Testing for Sphericity in a Fixed Effects Panel Data Model*

Badi H. Baltagi
Syracuse University

Qu Feng
Nanyang Technological University

Chihwa Kao
Syracuse University

November 8, 2009

Abstract

This paper proposes a test for the null of sphericity in a fixed effects panel data model. It uses the *Random Matrix Theory* based approach of Ledoit and Wolf (2002) to test for the null of sphericity of the error terms in a fixed effects panel model with a large number of cross-sectional units and time series observations. Since the errors are unobservable, the residuals from the fixed effects regression are used. The limiting distribution of the proposed test statistic is derived. Additionally, its finite sample properties are examined using Monte Carlo simulations.

Keywords: Sphericity; Panel Data; Cross-sectional Dependence; John Test.
JEL Classification: C13; C33.

*Email Addresses: Badi H. Baltagi, bbaltagi@maxwell.syr.edu; Qu Feng, qfeng@ntu.edu.sg; Chihwa Kao, cdkao@maxwell.syr.edu. We would like to thank the editor Jaap Abbring and two anonymous referees for their useful comments and suggestions.

1 Introduction

This paper proposes a new test for the null of sphericity of the remainder disturbances in a fixed effects panel data regression model with large n and T based on a random matrix theory (RMT) approach. This is important for applied panel data work given the numerous applications that report fixed effects estimates ignoring cross-section dependence or heteroskedasticity. We offer this test as a diagnostic.[1] One can always report robust HAC type options for the fixed effects estimates perhaps corrected for finite samples, see the simulation results in Long and Ervin (2000) in case of heteroskedasticity and the recommended finite sample correction by MacKinnon and White (1985). The popular method for robustifying the fixed effects estimator to account for heteroskedasticity in panels is based on Arellano (1987). This is programmed by Stata, see also Hansen (2007) and Stock and Watson (2008). For spatial HAC corrections, see Conley (1999), Driscoll and Kraay (1998) and Kelejian and Prucha (2007). More recently, the HAC suggested by Bester, Conley, and Hansen (2009) using dependent data in time series, spatial and panel data, seem to be promising.[2] Of course, if one is willing to put more formal structure on the form of cross-section dependence or heteroskedasticity, one can re-estimate the model using this structure. Conley and Molinari (2007) investigated the impact of location/distance measurement errors upon the accuracy of parametric and nonparametric estimators of asymptotic variances. They also suggested a specification test based on a parametric bootstrap that has good power properties for the types of measurement error they considered. The asymptotic results for our test statistic require large dimensional panels and these are becoming more available, especially in finance and marketing, e.g., scanner data on thousands of customers over a long period of time in marketing research, and stock purchases for thousands of firms over a long period of time in finance. We base our test on the statistical literature which assumes normality and typically focuses on the raw data. In order not to impose any structure on the covariance matrix using the raw data, statisticians have based their tests for the null of sphericity on the $n \times n$ sample covariance matrix, denoted by S. However, with the availability of more data, the dimension of the sample covariance matrix increases, and the

[1] Another diagnostic test, in the same spirit, is given by Hsiao, Pesaran and Pick (2009) who test for cross-sectional independence in nonlinear panel data models.

[2] One important warning however. Unlike heteroskedasticity, certain types of cross-sectional dependence in the data may result in the inconsistency of OLS, see Andrews (2005) for a factor model example, and Lee (2002) for a spatial autoregressive model example. See also, the more recent paper by Bai (2009) who considers a panel data model where the cross-section dependence is generated by interactive fixed effects. For this model, the usual fixed effects estimator would be inconsistent, and hence a HAC based on standard fixed effects residuals may lead to missleading inference. The point is that in these cases, a HAC estimate may not be sufficient to correct the effect of cross-sectional dependence.

researcher is soon faced with the "curse of dimensionality". Also, when n exceeds T, the sample covariance matrix S becomes singular. Even when n/T is smaller than 1, but n is still large, the $n \times n$ sample covariance matrix S will be ill-conditioned. These, in turn, cast doubt on any test involving S including the likelihood ratio test (Ledoit and Wolf, 2004). More importantly, the RMT literature shows that the sample covariance matrix is not a consistent estimator of the population covariance matrix when n is large and comparable with T. Since S is a random matrix, not a random variable or a random vector, a different concept of consistency is applied to account for the change of dimensionality. The spectral norm and Frobenius norm of a matrix are often used in this literature. In fact, Geman (1980) shows that when the sample is from an iid normal distribution with zero mean and an identity variance-covariance matrix, the spectral norm of S does not converge to that of the identity matrix.[3] In addition, the largest eigenvalue of the sample covariance matrix follows a Tracy-Widom distribution asymptotically, e.g., Johnstone (2001). These results are very different from the textbook results in multivariate statistics with fixed n and large T. In the latter case, the eigenvalues of S are consistent estimators of the population eigenvalues (Theorem 13.5.1 in Anderson (2003)). Additionally, Ledoit and Wolf (2004) show that the scaled Frobenius norm of the sample covariance matrix does not converge to that of the population covariance matrix. The intuition behind these results is straightforward. In large n and T panels, the noise contained in each element of S is the same as in the case of fixed n, and its magnitude is $1/T$. However, in a setup with comparably large n and T, the noise involving all the elements of S accumulates with increasing n. This can not be smoothed away by a large T as in the case with fixed n.

To account for the "curse of dimensionality" in testing for the null of sphericity with large panels, a RMT approach is introduced. The RMT literature provides useful asymptotic results in this setting with comparably large n and T. Based on the work of John (1971) and Ledoit and Wolf (2002), this paper proposes a test for the null of sphericity of the disturbances of a fixed effects panel data model with comparably large n and T. We call it the John test. Its limiting distribution under the null is derived and its finite sample properties are studied using Monte Carlo experiments.

The organization of the paper is as follows. The next section briefly discusses the fixed effects panel regression model and the hypothesis to be tested. Section 3 studies the proposed John test,

[3] If $\frac{n}{T}$ converges to a non-zero constant c as $(n, T) \to \infty$ then the eigenvalues of the sample covariance matrix will spread from $(1-\sqrt{c})^2$ to $(1+\sqrt{c})^2$, while all the eigenvalues of the population covariance matrix (identity matrix) are 1. See Bai (1999) for a survey. This paper follows the concept of asymptotics in Ledoit and Wolf (2002): n and T go to infinite jointly with comparable convergence rate. Please see Assumption 3 below.

while Section 4 derives its limiting distribution. Section 5 discusses the finite sample bias of this test. Section 6 compares the size and power of the proposed test as well as the traditional tests for cross-sectional dependence using Monte Carlo experiments. We include these tests because under homoskedasticity, testing the null of sphericity is equivalent to testing the null of no cross-section dependence. Section 7 concludes. The appendix contains all the proofs and the technical material.

Notation: The Frobenius norm of a matrix A is denoted as $\|A\|_F = (tr(A'A))^{1/2}$, where $tr(A)$ denotes the trace of A. $\xrightarrow{d}$ denotes convergence in distribution and $\xrightarrow{p}$ denotes convergence in probability.

2 The Model and Assumptions

Consider a fixed effects panel regression model:

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

where i indexes the cross-sectional units and t the time series observations. y_{it} is the dependent variable and x_{it} denotes the exogenous regressors of dimension $k \times 1$, β denotes the corresponding slope parameters. μ_i denotes the time-invariant individual effects. The individual effects are allowed to be fixed or random and they could be correlated with the regressors. For a detailed discussion, see Baltagi (2008). The idiosyncratic error v_{it} is assumed to be not serially correlated over time and independent of the x_{it}, but otherwise is allowed to have a general variance-covariance matrix across the individual units. Let $v_t = (v_{1t}, \cdots, v_{nT})'$.

Assumption 1 *The $n \times 1$ vectors $v_1, v_{2,} \cdots, v_T$ are assumed to be iid $N(0, \Sigma_n)$.*

Here, the $n \times n$ population variance-covariance matrix Σ_n allows for heteroskedasticity as well as a general form of cross-sectional dependence structure which is assumed to be stable over time. The null hypothesis of sphericity is given by:

$$H_0 : \Sigma_n = \sigma_v^2 I_n. \tag{2}$$

The alternative hypothesis $H_a : \Sigma_n \neq \sigma_v^2 I_n$ implies cross-sectional dependence or heteroskedasticity or both.

For panel data with fixed n and large T, classical multivariate statistics shows that Σ_n can be consistently estimated by the $n \times n$ sample covariance matrix S. It follows that S reveals information on cross-sectional dependence. Hence, tests can be constructed based on S or its sample correlation coefficients matrix counterpart, see Ng (2006) and Breusch and Pagan (1980).

In the statistics literature, (2) is usually tested using the raw data and not in a regression context, see John (1971). Ledoit and Wolf (2002) extend John's (1971) work and study a test for the null of sphericity of large-dimensional covariance matrices. They show that with large n and T, the test statistic proposed by John (1971) still works but follows a different limiting distribution. This paper extends their work to test for the null of sphericity of the disturbances of a fixed effects panel data regression model. [4]

Since v_{it} is unobservable, the test statistic is based on consistent estimates of the regression residuals. We use the within estimator for the slope parameter β:

$$\tilde{\beta} = \left[\sum_{i=1}^{n}\sum_{t=1}^{T}\widetilde{x}_{it}\widetilde{x}_{it}'\right]^{-1}\left[\sum_{i=1}^{n}\sum_{t=1}^{T}\widetilde{x}_{it}\widetilde{y}_{it}\right] \tag{3}$$

where $\widetilde{x}_{it} = x_{it} - \overline{x}_{i\cdot}$, and $\overline{x}_{i\cdot} = \frac{1}{T}\sum_{t=1}^{T} x_{it}$. The variable $\widetilde{y}_{it}$ is defined similarly. It is well established that under H_0, $\tilde{\beta}$ is consistent. The within estimator wipes out the time-invariant variables whether observed or not. In this sense, this estimator is robust to the omission of time-invariant variables from (1) that are unobserved. Simultaneously, this estimator guards against possible endogeneity between time-invariant regressors and the error term. The residuals $\widehat{v}_{it}$ can be obtained as follows:

$$\widehat{v}_{it} = \widetilde{y}_{it} - \widetilde{x}_{it}'\tilde{\beta}. \tag{4}$$

Having $\widehat{v}_{it}$, the residual-based sample covariance matrix can be obtained as $\hat{S} = \frac{1}{T}\sum_{t=1}^{T}\hat{v}_t\hat{v}_t'$ where $\hat{v}_t = (\hat{v}_{1t}, \cdots, \hat{v}_{nt})'$ for $t = 1, \cdots, T$. Unlike the multivariate analysis and RMT literature setting, this paper considers a panel fixed effects regression model. Consequently, the effect of replacing the error v_t with the residual $\hat{v}_t$ on the asymptotics is examined. The other assumptions needed to derive the asymptotics are given below.

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

The normality assumption may be strict but it is standard in this literature. Assumption 2 is required for the consistency of the fixed effects estimator.

[4] Kapetanios (2004) relaxes the assumption of homoskedasticity in Ledoit and Wolf (2002). Unlike the raw data setup in Kapetanios (2004), this paper considers a fixed effects panel data model, and applies the sphericity test on fixed effects residuals.

Assumption 3 $\frac{n}{T} \to c \in [0, \infty)$ *as* $(n, T) \to \infty$.

We consider an asymptotic framework employed by Ledoit and Wolf (2004). Unlike the standard asymptotics, where only T increases, or n increases, the framework $(n, T) \to \infty$ considered here regards n as a sequence indexed by T, denoted as n_T. As T goes to infinity, n_T/T approaches a constant c. For simplicity, we suppress the subscript T of n in the rest of the paper.

3 John Test

We propose the following test statistic for testing the null of sphericity described in (2) based on the sample covariance matrix of the fixed effects residuals:

$$J = \frac{T(\frac{1}{n}tr\hat{S})^{-2}\frac{1}{n}tr(\hat{S}^2) - T - n}{2} - \frac{1}{2} - \frac{n}{2(T-1)} \tag{5}$$

where $\hat{S} = \frac{1}{T}\sum_{t=1}^{T} \hat{v}_t\hat{v}_t'$ is the $n \times n$ sample covariance matrix computed using the within residuals $\hat{v}_t$. Since this test is based upon the raw data test proposed by John (1971), it is referred to as the John test in this paper and (5) is thus called the John statistic. The limiting distribution of the John statistic (5) is a standard normal under the null, as shown in the next section .

The discussion begins with the simple raw data case, and then extends it to the fixed effects model (1). Under the normality assumption, the $n \times n$ sample covariance matrix $S = \frac{1}{T}\sum_{t=1}^{T} v_t v_t'$ follows a Wishart distribution with degrees of freedom T. In the traditional case with fixed dimension n and $T \to \infty$, the most commonly used test statistic to test the null (2) is the likelihood ratio test. However, when $n > T$, S is singular and the likelihood ratio test is not feasible. John (1971) proposes a test for the null hypothesis of sphericity described in (2) in the case of fixed n and large T. In this context, this test statistic is given by:

$$U = \frac{1}{n}tr\left[(\frac{1}{n}trS)^{-1}S - I_n\right]^2 = (\frac{1}{n}trS)^{-2}[\frac{1}{n}tr(S^2)] - 1 \tag{6}$$

where S is the $n \times n$ sample covariance matrix and I_n is the $n \times n$ identity matrix. Note that $\frac{1}{n}trS$ is the average of the eigenvalues of S, and as shown below, $\frac{1}{n}trS$ converges to σ_v^2 under the null. Thus, $\left(\frac{1}{n}trS\right)^{-1} S$ can be regarded as a sample version of $(\sigma_v^2)^{-1}\Sigma_n$. Since the null hypothesis (2) can be written as $(\sigma_v^2)^{-1}\Sigma_n - I_n = 0$, using the matrix norm notation introduced above, $U = \frac{1}{n}\left\|\left(\frac{1}{n}trS\right)^{-1} S - I_n\right\|_F^2$ can be regarded as the Frobenius norm of a sample version of $(\sigma_v^2)^{-1}\Sigma_n - I_n$, measuring the deviation of the scaled S from the identity matrix. A large value of U indicates a significant deviation of Σ_n from sphericity.

John (1972) shows that under the null, and using the normality assumption,

$$\frac{nT}{2}U \xrightarrow{d} \chi^2_{n(n+1)/2-1} \text{ for fixed } n, \text{ as } T \to \infty. \tag{7}$$

However, as $n \to \infty$, $\frac{nT}{2}U$ diverges. Ledoit and Wolf (2002) derive the limiting distribution of a modified test statistic under the null, as $(n, T) \to \infty$ with $n/T \to c \in (0, \infty)$[5]

$$TU - n \xrightarrow{d} N(1, 4). \tag{8}$$

Define the statistic $J_0 = \frac{TU-n}{2} - \frac{1}{2}$, then[6]

$$J_0 \xrightarrow{d} N(0, 1). \tag{9}$$

A similar test was proposed by Srivastava (2005).[7]

In the fixed effects panel data model, we replace the raw data S by its counterpart $\hat{S} = \frac{1}{T}\sum_{t=1}^{T} \hat{v}_t\hat{v}_t'$, where $\hat{v}_t$ is the within residual in (4). Thus, the residual-based $\hat{U}$ is

$$\hat{U} = (\frac{1}{n}tr\hat{S})^{-2}\frac{1}{n}tr(\hat{S}^2) - 1 \tag{10}$$

and the residual-based statistic of J_0 is defined as:

$$\hat{J}_0 = \frac{T\hat{U} - n}{2} - \frac{1}{2}. \tag{11}$$

[5]Birke and Dette (Theorem 3.7, 2005) point out that (8) holds in cases of $n/T \to 0$ and $n/T \to \infty$. Therefore, the case of $c = 0$ is included in Assumption 3.

[6]Equation (9) can be regarded as a normalized version of (7) by substracting its asymptotic mean and dividing by its asymptotic standard deviation. For large n,

$$\sqrt{\frac{1}{n(n+1)-2}}\left[\frac{nT}{2}U - (\frac{n(n+1)}{2} - 1)\right] \approx J_0.$$

[7]The test statistic proposed by Srivastava (2005) for the null (2) is defined as:

$$W = \frac{T}{2}(\widehat{\gamma} - 1)$$

where $\widehat{\gamma} = [T^2/((T-1)(T+2))]\left(\frac{1}{n}tr(S)\right)^{-2}\frac{1}{n}\left[tr\left(S^2\right) - \frac{1}{T}(tr(S))^2\right]$. Srivastava (2005) showed that under the null

$$W \xrightarrow{d} N(0, 1)$$

for $T = O\left(n^{\delta}\right)$ with $0 < \delta \leq 1$. In fact, this test statistic is asymptotically equivalent to John test (6). We can write W as

$$W = \frac{T^2}{(T-1)(T+2)}\frac{TU-n}{2} - \frac{1}{2}\frac{T(T-2)}{(T-1)(T+2)}.$$

Hence, $W \approx \frac{TU-n}{2} - \frac{1}{2} = J_0$ for large T, implying the equivalence of the statistic W and that in (9).

However, $\hat{J}_0$ can not be used directly to test for the null of sphericity in the fixed effects panel model (1). As shown in the next section, a bias occurs using the residual-based statistic $\hat{J}_0$ by replacing v_t with the within residuals $\hat{v}_t$. Consequently, we propose the John statistic J. Comparing (5) and (11), we see that the John statistic is just a bias-corrected $\hat{J}_0$:

$$J = \hat{J}_0 - \frac{n}{2(T-1)}. \tag{12}$$

4 Asymptotics of the John Test

This section shows that for the fixed effects regression model (1) the John statistic J in (5) follows asymptotically a standard normal distribution under the null.

Equation (12) shows that the John statistic J is a bias-corrected $\hat{J}_0$. J can be written as the sum of three terms

$$J = \hat{J}_0 - \frac{n}{2(T-1)} = J_0 + (\hat{J}_0 - J_0) - \frac{n}{2(T-1)} = J_0 + \frac{T(\hat{U} - U)}{2} - \frac{n}{2(T-1)}. \tag{13}$$

The first term J_0 is asymptotically standard normal. The second term $\hat{J}_0 - J_0$ is just the scaled difference between the residual-based $\hat{U}$ and the true U. We will show that the sum of the second and the third terms vanishes asymptotically. Therefore, the John statistic J has the same limiting distribution as J_0 which is a standard normal. By (6) and (10), the bias term $(\hat{J}_0 - J_0) = T(\hat{U} - U)/2$ can be written as:

$$\frac{T(\hat{U} - U)}{2} = \frac{T}{2}\left[\left(\frac{1}{n}trS\right)^2 \frac{1}{n}tr(\hat{S}^2) - \left(\frac{1}{n}tr\hat{S}\right)^2 \frac{1}{n}tr(S^2)\right]\left(\frac{1}{n}tr\hat{S}\right)^{-2}\left(\frac{1}{n}trS\right)^{-2}. \tag{14}$$

It is clear that this bias term depends on the following differences: $\frac{1}{n}tr\widehat{S} - \frac{1}{n}trS$ and $\frac{1}{n}tr\widehat{S}^2 - \frac{1}{n}tr(S^2)$, which are studied in the next two propositions.

Proposition 1 *Under Assumptions 1 and 2,*

$$\frac{1}{n}tr\widehat{S} - \frac{1}{n}trS = O_p(\frac{1}{T}). \tag{15}$$

Proposition 2 *Under Assumptions 1 and 2,*

$$\frac{1}{n}tr\widehat{S}^2 - \frac{1}{n}tr(S^2) = O_p(\frac{1}{T}) + O_p(\frac{n}{T^2}). \tag{16}$$

The proofs of Propositions 1 and 2 are given in the Appendix. Due to the fact that the trace is equal to the sum of the eigenvalues, Proposition 1 gives the magnitude of the distance between the average of the sample eigenvalues and that of the true eigenvalues. In RMT, the average of the eigenvalues is defined as the first moment of the empirical spectral distribution (ESD) of S (e.g., Bai and Silverstein, 2006, p.9), so Proposition 1 provides the consequence, on the first moment, of replacing the unobservable sample covariance matrix with the residual-based one. Similarly, Proposition 2 shows the consequence of replacing S with $\hat{S}$ on the second moment of ESD.

Under Assumption 3**,** $n/T \to c \in [0, \infty)$, the distance between $\hat{S}$ and S is of order $1/T$. For the case of fixed n, it is easy to show that the distance is of $O_p(1/T)$.

Propositions 1 and 2 show that as $T \to \infty$, the difference of the first and second moments of eigenvalues between the residuals and the unobservable disturbances vanishes. Proposition 1 of Ledoit and Wolf (2002) shows that $\frac{1}{n}trS$ and $\frac{1}{n}tr(S^2)$ converge to σ_v^2 and $(1+c)\,\sigma_v^4$, respectively, as $(n,T) \to \infty$ with $n/T \to c \in (0, \infty)$.[8] Consequently, it is straightforward to obtain the limits of $\frac{1}{n}tr\widehat{S}$ and $\frac{1}{n}tr\widehat{S}^2$.

Corollary 1 *Under Assumptions 1 and 2, as* $(n,T) \to \infty$*,*

$$\frac{1}{n}tr\widehat{S} \xrightarrow{p} \sigma_v^2.$$

The limit of $\frac{1}{n}tr\widehat{S}$ requires no restriction on the relationship between n and T. However, this is not the case for $\frac{1}{n}tr\widehat{S}^2$. Assumption 3 is needed.

Corollary 2 *Under Assumptions 1, 2 and 3, as* $(n,T) \to \infty$*,*

$$\frac{1}{n}tr\widehat{S}^2 - (1 + \frac{n}{T})\sigma_v^4 \xrightarrow{p} 0.$$

The results above show that the scaled trace of $\hat{S}$ and $\hat{S}^2$ are bounded. The limit of the trace of $\hat{S}^2$ is related to the ratio n/T. When n increases with T, the noise accumulates and the result is affected by the ratio of n/T. Corollary 2 follows directly from the proof of Proposition 2.

With the results on the difference of trace between $\widehat{S}$ and S in Propositions 1 and 2, it is straightforward to calculate the probability limit of $\frac{T(\hat{U}-U)}{2}$. This is summarized in the following proposition:

[8] Proposition 1 of Ledoit and Wolf (2002, p. 1083) presents a slightly different form. The proof of these results with higher order terms can be found in Lemma 2 in the Appendix of this paper.

Proposition 3 *Under Assumptions 1, 2 and 3,*

$$\frac{T(\hat{U}-U)}{2}-\frac{n}{2(T-1)}=O_p(\frac{1}{n}). \tag{17}$$

Proposition 3 indicates that the bias term $\hat{J}_0 - J_0 = T(\hat{U}-U)/2$ vanishes as $(n,T)\to\infty$ in the case of no fixed effects. However, in the presence of fixed effects, this bias does not vanish and the bias converges to $c/2$, where $n/T\to c\in[0,\infty)$ as $(n,T)\to\infty$. Hence, for the fixed effects model (1), the residual-based statistic $\hat{J}_0$ is biased. The histogram of $\hat{J}_0$ is illustrated in Figure 1.

Proposition 3 suggests a bias adjustment term, $\frac{n}{2(T-1)}$, for the residual-based statistic $\hat{J}_0$ in a fixed effects panel regression model. It follows that under the null, the John statistic in the fixed effects panel regression model (1) $J = J_0 + (\hat{J}_0 - J_0) - \frac{n}{2(T-1)}$ has the same limiting distribution as J_0 which is standard normal.

Theorem 1 *Under Assumptions 1, 2 and 3, in the fixed effects regression model (1), as* $(n,T)\to\infty$

$$J \xrightarrow{d} N(0,1). \tag{18}$$

5 Finite Sample Bias Adjustment

The bias term $\frac{n}{2(T-1)}$ for $\hat{J}_0 - J_0$ in the fixed effects panel model is derived under the assumption $n/T\to c\in[0,\infty)$ as $(n,T)\to\infty$.[9]

Corollary 3 *Under Assumptions 1, 2 and 3, in the fixed effects model (1)*

$$\hat{J}_0 - J_0 = \frac{T(\hat{U}-U)}{2} = \frac{\frac{n}{T}\sigma_v^8 - \frac{n}{T^2}\sigma_v^8 + O_p(\frac{1}{\sqrt{n}}) + O_p(\frac{\sqrt{n}}{T}) + O_p(\frac{1}{\sqrt{T}})}{\frac{2(T-1)^2}{T^2}\sigma_v^8 + O_p(\frac{1}{\sqrt{nT}})}. \tag{19}$$

Corollary 3 is from (39) in the proof of Proposition 3 in the Appendix. When $\sqrt{n}/T\to 0$, a good approximation for the bias (19) becomes

$$\frac{\frac{n}{T}\sigma_v^8 - \frac{n}{T^2}\sigma_v^8}{\frac{2(T-1)^2}{T^2}\sigma_v^8} = \frac{n}{2(T-1)}. \tag{20}$$

Consequently, we propose the following John statistic for the fixed effects panel regression model:

$$J = J_0 + (\hat{J}_0 - J_0) - \frac{n}{2(T-1)} \tag{21}$$

$$= \frac{T(\frac{1}{n}tr\hat{S})^{-2}\frac{1}{n}tr(\hat{S}^2) - T - n}{2} - \frac{1}{2} - \frac{n}{2(T-1)}. \tag{22}$$

[9] This paper derives the bias term for the standard fixed individual effects model. If time effects are also included, a similar bias-term can be derived. The derivation is tedious, but straightforward.

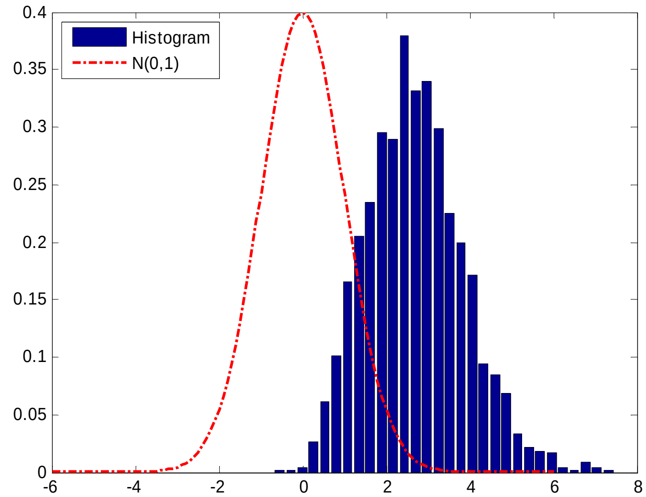

Figure 1: The Histogram of $\widehat{J}_0$ under the null of sphericity in the Fixed Effects Model

Note: The design is described in Section 6 and the replication number is 2000, n = 50, T = 10.

The bias term $\frac{n}{2(T-1)}$ is valid for any n and T satisfying $n/T \to c \in [0, \infty)$. Obviously, $\sqrt{n}/T \to 0$ can be regarded as special case of $n/T \to c \in [0, \infty)$. The finite sample properties of the John test will be explored using Monte Carlo experiments in the next section.

Proposition 3 and Figure 1 show that the bias term $\hat{J}_0 - J_0$ can not be ignored in a fixed effects model when n is large. To test for the null of sphericity of the idiosyncratic errors v_{it}'s, consistent estimates of v_{it}'s are needed. The within residuals in the fixed effects model are consistent as $T \to \infty$, and the convergence rate depends on T. When T is not very large, within residuals are inaccurate. It is this inaccuracy that accumulates in the bias term $\hat{J}_0 - J_0$. Consequently, the bias term $\hat{J}_0 - J_0$ lingers in the fixed effects model.

Specifically, in a fixed effects panel regression model in (1)

$$y_{it} = \alpha + x'_{it}\beta + \mu_i + v_{it}.$$

The distance between the within residuals $\widehat{v}_{it}$ and the idiosyncratic error v_{it} is

$$\widehat{v}_{it} - v_{it} = \bar{v}_{i.} - \tilde{x}'_{it}(\tilde{\beta} - \beta) = O_p(T^{-1/2}) + O_p\left((nT)^{-1/2}\right) = O_p(T^{-1/2}) \tag{23}$$

where $\tilde{\beta}$ is the within estimate, $\bar{v}_{i.} = \frac{1}{T}\sum_{t=1}^{T} v_{it}$ and $\tilde{x}_{it} = x_{it} - \frac{1}{T}\sum_{t=1}^{T} x_{it}$.

In (14), since the denominator is always bounded, the magnitude of this distance is determined by the numerator, which in turn depends on differences involving first and second moments of $\hat{S}$ and S.

$$\widehat{S} - S = \frac{1}{T}\sum_{t=1}^{T} \widehat{v}_t \widehat{v}'_t - \frac{1}{T}\sum_{t=1}^{T} v_t v'_t$$

where the $\widehat{v}_t$'s are within residuals. After some algebra, we get

$$\widehat{S} - S = -\frac{1}{T}\sum_{t=1}^{T} \tilde{x}_t(\tilde{\beta} - \beta)\tilde{v}'_t - \frac{1}{T}\sum_{t=1}^{T} \tilde{v}_t(\tilde{\beta} - \beta)'\tilde{x}'_t + \frac{1}{T}\sum_{t=1}^{T} \tilde{x}_t(\tilde{\beta} - \beta)(\tilde{\beta} - \beta)'\tilde{x}'_t - \bar{v}_{.}\bar{v}'_{.}. \tag{24}$$

Here $\tilde{x}_t$ is the within transformation on x_t, and $\tilde{v}_t$ is similarly defined. (24) has an additional term $-\bar{v}_{.}\bar{v}'_{.}$, which comes from the within transformation. As shown in the Appendix,

$$\frac{1}{n}tr\left[-\frac{1}{T}\sum_{t=1}^{T} \tilde{x}_t(\tilde{\beta} - \beta)\tilde{v}'_t - \frac{1}{T}\sum_{t=1}^{T} \tilde{v}_t(\tilde{\beta} - \beta)'\tilde{x}'_t + \frac{1}{T}\sum_{t=1}^{T} \tilde{x}_t(\tilde{\beta} - \beta)(\tilde{\beta} - \beta)'\tilde{x}'_t\right] = O_p(\frac{1}{nT}). \tag{25}$$

However, the scaled trace of the extra term from the within transformation has a bigger magnitude,

$$\frac{1}{n}tr(-\bar{v}_{.}\bar{v}'_{.}) = -\frac{1}{n}\sum_{i=1}^{n} \bar{v}_{i.}^2 = -\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{T}\sum_{t=1}^{T} v_{it}\right)^2 = O_p(\frac{1}{T}).$$

which dominates $O_p(\frac{1}{nT})$ for comparably large n and T. Therefore,

$$\frac{1}{n}tr(\widehat{S}-S)=O_p(\frac{1}{nT})+O_p(\frac{1}{T})=O_p(\frac{1}{T}).$$

Consequently, the distance of the first and second moments (between $\hat{S}$ and S) is so large that the bias term $\hat{J}_0-J_0$ can not be ignored asymptotically as shown in (17). As a result, the residual-based statistic $\hat{J}_0$ exhibits a shift and is subject to asymptotic bias.

6 Monte Carlo Simulations

In this section, we conduct Monte Carlo experiments to assess the empirical size and power of the John test proposed in this paper. Throughout the experiments, we assume homoskedasticity on the remainder error term. This means that testing the null of sphericity is equivalent to testing the null of no cross-section dependence. Hence, we include in our experiments Pesaran's (2004) CD test and the bias-adjusted LM test derived by Pesaran, Ullah, and Yamagata (2008), denoted as PUY's LM test. The latter tests are based on sample correlation coefficients and test for zero cross-section dependence. The John test is given by (5), while PUY's LM test is defined as:

$$\text{PUY's } LM=\sqrt{\frac{2}{n(n-1)}}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\frac{(T-k)\hat{\rho}_{ij}^2-\mu_{Tij}}{\upsilon_{Tij}}$$

where

$$\mu_{Tij}=\frac{1}{T-k}tr(M_iM_j)$$

and

$$\upsilon_{Tij}^2=[tr(M_iM_j)]^2a_{1T}+2tr[(M_iM_j)^2]a_{2T}$$

with

$$\begin{aligned} a_{1T} &= a_{2T}-\frac{1}{(T-k)^2}, \\ a_{2T} &= 3\left[\frac{(T-k-8)(T-k+2)+24}{(T-k+2)(T-k-2)(T-k-4)}\right]^2. \end{aligned}$$

Note that M_i is the residual maker matrix of the individual regression i, while k is the number of the regressors.[10] Pesaran's CD test statistic is defined as:

$$\text{Pesaran's } CD=\sqrt{\frac{2T}{n(n-1)}}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\hat{\rho}_{ij}$$

[10] $M_i=I-X_i(X_i'X_i)^{-1}X_i'$, where $X_i=(x_{i1},\cdots,x_{iT})'$ contains T observations of the regressors in the individual regression i.

where $\hat{\rho}_{ij}$ is the sample correlation of the residuals $\widehat{v}_{it}$. Specifically,

$$\hat{\rho}_{ij} = \left(\sum_{t=1}^{T} \widehat{v}_{it}^2\right)^{-1/2} \left(\sum_{t=1}^{T} \widehat{v}_{jt}^2\right)^{-1/2} \sum_{t=1}^{T} \widehat{v}_{it}\widehat{v}_{jt}.$$

We do not include the traditional Breusch and Pagan (1980) LM test statistic since it is not applicable when $n \to \infty$.

6.1 Experiment Design

The experiments use the following data generating process:

$$y_{it} = \alpha + \beta x_{it} + \mu_i + v_{it}, \; i = 1, \cdots, n; \; t = 1, \cdots, T, \tag{26}$$

$$x_{it} = \rho x_{i,t-1} + \mu_i + \eta_{it}, \tag{27}$$

where μ_i is the fixed effect. The regressor x_{it} is generated in a similar way to that of Im, Ahn, Schmidt and Wooldridge (1999). Since x_{it} is correlated with the $\mu_i's$, it is endogenous. $\eta_{it} \sim iidN(\phi_\eta, \sigma_\eta^2)$, and v_{it} is the idiosyncratic error.

Under the null, v_{it} is assumed to be $iidN(0, \sigma_v^2)$ across individuals and over time. To calculate the power of the tests considered, two different models of the cross-sectional dependence in the $v_{it}'s$ are used: a factor model and a spatial model. In the former, see Pesaran (2004) and Pesaran and Tosetti (2008) to mention a few, it is assumed that

$$v_{it} = \gamma_i f_t + \varepsilon_{it} \tag{28}$$

where f_t $(t = 1, \cdots, T)$ are the factors and γ_i $(i = 1, \cdots, n)$ are the loadings. In a spatial model, see Anselin and Bera (1998) and Baltagi, Song, and Koh (2003), to mention a few, we consider both a first-order spatial autocorrelation (SAR(1)) and a spatial moving average (SMA(1)) model as follows:

$$v_{it} = \delta(0.5 v_{i-1,t} + 0.5 v_{i+1,t}) + \varepsilon_{it}, \tag{29}$$

$$v_{it} = \delta(0.5 \varepsilon_{i-1,t} + 0.5 \varepsilon_{i+1,t}) + \varepsilon_{it}. \tag{30}$$

For these specifications (28), (29) and (30), ε_{it} is assumed to be $iidN(0, \sigma_\varepsilon^2)$ across individuals and over time. The null can be regarded as a special case of $\gamma_i = 0$ in the factor model (28) and $\delta = 0$ in the spatial model (29) and (30).

The parameters α and β are set arbitrarily to 1 and 2 respectively. The $\mu_i's$ are assumed to be $iidN(\phi_\mu, \sigma_\mu^2)$ for $i = 1, \cdots, n$. We set $\phi_\mu = 0$ and $\sigma_\mu^2 = 0.25$. For the regressor in (27),

$\rho = 0.7$ and $\phi_\eta = 0$ and $\sigma_\eta^2 = 1$. v_{it} (under the null) and ε_{it} (under the alternative) are from $N(0, \sigma_v^2)$ and $N(0, \sigma_\varepsilon^2)$ with $\sigma_v^2 = \sigma_\varepsilon^2 = 0.5$. For the factor model in (28) $\gamma_i \sim iidU(-0.5, 0.55)$ and $f_t \sim iidN(0, 1)$. For the spatial model $\delta = 0.4$ in (29) and (30).

The Monte Carlo experiments are conducted for $n = 5, 10, 20, , 30, 50, 100, 200$ and $T = 10, 20, 30, 50$. For each replication, we compute the John, CD and PUY's LM test statistics. 2,000 replications are performed. To obtain the empirical size, the John test is conducted at the two-sided 5% nominal significance level.

6.2 Results

Table 1 gives the empirical size of the John test under the null of sphericity. Hence, in these experiments we assume that there is no cross-section dependence of the factor model or spatial correlation type. By and large, the size of the John test is close to 5% for comparably large n and T and for small n and large T. This is consistent with the theoretical results derived in Theorem 1 of Section 4. For large n and small T, the John test is slightly oversized. Similarly, the size of PUY's LM test is close to 5% except for large n and small T, while the CD test has the correct size for all combinations of n and T.

Table 2 presents the size adjusted power of these tests under the alternative specification of a factor model. Note that the size adjusted power of the John and PUY's LM tests increase with n and T. However, the CD test lacks power in this case. The power of the CD test is less than 23% even for $n = 200$ and $T = 50$.

Tables 3 and 4 show the size adjusted power of these tests under the alternative specification of SAR(1) and SMA(1), respectively. Note that all the tests have low power for small $T = 10$, but this improves considerably when $T = 20$. In these cases, the size adjusted power of John test is slightly better than that of PUY's LM test for all combinations of n and T. The size adjusted power of the CD test performs much better than in the case of a factor model.[11]

Table 5 checks the sensitivity of our tests to non-normal disturbances. The CD and PUY tests seem robust to non-normality. However, the proposed John test seems to be very sensitive to non-normality. For the uniform distributions U[-2,2] and U[1,2], John's test has similar size as reported for the normal distribution. However, it suffers from big size distortion when the error distribution is Chi-square or a t-distribution. In this sense, the normality assumption (Assumption 1) is very

[11] See Pesaran and Tosetti (2008) who distinguish between factor models and spatial models in terms of time specific weak versus strong cross-sectional dependence.

Table 1: Size of Tests

Size	$T \setminus n$	5	10	20	30	50	100	200
John	10	5.5	7.3	7.5	8.3	8.2	7.3	9.7
	20	5.6	6.8	5.5	5.6	7.1	7.2	6.2
	30	6.0	6.1	5.6	7.0	5.2	6.3	6.4
	50	4.9	6.2	5.1	5.9	6.8	5.1	5.2
PUY's LM	10	5.0	5.7	3.7	6.4	5.5	8.3	24.8
	20	4.3	4.7	4.2	4.1	5.7	5.8	6.5
	30	4.6	4.1	3.9	6.3	4.8	5.4	6.2
	50	4.4	5.0	5.5	4.8	6.1	4.3	4.6
Pesaran's CD	10	5.4	5.2	5.9	5.7	4.4	5.0	6.2
	20	5.4	4.2	4.9	5.3	5.0	6.1	5.8
	30	4.4	4.9	5.3	4.7	6.5	5.4	5.0
	50	4.6	5.4	4.8	5.5	4.1	4.9	4.3

Note: In a fixed effects panel data model specified in Section 6, this table reports the size of the John test, Pesaran, Ullah and Yamagata (2008) (PUY) LM test and Pesaran's (2004) CD test. Homoskedasticity and normality of the idiosyncratic errors are assumed. These tests are conducted at the two-sided 5% nominal significance level.

Table 2: Size Adjusted Power of Tests: Factor Model

Size Adjusted Power	$T \setminus n$	5	10	20	30	50	100	200
John	10	9.4	14.8	37.8	47.3	71.7	88.0	94.8
	20	18.7	35.1	67.4	86.7	95.5	99.8	100.0
	30	24.6	55.3	88.1	94.7	99.6	100.0	100.0
	50	45.4	80.1	97.8	99.8	100.0	100.0	100.0
PUY's LM	10	7.1	12.2	32.2	37.6	62.9	84.2	92.0
	20	17.7	32.7	61.3	80.2	93.1	99.5	100.0
	30	23.3	51.3	84.6	93.5	99.4	100.0	100.0
	50	39.6	75.0	96.3	99.5	99.9	100.0	100.0
Pesaran's CD	10	4.9	6.9	5.2	6.8	6.6	7.0	9.8
	20	8.2	9.7	8.3	9.5	10.5	11.8	13.5
	30	11.0	8.6	9.9	12.6	9.0	12.1	15.8
	50	14.0	12.2	13.4	14.4	16.4	18.5	22.6

Note: In order to calculate the size adjusted power, a factor structure model is assumed to allow for cross-sectional dependence in the errors, see Section 6.

Table 3: Size Adjusted Power of Tests: SAR (1) Model

Size Adjusted Power	$T \setminus n$	5	10	20	30	50	100	200
John	10	33.9	34.6	40.2	34.4	42.0	44.5	45.0
	20	72.6	77.1	87.2	89.9	88.0	91.2	94.5
	30	91.5	96.9	99.6	98.9	99.1	99.6	99.8
	50	99.9	100.0	100.0	100.0	100.0	100.0	100.0
PUY's LM	10	26.6	22.5	29.4	20.0	23.9	27.4	24.6
	20	69.5	70.2	78.0	79.4	74.6	75.1	79.9
	30	91.7	95.8	98.1	97.3	97.4	98.6	99.8
	50	99.8	100.0	100.0	100.0	100.0	100.0	100.0
Pesaran's CD	10	58.9	50.4	46.2	47.8	45.3	41.5	38.9
	20	84.3	78.5	72.3	72.3	68.1	66.8	67.6
	30	96.3	87.5	87.7	85.3	85.0	82.9	82.3
	50	99.6	98.7	98.0	97.4	97.1	96.0	97.3

Note: In order to calculate the size adjusted power, a SAR (1) error structure is assumed to allow for cross-sectional dependence, see Section 6.

Table 4: Size Adjusted Power of Tests: SMA (1) Model

Size Adjusted Power	$T \setminus n$	5	10	20	30	50	100	200
John	10	16.6	19.0	22.4	16.9	22.9	24.1	25.6
	20	45.0	50.6	63.8	65.5	66.8	66.3	65.6
	30	75.2	84.2	90.2	92.1	94.6	95.2	94.8
	50	98.6	99.9	100.0	100.0	100.0	100.0	100.0
PUY's LM	10	15.9	16.6	21.2	12.6	16.2	20.6	15.4
	20	52.3	52.5	61.5	62.1	61.1	59.3	63.5
	30	85.6	85.7	90.0	90.5	91.0	91.0	89.4
	50	99.6	99.9	100.0	100.0	100.0	100.0	100.0
Pesaran's CD	10	38.2	32.2	33.6	32.3	31.7	30.6	26.8
	20	62.5	54.7	54.1	53.0	45.5	47.1	47.7
	30	83.9	70.4	66.6	67.7	68.5	64.7	67.6
	50	96.5	91.2	88.9	88.7	85.5	85.1	87.2

Note: In order to calculate the size adjusted power, a SMA (1) error structure is assumed to allow for cross-sectional dependence, see Section 6.

Table 5: Size of Tests with Non-normal Errors

Size	$T \setminus n$	N(0,0.5)			U[-2,2]			U[1,2]			Chi2(1)			t(4)		
		20	50	100	20	50	100	20	50	100	20	50	100	20	50	100
John	10	7.5	8.2	7.3	9.8	11.1	7.7	10.6	11.6	8.3	65.0	77.9	81.2	37.5	44.8	56.0
	30	5.6	5.2	6.3	8.9	9.3	7.7	9.1	9.4	7.6	83.8	92.5	94.8	52.2	63.6	75.0
PUY's LM	10	3.7	5.5	8.3	4.8	6.5	4.0	4.6	6.5	4.1	6.8	6.4	12.3	4.6	6.0	9.3
	30	3.9	4.8	5.4	5.0	4.6	4.0	4.8	4.0	4.6	8.7	8.5	8.8	6.1	6.8	6.8
Pesaran's CD	10	5.9	4.4	5.0	5.1	5.9	4.3	5.6	5.9	4.4	4.8	5.1	5.3	5.5	4.8	4.5
	30	5.3	6.5	5.4	6.1	5.4	4.5	6.3	5.7	5.2	5.4	4.8	4.3	4.7	5.9	6.5

Note: In order to check the sensitivity of the tests to non-normal disturbances, uniform distributions U[-2,2] and U[1,2], Chi-square distribution with 1 degree of freedom, Chi2(1) and t-distribution with 4 degrees of freedom, t(4) are considered. The normal case is also presented for comparison.

crucial, and our test is not robust to non-normality.

7 Conclusion

This paper proposes a new test for the null of sphericity of the disturbances of a fixed effects panel data regression model. Under homoskedasticity of the disturbances, this is equivalent to testing for no cross-sectional dependence. There are several econometric methods that model cross-section dependence, including the popular spatial correlation models and the factor models. To avoid the *ad hoc* specifications imposed on the covariance matrix, a test based on the sample covariance matrix of a fixed effects panel data regression model is proposed. Following Ledoit and Wolf (2002), we propose a John test using the fixed effects residuals. The limiting distribution of the proposed test is derived and its finite sample properties are examined using Monte Carlo experiments. The simulation results show that the John test performs better than the PUY's LM test and Pesaran's CD test and can be applied in empirical panel data studies using fixed effects residuals. However, the John test remains oversized for panels with large n and small T. Some of the limitations of our proposed test is that it is sensitive to non-normality of the disturbances, and its derivation and asymptotic distribution is heavily reliant on the normality assumption. Another limitation of our test statistic is the assumption that there is no time series dependence and full independence between time-varying regressors and time-varying unobservables. The reason we impose these restrictive assumptions is that we rely heavily on results from random matrix theory which to our knowledge has not been extended to deal with non-normality, serial correlation or endogeneity.

References

[1] Anderson, T. W. (2003), *An Introduction to Multivariate Statistical Analysis*, Wiley.

[2] Andrews, D. W. K. (2005), "Cross-section Regression with Common Shocks," *Econometrica*, 73, 1551-1585.

[3] Anselin, L., and Bera, A. K. (1998), "Spatial Dependence in Linear Regression Models with an Introduction to Spatial Econometrics," *Handbook of Applied Economic Statistics*, Amman Ullah and David E. A. Giles, eds., Marcel Dekker, New York.

[4] Arellano, M. (1987), "Computing Robust Standard Errors for Within-groups Estimators," *Oxford Bulletin of Economics and Statistics,* 49, 431-434.

[5] Bai, J. (2009), "Panel Data Models With Interactive Fixed Effects," *Econometrica*, 77, 1229–1279.

[6] Bai, Z. D. (1999), "Methodologies in Spectral Analysis of Large Dimensional Random Matrices, A Review," *Statistica Sinica*, 9, 611-677.

[7] Bai, Z. D., and Silverstein, J. (2006), *Spectral Analysis of Large Dimensional Random Matrices*, Science Press, Beijing.

[8] Baltagi, B. H. (2008), *Econometric Analysis of Panel Data*, John Wiley and Sons, Chichester.

[9] Baltagi, B. H., Song, S. H., and Koh, W. (2003), "Testing Panel Data Regression Models with Spatial Error Correlation," *Journal of Econometrics,* 117, 123-150.

[10] Baltagi, B. H., Feng, Q., and Kao, C. (2009), "Testing for Sphericity in a Fixed Effects Panel Data Model: Supplementary Appendix".

[11] Bester, C. A., Conley, T. G., and Hansen, C. B. (2009), "Inference with Depndent Data Using Cluster Covariance Estimators," Working paper, University of Chicago, Graduate School of Business, Chicago, Illinois.

[12] Birke, M., and Dette, H. (2005), "A Note on Testing the Covariance Matrix for Large Dimension," *Statistics and Probability Letters*, 74, 281-289.

[13] Breusch, T. S., and Pagan, A. R. (1980), "The Lagrange Multiplier Test and Its Application to Model Specification in Econometrics," *Review of Economic Studies*, 47, 239-254.

[14] Conley, T. G. (1999), "GMM Estimation with Cross Sectional Dependence," *Journal of Econometrics,* 92, 1-45.

[15] Conley, T. G., and Molinari, F. (2007), "Spatial Correlation Robust Inference with Errors in Location or Distance," *Journal of Econometrics,* 140, 76-96.

[16] Driscoll, J. C. and Kraay, A. C. (1998), "Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data," *Review of Economics and Statistics,* 80, 549–560.

[17] Geman, S. (1980), "A Limit Theorem for the Norm of Random Matrices," *Annals of Probability*, 8, 252-261.

[18] Hansen, C. B. (2007), "Asymptotic Properties of a Robust Variance Matrix Estimator for Panel Data when T is Large," *Journal of Econometrics,* 141, 597-620.

[19] Hsiao, C., Pesaran M. H., and Pick, A. (2009), "Diagnostic Tests of Cross Section Independence for Nonlinear Panel Data Models," Working Paper, University of Cambridge.

[20] Im, K. S., Ahn, S. C., Schmidt, P., and Wooldridge, J. M. (1999), "Efficient Estimation of Panel Data Models with Strictly Exogenous Explanatory Variables," *Journal of Econometrics*, 93, 177-201.

[21] John, S. (1971), "Some Optimal Multivariate Test," *Biometrika*, 58, 123-127.

[22] John, S. (1972), "The Distribution of a Statistic Used for Testing Sphericity of Normal Distributions," *Biometrika*, 59, 169–173.

[23] Johnstone, I. (2001), "On the Distribution of the Largest Principal Component," *Annals of Statistics*, 29, 295-327.

[24] Kapetanios, G. (2004), "On Testing for Diagonality of Large Dimensional Covariance Matrices," Working Paper, Queen Mary, University of London.

[25] Kelejian, H. H., and Prucha, I. (2007), "HAC Estimation in a Spatial Framework," *Journal of Econometrics,* 140, 131-154.

[26] Ledoit, O., and Wolf, M. (2002), "Some Hypothesis Tests for the Covariance Matrix When the Dimension is Large Compared to the Sample Size," *Annals of Statistics*, 30, 1081-1102.

[27] Ledoit, O., and Wolf, M. (2004), "A Well-conditioned Estimator for Large-Dimensional Covariance Matrices," *Journal of Multivariate Analysis*, 88, 365-411.

[28] Lee, L. (2002), "Consistency and Efficiency of Least Squares Estimation for Mixed Regressive, Spatial Autoregressive Models," *Econometric Theory*, 18, 252-277.

[29] Long, J. S., and Ervin, H. (2000), "Using Heteroskedasticity Consistent Standard Errors in the Linear Regression Model," *The American Statistician*, 54, 217-224.

[30] MacKinnon, J. G., and White, H. (1985), "Some Heteroskedasticity Consistent Covariance Matrix Estimators with Improved Finite Sample Properties," *Journal of Econometrics,* 29, 53-57.

[31] Ng, S. (2006), "Testing Cross-Section Correlation in Panel Data Using Spacings," *Journal of Business and Economic Statistics,* 24, 12-23.

[32] Pesaran, M. H. (2004), "General Diagnostic Test for Cross Section Dependence in Panels," Working Paper, University of Cambridge.

[33] Pesaran, M. H., and Tosetti, E. (2008), "Large Panels with Common Factors and Spatial Correlations," Working Paper, University of Cambridge.

[34] Pesaran, M. H., Ullah, A., and Yamagata, T. (2008), "A Bias-Adjusted LM Test of Error Cross Section Independence," *Econometrics Journal*, 11, 105-127.

[35] Srivastava, M. S. (2005), "Some Test Concerning the Covariance Matrix in High Dimensional Data," *Journal of Japanese Statistical Society*, 35, 251-272.

[36] Stock, J. H., and Watson, M. W. (2008), "Heteroskedasticity-Robust Standard Errors for Fixed Effects Panel Data Regression," *Econometrica*, 76, 155–174.

Appendix: Proofs of Propositions and Theorems

This appendix includes proofs of the main results in the text. The following lemma is frequently used in the proofs:

Lemma 1 *For a random sequence* $\{Z_n\}$*, if* $EZ_n^2 = O(n^{\nu})$*, where* ν *is a constant, then* $Z_n = O_p(n^{\nu/2})$*.*

In the fixed effects model, $y_{it} = x_{it}'\beta + \mu_i + v_{it}$, $\tilde{\beta}$ is the within estimator and the within residuals are given by $\widehat{v}_{it} = \tilde{y}_{it} - \tilde{x}_{it}'\tilde{\beta}$, where $\tilde{y}_{it} = y_{it} - \bar{y}_{i\cdot}$ and $\tilde{x}_{it} = x_{it} - \bar{x}_{i\cdot}$. Define $\tilde{v}_{it} = v_{it} - \bar{v}_{i\cdot}$, then, the residuals $\hat{v}_{it} = \tilde{v}_{it} - \tilde{x}_{it}'(\tilde{\beta} - \beta)$ and in vector form we have $\widehat{v}_t = \tilde{v}_t - \tilde{x}_t(\tilde{\beta} - \beta)$. To obtain $T(\hat{U} - U)$ in (14) we need to calculate four terms: $\frac{1}{n}trS$, $\frac{1}{n}tr(S^2)$, $\frac{1}{n}tr\hat{S}$ and $\frac{1}{n}tr(\hat{S}^2)$. The following lemma is needed to prove Theorem 1.

Lemma 2 *Under Assumptions 1 and 2,*

(a) $\frac{1}{n}trS = \sigma_v^2 + O_p\left(\frac{1}{\sqrt{nT}}\right)$;

(b) $\frac{1}{n}tr(S^2) = (\frac{n}{T} + 1)\sigma_v^4 + O_p(\frac{1}{T})$.

This lemma verifies Proposition 1 in Ledoit and Wolf (2002). Moreover, the results above provide the order of the higher order terms, which are used in the calculation of the asymptotic bias of the John statistic.

Proof. Consider (a),

$$\begin{aligned}
\frac{1}{n}trS &= \frac{1}{n}tr\left[\frac{1}{T}\sum_{t=1}^{T} v_t v_t'\right] = \frac{1}{nT}\sum_{t=1}^{T} tr\left(v_t v_t'\right) = \frac{1}{nT}\sum_{t=1}^{T} v_t' v_t = \frac{1}{nT}\sum_{t=1}^{T}\sum_{i=1}^{n} v_{it}^2 \\
&= \sigma_v^2 + \frac{1}{nT}\sum_{t=1}^{T}\sum_{i=1}^{n}(v_{it}^2 - \sigma_v^2) \\
&= \sigma_v^2 + O_p\left(\frac{1}{\sqrt{nT}}\right)
\end{aligned}$$

since $\frac{1}{\sqrt{nT}}\sum_{t=1}^{T}\sum_{i=1}^{n}(v_{it}^2 - \sigma_v^2) = O_p(1)$. This proves (a).

For (b),

$$\begin{aligned}
\frac{1}{n}tr(S^2) &= \frac{1}{n}tr\left[\left(\frac{1}{T}\sum_{t=1}^{T}v_t v_t'\right)\left(\frac{1}{T}\sum_{s=1}^{T}v_s v_s'\right)\right] = \frac{1}{nT^2}\sum_{t=1}^{T}\sum_{s=1}^{T}tr\left(v_t v_t' v_s v_s'\right) \\
&= \frac{1}{nT^2}\sum_{t=1}^{T}\sum_{s=1}^{T}v_t' v_s v_s' v_t \\
&= \frac{1}{nT^2}\sum_{t=1}^{T}v_t' v_t v_t' v_t + \frac{1}{nT^2}\sum_{s\neq t}^{T}\sum_{t=1}^{T}v_t' v_s v_s' v_t \\
&= \frac{1}{nT^2}\sum_{t=1}^{T}\sum_{j=1}^{n}\sum_{i=1}^{n}v_{it}^2 v_{jt}^2 + \frac{1}{nT^2}\sum_{s\neq t}^{T}\sum_{t=1}^{T}\sum_{j=1}^{n}\sum_{i=1}^{n}v_{it}v_{is}v_{js}v_{jt} \\
&= \frac{1}{nT^2}\sum_{t=1}^{T}\sum_{i=1}^{n}v_{it}^4 + \frac{1}{nT^2}\sum_{t=1}^{T}\sum_{j\neq i}^{n}\sum_{i=1}^{n}v_{it}^2 v_{jt}^2 \\
&\quad + \frac{1}{nT^2}\sum_{s\neq t}^{T}\sum_{t=1}^{T}\sum_{i=1}^{n}v_{it}^2 v_{is}^2 + \frac{1}{nT^2}\sum_{s\neq t}^{T}\sum_{t=1}^{T}\sum_{j\neq i}^{n}\sum_{i=1}^{n}v_{it}v_{is}v_{js}v_{jt} \\
&= I + II + III + IV.
\end{aligned}$$

It is easy to show that

$$I = \frac{1}{nT^2}\sum_{t=1}^{T}\sum_{i=1}^{n}v_{it}^4 = \frac{E(v_{it}^4)}{T} + \frac{1}{nT^2}\sum_{t=1}^{T}\sum_{i=1}^{n}\left[v_{it}^4 - E(v_{it}^4)\right] = O_p(\frac{1}{T}),$$

$$II = \frac{1}{nT^2}\sum_{t=1}^{T}\sum_{j\neq i}^{n}\sum_{i=1}^{n}v_{it}^2 v_{jt}^2 = \frac{n-1}{T}\sigma_v^4 + \frac{1}{nT^2}\sum_{t=1}^{T}\sum_{j\neq i}^{n}\sum_{i=1}^{n}(v_{it}^2 v_{jt}^2 - \sigma_v^4) = \frac{n-1}{T}\sigma_v^4 + O_p\left(\frac{1}{T\sqrt{T}}\right),$$

$$III = \frac{1}{nT^2}\sum_{s\neq t}^{T}\sum_{t=1}^{T}\sum_{i=1}^{n}v_{it}^2 v_{is}^2 = \frac{T-1}{T}\sigma_v^4 + \frac{1}{nT^2}\sum_{s\neq t}^{T}\sum_{t=1}^{T}\sum_{i=1}^{n}(v_{it}^2 v_{is}^2 - \sigma_v^4) = \frac{T-1}{T}\sigma_v^4 + O_p\left(\frac{1}{T\sqrt{n}}\right),$$

and

$$IV = \frac{1}{nT^2}\sum_{s\neq t}^{T}\sum_{t=1}^{T}\sum_{j\neq i}^{n}\sum_{i=1}^{n}v_{it}v_{is}v_{js}v_{jt} = O_p(\frac{1}{T})$$

using

$$\frac{1}{\sqrt{nT}}\sum_{t=1}^{T}\sum_{i=1}^{n}(v_{it}^4 - E(v_{it}^4)) = O_p\left(1\right),$$

$$\frac{1}{n\sqrt{T}}\sum_{t=1}^{T}\sum_{j\neq i}^{n}\sum_{i=1}^{n}(v_{it}^2 v_{jt}^2 - \sigma_v^4) = O_p\left(1\right),$$

and

$$\frac{1}{\sqrt{n}T}\sum_{s\neq t}^{T}\sum_{t=1}^{T}\sum_{i=1}^{n}(v_{it}^2 v_{is}^2 - \sigma_v^4) = O_p\left(1\right).$$

Hence,

$$\begin{aligned}
\frac{1}{n}tr(S^2) &= O_p(\frac{1}{T}) + \frac{n-1}{T}\sigma_v^4 + O_p(\frac{1}{T\sqrt{T}}) + \frac{T-1}{T}\sigma_v^4 + O_p(\frac{1}{T\sqrt{n}}) + O_p(\frac{1}{T}) \\
&= \left(\frac{n}{T}+1\right)\sigma_v^4 + O_p(\frac{1}{T})
\end{aligned}$$

as required. ■

A Proof of Proposition 1

Proof. Recall $\tilde{y}_{it} = \tilde{x}_{it}'\beta + \tilde{v}_{it}$ and

$$\tilde{\beta} - \beta = \left(\sum_{t=1}^{T} \sum_{i=1}^{n} \tilde{x}_{it}\tilde{x}_{it}' \right)^{-1} \sum_{t=1}^{T} \sum_{i=1}^{n} \tilde{x}_{it}\tilde{v}_{it}.$$

It is easy to show that

$$\tilde{\beta} - \beta = O_p\left(\frac{1}{\sqrt{nT}} \right).$$

Now $\widehat{v}_{it} = \tilde{y}_{it} - \tilde{x}_{it}'\tilde{\beta} = \tilde{x}_{it}'\beta + \tilde{v}_{it} - \tilde{x}_{it}'\tilde{\beta} = \tilde{v}_{it} - \tilde{x}_{it}'(\tilde{\beta} - \beta)$. In vector form $\hat{v}_t = \tilde{v}_t - \tilde{x}_t(\tilde{\beta} - \beta)$, where $\tilde{v}_t = v_t - \bar{v}_.$. It follows that

$$\begin{aligned} \widehat{S} - S &= \frac{1}{T}\sum_{t=1}^{T} \hat{v}_t\hat{v}_t' - \frac{1}{T}\sum_{t=1}^{T} v_t v_t' \\ &= \frac{1}{T}\sum_{t=1}^{T} [\tilde{v}_t - \tilde{x}_t(\tilde{\beta} - \beta)][\tilde{v}_t - \tilde{x}_t(\tilde{\beta} - \beta)]' - \frac{1}{T}\sum_{t=1}^{T} v_t v_t' \\ &= H + G, \end{aligned}$$

where

$$H = \frac{1}{T}\sum_{t=1}^{T} \tilde{v}_t\tilde{v}_t' - \frac{1}{T}\sum_{t=1}^{T} v_t v_t' \tag{31}$$

and

$$G = -\frac{1}{T}\sum_{t=1}^{T} \tilde{x}_t \left(\tilde{\beta} - \beta\right) \tilde{v}_t' - \frac{1}{T}\sum_{t=1}^{T} \tilde{v}_t \left(\tilde{\beta} - \beta\right)' \tilde{x}_t' + \frac{1}{T}\sum_{t=1}^{T} \tilde{x}_t \left(\tilde{\beta} - \beta\right) \left(\tilde{\beta} - \beta\right)' \tilde{x}_t'. \tag{32}$$

Hence,

$$\frac{1}{n}tr\widehat{S} - \frac{1}{n}trS = \frac{1}{n}tr(\widehat{S} - S) = \frac{1}{n}tr(H) + \frac{1}{n}tr(G). \tag{33}$$

It is easy to verify that $\sum_{t=1}^{T}\sum_{i=1}^{n} \tilde{x}_{it}'\tilde{x}_{it} = O_p(nT)$, $\sum_{t=1}^{T}\sum_{i=1}^{n} \tilde{x}_{it}\tilde{v}_{it}' = O_p\left(\sqrt{nT}\right)$. Then,

$$\begin{aligned} \frac{1}{n}tr(G) &= -\frac{1}{nT}\left(\tilde{\beta} - \beta\right)\sum_{t=1}^{T} \tilde{v}_t'\tilde{x}_t - \frac{1}{nT}\left(\tilde{\beta} - \beta\right)'\sum_{t=1}^{T} \tilde{x}_t'\tilde{v}_t + \frac{1}{nT}\left(\tilde{\beta} - \beta\right)\left(\tilde{\beta} - \beta\right)'\sum_{t=1}^{T} \tilde{x}_t'\tilde{x}_t \quad (34) \\ &= O_p\left(\frac{1}{nT}\right). \quad (35) \end{aligned}$$

We need to compute the first term of (33). Using the fact that

$$\begin{aligned} H &= \frac{1}{T}\sum_{t=1}^{T} \tilde{v}_t\tilde{v}_t' - \frac{1}{T}\sum_{t=1}^{T} v_t v_t' = \frac{1}{T}\sum_{t=1}^{T}(v_t - \bar{v}_.)(v_t - \bar{v}_.)' - \frac{1}{T}\sum_{t=1}^{T} v_t v_t' \\ &= \frac{1}{T}\sum_{t=1}^{T}(v_t v_t' - v_t\bar{v}_.' - \bar{v}_.v_t' + \bar{v}_.\bar{v}_.') - \frac{1}{T}\sum_{t=1}^{T} v_t v_t' \\ &= -\bar{v}_.\bar{v}_.', \end{aligned}$$

it follows that

$$\begin{aligned}
\frac{1}{n}tr(H) &= \frac{1}{n}tr(-\bar{v}_{\cdot}\bar{v}_{\cdot}') = -\frac{1}{n}\bar{v}_{\cdot}'\bar{v}_{\cdot} = -\frac{1}{n}\sum_{i=1}^{n}\bar{v}_{i\cdot}^{2} = -\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{T}\sum_{t=1}^{T}v_{it}\right)^{2} \\
&= -\frac{1}{nT^2}\sum_{i=1}^{n}\sum_{s=1}^{T}\sum_{t=1}^{T}v_{is}v_{it} \\
&= -\frac{1}{nT^2}\sum_{i=1}^{n}\sum_{t=1}^{T}v_{it}^2 - \frac{1}{nT^2}\sum_{i=1}^{n}\sum_{s\neq t}^{T}\sum_{t=1}^{T}v_{is}v_{it} \\
&= -\frac{\sigma_v^2}{T} + O_p\left(\frac{1}{T\sqrt{n}}\right) = O_p\left(\frac{1}{T}\right). \qquad (36)
\end{aligned}$$

Collecting (33), (34) and (36), we obtain

$$\frac{1}{n}tr\widehat{S} - \frac{1}{n}trS = O_p\left(\frac{1}{T}\right).$$

■

B Proof of Proposition 2

We define

$$\begin{aligned}
A_0 &= \bar{v}_{\cdot}\bar{v}_{\cdot}', \\
A_1 &= \frac{1}{T}\sum_{t=1}^{T}\tilde{x}_t\left(\tilde{\beta}-\beta\right)\tilde{v}_t', \\
A_2 &= A_1' = \frac{1}{T}\sum_{t=1}^{T}\tilde{v}_t\left(\tilde{\beta}-\beta\right)'\tilde{x}_t',
\end{aligned}$$

and

$$A_3 = \frac{1}{T}\sum_{t=1}^{T}\tilde{x}_t\left(\tilde{\beta}-\beta\right)\left(\tilde{\beta}-\beta\right)'\tilde{x}_t'.$$

Thus, $\widehat{S} - S = -A_0 - A_1 - A_2 + A_3$. To prove Proposition 2, we need the following lemma.

Lemma 3 *Under Assumptions 1 and 2,*

(a) $\frac{1}{n}tr(SA_1) = O_p(\frac{1}{T^2}) + O_p(\frac{1}{nT}) + O_p(\frac{1}{T\sqrt{nT}})$;

(b) $\frac{1}{n}tr(SA_3) = O_p(\frac{1}{nT})$;

(c) $\frac{1}{n}tr(A_1^2) = O_p(\frac{1}{nT^2})$;

(d) $\frac{1}{n}tr(A_1A_2) = O_p(\frac{1}{T^2})$;

(e) $\frac{1}{n}tr(A_1A_3) = O_p(\frac{1}{nT^2})$;

(f) $\frac{1}{n}tr(A_3^2) = O_p(\frac{1}{nT^2})$;

(g) $\frac{1}{n}tr(SA_0) = O_p(\frac{1}{T}) + O_p(\frac{n}{T^2})$;

(h) $\frac{1}{n}tr(A_0^2) = O_p(\frac{n}{T^2})$;

(i) $\frac{1}{n}tr(A_0A_1) = O_p(\frac{1}{T^2})$;

(j) $\frac{1}{n}tr(A_0A_3) = O_p(\frac{1}{nT^2})$.

The proofs are included in Baltagi et al. (2009).

Proof of Proposition 2. Using $2S(\hat{S}-S)+(\hat{S}-S)^2 = \hat{S}^2 - \hat{S}S + S\hat{S} - S^2$ and $tr(\hat{S}S) = tr(S\hat{S})$, we have

$$\frac{1}{n}tr\widehat{S}^2 - \frac{1}{n}tr(S^2) = \frac{1}{n}tr(\widehat{S}^2 - S^2) = 2\frac{1}{n}tr[S(\hat{S}-S)] + \frac{1}{n}tr(\hat{S}-S)^2.$$

Using the notation above,

$$\begin{aligned}
&\frac{1}{n}tr\widehat{S}^2 - \frac{1}{n}tr(S^2) \\
= \;& \frac{2}{n}tr\left(S(-A_0 - A_1 - A_2 + A_3)\right) + \frac{1}{n}tr\left((-A_0 - A_1 - A_2 + A_3)(-A_0 - A_1 - A_2 + A_3)\right) \\
= \;& \frac{2}{n}tr(-SA_0 - SA_1 - SA_2 + SA_3) \\
& + \frac{1}{n}tr(A_0^2 + A_0A_1 + A_0A_2 - A_0A_3 + A_1A_0 + A_1^2 + A_1A_2 - A_1A_3 \\
& + A_2A_0 + A_2A_1 + A_2^2 - A_2A_3 - A_3A_0 - A_3A_1 - A_3A_2 + A_3^2).
\end{aligned}$$

We have $tr(A_0A_1) = tr(A_1A_0) = tr(A_0A_2) = tr(A_2A_0)$, $tr(A_1A_2) = tr(A_2A_1)$, $tr(A_3A_1) = tr(A_1A_3) = tr(A_3A_2) = tr(A_2A_3)$, $tr(A_1^2) = tr(A_2^2)$, and $tr(SA_2) = tr(SA_1)$. Hence,

$$\begin{aligned}
&\frac{1}{n}tr\widehat{S}^2 - \frac{1}{n}tr(S^2) \\
= \;& -\frac{4}{n}tr(SA_1) + \frac{2}{n}tr(SA_3) + \frac{2}{n}tr(A_1^2) + \frac{2}{n}tr(A_1A_2) - \frac{4}{n}tr(A_1A_3) + \frac{1}{n}tr(A_3^2) \\
& -\frac{2}{n}tr(SA_0) + \frac{1}{n}tr(A_0^2) + \frac{4}{n}tr(A_0A_1) - \frac{2}{n}tr(A_0A_3).
\end{aligned}$$

Given these results, the proposition follows directly from Lemma 3. ■

C Proof of Proposition 3

Define $W_1 = \frac{1}{n}tr\hat{S} - \frac{1}{n}trS$ and $W_2 = \frac{1}{n}tr(\hat{S}^2) - \frac{1}{n}tr(S^2)$. To derive the asymptotics of $T(\hat{U} - U)$, we need to calculate the magnitudes of W_1 and W_2, which requires the following lemma.

Lemma 4 *Under Assumptions 1 and 2,*

(a) $\frac{1}{nT^2}\sum_{i=1}^{n}\sum_{t=1}^{T} v_{it}^2 = \frac{\sigma_v^2}{T} + O_p(\frac{1}{T\sqrt{nT}})$;

(b) $\frac{1}{nT^3}\sum_{i=1}^{n}\sum_{s\neq t}^{T}\sum_{t=1}^{T} v_{it}^2 v_{is}^2 = \frac{T-1}{T^2}\sigma_v^4 + O_p(\frac{1}{T^2\sqrt{n}})$;

(c) $\frac{1}{nT^3}\sum_{s=1}^{T}\sum_{j\neq i}^{n}\sum_{i=1}^{n} v_{is}^2 v_{js}^2 = \frac{(n-1)}{T^2}\sigma_v^4 + O_p(\frac{1}{T^2\sqrt{T}})$;

(d) $n\left(\frac{1}{nT^2}\sum_{i=1}^{n}\sum_{t=1}^{T} v_{it}^2\right)^2 = \frac{n}{T^2}\sigma_v^4 + O_p(\frac{\sqrt{n}}{T^2\sqrt{T}})$.

Proof. Consider (a).

$$\frac{1}{nT^2}\sum_{i=1}^{n}\sum_{t=1}^{T} v_{it}^2 = \frac{\sigma_v^2}{T} + \frac{1}{nT^2}\sum_{i=1}^{n}\sum_{t=1}^{T}(v_{it}^2-\sigma_v^2) = \frac{\sigma_v^2}{T} + O_p\left(\frac{1}{T\sqrt{nT}}\right).$$

For (b),

$$\frac{1}{nT^3}\sum_{i=1}^{n}\sum_{s\neq t}^{T}\sum_{t=1}^{T} v_{it}^2 v_{is}^2 = \frac{T-1}{T}\sigma_v^4 + \frac{1}{nT^3}\sum_{i=1}^{n}\sum_{s\neq t}^{T}\sum_{t=1}^{T}(v_{it}^2 v_{is}^2 - \sigma_v^4) = \frac{T-1}{T^2}\sigma_v^4 + O_p\left(\frac{1}{T^2\sqrt{n}}\right).$$

For (c),

$$\frac{1}{nT^3}\sum_{s=1}^{T}\sum_{j\neq i}^{n}\sum_{i=1}^{n} v_{is}^2 v_{js}^2 = \frac{n-1}{T^2}\sigma_v^4 + \frac{1}{nT^3}\sum_{s=1}^{T}\sum_{j\neq i}^{n}\sum_{i=1}^{n}(v_{is}^2 v_{js}^2 - \sigma_v^4) = \frac{n-1}{T^2}\sigma_v^4 + O_p\left(\frac{1}{T^2\sqrt{T}}\right).$$

Finally, we consider (d).

$$\begin{aligned} n\left(\frac{1}{nT^2}\sum_{i=1}^{n}\sum_{t=1}^{T} v_{it}^2\right)^2 &= \left(\frac{1}{T^2\sqrt{n}}\sum_{i=1}^{n}\sum_{t=1}^{T} v_{it}^2\right)^2 = \left[\frac{\sqrt{n}}{T}\sigma_v^2 + \frac{1}{T^2\sqrt{n}}\sum_{i=1}^{n}\sum_{t=1}^{T}(v_{it}^2-\sigma_v^2)\right]^2 \\ &= \left[\frac{\sqrt{n}}{T}\sigma_v^2 + O_p(\frac{1}{T\sqrt{T}})\right]^2 = \frac{n}{T^2}\sigma_v^4 + O_p\left(\frac{\sqrt{n}}{T^2\sqrt{T}}\right). \end{aligned}$$

This proves the lemma. ■

The magnitudes and high order terms of W_1 and W_2 are summarized in the following lemma.

Lemma 5 *Under Assumptions 1 and 2,*

(a) $W_1 = -\frac{\sigma_v^2}{T} + O_p(\frac{1}{T\sqrt{n}})$;

(b) $W_2 = -\frac{2}{T}\sigma_v^4 - \frac{n}{T^2}\sigma_v^4 + O_p\left(\frac{1}{T\sqrt{n}}\right) + O_p\left(\frac{\sqrt{n}}{T^2}\right) + O_p\left(\frac{1}{T\sqrt{T}}\right)$.

Proof. From the proof of Proposition 1 and Lemma 4, we obtain $W_1 = -\frac{\sigma_v^2}{T} + O_p(\frac{1}{T\sqrt{n}})$.

By Lemma 3, the leading terms of W_2 are from parts (g) and (h) of Lemma 3,

$$W_2 = -\frac{2}{n}tr(SA_0) + \frac{1}{n}tr(A_0^2) + O_p(\frac{1}{nT}) + O_p(\frac{1}{T^2}) + O_p\left(\frac{1}{T\sqrt{nT}}\right).$$

Lemma 3 in the supplementary appendix of Bai el al. (2009) shows that

$$\frac{1}{n}tr(SA_0) = \frac{1}{nT^3}\sum_{i=1}^{n}\sum_{s\neq t}^{T}\sum_{t=1}^{T} v_{it}^2 v_{is}^2 + \frac{1}{nT^3}\sum_{s=1}^{T}\sum_{j\neq i}^{n}\sum_{i=1}^{n} v_{is}^2 v_{js}^2 + O_p\left(\frac{1}{T\sqrt{n}}\right) + O_p\left(\frac{\sqrt{n}}{T^2}\right) + O_p\left(\frac{1}{T\sqrt{T}}\right)$$

and

$$\frac{1}{n}tr(A_0^2) = n\left(\frac{1}{nT^2}\sum_{i=1}^{n}\sum_{t=1}^{T} v_{it}^2\right)^2 + O_p\left(\frac{\sqrt{n}}{T^2}\right).$$

From Lemma 4, we obtain

$$\begin{aligned} \frac{1}{n}tr(SA_0) &= \frac{1}{T}\sigma_v^4 + \frac{n}{T^2}\sigma_v^4 + O_p\left(\frac{1}{T\sqrt{n}}\right) + O_p\left(\frac{\sqrt{n}}{T^2}\right) + O_p\left(\frac{1}{T\sqrt{T}}\right); \\ \frac{1}{n}tr(A_0^2) &= \frac{n}{T^2}\sigma_v^4 + O_p\left(\frac{\sqrt{n}}{T^2}\right). \end{aligned}$$

It follows that

$$\begin{aligned} W_2 &= \frac{1}{n}tr(\hat{S}^2) - \frac{1}{n}tr(S^2) \\ &= -2\cdot\frac{1}{n}tr(SA_0) + \frac{1}{n}tr(A_0^2) + O_p(\frac{1}{nT}) + O_p(\frac{1}{T^2}) + O_p(\frac{1}{T\sqrt{nT}}) \\ &= -2\cdot\left[\frac{1}{T}\sigma_v^4 + \frac{n}{T^2}\sigma_v^4 + O_p\left(\frac{1}{T\sqrt{n}}\right) + O_p\left(\frac{\sqrt{n}}{T^2}\right) + O_p\left(\frac{1}{T\sqrt{T}}\right)\right] \\ &\quad + \frac{n}{T^2}\sigma_v^4 + O_p\left(\frac{\sqrt{n}}{T^2}\right) + O_p(\frac{1}{nT}) + O_p(\frac{1}{T^2}) + O_p(\frac{1}{T\sqrt{nT}}) \qquad (37) \\ &= -\frac{2}{T}\sigma_v^4 - \frac{n}{T^2}\sigma_v^4 + O_p\left(\frac{1}{T\sqrt{n}}\right) + O_p\left(\frac{\sqrt{n}}{T^2}\right) + O_p\left(\frac{1}{T\sqrt{T}}\right). \qquad (38) \end{aligned}$$

■

Having the results on $\frac{1}{n}trS$, $\frac{1}{n}tr(S^2)$, W_1 and W_2, we are in good position to prove Proposition 3.

Proof of Proposition 3. Plugging the results in Lemma 2 and Lemma 5 in (14), we obtain

$$\begin{aligned} T(\hat{U} - U) &= \frac{T\left[\left(\frac{1}{n}trS\right)^2 \frac{1}{n}tr(\hat{S}^2) - \left(\frac{1}{n}tr\hat{S}\right)^2 \frac{1}{n}tr(S^2)\right]}{\left(\frac{1}{n}tr\hat{S}\right)^2 \left(\frac{1}{n}trS\right)^2} \\ &= \frac{T\left[\left(\frac{1}{n}trS\right)^2 \left(\frac{1}{n}tr(S^2) + W_2\right) - \left(\frac{1}{n}trS + W_1\right)^2 \frac{1}{n}tr(S^2)\right]}{\left(\frac{1}{n}tr\hat{S}\right)^2 \left(\frac{1}{n}trS\right)^2} \\ &= \frac{TW_2\left(\frac{1}{n}trS\right)^2 - 2TW_1\frac{1}{n}trS\frac{1}{n}tr(S^2) - TW_1^2\frac{1}{n}tr(S^2)}{\left(\frac{1}{n}tr\hat{S}\right)^2 \left(\frac{1}{n}trS\right)^2}. \end{aligned}$$

Consider the numerator,

$$
\begin{aligned}
& TW_2\left(\frac{1}{n}trS\right)^2 - 2TW_1\frac{1}{n}trS\frac{1}{n}tr(S^2) - TW_1^2\frac{1}{n}tr(S^2) \\
= \;& T\left[-\frac{2}{T}\sigma_v^4 - \frac{n}{T^2}\sigma_v^4 + O_p\left(\frac{1}{T\sqrt{n}}\right) + O_p\left(\frac{\sqrt{n}}{T^2}\right) + O_p\left(\frac{1}{T\sqrt{T}}\right)\right]\left[\sigma_v^2 + O_p(\frac{1}{\sqrt{nT}})\right]^2 \\
& -2T\left[-\frac{\sigma_v^2}{T} + O_p(\frac{1}{T\sqrt{n}})\right]\left[\sigma_v^2 + O_p(\frac{1}{\sqrt{nT}})\right]\left[(\frac{n}{T}+1)\sigma_v^4 + O_p(\frac{1}{T})\right] \\
& -T\left[-\frac{\sigma_v^2}{T} + O_p(\frac{1}{T\sqrt{n}})\right]^2\left[(\frac{n}{T}+1)\sigma_v^4 + O_p(\frac{1}{T})\right] \\
= \;& \left[-2\sigma_v^4 - \frac{n}{T}\sigma_v^4 + O_p\left(\frac{1}{\sqrt{n}}\right) + O_p\left(\frac{\sqrt{n}}{T}\right) + O_p\left(\frac{1}{\sqrt{T}}\right)\right]\left[\sigma_v^4 + O_p(\frac{1}{\sqrt{nT}})\right] \\
& + \left[2\sigma_v^2 + O_p(\frac{1}{\sqrt{n}})\right]\left[\sigma_v^2 + O_p(\frac{1}{\sqrt{nT}})\right]\left[(\frac{n}{T}+1)\sigma_v^4 + O_p(\frac{1}{T})\right] \\
& + \left[-\frac{\sigma_v^4}{T} + O_p(\frac{1}{T\sqrt{n}})\right]\left[(\frac{n}{T}+1)\sigma_v^4 + O_p(\frac{1}{T})\right] \\
= \;& -2\sigma_v^8 - \frac{n}{T}\sigma_v^8 + O_p\left(\frac{1}{\sqrt{n}}\right) + O_p\left(\frac{\sqrt{n}}{T}\right) + O_p\left(\frac{1}{\sqrt{T}}\right) \\
& +2(\frac{n}{T}+1)\sigma_v^8 + O_p(\frac{1}{\sqrt{n}}) + O_p(\frac{\sqrt{n}}{T}) + O_p(\frac{1}{T})) \\
& -(\frac{n}{T^2}+\frac{1}{T})\sigma_v^8 + O_p(\frac{\sqrt{n}}{T^2}) + O_p(\frac{1}{T\sqrt{n}}) + O_p(\frac{1}{T^2}) \\
= \;& \frac{n}{T}\sigma_v^8 - \frac{n}{T^2}\sigma_v^8 + O_p(\frac{1}{\sqrt{n}}) + O_p(\frac{\sqrt{n}}{T}) + O_p(\frac{1}{\sqrt{T}}).
\end{aligned}
$$

Similarly, the denominator is

$$
\begin{aligned}
& \left(\frac{1}{n}tr\hat{S}\right)^2\left(\frac{1}{n}trS\right)^2 = (\frac{1}{n}trS + W_1)^2\left(\frac{1}{n}trS\right)^2 \\
= \;& \left[\sigma_v^2 + O_p(\frac{1}{\sqrt{nT}}) - \frac{\sigma_v^2}{T} + O_p(\frac{1}{T\sqrt{n}})\right]^2\left[\sigma_v^2 + O_p(\frac{1}{\sqrt{nT}})\right]^2 \\
= \;& \left[\frac{(T-1)^2}{T^2}\sigma_v^4 + O_p(\frac{1}{\sqrt{nT}})\right]\left[\sigma_v^4 + O_p(\frac{1}{\sqrt{nT}})\right] \\
= \;& \frac{(T-1)^2}{T^2}\sigma_v^8 + O_p(\frac{1}{\sqrt{nT}}).
\end{aligned}
$$

It follows directly that

$$
T(\hat{U}-U) = \frac{\frac{n}{T}\sigma_v^8 - \frac{n}{T^2}\sigma_v^8 + O_p(\frac{1}{\sqrt{n}}) + O_p(\frac{\sqrt{n}}{T}) + O_p(\frac{1}{\sqrt{T}})}{\frac{(T-1)^2}{T^2}\sigma_v^8 + O_p(\frac{1}{\sqrt{nT}})}. \tag{39}
$$

Obviously, $\frac{T(\hat{U}-U)}{2} - \frac{n}{2(T-1)} \to 0$ as $(n,T)\to\infty$ with $n/T \to c \in [0,\infty)$. ■