0.008)	0.876(0.025)	0.920(0.009)	0.973(0.007)	0.956(0.005)	0.924(0.013)	0.952(0.006)
		0.975	0.987(0.006)	0.966(0.004)	0.923(0.017)	0.958(0.005)	0.987(0.004)	0.979(0.003)	0.956(0.010)	0.978(0.003)
B	0	0.925	0.953(0.016)	0.927(0.005)	0.812(0.036)	0.874(0.008)	0.950(0.009)	0.931(0.003)	0.894(0.014)	0.917(0.004)
		0.950	0.973(0.010)	0.957(0.005)	0.863(0.028)	0.918(0.007)	0.969(0.008)	0.956(0.006)	0.925(0.013)	0.947(0.007)
		0.975	0.986(0.004)	0.979(0.002)	0.918(0.020)	0.965(0.004)	0.989(0.004)	0.981(0.004)	0.961(0.008)	0.978(0.004)
	0.3	0.925	0.950(0.019)	0.898(0.011)	0.772(0.039)	0.815(0.018)	0.950(0.011)	0.933(0.005)	0.880(0.017)	0.915(0.007)
		0.950	0.971(0.011)	0.930(0.007)	0.826(0.031)	0.873(0.012)	0.968(0.007)	0.956(0.003)	0.913(0.012)	0.943(0.004)
		0.975	0.987(0.004)	0.970(0.005)	0.885(0.021)	0.938(0.008)	0.985(0.004)	0.972(0.003)	0.943(0.010)	0.964(0.004)

Table 2
Averages of empirical coverages and their standard deviations (in parentheses) for $p = 500$.

Covariance structure	ρ	$1-\alpha$	$n = 150$				$n = 300$			
			$\mathcal{S} = \{(j_1,j_2) : \omega_{j_1,j_2} = 0\}$		$\mathcal{S} = \{(j_1,j_2) : j_1 \neq j_2\}$		$\mathcal{S} = \{(j_1,j_2) : \omega_{j_1,j_2} = 0\}$		$\mathcal{S} = \{(j_1,j_2) : j_1 \neq j_2\}$	
			KMB	SKMB	KMB	SKMB	KMB	SKMB	KMB	SKMB
A	0	0.925	0.967(0.006)	0.891(0.010)	0.872(0.017)	0.873(0.009)	0.971(0.003)	0.935(0.003)	0.924(0.008)	0.799(0.003)
		0.950	0.978(0.004)	0.934(0.007)	0.903(0.011)	0.923(0.009)	0.977(0.002)	0.954(0.002)	0.939(0.006)	0.822(0.004)
		0.975	0.987(0.003)	0.975(0.003)	0.933(0.012)	0.968(0.004)	0.983(0.002)	0.977(0.002)	0.956(0.004)	0.856(0.004)
	0.3	0.925	0.961(0.010)	0.871(0.010)	0.786(0.027)	0.833(0.011)	0.973(0.004)	0.937(0.005)	0.867(0.011)	0.905(0.007)
		0.950	0.979(0.006)	0.918(0.010)	0.842(0.021)	0.890(0.011)	0.982(0.004)	0.959(0.003)	0.899(0.011)	0.934(0.004)
		0.975	0.991(0.004)	0.966(0.005)	0.890(0.014)	0.949(0.006)	0.991(0.001)	0.973(0.003)	0.936(0.007)	0.950(0.003)
B	0	0.925	0.961(0.007)	0.884(0.009)	0.713(0.027)	0.746(0.015)	0.966(0.006)	0.921(0.005)	0.884(0.011)	0.814(0.007)
		0.950	0.974(0.004)	0.934(0.008)	0.780(0.030)	0.831(0.015)	0.980(0.003)	0.938(0.004)	0.915(0.009)	0.840(0.006)
		0.975	0.985(0.003)	0.974(0.004)	0.869(0.019)	0.912(0.010)	0.988(0.002)	0.970(0.003)	0.952(0.006)	0.887(0.007)
	0.3	0.925	0.954(0.007)	0.856(0.014)	0.641(0.034)	0.673(0.019)	0.964(0.005)	0.928(0.004)	0.853(0.016)	0.850(0.007)
		0.950	0.968(0.006)	0.908(0.008)	0.716(0.036)	0.767(0.018)	0.979(0.004)	0.950(0.003)	0.900(0.012)	0.889(0.006)
		0.975	0.983(0.004)	0.954(0.005)	0.821(0.028)	0.878(0.014)	0.988(0.002)	0.971(0.002)	0.940(0.010)	0.925(0.005)

Table 3
Averages of empirical coverages and their standard deviations (in parentheses) for $p = 1500$.

Covariance structure	ρ	$1-\alpha$	$n = 150$				$n = 300$			
			$\mathcal{S} = \{(j_1,j_2) : \omega_{j_1,j_2} = 0\}$		$\mathcal{S} = \{(j_1,j_2) : j_1 \neq j_2\}$		$\mathcal{S} = \{(j_1,j_2) : \omega_{j_1,j_2} = 0\}$		$\mathcal{S} = \{(j_1,j_2) : j_1 \neq j_2\}$	
			KMB	SKMB	KMB	SKMB	KMB	SKMB	KMB	SKMB
A	0	0.925	0.976(0.005)	0.854(0.013)	0.826(0.017)	0.834(0.013)	0.979(0.002)	0.959(0.003)	0.913(0.009)	0.948(0.004)
		0.950	0.987(0.003)	0.908(0.010)	0.866(0.011)	0.892(0.010)	0.991(0.002)	0.974(0.001)	0.945(0.007)	0.963(0.001)
		0.975	0.991(0.002)	0.954(0.005)	0.903(0.009)	0.944(0.006)	0.997(0.001)	0.987(0.003)	0.967(0.003)	0.979(0.003)
	0.3	0.925	0.967(0.010)	0.823(0.013)	0.674(0.031)	0.758(0.016)	0.981(0.002)	0.951(0.004)	0.822(0.011)	0.933(0.004)
		0.950	0.983(0.004)	0.887(0.011)	0.754(0.030)	0.840(0.012)	0.987(0.002)	0.972(0.004)	0.861(0.012)	0.958(0.005)
		0.975	0.994(0.002)	0.952(0.010)	0.841(0.019)	0.922(0.011)	0.996(0.001)	0.988(0.002)	0.926(0.010)	0.978(0.003)
B	0	0.925	0.964(0.008)	0.852(0.013)	0.638(0.031)	0.631(0.019)	0.973(0.004)	0.944(0.005)	0.882(0.010)	0.912(0.006)
		0.950	0.981(0.004)	0.915(0.008)	0.729(0.031)	0.738(0.021)	0.987(0.003)	0.967(0.003)	0.915(0.009)	0.946(0.004)
		0.975	0.991(0.002)	0.961(0.007)	0.831(0.017)	0.860(0.015)	0.995(0.001)	0.984(0.001)	0.952(0.006)	0.968(0.003)
	0.3	0.925	0.958(0.008)	0.781(0.025)	0.528(0.047)	0.417(0.031)	0.978(0.003)	0.930(0.006)	0.813(0.015)	0.867(0.010)
		0.950	0.977(0.006)	0.870(0.009)	0.643(0.040)	0.564(0.023)	0.985(0.002)	0.956(0.005)	0.866(0.013)	0.912(0.009)
		0.975	0.989(0.002)	0.939(0.007)	0.787(0.031)	0.737(0.023)	0.997(0.001)	0.980(0.002)	0.932(0.011)	0.954(0.005)

the supercomputer Raijin at the NCI National Facility systems supported by the Australian Government. The supercomputer Raijin comprises 57,864 cores, which helped us parallel process a large number of simulations simultaneously.

From Tables 1–3, we observe that, for both KMB and SKMB procedures, the overall differences between the empirical coverage rates and the corresponding nominal levels are small, which demonstrates that the KMB and SKMB procedures can provide accurate approximations to the distributions of $n^{1/2}|\widehat{\boldsymbol{\Omega}}_{\mathcal{S}} - \boldsymbol{\Omega}_{\mathcal{S}}|_\infty$ and $n^{1/2}|\widehat{\mathbf{D}}^{-1}(\widehat{\boldsymbol{\Omega}}_{\mathcal{S}} - \boldsymbol{\Omega}_{\mathcal{S}})|_\infty$, respectively. Also note that the coverage rates improve as n increases. And, our results are robust to the temporal dependence parameter ρ, which indicates the proposed procedures are adaptive to time dependent observations.

Please cite this article in press as: Chang J., et al., Confidence regions for entries of a large precision matrix. Journal of Econometrics (2018), https://doi.org/10.1016/j.jeconom.2018.03.020.

Comparing the simulation results indicated by KMB and SKMB in the category $\mathcal{S} = \{(j_1, j_2) : j_1 \neq j_2\}$ of Tables 1–3, when the dimension is less than the sample size ($p = 100, n = 150, 300$), we can see that the SKMB procedure has better accuracy than the KMB procedure if the heteroscedastic issue exists. This finding also exists when the dimension is over the sample size and both of them are large ($n = 300, p = 1500$). For the homogeneous case $\mathcal{S} = \{(j_1, j_2) : \omega_{j_1,j_2} = 0\}$, the KMB procedure provides better accuracy than the SKMB procedure when sample size is small ($n = 150$). However, when the sample size becomes larger ($n = 300$), the accuracy of the SKMB procedure can be significantly improved and it will outperform the KMB procedure. The phenomenon that the SKMB procedure sometimes cannot beat the KMB procedure might be caused by incorporating the estimated standard deviations of $\widehat{\omega}_{j_1,j_2}$'s in the denominator of the Studentized-type statistic, which suffers from high variability when the sample size is small. The simulation results suggest us that: (i) when the dimension is less than the sample size or both the dimension and the sample size are very large, the SKMB procedure should be used to construct the confidence regions of $\boldsymbol{\Omega}_{\mathcal{S}}$ if the heteroscedastic issue exists; (ii) if the sample size is small, and we have some previous information that there does not exist heteroscedastic issue, then the KMB procedure should be used to construct the confidence regions of $\boldsymbol{\Omega}_{\mathcal{S}}$. However, even in the homogeneous case, the SKMB procedure should still be employed when the sample size is large.

6. Real data analysis

In this section, we follow Example 3 in Section 2 to study the partial correlation networks of the Standard and Poors (S&P) 500 Component Stocks in 2005 (252 trading days, preceding the crisis) and in 2008 (253 trading days, during the crisis), respectively. The reason to analyze those two periods is to understand the structure and dynamic of financial networks affected by the global financial crisis (Schweitzer et al., 2009). Aït-Sahalia and Xiu (2015) analyzed the data in 2005 and 2008 as well in order to investigate the influence of the financial crisis.

We analyzed the data from http://quote.yahoo.com/ via the R package `tseries`, which contains the daily closing prices of S&P 500 stocks. The R command `get.hist.quote` can be used to acquire the data. We kept 402 stocks in our analysis whose closing prices were capable of being downloaded by the R command and did not have any missing values during 2005 and 2008. Let $y_{j,t}$ be the jth stock price at day t. We considered the log return of the stocks, which is defined by $\log(y_{j,t}) - \log(y_{j,t-1})$. As kindly pointed out by a referee that the log return data usually exhibit volatility clustering, we utilized the R package `fGarch` to obtain the conditional standard deviation for the mean centered log return of each stock via fitting a GARCH(1,1) model, and then we standardized the log return by its mean and conditional standard deviation. Ultimately, we had the standardized log returns $\mathbf{R}_t = (R_{1,t}, \ldots, R_{402,t})^{\mathrm{T}}$ of all the 402 assets at day t.

Let $\boldsymbol{\Omega} = (\omega_{j_1,j_2})_{p\times p}$ be the precision matrix of $\mathbf{R}_t$. By the relationship between partial correlation and precision matrix, the partial correlation network can be constructed by the non-zero precision coefficients ω_{j_1,j_2} as demonstrated in Example 3 in Section 2. To learn the structures of $\boldsymbol{\Omega}$, we focused on the Global Industry Classification Standard (GICS) sectors and their sub industries of the S&P 500 companies, and aimed to discover the sub blocks of $\boldsymbol{\Omega}$ which had nonzero entries. Those blocks could help us build the partial correlation networks of the sectors and sub industries for the S&P 500 stocks in 2005 and 2008, respectively.

The advantage of investigating the complex financial network system by partial correlation is to overcome the issue that the marginal correlation between two stocks might be a result of their correlations to other mediating stocks (Kenett et al., 2010). For example, if two stocks $R_{j_1,t}$ and $R_{j_2,t}$ are both correlated with some stocks in the set $\mathbf{R}_{-(j_1,j_2),t} = \{R_{j,t} : j \neq j_1, j_2\}$, the partial correlation can suitably remove the linear effect of $\mathbf{R}_{-(j_1,j_2),t}$ on $R_{j_1,t}$ and $R_{j_2,t}$. Hence, it measures a "direct" relationship between j_1 and j_2 (de la Fuente et al., 2004). The partial correlation analysis is widely used in the study of financial networks (Shapira et al., 2009; Kenett et al., 2010), as well as the study of gene networks (de la Fuente et al., 2004; Reverter and Chan, 2008; Chen and Zheng, 2009).

Based on the information on bloomberg and "List of S&P 500 companies" on wikipedia, we identified 10 major sectors with 54 sub industries of the S&P 500 companies (see Tables 4 and 5 for detailed categories). The 10 sectors were Consumer Discretionary, Consumer Staples, Energy, Financials, Health Care, Industrials, Information Technology, Materials, Telecommunication Services and Utilities. There were one company with the unidentified sector and eight companies with unidentified sub industries due to acquisition or ticket change (represented by "NA" in Tables 4 and 5).

To explore the partial correlation networks of different sectors and sub industries, we were interested in a set of hypotheses

$$H_{h_1h_2,0} : \omega_{j_1,j_2} = 0 \text{ for any } (j_1, j_2) \in I_{h_1} \times I_{h_2} \text{ vs. } H_{h_1h_2,1} : \omega_{j_1,j_2} \neq 0 \text{ for some } (j_1, j_2) \in I_{h_1} \times I_{h_2} \tag{20}$$

for disjoint index sets $\{I_1, \ldots, I_H\}$, which represented different sub industries. For each of the hypotheses in (20), we calculated the Studentized-type statistic $n^{1/2}|\widehat{\mathbf{D}}^{-1}\widehat{\boldsymbol{\Omega}}_{\mathcal{S}}|_\infty$ in (13) with $\mathcal{S} = I_{h_1} \times I_{h_2}$ and apply the SKMB procedure to obtain $M = 10000$ parametric bootstrap samples $\widehat{\boldsymbol{\xi}}_1, \ldots, \widehat{\boldsymbol{\xi}}_M$. The P-value of the hypothesis (20) was

$$\text{P-value}_{h_1,h_2} = \frac{1}{M}\sum_{m=1}^{M} \mathbb{I}\{|\widehat{\boldsymbol{\xi}}_m|_\infty \geq n^{1/2}|\widehat{\mathbf{D}}^{-1}\widehat{\boldsymbol{\Omega}}_{\mathcal{S}}|_\infty\} \text{ for } \mathcal{S} = I_{h_1} \times I_{h_2}.$$

To identify the significant blocks, we applied the Benjamini and Hochberg (1995)'s multiple testing procedure that controls the false discovery rate (FDR) of (20) at the rate $\alpha = 0.1$. Let $\text{pvalue}_{(1)} \leq \cdots \leq \text{pvalue}_{(K)}$ be the ordered P-values and $H_{(1),0}, \ldots, H_{(K),0}$ be the corresponding null hypotheses, where $K = H(H-1)/2$ is the number of hypotheses under

Please cite this article in press as: Chang J., et al., Confidence regions for entries of a large precision matrix. Journal of Econometrics (2018), https://doi.org/10.1016/j.jeconom.2018.03.020.

Table 4
Sectors and sub industries of the 402 S&P 500 stocks. "NA" represents that the sector or sub industry of the corresponding stock cannot be identified due to acquisition or ticket change.

Stock Symbols	Sectors	Sector No.	Sub Industries	Industry No.
IPG	Consumer Discretionary	1	Advertising	1
ANF, COH, NKE, TIF, VFC	Consumer Discretionary	1	Apparel, Accessories & Luxury Goods	2
F, HOG, JCI	Consumer Discretionary	1	Auto Parts & Equipment	3
CBS, CMCSA, DIS, DTV, TWC, TWX	Consumer Discretionary	1	Broadcasting & Cable TV	4
IGT, WYNN	Consumer Discretionary	1	Casinos & Gaming	5
JCP, JWN, KSS, M	Consumer Discretionary	1	Department Stores	6
APOL, DV	Consumer Discretionary	1	Educational Services	7
DHI, KBH, LEN, LOW, PHM	Consumer Discretionary	1	Homebuilding	8
EXPE, HOT, MAR, WYN	Consumer Discretionary	1	Hotels, Resorts & Cruise Lines	9
BDK, NWL, SNA, SWK, WHR	Consumer Discretionary	1	Household Appliances	10
AMZN	Consumer Discretionary	1	Internet Retail	11
HAS, MAT, ODP, RRD	Consumer Discretionary	1	Printing Services	12
GCI, MDP, NYT	Consumer Discretionary	1	Publishing	13
DRI, SBUX, YUM	Consumer Discretionary	1	Restaurants	14
AN, AZO, BBBY, GPC, GPS, HAR, LTD, SPLS	Consumer Discretionary	1	Specialty Stores	15
FPL, WPO	Consumer Discretionary	1	NA	NA
ADM	Consumer Staples	2	Agricultural Products	16
CVS, SVU, SWY, WAG	Consumer Staples	2	Food & Drug Stores	17
AVP, CL, KMB	Consumer Staples	2	Household Products	18
TGT, FDO, WMT	Consumer Staples	2	Hypermarkets & Super Centers	19
CAG, CCE, CPB, DF, GIS, HNZ, HRL, HSY, K, KFT, KO, MKC, PBG, PEP, SJM, SLE, STZ, TAP, TSN	Consumer Staples	2	Packaged Food	20
EL, PG	Consumer Staples	2	Personal Products	21
MO, RAI	Consumer Staples	2	Tobacco	22
BTU, CNX, MEE	Energy	3	Coal Operations	23
APA, CHK, COG, COP, CTX, CVX, DNR, DO, DVN, EOG, EP, EQT, ESV, FO, HES, MRO, MUR, NBL, OXY, PXD, RRC, SE, SWN, TSO, VLO, WMB, XTO	Energy	3	Oil & Gas Exploration & Production	24
BHI, BJS, CAM, FTI, NBR, NOV, RDC, SII, SLB	Energy	3	Oil & Gas Equipment & Services	25
BAC, BBT, BK, C, CIT, CMA, COF, FHN, FITB, HCBK, HRB, IVZ, KEY, LM, MI, MTB, NTRS, PNC, SLM, STI, USB, WFC	Financials	4	Banks	26
CME, EFX, ICE, NYX, PFG, PRU, RF, STT, TROW, UNM, VTR	Financials	4	Diversified Financial Services	27
ETFC, FII, JNS, LUK, MS, SCHW	Financials	4	Investment Banking & Brokerage	28
AFL, AIG, AIZ, CB, CINF, GNW, HIG, L, LNC, MBI, MET, MMC, PGR, TMK, TRV, XL	Financials	4	Property & Casualty Insurance	29
AMT, AVB, BXP, CBG, HCN, HCP, HST, IRM, KIM, PBCT, PCL, PSA, SPG, VNO, WY	Financials	4	REITs	30

our consideration. Note that we had $K = 1431$ for testing sub industry blocks. We rejected $H_{(1),0}, \ldots, H_{(v),0}$ in (20) for $v = \max\{1 \le j \le K : \text{pvalue}_{(j)} \le \alpha j/K\}$.

We constructed the partial correlation networks based on the significant blocks from the above multiple testing procedure. The estimated partial correlation networks of the 54 sub industries, labeled by numbers from 1 to 54, are shown in the right panels of Figs. 1 and 2, corresponding to 2005 and 2008, respectively. The name of each sub industry and the stocks included can be found in Tables 4 and 5. The shaded areas with different colors represent the 10 major sectors, respectively. The left panels in Figs. 1 and 2 give the partial correlation networks of the sectors, where the nodes represent the 10 sectors, and two nodes (sectors) $\tilde{h}_1$ and $\tilde{h}_2$ are connected if and only if there exists a connection between one of sub industries belonging to sector $\tilde{h}_1$ and one of sub industries belonging to sector $\tilde{h}_2$ in the right panel.

We observed from the left panel of Fig. 1 that preceding the crisis in 2005, the Consumer Discretionary sector was likely to be a hub connecting to all the other 9 sectors. It was the most influential sector with the largest degree, i.e., the total number of directed links connecting to the Consumer Discretionary sector in the network. During the crisis in 2008, the Consumer Discretionary sector was still the most influential sector as shown by the left panel of Fig. 2, but it had less connections compared to 2005. The Financials sector was a little bit separated from the other sectors in 2008, with only half connections in contrast with the network connectivity in 2005. The similar situation also appeared in the partial correlations networks of S&P 500 sub industries as shown in the right panels of Figs. 1 and 2. More specifically, both the numbers of the edges within and between most sectors for the network of S&P 500 sub industries in 2008 were significantly less than those in 2005 (see Table 6 for details), which indicated that the market fear in the crisis broke the connections of stock sectors and sub industries. From the perspective of financial network studies, the above analysis confirmed that fear froze the market in the 2008 crisis (Reavis, 2012).

Please cite this article in press as: Chang J., et al., Confidence regions for entries of a large precision matrix. Journal of Econometrics (2018), https://doi.org/10.1016/j.jeconom.2018.03.020.

Table 5
Sectors and sub industries of the 402 S&P 500 stocks (continued). “NA” represents that the sector or sub industry of the corresponding stock cannot be identified due to acquisition or ticket change.

Stock Symbols	Sectors	Sector No.	Sub Industries	Industry No.
AOC	Financials	4	NA	NA
AMGN, BIIB, CELG, FRX, GENZ, GILD, HSP, KG, LIFE, LLY, MRK, MYL, WPI	Health Care	5	Pharmaceuticals	31
ABC, AET, BMY, CAH, CI, DGX, DVA, ESRX, HUM, MCK, MHS, PDCO, THC, UNH, WAT, WLP, XRAY	Health Care	5	Health Care Supplies	32
ABT, BAX, BCR, BDX, ISRG, JNJ, MDT, MIL, PKI, PLL, STJ, SYK, TMO, VAR	Health Care	5	Health Care Equipment & Services	33
BA, RTN	Industrials	6	Aerospace & Defense	34
CHRW, EXPD, FDX, UPS	Industrials	6	Air Freight & Logistics	35
LUV	Industrials	6	Airlines	36
DE, FAST, GLW, MAS, MTW, PCAR	Industrials	6	Construction & Farm Machinery & Heavy Trucks	37
COL, EMR, ETN, GE, HON, IR, JEC, LEG, LLL, MMM, PH, ROK, RSG, TXT, TYC	Industrials	6	Industrial Conglomerates	38
CMI, DHR, DOV, FLS, GWW, ITT, ITW	Industrials	6	Industrial Machinery	39
CSX, NSC, UNP	Industrials	6	Railroads	40
ACS, CTAS, FLR, RHI	Industrials	6	NA	NA
CBE, MOLX, JBL, LXK	Information Technology	7	Office Electronics	41
ADBE, ADSK, BMC, CA, ERTS, MFE, MSFT, NOVL, ORCL, TDC	Information Technology	7	Application Software	42
CIEN, HRS, JDSU, JNPR, MOT	Information Technology	7	Communications Equipment	43
AAPL, AMD, HPQ, JAVA, QLGC, SNDK	Information Technology	7	Computer Storage & Peripherals	44
ADP, AKAM, CRM, CSC, CTSH, CTXS, CVG, EBAY, FIS, GOOG, IBM, INTU, MA, MWW, PAYX, TSS, XRX, YHOO, DNB	Information Technology	7	Information Services	45
ALTR, AMAT, BRCM, INTC, KLAC, LLTC, LSI, MCHP, MU, NSM, NVDA, NVLS, QCOM, TXN, XLNX	Information Technology	7	Semiconductors	46
ATI, BLL, FCX, NEM, OI	Materials	8	Metal & Glass Containers	47
DD, DOW, ECL, EMN, IFF, MON, PPG, PX, SHW, SIAL	Materials	8	Specialty Chemicals	48
BMS, MWV, PTV	Materials	8	Containers & Packaging	49
AKS, TIE, X	Materials	8	Iron & Steel	50
AVY, IP, SEE	Materials	8	Paper Packaging	51
VMC	Materials	8	NA	NA
CTL, EQ, FTR, Q, S, T, VZ, WIN	Telecommunications Services	9	Telecom Carriers	52
AEE, AEP, AES, AYE, CMS, CNP, D, DYN, ETR, FE, PEG, POM, PPL, SCG, SO, SRE, TE, WEC, XEL	Utilities	10	MultiUtilities	53
STR, TEG	Utilities	10	Utility Networks	54
RX	NA	NA	NA	NA

Table 6
The numbers of edges within and between sectors for the partial correlation networks of the S&P 500 sub industries in Figs. 1 and 2.

Sectors	2005		2008	
	Within	Between	Within	Between
Consumer Discretionary	13	37	9	12
Consumer Staples	4	16	1	6
Energy	0	8	1	4
Financials	3	14	5	5
Health Care	2	10	2	8
Industrials	5	19	3	5
Information Technology	5	13	6	9
Materials	2	12	2	10
Telecommunication Services	0	3	0	1
Utilities	0	4	1	2

Acknowledgments

The authors are grateful to the Co-Editor, an Associate Editor and two anonymous referees for constructive comments and suggestions. This research was undertaken with the assistance of resources provided at the NCI National Facility systems at the Australian National University through the ANU Allocation Scheme supported by the Australian Government. The

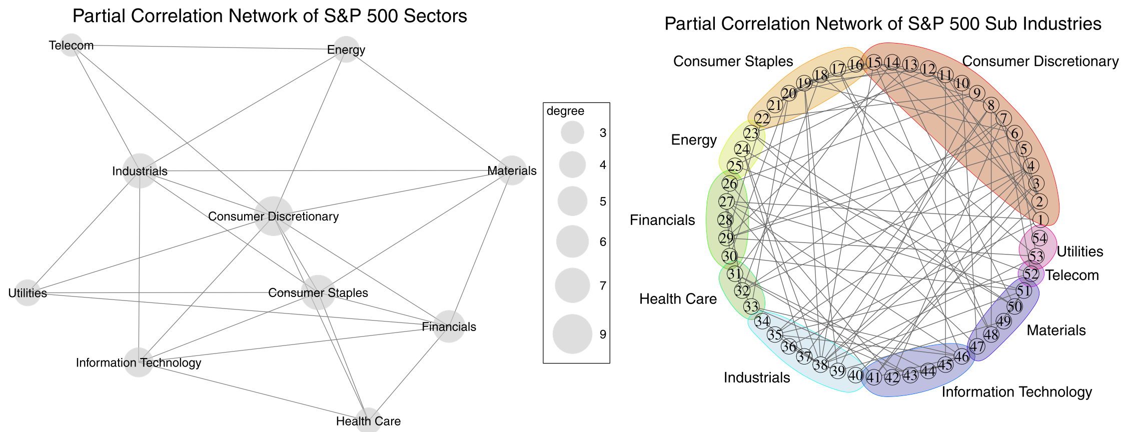

Fig. 1. Partial correlation networks of S&P 500 sectors and sub industries in 2005 (preceding the crisis). The detailed information of the sub industries represented by numbers 1–54 in the right panel can be correspondingly found in Tables 4 and 5.

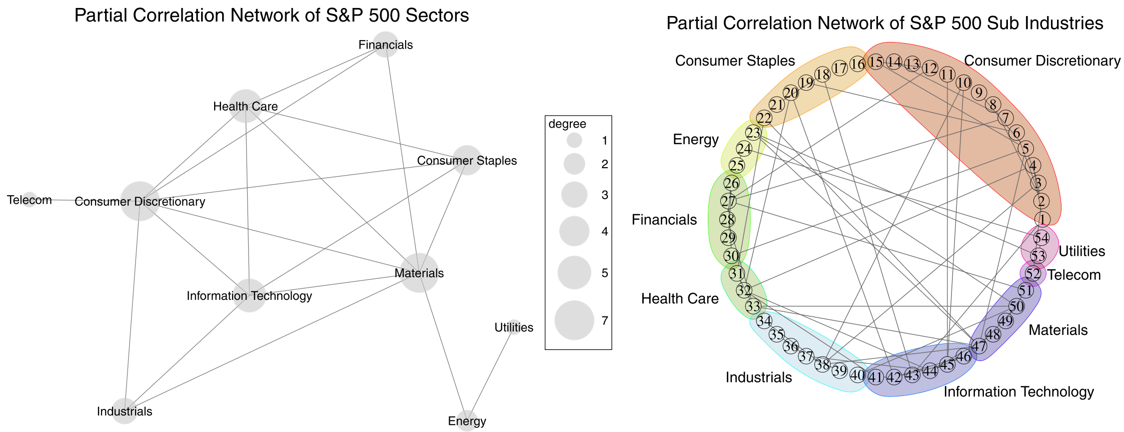

Fig. 2. Partial correlation networks of S&P 500 sectors and sub industries in 2008 (during the crisis). The detailed information of the sub industries represented by numbers 1–54 in the right panel can be correspondingly found in Tables 4 and 5.

authors would like to thank Pauline O'Shaughnessy for her great help in getting us started with the supercomputer Raijin. Jinyuan Chang was supported in part by the Fundamental Research Funds for the Central Universities (Grant No. JBK1802069, JBK171121), NSFC (Grant No. 11501462), the Fok Ying-Tong Education Foundation, the Center of Statistical Research at SWUFE, and the Joint Lab of Data Science and Business Intelligence at SWUFE. Qiwei Yao was supported in part by an EPSRC research grant.

Appendix

Throughout the Appendix, let C denote a generic positive constant depending only on the constants specified in Conditions 1–4, which may be different in different cases. Let $\rho_1^{-1}=2\gamma_1^{-1}+\gamma_3^{-1}$, $\rho_2^{-1}=2\gamma_2^{-1}+\gamma_3^{-1}$, $\rho_3^{-1}=\gamma_1^{-1}+\gamma_2^{-1}+\gamma_3^{-1}$ and $\rho_4^{-1}=\max\{\rho_2^{-1},\rho_3^{-1}\}+\gamma_3^{-1}$. Define $\zeta=\min\{\rho_1,\rho_2,\rho_3,\rho_4\}$ and $\boldsymbol{\Delta}=n^{-1}\sum_{t=1}^{n}\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^{\mathrm{T}}-\mathbf{V}=:(\delta_{j_1,j_2})$.

Lemma 1. *Assume Conditions 1–3 hold. If $\log p=o\{n^{\zeta/(2-\zeta)}\}$, there exists a uniform constant $A_0>1$ independent of n and p such that*

$$\mathbb{P}\big\{|\widehat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma}|_\infty>A_1(n^{-1}\log p)^{1/2}\big\}\le\exp\{-CA_1^{\rho_1}(n\log p)^{\rho_1/2}\}+\exp(-CA_1^2\log p),$$

$$\mathbb{P}\big\{|\boldsymbol{\Delta}|_\infty>A_2(n^{-1}\log p)^{1/2}\big\}\le\exp\{-CA_2^{\rho_2}(n\log p)^{\rho_2/2}\}+\exp(-CA_2^2\log p),$$

$$\sup_{1\le j\le p}\mathbb{P}\bigg(\frac{1}{n}\sum_{t=1}^{n}\epsilon_{j,t}^2>A_3v_{j,j}\bigg)\le\exp(-CA_3^{\rho_2}n^{\rho_2}),$$

Please cite this article in press as: Chang J., et al., Confidence regions for entries of a large precision matrix. Journal of Econometrics (2018), https://doi.org/10.1016/j.jeconom.2018.03.020.

$$\sup_{1\le j\le p}\mathbb{P}\left\{\max_{k\ne j}\left|\frac{1}{n}\sum_{t=1}^{n}\epsilon_{j,t}y_{k,t}\right|>A_4(n^{-1}\log p)^{1/2}\right\}\le\exp\{-CA_4^{\rho_3}(n\log p)^{\rho_3/2}\}+\exp(-CA_4^2\log p),$$

$$\sup_{1\le j\le p}\mathbb{P}\left\{\left|\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{\alpha}_{j,-j}^{\mathrm{T}}\mathbf{y}_{-j,t}\epsilon_{j,t}\right|>A_5(n^{-1}\log p)^{1/2}\right\}\le\exp\{-CA_5^{\rho_4}(n\log p)^{\rho_4/2}\}+\exp(-CA_5^2\log p)$$

for any $A_1,A_2,A_3,A_4,A_5>A_0$.

Proof. For any given j_1 and j_2, based on the first part of Condition 1, Lemma 2 of Chang et al. (2013) leads to $\sup_{1\le t\le n}\mathbb{P}(|y_{j_1,t}y_{j_2,t}-\sigma_{j_1,j_2}|>x)\le C\exp(-Cx^{\gamma_1/2})$ for any $x>0$. Hence, for any $x>0$ such that $nx\to\infty$, Theorem 1 of Merlevède et al. (2011) leads to

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{t=1}^{n}y_{j_1,t}y_{j_2,t}-\sigma_{j_1,j_2}\right|>x\right)\le n\exp(-Cn^{\rho_1}x^{\rho_1})+\exp(-Cnx^2).$$

By Bonferroni inequality, we have $\mathbb{P}(|\widehat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma}|_\infty>x)\le np^2\exp(-Cn^{\rho_1}x^{\rho_1})+p^2\exp(-Cnx^2)$. Let $x=A_1(n^{-1}\log p)^{1/2}$, we obtain the first conclusion. Following the same arguments, we can establish the other inequalities. □

Lemma 2. *Assume Conditions 1–3 hold. Let* $s=\max_{1\le j\le p}|\boldsymbol{\alpha}_j|_0$. *For some suitable* $\lambda_j\asymp(n^{-1}\log p)^{1/2}$ *for each* $j=1,\ldots,p$, *we have* $\max_{1\le j\le p}|\widehat{\boldsymbol{\alpha}}_j-\boldsymbol{\alpha}_j|_1=o_p\{(\log p)^{-1}\}$ *and* $\max_{1\le j\le p}|\widehat{\boldsymbol{\alpha}}_j-\boldsymbol{\alpha}_j|_2=o_p\{(n\log p)^{-1/4}\}$ *provided that* $\log p=o\{n^{\zeta/(2-\zeta)}\}$ *and* $s^2(\log p)^3n^{-1}=o(1)$.

Proof. Define

$$\mathscr{T}=\left\{\max_{1\le j\le p}\max_{k\ne j}\left|\frac{1}{n}\sum_{t=1}^{n}\epsilon_{j,t}y_{k,t}\right|\le A_4(n^{-1}\log p)^{1/2}\right\}$$

for some $A_4>A_0$, where A_0 is given in Lemma 1. Selecting $\lambda_j\ge 4A_4(n^{-1}\log p)^{1/2}$ for any j, Theorem 6.1 and Corollary 6.8 of Bühlmann and van de Geer (2011) imply that, restricted on $\mathscr{T}$, we have

$$\max_{1\le j\le p}|\widehat{\boldsymbol{\alpha}}_j-\boldsymbol{\alpha}_j|_1\le Cs(n^{-1}\log p)^{1/2}\tag{21}$$

and

$$(\widehat{\boldsymbol{\alpha}}_j-\boldsymbol{\alpha}_j)^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_{-j,-j}(\widehat{\boldsymbol{\alpha}}_j-\boldsymbol{\alpha}_j)\le Csn^{-1}\log p\tag{22}$$

with probability approaching 1. By Bonferroni inequality and Lemma 1,

$$\mathbb{P}(\mathscr{T}^c)\le\sum_{j=1}^{p}\mathbb{P}\left\{\sum_{k\ne j}\left|\frac{1}{n}\sum_{t=1}^{n}\epsilon_{j,t}y_{k,t}\right|>A_4(n^{-1}\log p)^{1/2}\right\}\le\ p\exp\{-CA_4^{\rho_3}(n\log p)^{\rho_3/2}\}+p\exp(-CA_4^2\log p).$$

For suitable selection of A_4, we have $\mathbb{P}(\mathscr{T}^c)\to 0$ as $n\to\infty$. Thus, from (21), it holds that

$$\max_{1\le j\le p}|\widehat{\boldsymbol{\alpha}}_j-\boldsymbol{\alpha}_j|_1=\ O_p\{s(n^{-1}\log p)^{1/2}\}=o_p\{(\log p)^{-1}\}.\tag{23}$$

On the other hand, notice that

$$(\widehat{\boldsymbol{\alpha}}_j-\boldsymbol{\alpha}_j)^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_{-j,-j}(\widehat{\boldsymbol{\alpha}}_j-\boldsymbol{\alpha}_j)\ge\lambda_{\min}(\boldsymbol{\Sigma}_{-j,-j})|\widehat{\boldsymbol{\alpha}}_j-\boldsymbol{\alpha}_j|_2^2-|\widehat{\boldsymbol{\Sigma}}_{-j,-j}-\boldsymbol{\Sigma}_{-j,-j}|_\infty|\widehat{\boldsymbol{\alpha}}_j-\boldsymbol{\alpha}_j|_1^2,$$

by Condition 2, Lemma 1, (22) and (23), we have

$$\max_{1\le j\le p}|\widehat{\boldsymbol{\alpha}}_j-\boldsymbol{\alpha}_j|_2=\ O_p\{(sn^{-1}\log p)^{1/2}\}=o_p\{(n\log p)^{-1/4}\}.$$

Hence, we complete the proof. □

Lemma 3. *Assume the conditions for Lemmas 1 and 2 hold, then*

$$\frac{1}{n}\sum_{t=1}^{n}\widehat{\epsilon}_{j_1,t}\widehat{\epsilon}_{j_2,t}-\frac{1}{n}\sum_{t=1}^{n}\epsilon_{j_1,t}\epsilon_{j_2,t}=-(\widehat{\alpha}_{j_1,j_2}-\alpha_{j_1,j_2})\left(\frac{1}{n}\sum_{t=1}^{n}\epsilon_{j_2,t}^2\right)\mathbb{I}(j_1\ne j_2)$$
$$-(\widehat{\alpha}_{j_2,j_1}-\alpha_{j_2,j_1})\left(\frac{1}{n}\sum_{t=1}^{n}\epsilon_{j_1,t}^2\right)\mathbb{I}(j_1\ne j_2)+o_p\{(n\log p)^{-1/2}\}.$$

Here the remainder term $o_p\{(n\log p)^{-1/2}\}$ is uniform over all j_1 and j_2.

Proof. Notice that $\epsilon_{j,t}=-\boldsymbol{\alpha}_j^{\mathrm{T}}\mathbf{y}_t$ and $\widehat{\epsilon}_{j,t}=-\widehat{\boldsymbol{\alpha}}_j^{\mathrm{T}}\mathbf{y}_t$ for any t, then

$$\frac{1}{n}\sum_{t=1}^{n}\widehat{\epsilon}_{j_1,t}\widehat{\epsilon}_{j_2,t}-\frac{1}{n}\sum_{t=1}^{n}\epsilon_{j_1,t}\epsilon_{j_2,t}=-\frac{1}{n}\sum_{t=1}^{n}(\widehat{\boldsymbol{\alpha}}_{j_1}-\boldsymbol{\alpha}_{j_1})^{\mathrm{T}}\mathbf{y}_t\epsilon_{j_2,t}-\frac{1}{n}\sum_{t=1}^{n}(\widehat{\boldsymbol{\alpha}}_{j_2}-\boldsymbol{\alpha}_{j_2})^{\mathrm{T}}\mathbf{y}_t\epsilon_{j_1,t}$$
$$+\frac{1}{n}\sum_{t=1}^{n}(\widehat{\boldsymbol{\alpha}}_{j_1}-\boldsymbol{\alpha}_{j_1})^{\mathrm{T}}\mathbf{y}_t\mathbf{y}_t^{\mathrm{T}}(\widehat{\boldsymbol{\alpha}}_{j_2}-\boldsymbol{\alpha}_{j_2}).$$

Condition 2, Lemmas 1 and 2 imply that

$$\begin{aligned}&\max_{1\le j_1,j_2\le p}\left|\frac{1}{n}\sum_{t=1}^{n}(\widehat{\boldsymbol{\alpha}}_{j_1}-\boldsymbol{\alpha}_{j_1})^{\mathrm{T}}\mathbf{y}_t\mathbf{y}_t^{\mathrm{T}}(\widehat{\boldsymbol{\alpha}}_{j_2}-\boldsymbol{\alpha}_{j_2})\right|\\&\le\max_{1\le j_1,j_2\le p}|(\widehat{\boldsymbol{\alpha}}_{j_1}-\boldsymbol{\alpha}_{j_1})^{\mathrm{T}}\boldsymbol{\Sigma}(\widehat{\boldsymbol{\alpha}}_{j_2}-\boldsymbol{\alpha}_{j_2})|+\max_{1\le j_1,j_2\le p}|(\widehat{\boldsymbol{\alpha}}_{j_1}-\boldsymbol{\alpha}_{j_1})^{\mathrm{T}}(\widehat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma})(\widehat{\boldsymbol{\alpha}}_{j_2}-\boldsymbol{\alpha}_{j_2})|\\&\le C\max_{1\le j\le p}|\widehat{\boldsymbol{\alpha}}_j-\boldsymbol{\alpha}_j|_2^2+|\widehat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma}|_\infty\max_{1\le j\le p}|\widehat{\boldsymbol{\alpha}}_j-\boldsymbol{\alpha}_j|_1^2\\&=o_p\{(n\log p)^{-1/2}\}.\end{aligned}$$

Meanwhile, by Lemma 1, we have $\max_{1\le j\le p}\max_{k\ne j}|n^{-1}\sum_{t=1}^{n}\epsilon_{j,t}y_{k,t}|=O_p\{(n^{-1}\log p)^{1/2}\}$, which implies that

$$\begin{aligned}\max_{1\le j_1,j_2\le p}\left|\sum_{k\ne j_1,j_2}(\widehat{\alpha}_{j_1,k}-\alpha_{j_1,k})\left(\frac{1}{n}\sum_{t=1}^{n}y_{k,t}\epsilon_{j_2,t}\right)\right|&\le\max_{1\le j\le p}|\widehat{\boldsymbol{\alpha}}_j-\boldsymbol{\alpha}_j|_1\cdot\max_{1\le j\le p}\max_{k\ne j}\left|\frac{1}{n}\sum_{t=1}^{n}y_{k,t}\epsilon_{j,t}\right|\\&=o_p\{(n\log p)^{-1/2}\}.\end{aligned}$$

Therefore, we have

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^{n}(\widehat{\boldsymbol{\alpha}}_{j_1}-\boldsymbol{\alpha}_{j_1})^{\mathrm{T}}\mathbf{y}_t\epsilon_{j_2,t}&=(\widehat{\alpha}_{j_1,j_2}-\alpha_{j_1,j_2})\left(\frac{1}{n}\sum_{t=1}^{n}y_{j_2,t}\epsilon_{j_2,t}\right)\mathbb{I}(j_1\ne j_2)\\&\quad+\sum_{k\ne j_1,j_2}(\widehat{\alpha}_{j_1,k}-\alpha_{j_1,k})\left(\frac{1}{n}\sum_{t=1}^{n}y_{k,t}\epsilon_{j_2,t}\right)\\&=(\widehat{\alpha}_{j_1,j_2}-\alpha_{j_1,j_2})\left(\frac{1}{n}\sum_{t=1}^{n}y_{j_2,t}\epsilon_{j_2,t}\right)\mathbb{I}(j_1\ne j_2)+o_p\{(n\log p)^{-1/2}\}.\end{aligned}\tag{24}$$

Here the remainder term is uniform over any j_1 and j_2. On the other hand, $n^{-1}\sum_{t=1}^{n}y_{j,t}\epsilon_{j,t}=n^{-1}\sum_{t=1}^{n}\epsilon_{j,t}^2+n^{-1}\sum_{t=1}^{n}\boldsymbol{\alpha}_{j,-j}^{\mathrm{T}}\mathbf{y}_{-j,t}\epsilon_{j,t}$. By the fourth result of Lemma 1, it yields that $n^{-1}\sum_{t=1}^{n}y_{j,t}\epsilon_{j,t}=n^{-1}\sum_{t=1}^{n}\epsilon_{j,t}^2+O_p\{(n^{-1}\log p)^{1/2}\}$. Here the remainder term is uniform over all j. Together with (24), we have

$$\frac{1}{n}\sum_{t=1}^{n}(\widehat{\boldsymbol{\alpha}}_{j_1}-\boldsymbol{\alpha}_{j_1})^{\mathrm{T}}\mathbf{y}_t\epsilon_{j_2,t}=(\widehat{\alpha}_{j_1,j_2}-\alpha_{j_1,j_2})\left(\frac{1}{n}\sum_{t=1}^{n}\epsilon_{j_2,t}^2\right)\mathbb{I}(j_1\ne j_2)+o_p\{(n\log p)^{-1/2}\}.$$

Here the remainder term is also uniform over all j_1 and j_2. Hence,

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^{n}\widehat{\epsilon}_{j_1,t}\widehat{\epsilon}_{j_2,t}-\frac{1}{n}\sum_{t=1}^{n}\epsilon_{j_1,t}\epsilon_{j_2,t}&=-(\widehat{\alpha}_{j_1,j_2}-\alpha_{j_1,j_2})\left(\frac{1}{n}\sum_{t=1}^{n}\epsilon_{j_2,t}^2\right)\mathbb{I}(j_1\ne j_2)\\&\quad-(\widehat{\alpha}_{j_2,j_1}-\alpha_{j_2,j_1})\left(\frac{1}{n}\sum_{t=1}^{n}\epsilon_{j_1,t}^2\right)\mathbb{I}(j_1\ne j_2)+o_p\{(n\log p)^{-1/2}\}.\end{aligned}$$

We complete the proof. □

Proof of Proposition 1. Notice that $v_{j_1,j_2} = \frac{\omega_{j_1,j_2}}{\omega_{j_1,j_1}\omega_{j_2,j_2}}$, $\alpha_{j_1,j_2} = -\frac{\omega_{j_1,j_2}}{\omega_{j_1,j_1}}$ and $\widetilde{v}_{j_1,j_2} = n^{-1}\sum_{t=1}^n \widehat{\epsilon}_{j_1,t}\widehat{\epsilon}_{j_2,t}$ for any j_1 and j_2, Lemma 3 implies that

$$\begin{aligned}-\widehat{v}_{j_1,j_2} + v_{j_1,j_2} &= \widetilde{v}_{j_1,j_2} + \frac{\widehat{\alpha}_{j_1,j_2}}{n}\sum_{t=1}^n \widehat{\epsilon}_{j_2,t}^2 + \frac{\widehat{\alpha}_{j_2,j_1}}{n}\sum_{t=1}^n \widehat{\epsilon}_{j_1,t}^2 + v_{j_1,j_2} \\ &= \frac{1}{n}\sum_{t=1}^n(\epsilon_{j_1,t}\epsilon_{j_2,t} - v_{j_1,j_2}) + \frac{\alpha_{j_1,j_2}}{n}\sum_{t=1}^n(\epsilon_{j_2,t}^2 - v_{j_2,j_2}) \\ &\quad + \frac{\alpha_{j_2,j_1}}{n}\sum_{t=1}^n(\epsilon_{j_1,t}^2 - v_{j_1,j_1}) + o_p\{(n\log p)^{-1/2}\}\end{aligned}$$

for any $j_1 \neq j_2$. Recall $\boldsymbol{\Delta} = n^{-1}\sum_{t=1}^n \boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^{\mathrm{T}} - \mathbf{V} =: (\delta_{j_1,j_2})$. It follows from Lemma 1 that $\max_{1\le j_1,j_2\le p}|\delta_{j_1,j_2}| = O_p\{(n^{-1}\log p)^{1/2}\}$. Recall $\widehat{\omega}_{j_1,j_2} = \frac{\widehat{v}_{j_1,j_2}}{\widehat{v}_{j_1,j_1}\widehat{v}_{j_2,j_2}}$, if $\log p = o\{n^{\zeta/(2-\zeta)}\}$ for ζ specified in Lemma 1 and $s^2(\log p)^3 n^{-1} = o(1)$, it holds that

$$\widehat{\omega}_{j_1,j_2} - \omega_{j_1,j_2} = -\frac{\delta_{j_1,j_2}}{v_{j_1,j_1}v_{j_2,j_2}} + o_p\{(n\log p)^{-1/2}\}$$

for any $j_1 \neq j_2$. Meanwhile, by the same arguments, for each $j = 1, \ldots, p$, it holds that $\widehat{\omega}_{j,j} - \omega_{j,j} = -\frac{\delta_{j,j}}{v_{j,j}^2} + o_p\{(n\log p)^{-1/2}\}$. This proves Proposition 1. □

Proof of Theorem 1. Define $d_1 = \sup_{x>0}|\mathbb{P}(n^{1/2}|\boldsymbol{\Pi}_{\mathcal{S}}|_\infty > x) - \mathbb{P}(|\boldsymbol{\xi}|_\infty > x)|$. For any $x > 0$ and $\varepsilon_1 > 0$, it yields that

$$\begin{aligned}\mathbb{P}\big(n^{1/2}|\widehat{\boldsymbol{\Omega}}_{\mathcal{S}} - \boldsymbol{\Omega}_{\mathcal{S}}|_\infty > x\big) &\le \mathbb{P}(n^{1/2}|\boldsymbol{\Pi}_{\mathcal{S}}|_\infty > x - \varepsilon_1) + \mathbb{P}(n^{1/2}|\boldsymbol{\Upsilon}_{\mathcal{S}}|_\infty > \varepsilon_1) \\ &\le \mathbb{P}(|\boldsymbol{\xi}|_\infty > x - \varepsilon_1) + d_1 + \mathbb{P}(n^{1/2}|\boldsymbol{\Upsilon}_{\mathcal{S}}|_\infty > \varepsilon_1) \\ &= \mathbb{P}(|\boldsymbol{\xi}|_\infty > x) + \mathbb{P}(x - \varepsilon_1 < |\boldsymbol{\xi}|_\infty \le x) + d_1 + \mathbb{P}(n^{1/2}|\boldsymbol{\Upsilon}_{\mathcal{S}}|_\infty > \varepsilon_1).\end{aligned}$$

On the other hand, notice that $\mathbb{P}\big(n^{1/2}|\widehat{\boldsymbol{\Omega}}_{\mathcal{S}} - \boldsymbol{\Omega}_{\mathcal{S}}|_\infty > x\big) \ge \mathbb{P}(n^{1/2}|\boldsymbol{\Pi}_{\mathcal{S}}|_\infty > x + \varepsilon_1) - \mathbb{P}(n^{1/2}|\boldsymbol{\Upsilon}_{\mathcal{S}}|_\infty > \varepsilon_1)$, following the same arguments, we have

$$\sup_{x>0}\big|\mathbb{P}\big(n^{1/2}|\widehat{\boldsymbol{\Omega}}_{\mathcal{S}} - \boldsymbol{\Omega}_{\mathcal{S}}|_\infty > x\big) - \mathbb{P}(|\boldsymbol{\xi}|_\infty > x)\big| \le d_1 + \sup_{x>0}\mathbb{P}(x - \varepsilon_1 < |\boldsymbol{\xi}|_\infty \le x + \varepsilon_1) + \mathbb{P}(n^{1/2}|\boldsymbol{\Upsilon}_{\mathcal{S}}|_\infty > \varepsilon_1). \quad (25)$$

By the Anti-concentration inequality for Gaussian random vector [Corollary 1 of Chernozhukov et al. (2015)], it holds that

$$\sup_{x>0}\mathbb{P}(x - \varepsilon_1 < |\boldsymbol{\xi}|_\infty \le x + \varepsilon_1) \le C\varepsilon_1(\log p)^{1/2} \quad (26)$$

for any $\varepsilon_1 \to 0$. From the proofs of Lemmas 2 and 3, we know $n^{1/2}|\boldsymbol{\Upsilon}_{\mathcal{S}}|_\infty = O_p(sn^{-1/2}\log p)$. Thus, if $s^2(\log p)^3 n^{-1} = o(1)$, we can select a suitable ε_1 to guarantee $\varepsilon_1(\log p)^{1/2} \to 0$ and $n^{1/2}|\boldsymbol{\Upsilon}_{\mathcal{S}}|_\infty = o_p(\varepsilon_1)$. Therefore, for such selected ε_1, (25) leads to

$$\sup_{x>0}\big|\mathbb{P}\big(n^{1/2}|\widehat{\boldsymbol{\Omega}}_{\mathcal{S}} - \boldsymbol{\Omega}_{\mathcal{S}}|_\infty > x\big) - \mathbb{P}(|\boldsymbol{\xi}|_\infty > x)\big| \le d_1 + o(1). \quad (27)$$

To prove Theorem 1, it suffices to show $d_1 \to 0$ as $n \to \infty$. We will show this below.

Write $\boldsymbol{\Pi}_{\mathcal{S}} = -(\bar{\varsigma}_1, \ldots, \bar{\varsigma}_r)^{\mathrm{T}}$ where $\bar{\varsigma}_j = n^{-1}\sum_{t=1}^n \varsigma_{j,t}$ and $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_r)^{\mathrm{T}}$. Given a $D_n \to \infty$, define $\varsigma_{j,t}^+ = \varsigma_{j,t}\mathbb{I}\{|\varsigma_{j,t}| \le D_n\} - \mathbb{E}[\varsigma_{j,t}\mathbb{I}\{|\varsigma_{j,t}| \le D_n\}]$ and $\varsigma_{j,t}^- = \varsigma_{j,t}\mathbb{I}\{|\varsigma_{j,t}| > D_n\} - \mathbb{E}[\varsigma_{j,t}\mathbb{I}\{|\varsigma_{j,t}| > D_n\}]$. Write $\boldsymbol{\varsigma}_t^+ = (\varsigma_{1,t}^+, \ldots, \varsigma_{r,t}^+)^{\mathrm{T}}$ and $\boldsymbol{\varsigma}_t^- = (\varsigma_{1,t}^-, \ldots, \varsigma_{r,t}^-)^{\mathrm{T}}$ for each t. The diverging rate of D_n will be specified later. Let L be a positive integer satisfying $L \le n/2$, $L \to \infty$ and $L = o(n)$. We decompose the sequence $\{1, \ldots, n\}$ to the following $m + 1$ blocks where $m = \lfloor n/L \rfloor$ and $\lfloor\cdot\rfloor$ is the integer truncation operator: $\mathcal{G}_\ell = \{(\ell-1)L+1, \ldots, \ell L\}$ $(\ell = 1, \ldots, m)$ and $\mathcal{G}_{m+1} = \{mL+1, \ldots, n\}$. Additionally, let $b > h$ be two positive integers such that $L = b + h$, $h \to \infty$ and $h = o(b)$. We decompose each $\mathcal{G}_\ell$ $(\ell = 1, \ldots, m)$ to a "large" block with length b and a "small" block with length h. Specifically, $\mathcal{I}_\ell = \{(\ell-1)L+1, \ldots, (\ell-1)L+b\}$ and $\mathcal{J}_\ell = \{(\ell-1)L+b+1, \ldots, \ell L\}$ for any $\ell = 1, \ldots, m$, and $\mathcal{J}_{m+1} = \mathcal{G}_{m+1}$. Assume $\mathbf{u}$ is a centered normal random vector such that

$$\mathbf{u} = (u_1, \ldots, u_r)^{\mathrm{T}} \sim N\bigg[\mathbf{0}, \frac{1}{mb}\sum_{\ell=1}^m \mathbb{E}\bigg\{\bigg(\sum_{t\in\mathcal{I}_\ell}\boldsymbol{\varsigma}_t^+\bigg)\bigg(\sum_{t\in\mathcal{I}_\ell}\boldsymbol{\varsigma}_t^+\bigg)^{\mathrm{T}}\bigg\}\bigg].$$

Our following proof includes two steps. The first step is to show

$$d_2 := \sup_{x>0}\big|\mathbb{P}\big(n^{1/2}|\boldsymbol{\Pi}_{\mathcal{S}}|_\infty > x\big) - \mathbb{P}(|\mathbf{u}|_\infty > x)\big| = o(1). \quad (28)$$

And the second step is to show

$$\sup_{x>0}\left|\mathbb{P}(|\mathbf{u}|_\infty > x) - \mathbb{P}(|\boldsymbol{\xi}|_\infty > x)\right| = o(1). \tag{29}$$

From (28) and (29), we have $d_1 = o(1)$.

We first show (28). Define $d_3 = \sup_{x>0}|\mathbb{P}(|n^{-1/2}\sum_{t=1}^n \boldsymbol{\varsigma}_t^+|_\infty > x) - \mathbb{P}(|\mathbf{u}|_\infty > x)|$. Notice that $n^{1/2}\boldsymbol{\Pi}_{\mathcal{S}} = n^{-1/2}\sum_{t=1}^n \boldsymbol{\varsigma}_t^+ + n^{-1/2}\sum_{t=1}^n \boldsymbol{\varsigma}_t^-$, by the triangle inequality, it holds that $|n^{1/2}|\boldsymbol{\Pi}_{\mathcal{S}}|_\infty - |n^{-1/2}\sum_{t=1}^n \boldsymbol{\varsigma}_t^+|_\infty| \le |n^{-1/2}\sum_{t=1}^n \boldsymbol{\varsigma}_t^-|_\infty$. Similar to (25), we have

$$d_2 \le d_3 + \sup_{x>0}\mathbb{P}(x-\varepsilon_2 < |\mathbf{u}|_\infty \le x+\varepsilon_2) + \mathbb{P}\left(\left|\frac{1}{n^{1/2}}\sum_{t=1}^n \boldsymbol{\varsigma}_t^-\right|_\infty > \varepsilon_2\right) \tag{30}$$

for any $\varepsilon_2 > 0$. For each j, it follows from Davydov inequality (Davydov, 1968) that

$$\begin{aligned}\mathbb{E}\left(\left|\frac{1}{\sqrt{n}}\sum_{t=1}^n \varsigma_{j,t}^-\right|^2\right) &= \frac{1}{n}\sum_{t=1}^n \mathbb{E}\{(\varsigma_{j,t}^-)^2\} + \frac{1}{n}\sum_{t_1\ne t_2}\mathbb{E}(\varsigma_{j,t_1}^-\varsigma_{j,t_2}^-)\\ &\le \frac{1}{n}\sum_{t=1}^n \mathbb{E}\{(\varsigma_{j,t}^-)^2\} + \frac{C}{n}\sum_{t_1\ne t_2}[\mathbb{E}\{(\varsigma_{j,t_1}^-)^4\}]^{1/4}[\mathbb{E}\{(\varsigma_{j,t_2}^-)^4\}]^{1/4}\exp(-C|t_1-t_2|^{\gamma_3}).\end{aligned}$$

Applying Lemma 2 of Chang et al. (2013), Conditions 1 and 4 imply that $\sup_{j,t}\mathbb{P}(|\varsigma_{j,t}| > x) \le C\exp(-Cx^{\gamma_2/2})$ for any $x > 0$. Then

$$\mathbb{E}\{\varsigma_{j,t}^4\mathbb{I}(|\varsigma_{j,t}| > D_n)\} = 4\int_0^{D_n} x^3\mathbb{P}(|\varsigma_{j,t}| > D_n)\,dx + 4\int_{D_n}^\infty x^3\mathbb{P}(|\varsigma_{j,t}| > x)\,dx \le CD_n^4\exp(-CD_n^{\gamma_2/2}). \tag{31}$$

By the triangle inequality and Jensen's inequality,

$$\mathbb{E}\{(\varsigma_{j,t}^-)^4\} \le C\mathbb{E}\{\varsigma_{j,t}^4\mathbb{I}(|\varsigma_{j,t}| > D_n)\} + C[\mathbb{E}\{\varsigma_{j,t}\mathbb{I}(|\varsigma_{j,t}| > D_n)\}]^4 \le CD_n^4\exp(-CD_n^{\gamma_2/2}), \tag{32}$$

which implies that

$$\sup_{1\le j\le r}\mathbb{E}\left(\left|\frac{1}{n^{1/2}}\sum_{t=1}^n \varsigma_{j,t}^-\right|^2\right) \le CD_n^2\exp(-CD_n^{\gamma_2/2}) + CD_n^2\exp(-CD_n^{\gamma_2/2})\sum_{k=1}^{n-1}\exp(-Ck^{\gamma_3}) \le CD_n^2\exp(-CD_n^{\gamma_2/2}).$$

Thus, it follows from Markov inequality that

$$\mathbb{P}\left(\left|\frac{1}{n^{1/2}}\sum_{t=1}^n \boldsymbol{\varsigma}_t^-\right|_\infty > \varepsilon_2\right) \le \frac{r}{\varepsilon_2^2}\sup_{1\le j\le r}\mathbb{E}\left(\left|\frac{1}{n^{1/2}}\sum_{t=1}^n \varsigma_{j,t}^-\right|^2\right) \le Cr\varepsilon_2^{-2}D_n^2\exp(-CD_n^{\gamma_2/2}).$$

Similar to (26), it holds that $\sup_{x>0}\mathbb{P}(x-\varepsilon_2 < |\mathbf{u}|_\infty \le x+\varepsilon_2) \le C\varepsilon_2(\log p)^{1/2}$ for any $\varepsilon_2 \to 0$. If we choose $\varepsilon_2 = (\log p)^{-1}$ and $D_n = C(\log p)^{2/\gamma_2}$ for some sufficiently large C, then $\sup_{x>0}\mathbb{P}(x-\varepsilon_2 < |\mathbf{u}|_\infty \le x+\varepsilon_2) + \mathbb{P}(|n^{-1/2}\sum_{t=1}^n \boldsymbol{\varsigma}_t^-|_\infty > \varepsilon_2) = o(1)$. Therefore, (30) implies $d_2 \le d_3 + o(1)$. To show (28) that $d_2 = o(1)$, it suffices to prove $d_3 = o(1)$.

Let $\boldsymbol{\varsigma}_t^{+,\mathrm{ext}} = (\boldsymbol{\varsigma}_t^{+,\mathrm{T}}, -\boldsymbol{\varsigma}_t^{+,\mathrm{T}})^\mathrm{T} = (\varsigma_{1,t}^{+,\mathrm{ext}},\ldots,\varsigma_{2r,t}^{+,\mathrm{ext}})^\mathrm{T}$ and $\mathbf{u}^{\mathrm{ext}} = (\mathbf{u}^\mathrm{T}, -\mathbf{u}^\mathrm{T})^\mathrm{T} = (u_1^{\mathrm{ext}},\ldots,u_{2r}^{\mathrm{ext}})^\mathrm{T}$. To prove $d_3 = \sup_{x>0}|\mathbb{P}(|n^{-1/2}\sum_{t=1}^n \boldsymbol{\varsigma}_t^+|_\infty > x) - \mathbb{P}(|\mathbf{u}|_\infty > x)| \to 0$, it is equivalent to show $\sup_{x>0}|\mathbb{P}(\max_{1\le j\le 2r} n^{-1/2}\sum_{t=1}^n \varsigma_{j,t}^{+,\mathrm{ext}} > x) - \mathbb{P}(\max_{1\le j\le 2r} u_j^{\mathrm{ext}} > x)| \to 0$. From Theorem B.1 of Chernozhukov et al. (2014), $\sup_{z\in\mathbb{R}}|\mathbb{P}(\max_{1\le j\le 2r} n^{-1/2}\sum_{t=1}^n \varsigma_{j,t}^{+,\mathrm{ext}} > z) - \mathbb{P}(\max_{1\le j\le 2r} u_j^{\mathrm{ext}} > z)| \to 0$ if $|\mathrm{Var}(n^{-1/2}\sum_{t=1}^n \boldsymbol{\varsigma}_t^{+,\mathrm{ext}}) - \mathrm{Var}(\mathbf{u}^{\mathrm{ext}})|_\infty \to 0$. Notice that $|\mathrm{Var}(n^{-1/2}\sum_{t=1}^n \boldsymbol{\varsigma}_t^{+,\mathrm{ext}}) - \mathrm{Var}(\mathbf{u}^{\mathrm{ext}})|_\infty = |\mathrm{Var}(n^{-1/2}\sum_{t=1}^n \boldsymbol{\varsigma}_t^+) - \mathrm{Var}(\mathbf{u})|_\infty$, thus to show $d_3 = o(1)$, it suffices to show

$$d_4 := \sup_{z\in\mathbb{R}}\left|\mathbb{P}\left(\max_{1\le j\le r} n^{-1/2}\sum_{t=1}^n \varsigma_{j,t}^+ > z\right) - \mathbb{P}\left(\max_{1\le j\le r} u_j > z\right)\right| \to 0.$$

By Theorem B.1 of Chernozhukov et al. (2014), it holds that $d_4 \le Cn^{-C} + Cm\exp(-Ch^{\gamma_3})$ provided that

$$hb^{-1}(\log p)^2 \le Cn^{-\varpi} \quad\text{and}\quad b^2D_n^2\log p + bD_n^2(\log p)^7 \le Cn^{1-2\varpi} \tag{33}$$

for some $\varpi \in (0, 1/4)$. As we mentioned above, $D_n \asymp (\log p)^{2/\gamma_2}$. To make p diverge as fast as possible, we can take $h \asymp (\log n)^\vartheta$ for some $\vartheta > 0$. Then (33) becomes

$$\begin{cases} C(\log n)^\vartheta n^\varpi(\log p)^2 \le b;\\ C(\log n)^{2\vartheta}(\log p)^{4/\gamma_2+5} \le n^{1-4\varpi};\\ C(\log n)^\vartheta(\log p)^{4/\gamma_2+9} \le n^{1-3\varpi}.\end{cases}$$

Please cite this article in press as: Chang J., et al., Confidence regions for entries of a large precision matrix. Journal of Econometrics (2018), https://doi.org/10.1016/j.jeconom.2018.03.020.

Therefore, $\log p = o(n^{\varphi})$ where $\varphi = \min\{\frac{(1-4\varpi)\gamma_2}{4+5\gamma_2}, \frac{(1-3\varpi)\gamma_2}{4+9\gamma_2}\}$. Notice that φ takes the supremum when $\varpi = 0$. Hence, if $\log p = o\{n^{\gamma_2/(4+9\gamma_2)}\}$, it holds that $d_4 \to 0$. Then we construct the result (28).

Analogously, to show (29), it suffices to show $\sup_{z\in\mathbb{R}}|\mathbb{P}(\max_{1\le j\le r} u_j > z) - \mathbb{P}(\max_{1\le j\le r}\xi_j > z)| \to 0$. Write $\widetilde{\mathbf{W}}$ as the covariance of $\mathbf{u}$. Recall $\mathbf{W}$ denotes the covariance of $\boldsymbol{\xi}$. Lemma 3.1 of Chernozhukov et al. (2013) leads to

$$\sup_{z\in\mathbb{R}}\bigg|\mathbb{P}\bigg(\max_{1\le j\le r} u_j > z\bigg) - \mathbb{P}\bigg(\max_{1\le j\le r}\xi_j > z\bigg)\bigg| \le C|\widetilde{\mathbf{W}} - \mathbf{W}|_\infty^{1/3}\{1 \vee \log(r/|\widetilde{\mathbf{W}} - \mathbf{W}|_\infty)\}^{2/3}. \tag{34}$$

We will specify the convergence rate of $|\widetilde{\mathbf{W}} - \mathbf{W}|_\infty$ below. Notice that, for any $1 \le j_1, j_2 \le r$, we have

$$\begin{aligned}
&\frac{1}{mb}\sum_{\ell=1}^m \mathbb{E}\bigg\{\bigg(\sum_{t\in\mathcal{I}_\ell}\varsigma_{j_1,t}^+\bigg)\bigg(\sum_{t\in\mathcal{I}_\ell}\varsigma_{j_2,t}^+\bigg)\bigg\} - \frac{1}{mb}\sum_{\ell=1}^m \mathbb{E}\bigg\{\bigg(\sum_{t\in\mathcal{I}_\ell}\varsigma_{j_1,t}\bigg)\bigg(\sum_{t\in\mathcal{I}_\ell}\varsigma_{j_2,t}\bigg)\bigg\}\\
&\quad = -\frac{1}{mb}\sum_{\ell=1}^m \mathbb{E}\bigg\{\bigg(\sum_{t\in\mathcal{I}_\ell}\varsigma_{j_1,t}^-\bigg)\bigg(\sum_{t\in\mathcal{I}_\ell}\varsigma_{j_2,t}^-\bigg)\bigg\} - \frac{1}{mb}\sum_{\ell=1}^m \mathbb{E}\bigg\{\bigg(\sum_{t\in\mathcal{I}_\ell}\varsigma_{j_1,t}^+\bigg)\bigg(\sum_{t\in\mathcal{I}_\ell}\varsigma_{j_2,t}^-\bigg)\bigg\}\\
&\qquad - \frac{1}{mb}\sum_{\ell=1}^m \mathbb{E}\bigg\{\bigg(\sum_{t\in\mathcal{I}_\ell}\varsigma_{j_1,t}^-\bigg)\bigg(\sum_{t\in\mathcal{I}_\ell}\varsigma_{j_2,t}^+\bigg)\bigg\}.
\end{aligned}$$

The triangle inequality yields

$$\begin{aligned}
&\bigg|\frac{1}{mb}\sum_{\ell=1}^m \mathbb{E}\bigg\{\bigg(\sum_{t\in\mathcal{I}_\ell}\varsigma_{j_1,t}^+\bigg)\bigg(\sum_{t\in\mathcal{I}_\ell}\varsigma_{j_2,t}^+\bigg)\bigg\} - \frac{1}{mb}\sum_{\ell=1}^m \mathbb{E}\bigg\{\bigg(\sum_{t\in\mathcal{I}_\ell}\varsigma_{j_1,t}\bigg)\bigg(\sum_{t\in\mathcal{I}_\ell}\varsigma_{j_2,t}\bigg)\bigg\}\bigg|\\
&\quad \le \frac{1}{mb}\sum_{\ell=1}^m \bigg|\mathbb{E}\bigg\{\bigg(\sum_{t\in\mathcal{I}_\ell}\varsigma_{j_1,t}^-\bigg)\bigg(\sum_{t\in\mathcal{I}_\ell}\varsigma_{j_2,t}^-\bigg)\bigg\}\bigg| + \frac{1}{mb}\sum_{\ell=1}^m \bigg|\mathbb{E}\bigg\{\bigg(\sum_{t\in\mathcal{I}_\ell}\varsigma_{j_1,t}^+\bigg)\bigg(\sum_{t\in\mathcal{I}_\ell}\varsigma_{j_2,t}^-\bigg)\bigg\}\bigg|\\
&\qquad + \frac{1}{mb}\sum_{\ell=1}^m \bigg|\mathbb{E}\bigg\{\bigg(\sum_{t\in\mathcal{I}_\ell}\varsigma_{j_1,t}^-\bigg)\bigg(\sum_{t\in\mathcal{I}_\ell}\varsigma_{j_2,t}^+\bigg)\bigg\}\bigg|.
\end{aligned}$$

For each $\ell = 1, \ldots, m$, the following identities hold:

$$\begin{aligned}
\mathbb{E}\bigg\{\bigg(\sum_{t\in\mathcal{I}_\ell}\varsigma_{j_1,t}^-\bigg)\bigg(\sum_{t\in\mathcal{I}_\ell}\varsigma_{j_2,t}^-\bigg)\bigg\} &= \sum_{t\in\mathcal{I}_\ell}\mathbb{E}(\varsigma_{j_1,t}^-\varsigma_{j_2,t}^-) + \sum_{t_1\ne t_2}\mathbb{E}(\varsigma_{j_1,t_1}^-\varsigma_{j_2,t_2}^-),\\
\mathbb{E}\bigg\{\bigg(\sum_{t\in\mathcal{I}_\ell}\varsigma_{j_1,t}^+\bigg)\bigg(\sum_{t\in\mathcal{I}_\ell}\varsigma_{j_2,t}^-\bigg)\bigg\} &= \sum_{t\in\mathcal{I}_\ell}\mathbb{E}(\varsigma_{j_1,t}^+\varsigma_{j_2,t}^-) + \sum_{t_1\ne t_2}\mathbb{E}(\varsigma_{j_1,t_1}^+\varsigma_{j_2,t_2}^-),\\
\mathbb{E}\bigg\{\bigg(\sum_{t\in\mathcal{I}_\ell}\varsigma_{j_1,t}^-\bigg)\bigg(\sum_{t\in\mathcal{I}_\ell}\varsigma_{j_2,t}^+\bigg)\bigg\} &= \sum_{t\in\mathcal{I}_\ell}\mathbb{E}(\varsigma_{j_1,t}^-\varsigma_{j_2,t}^+) + \sum_{t_1\ne t_2}\mathbb{E}(\varsigma_{j_1,t_1}^-\varsigma_{j_2,t_2}^+).
\end{aligned}$$

Together with the triangle inequality, Davydov inequality and Cauchy–Schwarz inequality, we have

$$\begin{aligned}
\bigg|\mathbb{E}\bigg\{\bigg(\sum_{t\in\mathcal{I}_\ell}\varsigma_{j_1,t}^-\bigg)\bigg(\sum_{t\in\mathcal{I}_\ell}\varsigma_{j_2,t}^-\bigg)\bigg\}\bigg| &\le Cb\sup_{j,t}[\mathbb{E}\{(\varsigma_{j,t}^-)^4\}]^{1/2},\\
\bigg|\mathbb{E}\bigg\{\bigg(\sum_{t\in\mathcal{I}_\ell}\varsigma_{j_1,t}^+\bigg)\bigg(\sum_{t\in\mathcal{I}_\ell}\varsigma_{j_2,t}^-\bigg)\bigg\}\bigg| &\le Cb\sup_{j,t}[\mathbb{E}\{(\varsigma_{j,t}^+)^4\}]^{1/4}\sup_{j,t}[\mathbb{E}\{(\varsigma_{j,t}^-)^4\}]^{1/4},\\
\bigg|\mathbb{E}\bigg\{\bigg(\sum_{t\in\mathcal{I}_\ell}\varsigma_{j_1,t}^-\bigg)\bigg(\sum_{t\in\mathcal{I}_\ell}\varsigma_{j_2,t}^+\bigg)\bigg\}\bigg| &\le Cb\sup_{j,t}[\mathbb{E}\{(\varsigma_{j,t}^+)^4\}]^{1/4}\sup_{j,t}[\mathbb{E}\{(\varsigma_{j,t}^-)^4\}]^{1/4}.
\end{aligned}$$

From (32), it holds that

$$\sup_{1\le j_1,j_2\le r}\bigg|\frac{1}{mb}\sum_{\ell=1}^m \mathbb{E}\bigg\{\bigg(\sum_{t\in\mathcal{I}_\ell}\varsigma_{j_1,t}^+\bigg)\bigg(\sum_{t\in\mathcal{I}_\ell}\varsigma_{j_2,t}^+\bigg)\bigg\} - \frac{1}{mb}\sum_{\ell=1}^m \mathbb{E}\bigg\{\bigg(\sum_{t\in\mathcal{I}_\ell}\varsigma_{j_1,t}\bigg)\bigg(\sum_{t\in\mathcal{I}_\ell}\varsigma_{j_2,t}\bigg)\bigg\}\bigg| \le CD_n\exp(-CD_n^{\gamma_2/2}).$$

By the proof of Lemma 2 in Chang et al. (2015), we can prove that

$$\sup_{1\le j_1,j_2\le r}\bigg|\frac{1}{mb}\sum_{\ell=1}^m \mathbb{E}\bigg\{\bigg(\sum_{t\in\mathcal{I}_\ell}\varsigma_{j_1,t}\bigg)\bigg(\sum_{t\in\mathcal{I}_\ell}\varsigma_{j_2,t}\bigg)\bigg\} - \frac{1}{n}\mathbb{E}\bigg\{\bigg(\sum_{t=1}^n\varsigma_{j_1,t}\bigg)\bigg(\sum_{t=1}^n\varsigma_{j_2,t}\bigg)\bigg\}\bigg| \le Ch^{1/2}b^{-1/2} + Cbn^{-1}. \tag{35}$$

Specifically, notice that

$$\begin{aligned}\mathbb{E}\bigg\{\bigg(\sum_{t=1}^{n}\varsigma_{j_1,t}\bigg)\bigg(\sum_{t=1}^{n}\varsigma_{j_2,t}\bigg)\bigg\} = &\sum_{\ell=1}^{m}\mathbb{E}\bigg\{\bigg(\sum_{t\in\mathcal{I}_\ell}\varsigma_{j_1,t}\bigg)\bigg(\sum_{t\in\mathcal{I}_\ell}\varsigma_{j_2,t}\bigg)\bigg\} + \sum_{\ell_1\neq\ell_2}\mathbb{E}\bigg\{\bigg(\sum_{t\in\mathcal{I}_{\ell_1}}\varsigma_{j_1,t}\bigg)\bigg(\sum_{t\in\mathcal{I}_{\ell_2}}\varsigma_{j_2,t}\bigg)\bigg\}\\ &+\sum_{\ell=1}^{m+1}\mathbb{E}\bigg\{\bigg(\sum_{t\in\mathcal{I}_\ell}\varsigma_{j_1,t}\bigg)\bigg(\sum_{t\in\mathcal{J}_\ell}\varsigma_{j_2,t}\bigg)\bigg\} + \sum_{\ell_1\neq\ell_2}\mathbb{E}\bigg\{\bigg(\sum_{t\in\mathcal{I}_{\ell_1}}\varsigma_{j_1,t}\bigg)\bigg(\sum_{t\in\mathcal{J}_{\ell_2}}\varsigma_{j_2,t}\bigg)\bigg\}\\ &+\sum_{\ell=1}^{m+1}\mathbb{E}\bigg\{\bigg(\sum_{t\in\mathcal{J}_\ell}\varsigma_{j_1,t}\bigg)\bigg(\sum_{t\in\mathcal{I}_\ell}\varsigma_{j_2,t}\bigg)\bigg\} + \sum_{\ell_1\neq\ell_2}\mathbb{E}\bigg\{\bigg(\sum_{t\in\mathcal{J}_{\ell_1}}\varsigma_{j_1,t}\bigg)\bigg(\sum_{t\in\mathcal{I}_{\ell_2}}\varsigma_{j_2,t}\bigg)\bigg\}\\ &+\sum_{\ell=1}^{m+1}\mathbb{E}\bigg\{\bigg(\sum_{t\in\mathcal{J}_\ell}\varsigma_{j_1,t}\bigg)\bigg(\sum_{t\in\mathcal{J}_\ell}\varsigma_{j_2,t}\bigg)\bigg\} + \sum_{\ell_1\neq\ell_2}\mathbb{E}\bigg\{\bigg(\sum_{t\in\mathcal{J}_{\ell_1}}\varsigma_{j_1,t}\bigg)\bigg(\sum_{t\in\mathcal{J}_{\ell_2}}\varsigma_{j_2,t}\bigg)\bigg\}\end{aligned} \tag{36}$$

where we set $\mathcal{I}_{m+1}=\emptyset$. By Cauchy–Schwarz inequality and Davydov inequality, we have

$$\begin{aligned}&\bigg|\frac{1}{mb}\sum_{\ell=1}^{m}\mathbb{E}\bigg\{\bigg(\sum_{t\in\mathcal{I}_\ell}\varsigma_{j_1,t}\bigg)\bigg(\sum_{t\in\mathcal{I}_\ell}\varsigma_{j_2,t}\bigg)\bigg\} - \frac{1}{n}\sum_{\ell=1}^{m}\mathbb{E}\bigg\{\bigg(\sum_{t\in\mathcal{I}_\ell}\varsigma_{j_1,t}\bigg)\bigg(\sum_{t\in\mathcal{I}_\ell}\varsigma_{j_2,t}\bigg)\bigg\}\bigg|\\ &\quad=\frac{n-mb}{nm}\sum_{\ell=1}^{m}\bigg|\mathbb{E}\bigg\{\bigg(\frac{1}{\sqrt{b}}\sum_{t\in\mathcal{I}_\ell}\varsigma_{j_1,t}\bigg)\bigg(\frac{1}{\sqrt{b}}\sum_{t\in\mathcal{I}_\ell}\varsigma_{j_2,t}\bigg)\bigg\}\bigg|\\ &\quad\le\frac{mh+b}{nm}\times Cm\le Chb^{-1}+Cbn^{-1}\end{aligned}$$

and

$$\begin{aligned}\bigg|\frac{1}{n}\sum_{\ell_1\neq\ell_2}\mathbb{E}\bigg\{\bigg(\sum_{t\in\mathcal{I}_{\ell_1}}\varsigma_{j_1,t}\bigg)\bigg(\sum_{t\in\mathcal{I}_{\ell_2}}\varsigma_{j_2,t}\bigg)\bigg\}\bigg| &\le \frac{b}{n}\sum_{\ell_1\neq\ell_2}\bigg|\mathbb{E}\bigg\{\bigg(\frac{1}{\sqrt{b}}\sum_{t\in\mathcal{I}_{\ell_1}}\varsigma_{j_1,t}\bigg)\bigg(\frac{1}{\sqrt{b}}\sum_{t\in\mathcal{I}_{\ell_2}}\varsigma_{j_2,t}\bigg)\bigg\}\bigg|\\ &\le Cbn^{-1}\sum_{\ell_1\neq\ell_2}\exp\{-C|(\ell_1-\ell_2)b|^{\gamma_3}\}\le Cbn^{-1}.\end{aligned}$$

Similarly, we can bound the other terms in (36). Therefore, we have (35) holds which implies that $|\widetilde{\mathbf{W}}-\mathbf{W}|_\infty\le Ch^{1/2}b^{-1/2}+Cbn^{-1}+CD_n\exp(-CD_n^{\gamma_2/2})$. For b, h and D_n specified above, (34) implies $\sup_{z\in\mathbb{R}}|\mathbb{P}(\max_{1\le j\le r}u_j>z)-\mathbb{P}(\max_{1\le j\le r}\xi_j>z)|\to 0$. Then we construct the result (29). Hence, we complete the proof of Theorem 1. □

Lemma 4. *Assume Conditions 1 and 3 hold, the kernel function $\mathcal{K}(\cdot)$ satisfies $|\mathcal{K}(x)|\asymp|x|^{-\tau}$ as $x\to\infty$ for some $\tau>1$, and the bandwidth $S_n\asymp n^\rho$ for some $0<\rho<\min\{\frac{\tau-1}{3\tau},\frac{\gamma_3}{2\gamma_3+1}\}$. Let $\kappa=\max\{\frac{1}{2\gamma_3+1},\frac{\rho\tau-\rho+2}{\tau+1+\gamma_3},\frac{\rho\tau+1}{\tau}\}$, and α_0 be the maximizer for the function $f(\alpha)=\min\{1-\alpha-2\rho,2(\alpha-\rho)\tau-2\}$ over $\kappa<\alpha<1-2\rho$. Then*

$$\bigg|\sum_{k=0}^{n-1}\mathcal{K}\bigg(\frac{k}{S_n}\bigg)\bigg[\frac{1}{n}\sum_{t=k+1}^{n}\{\boldsymbol{\eta}_t\boldsymbol{\eta}_{t-k}^{\mathrm{T}}-\mathbb{E}(\boldsymbol{\eta}_t\boldsymbol{\eta}_{t-k}^{\mathrm{T}})\}\bigg]\bigg|_\infty = O_p\big(\{\log(pn)\}^{4/\gamma_2}n^{-f(\alpha_0)/2}\big)$$

provided that $\log p\le Cn^{C\delta}$ where $\delta=\min[\frac{\gamma_2}{\gamma_2+8}(2\alpha_0\gamma_3+\alpha-1),\frac{\gamma_2}{8}\{(\alpha_0-\rho)\tau+\alpha_0+\alpha_0\gamma_3+\rho-2\}]$.

Proof. We first construct an upper bound for $\sup_{1\le j_1,j_2\le r}\mathbb{P}\{|\sum_{k=0}^{n-1}\mathcal{K}(k/S_n)[n^{-1}\sum_{t=k+1}^{n}\{\eta_{j_1,t}\eta_{j_2,t-k}-\mathbb{E}(\eta_{j_1,t}\eta_{j_2,t-k})\}]|>x\}$. For any j_1 and j_2, it holds that

$$\begin{aligned}&\mathbb{P}\bigg\{\bigg|\sum_{k=0}^{n-1}\mathcal{K}\bigg(\frac{k}{S_n}\bigg)\bigg[\frac{1}{n}\sum_{t=k+1}^{n}\{\eta_{j_1,t}\eta_{j_2,t-k}-\mathbb{E}(\eta_{j_1,t}\eta_{j_2,t-k})\}\bigg]\bigg|>x\bigg\}\\ &\quad\le\mathbb{P}\bigg\{\sum_{k=0}^{\lfloor Cn^\alpha\rfloor}\bigg|\mathcal{K}\bigg(\frac{k}{S_n}\bigg)\bigg|\bigg|\frac{1}{n}\sum_{t=1}^{n-k}\psi_{t,k}\bigg|>\frac{x}{2}\bigg\}+\mathbb{P}\bigg\{\sum_{k=\lfloor Cn^\alpha\rfloor+1}^{n-1}\bigg|\mathcal{K}\bigg(\frac{k}{S_n}\bigg)\bigg|\bigg|\frac{1}{n}\sum_{t=1}^{n-k}\psi_{t,k}\bigg|>\frac{x}{2}\bigg\}\end{aligned} \tag{37}$$

for any $\alpha\in(0,1)$, where $\psi_{t,k}=\eta_{j_1,t+k}\eta_{j_2,t}-\mathbb{E}(\eta_{j_1,t+k}\eta_{j_2,t})$. Following Lemma 2 of Chang et al. (2013), it holds that

$$\sup_{0\le k\le n-1}\sup_{1\le t\le n-k}\mathbb{P}\left(|\psi_{t,k}|>x\right)\le C\exp(-Cx^{\gamma_2/4}) \tag{38}$$

for any $x > 0$. Notice that $S_n \asymp n^\rho$, we have $\max_{\lfloor Cn^\alpha \rfloor+1\le k\le n-1}|\mathcal{K}(k/S_n)| \le Cn^{-(\alpha-\rho)\tau}$ if $\alpha > \rho$. Then, (38) leads to

$$\begin{aligned}\mathbb{P}\bigg\{\sum_{k=\lfloor Cn^\alpha\rfloor+1}^{n-1}\bigg|\mathcal{K}\bigg(\frac{k}{S_n}\bigg)\bigg|\bigg|\frac{1}{n}\sum_{t=1}^{n-k}\psi_{t,k}\bigg| > \frac{x}{2}\bigg\} &\le \sum_{k=\lfloor Cn^\alpha\rfloor+1}^{n-1}\mathbb{P}\bigg\{\bigg|\frac{1}{n}\sum_{t=1}^{n-k}\psi_{t,k}\bigg| > Cxn^{(\alpha-\rho)\tau-1}\bigg\}\\ &\le \sum_{k=\lfloor Cn^\alpha\rfloor+1}^{n-1}\sum_{t=1}^{n-k}\mathbb{P}\{|\psi_{t,k}| > Cxn^{(\alpha-\rho)\tau-1}\}\\ &\le Cn^2\exp[-C\{xn^{(\alpha-\rho)\tau-1}\}^{\gamma_2/4}].\end{aligned} \tag{39}$$

We will specify the upper bound for $\mathbb{P}\{\sum_{k=0}^{\lfloor Cn^\alpha\rfloor}|\mathcal{K}(k/S_n)||n^{-1}\sum_{t=1}^{n-k}\psi_{t,k}| > x/2\}$ below. Similar to (38), we have that

$$\sup_{1\le j_1,j_2\le r}\ \sup_{0\le k\le n-1}\ \sup_{1\le t\le n-k}\ \mathbb{P}(|\eta_{j_1,t+k}\eta_{j_2,t}| > x) \le C\exp(-Cx^{\gamma_2/4}) \tag{40}$$

for any $x > 0$. Denote by $\mathcal{T}$ the event $\{\sup_{0\le k\le n-1}\sup_{1\le t\le n-k}|\eta_{j_1,t+k}\eta_{j_2,t}| > M\}$. For each $k = 0,\ldots,\lfloor Cn^\alpha\rfloor$, let $\psi_{t,k}^+ = \eta_{j_1,t+k}\eta_{j_2,t}\mathbb{I}\{|\eta_{j_1,t+k}\eta_{j_2,t}| \le M\} - \mathbb{E}[\eta_{j_1,t+k}\eta_{j_2,t}\mathbb{I}\{|\eta_{j_1,t+k}\eta_{j_2,t}| \le M\}]$ for $t = 1,\ldots,n-k$. Write $D = \sum_{k=0}^{\lfloor Cn^\alpha\rfloor}|\mathcal{K}(k/S_n)|$, then

$$\begin{aligned}\mathbb{P}\bigg\{\sum_{k=0}^{\lfloor Cn^\alpha\rfloor}\bigg|\mathcal{K}\bigg(\frac{k}{S_n}\bigg)\bigg|\bigg|\frac{1}{n}\sum_{t=1}^{n-k}\psi_{t,k}\bigg| > \frac{x}{2}\bigg\} &\le \sum_{k=0}^{\lfloor Cn^\alpha\rfloor}\mathbb{P}\bigg(\bigg|\frac{1}{n}\sum_{t=1}^{n-k}\psi_{t,k}\bigg| > \frac{x}{2D},\ \mathcal{T}^c\bigg) + \mathbb{P}(\mathcal{T})\\ &\le \sum_{k=0}^{\lfloor Cn^\alpha\rfloor}\mathbb{P}\bigg(\bigg|\frac{1}{n}\sum_{t=1}^{n-k}\psi_{t,k}^+\bigg| > \frac{x}{4D}\bigg) + \mathbb{P}(\mathcal{T})\\ &\quad + \sum_{k=0}^{\lfloor Cn^\alpha\rfloor}\mathbb{P}\bigg(\frac{1}{n}\sum_{t=1}^{n-k}\mathbb{E}[|\eta_{j_1,t+k}\eta_{j_2,t}|\mathbb{I}\{|\eta_{j_1,t+k}\eta_{j_2,t}| > M\}] > \frac{x}{4D}\bigg).\end{aligned} \tag{41}$$

From (40), we have $\mathbb{P}(\mathcal{T}) \le Cn^2\exp(-CM^{\gamma_2/4})$. Similar to (31), we have

$$\sup_{1\le j_1,j_2\le r}\ \sup_{0\le k\le n-1}\ \sup_{1\le t\le n-k}\ \mathbb{E}[|\eta_{j_1,t+k}\eta_{j_2,t}|\mathbb{I}\{|\eta_{j_1,t+k}\eta_{j_2,t}| > M\}] \le CM\exp(-CM^{\gamma_2/4}).$$

If $DMx^{-1}\exp(-CM^{\gamma_2/4}) \to 0$, then (41) yields that

$$\mathbb{P}\bigg\{\sum_{k=0}^{\lfloor Cn^\alpha\rfloor}\bigg|\mathcal{K}\bigg(\frac{k}{S_n}\bigg)\bigg|\bigg|\frac{1}{n}\sum_{t=1}^{n-k}\psi_{t,k}\bigg| > \frac{x}{2}\bigg\} \le \sum_{k=0}^{\lfloor Cn^\alpha\rfloor}\mathbb{P}\bigg(\bigg|\frac{1}{n}\sum_{t=1}^{n-k}\psi_{t,k}^+\bigg| > \frac{x}{4D}\bigg) + Cn^2\exp(-CM^{\gamma_2/4}). \tag{42}$$

For each $k = 0,\ldots,\lfloor Cn^\alpha\rfloor$, we first consider $\mathbb{P}\{n^{-1}\sum_{t=1}^{n-k}\psi_{t,k}^+ > x/(4D)\}$. By Markov inequality, it holds that

$$\mathbb{P}\bigg(\frac{1}{n}\sum_{t=1}^{n-k}\psi_{t,k}^+ > \frac{x}{4D}\bigg) \le \exp\bigg(-\frac{unx}{4D}\bigg)\mathbb{E}\bigg\{\exp\bigg(\sum_{t=1}^{n-k}u\psi_{t,k}^+\bigg)\bigg\} \tag{43}$$

for any $u > 0$. Let L be a positive integer such that $L \asymp n^\alpha$ and $L \ge 3\lfloor Cn^\alpha\rfloor$ for C specified in (37). We decompose the sequence $\{1,\ldots,n-k\}$ to the following $m+1$ blocks where $m = \lfloor(n-k)/L\rfloor$: $\mathcal{G}_\ell = \{(\ell-1)L+1,\ldots,\ell L\}$ $(\ell = 1,\ldots,m)$ and $\mathcal{G}_{m+1} = \{mL+1,\ldots,n-k\}$. Additionally, let $b = \lfloor L/2\rfloor$ and $h = L-b$. We then decompose each $\mathcal{G}_\ell$ $(\ell = 1,\ldots,m)$ to a block with length b and a block with length h. Specifically, $\mathcal{I}_\ell = \{(\ell-1)L+1,\ldots,(\ell-1)L+b\}$ and $\mathcal{J}_\ell = \{(\ell-1)L+b+1,\ldots,\ell L\}$ for any $\ell = 1,\ldots,m$, and $\mathcal{I}_{m+1} = \mathcal{G}_{m+1}$. Based on these notations and Cauchy–Schwarz inequality, it holds that

$$\mathbb{E}\bigg\{\exp\bigg(\sum_{t=1}^{n-k}u\psi_{t,k}^+\bigg)\bigg\} \le \bigg[\mathbb{E}\bigg\{\exp\bigg(\sum_{\ell=1}^{m+1}\sum_{t\in\mathcal{I}_\ell}2u\psi_{t,k}^+\bigg)\bigg\}\bigg]^{1/2}\bigg[\mathbb{E}\bigg\{\exp\bigg(\sum_{\ell=1}^{m}\sum_{t\in\mathcal{J}_\ell}2u\psi_{t,k}^+\bigg)\bigg\}\bigg]^{1/2}.$$

By Lemma 2 of Merlevède et al. (2011), noticing that $b(m+1) \le 2n$, we have

$$\mathbb{E}\bigg\{\exp\bigg(\sum_{\ell=1}^{m+1}\sum_{t\in\mathcal{I}_\ell}2u\psi_{t,k}^+\bigg)\bigg\} \le \prod_{\ell=1}^{m+1}\mathbb{E}\bigg\{\exp\bigg(\sum_{t\in\mathcal{I}_\ell}2u\psi_{t,k}^+\bigg)\bigg\} + CuMn\exp(8uMn - C|b-k|_+^{\gamma_3}). \tag{44}$$

Following the inequality $e^x \le 1 + x + x^2e^{x\vee 0}/2$ for any $x \in \mathbf{R}$, we have that

$$\mathbb{E}\bigg\{\exp\bigg(\sum_{t\in\mathcal{I}_\ell}2u\psi_{t,k}^+\bigg)\bigg\} \le 1 + 2u^2\mathbb{E}\bigg\{\bigg(\sum_{t\in\mathcal{I}_\ell}\psi_{t,k}^+\bigg)^2\bigg\}\exp(4ubM) \le\ 1 + Cu^2b^2\exp(4ubM).$$

Together with (44), following the inequality $(1+x)^{m+1} \le e^{(m+1)x}$ for any $x > 0$, and $bm \le n/2$, it holds that

$$\mathbb{E}\bigg\{\exp\bigg(\sum_{\ell=1}^{m+1}\sum_{t\in\mathcal{I}_\ell} 2u\psi_{t,k}^+\bigg)\bigg\} \le \exp\{Cu^2nb\exp(4ubM)\} + CuMn\exp(8uMn - C|b-k|_+^{\gamma_3}).$$

Similarly, we can obtain the same upper bound for $\mathbb{E}\{\exp(\sum_{\ell=1}^m\sum_{t\in\mathcal{J}_\ell} 2u\psi_{t,k}^+)\}$. Hence,

$$\mathbb{E}\bigg\{\exp\bigg(\sum_{t=1}^{n-k} u\psi_{t,k}^+\bigg)\bigg\} \le \exp\{Cu^2nb\exp(4ubM)\} + CuMn\exp\{8uMn - C|b-k|_+^{\gamma_3}).$$

We restrict $|ubM| \le C$. Notice that $b - k \ge \lfloor Cn^\alpha\rfloor/2 - 1$, then

$$\mathbb{E}\bigg\{\exp\bigg(\sum_{t=1}^{n-k} u\psi_{t,k}^+\bigg)\bigg\} \le C\exp(Cu^2nb) + CuMn\exp(8uMn - Cn^{\alpha\gamma_3}).$$

Together with (43), notice that $D \asymp S_n \asymp n^\rho$ and $b \asymp n^\alpha$, it holds that

$$\mathbb{P}\bigg(\frac{1}{n}\sum_{t=1}^{n-k}\psi_{t,k}^+ > \frac{x}{4D}\bigg) \le C\exp(-Cun^{1-\rho}x + Cu^2n^{1+\alpha}) + CuMn\exp(-Cun^{1-\rho}x + 8uMn - Cn^{\alpha\gamma_3}). \tag{45}$$

To make the upper bound in above inequality decay to zero for some $x \to 0^+$ and $M \to \infty$, we need to require $uMn^{1-\alpha\gamma_3} \le C$. For the first term on the right-hand side of above inequality, the optimal selection of u is $u \asymp xn^{-\alpha-\rho}$. Therefore, (45) can be simplified to

$$\mathbb{P}\bigg(\frac{1}{n}\sum_{t=1}^{n-k}\psi_{t,k}^+ > \frac{x}{4D}\bigg) \le C\exp(-Cn^{1-\alpha-2\rho}x^2) + C\exp(-Cn^{\alpha\gamma_3})$$

if $xMn^{1-\alpha-\alpha\gamma_3-\rho} \le C$. The same inequality also holds for $\mathbb{P}\{n^{-1}\sum_{t=1}^{n-k}\psi_{t,k}^+ < -x/(4D)\}$. Combining with (37), (39) and (42),

$$\begin{aligned}&\mathbb{P}\bigg\{\bigg|\sum_{k=0}^{n-1}\mathcal{K}\bigg(\frac{k}{S_n}\bigg)\bigg[\frac{1}{n}\sum_{t=k+1}^{n}\{\eta_{j_1,t}\eta_{j_2,t-k} - \mathbb{E}(\eta_{j_1,t}\eta_{j_2,t-k})\}\bigg]\bigg| > x\bigg\}\\ &\quad\le Cn^\alpha\exp(-Cn^{1-\alpha-2\rho}x^2) + Cn^\alpha\exp(-Cn^{\alpha\gamma_3}) + Cn^2\exp[-C\{xn^{(\alpha-\rho)\tau-1}\}^{\gamma_2/4}] + Cn^2\exp(-CM^{\gamma_2/4})\end{aligned}$$

for any $x > 0$ such that $xMn^{1-\alpha-\alpha\gamma_3-\rho} \le C$. Notice that above inequality is uniform for any j_1 and j_2, thus

$$\begin{aligned}&\mathbb{P}\bigg\{\bigg|\sum_{k=0}^{n-1}\mathcal{K}\bigg(\frac{k}{S_n}\bigg)\bigg[\frac{1}{n}\sum_{t=k+1}^{n}\{\boldsymbol{\eta}_t\boldsymbol{\eta}_{t-k}^{\mathrm{T}} - \mathbb{E}(\boldsymbol{\eta}_t\boldsymbol{\eta}_{t-k}^{\mathrm{T}})\}\bigg]\bigg|_\infty > x\bigg\}\\ &\quad\le Cp^2n^\alpha\exp(-Cn^{1-\alpha-2\rho}x^2) + Cp^2n^\alpha\exp(-Cn^{\alpha\gamma_3}) + Cp^2n^2\exp[-C\{xn^{(\alpha-\rho)\tau-1}\}^{\gamma_2/4}] + Cp^2n^2\exp(-CM^{\gamma_2/4}).\end{aligned}$$

To make the upper bound of above inequality converge to zero, x and M should satisfy the following restrictions:

$$\begin{cases} x \ge C\bigg[\sqrt{\dfrac{\log(pn)}{n^{1-\alpha-2\rho}}} \vee \dfrac{\{\log(pn)\}^{4/\gamma_2}}{n^{(\alpha-\rho)\tau-1}}\bigg],\\ M \ge C\{\log(pn)\}^{4/\gamma_2}.\end{cases} \tag{46}$$

Notice that $xMn^{1-\alpha-\alpha\gamma_3-\rho} \le C$, (46) implies that $\log p \le Cn^{C\delta}$ where $\delta = \min\{\frac{\gamma_2}{\gamma_2+8}(2\alpha\gamma_3+\alpha-1), \frac{\gamma_2}{8}\{(\alpha-\rho)\tau+\alpha+\alpha\gamma_3+\rho-2\}\}$. To make x can decay to zero and p can diverge at exponential rate of n, we need to assume $0 < \rho < \min\{\frac{\tau-1}{3\tau}, \frac{\gamma_3}{2\gamma_3+1}\}$ and $\kappa < \alpha < 1-2\rho$. Let $f(\alpha) = \min\{1-\alpha-2\rho, 2(\alpha-\rho)\tau-2\}$ and $\alpha_0 = \arg\max_{\kappa<\alpha<1-2\rho} f(\alpha)$. We select $\alpha = \alpha_0$ and $x = C\{\log(pn)\}^{4/\gamma_2}n^{-f(\alpha_0)/2}$, then

$$\mathbb{P}\bigg\{\bigg|\sum_{k=0}^{n-1}\mathcal{K}\bigg(\frac{k}{S_n}\bigg)\bigg[\frac{1}{n}\sum_{t=k+1}^{n}\{\boldsymbol{\eta}_t\boldsymbol{\eta}_{t-k}^{\mathrm{T}} - \mathbb{E}(\boldsymbol{\eta}_t\boldsymbol{\eta}_{t-k}^{\mathrm{T}})\}\bigg]\bigg|_\infty > x\bigg\} \to 0.$$

Hence, we complete the proof of Lemma 4. □

Proof of Theorem 2. Similar to the proof of (29), it suffices to prove $|\widehat{\mathbf{W}} - \mathbf{W}|_\infty = o_p(1)$. By Lemmas 1 and 3, we have $\max_{1\le j\le p}|\widehat{v}_{j,j} - v_{j,j}| = O_p\{(n^{-1}\log p)^{1/2}\}$. Notice that $v_{j,j}$'s are uniformly bounded away from zero, then $\widehat{v}_{j,j}^{-1}$'s are uniformly bounded away from infinity with probability approaching one. Thus,

$$|\widehat{\mathbf{W}} - \mathbf{W}|_\infty \le C|\widehat{\boldsymbol{\Xi}} - \boldsymbol{\Xi}|_\infty + C|\widehat{\mathbf{H}} - \mathbf{H}|_\infty = C|\widehat{\boldsymbol{\Xi}} - \boldsymbol{\Xi}|_\infty + O_p\{(n^{-1}\log p)^{1/2}\}. \tag{47}$$

Please cite this article in press as: Chang J., et al., Confidence regions for entries of a large precision matrix. Journal of Econometrics (2018), https://doi.org/10.1016/j.jeconom.2018.03.020.

We will show $|\widehat{\Xi} - \Xi|_\infty = o_p(1)$ below.

Define

$$\widetilde{\Xi} = \sum_{k=-n+1}^{n-1} \mathcal{K}\bigg(\frac{k}{S_n}\bigg)\boldsymbol{\Gamma}_k$$

where

$$\boldsymbol{\Gamma}_k = \begin{cases} \dfrac{1}{n}\displaystyle\sum_{t=k+1}^{n} \mathbb{E}(\boldsymbol{\eta}_t\boldsymbol{\eta}_{t-k}^{\mathrm{T}}), & k \geq 0; \\ \dfrac{1}{n}\displaystyle\sum_{t=-k+1}^{n} \mathbb{E}(\boldsymbol{\eta}_{t+k}\boldsymbol{\eta}_{t}^{\mathrm{T}}), & k < 0. \end{cases}$$

We will specify the convergence rates of $|\widehat{\Xi} - \widetilde{\Xi}|_\infty$ and $|\widetilde{\Xi} - \Xi|_\infty$, respectively. Notice that

$$\widehat{\Xi} - \widetilde{\Xi} = \sum_{k=0}^{n-1} \mathcal{K}\bigg(\frac{k}{S_n}\bigg)(\widehat{\boldsymbol{\Gamma}}_k - \boldsymbol{\Gamma}_k) + \sum_{k=-n+1}^{-1} \mathcal{K}\bigg(\frac{k}{S_n}\bigg)(\widehat{\boldsymbol{\Gamma}}_k - \boldsymbol{\Gamma}_k).$$

For any $k \geq 0$, it holds that

$$\widehat{\boldsymbol{\Gamma}}_k = \frac{1}{n}\sum_{t=k+1}^{n} \boldsymbol{\eta}_t\boldsymbol{\eta}_{t-k}^{\mathrm{T}} + \frac{1}{n}\sum_{t=k+1}^{n} (\widehat{\boldsymbol{\eta}}_t - \boldsymbol{\eta}_t)\boldsymbol{\eta}_{t-k}^{\mathrm{T}} + \frac{1}{n}\sum_{t=k+1}^{n} \boldsymbol{\eta}_t(\widehat{\boldsymbol{\eta}}_{t-k} - \boldsymbol{\eta}_{t-k})^{\mathrm{T}} + \frac{1}{n}\sum_{t=k+1}^{n} (\widehat{\boldsymbol{\eta}}_t - \boldsymbol{\eta}_t)(\widehat{\boldsymbol{\eta}}_{t-k} - \boldsymbol{\eta}_{t-k})^{\mathrm{T}},$$

which implies

$$\begin{aligned}\sum_{k=0}^{n-1} \mathcal{K}\bigg(\frac{k}{S_n}\bigg)(\widehat{\boldsymbol{\Gamma}}_k - \boldsymbol{\Gamma}_k) &= \sum_{k=0}^{n-1} \mathcal{K}\bigg(\frac{k}{S_n}\bigg)\bigg[\frac{1}{n}\sum_{t=k+1}^{n} \{\boldsymbol{\eta}_t\boldsymbol{\eta}_{t-k}^{\mathrm{T}} - \mathbb{E}(\boldsymbol{\eta}_t\boldsymbol{\eta}_{t-k}^{\mathrm{T}})\}\bigg] + \sum_{k=0}^{n-1} \mathcal{K}\bigg(\frac{k}{S_n}\bigg)\bigg\{\frac{1}{n}\sum_{t=k+1}^{n} (\widehat{\boldsymbol{\eta}}_t - \boldsymbol{\eta}_t)\boldsymbol{\eta}_{t-k}^{\mathrm{T}}\bigg\} \\ &\quad + \sum_{k=0}^{n-1} \mathcal{K}\bigg(\frac{k}{S_n}\bigg)\bigg\{\frac{1}{n}\sum_{t=k+1}^{n} \boldsymbol{\eta}_t(\widehat{\boldsymbol{\eta}}_{t-k} - \boldsymbol{\eta}_{t-k})^{\mathrm{T}}\bigg\} + \sum_{k=0}^{n-1} \mathcal{K}\bigg(\frac{k}{S_n}\bigg)\bigg\{\frac{1}{n}\sum_{t=k+1}^{n} (\widehat{\boldsymbol{\eta}}_t - \boldsymbol{\eta}_t)(\widehat{\boldsymbol{\eta}}_{t-k} - \boldsymbol{\eta}_{t-k})^{\mathrm{T}}\bigg\}.\end{aligned} \quad (48)$$

We will prove the $|\cdot|_\infty$-norm of the last three terms on the right-hand side of above identity are $O_p\{sS_n(n^{-1}\log p)^{1/2}\}$. We only need to show this rate for one of them and the proofs for the other two are similar. For any j and t,

$$\begin{aligned}\widehat{\eta}_{j,t} - \widehat{\eta}_{j,t} &= \big\{\widehat{\epsilon}_{\chi_1(j),t}\widehat{\epsilon}_{\chi_2(j),t} - \epsilon_{\chi_1(j),t}\epsilon_{\chi_2(j),t}\big\} - \big\{\widehat{v}_{\boldsymbol{\chi}(j)} - v_{\boldsymbol{\chi}(j)}\big\} \\ &= \widehat{\epsilon}_{\chi_1(j),t}\widehat{\epsilon}_{\chi_2(j),t} - \epsilon_{\chi_1(j),t}\epsilon_{\chi_2(j),t} + O_p\{(n^{-1}\log p)^{1/2}\} \\ &= \big\{\widehat{\boldsymbol{\alpha}}_{\chi_1(j)} - \boldsymbol{\alpha}_{\chi_1(j)}\big\}^{\mathrm{T}}\mathbf{y}_t\mathbf{y}_t^{\mathrm{T}}\big\{\widehat{\boldsymbol{\alpha}}_{\chi_2(j)} - \boldsymbol{\alpha}_{\chi_2(j)}\big\} - \epsilon_{\chi_2(j),t}\big\{\widehat{\boldsymbol{\alpha}}_{\chi_1(j)} - \boldsymbol{\alpha}_{\chi_1(j)}\big\}^{\mathrm{T}}\mathbf{y}_t - \epsilon_{\chi_1(j),t}\big\{\widehat{\boldsymbol{\alpha}}_{\chi_2(j)} - \boldsymbol{\alpha}_{\chi_2(j)}\big\}^{\mathrm{T}}\mathbf{y}_t \\ &\quad + O_p\{(n^{-1}\log p)^{1/2}\}.\end{aligned}$$

Here the term $O_p\{(n^{-1}\log p)^{1/2}\}$ is uniform for any j and t. Then the (j_1, j_2)th component of $\sum_{k=0}^{n-1}\mathcal{K}(k/S_n)\{n^{-1}\sum_{t=k+1}^{n}(\widehat{\boldsymbol{\eta}}_t - \boldsymbol{\eta}_t)\boldsymbol{\eta}_{t-k}^{\mathrm{T}}\}$ is

$$\begin{aligned}&\big\{\widehat{\boldsymbol{\alpha}}_{\chi_1(j_1)} - \boldsymbol{\alpha}_{\chi_1(j_1)}\big\}^{\mathrm{T}}\bigg\{\sum_{k=0}^{n-1}\mathcal{K}\bigg(\frac{k}{S_n}\bigg)\bigg(\frac{1}{n}\sum_{t=k+1}^{n}\eta_{j_2,t-k}\mathbf{y}_t\mathbf{y}_t^{\mathrm{T}}\bigg)\bigg\}\big\{\widehat{\boldsymbol{\alpha}}_{\chi_2(j_2)} - \boldsymbol{\alpha}_{\chi_2(j_2)}\big\} \\ &- \big\{\widehat{\boldsymbol{\alpha}}_{\chi_1(j_1)} - \boldsymbol{\alpha}_{\chi_1(j_1)}\big\}^{\mathrm{T}}\bigg\{\sum_{k=0}^{n-1}\mathcal{K}\bigg(\frac{k}{S_n}\bigg)\bigg(\frac{1}{n}\sum_{t=k+1}^{n}\mathbf{y}_t\eta_{j_2,t-k}\epsilon_{\chi_2(j_1),t}\bigg)\bigg\} \\ &- \big\{\widehat{\boldsymbol{\alpha}}_{\chi_2(j_1)} - \boldsymbol{\alpha}_{\chi_2(j_1)}\big\}^{\mathrm{T}}\bigg\{\sum_{k=0}^{n-1}\mathcal{K}\bigg(\frac{k}{S_n}\bigg)\bigg(\frac{1}{n}\sum_{t=k+1}^{n}\mathbf{y}_t\eta_{j_2,t-k}\epsilon_{\chi_1(j_1),t}\bigg)\bigg\} \\ &+ \widetilde{R}_{j_1,j_2},\end{aligned} \quad (49)$$

where

$$
\begin{aligned}
|\widetilde{R}_{j_1,j_2}| &\le \left\{\sum_{k=0}^{n-1}\left|\mathcal{K}\left(\frac{k}{S_n}\right)\right|\left(\frac{1}{n}\sum_{t=k+1}^{n}|\eta_{j_2,t-k}|\right)\right\}\cdot O_p\{(n^{-1}\log p)^{1/2}\} \\
&\le \left\{\sum_{k=0}^{n}\left|\mathcal{K}\left(\frac{k}{S_n}\right)\right|\right\}\left(\frac{1}{n}\sum_{t=1}^{n}|\eta_{j_2,t}|\right)\cdot O_p\{(n^{-1}\log p)^{1/2}\} \\
&= O_p\{S_n(n^{-1}\log p)^{1/2}\}.
\end{aligned}
$$

Here the term $O_p\{S_n(n^{-1}\log p)^{1/2}\}$ is uniform for any j_1 and j_2. Following the same arguments, we have

$$
\sup_{1\le j_1,j_2\le p}\left|\sum_{k=0}^{n-1}\mathcal{K}\left(\frac{k}{S_n}\right)\left(\frac{1}{n}\sum_{t=k+1}^{n}\eta_{j_2,t-k}\mathbf{y}_t\mathbf{y}_t^{\mathrm{T}}\right)\right|_\infty \le CS_n,
$$

$$
\sup_{1\le j_1,j_2\le p}\left|\sum_{k=0}^{n-1}\mathcal{K}\left(\frac{k}{S_n}\right)\left(\frac{1}{n}\sum_{t=k+1}^{n}\mathbf{y}_t\eta_{j_2,t-k}\epsilon_{\chi_2(j_1),t}\right)\right|_\infty \le CS_n,
$$

$$
\sup_{1\le j_1,j_2\le p}\left|\sum_{k=0}^{n-1}\mathcal{K}\left(\frac{k}{S_n}\right)\left(\frac{1}{n}\sum_{t=k+1}^{n}\mathbf{y}_t\eta_{j_2,t-k}\epsilon_{\chi_1(j_1),t}\right)\right|_\infty \le CS_n.
$$

Therefore, the (j_1,j_2)th component of $\sum_{k=0}^{n-1}\mathcal{K}(k/S_n)\{n^{-1}\sum_{t=k+1}^{n}(\widehat{\boldsymbol{\eta}}_t-\boldsymbol{\eta}_t)\boldsymbol{\eta}_{t-k}^{\mathrm{T}}\}$ can be bounded by $CS_n\sup_{1\le j\le p}|\widehat{\boldsymbol{\alpha}}_j-\boldsymbol{\alpha}_j|_1+O_p\{S_n(n^{-1}\log p)^{1/2}\}=O_p\{sS_n(n^{-1}\log p)^{1/2}\}$, where the last identity in above equation is based on (23). Therefore, from (48), by Lemma 4, we have

$$
\begin{aligned}
\left|\sum_{k=0}^{n-1}\mathcal{K}\left(\frac{k}{S_n}\right)(\widehat{\boldsymbol{\Gamma}}_k-\boldsymbol{\Gamma}_k)\right|_\infty &\le \left|\sum_{k=0}^{n-1}\mathcal{K}\left(\frac{k}{S_n}\right)\left[\frac{1}{n}\sum_{t=k+1}^{n}\{\boldsymbol{\eta}_t\boldsymbol{\eta}_{t-k}^{\mathrm{T}}-\mathbb{E}(\boldsymbol{\eta}_t\boldsymbol{\eta}_{t-k}^{\mathrm{T}})\}\right]\right|_\infty + O_p\{sS_n(n^{-1}\log p)^{1/2}\} \\
&= O_p[\{\log(pn)\}^{4/\gamma_2}n^{-f(\alpha_0)/2}]+O_p\{sS_n(n^{-1}\log p)^{1/2}\}.
\end{aligned}
$$

Analogously, we can prove the same result for $|\sum_{k=-n+1}^{-1}\mathcal{K}(k/S_n)(\widehat{\boldsymbol{\Gamma}}_k-\boldsymbol{\Gamma}_k)|_\infty$. Therefore, $|\widehat{\boldsymbol{\Xi}}-\widetilde{\boldsymbol{\Xi}}|_\infty = O_p[\{\log(pn)\}^{4/\gamma_2}n^{-f(\alpha_0)/2}]+O_p\{sS_n(n^{-1}\log p)^{1/2}\}$. Repeating the proof of Proposition 1(b) in Andrews (1991), we know the convergence in Proposition 1(b) is uniformly for each component of $\widetilde{\boldsymbol{\Xi}}-\boldsymbol{\Xi}$. Thus, $|\widetilde{\boldsymbol{\Xi}}-\boldsymbol{\Xi}|_\infty=o(1)$. Then $|\widehat{\boldsymbol{\Xi}}-\boldsymbol{\Xi}|_\infty=o_p(1)$. Similar to (34), we complete the proof. □

Proof of Corollary 1. From Theorem 2, it holds that $\mathbb{P}_{H_0}(\mathbf{c}\in\mathcal{C}_{\mathcal{S},1-\alpha,1})\to 1-\alpha$. Therefore, $\mathbb{P}_{H_0}(\Psi_\alpha=1)=\mathbb{P}_{H_0}(\mathbf{c}\notin\mathcal{C}_{\mathcal{S},1-\alpha,1})\to\alpha$ which establishes part (i). For part (ii), the following standard results on Gaussian maximum hold:

$$
\mathbb{E}\big(|\widehat{\boldsymbol{\xi}}|_\infty|\mathcal{Y}_n\big)\le\{1+(2\log p)^{-1}\}(2\log p)^{1/2}\max_{1\le j\le r}\widehat{w}_{j,j}^{1/2}
$$

and

$$
\mathbb{P}\Big\{|\widehat{\boldsymbol{\xi}}|_\infty\ge\mathbb{E}\big(|\widehat{\boldsymbol{\xi}}|_\infty|\mathcal{Y}_n\big)+u|\mathcal{Y}_n\Big\}\le\exp\left(-\frac{u^2}{2\max_{1\le j\le p}\widehat{w}_{j,j}}\right)
$$

for any $u>0$. Then, $\widehat{q}_{\mathcal{S},1-\alpha,1}\le[\{1+(2\log p)^{-1}\}(2\log p)^{1/2}+\{2\log(1/\alpha)\}^{1/2}]\max_{1\le j\le r}\widehat{w}_{j,j}^{1/2}$. Let $\mathscr{T}_\varepsilon=\{\max_{1\le j\le r}|\widehat{w}_{j,j}^{1/2}-w_{j,j}^{1/2}|/w_{j,j}^{1/2}\le\varepsilon\}$ for some $\varepsilon>0$. Restricted on $\mathscr{T}_\varepsilon$, $\widehat{q}_{\mathcal{S},1-\alpha,1}\le(1+\varepsilon)[\{1+(2\log p)^{-1}\}(2\log p)^{1/2}+\{2\log(1/\alpha)\}^{1/2}]$ $\max_{1\le j\le r}w_{j,j}^{1/2}$. Let $(\tilde{j}_1,\tilde{j}_2)=\arg\max_{(j_1,j_2)\in\mathcal{S}}|\omega_{j_1,j_2}-c_{j_1,j_2}|$. Without lose of generality, we assume $\omega_{\tilde{j}_1,\tilde{j}_2}-c_{\tilde{j}_1,\tilde{j}_2}>0$. Therefore,

$$
\begin{aligned}
\mathbb{P}_{H_1}(\Psi_\alpha=1)&=\mathbb{P}_{H_1}\left\{\max_{(j_1,j_2)\in\mathcal{S}}n^{1/2}|\widehat{\omega}_{j_1,j_2}-c_{j_1,j_2}|>\widehat{q}_{\mathcal{S},1-\alpha,1}\right\} \\
&\ge\mathbb{P}_{H_1}\left\{n^{1/2}(\widehat{\omega}_{\tilde{j}_1,\tilde{j}_2}-c_{\tilde{j}_1,\tilde{j}_2})>\widehat{q}_{\mathcal{S},1-\alpha,1}\right\} \\
&=1-\mathbb{P}_{H_1}\left\{n^{1/2}(\widehat{\omega}_{\tilde{j}_1,\tilde{j}_2}-c_{\tilde{j}_1,\tilde{j}_2})\le\widehat{q}_{\mathcal{S},1-\alpha,1},\ \mathscr{T}_\varepsilon\right\}-\mathbb{P}(\mathscr{T}_\varepsilon^c).
\end{aligned}
$$

Restricted on $\mathscr{T}_\varepsilon$, if $\varepsilon\to 0$, it holds that $\widehat{q}_{\mathcal{S},1-\alpha,1}-(\omega_{\tilde{j}_1,\tilde{j}_2}-c_{\tilde{j}_1,\tilde{j}_2})\le -C(\log p)^{1/2}\max_{1\le j\le r}w_{j,j}^{1/2}$ for some $C>0$, which implies

$$
\mathbb{P}_{H_1}\left\{n^{1/2}(\widehat{\omega}_{\tilde{j}_1,\tilde{j}_2}-c_{\tilde{j}_1,\tilde{j}_2})\le\widehat{q}_{\mathcal{S},1-\alpha,1},\ \mathscr{T}_\varepsilon\right\}\le\mathbb{P}_{H_1}\left\{n^{1/2}(\widehat{\omega}_{\tilde{j}_1,\tilde{j}_2}-\omega_{\tilde{j}_1,\tilde{j}_2})\le -C(\log p)^{1/2}\max_{1\le j\le r}w_{j,j}^{1/2}\right\}\to 0.
$$

From Lemma 4, we know that $\max_{1\le j\le r}|\widehat{w}_{j,j}-w_{j,j}|=o_p(1)$ which also implies that $\max_{1\le j\le r}|\widehat{w}_{j,j}^{1/2}-w_{j,j}^{1/2}|/w_{j,j}^{1/2}=o_p(1)$. Then we can choose suitable $\varepsilon\to 0$ such that $\mathbb{P}(\mathscr{T}_\varepsilon^c)\to 0$. Hence, we complete part (ii). □

Proof of Corollary 2. Our proof includes two steps: (i) to show $\mathbb{P}(\widehat{\mathcal{M}}_{n,\alpha}\subset\mathcal{M}_0)\to 1$, and (ii) to show $\mathbb{P}(\mathcal{M}_0\subset\widehat{\mathcal{M}}_{n,\alpha})\to 1$. Result (i) is equivalent to $\mathbb{P}(\mathcal{M}_0^c\subset\widehat{\mathcal{M}}_{n,\alpha}^c)\to 1$. The latter one is equivalent to $\mathbb{P}\{\max_{(j_1,j_2)\in\mathcal{M}_0^c}n^{1/2}|\widehat{\omega}_{j_1,j_2}|\ge\widehat{q}_{\mathcal{S},1-\alpha,1}\}\to 0$. Notice that $\mathcal{S}=\{1,\ldots,p\}^2$, it holds that

$$\mathbb{P}\bigg\{\max_{(j_1,j_2)\in\mathcal{M}_0^c}n^{1/2}|\widehat{\omega}_{j_1,j_2}|\ge\widehat{q}_{\mathcal{S},1-\alpha,1}\bigg\}\le\mathbb{P}\bigg\{\max_{(j_1,j_2)\in\mathcal{S}}n^{1/2}|\widehat{\omega}_{j_1,j_2}-\omega_{j_1,j_2}|\ge\widehat{q}_{\mathcal{S},1-\alpha,1}\bigg\}\le\alpha+o(1),$$

which implies $\mathbb{P}\{\max_{(j_1,j_2)\in\mathcal{M}_0^c}n^{1/2}|\widehat{\omega}_{j_1,j_2}|\ge\widehat{q}_{\mathcal{S},1-\alpha,1}\}\to 0$. Then we construct result (i). Result (ii) is equivalent to $\mathbb{P}\{\min_{(j_1,j_2)\in\mathcal{M}_0}n^{1/2}|\widehat{\omega}_{j_1,j_2}|\le\widehat{q}_{\mathcal{S},1-\alpha,1}\}\to 0$. Let $(\tilde{j}_1,\tilde{j}_2)=\arg\min_{(j_1,j_2)\in\mathcal{M}_0}|\omega_{j_1,j_2}|$. Without lose of generality, we assume $\omega_{\tilde{j}_1,\tilde{j}_2}>0$. Notice that

$$\mathbb{P}\bigg\{\min_{(j_1,j_2)\in\mathcal{M}_0}n^{1/2}|\widehat{\omega}_{j_1,j_2}|\le\widehat{q}_{\mathcal{S},1-\alpha,1}\bigg\}\le\mathbb{P}\big\{n^{1/2}(\widehat{\omega}_{\tilde{j}_1,\tilde{j}_2}-\omega_{\tilde{j}_1,\tilde{j}_2})\le\widehat{q}_{\mathcal{S},1-\alpha,1}-n^{1/2}\omega_{\tilde{j}_1,\tilde{j}_2}\big\},$$

we can construct result (ii) following the arguments for the proof of Corollary 1. □

References

Aït-Sahalia, Y., Xiu, D., 2015. Principal component analysis of high frequency data. National Bureau of Economic Research. Working paper, No. w21584.
Andrews, D.W.K., 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica 59, 817–858.
Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B 57, 289–300.
Bickel, P., Levina, E., 2008a. Regularized estimation of large covariance matrices. Ann. Statist. 36, 199–227.
Bickel, P., Levina, E., 2008b. Covariance regularization by thresholding. Ann. Statist. 36, 2577–2604.
Bühlmann, P., van de Geer, S., 2011. Statistics for High-Dimensional Data: Methods, Theory and Applications. Springer, Heidelberg.
Cai, T.T., Liu, W., Luo, X., 2011. A constrained l_1 minimization approach to sparse precision matrix estimation. J. Amer. Statist. Assoc. 106, 594–607.
Cai, T.T., Liu, W., Xia, Y., 2013. Two-sample covariance matrix testing and support recovery in high-dimensional and sparse settings. J. Amer. Statist. Assoc. 108, 265–277.
Candes, E., Tao, T., 2007. The Dantzig selector: Statistical estimation when p is much larger than n. Ann. Statist. 35, 2313–2351.
Carrasco, M., Chen, X., 2002. Mixing and moment properties of various GARCH and stochastic volatility models. Econometric Theory 18, 17–39.
Chang, J., Chen, S.X., Chen, X., 2015. High dimensional generalized empirical likelihood for moment restrictions with dependent data. J. Econometrics 185, 283–304.
Chang, J., Guo, B., Yao, Q., 2018. Principal component analysis for second-order stationary vector time series. Ann. Statist. http://dx.doi.org/10.1214/17-AOS1613. (in press).
Chang, J., Tang, C.Y., Wu, Y., 2013. Marginal empirical likelihood and sure independence feature screening. Ann. Statist. 41, 2123–2148.
Chang, J., Zheng, C., Zhou, W.-X., Zhou, W., 2017a. Simulation-based hypothesis testing of high dimensional means under covariance heterogeneity. Biometrics 73, 1300–1310.
Chang, J., Zhou, W., Zhou, W.-X., Wang, L., 2017b. Comparing large covariance matrices under weak conditions on the dependence structure and its application to gene clustering. Biometrics 73, 31–41.
Chen, L., Zheng, S., 2009. Studying alternative splicing regulatory networks through partial correlation analysis. Genome Biol. R3.
Chen, X., Xu, M., Wu, W.B., 2013. Covariance and precision matrix estimation for high-dimensional time series. Ann. Statist. 41, 2994–3021.
Chernozhukov, V., Chetverikov, D., Kato, K., 2013. Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. Ann. Statist. 41, 2786–2819.
Chernozhukov, V., Chetverikov, D., Kato, K., 2014. Testing many moment inequalities. arXiv:1312.7614.
Chernozhukov, V., Chetverikov, D., Kato, K., 2015. Comparison an anti-concentration bounds for maxima of Gaussian random vectors. Probab. Theory Related Fields 162, 47–70.
Davydov, Y.A., 1968. Convergence of distributions generated by stationary stochastic processes. Theory Probab. Appl. 13, 691–696.
Fan, J., Yao, Q., 2003. Nonlinear Time Series. Springer, New York.
Friedman, J., Hastie, T., Tibshirani, R., 2008. Sparse inverse covariance estimation with the graphical lasso. Biostatistics 9, 432–441.
de la Fuente, A., Bing, N., Hoeschele, I., Mendes, P., 2004. Discovery of meaningful associations in genomic data using partial correlation coefficients. Bioinformatics 20, 3565–3574.
Geer, S.van.de., Bühlmann, P., Ritov, Y., Dezeure, R., 2014. On asymptotically optimal confidence regions and tests for high-dimensional models. Ann. Statist. 42, 1166–1202.
den Haan, W.J., Levin, A., 1997. A practitioner's guide to robust covariance matrix estimation. In: Handbook of Statistics, vol. 15, pp. 291–341 (Chapter 12).
Hochberg, Y., Tamhane, A.C., 2009. Multiple Comparison Procedures. Wiley, New York.
Huang, D., Yao, Q., Zhang, R., 2017. Krigings over space and time based on latent low-dimensional structures. A preprint.
Huang, S., Li, J., Sun, L., Liu, J., Wu, T., Chen, K., Reiman, E., 2010. Learning brain connectivity of Alzheimer's disease by sparse inverse covariance estimation. Neuroimage 50, 935–949.
Kenett, D.Y., Tumminello, M., Madi, A., Gur-Gershgoren, G., Mantegna, R.N., Ben-Jacob, E., 2010. Dominating clasp of the financial sector revealed by partial correlation analysis of the stock market. PloS One 5, e15032.
Kiefer, N.M., Vogelsang, T.J., Bunzel, H., 2000. Simple roubust testing of regression hypothesis. Econometrica 68, 695–714.
Lahiri, S.N., 2003. Resampling Methods for Dependent Data. Springer, Berlin.
Liu, W., 2013. Gaussian graphical model estimation with false discovery rate control. Ann. Statist. 41, 2948–2978.
Meinshausen, N., Bühlmann, P., 2006. High-dimensional graphs and variable selection with the Lasso. Ann. Statist. 34, 1436–1462.
Merlevède, F., Peligrad, M., Rio, E., 2011. A Bernstein type inequality and moderate deviations for weakly dependent sequences. Probab. Theory Related Fields 151, 435–474.
Nordman, D.J., Lahiri, S.N., 2005. Validity of sampling window method for linear long-range dependent processes. Econometric Theory 21, 1087–1111.
Peng, J., Wang, P., Zhou, N., Zhu, J., 2009. Partial correlation estimation by joint sparse regression models. J. Amer. Statist. Assoc. 104, 735–746.

Qiu, Y., Chen, S.X., 2012. Test for bandedness of high-dimensional covariance matrices and bandwidth estimation. Ann. Statist. 40, 1285–1314.
Reavis, C., 2012. The global financial crisis of 2008: The role of greed, fear, and oligarchs. MIT Sloan Manag. Rev. 16, 1–22.
Ren, Z., Sun, T., Zhang, C.H., Zhou, H., 2015. Asymptotic normality and optimalities in estimation of large Gaussian graphical models. Ann. Statist. 43, 991–1026.
Reverter, A., Chan, E.K.F., 2008. Combining partial correlation and an information theory approach to the reversed engineering of gene co-expression networks. Bioinformatics 24, 2491–2497.
Schweitzer, F., Fagiolo, G., Sornette, D., Vega-Redondo, F., Vespignani, A., White, D.R., 2009. Economic networks: The new challenges. Science 325, 422–425.
Shapira, Y., Kenett, D.Y., Ben-Jacob, E., 2009. The index cohesive effect on stock market correlations. Eur. Phys. J. B 72, 657–669.
Sun, T., Zhang, C.H., 2012. Scaled sparse linear regression. Biometrika 99, 879–898.
Sun, T., Zhang, C.H., 2013. Sparse matrix inversion with scaled Lasso. J. Machine Learning Res. 14, 3385–3418.
Tibshirani, R., 1996. Regression shrinkage and selection via the Lasso. J. R. Stat. Soc. Ser. B 58, 267–288.
Xue, L., Zou, H., 2012. Regularized rank-based estimation of high-dimensional nonparanormal graphical models. Ann. Statist. 40, 2541–2571.
Yuan, M., Lin, Y., 2007. Model selection and estimation in the Gaussian graphical model. Biometrika 94, 19–35.

Please cite this article in press as: Chang J., et al., Confidence regions for entries of a large precision matrix. Journal of Econometrics (2018), https://doi.org/10.1016/j.jeconom.2018.03.020.