Statistics and Probability Letters 98 (2015) 123–130

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

On testing for sphericity with non-normality in a fixed effects panel data model[☆]

Badi H. Baltagi [a,*], Chihwa Kao [a], Bin Peng [b]

[a] Department of Economics and Center for Policy Research, 426 Eggers Hall, Syracuse, NY 13244-1020, USA
[b] Department of Economics, 110 Eggers Hall, Syracuse, NY 13244-1020, USA

ARTICLE INFO

Article history:
Received 4 December 2014
Received in revised form 15 December 2014
Accepted 15 December 2014
Available online 29 December 2014

Keywords:
Sphericity
Panel data
Cross-sectional dependence
John test

ABSTRACT

Building upon the work of Chen et al. (2010), this paper proposes a test for sphericity of the variance–covariance matrix in a fixed effects panel data regression model without the normality assumption on the disturbances.

1. Introduction

This paper proposes testing the null of sphericity of the variance–covariance matrix in a fixed effects panel data model which does not require the normality assumption on the disturbances. This builds on the paper by Chen et al. (2010) who use U-statistics to test for sphericity of the variance–covariance matrix in statistics. The null of sphericity means that the variance–covariance matrix is proportional to the identity matrix. Rejecting the null means having cross-sectional dependence among the individual units of observation or heteroskedasticity or both. In empirical economic studies, individuals are affected by common shocks. For example, investors' decisions may be influenced by the way they interact with each other and also by common macro-economic shocks or public policies. These potentially cause cross-sectional dependence among the units.

In statistics, the $n \times n$ sample covariance matrix S_n is widely used for tests of sphericity since it is a consistent estimator for the variance–covariance matrix Σ_n. One could either use the likelihood ratio test, see Anderson (2003), or test the Frobenius norm of the difference between S_n and Σ_n, see John (1971, 1972). However, with panel data sets where n the number of individuals is larger than the time series dimension of the data T, the sample covariance matrix becomes singular. This causes problems for the likelihood ratio test which is based on the inverse of S_n. Even when n is smaller than T, the sample covariance matrix S_n is ill-conditioned as shown in the Random Matrix Theory (RMT) literature. In fact, the eigenvalues of the sample covariance matrix S_n are no longer consistent for their population counterpart, see Johnstone (2001). Ledoit and Wolf (2004) show that the scaled Frobenius norm of S_n does not converge to that of Σ_n with $n/T \to c \in (0, \infty)$. As a result, John's test, see John (1971, 1972), is no longer applicable. Hence, Ledoit and Wolf (2002) propose a new test for the null of sphericity which could be applied even when n is relatively as large as T. However, these statistical tests for raw data are

☆ We would like to thank the Associate editor and two anonymous referees for their helpful comments and suggestions.
* Corresponding author.
E-mail addresses: bbaltagi@maxwell.syr.edu (B.H. Baltagi), dkao@maxwell.syr.edu (C. Kao), bpeng01@syr.edu (B. Peng).

http://dx.doi.org/10.1016/j.spl.2014.12.017
0167-7152/

not directly applicable to testing sphericity in panel data regressions since the disturbances are unobservable. Baltagi et al. (2011) extend the Ledoit and Wolf (2002)'s John test to the fixed effects panel data model and correct for the bias due to substituting within residuals for the actual disturbances. However, their test relies on the normality assumption and their simulation results show that the test has size distortion under non-normality of the disturbances.

To account for the possible "non-normality" of the disturbances as well as the "large n, small T" issues in testing the null of sphericity, Chen et al. (2010) propose a modified John test by constructing U-statistics of observable samples for estimating $\mathrm{tr}\,\Sigma_n$ and $\mathrm{tr}\,\Sigma_n^2$. Based on their work, this paper proposes a new test for the null of sphericity of the disturbances in a fixed effects regression panel data model. This test does not require the assumption of normality of the disturbances, and can be applied to the case where n is larger than T. The limiting distribution of this test statistic under the null is derived. Also, its finite sample properties are studied using Monte Carlo simulations.

The paper is organized as follows. Section 2 specifies the fixed effects panel data regression model and the assumptions required. Section 3 introduces the test statistic. Section 4 derives the limiting distribution of this test statistic under the null and discusses its power properties. Section 5 reports the results of Monte Carlo simulations, while Section 6 concludes. All the proofs and technical details can be found in an Appendix available upon request from the authors.

Notation. $\|B\| = \left(\mathrm{tr}(B'B)\right)^{1/2}$ is the Frobenius norm of a matrix B or the Euclidean norm of a vector B, and $\mathrm{tr}(B)$ is the trace of B. $\xrightarrow{d}$ denotes convergence in distribution and $\xrightarrow{p}$ denotes convergence in probability. For two matrices $B = (b_{ij})$ and $C = (c_{ij})$, we define $B \circ C = (b_{ij}c_{ij})$.

2. The model and assumptions

Consider the following fixed effects panel data regression model

$$y_{it} = \alpha + x_{it}'\beta + \mu_i + v_{it}, \quad \text{for } i = 1, 2, \ldots, n;\ t = 1, 2, \ldots, T, \tag{2.1}$$

where i indexes the cross-sectional dimension and t indexes the time series dimension. y_{it} is the dependent variable, x_{it} denotes the $k \times 1$ vector of exogenous regressors, and β is the corresponding $k \times 1$ vector of parameters. μ_i denotes the time-invariant individual effects which can be fixed or random and could be correlated with the regressors. Define the vector of disturbances $v_t = (v_{1t}, \ldots, v_{nt})'$ and its corresponding variance–covariance matrix Σ_n. The null hypothesis of interest is sphericity:

$$H_0 : \Sigma_n = \sigma_v^2 I_n \quad \text{vs} \quad H_1 : \Sigma_n \neq \sigma_v^2 I_n. \tag{2.2}$$

The alternative hypothesis allows cross-sectional dependence or heteroskedasticity or both.

For the panel data regression model, v_{it} is unobserved, and the test statistic is based upon consistent estimates of variance–covariance matrix, denoted by S_n or its correlation coefficients matrix counterpart, see Breusch and Pagan (1980). Baltagi et al. (2011) extend the Ledoit and Wolf (2002) test to a fixed effects panel data model with large n and large T. They show that the noise resulting from using within residuals rather than the actual disturbances accumulates and causes bias for the proposed test statistic. However, their simulations show that their test is oversized under non-normality of the disturbances. This paper extends Chen et al. (2010) to test the null of sphericity of the variance–covariance matrix of the disturbances in a fixed effects panel data regression model without assuming normality of the disturbances. We use the within residuals which are given by

$$\hat{v}_{it} = \tilde{y}_{it} - \tilde{x}_{it}'\tilde{\beta} = v_{it} - \bar{v}_{i.} - \tilde{x}_{it}'\left(\tilde{\beta} - \beta\right), \tag{2.3}$$

where $\tilde{x}_{it} = x_{it} - \bar{x}_{i.}$ and $\bar{x}_{i.} = \frac{1}{T}\sum_{t=1}^{T} x_{it}$. Similarly, $\tilde{y}_{it} = y_{it} - \bar{y}_{i.}$, $\bar{y}_{i.} = \frac{1}{T}\sum_{t=1}^{T} y_{it}$, and $\bar{v}_{i.} = \frac{1}{T}\sum_{t=1}^{T} v_{it}$. The within estimator of β is given by $\tilde{\beta} = \left(\sum_{t=1}^{T}\sum_{i=1}^{n}\tilde{x}_{it}\tilde{x}_{it}'\right)^{-1}\left(\sum_{t=1}^{T}\sum_{i=1}^{n}\tilde{x}_{it}\tilde{y}_{it}\right)$. Let $\tilde{y}_t = (\tilde{y}_{1t}, \ldots, \tilde{y}_{nt})'$, $\hat{v}_t = (\hat{v}_{1t}, \ldots, \hat{v}_{nt})'$, $\bar{v}_. = \left(\bar{v}_{1.}, \ldots, \bar{v}_{n.}\right)'$, and $\tilde{x}_t = (\tilde{x}_{1t}, \ldots, \tilde{x}_{nt})$. The within residuals can be rewritten in matrix form as $\hat{v}_t = v_t - \bar{v}_. - \tilde{x}_t'\left(\tilde{\beta} - \beta\right)$. To facilitate our analysis, we require the following assumptions:

Assumption 1. The $n \times 1$ vectors $v_1, v_2, \ldots, v_T$ are independent and identically distributed (*i.i.d.*) with mean vector 0 and covariance matrix $\Sigma_n = \Gamma\Gamma'$, where Γ is an $n \times m$ ($m \leq \infty$) matrix, v_t can be written as $v_t = \Gamma Z_t$, where $Z_t = (z_{t1}, \ldots, z_{tm})$ are *i.i.d.*random vectors with mean vector 0 and covariance matrix I_m. We also assume that each v_{it}, for $i = 1, \ldots, n$ has uniformly bounded 8th moment and there exists a finite constant Δ such that $E(z_{1l}^4) = 3 + \Delta$, for $l = 1, \ldots, m$.

Assumption 2. The regressors x_{it}, $i = 1, \ldots, n$, $t = 1, \ldots, T$ are independent of the idiosyncratic disturbances v_{it}, $i = 1, \ldots, n$, $t = 1, \ldots, T$. The regressors x_{it} have finite fourth moments: $E[\|x_{it}\|^4] \leq K < \infty$, where K is a positive constant.

Assumption 3. As $(n, T) \to \infty$, $\mathrm{tr}(\Sigma_n^2) \to \infty$, $\mathrm{tr}(\Sigma_n^4)/\mathrm{tr}^2(\Sigma_n^2) \to 0$.

The asymptotics follow the framework employed by Chen et al. (2010). Assumption 3 requires $\mathrm{tr}(\Sigma_n^4)$ to grow at a slower rate than $\mathrm{tr}^2(\Sigma_n^2)$. This assumption is flexible. In fact, if all the eigenvalues of Σ_n are bounded away from zero and infinity,