UNIVERSITY OF CAMBRIDGE
Faculty of Economics

Cambridge Working Papers in Economics

Cambridge Working Papers in Economics: 1910

TESTING FOR CORRELATION IN ERROR-COMPONENT MODELS

Koen Jochmans

25 January 2019

This paper concerns linear models for grouped data with group-specific effects. We construct a test for the null of no within-group correlation beyond that induced by the group-specific effect. The approach tests against correlation of arbitrary form while allowing for (conditional) heteroskedasticity. Our setup covers models with exogenous, predetermined, or endogenous regressors. We provide theoretical results on size and power under asymptotics where the number of groups grows but their size is held fixed. In simulation experiments we find good size control and high power in a wide range of designs. We also find that our test is more powerful than the popular test developed by Arellano and Bond (1991), which uses only a subset of the information used by our procedure.

TESTING FOR CORRELATION IN ERROR-COMPONENT MODELS

KOEN JOCHMANS*

UNIVERSITY OF CAMBRIDGE

THIS VERSION: JANUARY 22, 2019

Abstract

This paper concerns linear models for grouped data with group-specific effects. We construct a test for the null of no within-group correlation beyond that induced by the group-specific effect. The approach tests against correlation of arbitrary form while allowing for (conditional) heteroskedasticity. Our setup covers models with exogenous, predetermined, or endogenous regressors. We provide theoretical results on size and power under asymptotics where the number of groups grows but their size is held fixed. In simulation experiments we find good size control and high power in a wide range of designs. We also find that our test is more powerful than the popular test developed by Arellano and Bond (1991), which uses only a subset of the information used by our procedure.

Keywords: analysis of variance, clustered standard errors, error components, fixed effects, heteroskedasticity, within-group correlation, Portmanteau test, short panel data.

JEL classification: C12, C23.

*Address: University of Cambridge, Faculty of Economics, Austin Robinson Building, Sidgwick Avenue, Cambridge CB3 9DD, U.K. E-mail: kj345@cam.ac.uk. Support of the European Research Council through grant n^{o} 715787 (MiMo) is gratefully acknowledged.

1 Introduction

The standard linear model for stratified observations on many small independent groups is

$$y_{g,i} = \boldsymbol{x}'_{g,i}\boldsymbol{\beta} + u_{g,i}, \qquad g = 1, \ldots, G, \qquad i = 1, \ldots, N.$$

The errors for a given group g are likely to be correlated. A standard approach is to model such dependence through the error-component model

$$u_{g,i} = \alpha_g + \varepsilon_{g,i},$$

where α_g is a group-specific effect and the errors $\varepsilon_{g,i}$ are assumed uncorrelated within each group (see, e.g., Moulton 1986). This formulation restricts the pairwise within-group correlation to be constant, a restriction that is seldomly tested. There are several formal tests for the presence of a group effect (King and Evans 1986, Moulton and Randolph 1989, Akritas and Arnold 2000, Akritas and Papadatos 2004). These tests all break down when the $\varepsilon_{g,i}$ are correlated within groups, however. Our goal here is to develop a test of the null of no within-group correlation beyond that induced by the group-specific effect. Aside from a specification test for the error-component specification, the test can also serve to evaluate whether cluster-robust standard errors (Liang and Zeger 1986, Arellano 1987) should be used for fixed-effect estimators of $\boldsymbol{\beta}$.

For our purposes it is important to construct a test that has non-trivial power against any alternative. As such, we aim for a Portmanteau test. Inoue and Solon (2006) proposed such a test under the additional assumption that the covariates are strictly exogenous and that the errors are homoskedastic.[1] The approach proposed here allows for unspecified forms of (conditional) heteroskedasticity and only requires the estimator of $\boldsymbol{\beta}$ used to be asymptotically linear under the null. As such, it can be applied to models with exogenous, predetermined, or endogenous regressors (provided, of course, that suitable instrumental variables are available).

[1] Under these conditions several tests against specific alternatives—typically (first-order) autoregressive errors—have also been proposed. Born and Breitung (2016) provide references, discussion, and several refinements.

The test statistic we construct uses (estimators of) all linearly-independent differences between pairwise within-group covariances. Linear combinations of a subset of the moment conditions underlying our test statistic yield the m_n-statistics of Arellano and Bond (1991), which can be used to test against non-zero nth-order autocorrelation in the first-differenced errors $u_{g,i} - u_{g,i-1}$; see Yamagata (2008) for a joint test. Because first-differencing introduces first-order autocorrelation also under the null such a test can only be constructed for $n \geq 2$. Furthermore, at least $n + 2$ observations per group are needed to construct a meaningfull m_n-statistic. Hence, a four-wave panel is needed to construct the m_2-statistic. In contrast, our test can be applied as soon as three observations per group are available. Observe that an error-component model for two-wave data will always satisfy the null of no remaining within-group correlation; see below.

In the next section we formally introduce our test and presents its asymptotic properties as the number of groups, G, grows large but their size, N, is held fixed. We then present the results from various simulation experiments to assess the size and power of the test for realistic sample sizes. The test is found to be near size-correct and powerful against common deviations from the null. In our simulations the power of our test is also uniformly larger than the power of the m_n-test. All proofs are collected in the appendix to the paper.

2 Testing for within-group correlation

We initially consider the error-component model

$$u_{g,i} = \alpha_g + \varepsilon_{g,i}, \qquad g = 1, \ldots, G, \qquad i = 1, \ldots, N, \tag{2.1}$$

where $u_{g,i}$ is directly observable. Later we will replace $u_{g,i}$ by a suitable estimator. In (2.1), α_g represents a group-specific unobserved effect while $\varepsilon_{g,i}$ is a latent idiosyncratic error that varies both across and within groups. The standard error-component formulation assumes that all variables are independent and identically distributed, both across and within groups (as in Arellano 2003, Chapter 3). We will maintain this assumption across groups but will only impose $E(\varepsilon_{g,i}|\alpha_g) = 0$ for each group.[2] Our aim is to test the (multiple) null hypothesis

$$E(\varepsilon_{g,i_1}\varepsilon_{g,i_2}) = 0 \text{ for all } i_1 \neq i_2, \tag{2.2}$$

which states that there is no within-group correlation beyond the correlation induced by the group-specific effect.

The presence of α_g implies that a test of (2.2) based on covariances of the levels of $u_{g,i}$ will not be suitable. However, when iterating expectations using $E(\varepsilon_{g,i}|\alpha_g) = 0$ we see that

$$E(u_{g,i_1}(u_{g,i_2} - u_{g,i_3})) = E(\varepsilon_{g,i_1}(\varepsilon_{g,i_2} - \varepsilon_{g,i_3})) = E(\varepsilon_{g,i_1}\varepsilon_{g,i_2}) - E(\varepsilon_{g,i_1}\varepsilon_{g,i_3}).$$

For any $i_1 \neq i_2 \neq i_3$ this is the difference between two covariances. There are $N(N-1)/2$ different covariances and, hence,

$$r := \frac{N(N-1)}{2} - 1 = \frac{(N+1)(N-2)}{2}$$

unique such differences. These differences are all zero if and only if $E(\varepsilon_{g,i_1}\varepsilon_{g,i_2}) = \sigma^2$ for all $i_1 \neq i_2$ and some constant σ^2. In this case, $\varepsilon_{g,i}$ itself follows an error-component model, say $\varepsilon_{g,i} = \eta_g + \epsilon_{g,i}$ with $\operatorname{var}(\eta_g) = \sigma^2$ and $\operatorname{cov}(\epsilon_{g,i_1}, \epsilon_{g,i_2}) = 0$ for $i_1 \neq i_2$. This is observationally equivalent to (2.1) under the null with the group-specific effect redefined as $\alpha_g + \eta_g$. Consequently, we may set $\sigma^2 = 0$ without loss of generality. It follows that testing (2.2) is equivalent to jointly testing that all r differences are equal to zero. A convenient way to re-write the null is as follows. Let Δ be the first-differencing operator, so $\Delta u_{g,i} = u_{g,i} - u_{g,i-1}$, and let $\boldsymbol{u}_{g,i} := (u_{g,1}, \ldots, u_{g,i-2}, u_{g,i+1})'$.[3] Then testing (2.1) is equivalent to testing

$$E(\boldsymbol{u}_{g,i}\Delta u_{g,i}) = \mathbf{0} \text{ for all } 1 < i \leq N. \tag{2.3}$$

[2] Random sampling at the group level can be relaxed. It suffices to assume that the $u_{g,i}$ are independent but not identically distributed across groups for our approach to go through—under suitable strengthening of the assumptions required for a law of large numbers and central limit theorem to apply. We refrain from such a sampling assumption here for ease of exposition.

[3] The null (2.2) is equivalent to the set of moment conditions $E(u_{g,i_1}(u_{g,i_2} - u_{g,i_3})) = 0$, for $i_1 \neq i_2 \neq i_3$ but only r of these are linearly independent. Our formulation in (2.3) is not the only way of selecting r such moments but is notationaly convenient. Note that any other way would yield (numerically) the same test statistic.

This approach delivers testable moments as soon as more than two observations per group are available.[4]

Observe that moments of the form

$$E(\Delta u_{g,i}\Delta u_{g,i-n}) = 0 \text{ for } 1 < n \leq i-2,$$

are linear combinations of a subset of those in (2.3). These are nth-order autocovariances of $\Delta\varepsilon_{g,i}$. Arellano and Bond (1991) suggested testing for nth-order autocorrelation by evaluating whether the corresponding sample moment can be considered large relative to its standard error. The resulting test statistic is know as the m_n-statistic. Yamagata (2008) proposed to combine all available such m_n-statistics into a single test procedure. Notice that, as first-differencing introduces autocorrelation of order one, an m_n-statistic can only be used for $n \geq 2$. Furthermore, m_2-statistic requires $N \geq 4$; more generally, the m_n-statistic is available if $N \geq n+2$.

Moving on, to state our test statistic compactly it is useful to introduce the $r \times 1$ vector

$$\boldsymbol{v}_g := (\boldsymbol{v}'_{g,2}, \ldots, \boldsymbol{v}'_{g,N})', \qquad \boldsymbol{v}_{g,i} := \boldsymbol{u}_{g,i}\Delta u_{g,i}.$$

The null (2.3) can then be written as the moment condition $E(\boldsymbol{v}_g) = \mathbf{0}$. The corresponding sample moment is $\sum_{g=1}^G \boldsymbol{v}_g$ and the sample variance is $\sum_{g=1}^G \boldsymbol{v}_g\boldsymbol{v}'_g$. Our test statistic is the quadratic form

$$s_G := \left(\sum_{g=1}^G \boldsymbol{v}_g\right)' \left(\sum_{g=1}^G \boldsymbol{v}_g\boldsymbol{v}'_g\right)^{-1} \left(\sum_{g=1}^G \boldsymbol{v}_g\right),$$

and its large-sample behavior, as the number of groups G grows, is summarized in Theorem 1 below.

[4] An alternative way to arrive at (2.3) is by noting that

$$E(u_{g,i_1}u_{g,i_2}) = E(\alpha_g^2) + E(\varepsilon_{g,i_1}\varepsilon_{g,i_2}).$$

Because the distribution of α_g is left unrestricted this equation, in itself, is not of direct use. However, the panel dimension allows to difference-out the second moment of the group-specific effect, yielding differences of the form

$$E(u_{g,i_1}\Delta u_{g,i_2}) = E(\varepsilon_{g,i_1}\Delta\varepsilon_{g,i_2}),$$

which lead to (2.3).

In the theorem we consider sequences of local alternatives where

$$E(\varepsilon_{g,i_1}\varepsilon_{g,i_2}) = \frac{\sigma_{i_1,i_2}}{\sqrt{G}} \tag{2.4}$$

and σ_{i_1,i_2} is non-zero for at least one pair of indices $i_1 \neq i_2$. We write the resulting Pitman drift in the moment condition as $E(\boldsymbol{v}_g) = \boldsymbol{\sigma}/\sqrt{G}$ where the vector $\boldsymbol{\sigma}$ collects all the relevant differences of the form $\sigma_{i_1,i_2} - \sigma_{i_1,i_2-1}$ according to our specification of $\boldsymbol{v}_g$ above. We let $\chi^2(n;m)$ denote the non-central χ^2-distribution with n degrees of freedom and non-centrality parameter m.

Theorem 1. *Suppose that $E(\alpha_g^4) < \infty$, $E(\varepsilon_{g,i}^4) < \infty$, and that $\boldsymbol{V} := E(\boldsymbol{v}_g\boldsymbol{v}_g')$ has maximal rank r.*

(i) If the null (2.3) holds $s_G \xrightarrow{d} \chi^2(r,0)$.

(ii) Under a sequence of local alternatives as in (2.4) $s_G \xrightarrow{d} \chi^2(r,\boldsymbol{\sigma}'\boldsymbol{V}^{-1}\boldsymbol{\sigma})$.

The result implies that a test that has size $\alpha \in (0,1)$ in large samples can be constructed by comparing s_G to the $(1-\alpha)$th quantile of the $\chi^2(r,0)$ distribution, rejecting the null if the statistic is larger than the quantile in question. Such a test is asymptotically unbiased, consistent against any fixed alternative, and has non-trivial asymptotic power against any Pitman sequence.

We now generalize (2.1) to

$$y_{g,i} = \boldsymbol{x}_{g,i}'\boldsymbol{\beta} + u_{g,i}, \qquad u_{g,i} = \alpha_g + \varepsilon_{g,i},$$

where $y_{g,i}$ and $\boldsymbol{x}_{g,i}$ are an observable outcome and vector of covariates, respectively, and $u_{g,i}$ is now the latent error term. Suppose that an estimator $\hat{\boldsymbol{\beta}}$ of the coefficient vector $\boldsymbol{\beta}$ is available. Then we may use the residuals

$$\hat{u}_{g,i} := y_{g,i} - \boldsymbol{x}_{g,i}'\hat{\boldsymbol{\beta}}$$

as estimators of the $u_{g,i}$ and construct a test statistic for our null based on these residuals.

Compared to the error-component formulation the test statistic needs to be modified slightly to take into account the sampling noise in $\hat{\boldsymbol{\beta}}$. To do so we impose the requirement that $\hat{\boldsymbol{\beta}}$ is asymptotically linear under the null, i.e., that

$$\sqrt{G}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \frac{1}{\sqrt{G}} \sum_{g=1}^{G} \boldsymbol{\omega}_g + o_p(1) \tag{2.5}$$

for a random variable $\boldsymbol{\omega}_g$ that has zero mean and finite variance. This is a very mild requirement as all common estimators satisfy this condition (of course, under suitable regularity conditions). When the covariates are strictly exogenous, for example, an obvious estimator of $\boldsymbol{\beta}$ would be within-group least squares. In that case, $\boldsymbol{\omega}_{g,i} = \boldsymbol{Q}^{-1} \sum_{i=1}^{N} (\tilde{\boldsymbol{x}}_{g,i}\, u_{g,i})$ for $\boldsymbol{Q} := \sum_{i=1}^{N} E(\tilde{\boldsymbol{x}}_{g,i} \tilde{\boldsymbol{x}}'_{g,i})$ and $\tilde{\boldsymbol{x}}_{g,i} := \boldsymbol{x}_{g,i} - \overline{\boldsymbol{x}}_g$, where $\overline{\boldsymbol{x}}_g$ denotes the within-group mean of the covariates. This estimator is robust to within-group correlation and, hence, remains asymptotically linear when our null is false. On the other hand, when the covariates are merely pre-determined, instrumental-variable estimators of the form in Holtz-Eakin, Newey and Rosen (1988), which are based on moment conditions of the form $E(\boldsymbol{z}_{g,i} \Delta u_{g,i}) = \boldsymbol{0}$ for $\boldsymbol{z}_{g,i} := (\boldsymbol{x}'_{g,i-2}, \ldots, \boldsymbol{x}'_{g,1})'$ (or a subvector thereof) will generally break down when the errors are correlated.

If we let $\hat{\boldsymbol{v}}_g$ denote the plug-in estimator of $\boldsymbol{v}_g$ using the residuals then, under the regularity conditions stated in the theorem below, we have that

$$\sum_{g=1}^{G} \hat{\boldsymbol{v}}_g = \sum_{g=1}^{G} (\boldsymbol{v}_g - \boldsymbol{\Omega} \boldsymbol{\omega}_g) + o_p(\sqrt{G}),$$

where $\boldsymbol{\Omega} := (\boldsymbol{\Omega}'_2, \ldots, \boldsymbol{\Omega}'_N)'$ for $\boldsymbol{\Omega}_i := E(\boldsymbol{u}_{g,i} \Delta \boldsymbol{x}'_{g,i}) + E(\boldsymbol{X}'_{g,i} \Delta u_{g,i})$ and we have introduced the matrix $\boldsymbol{X}_{g,i} := (\boldsymbol{x}_{g,1}, \ldots, \boldsymbol{x}_{g,i-2}, \boldsymbol{x}_{g,i+1})$. Our test statistic in the presence of covariates can then be written as

$$\dot{s}_G := \left(\sum_{g=1}^{G} \hat{\boldsymbol{v}}_g \right)' \left(\sum_{g=1}^{G} \dot{\boldsymbol{v}}_g \dot{\boldsymbol{v}}'_g \right)^{-1} \left(\sum_{g=1}^{G} \hat{\boldsymbol{v}}_g \right),$$

for $\dot{\boldsymbol{v}}_g := \hat{\boldsymbol{v}}_g - \hat{\boldsymbol{\Omega}} \hat{\boldsymbol{\omega}}_g$, where $\hat{\boldsymbol{\Omega}}$ denotes the plug-in estimator of $\boldsymbol{\Omega}$ and $\hat{\boldsymbol{\omega}}_g$ is an estimator of the influence function of $\hat{\boldsymbol{\beta}}$. The latter estimator depends on the problem at hand. For the within-group least-squares estimator, for example, we have $\hat{\boldsymbol{\omega}}_g = \hat{\boldsymbol{Q}}^{-1} \sum_{i=1}^{N} (\tilde{\boldsymbol{x}}_{g,i} \hat{u}_{g,i})$, where $\hat{\boldsymbol{Q}} := G^{-1} \sum_{g=1}^{G} \sum_{i=1}^{N} (\tilde{\boldsymbol{x}}_{g,i} \tilde{\boldsymbol{x}}'_{g,i})$.

Theorem 2 summarizes the large-sample properties of the test based on $\dot{s}_G$. Note that, when the estimator of $\boldsymbol{\beta}$ is not robust to violation of our null, its influence function under the Pitman drift in (2.4) will have

$$E(\boldsymbol{\omega}_g) = \frac{\boldsymbol{b}_{\boldsymbol{\sigma}}}{\sqrt{G}}$$

for some non-zero vector $\boldsymbol{b}_{\boldsymbol{\sigma}}$. The exact form of $\boldsymbol{b}_{\boldsymbol{\sigma}}$ will depend both on the estimator used and the alternative hypothesis under consideration. Some calculations and discussion are provided in the next section. We let $\|\cdot\|$ denote both the Euclidean norm and the Frobenius norm.

Theorem 2. *Suppose that $E(\alpha_g^4) < \infty$, $E(\varepsilon_{g,i}^4) < \infty$, and $E(\|\boldsymbol{x}_{g,i}\|^4) < \infty$, that (2.5) holds and that $G^{-1} \sum_{g=1}^{G} \|\hat{\boldsymbol{\omega}}_g - \boldsymbol{\omega}_g\|^2 = o_p(1)$, and that $\dot{\boldsymbol{V}} := E((\boldsymbol{v}_g - \boldsymbol{\Omega}\boldsymbol{\omega}_g)(\boldsymbol{v}_g - \boldsymbol{\Omega}\boldsymbol{\omega}_g)')$ has maximal rank r.*

(i) If the null (2.3) holds $\dot{s}_G \xrightarrow{d} \chi^2(r, 0)$.

(ii) Under a sequence of local alternatives as in (2.4) $\dot{s}_G \xrightarrow{d} \chi^2(r, \dot{\boldsymbol{\sigma}}' \dot{\boldsymbol{V}}^{-1} \dot{\boldsymbol{\sigma}})$ where we let $\dot{\boldsymbol{\sigma}} := \boldsymbol{\sigma} - \boldsymbol{\Omega}\boldsymbol{\beta}$.

Theorem 2 differs from Theorem 1 only in the local-power result. Estimation noise in $\hat{\boldsymbol{\beta}}$ changes the weight matrix in the non-centrality parameter in a way that is independent of the alternative under consideration. Local power will be further affected if $\hat{\boldsymbol{\beta}}$ suffers from asymptotic bias under the alternative. The extent to which both channels matter depends on how sensitive the moment in (2.3) are to changes in $\boldsymbol{\beta}$. This is measured by the Jacobian matrix $\boldsymbol{\Omega}$. Consequently, a sufficient condition for the estimation of $\boldsymbol{\beta}$ to have no (asymptotic) impact on our test is that $\boldsymbol{\Omega}$ is equal to the zero matrix. This would happen, for example, when the covariates are strictly exogenous and α_g and the $\boldsymbol{x}_{g,i}$ are all uncorrelated.

3 Simulations

We next present results from numerical experiments on the size and power of our test. We consider a model without regressors, i.e., $y_{g,i} = \alpha_g + \varepsilon_{g,i}$. We generate data by setting $\alpha_g = 0$ for all g and draw the errors from a multivariate normal distribution with mean zero and covariance matrix $\boldsymbol{\Sigma}$. We provide results for three configurations of this covariance matrix. In the first configuration all observations have unit variance, the first two observations have correlation ρ, and all other pairwise correlations are zero. In the second configuration the errors follow a (stationary) first-order autoregressive process $\varepsilon_{g,i} = \rho\, \varepsilon_{g,i-1} + \eta_{g,i}$ for independent standard-normal innovations $\eta_{g,i}$. In the third configuration the errors follow a first-order moving-average process $\varepsilon_{g,i} = \eta_{g,i} + \theta_\rho \eta_{g,i-1}$, where the innovations are again independent standard normal. Here, the parameter θ_ρ can be parameterized in terms of the (first-order) autocorrelation coefficient ρ.[5] All configurations depend on a single correlation parameter, ρ, and all yield an identity matrix as covariance matrix under the null, i.e., when $\rho = 0$.

Figure 1 provides histograms of the sampling distribution of our test statistic under the null for $G \in \{100, 250, 500\}$ groups of size three (upper panels) and four (lower panels). Each histogram is accompanied by its limit distribution. The figure reveals our asymptotic approximation under the null to be quite accurate even for a relatively small number of groups.

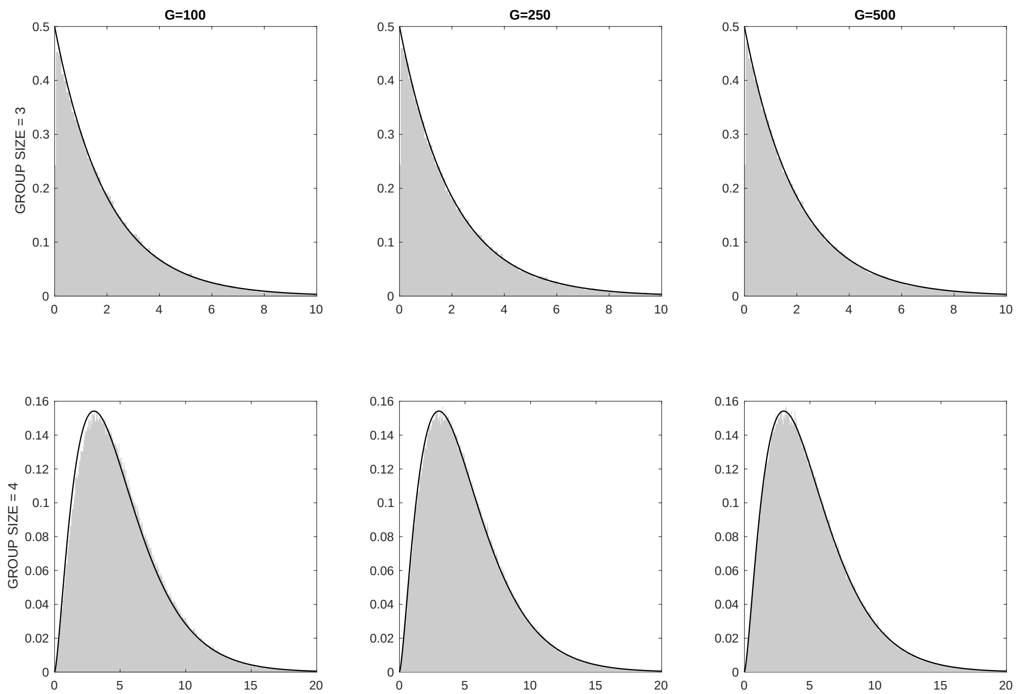
The upper panels in Figure 3 plot the power of our test against the three alternatives discussed above (as a function of ρ) for $G = 100$ (solid line), $G = 250$ (dashed line), and $G = 500$ (dashed-dotted line) groups of size three. This is the smallest number of

[5]For a given parameter θ the variance and first-order autocovariance of the moving-average process are $(1+\theta^2)$ and θ, respectively. The first-order autocorrelation of the process then is $\rho := \theta/(1+\theta^2)$. Provided that $\rho \neq 0$, inversion gives

$$\theta = \frac{1 \pm \sqrt{1 - \frac{\rho^2}{4}}}{2\rho}$$

which is well defined for all $-\frac{1}{2} < \rho < \frac{1}{2}$. If $\rho = 0$ then $\theta = 0$ and vice versa. We generate our data using the largest of the two roots. The results are invariant to this choice.

Figure 1: Sampling distributions under the null



Sampling distributions computed over 100,000 Monte Carlo replications. Reference distribution is χ^2_2 in the upper panels and χ^2_5 in lower panels.

within-group observations for which any test that allows for group-specific effects can hope to have non-trivial power. Each panel also marks the significance level of the test of .05 (horizontal dotted line). The test does well in rejecting the null when it is false, with rejection frequencies moving up quite rapidely as the parameter ρ moves away from zero. The power functions are monotone increasing except in the autoregressive case. There, the test has some difficulty against near unit-root alternatives. This is most noticeable when $G = 100$.

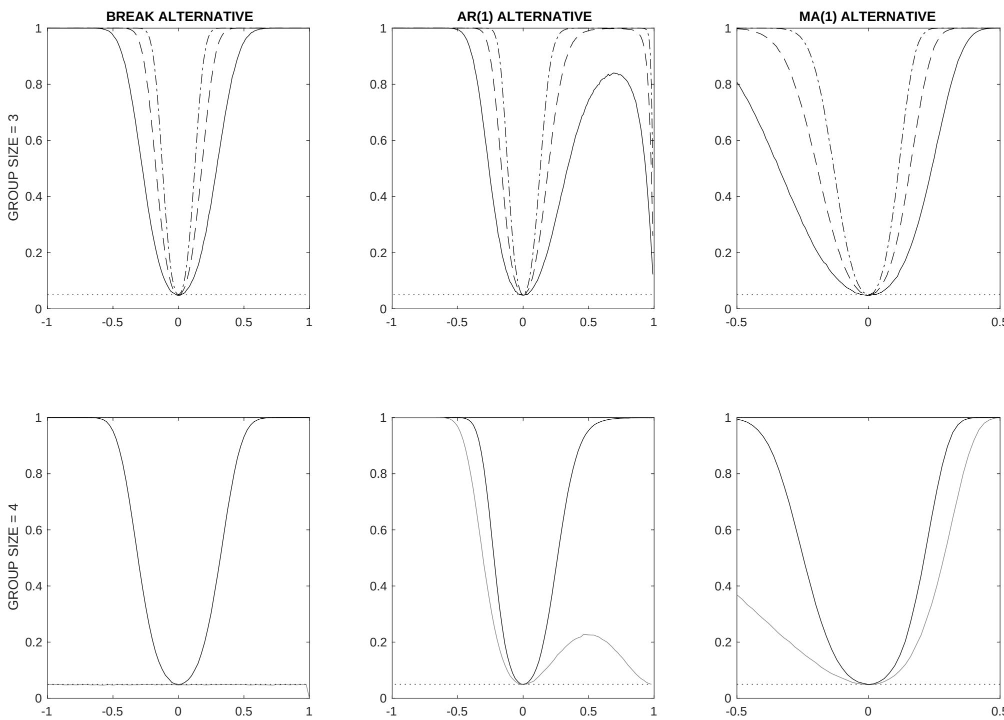
The lower panels in Figure 3 illustrate power for $G = 100$ groups of size four. Here we can compare our approach (black solid line) to the test of Yamagata (2008) (grey solid line). The latter approach is a test of the single moment condition $E[\Delta\varepsilon_{g,4}\Delta\varepsilon_{g,2}] = 0$ and, hence, co-incides with the m_2-test developed by Arellano and Bond (1991). The m_2-test has no power against any alternative in our first configuration; its power curve is flat. This is not surprising as the moment condition being tested continues to hold under all alternatives in this configuration. The m_2-test does have power against autoregressive and moving-average alternatives. However, it is uniformly (in ρ) less powerful than our test in both cases and the power curves are highly asymmetric around zero. It has great difficulty detecting autoregressive correlation patterns when the autocorrelation coefficient is positive and also struggles with moving-average errors when their (first-order) autocorrelation is negative. Our test has high power against all these alternatives. Also note that, now, our test does not suffer from power loss against near unit-root scenarios, even with data on only 100 groups.

To illustrate Theorem 2 we present some further simulation results for the dynamic specification

$$y_{g,i} = y_{g,i-1}\,\beta + u_{g,i},$$

where β is estimated by an instrumental-variable estimator that uses suitably-lagged levels as instruments for the equation in first differences. Such estimators are given in Anderson and Hsiao (1981), Holtz-Eakin, Newey and Rosen (1988), and Arellano and Bond (1991)

Figure 2: Power functions



Power functions computed over 100,000 Monte Carlo replications. Upper panels: power (as a function of ρ) for $G = 100$ (solid line), $G = 250$ (dashed line), and $G = 500$ (dotted line). Lower panels: power of our test for $G = 100$ (solid black line) and of the m_2-test (solid grey line). Horizontal dotted lines mark the significance level of the test.

and are based on sequential moment conditions of the form

$$E(y_{g,i-j}\Delta u_{g,i}) = 0, \qquad \text{for all } 1 < j \leq i \text{ and } 1 < i \leq N. \tag{3.6}$$

The validity of these moments comes from an assumption of no within-group correlation beyond a certain lag. Hence, such estimators will be inconsistent under fixed alternatives and asymptotically biased under Pitman sequences. To see this, consider (stationary) autoregressive alternatives of the form

$$\varepsilon_{g,i} = \rho\,\varepsilon_{g,i-1} + \eta_{g,i}$$

for $\eta_{g,i} \sim (0, \sigma^2)$ white noise. Using backward substitution gives $y_{g,i} = \frac{\alpha_g}{1-\beta} + \sum_{j=0}^{\infty} \beta^j \varepsilon_{g,i-j}$ and so, for any $j > 1$,

$$E(y_{g,i-j}\Delta u_{g,i}) = -\rho^{j-1} \frac{\sigma^2}{(1+\rho)(1-\rho\beta)}$$

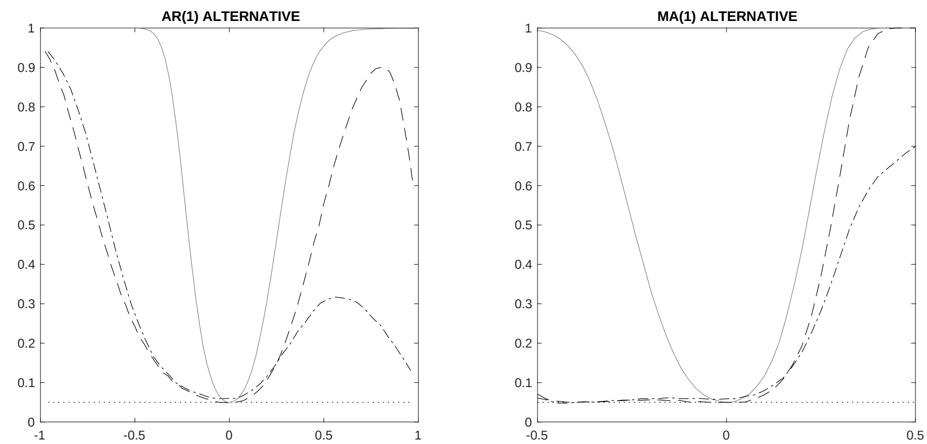
follows from standard results on autoregressive processes. On the other hand, if we consider moving-average alternatives like

$$\varepsilon_{g,i} = \eta_{g,i} + \theta_\rho \eta_{g,i-1},$$

again for $\eta_{g,i} \sim (0, \sigma^2)$ white noise, we have $E(y_{g,i-2}\Delta u_{g,i}) = -\theta_\rho\,\sigma^2$ but $E(y_{g,i-j}\Delta u_{g,i}) = 0$ for all $j > 2$. As instrumental-variable estimators based on the moments in (3.6) are linear, their asymptotic bias, $\boldsymbol{b_\sigma}$, is a linear transformation of the bias in said moments just derived.

We applied our test to the autoregressive specification against the autoregressive and moving-average alternatives. We generated the short time series with $\beta = \frac{1}{2}$ and initialized each processes by drawing from its steady-state distribution. Figure 3 contains the power plots for each of the alternatives for samples on $G = 100$ groups of size four. To illustrate the impact of the first-step estimator we present results for our test when β is estimated by the Arellano and Bond (1991) estimator (dashed line) and by the Anderson and Hsiao (1981) estimator (dashed-dotted line); see Arellano (2003, Chapter 6). The former uses all lagged levels of the outcome variable that are valid under the null of no correlation.

Figure 3: Power functions for the dynamic specification



Power functions computed over 100,000 Monte Carlo replications. Each plot contains the power function based on the Arellano-Bond estimator (dashed line), the Anderson-Hsiao estimator (dashed-dotted line), and the oracle that knowns $\boldsymbol{\beta}$ (solid grey line). Horizontal dotted lines mark the significance level of the test.

The latter estimator only uses the first such lag and, consequently, is inefficient relative to the former. Each plot also contains the power of the (now infeasible) test that presumes knowledge of $\boldsymbol{\beta}$ (solid grey line). These power curves co-incide with those in the lower panels of Figure 2 and serve merely as a comparison.

The results show power loss relative to the oracle case. The difference in the power curves for the two feasible test statistics further highlights the dependence of power on the first-step estimator used. As discussed below Theorem 2, power is affected both by the variance and the bias in the estimator. In our simulations, using either the Arellano and Bond (1991) estimator or the Anderson and Hsiao (1981) estimator creates large power differences primarily when the errors have positive autocorrelation. On the other hand, the results also show that our test has non-trivial power even in cases where $\boldsymbol{\beta}$ is not identified. Indeed, with autoregressive errors, none of the moment conditions in (3.6) are valid away from the null.

Appendix

Proof of Theorem 1 The proof is standard. Consider first the limit result under the null. The moment conditions stated in the theorem imply that

$$\frac{1}{\sqrt{G}}\sum_{g=1}^{G}\boldsymbol{v}_g \xrightarrow{d} N(\mathbf{0},\boldsymbol{V}), \qquad \text{and} \qquad \frac{1}{G}\sum_{g=1}^{G}\boldsymbol{v}_g\boldsymbol{v}_g' \xrightarrow{p} \boldsymbol{V}.$$

Hence,

$$\boldsymbol{z} := \left(\sum_{g=1}^{G}\boldsymbol{v}_g\boldsymbol{v}_g'\right)^{-1/2}\sum_{g=1}^{G}\boldsymbol{v}_g \xrightarrow{d} N(\mathbf{0},\boldsymbol{I}_r),$$

where, here and later, $\boldsymbol{I}_r$ denotes the $r\times r$ identity matrix. Thus, $\boldsymbol{z}'\boldsymbol{z} \xrightarrow{d} \chi^2(r,0)$ follows.

Under the sequence of local alternatives

$$E(\boldsymbol{v}_g) = \frac{\boldsymbol{\sigma}}{\sqrt{G}},$$

the plug-in estimator of $\boldsymbol{V}$ is still consistent but we have an asymptotic-bias term, i.e, now

$$\boldsymbol{z} := \left(\sum_{g=1}^{G}\boldsymbol{v}_g\boldsymbol{v}_g'\right)^{-1/2}\sum_{g=1}^{G}\boldsymbol{v}_g \xrightarrow{d} N(\boldsymbol{V}^{-1/2}\boldsymbol{\sigma},\boldsymbol{I}_r),$$

and so $\boldsymbol{z}'\boldsymbol{z} \xrightarrow{d} \chi^2(r,\boldsymbol{\sigma}'\boldsymbol{V}^{-1}\boldsymbol{\sigma})$ follows. □

Proof of Theorem 2 The main difference with the proof of Theorem 1 is accounting for the estimation noise in $\hat{\boldsymbol{\beta}}$. Because

$$\hat{u}_{g,i} = y_{g,i} - \boldsymbol{x}_{g,i}'\hat{\boldsymbol{\beta}} = u_{g,i} - \boldsymbol{x}_{g,i}'(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}),$$

$\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\|^2 = O_p(G^{-1})$ by (2.5), and the covariates have finite second moments the expansion

$$\hat{\boldsymbol{v}}_{g,i} = \hat{\boldsymbol{u}}_{g,i}\Delta\hat{u}_{g,i} = \boldsymbol{v}_{g,i} - \boldsymbol{A}_{g,i}\,(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}) + O_p(G^{-1})$$

holds with

$$\boldsymbol{A}_{g,i} := \boldsymbol{u}_{g,i}\Delta\boldsymbol{x}_{g,i}' + \boldsymbol{X}_{g,i}'\Delta u_{g,i}.$$

Further, our moment assumptions equally allow the application of a law of large numbers to establish that

$$\frac{1}{G}\sum_{g=1}^{G}\boldsymbol{A}_{g,i} \xrightarrow{p} \boldsymbol{\Omega}_i, \qquad 1 < i \leq N.$$

Put together with the influence-function representation of $\hat{\boldsymbol{\beta}}$ as stated in (2.5) this yields

$$\frac{1}{G}\sum_{g=1}^{G}\hat{\boldsymbol{v}}_g = \frac{1}{G}\sum_{g=1}^{G}(\boldsymbol{v}_g - \boldsymbol{\Omega}\boldsymbol{\omega}_g) + o_p(G^{-1/2}).$$

Under the null the summand has zero mean and variance $\dot{\boldsymbol{V}}$. Therefore, by a central limit theorem,

$$\boldsymbol{z} := \left(\sum_{g=1}^{G}(\boldsymbol{v}_g - \boldsymbol{\Omega}\boldsymbol{\omega}_g)(\boldsymbol{v}_g - \boldsymbol{\Omega}\boldsymbol{\omega}_g)'\right)^{-1/2}\sum_{g=1}^{G}\hat{\boldsymbol{v}}_g \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{I}_r),$$

and so equally $\boldsymbol{z}'\boldsymbol{z} \to \chi^2(r, 0)$. Now, it remains only to show that instead using the estimator

$$\frac{1}{G}\sum_{g=1}^{G}(\hat{\boldsymbol{v}}_g - \hat{\boldsymbol{\Omega}}\hat{\boldsymbol{\omega}}_g)(\hat{\boldsymbol{v}}_g - \hat{\boldsymbol{\Omega}}\hat{\boldsymbol{\omega}}_g)'$$

in the denominator of the above expression does not affect the limit behavior of the statistic. For this it suffices to show that $\hat{\boldsymbol{\Omega}} \xrightarrow{p} \boldsymbol{\Omega}$ and that

$$\frac{1}{G}\sum_{g=1}^{G}\|\hat{\boldsymbol{v}}_g - \boldsymbol{v}_g\|^2 = o_p(1);$$

recall that our regularity conditions already include that $G^{-1}\sum_{g=1}^{G}\|\hat{\boldsymbol{\omega}}_g - \boldsymbol{\omega}_g\|^2 = o_p(1)$, i.e., the equivalent requirement for the influence function. To see consistency for the Jacobian matrix observe that its ith submatrix is

$$\begin{aligned}\frac{1}{G}\sum_{g=1}^{G}(\hat{\boldsymbol{u}}_{g,i}\Delta\boldsymbol{x}'_{g,i} + \boldsymbol{X}'_{g,i}\Delta\hat{u}_{g,i}) &= \frac{1}{G}\sum_{g=1}^{G}\boldsymbol{A}_{g,i} \\ &\quad - \frac{1}{G}\sum_{g=1}^{G}(\boldsymbol{X}'_{g,i}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})\,\Delta\boldsymbol{x}'_{g,i}) + (\boldsymbol{X}'_{g,i}\,\Delta\boldsymbol{x}'_{g,i}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})).\end{aligned}$$

From above, the first term on the right-hand side converges in probability to $\boldsymbol{\Omega}_i$. The second right-hand side term is $O_p(G^{-1/2})$ by the existence of second-order moments on the covariates and the asymptotic-linearity of $\hat{\boldsymbol{\beta}}$. Hence, $\hat{\boldsymbol{\Omega}} \xrightarrow{p} \boldsymbol{\Omega}$ follows. Next, we have that

$$\frac{1}{G}\sum_{g=1}^{G}\|\hat{\boldsymbol{v}}_g - \boldsymbol{v}_g\|^2 = \frac{1}{G}\sum_{g=1}^{G}\sum_{i=2}^{N}\|\hat{\boldsymbol{v}}_{g,i} - \boldsymbol{v}_{g,i}\|^2 = \frac{1}{G}\sum_{g=1}^{G}\sum_{i=2}^{N}\|\boldsymbol{A}_{g,i}\,(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})\|^2 + o_p(1).$$

Because

$$\frac{1}{G}\sum_{g=1}^{G}\sum_{i=2}^{N}\|\boldsymbol{A}_{g,i}\,(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})\|^2 = (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\left(\frac{1}{G}\sum_{g=1}^{G}\sum_{i=2}^{N}(\boldsymbol{A}'_{g,i}\boldsymbol{A}_{g,i})\right)(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$$

and $\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|^2 = O_p(G^{-1})$ it remains only to verify that the weight matrix in this quadratic form is $O_p(1)$. Using the definition of $\boldsymbol{A}_{g,i}$, this matrix is equal to the sum of the four terms

$$\frac{1}{G}\sum_{g=1}^{G}\sum_{i=2}^{N}(\Delta u_{g,i}\boldsymbol{X}_{g,i}\boldsymbol{X}'_{g,i}\Delta u_{g,i}), \qquad \frac{1}{G}\sum_{g=1}^{G}\sum_{i=2}^{N}(\Delta u_{g,i}\boldsymbol{X}_{g,i}\boldsymbol{u}_{g,i}\Delta\boldsymbol{x}'_{g,i}),$$

$$\frac{1}{G}\sum_{g=1}^{G}\sum_{i=2}^{N}(\Delta\boldsymbol{x}_{g,i}\boldsymbol{u}'_{g,i}\boldsymbol{X}'_{g,i}\Delta u_{g,i}), \qquad \frac{1}{G}\sum_{g=1}^{G}\sum_{i=2}^{N}(\Delta\boldsymbol{x}_{g,i}\boldsymbol{u}'_{g,i}\boldsymbol{u}_{g,i}\Delta\boldsymbol{x}'_{g,i}),$$

and each of these terms is bounded in probability. To see this observe that for the first term

$$E(\|\Delta u_{g,i}\boldsymbol{X}_{g,i}\boldsymbol{X}'_{g,i}\Delta u_{g,i}\|) = E(\Delta u_{g,i}^2\|\boldsymbol{X}_{g,i}\boldsymbol{X}'_{g,i}\|) \leq \sqrt{E(\Delta u_{g,i}^4)}\,\sqrt{E(\|\boldsymbol{X}'_{g,i}\boldsymbol{X}_{g,i}\|^2)} < \infty$$

because $u_{g,i}$ and $\boldsymbol{x}_{g,i}$ have finite fourth-order moments and $\|\boldsymbol{X}'_{g,i}\boldsymbol{X}_{g,i}\|^2 = \sum_i\sum_j(\boldsymbol{x}'_{g,i}\boldsymbol{x}_{g,j})^2$. A law of large numbers then ensures that the sample average is $O_p(1)$. For the fourth term we similarly have that

$$E(\|\Delta\boldsymbol{x}_{g,i}\boldsymbol{u}'_{g,i}\boldsymbol{u}_{g,i}\Delta\boldsymbol{x}'_{g,i}\|) \leq \sqrt{E((\boldsymbol{u}'_{g,i}\boldsymbol{u}_{g,i})^2)}\,\sqrt{E(\|\Delta\boldsymbol{x}_{g,i}\Delta\boldsymbol{x}'_{g,i}\|^2)} < \infty.$$

The second and third term are each others transpose and it suffices to establish the result for the former. Let $S_i = \{1, \ldots, i-2, i+1\}$. Then

$$E(\|\Delta u_{g,i}\boldsymbol{X}_{g,i}\boldsymbol{u}_{g,i}\Delta\boldsymbol{x}'_{g,i}\|) \leq \sum_{j\in S_i} E\left(|u_{g,j}\Delta u_i|\,\|\boldsymbol{x}_{g,j}\Delta\boldsymbol{x}'_{g,i}\|\right)$$

is bounded from above by

$$\sum_{j\in S_i}\sqrt{E(|u_{g,j}\Delta u_i|^2)}\,\sqrt{E(\|\boldsymbol{x}_{g,j}\Delta\boldsymbol{x}'_{g,i}\|^2)} < \infty,$$

where the last equality again follows from our fourth-order moment assumptions. This then yields Theorem 2(i).

Theorem 2(ii) follows in the same way as did Theorem 1(ii) once the asymptotic bias in the estimator of $\boldsymbol{\beta}$ is accounted for. To do so observe that, under a sequence of local alternatives where

$$E(\boldsymbol{v}_g) = \frac{\boldsymbol{\sigma}}{\sqrt{G}}$$

we still have an asymptotic-equivalence result of the form

$$\frac{1}{\sqrt{G}}\sum_{g=1}^{G}\hat{\boldsymbol{v}}_g = \frac{1}{\sqrt{G}}\sum_{g=1}^{G}(\boldsymbol{v}_g - \boldsymbol{\Omega}\boldsymbol{\omega}_g) + o_p(1).$$

Now, however, the summand on the right-hand side is no longer mean zero but, rather, has expectation

$$E(\boldsymbol{v}_g - \boldsymbol{\Omega}\boldsymbol{\omega}_g) = \frac{\boldsymbol{\sigma} - \boldsymbol{\Omega}\boldsymbol{b}_{\boldsymbol{\sigma}}}{\sqrt{G}} = \frac{\dot{\boldsymbol{\sigma}}}{\sqrt{G}}.$$

Accounting for the bias yields $\sum_{g=1}^{G}\hat{\boldsymbol{v}}_g/\sqrt{G} \overset{d}{\to} N(\dot{\boldsymbol{\sigma}}, \dot{\boldsymbol{V}})$. The remainder of the proof is identical to that of Theorem 2(ii). □

References

Akritas, M. G. and S. F. Arnold (2000). Asymptotics for analysis of variance when the number of levels is large. *Journal of the American Statistical Association 95*, 212–226.

Akritas, M. G. and N. Papadatos (2004). Heteroskedastic one-way ANOVA and lack-of-fit tests. *Journal of the American Statistical Association 99*, 368–382.

Anderson, T. W. and C. Hsiao (1981). Estimation of dynamic models with error components. *Journal of the American Statistical Association 76*, 598–606.

Arellano, M. (1987). Computing robust standard errors for within-group estimators. *Oxford Bulletin of Economics and Statistics 49*, 431–434.

Arellano, M. (2003). *Panel Data Econometrics*. Oxford University Press.

Arellano, M. and S. Bond (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies 58*, 277–297.

Born, B. and J. Breitung (2016). Testing for serial correlation in fixed-effects panel data models. *Econometric Reviews 35*, 1290–1316.

Holtz-Eakin, D., W. K. Newey, and H. Rosen (1988). Estimating vector autoregressions with panel data. *Econometrica 56*, 1371–1395.

Inoue, A. and G. Solon (2006). A Portmanteau test for serially correlated errors in fixed effects models. *Econometric Theory 22*, 835–851.

King, M. L. and M. A. Evans (1986). Testing for block effects in regression models based on survey data. *Journal of the American Statistical Association 81*, 677–679.

Liang, K.-Y. and S. L. Zeger (1986). Longitudinal data analysis using generalized linear models. *Biometrika 73*, 13–22.

Moulton, B. R. (1986). Random group effects and the precision of regression estimates. *Journal of Econometrics 32*, 385–397.

Moulton, B. R. and W. C. Randolph (1989). Alternative tests of the error components model. *Econometrica 57*, 685–693.

Yamagata, T. (2008). A joint serial correlation test for linear panel data models. *Journal of Econometrics 146*, 135–145.