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

Testing super-diagonal structure in high dimensional covariance matrices[☆]

Jing He[a,c], Song Xi Chen[a,b,*]

[a] Department of Business Statistics and Econometrics, Guanghua School of Management, Peking University, Beijing 100871, China
[b] The Center for Statistical Science, Peking University, Beijing 100871, China
[c] School of Statistics, Southwestern University of Finance and Economics, Chengdu, 611130, China

ARTICLE INFO

Article history:
Available online xxxx

JEL classification:
C12
C14
G10

Keywords:
Bandable covariance
High dimensionality
Sparse covariance matrix
Multiple testing

ABSTRACT

The covariance matrices are essential quantities in econometric and statistical applications including portfolio allocation, asset pricing and factor analysis. Testing the entire covariance under high dimensionality endures large variability and causes a dilution of the signal-to-noise ratio and hence a reduction in the power. We consider a more powerful test procedure that focuses on testing along the super-diagonals of the high dimensional covariance matrix, which can infer more accurately on the structure of the covariance. We show that the test is powerful in detecting sparse signals and parametric structures in the covariance. The properties of the test are demonstrated by theoretical analyses, simulation and empirical studies.

1. Introduction

The covariance matrix of a random vector or a multivariate estimating function is a basic ingredient in multivariate analysis and econometrics in gaining information on the dependence between the components of the random vectors and the estimating functions. The celebrated Markowitz theory for optimal portfolio selection (Markowitz, 1952) is based on consistent estimation of the covariance matrix whose dimension is the number of assets of the portfolio. The sample variance is actively employed in an array of multivariate procedures such as the principal component analysis (PCA), the discrimination analysis and the factor analysis. In econometrics, the generalized method of moment (GMM) requires inversion of the covariance matrix of the multivariate moments as the weighting matrix. When the dimension of the data vector or the moments is fixed, the sample/empirical covariance is known to be consistent to the underlying covariance matrix.

Data with dimensions comparable to or larger than the sample size are increasingly encountered in econometric and statistical analyses. They include analyses of large panels of financial portfolios, on-line prices of consumer goods, macro-economic data that measure a large number of features of an economy; see Stock and Watson (2005), Bai and Ng (2002), Lam et al. (2011), Bai and Li (2012), Lam and Yao (2012) and Chang et al. (2015) for over-views and specific results. Fan et al. (2008) considered a covariance matrix estimator for a multi-factor model where the number of factors is allowed to grow with dimension p when p tends to infinity as the sample size n increases.

Extensive research in obtaining consistent estimators of high dimensional covariance matrix has been made. Bickel and Levina (2008a,b) proposed, respectively, the banding and the thresholding estimator of the covariance matrix by either banding or thresholding the sample covariance matrix. Wu and Pourahmadi (2003) and Rothman et al. (2010) studied methods based on the Cholesky decomposition. Cai et al. (2010) proposed a tapering estimator. The banding and tapering estimators are operational when the underlying covariance matrix $\Sigma = (\sigma_{i,j})_{p\times p}$ belongs to the so-called bandable class, which prescribes that $\sigma_{i,j}$ diminishes to zero at certain rates as either j or i increases. There is a set of high dimensionality tests on the covariance Σ. Testing for the identity or sphericity hypotheses of Σ has been considered in Ledoit and

☆ We would like to thank the editor of the Special Issue, Professor Rong Chen, and two referees for helpful comments and suggestions which have substantially improved the presentation of the paper. The research was partially supported by Natural Science Foundation of China grants 11131002, 71371016 and 71532001, National Key Basic Research Program of China Grant 2015CB856000, Center for Statistical Science and LMEQF at Peking University.

* Corresponding author at: Department of Business Statistics and Econometrics, Guanghua School of Management, Peking University, Beijing 100871, China.
E-mail addresses: hejing1790@pku.edu.cn (J. He), csx@gsm.pku.edu.cn (S.X. Chen).

http://dx.doi.org/10.1016/j.jeconom.2016.05.007
0304-4076/

Wolf (2002) and Chen et al. (2010). Cai and Jiang (2011) and Qiu and Chen (2012) proposed tests for the bandedness of a covariance matrix. See also Schott (2005) and Srivastava (2005) for other formulations.

We propose a test regarding the super-diagonals of Σ, which has much smaller scale than the existing tests, and targets on global features of Σ, for instance the bandedness or specific parametric structure. The smaller scale of the super-diagonal as compared to the entire Σ does pose theoretical challenges when establishing the asymptotic properties of the test statistic. This is because the variation of the test statistic is much smaller, which requires finer derivations in the asymptotic analysis. The benefits of working with a test statistic being a smaller magnitude is a reduced variance and an increased signal-to-noise ratio of the test, which can produce more power than those targeting on the entire covariance matrix Σ. Tests for overall structures of Σ can be made by multiple testing on the super-diagonals in conjunction with the false discovery rate or the Bonferroni procedure. We demonstrate in the paper that the proposed test is useful to the inference of spatial econometrical and statistical models on covariance, which tends to be written in terms of the super-diagonals (Kapoor et al., 2007; Baltagi et al., 2003; Lee and Yu, 2010; Rodríguez and Bárdossy, 2014).

The paper is organized as follows. We outline the framework of the testing problem, including the hypotheses, assumptions and the proposed test statistics in Section 2. Section 3 provides the theoretical properties of the test statistics and the multiple testing procedure. In Sections 4 and 5, we discuss tests for bandedness and parameter structures of Σ, respectively. Results of simulation studies are provided in Section 6. An empirical analysis is reported in Section 7. All technical details are given in Appendix.

2. Preliminaries

Consider a p-dimensional generic random vector $X = (X_1, X_2, \ldots, X_p)^T$, which has mean $\mu = (\mu_1, \mu_2, \ldots, \mu_p)^T$ and covariance matrix $\Sigma = (\sigma_{i,j})_{p\times p}$. The observed data $X_i = (X_{i,1}, \ldots, X_{i,p})^T$, $i = 1, \ldots, n$, are independent copies of X. For $q = 0, 1, \ldots, p-1$, let $D_q = \sum_{l=1}^{p-q} \sigma_{l,l+q}^2$ be the sum of the $\sigma_{i,j}^2$ along the qth super-diagonal, where D_0 represents that on the main diagonal.

We consider testing covariance structures with respect to the super- or sub-diagonals of Σ via D_q. We have two specific covariance structures in mind. One is the nonparametric banded structure in that $\sigma_{i,j} = 0$ for any $|i-j| > k$ for an integer k. The smallest such k is called the bandwidth of Σ. And the other is an isomorphic parametric structure where $\sigma_{i,j} = \sigma(|i-j|; \boldsymbol{\theta})$ for a finite dimensional parameter $\boldsymbol{\theta}$, which is a popular form in spatial econometrics.

The banding structure can be produced by a moving average structure such that, for $i = 1, \ldots, n$,

$$X_{i,l} = \mu_l + \sum_{j=0}^{k} \gamma_j Z_{i,l-j},$$

where for each given i, μ_l is the mean of $X_{i,l}$ and $\{Z_{i,1}, Z_{i,2}, \ldots\}$ is a sequence of independent white noise with zero mean and unit variance, $Z_{i,j} = 0$ for $j \le 0$, and $\gamma_0 = 1$. The integer k is the bandwidth of Σ.

More generally, we consider testing certain parametric model regarding the super-diagonal structure of Σ:

$$H_{0,q}: D_q = D_q(\boldsymbol{\theta}) \quad \text{vs } H_{1,q}: D_q \ne D_q(\boldsymbol{\theta})$$

where $D_q(\boldsymbol{\theta})$ is a parametric form, for $q = 1, 2, \ldots, p-1$, and $\boldsymbol{\theta}$ is a finite dimensional parameter. For bandedness test, $D_q(\boldsymbol{\theta}) \equiv 0$ for $q > k$. A motivation for such model comes from the spatial econometrics or statistics where the X_i consists of recordings at p locations. If $\{X_{i,j}\}_{j=1}^p$ is weakly stationary, $\sigma_{j,j+h} = \mathrm{Cov}(X_{i,j}, X_{i,j+h}) = C(h)$ defines a covariance function $C(\cdot)$. Let $\boldsymbol{\theta} = (\sigma^2, \phi)^T$ and commonly used spatial models for $C(\cdot)$ include the spherical model

$$C(h; \boldsymbol{\theta}) = \sigma^2 \left(1 - 1.5(h/\phi) + 0.5(h/\phi)^3\right), \phi > 0, h < \phi;$$

the wave model

$$C(h; \boldsymbol{\theta}) = \sigma^2 \phi \sin(h/\phi)/h;$$

the exponential model

$$C(h; \boldsymbol{\theta}) = \sigma^2 \exp\left(-h/\phi\right) \quad \text{and}$$

the Gaussian model

$$C(h; \boldsymbol{\theta}) = \sigma^2 \exp\left(-h^2/\phi\right).$$

See Cressie (1993), Kapoor et al. (2007), Baltagi et al. (2003), Lee and Yu (2010) and Rodríguez and Bárdossy (2014) for more details.

The proposed test statistics for super-diagonals are based on an unbiased estimator of D_q:

$$\hat{D}_q = \sum_{l=1}^{p-q} \Bigg\{ \frac{1}{P_n^2} \sum_{i,j}^{*} (X_{i,l}X_{i,l+q})(X_{j,l}X_{j,l+q}) - \frac{2}{P_n^3} \sum_{i,j,k}^{*} X_{i,l}X_{k,l+q}(X_{j,l}X_{j,l+q}) + \frac{1}{P_n^4} \sum_{i,j,k,m}^{*} X_{i,l}X_{j,l+q}X_{k,l}X_{m,l+q} \Bigg\},$$

where $\sum^*$ denotes summation over mutually different subscripts, and $P_n^b = n!/(n-b)!$. It is clear that $\hat{D}_q$ is a linear combinations of U-statistics. Without loss of generality, we assume $\mu = 0$ since $\hat{D}_q$ is invariant to the location shift.

To quantify the dependence among components of the data vector, we invoke the notion of α-mixing. The α-mixing coefficient of the generic $X = (X_1, \ldots, X_p)^T$ is defined as

$$\alpha_X(k) = \sup_{m\in\mathbb{Z}} \alpha(\mathcal{G}_1^m, \mathcal{G}_{m+k}^p), \tag{2.1}$$

where $\alpha(\mathcal{G}_1^m, \mathcal{G}_{m+k}^p) = \sup\{|P(A\cap B) - P(A)P(B)| : A \in \mathcal{G}_1^m, B \in \mathcal{G}_{m+k}^p\}$, $\mathcal{G}_1^m$ and $\mathcal{G}_{m+k}^p$ are the σ-fields generated by $\{X_1, \ldots, X_m\}$ and $\{X_{m+k}, \ldots, X_p\}$, respectively. If $\lim_{k\to\infty} \alpha_X(k) = 0$, the sequence of components in X is said to be α-mixing. Furthermore, we denote the eigenvalues of Σ as $\lambda_{max}(\Sigma) = \lambda_1(\Sigma) \ge \lambda_2(\Sigma) \ge \cdots \ge \lambda_p(\Sigma) = \lambda_{min}(\Sigma)$.

Our test procedure does not require any explicit relationship between the sample size n and the dimension p other than that they both diverges to infinity. It allows p to be much larger than n, that is, the "large p, small n" situation. We assume the following conditions in our analysis.

A1 There are positive constants c and $a \in (0, 1)$ such that $\alpha_X(k) \le ca^k$.

A2 The eighth moment of X_ℓ is uniformly bounded, i.e. $\sup_{1\le\ell\le p} \mathrm{E}|X_\ell|^8 \le M$, for a positive constant M. There exists a positive constant ϵ_0, such that $\lambda_{min}(\Sigma) \ge \epsilon_0 > 0$.

A3 Data vectors X_i are generated by $X_i = \Gamma Z_i$ for $i = 1, 2, \cdots, n$, where $\Gamma = \left(\Gamma_{i,j}\right)_{p\times m}$ is a $p \times m$ constant matrix, satisfying $\Gamma\Gamma' = \Sigma$ and $m \ge p$, and $Z_1, Z_2, \cdots, Z_n$ are independently and identically distributed (IID) m-dimensional random vectors such that $\mathrm{E}(Z_i) = 0$ and $\mathrm{Var}(Z_i) = I_m$. Write $Z_i = (Z_{i,1}, \ldots, Z_{i,m})^T$. We assume $Z_{i,j}$ have uniformly bounded 8th moment, and there exists a finite constant Δ such that $\mathrm{E}(Z_{i,j}^4) = 3 + \Delta$ for $j = 1, \ldots, m$, and $\mathrm{E}(Z_{i,j_1}^{\ell_1} Z_{i,j_2}^{\ell_2} \cdots Z_{i,j_q}^{\ell_q}) = \mathrm{E}(Z_{i,j_1}^{\ell_1})\mathrm{E}(Z_{i,j_2}^{\ell_2}) \cdots \mathrm{E}(Z_{i,j_q}^{\ell_q})$ for any integers $\ell_\nu \ge 0$ with $\sum_{\nu=1}^{q} \ell_\nu \le 8$ and distinct subscripts $j_1, \ldots, j_q$.

The α-mixing coefficient in Assumption A1 can be relaxed to be polynomial decay such that $\alpha_X(k) \le ck^{-\beta}$ for positive constants

Please cite this article in press as: He, J., Chen, S.X., Testing super-diagonal structure in high dimensional covariance matrices. Journal of Econometrics (2016), http://dx.doi.org/10.1016/j.jeconom.2016.05.007