Testing for Sphericity in Panels

Simon A. Broda*

Swiss Banking Institute, University of Zurich, Switzerland

February 2010

Abstract

This manuscript considers locally best invariant tests for sphericity in heterogeneous panel models, which are given as weighted sums of individual tests. An exact integral expression and a saddlepoint approximation for their null distributions are provided. The saddlepoint approximation generalizes the earlier result of Broda, Paolella, and Tchopourian (2006). Despite having been derived by an entirely different method of proof, the new approximation reduces to the earlier result under the proper assumptions, thus shedding some new light on the nature of the approximation. A panel stationarity test serves as a numerical example.

Keywords: Saddlepoint Approximation, Panel Data, Locally Best Test.

JEL Classification: C33, C63

**E-mail address:* broda@isb.uzh.ch.

Part of this research has been carried out within the National Centre of Competence in Research "Financial Valuation and Risk Management" (NCCR FINRISK), which is a research program supported by the *Swiss National Science Foundation*.

1 Introduction

Testing for spherically symmetric errors in the linear model has a long history. For example, the problems of testing for serial correlation, spatial correlation, unit roots, and stationarity can all be cast in this framework. In most of these problems, no uniformly most powerful test exists, which has lead to the adoption of weaker optimality criteria. The locally best invariant (LBI) test, for example, is that which maximizes the slope of the power function at the null hypothesis among all invariant tests. In the aforementioned problems, Cliff and Ord's (1973) test for spatial correlation, Dufour and King's (1991) test for serial correlation and unit roots, and Kwiatkowski, Phillips, Schmidt, and Shin's (1992) test for stationarity are all locally best invariant. The statistics associated with these tests are in the form of ratios in quadratic forms in elliptically symmetric random vectors, and, consequently, a sizeable body of literature has emerged dealing with the computation of tail probabilities of such ratios, both exactly (Grad and Solomon, 1955; Imhof, 1961) and approximately (Lieberman, 1994; Marsh, 1998; Butler and Paolella, 2008).

Recently, there has been considerable interest in the corresponding test procedures for panel data models, see, for example, Baltagi (2005) and the references therein. Typically, the tests considered in the literature are based on the assumptions of either homogeneous, or known heterogenous variances. The former assumption is unrealistic; the latter requires that a feasible test be constructed, which, in general, does not share the optimality properties of the exact test. The stationarity test of Hadri and Larsson (2005) is an exception, and we shall consider it as an example below.

In this manuscript, we consider an exact locally best invariant test for spherical symmetry in heterogeneous panel models. In the absence of cross-sectional correlation, the test statistic is given by a weighted sum of the corresponding individual test statistics, each of which is in the form of a ratio of quadratic forms in an elliptically symmetric random vector. In the limit as the number of individuals tends to infinity, the null distribution converges to a Gaussian; if, however, one is to control the size of the test in finite samples, then the finite-sample null distribution is required, which, unlike in the pure time-series case, appears to have eluded computation to date.

This paper provides both an exact expression and a remarkably accurate saddlepoint approximation for the null distribution. The latter is a generalization of a result of Broda et al. (2006), which was based on the assumption of identically distributed summands. The present manuscript removes this assumption; in particular, the matrices appearing in the numerators of the summands are no longer assumed to be identical. Broda et al. derived their approximation by applying a Laplace approximation to the marginalizing integral over the saddlepoint approximation to the joint density, and then further approximating the distribution function by a Temme (1982) approximation. In the present setting, it appears more straightforward to derive the approximation via a different route, similar to the approach used in Butler and Paolella (1998) to obtain the joint density of the serial correlogram from a linear regression. Surprisingly however, in the i.i.d. case, the two approximations coincide.

The remainder of this manuscript is organized as follows. Section 2 details the theory of the locally best invariant test. Section 3 provides an exact expression for its null distribution. Section 4 derives the saddlepoint approximation. Section 5 proves that under the assumptions made there, the approximation coincides with that of Broda et al. (2006). In Section 6, the accuracy of the approximation is exemplified by applying it to the panel stationarity test of Hadri and Larsson (2005). Section 7 concludes.

2 Locally Best Tests in Heterogenous Panels

Consider the heterogenous panel data model

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta}_i + \mathbf{u}_i, \quad i \in \{1 \ldots N\}, \tag{1}$$

where $\mathbf{X}_i$ is a known constant $T_i \times k_i$ matrix of rank k_i, and the $\mathbf{u}_i$ are mutually independent $T_i \times 1$ random vectors. Each $\mathbf{u}_i$ is distributed according to an elliptically symmetric law, with density

$$f_i(\mathbf{u}_i; \theta, \sigma_i) = \left|\sigma_i^2 \boldsymbol{\Sigma}_i(\theta)\right|^{-1/2} g_i\left(\mathbf{u}_i' \left(\sigma_i^2 \boldsymbol{\Sigma}_i(\theta)\right)^{-1} \mathbf{u}_i\right), \tag{2}$$

for some known function $g_i : \mathbb{R}_+ \to \mathbb{R}_+$ such that f_i is a valid density with respect to Lebesgue measure on $\mathbb{R}_{T_i}$. In (2), each $\boldsymbol{\Sigma}_i(\theta)$ is a known, symmetric, differentiable matrix function of the common, but unknown parameter $\theta \in \Theta \equiv \{\theta \in \mathbb{R} : \boldsymbol{\Sigma}_i > 0 \forall i\}$, and without any loss of generality, we assume that $\boldsymbol{\Sigma}_i(0) = \mathbf{I}_{T_i}$. (Typically, matrices $\boldsymbol{\Sigma}_i(\theta)$ will have the same structure, but possibly different dimensions.) We consider the problem of testing $H_0 : \theta = 0$ against $H_a : \theta = a \in \Theta \cap \mathbb{R}_+$.

Let

$$\mathbf{M}_i \equiv \begin{cases} \mathbf{I}_{T_i} - \mathbf{X}_i \left(\mathbf{X}_i'\mathbf{X}_i\right)^{-1} \mathbf{X}_i', & \text{if } k_i > 0, \\ \mathbf{I}_{T_i}, & \text{otherwise,} \end{cases}$$

and for each i, choose a matrix $\mathbf{P}_i$ such that $\mathbf{P}_i\mathbf{P}_i' = \mathbf{I}_{T_i - k_i}$ and $\mathbf{P}_i'\mathbf{P}_i = \mathbf{M}_i$. Let $\mathbf{y} = [\mathbf{y}_1', \ldots, \mathbf{y}_N']'$; then the vector $\mathbf{v} = [\mathbf{v}_1', \ldots, \mathbf{v}_N']'$, where $\mathbf{v}_i = \mathbf{P}_i\mathbf{y}_i / \|\mathbf{P}_i\mathbf{y}_i\|$, is a maximal invariant with respect to transformations of the form

$$\mathbf{y} \to \tilde{\mathbf{y}}, \quad \tilde{\mathbf{y}} = [\tilde{\mathbf{y}}_1', \ldots, \tilde{\mathbf{y}}_N']', \quad \tilde{\mathbf{y}}_i = a_i\mathbf{y}_i + \mathbf{X}_i\mathbf{b}_i. \tag{3}$$

The density (with respect to the uniform measure on the unit m_i-sphere) of $\mathbf{v}_i$ under H_θ is

$$h_i(\mathbf{v}_i; \theta) = \left|\mathbf{P}_i\boldsymbol{\Sigma}_i(\theta)\mathbf{P}_i'\right|^{-1/2} \left[\mathbf{v}_i' \left(\mathbf{P}_i\boldsymbol{\Sigma}_i(\theta)\mathbf{P}_i'\right)^{-1} \mathbf{v}_i\right]^{-\frac{m_i}{2}},$$

where $m_i = T_i - k_i$; see Kariya (1980) and King (1980). Thus, from independence, the density of $\mathbf{v}$ is

$$h(\mathbf{v}; \theta) = \prod_{i=1}^{N} h_i(\mathbf{v}_i; \theta), \tag{4}$$

based on which optimal invariant tests can be constructed. From Ferguson (1967, p. 236), a locally best invariant test of size α of $H_0 : \theta = 0$ against $H_a : \theta = a \in \Theta \cap \mathbb{R}_+$ is to reject H_0 whenever

$$\frac{\partial}{\partial\theta} \log h(\mathbf{v}; \theta)\Big|_{\theta=0} > d_1,$$

where d_1 is a constant chosen such that the size of the test is α. Applied to the density (4), this yields critical regions of the form

$$\tau(\mathbf{y}) \equiv \sum_{i=1}^{N} m_i \mathbf{v}_i' \mathbf{P}_i \dot{\mathbf{\Sigma}}_i(0) \mathbf{P}_i' \mathbf{v}_i = \sum_{i=1}^{N} m_i \frac{\mathbf{y}_i' \mathbf{M}_i \dot{\mathbf{\Sigma}}_i(0) \mathbf{M}_i \mathbf{y}_i}{\mathbf{y}_i' \mathbf{M}_i \mathbf{y}_i} > d,$$

where $\dot{\mathbf{\Sigma}}_i(\theta)$ denotes the elementwise derivative of $\mathbf{\Sigma}_i(\theta)$ with respect to θ. The constant d satisfies

$$\begin{aligned} 1 - \alpha = \Pr_{\theta=0}[\tau(\mathbf{y}) < d] &= \Pr_{\theta=0}\left[\sum_{i=1}^{N} m_i \frac{\mathbf{u}_i' \mathbf{M}_i \dot{\mathbf{\Sigma}}_i(0) \mathbf{M}_i \mathbf{u}_i}{\mathbf{u}_i' \mathbf{M}_i \mathbf{u}_i} < d\right] \\ &= \Pr_{\theta=0}\left[\sum_{i=1}^{N} m_i \frac{\tilde{\mathbf{u}}_i' \mathbf{A}_i \tilde{\mathbf{u}}_i}{\tilde{\mathbf{u}}_i' \mathbf{I}_{m_i} \tilde{\mathbf{u}}_i} < d\right], \end{aligned} \tag{5}$$

where $\mathbf{A}_i = \mathbf{P}_i \dot{\mathbf{\Sigma}}_i(0) \mathbf{P}_i'$, $\tilde{\mathbf{u}}_i \equiv \mathbf{P}_i \mathbf{u}_i$, and under H_0, $\tilde{\mathbf{u}}_i$, like $\mathbf{u}_i$, has a spherically symmetric density

$$\tilde{f}_i(\tilde{\mathbf{u}}_i) = \tilde{g}_i\left(\tilde{\mathbf{u}}_i' \tilde{\mathbf{u}}_i\right), \tag{6}$$

see Kelker (1970).

3 Exact Null Distribution

In order to determine the critical value d in (5), we require the distribution function of statistics of the form

$$\bar{R} \equiv \frac{1}{N} \sum_{i=1}^{N} R_i, \quad \text{where } R_i \equiv \frac{U_i}{D_i} \equiv \frac{\mathbf{u}_i' \mathbf{A}_i \mathbf{u}_i}{\mathbf{u}_i' \mathbf{u}_i}, \tag{7}$$

each $\mathbf{A}_i$ is a symmetric $m_i \times m_i$ matrix, $\mathbf{u}_i$ has density (6), and, for notational convenience, we write $\mathbf{u}$ and g instead of $\tilde{\mathbf{u}}$ and $\tilde{g}$. The characteristic function of R_i is

$$\psi_{R_i}(t) \equiv \mathbb{E}[\exp\{\imath t R_i\}] = {}_1F_1(1/2, m_i/2, \imath t \mathbf{A}_i), \tag{8}$$

where $\imath^2 \equiv -1$, see, e.g., Hillier (2001). In (8), ${}_1F_1(a, b, \mathbf{Z})$ denotes the confluent hypergeometric function of matrix argument. The fact that this function is notoriously difficult to evaluate with high precision makes it doubtful that the distribution function of $\bar{R}$ could be efficiently obtained by numerical inversion of the characteristic function. When $N = 1$, this difficulty is usually circumvented by exploiting the relation

$$\Pr\left(\frac{\mathbf{u}_1' \mathbf{A}_1 \mathbf{u}_1}{\mathbf{u}_1' \mathbf{u}_1} \le r\right) = \Pr\left(\mathbf{u}_1' [\mathbf{A}_1 - r \mathbf{I}_{m_1}] \mathbf{u}_1 \le 0\right),$$

but this method fails when $N > 1$. As such, we shall consider a different approach in what follows.

In order to simplify the derivations, suppose for the moment that $m_i = T$ for all i; if $m_i \neq m_j$ for some i, j, all occurrences of T should be replaced by m_i in Theorems 1 and 2; such would be the situation in an unbalanced panel, or if a different number of regressors is included for at least some individuals. Noting that the null distribution of $\bar{R}$ is independent of the specific choice of g_i in (6), we shall further assume, without loss of generality, that

$$g_i(\mathbf{u}_i'\mathbf{u}_i) = (2\pi)^{-T/2} \exp\{\mathbf{u}_i'\mathbf{u}_i\} \quad \forall i,$$

i.e., the $\mathbf{u}_i$ are independently distributed as $\mathrm{N}(\mathbf{0}, \mathbf{I}_T)$.

In order to derive an expression for the distribution of $\bar{R}$ in (7), we will exploit the fact that

$$R_i \perp D_j \quad \forall i, j \in \{1, \ldots, N\},$$

i.e., each ratio R_i in (7) is independent of its own denominator D_i, and, trivially, also of the denominators of the remaining ratios in the sum. We thus have that

$$\begin{aligned}\Pr\left(\bar{R} \leq \bar{r}\right) &= \Pr\left(\frac{1}{N}\sum_{i=1}^{N}\frac{U_i}{D_i} \leq \bar{r}\right) = \Pr\left(\frac{1}{N}\sum_{i=1}^{N}\frac{U_i}{D_i} \leq \bar{r} \,\middle|\, D_i = 1 \; \forall i \in \{1, \ldots, N\}\right) \\ &= \Pr\left(\sum_{i=1}^{N} U_i \leq r \,\middle|\, D_i = 1 \; \forall i \in \{1, \ldots, N\}\right), \end{aligned} \tag{9}$$

where $r = \bar{r}N$.

The inversion formula for the joint density of $X := \sum_{i=1}^{N} U_i$ and $\mathbf{D} = [D_1, \ldots, D_N]'$ is

$$f_{X,\mathbf{D}}(x, \mathbf{d}) = \frac{1}{(2\pi\imath)^{N+1}} \int_{c-\imath\infty}^{c+\imath\infty} \int_{-\imath\infty}^{\imath\infty} \cdots \int_{-\imath\infty}^{\imath\infty} e^{\mathbb{K}(s,\mathbf{t}) - sx - \mathbf{t}'\mathbf{d}} \mathrm{d}t_1 \cdots \mathrm{d}t_N \mathrm{d}s,$$

where $\mathbf{t} = (t_1, \ldots, t_N)'$, $\mathbb{K}(s, \mathbf{t})$ denotes the joint cumulant generating function of X and $\mathbf{D}$, and the constant $c < 0$ lying in the convergence strip of $\mathbb{K}$ is required to justify the application of Fubini's theorem later. Hence,

$$\begin{aligned} F_{\bar{R}}(\bar{r}) = F_{X|\mathbf{D}}(r|\mathbf{D} = \mathbf{1}) &= \int_{-\infty}^{r} f_{X|\mathbf{D}}(x|\mathbf{D} = \mathbf{1})\mathrm{d}x = \frac{1}{f_{\mathbf{D}}(\mathbf{1})} \int_{-\infty}^{r} f_{X,\mathbf{D}}(x, \mathbf{1})\mathrm{d}x \\ &= \frac{1}{(2\pi\imath)^{N+1} f_{\mathbf{D}}(\mathbf{1})} \int_{-\infty}^{r} \int_{c-\imath\infty}^{c+\imath\infty} \int_{-\imath\infty}^{\imath\infty} \cdots \int_{-\imath\infty}^{\imath\infty} e^{K(s,\mathbf{t}) - sx - \sum_{i=1}^{N} t_i} \mathrm{d}t_1 \cdots \mathrm{d}t_N \mathrm{d}s \mathrm{d}x \\ &= \frac{-1}{(2\pi\imath)^{N+1} f_{\mathbf{D}}(\mathbf{1})} \int_{c-\imath\infty}^{c+\imath\infty} \int_{-\imath\infty}^{\imath\infty} \cdots \int_{-\imath\infty}^{\imath\infty} e^{K(s,\mathbf{t}) - sr - \sum_{i=1}^{N} t_i} \mathrm{d}t_1 \cdots \mathrm{d}t_N \frac{\mathrm{d}s}{s}. \end{aligned} \tag{10}$$

The D_i are independently distributed as χ^2_T, so that

$$f_{\mathbf{D}}(\mathbf{1}) = \left[f_{\chi^2}(1; T)\right]^N = \left[\frac{2^{-T/2} e^{-1/2}}{\Gamma(T/2)}\right]^N,$$

and it remains to find an expression for $\mathbb{K}(s,\mathbf{t})$. Upon defining $\mathbf{u}=(\mathbf{u}_1,\ldots,\mathbf{u}_N)'$, a block diagonal matrix $\mathbf{A}$ with ith block diagonal element $\mathbf{A}_i$, $i\in\{1,\ldots,N\}$, and the block diagonal matrices $\mathbf{S}_i$, $i\in\{1,\ldots,N\}$, with ith block diagonal element $\mathbf{I}_T$ and zeroes everywhere else, we may write $X=\sum_{i=1}^N U_i=\mathbf{u}'\mathbf{A}\mathbf{u}$ and $D_i=\mathbf{u}'\mathbf{S}_i\mathbf{u}$, $i\in\{1,\ldots,N\}$, and it is easily seen that

$$\begin{aligned}\mathbb{K}(s,\mathbf{t}) &\equiv \log\mathbb{E}\left[\exp\left\{s\sum_{i=1}^N U_i+\sum_{i=1}^N t_iD_i\right\}\right]\\ &= -\frac{1}{2}\log\left|\mathbf{I}_{NT}-2s\mathbf{A}-2\sum_{i=1}^N t_i\mathbf{S}_i\right|\\ &= \frac{1}{2}\sum_{i=1}^N\sum_{j=1}^T\log\nu_{ij},\end{aligned}$$

where $\nu_{ij}\equiv(1-2s\omega_{ij}-2t_i)^{-1}$, and for each value of i, $\{\omega_{ij}\}_{j\in\{1,\ldots,T\}}$ are the eigenvalues of $\mathbf{A}_i$. We thus have the following result.

THEOREM 1. *Let $R_i\equiv\mathbf{u}_i'\mathbf{A}_i\mathbf{u}_i/\mathbf{u}_i'\mathbf{u}_i, i\in\{1,\ldots,N\}$, where the $\mathbf{u}_i$ are independently distributed with densities of the form $g_i(\mathbf{u}_i'\mathbf{u}_i)$. Then the distribution function of $\bar{R}\equiv N^{-1}\sum_{i=1}^N R_i$ can be computed as*

$$F_{\bar{R}}(\bar{r})=\frac{\imath}{2\pi}\int_{c-\imath\infty}^{c+\imath\infty}\frac{e^{-sN\bar{r}}}{s}\prod_{i=1}^N M_i(s)\mathrm{d}s, \tag{11}$$

where

$$M_i(s)\equiv\frac{2^{T/2}\Gamma(T/2)e^{1/2}}{2\pi\imath}\int_{-\imath\infty}^{\imath\infty}e^{-\frac{1}{2}\sum_{j=1}^T\log(1-2s\omega_{ij}-2t)-t}\mathrm{d}t,$$

$\{\omega_{ij}\}_{j\in\{1,\ldots,T\}}$ are the eigenvalues of $(\mathbf{A}_i+\mathbf{A}_i')/2$, and c lies to the left of the origin in the interval $\{s\in\mathbb{R}:s\omega_{ij}<1/2\forall i,j\}$.

We have thus reduced the $N+1$ fold integral to a double integral, which lends itself to numeric evaluation.

4 Saddlepoint Approximation

Even though Theorem 1 provides a computable expression for the critical values of the panel LBI test, its computational complexity may be unacceptably high for routine applications. As such, it will be useful to have available an approximation which improves upon the accuracy of the normal approximation, while at the same time maintaining relative computational simplicity. Such a trade-off is afforded by the saddlepoint approximation. However, because the moment generating function of $\bar{R}$ — as a product of confluent hypergeometric functions of matrix argument — is numerically intractable, it appears infeasible to obtain an asymptotic expansion of its distribution function by a direct application of saddlepoint methods. Instead, the idea is to approximate the conditional distribution appearing in (9) by the double saddlepoint approximation of Skovgaard (1987), which we briefly discuss next.

Let X and $\boldsymbol{Y}$ have dimensions 1×1 and $d \times 1$, respectively, and assume that the random vector $(X, \boldsymbol{Y}')'$ possesses a joint density and a joint cumulant generating function $\mathbb{K}(s, \mathbf{t}) \equiv \log \mathbb{E}[\exp(sX + \mathbf{t}'\boldsymbol{Y})]$. Denote by $\mathbb{K}_{\mathfrak{s}}, \mathfrak{s} \in \{s, \mathbf{t}\}$, the vector of partial derivatives of $\mathbb{K}$ with respect to the elements of $\mathfrak{s}$, and by $\mathbb{K}''(s, \mathbf{t})$ its Hessian. Skovgaard (1987) shows that a saddlepoint approximation for the conditional distribution of X given $\boldsymbol{Y} = \mathbf{y}$ is given by

$$\Pr\left(X \leq x, \boldsymbol{Y} = \mathbf{y}\right) \approx \Phi(\hat{w}) + \phi(\hat{w})\left(\hat{w}^{-1} - \hat{u}^{-1}\right), \tag{12}$$

where

$$\hat{w} \equiv \operatorname{sgn}(\hat{s})\sqrt{2\left(\hat{s}x + \hat{\mathbf{t}}'\mathbf{y} - K\left(\hat{s}, \hat{\mathbf{t}}\right) - \hat{\mathbf{t}}_0'\mathbf{y} + \mathbb{K}\left(0, \hat{\mathbf{t}}_0\right)\right)} \quad \text{and} \quad \hat{u} \equiv \hat{s}\sqrt{\left|\mathbb{K}''\left(\hat{s}, \hat{\mathbf{t}}\right)\right| \Big/ \left|\mathbb{K}''\left(0, \hat{\mathbf{t}}_0\right)\right|}.$$

The quantity $\left(\hat{s}, \hat{\mathbf{t}}\right)$ appearing in (12) is commonly referred to as the numerator saddlepoint. It solves the system

$$\begin{aligned} \mathbb{K}_s\left(\hat{s}, \hat{\mathbf{t}}\right) &= x \\ \mathbb{K}_{\mathbf{t}}\left(\hat{s}, \hat{\mathbf{t}}\right) &= \mathbf{y}. \end{aligned}$$

The denominator saddlepoint $\hat{\mathbf{t}}_0$ solves $\mathbb{K}_{\mathbf{t}}(0, \hat{\mathbf{t}}_0) = \mathbf{y}$. Approximation (12) is the leading term in an asymptotic expansion, the second-order term in which has been derived in Kolassa (1996).

In order to apply this result to the problem at hand, we require derivatives of the cumulant generating function. In obvious notation, they are given by

$$\mathbb{K}_s(s, \mathbf{t}) = \sum_{i=1}^{N}\sum_{j=1}^{T}\omega_{ij}\nu_{ij}, \qquad \mathbb{K}_{t_i}(s, \mathbf{t}) = \sum_{j=1}^{T}\nu_{ij}, \qquad \mathbb{K}_{ss}(s, \mathbf{t}) = 2\sum_{i=1}^{N}\sum_{j=1}^{T}\omega_{ij}^2\nu_{ij}^2,$$

$$\mathbb{K}_{st_i}(s, \mathbf{t}) = 2\sum_{j=1}^{T}\omega_{ij}\nu_{ij}^2, \qquad \mathbb{K}_{t_it_k}(s, \mathbf{t}) = 0,\ i \neq k,\ \text{and} \qquad \mathbb{K}_{t_it_i}(s, \mathbf{t}) = 2\sum_{j=1}^{T}\nu_{ij}^2,$$

so that the numerator saddlepoint solves

$$\begin{aligned} r &= \sum_{i=1}^{N}\sum_{j=1}^{T}\omega_{ij}\hat{\nu}_{ij} \\ 1 &= \sum_{j=1}^{T}\hat{\nu}_{ij}, \quad i \in \{1, \ldots, N\}, \end{aligned} \tag{13}$$

where hatted quantities depend on $(\hat{s}, \hat{\mathbf{t}})$ rather than $(s, \mathbf{t})$. This system of $N + 1$ equations must be solved numerically for each value of $\bar{r}$, which, owing to the sparsity of the problem's Jacobian, is a less daunting task than may at first appear.

The denominator saddlepoint $\hat{\mathbf{t}}_0 = (t_{0,1}, \ldots t_{0,N})'$ is given analytically as $\hat{t}_{i,0} = (1 - T)/2,\ i \in \{1, \ldots, N\}$, so that

$$\mathbb{K}\left(0, \hat{\mathbf{t}}_0\right) = -(NT/2)\log T \quad \text{and} \quad \left|\mathbb{K}''\left(0, \hat{\mathbf{t}}_0\right)\right| = (2/T)^N. \tag{14}$$

A further simplification occurs in (12) by noting that it follows from (13) that

$$\begin{aligned}
\sum_{i=1}^{N}\sum_{j=1}^{T}(\omega_{ij}-\bar{r})\hat{\nu}_{ij} = 0 &\Leftrightarrow \qquad (15)\\
\hat{s}\sum_{i=1}^{N}\sum_{j=1}^{T}(\omega_{ij}-\bar{r})\hat{\nu}_{ij} + \sum_{i=1}^{N}\hat{t}_i\sum_{j=1}^{T}\hat{\nu}_{ij} = \hat{s}\cdot 0 + \hat{\mathbf{t}}'\mathbf{1} &\Leftrightarrow\\
\sum_{i=1}^{N}\sum_{j=1}^{T}(\hat{s}\omega_{ij}-\hat{s}\bar{r}+\hat{t}_i)\hat{\nu}_{ij} = \hat{\mathbf{t}}'\mathbf{1} &\Leftrightarrow\\
-\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{T}(1-2\hat{s}\omega_{ij}-2\hat{t}_i+2\hat{s}\bar{r}-1)\hat{\nu}_{ij} = \hat{\mathbf{t}}'\mathbf{1} &\Leftrightarrow\\
-\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{T}(\hat{\nu}_{ij}^{-1}+2\hat{s}\bar{r}-1)\hat{\nu}_{ij} = \hat{\mathbf{t}}'\mathbf{1} &\Leftrightarrow\\
\frac{N(1-T)}{2} = \hat{\mathbf{t}}_0'\mathbf{1} = \hat{s}r + \hat{\mathbf{t}}'\mathbf{1}&, \qquad (16)
\end{aligned}$$

so that

$$\hat{w} = \operatorname{sgn}(\hat{s})\sqrt{2\Big(\mathbb{K}\big(0,\hat{\mathbf{t}}_0\big) - K\big(\hat{s},\hat{\mathbf{t}}\big)\Big)} = \operatorname{sgn}(\hat{s})\sqrt{-\sum_{i=1}^{N}\sum_{j=1}^{T}\log(T\hat{\nu}_{ij})}.$$

Next, we will simplify the expression for $\hat{u}$. In order to economize the notation, let $\hat{\kappa}_{00} \equiv \mathbb{K}_{ss}(\hat{s},\hat{\mathbf{t}})$, $\hat{\kappa}_{0i} \equiv \mathbb{K}_{st_i}(\hat{s},\hat{\mathbf{t}}), i \in \{1,\ldots,N\}$, and $\hat{\kappa}_i \equiv \mathbb{K}_{t_it_i}(\hat{s},\hat{\mathbf{t}}), i \in \{1,\ldots,N\}$. Then differentiating equations (13) and (16) with respect to r gives

$$\frac{\mathrm{d}\hat{s}}{\mathrm{d}r} = \left(\hat{\kappa}_{00} - \sum_{i=1}^{N}\hat{\kappa}_{0i}^2\hat{\kappa}_i^{-1}\right)^{-1}, \quad \frac{\mathrm{d}\hat{t}_i}{\mathrm{d}r} = -\frac{\mathrm{d}\hat{s}}{\mathrm{d}r}\frac{\hat{\kappa}_{0i}}{\hat{\kappa}_i}, \text{ and } \quad 0 = \frac{\mathrm{d}\hat{s}}{\mathrm{d}r}r + \hat{s} + \sum_{i=1}^{N}\frac{\mathrm{d}\hat{t}_i}{\mathrm{d}r},$$

so that

$$\begin{aligned}
\hat{s}\left(\hat{\kappa}_{00} - \sum_{i=1}^{N}\hat{\kappa}_{0i}^2\hat{\kappa}_i^{-1}\right) &= -r + \sum_{i=1}^{N}\hat{\kappa}_{0i}\hat{\kappa}_i^{-1}\\
&= \hat{s}^{-1}\left(-\hat{s}r + \sum_{i=1}^{N}\frac{2\sum_{j=1}^{T}\hat{s}\omega_{ij}\nu_{ij}^2}{2\sum_{j=1}^{T}\nu_{ij}^2}\right)\\
&= \hat{s}^{-1}\left(-\hat{s}r + \sum_{i=1}^{N}\frac{1}{2} - \hat{t}_i + \frac{\sum_{j=1}^{T}-\hat{\nu}_{ij}^2 + 2\hat{t}_i\nu_{ij}^2 + 2\hat{s}\omega_{ij}\nu_{ij}^2}{2\sum_{j=1}^{T}\nu_{ij}^2}\right)\\
&= \hat{s}^{-1}\left(-\hat{s}r + \frac{N}{2} - \left(\frac{N(1-T)}{2} - \hat{s}r\right) + \sum_{i=1}^{N}\frac{-\sum_{j=1}^{T}\hat{\nu}_{ij}}{2\sum_{j=1}^{T}\nu_{ij}^2}\right)\\
&= \hat{s}^{-1}\left(\frac{NT}{2} - \sum_{i=1}^{N}\hat{\kappa}_i^{-1}\right), \qquad (17)
\end{aligned}$$

where the penultimate and last equalities follow from (16) and (13), respectively. The determinant of $\mathbb{K}''(\hat{s}, \hat{\mathbf{t}})$ appearing in (12) evaluates to

$$\left|\mathbb{K}''\left(\hat{s}, \hat{\mathbf{t}}\right)\right| = \left(\hat{\kappa}_{00} - \sum_{i=1}^{N} \hat{\kappa}_{0i}^2 \hat{\kappa}_i^{-1}\right) \prod_{i=1}^{N} \hat{\kappa}_i,$$

which, together with (14) and (17), and with $\hat{\gamma}_i \equiv T\hat{\kappa}_i/2, i \in \{1, \ldots, N\}$, yields

$$\hat{u} = \operatorname{sgn}(\hat{s}) \sqrt{\frac{1}{2} \left[\prod_{i=1}^{N} \hat{\gamma}_i\right] \sum_{i=1}^{N} T(1 - \hat{\gamma}_i^{-1})}.$$

We collect the relevant formulae in the following theorem.

Theorem 2. *Let $R_i \equiv \mathbf{u}_i'\mathbf{A}_i\mathbf{u}_i/\mathbf{u}_i'\mathbf{u}_i, i \in \{1, \ldots, N\}$, where the $\mathbf{u}_i$ are independently distributed with densities of the form $g_i(\mathbf{u}_i'\mathbf{u}_i)$. Then a saddlepoint approximation to the distribution function of $\bar{R} \equiv N^{-1} \sum_{i=1}^{N} R_i$ is given by*

$$\hat{F}(\bar{r}) \equiv \Phi(\hat{w}) + \phi(\hat{w})\left(\hat{w}^{-1} - \hat{u}^{-1}\right), \tag{18}$$

where

$$\hat{w} = \operatorname{sgn}(\hat{s}) \sqrt{-\sum_{i=1}^{N} \sum_{j=1}^{T} \log(T\hat{\nu}_{ij})}, \quad \hat{u} = \operatorname{sgn}(\hat{s}) \sqrt{\frac{1}{2} \left[\prod_{i=1}^{N} \hat{\gamma}_i\right] \sum_{i=1}^{N} T(1 - \hat{\gamma}_i^{-1})},$$

$\hat{\nu}_{ij} = (1 - 2\hat{s}\omega_{ij} - 2\hat{t}_i)^{-1}$, $\hat{\gamma}_i = T \sum_{j=1}^{T} \hat{\nu}_{ij}^2$, $\{\omega_{ij}\}_{j \in \{1, \ldots, T\}}$ *are the eigenvalues of* $(\mathbf{A}_i + \mathbf{A}_i')/2$, *and the saddlepoint* $(\hat{s}, \hat{\mathbf{t}})$ *solves*

$$\bar{r} = N^{-1} \sum_{i=1}^{N} \sum_{j=1}^{T} \omega_{ij}\hat{\nu}_{ij}, \quad 1 = \sum_{j=1}^{T} \hat{\nu}_{ij}, \quad i \in \{1, \ldots, N\}.$$

5 Equivalence in the i. i. d. Case

In the i.i.d. case, i.e., under the assumption that $\mathbf{A}_i = \mathbf{A}\ \forall i$, Broda et al. (2006) derived an approximation to the distribution function of $\bar{R}$ as

$$\tilde{F}(\bar{r}) \equiv \Phi(\tilde{w}) + \phi(\tilde{w}) \left(\tilde{w}^{-1} - \tilde{u}^{-1}\right), \tag{19}$$

where

$$\tilde{w} = \operatorname{sgn}(\tilde{s}) \sqrt{-N \sum_{j=1}^{T} \log \tilde{\nu}_j}, \quad \tilde{u} = \tilde{s} \sqrt{2N \sum_{j=1}^{T} \lambda_j^2 \tilde{\nu}_j^2 \left[1 + 4\tilde{s}^2 \sum_{j=1}^{T} \lambda_j^2 \tilde{\nu}_j^2 / T\right]^{N-1}},$$

$\lambda_j = \omega_j - \bar{r}$, ω_j, $j \in \{1, \ldots, T\}$, are the eigenvalues of $\mathbf{A}$, $\tilde{\nu}_j = (1 - 2\tilde{s}\lambda_j)^{-1}$, and the saddlepoint $\tilde{s}$ solves

$$\sum_{j=1}^{T} \frac{\lambda_j}{1 - 2\tilde{s}\lambda_j} = 0. \tag{20}$$

In order to demonstrate the equivalence of (18) and (19) if $\mathbf{A}_i = \mathbf{A}\ \forall i$, we have to show that $\hat{w} \equiv \tilde{w}$ and $\hat{u} \equiv \tilde{u}$ for all values of $\bar{r}$. To that end, first note that $\omega_{ij} = \omega_j\ \forall i, j$. Thus, from (13), $\hat{t}_i = \hat{t}\ \forall i$, where, from (16)

$$\hat{t} = \frac{1-T}{2} - \hat{s}\bar{r},$$

so that $\hat{\nu}_{ij} = (T - 2\hat{s}\lambda_j)^{-1}\ \forall i, j$. Substituting this into (15) and simplifying yields

$$\sum_{j=1}^{T} \frac{\lambda_j}{1 - 2(\hat{s}/T)\lambda_j} = 0,$$

which, upon comparison with (20), gives $T\tilde{s} = \hat{s}$ and $T\hat{\nu}_{ij} = \tilde{\nu}_j\ \forall i, j$. It thus follows from the definition of $\hat{w}$ that $\hat{w} \equiv \tilde{w}$.

We now turn to proving that $\hat{u} \equiv \tilde{u}$, or equivalently (as $\mathrm{sgn}(\hat{u}) = \mathrm{sgn}(\tilde{u})$), that $\hat{u}^2 \equiv \tilde{u}^2$. Noting that, for all $i \in \{1, \ldots, N\}$, $\hat{\gamma}_i \equiv T\sum_{j=1}^{T} \hat{\nu}_{ij}^2 = \sum_{j=1}^{T} \tilde{\nu}_j^2/T =: \tilde{\gamma}$, one has that

$$\hat{u}^2 = \frac{T}{2}\left[\prod_{i=1}^{N} \hat{\gamma}_i\right] \sum_{i=1}^{N} 1 - \hat{\gamma}_i^{-1} = 2N\frac{T(\tilde{\gamma}-1)}{4}\tilde{\gamma}^{N-1},$$

and it remains to show that $T(\tilde{\gamma}-1) \equiv 4\tilde{s}^2 \sum_{j=1}^{T} \lambda_j^2 \tilde{\nu}_j^2$, for which we will need the following two identities, both of which follow from (20). First,

$$T = \sum_{j=1}^{T} 1 = \sum_{j=1}^{T} \tilde{\nu}_j(1 - 2\tilde{s}\lambda_j) = \sum_{j=1}^{T} \tilde{\nu}_j - 2\tilde{s}\sum_{j=1}^{T} \lambda_j\tilde{\nu}_j = \sum_{j=1}^{T} \tilde{\nu}_j.$$

Secondly, by a similar argument,

$$0 = \sum_{j=1}^{T} \lambda_j\tilde{\nu}_j = \sum_{j=1}^{T} \lambda_j\tilde{\nu}_j^2(1 - 2\tilde{s}\lambda_j) = \sum_{j=1}^{T} \lambda_j\tilde{\nu}_j^2 - 2\tilde{s}\sum_{j=1}^{T} \lambda_j^2\tilde{\nu}_j^2.$$

Using these,

$$T(\tilde{\gamma}-1) = \sum_{j=1}^{T} \tilde{\nu}_j^2 - \sum_{j=1}^{T} \tilde{\nu}_j = \sum_{j=1}^{T} \tilde{\nu}_j^2 - \tilde{\nu}_j^2(1 - 2\tilde{s}\lambda_j) = 2\tilde{s}\sum_{j=1}^{T} \lambda_j\tilde{\nu}_j^2 = 4\tilde{s}^2\sum_{j=1}^{T} \lambda_j^2\tilde{\nu}_j^2,$$

which completes the proof.

6 Application

In this section, we demonstrate the accuracy of the saddlepoint approximation by applying it to the stationarity test of Hadri and Larsson (2005). The intuition of the test, which is a generalization of the KPSS test for stationarity (Kwiatkowski et al., 1992) to panel data models, is to decompose the error term, for each individual, into a white noise component and a random walk. The null hypothesis is that the variance of the innovation sequence of the random walk is zero.

The results of this paper allow us to obtain the finite sample null distribution of the test, thus generalizing the results of Hornok and Larsson (2000), who considered the pure time-series case, i.e., the KPSS test. We consider Hadri and Larsson's Model (2), where under the alternative, some series are stationary around incidental trends. In terms of model (1), $\mathbf{X}_i = [\mathbf{1}_{T_i}\ \mathbf{t}_{T_i}]$, where $\mathbf{1}_{T_i}$ is a $T_i \times 1$ vector of ones and $\mathbf{t}_{T_i}$ is a $T_i \times 1$ vector of consecutive natural numbers starting at 1,

$$\sigma_i^2 \mathbf{\Sigma}_i(\theta) = \sigma_i^2 \left[\theta \mathbf{F}_{T_i} + \mathbf{I}_{T_i}\right],$$

and the (j,k)th element of $\mathbf{F}_{T_i}$ is $\min(j,k)$. It is immediate that

$$\dot{\mathbf{\Sigma}}_i(0) = \mathbf{F}_{T_i},$$

so that the locally best test of $H_0 : \theta = 0$ versus $H_a : \theta > 0$, invariant to transformations of the form (3), is the one which rejects for large values of

$$\tau(\mathbf{y}) = \sum_{i=1}^{N} \tau_i(\mathbf{y}_i) \equiv \sum_{i=1}^{N} (T_i - 2) \frac{\mathbf{y}_i' \mathbf{M}_i \mathbf{F}_{T_i} \mathbf{M}_i \mathbf{y}_i}{\mathbf{y}_i' \mathbf{M}_i \mathbf{y}_i}.$$

In order to ensure that its distribution converge to a standard Gaussian under the null, Hadri and Larsson define their test statistic in the slightly modified form

$$Z_{\tau NT} \equiv \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{\tau_i(\mathbf{y}_i) - (T_i^2 - 4)/15}{\sqrt{(T_i + 2)(T_i - 2)^2 (13T_i^2 + 23)/(2100T_i) - (T_i^2 - 4)^2/225}}.$$

Clearly, if $T_i = T \forall i$, then the tests based on $\tau(\mathbf{y})$ and on $Z_{\tau NT}$ are equivalent; if, however, the time series dimensions differ across individuals, then the latter is no longer locally best, but only approximately so. The intuition behind this is as follows: suppose the panel consists of $N - 1$ individuals of time series dimension T_1, and one — individual N, say — with time series dimension T_N, much larger than T_1. Suppose that the Nth series contains a unit root, which is correctly detected by the LBI test. The test based on $Z_{\tau NT}$ places a relatively lower weight on the Nth individual statistic, and, hence, may not detect the nonstationarity.

Figure 1 offers a comparison of the approximate null distributions of the $Z_{\tau NT}$-test obtained from the saddlepoint and normal approximations, respectively. Depicted is the relative error in percent, defined as $100(\hat{F} - F)/\min(F, 1 - F)$, where $\hat{F}$ denotes the respective approximate distribution function, and the exact values F have been computed from (11). For the example considered here ($N = 10$, $T_i = 10 \forall i$), the computation of the 29 values took just over 3 seconds using the saddlepoint approximation, whereas the exact values required more than 15 *minutes*, about 300 times longer. As expected, the normal approximation deteriorates in the tails of the distribution — the crucial part of the support in hypothesis testing —, whereas the relative error of the saddlepoint approximation never exceeds 20%. This is in stark contrast with the Cornish-Fisher expansion explored by Hadri and Larsson, who note on Page 60:

> We tried to improve the empirical size of our two tests particularly for $T < 10$ by using the Fisher-Cornish expansion [...]. However, the improvements were very marginal. This is not surprising as it is well known that the Fisher-Cornish expansion like the Edgeworth expansion from which it is derived deteriorates in the tails.

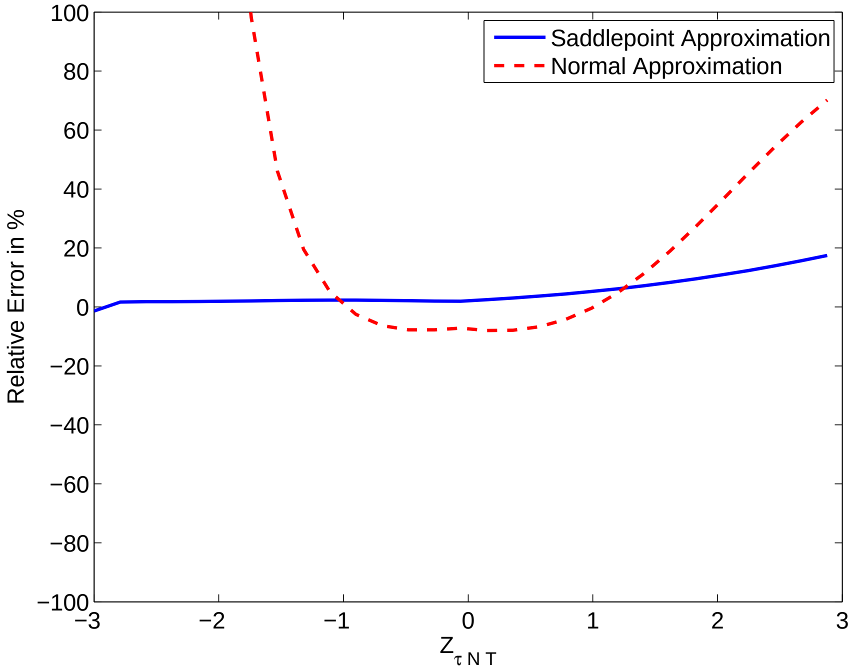

Figure 1: Relative Error in %, defined as $100(\hat{F}(x) - F(x))/\min(F(x), 1 - F(x))$, of approximations for the distribution of $Z_{\tau NT}$, where $N = T = 10$.

The saddlepoint approximation, on the other hand, can be thought of as an Edgeworth expansion applied to the exponentially tilted density, and does not suffer from this deficiency.

7 Conclusions

The locally best invariant test for sphericity in a heterogeneous panel model is given by a weighted sum of the individual tests. While the limiting (as $N \to \infty$) null distribution of the appropriately scaled test statistic is Gaussian, the finite sample distribution is of considerable interest, for the following reasons: Firstly, in macroeconomic panels, the cross-section dimension is typically small, so that use of asymptotic critical values will incur significant size distortions; secondly, even for moderate values of N, the quality of the Normal approximation, for any particular test, is unknown a priori. The results provided herein address both of these problems: the exact integral expression for the distribution function is useful for determining the accuracy of the Normal approximation when devising a test; the saddlepoint approximation, owing to its relative ease of computation, can serve as a routine tool for the computation of p-values. While also approximate in nature, the saddlepoint p-values have relative error, as opposed to the absolute error associated with the Normal approximation, and thus preserve good accuracy even in the extreme tails of the distribution.

References

BALTAGI, B. H. (2005): *Econometric Analysis of Panel Data*, Chichester: John Wiley & Sons. 1

BRODA, S., M. S. PAOLELLA, AND Y. TCHOPOURIAN (2006): "Approximately Exact Inference in Dynamic Panel Models: a QUEST for Unbiasedness," *submitted*. 1, 2, 8

BUTLER, R. AND M. PAOLELLA (2008): "Uniform Saddlepoint Approximations for Ratios of Quadratic Forms," *Bernoulli*, 14, 140–154. 1

BUTLER, R. W. AND M. S. PAOLELLA (1998): "Approximate Distributions for the Various Serial Correlograms," *Bernoulli*, 4, 497–518. 1

CLIFF, A. D. AND J. K. ORD (1973): *Spatial Correlation*, London: Pion. 1

DUFOUR, J. M. AND M. L. KING (1991): "Optimal Invariant Tests for the Autocorrelation Coefficient in Linear Regressions with Stationary or Nonstationary AR(1) Errors," *Journal of Econometrics*, 47, 115–143. 1

FERGUSON, T. S. (1967): *Mathematical Statistics — A Decision Theoretic Approach*, New York: Academic Press. 3

GRAD, A. AND H. SOLOMON (1955): "Distribution of Quadratic Forms and some Applications," *Annals of Mathematical Statistics*, 26, 464–477. 1

HADRI, K. AND R. LARSSON (2005): "Testing for Stationarity in Heterogeneous Panel Data where the Time Dimension is Finite," *The Econometrics Journal*, 8, 55–69. 1, 2, 9, 10

HILLIER, G. (2001): "The Density of a Quadratic Form in a Vector Uniformly Distributed on the n-Sphere," *Econometric Theory*, 17, 1–28. 3

HORNOK, A. AND R. LARSSON (2000): "The Finite Sample Distribution of the KPSS Test," *The Econometrics Journal*, 3, 108–121. 9

IMHOF, J. P. (1961): "Computing the Distribution of Quadratic Forms in Normal Variables," *Biometrika*, 48, 419–26. 1

KARIYA, T. (1980): "Locally Robust Tests for Serial Correlation in Least Squares Regression," *The Annals of Statistics*, 8, 1065–1070. 2

KELKER, D. (1970): "Distribution Theory of Spherical Distribution and a Location Scale Parameter Generalization," *Sankhyā A*, 32, 419–430. 3

KING, M. L. (1980): "Robust Tests for Spherical Symmetry and Their Application to Least Squares Regression," *The Annals of Statistics*, 8, 1265–1271. 2

Kolassa, J. E. (1996): "Higher Order Approximations to Conditional Distribution Functions," *The Annals of Statistics*, 24, 353–364. 6

Kwiatkowski, D., P. C. B. Phillips, P. Schmidt, and Y. Shin (1992): "Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root," *Journal of Econometrics*, 54, 91–115. 1, 9

Lieberman, O. (1994): "Saddlepoint Approximation for the Distribution of a Ratio of Quadratic Forms in Normal Variables," *Journal of the American Statistical Association*, 89, 924–928. 1

Marsh, P. W. N. (1998): "Saddlepoint Approximations for Noncentral Quadratic Forms," *Econometric Theory*, 14, 539–559. 1

Skovgaard, I. M. (1987): "Saddlepoint Expansions for Conditional Distributions," *Journal of Applied Probability*, 24, 875–887. 6

Temme, N. M. (1982): "The Uniform Asymptotic Expansion of a Class of Integrals Related to Cumulative Distribution Functions," *SIAM Journal on Mathematical Analysis*, 13, 239–253. 1