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

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

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.