TESTING FOR CORRELATION IN ERROR-COMPONENT MODELS

KOEN JOCHMANS*

UNIVERSITY OF CAMBRIDGE

June 10, 2020

Abstract

This paper concerns linear models for grouped data with group-specific effects. We construct a portmanteau test for the null of no within-group correlation beyond that induced by the group-specific effect. The approach allows for heteroskedasticity and is applicable to models with exogenous, predetermined, or endogenous regressors. The test can be implemented as soon as three observations per group are available and is applicable to unbalanced data. A test with such general applicability is not available elsewhere. We provide theoretical results on size and power under asymptotics where the number of groups grows but their size is held fixed. Extensive power comparisons with other tests available in the literature for special cases of our setup reveal that our test compares favorably. In a simulation study we find that, under heteroskedasticity, only our procedure yields a test that is both size-correct and powerful. In a large data set on mothers with multiple births we find that infant birthweight is correlated across children even after controlling for mother fixed effects and a variety of prenatal care factors. This suggests that such a strategy may be inadequate to take care of all confounding factors that correlate with the mother's decision to engage in activities that are detrimental to the infant's health, such as smoking.

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`. Financial support from the European Research Council through grant n° 715787 (MiMo) is gratefully acknowledged. Helmut Farbmacher gave much appreciated feedback on an earlier version of this paper and Vincenzo Verardi provided indispensable assistance with the Stata implementation. I am also most grateful to the referees for comments and suggestions. The test developed in this paper can be implemented in Stata using the command `xtserialpm`, which can be installed by typing `ssc install xtserialpm` in Stata's command window. The first version of this paper dates from January 22, 2019 and is available as Cambridge Working Paper in Economics CWPE 19/10.

1 Introduction

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

$$y_{g,i} = \boldsymbol{x}_{g,i}' \boldsymbol{\beta} + \eta_{g,i}, \qquad g = 1, \ldots, n, \qquad i = 1, \ldots, m.$$

Although we do not make it explicit in the notation, everything to follow extends readily to unbalanced data; a discussion on this follows. The errors are likely to be correlated within groups. A standard approach (Moulton, 1986) is to model such dependence through the error-component form

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

where α_g is a group-specific effect and the errors $\varepsilon_{g,i}$ are assumed uncorrelated within each group. We will allow for arbitrary dependence between the α_g and the $\boldsymbol{x}_{g,i}$ within the groups. This formulation restricts the pairwise within-group covariances to be constant, a restriction that is seldomly tested.

There are several tests for the presence of a group effect (Moulton and Randolph 1989, Akritas and Arnold 2000, Akritas and Papadatos 2004, Orme and Yamagata 2006, 2013) as well as direct estimators of the variance of group effects (Kline, Saggio and Sølvsten 2018) that are interesting in their own right and can be used to construct a fixed-effect version of the within-between variance decomposition, as is standard in the random-effect model. These procedures all break down when the $\varepsilon_{g,i}$ are correlated within groups. Likewise, the validity of many estimators of $\boldsymbol{\beta}$ hinges on the absence of serial correlation; examples include the popular instrumental-variable estimators of Anderson and Hsiao (1981) and Holtz-Eakin, Newey and Rosen (1988), where identification is usually tested using the procedures developed by Arellano and Bond (1991). Furthermore, even if the estimator of $\boldsymbol{\beta}$ is robust to serial correlation, the use of cluster-robust standard errors (Liang and Zeger 1986, Arellano 1987) for inference can lead to substantial power loss and much inflated confidence regions when the errors are, in fact, uncorrelated; see Wooldridge (2003) and Stock and Watson (2008) for discussion and numerical illustrations. Of course, the presence of correlation in the errors may also be a reflection of model misspecification that is of interest to detect, as in the time-series literature (using, say, the test of Ljung and Box 1978). Indeed, such correlation can arise when a fixed-effect approach to identification fails to adequately control for unobserved within-group heterogeneity. Our empirical illustration on the effect of smoking on infant birthweight is an example of this.

In this paper we develop a test of the null of no within-group correlation beyond that induced by the group-specific effect that has good sampling properties in settings where n is large and m is small. This fixed-m framework is the suitable asymptotic paradigm for micro data and is complicated by the inability to well-estimate the group-specific effects. This is a manifestation of the incidental-parameter problem (Neyman and Scott, 1948) and causes the standard correlation tests from the time-series literature to be inapplicable here. Using a portmanteau test is of interest if no strong stand can be taken on the particular form of correlation that should serve as the alternative. This is relevant in many applications, especially when the observations for a given group do not have a natural ordering (such as time, for example).

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 test statistics of Arellano and Bond (1991), which can be used to test against non-zero qth-order autocorrelation in the first-differenced errors, as well as the joint test for correlation at multiple lags as discussed in Yamagata (2008). Because first-differencing introduces first-order autocorrelation also under the null such a test can only be constructed for $q \geq 2$. Furthermore, at least $q + 2$ observations per group are needed to construct a meaningfull test for qth-order autocorrelation. Hence, a four-wave panel is needed to construct the statistic for $q = 2$, and a five-wave panel is needed for a joint test. In contrast, our test can be applied as soon as three observations per group are available.

Inoue and Solon (2006) proposed a portmanteau test for our null under the additional assumptions that the covariates are strictly exogenous and the errors are homoskedastic. The available tests against specific alternatives (still in the context of static models and under the maintained assumption of homoskedasticity) are discussed in Born and Breitung

(2016). The approach proposed here allows for heteroskedasticity of arbitrary form and only requires the estimator of $\boldsymbol{\beta}$ used to be asymptotically linear. As such, it can be applied to models with exogenous, predetermined, or endogenous regressors (provided, of course, that suitable instrumental variables are available). To the best of our knowledge, a comparable test does not exist in the literature. Relaxing the constant-variance requirement is not a mere technical detail, as it is a strong and often unrealistic restriction to impose. One situation where heteroskedasticity will arise is when the error process is not in its steady state; this is typical in short panels. A second situation is one where errors are conditionally heteroskedastic and some of the regressors are non-stationary. An example here would be a wage regression where the regressors include such characteristics as age, job tenure, and experience, all of which are non-stationary. Born and Breitung (2016) have developed a test for first-order autocorrelation that is size-correct under heteroskedasticity. Unfortunately, this test can have very low power in short panels, even against the first-order moving-average and autoregressive alternatives for which it was designed (this is apparent from our simulations discussed below).

We provide asymptotic-power calculations for three-wave and four-wave data. They give further insight in the behavior of our test and are subsequently used to compare its (theoretical) power to the regression-based test of Wooldridge (2002), the portmanteau test of Inoue and Solon (2006), and the m_2-statistic of Arellano and Bond (1991) (in cases where these test are asymptotically valid). The power comparisons reflect favorably on our test. We also provide results from a simulation study with heteroskedasticity in which we additionally compare power to the test of Born and Breitung (2016). Finally, using data from Abrevaya (2006), we find substantial evidence for correlation across the birthweight of infants from mothers with multiple children, after controlling for mother heterogeneity through a fixed effect and a set of prenatal-care indicators. This finding invalidates the estimates of the between component of a variance decomposition of birthweight and, more importantly, casts doubt on the ability of a fixed-effect approach to account for all latent factors that drive the mother's decision to engage in activities that can have detrimental effects on the infant's health, such as smoking.

2 Testing for within-group correlation

For notational simplicity we maintain a balanced panel. It will become apparent when the test statistic is introduced that unbalanced data (possibly with gaps) does not cause any complication. This, together with its portmanteau nature, makes the test well suited for general data with a group structure; examples are students in classrooms and members of households.

Testing in the one-way analysis of variance model. We initially consider the model

$$\eta_{g,i} = \alpha_g + \varepsilon_{g,i}, \qquad g = 1, \ldots, n, \qquad i = 1, \ldots, m, \tag{2.1}$$

where $\eta_{g,i}$ is directly observable, as in Cox and Solomon (1988), for example. Later we will replace $\eta_{g,i}$ 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 $\mathbb{E}(\varepsilon_{g,i}|\alpha_g) = 0$ for each group.[1] Our aim is to test the (composite) null hypothesis

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

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 $\eta_{g,i}$ will not be suitable. However, when iterating expectations using $\mathbb{E}(\varepsilon_{g,i}|\alpha_g) = 0$ we see that

$$\mathbb{E}(\eta_{g,i_1}(\eta_{g,i_2} - \eta_{g,i_3})) = \mathbb{E}(\varepsilon_{g,i_1}(\varepsilon_{g,i_2} - \varepsilon_{g,i_3})) = \mathbb{E}(\varepsilon_{g,i_1}\varepsilon_{g,i_2}) - \mathbb{E}(\varepsilon_{g,i_1}\varepsilon_{g,i_3}). \tag{2.3}$$

[1] Random sampling at the group level can be relaxed. It suffices to assume that the $\eta_{g,i}$ 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.

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

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

linearly-independent differences. These differences are all zero if and only if we have that $\mathbb{E}(\varepsilon_{g,i_1}\varepsilon_{g,i_2}) = c$ for all $i_1 \neq i_2$ and some constant c. Aside from this, testing (2.2) is equivalent to testing the hypothesis that all r linearly-independent differences of the form in (2.3) are equal to zero. Lack of power against constant within-group covariance is unavoidable in the presence of group-specific effects and is shared by the other available test, including those by Wooldridge (2002) and Drukker (2003) and by Inoue and Solon (2006).

A convenient way to re-write the null that all the differences between covariances are zero is as follows.[2] Introduce the $(m-1) \times r$ matrix

$$\boldsymbol{H}_g := \left(\begin{array}{cccccccc|cccc} 0 & 0 & 0 & 0 & \cdots & 0 & \cdots & 0 & \eta_{g,3} & 0 & \cdots & 0 \\ \eta_{g,1} & 0 & 0 & 0 & & 0 & & 0 & 0 & \eta_{g,4} & & \vdots \\ 0 & \eta_{g,1} & \eta_{g,2} & 0 & & 0 & & 0 & \vdots & & \ddots & 0 \\ \vdots & & & \ddots & & & & \vdots & 0 & 0 & & \eta_{g,m} \\ 0 & 0 & 0 & 0 & \cdots & \eta_{g,1} & \cdots & \eta_{g,m-2} & 0 & 0 & \cdots & 0 \end{array}\right),$$

and collect all errors for a given group in the vector $\boldsymbol{\eta}_g := (\eta_{g,1}, \ldots, \eta_{g,m})'$. Let $\boldsymbol{D}$ denote the $(m-1) \times m$ matrix first-difference operator, so $\boldsymbol{D}\,\boldsymbol{\eta}_g = (\Delta\eta_{g,2}, \ldots, \Delta\eta_{g,m})'$, where $\Delta\eta_{g,i} := \eta_{g,i} - \eta_{g,i-1}$. We then test the r-dimensional null

$$\mathbb{E}(\boldsymbol{v}_g) = \boldsymbol{0}, \qquad \boldsymbol{v}_g := \boldsymbol{H}_g'\boldsymbol{D}\,\boldsymbol{\eta}_g. \tag{2.4}$$

This approach delivers testable moments as soon as more than two observations per group

[2] Our null is equivalent to the set of moment conditions $\mathbb{E}(\eta_{g,i_1}(\eta_{g,i_2} - \eta_{g,i_3})) = 0$, for $i_1 \neq i_2 \neq i_3$, but only r of these equations are linearly independent. Our formulation in (2.4) is not the only way of selecting r such moments but is notationaly convenient. Any other way would yield (numerically) the same test statistic.

are available.[3] The moments in (2.4) are a vectorized version of the collection of moments $\mathbb{E}(\eta_{g,i-j}\,\Delta\eta_{g,i}) = 0$ (for $j = 2, \ldots, i-1$ and $i = 3, \ldots, m$), which correspond to the left block of $\boldsymbol{H}_g$, and $\mathbb{E}(\eta_{g,i+1}\,\Delta\eta_{g,i}) = 0$ ($i = 2, \ldots, m-1$), which correspond to the right block of the same matrix.

Observe that moments of the form

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

are linear combinations of a subset of those in (2.4). These are qth-order autocovariances of $\Delta\varepsilon_{g,i}$. Arellano and Bond (1991) suggested testing for qth-order autocorrelation by evaluating whether the corresponding sample moment can be considered large relative to its standard error. The resulting test statistic is known as the m_q-statistic. Yamagata (2008) proposed to combine all available m_q-statistics into a single test procedure. By consequence, his moment conditions are also nested in (2.4). Notice that, as first-differencing introduces autocorrelation of order one under the null a sensible m_q-statistic can only be constructed for $q \geq 2$. Furthermore, the m_2-statistic requires $m \geq 4$ observations per group. The m_q-statistic is available if $m \geq q+2$. Hence, for the joint approach of Yamagata (2008) to be different from the m_2-statistic one needs panel data that consists of at least five waves.

Our test statistic for the null (2.4) is the quadratic form

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

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

[3] An alternative way to arrive at (2.4) is by noting that

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

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

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

which lead to (2.4).

In the theorem we consider sequences of local alternatives where

$$\mathbb{E}(\varepsilon_{g,i_1}\varepsilon_{g,i_2}) = \frac{\sigma_{i_1,i_2}}{\sqrt{n}}$$

and σ_{i_1,i_2} is non-zero for at least one pair of indices $i_1 \neq i_2$ (and not the same constant for all distinct pairs $i_1 \neq i_2$). We write the resulting Pitman drift in the moment condition as

$$\mathbb{E}(\boldsymbol{v}_g) = \frac{\boldsymbol{\delta}}{\sqrt{n}}. \tag{2.5}$$

Here, $\boldsymbol{\delta}$ contains differences of the form $\sigma_{i_1,i_2} - \sigma_{i_1,i_3}$, according to our specification of $\boldsymbol{v}_g$. For example, when $m = 3$,

$$\boldsymbol{\delta} = \begin{pmatrix} \sigma_{1,3} - \sigma_{1,2} \\ \sigma_{3,2} - \sigma_{3,1} \end{pmatrix}.$$

Here and later we denote the non-central χ^2-distribution with q degrees of freedom and non-centrality parameter c by $\chi^2(q, c)$.

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

(i) If the null (2.4) *holds* $s_n \xrightarrow{L} \chi^2(r, 0)$*.*

(ii) Under a sequence of local alternatives as in (2.5) $s_n \xrightarrow{L} \chi^2(r, \boldsymbol{\delta}'\boldsymbol{V}^{-1}\boldsymbol{\delta})$*.*

Proof. See the Appendix. □

The result implies that a test that has size $\alpha \in (0, 1)$ in large samples can be constructed by comparing s_n 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 of the form in (2.5).[4]

[4]Asymptotic power is not invariant to reshuffling the observations within the groups unless the same permutation of the data points is applied to all groups. This lack of invariance, shared by other tests, is not an issue in time-series and related problems but could possibly be in some other situations. Nonetheless, often a natural ordering presents itself. When groups are households, for example, it seems natural to order their members so that parents are the initial observations and children (ordered by age, say) follow after.

When the panel is unbalanced some of the entries of the vector $\boldsymbol{v}_g$ will be missing for some groups. Setting such entries to zero (i.e., retaining only the non-missing data), our test remains consistent provided that the number of groups for which we observe $\eta_{g,i_1} \Delta\eta_{g,i_2}$ grows at the rate n for each pair (i_1, i_2) that features in (2.4) (of course, under the assumption that the missingness is at-random)[5].

Testing in the one-way regression model. We now generalize (2.1) to the regression setting

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

where $y_{g,i}$ and $\boldsymbol{x}_{g,i}$ are an observable outcome and p-vector of covariates, respectively, and $\eta_{g,i}$ 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

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

as estimators of the $\eta_{g,i}$ and construct the test statistic based on these residuals. We will require the estimator $\hat{\boldsymbol{\beta}}$ to be asymptotically linear under the null and under local alternatives of the form in (2.5), i.e., that

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \sum_{g=1}^{n} \frac{\boldsymbol{\omega}_g}{\sqrt{n}} + o_P(1) \tag{2.6}$$

for a random variable $\boldsymbol{\omega}_g$ that has finite variance and zero-mean under our null, but may have non-zero mean under local deviations from our null. This is a very mild requirement as all commonly-used estimators satisfy it (of course, under suitable regularity conditions, see Newey 1985). When the covariates are strictly exogenous, for example, a natural estimator of $\boldsymbol{\beta}$ would be the within-group least squares estimator. This estimator is robust to within-group correlation. In contrast, when the covariates are only pre-determined, the instrumental-variable estimators described in Holtz-Eakin, Newey and Rosen (1988), which are based on moment conditions of the form $\mathbb{E}(\boldsymbol{z}_{g,i}\Delta\eta_{g,i}) = \boldsymbol{0}$ for $\boldsymbol{z}_{g,i} := (\boldsymbol{x}'_{g,i-2}, \ldots, \boldsymbol{x}'_{g,1})'$

[5]An alternative approach would be to drop some observations to restore balancedness. Depending on the alternative under consideration this may or may not be power enhancing.

(or a subvector thereof, as in Anderson and Hsiao 1981) and all $1 < i \leq m$, will generally be asymptotically biased under local alternatives as in (2.5).

To set up our test statistic based on residuals we first define, in analogy to $\boldsymbol{H}_g$ and $\boldsymbol{\eta}_g$,

$$\boldsymbol{E}_g := \begin{pmatrix} 0 & 0 & 0 & 0 & \cdots & 0 & \cdots & 0 & e_{g,3} & 0 & \cdots & 0 \\ e_{g,1} & 0 & 0 & 0 & & 0 & & 0 & 0 & e_{g,4} & & \vdots \\ 0 & e_{g,1} & e_{g,2} & 0 & & 0 & & 0 & \vdots & & \ddots & 0 \\ \vdots & & & \ddots & & & & \vdots & 0 & 0 & & e_{g,m} \\ 0 & 0 & 0 & 0 & \cdots & e_{g,1} & \cdots & e_{g,m-2} & 0 & 0 & \cdots & 0 \end{pmatrix}$$

and $\boldsymbol{e}_g := (e_{g,1}, \ldots, e_{g,m})'$. These allow us to construct $\hat{\boldsymbol{v}}_g := \boldsymbol{E}_g' \boldsymbol{D}\, \boldsymbol{e}_g$, which is the plug-in estimator of $\boldsymbol{v}_g = \boldsymbol{H}_g' \boldsymbol{D}\, \boldsymbol{\eta}_g$. The use of residuals requires modifying the weight matrix in the quadratic form of the test statistic, however. In the proof to Theorem 2 we show that

$$\sum_{g=1}^{n} \hat{\boldsymbol{v}}_g = \sum_{g=1}^{n} (\boldsymbol{v}_g + \boldsymbol{\Omega} \boldsymbol{\omega}_g) + o_P(\sqrt{n}), \tag{2.7}$$

where the $r \times p$ Jacobian matrix

$$\boldsymbol{\Omega} := \mathbb{E}\left(\partial \boldsymbol{v}_g / \partial \boldsymbol{\beta}'\right) = \mathbb{E}(\dot{\boldsymbol{H}}_g'(\boldsymbol{I}_p \otimes \boldsymbol{D}\, \boldsymbol{\eta}_g)) - \mathbb{E}(\boldsymbol{H}_g' \boldsymbol{D}\, \boldsymbol{X}_g)$$

features $\dot{\boldsymbol{H}}_g := \left(\partial \boldsymbol{H}_g' / \partial \boldsymbol{\beta}_1, \ldots, \partial \boldsymbol{H}_g' / \partial \boldsymbol{\beta}_p\right)'$ and $\boldsymbol{X}_g := (\boldsymbol{x}_{g,1}, \ldots, \boldsymbol{x}_{g,m})'$, and we use $\boldsymbol{I}_p$ to denote the $p \times p$ identity matrix. Each of the matrices $\partial \boldsymbol{H}_g / \partial \boldsymbol{\beta}_q$ is of the same form as $\boldsymbol{H}_g$, only with $\eta_{g,i}$ replaced by the qth-entry of $\partial \eta_{g,i} / \partial \boldsymbol{\beta} = -\boldsymbol{x}_{g,i}$. A plug-in estimator of $\boldsymbol{\Omega}$ is easily constructed as

$$\boldsymbol{O} := \frac{1}{n} \sum_{g=1}^{n} \dot{\boldsymbol{H}}_g'(\boldsymbol{I}_p \otimes \boldsymbol{D}\, \boldsymbol{e}_g) - \frac{1}{n} \sum_{g=1}^{n} \boldsymbol{E}_g' \boldsymbol{D}\, \boldsymbol{X}_g.$$

Thus, with $\hat{\boldsymbol{\omega}}_g$ denoting an estimator of the influence function $\boldsymbol{\omega}_g$, our test statistic based on residuals is

$$\hat{s}_n := \left(\sum_{g=1}^{n} \hat{\boldsymbol{v}}_g \right)' \left(\sum_{g=1}^{n} (\hat{\boldsymbol{v}}_g + \boldsymbol{O} \hat{\boldsymbol{\omega}}_g)(\hat{\boldsymbol{v}}_g + \boldsymbol{O} \hat{\boldsymbol{\omega}}_g)' \right)^{-1} \left(\sum_{g=1}^{n} \hat{\boldsymbol{v}}_g \right).$$

Here, $\hat{\boldsymbol{\omega}}_g$ will depend on the problem at hand. If residuals are constructed using the within-group estimator, for example, then $\hat{\boldsymbol{\omega}}_g = (n^{-1} \sum_{g=1}^{n} \boldsymbol{X}_g' \boldsymbol{M} \boldsymbol{X}_g)^{-1} \boldsymbol{X}_g' \boldsymbol{M} \boldsymbol{e}_g$, where

$\boldsymbol{M} := \boldsymbol{I}_m - \boldsymbol{D}'(\boldsymbol{D}\,\boldsymbol{D}')^{-1}\boldsymbol{D}$, the matrix that transforms observations into deviations from their within-group mean.

Theorem 2 summarizes the large-sample properties of the test statistic $\hat{s}_n$. We let $\tilde{\boldsymbol{\delta}} := \boldsymbol{\delta} + \boldsymbol{\Omega}\,\mathbb{E}(\boldsymbol{\omega}_g)$ and use $\|\cdot\|$ to denote both the Euclidean norm and the Frobenius norm.

Theorem 2. *Suppose that $\mathbb{E}(\alpha_g^4) < \infty$, $\mathbb{E}(\varepsilon_{g,i}^4) < \infty$, and $\mathbb{E}(\|\boldsymbol{x}_{g,i}\|^4) < \infty$, that (2.6) holds and that $n^{-1}\sum_{g=1}^{n}\|\hat{\boldsymbol{\omega}}_g - \boldsymbol{\omega}_g\|^2 = o_P(1)$, and that $\tilde{\boldsymbol{V}} := \mathbb{E}((\boldsymbol{v}_g + \boldsymbol{\Omega}\boldsymbol{\omega}_g)(\boldsymbol{v}_g + \boldsymbol{\Omega}\boldsymbol{\omega}_g)')$ has maximal rank r.*

(i) If the null (2.4) holds $\hat{s}_n \xrightarrow{L} \chi^2(r, 0)$.

(ii) Under a sequence of local alternatives as in (2.5) $\hat{s}_n \xrightarrow{L} \chi^2(r, \tilde{\boldsymbol{\delta}}'\tilde{\boldsymbol{V}}^{-1}\tilde{\boldsymbol{\delta}})$.

Proof. See the Appendix. □

Theorem 2 differs from Theorem 1 in the local-power result. Estimation noise in $\hat{\boldsymbol{\beta}}$ changes the weight matrix in the non-centrality parameter. This change 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 this happens depends on the alternative in question. The degree to which both channels matter is governed by the Jacobian matrix $\boldsymbol{\Omega}$. Estimation of $\boldsymbol{\beta}$ will have no (asymptotic) impact on the properties of our test if $\boldsymbol{\Omega}$ is equal to the zero matrix. This would happen, for example, when the covariates are strictly exogenous and the effect α_g is orthogonal to all the $\Delta\boldsymbol{x}_{g,i}$, as in the random-effect model (Arellano, 2003, Chapter 3). In this case, Theorem 2 collapses to Theorem 1.

3 Power comparisons

In this section we provide power comparisons in the random-effect model for three-wave and four-wave data.

Power calculations. We first calculate asymptotic power in specific cases for three-wave panels. In this case we test two moment conditions. Suppose that $\alpha_g \sim$ independent $(0, \gamma^2)$

and that the errors are generated according to the (non-stationary) moving-average process of order one

$$\varepsilon_{g,i} = u_{g,i} + \theta\, u_{g,i-1}, \qquad u_{g,i} \sim \text{independent}\,(0, \sigma_i^2).$$

Then

$$\mathbb{E}(\boldsymbol{v}_g) = \mathbb{E}\begin{pmatrix} \eta_{g,1}\Delta\eta_{g,3} \\ \eta_{g,3}\Delta\eta_{g,2} \end{pmatrix} = \theta \begin{pmatrix} -\sigma_1^2 \\ \sigma_2^2 \end{pmatrix},$$

which is zero if and only if the errors are uncorrelated, i.e., $\theta = 0$. Further, under the null,

$$\boldsymbol{V} = \begin{pmatrix} (\gamma^2+\sigma_1^2)(\sigma_3^2+\sigma_2^2) & -\sigma_1^2\sigma_3^2 - \gamma^2\sigma_2^2 \\ -\sigma_1^2\sigma_3^2 - \gamma^2\sigma_2^2 & (\gamma^2+\sigma_3^2)(\sigma_1^2+\sigma_2^2) \end{pmatrix}.$$

The non-centrality parameter of the limit distribution under local alternatives then equals

$$\theta^2 \frac{\sigma_1^4(\gamma^2+\sigma_3^2)(\sigma_1^2+\sigma_2^2) - 2\sigma_1^2\sigma_2^2(\sigma_1^2\sigma_3^2+\gamma^2\sigma_2^2) + \sigma_2^4(\gamma^2+\sigma_1^2)(\sigma_2^2+\sigma_3^2)}{(\gamma^2+\sigma_1^2)(\gamma^2+\sigma_3^2)(\sigma_1^2+\sigma_2^2)(\sigma_2^2+\sigma_3^2) - (\sigma_1^2\sigma_3^2+\gamma^2\sigma_2^2)^2},$$

which is independent of σ_0^2 but otherwise complicated. When the errors are homoskedastic with common variance σ^2 the non-centrality parameter becomes

$$\frac{2}{3}\frac{\theta^2}{1+\gamma^2/\sigma^2}.$$

Consequently, power is monotone increasing in $|\theta|$ and decreasing in the ratio γ^2/σ^2. Both these findings are intuitive. When we set $\gamma^2 = 0$ but allow the errors to be heteroskedastic the non-centrality parameter equals

$$\theta^2 \frac{\sigma_2^2\left(\frac{\sigma_2^2}{\sigma_3^2}+1\right) + \sigma_1^2\left(\frac{\sigma_1^2}{\sigma_2^2}-1\right)}{\sigma_1^2+\sigma_2^2+\sigma_3^2}.$$

Power is sensitive to the variances of the innovations. For example, an increase in σ_3^2, ceteris paribus, will always cause power loss while changes in σ_1^2 and σ_2^2 can be both power enhancing and power reducing.

Given an expression for the non-centrality parameter we can use it to approximate the power function of our test for any given sample size n. In general, for a test with size α and an alternative for which the non-centrality parameter takes on the value μ, we do this by computing the $(1-\alpha)$th quantile of the $\chi^2(r, n\mu)$ distribution. The left plot in

Figure 1: Theoretical power calculations

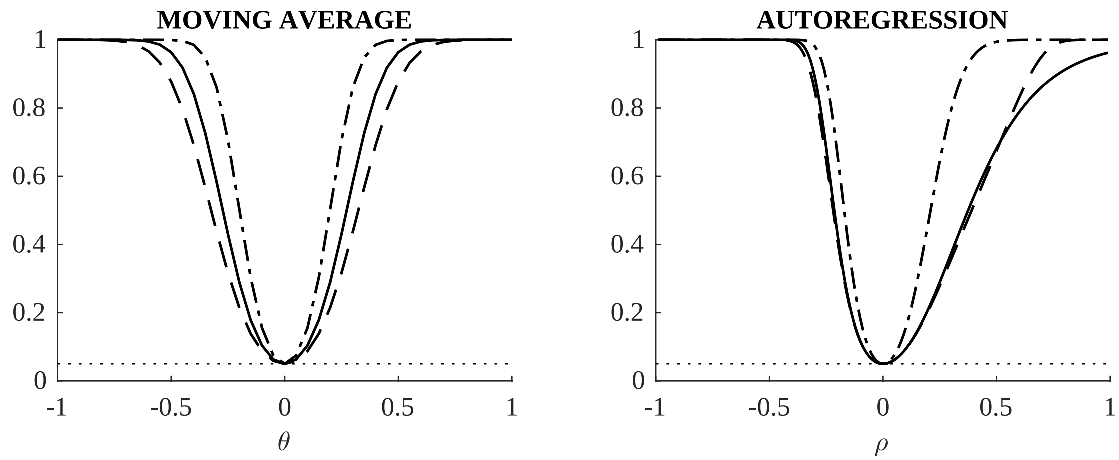

Figure notes. Left plot: stationary data (solid line), increasing heteroskedasticity (dashed line), and decreasing heteroskedasticity (dashed-dotted line). Right plot: stationary data (solid line), constant initial conditions (dashed line), variance jump in second wave (dashed-dotted line). The size of the test (.05) is indicated by a horizontal dotted line.

Figure 1 shows the (theoretical) power of a 5%-level test, plotted as a function of θ, for 100 observations, no group-specific effects, and three choices of the variance parameters. The horizontal dotted line marks the size of the test. The solid line corresponds to the stationary case where $\sigma^2 = 1$ and shows substantial power. The dashed line is for a case where the variances increase, with $\sigma_1^2 = 1$, $\sigma_2^2 = 2$, and $\sigma_3^2 = 3$. In this case power is slightly lower than in the stationary case. The dashed-dotted line is for the reverse case where the cross-sectional variances decrease from three down to one. This yields a uniformly-superior power curve relative to the stationary case.

Now suppose that the errors are generated via the first-order autoregressive specification

$$\varepsilon_{g,i} = \rho\, \varepsilon_{g,i-1} + u_{g,i}, \qquad u_{g,i} \sim \text{independent}\,(0, \sigma_i^2),$$

where $\varepsilon_{g,0} \sim$ independent $(0, \sigma_0^2)$. Then

$$\mathbb{E}(\boldsymbol{v}_g) = \mathbb{E}\begin{pmatrix} \eta_{g,1}\Delta\eta_{g,3} \\ \eta_{g,3}\Delta\eta_{g,2} \end{pmatrix} = -(1-\rho)\,\rho \begin{pmatrix} \rho^2\sigma_0^2 + \quad \sigma_1^2 \\ \rho^3\sigma_0^2 + \rho\sigma_1^2 \end{pmatrix} + \begin{pmatrix} 0 \\ \rho\sigma_2^2 \end{pmatrix}.$$

This will be non-zero for any value of the autoregressive parameter except zero. This includes the random-walk alternative. The non-centrality parameter in Theorem 1 can again be computed using the same matrix $\boldsymbol{V}$ as was used above. However, the resulting expression is long and difficult to interpret. When the process is stationary the expression reduces to

$$\mathbb{E}(\boldsymbol{v}_g) = \sigma^2 \varrho \begin{pmatrix} -1 \\ 1 \end{pmatrix},$$

with $\varrho := \rho/(1+\rho)$, which lives on $(-\infty, \frac{1}{2})$. Then the non-centrality parameter is equal to

$$\frac{2}{3} \frac{\varrho^2}{1+\gamma^2/\sigma^2}.$$

An alternative where $\varrho = h/\sqrt{n}$ for some h corresponds to an autoregressive parameter that satisfies $\rho = h/\sqrt{n} + O(n^{-1})$. This result implies that the moving-average and autoregressive alternatives are locally asymptotically equivalent. Godfrey (1981) and Yamagata (2008) have reached the same conclusion, albeit for different tests.

The right plot in Figure 1 provides power approximations for three configurations of the autoregressive process—again for $n = 100$ and no group-specific effects—plotted as a function of ρ. The solid line is again for the stationary setting. This power curve is asymmetric, with power against a given $\rho > 0$ being lower than against $-\rho$. This can be explained by the fact that $|\varrho|$ is asymmetric in ρ. It takes on large (negative) values for $\rho < 0$ and small values for $\rho > 0$. Non-stationarity can be a source of power against autoregressive alternatives. The dashed power curve corresponds to the case where $\sigma_0^2 = 0$ and all other variances are equal to unity. This corresponds to setting all initial conditions of the error process to zero and increases power relative to the stationary design. In the unit-root case the Pitman drift depends only on σ_2^2 and so a larger value would be expected to yield more power close to unity. The dashed-dotted curve illustrates this by showing power when $\sigma_1 = \sigma_3^2 = 1$, $\sigma_2^2 = 2$, and $\sigma_0^2 = 1/(1-\rho^2)$.

Power comparisons. Under the null of no within-group correlation the first-differenced errors are autocorrelated at first-order (but not beyond). When errors are homoskedastic this correlation equals $-\frac{1}{2}$. A simple and popular alternative approach in this setting is to test whether $t := \text{corr}(\varDelta\eta_{g,i}, \varDelta\eta_{g,i-1})$ is different from $-\frac{1}{2}$. This test was introduced in Wooldridge (2002, p. 282–283) and further discussed in Drukker (2003). Simple calculations reveal that

$$t = -\frac{1}{2} + \frac{1}{2}\frac{\theta}{1-\theta+\theta^2}$$

under moving-average alternatives, while

$$t = -\frac{1}{2} + \frac{1}{2}\rho$$

when the errors follow an autoregression of order one. In both classes of alternatives autocorrelation is most pronounced at first order. This makes Wooldridge-Drukker test the primary competitor. If we assume that $\varepsilon_{g,i} \sim \text{i.i.d. N}(0, \sigma^2)$, then, under local alternatives,

$$\left(\hat{t} + \frac{1}{2}\right)^2 \xrightarrow{L} \chi^2\left(1, \frac{4}{3}\left(t + \frac{1}{2}\right)^2\right),$$

where $\hat{t}$ denotes the first-order sample correlation coefficient. Inspection of how t varies with θ shows that power is asymmetric, with power being higher against $\theta > 0$ than against the corresponding $-\theta$. This asymmetry does not arise with autoregressive alternatives. Figure 2 presents power comparisons for sample sizes of $n = 100$. Power is plotted for our test with $\gamma^2/\sigma^2 = 0$ (solid line), $\gamma^2/\sigma^2 = 1$ (dashed line), and $\gamma^2/\sigma^2 = 2$ (dashed-dotted line) and for the test of Wooldridge (2002) and Drukker (2003) (dotted line). The curves clearly illustrate our findings.

The Wooldridge-Drukker test is not a portmanteau test. It will have no power when serial correlation manisfests itself only at higher order. It will also have no power against alternatives where $\mathbb{E}(\varepsilon_{g,i}\,\varepsilon_{g,i-1}) = \mathbb{E}(\varepsilon_{g,i}\,\varepsilon_{g,i-2})$, for example. The latter situation arises, for example, in the moving-average process of order two,

$$\varepsilon_{g,i} = u_{g,i} + \theta_1 u_{g,i-1} + \theta_2 u_{g,i-2}, \qquad u_{g,i} \sim \text{independent}(0, \sigma^2),$$

when $\theta_1 = \theta_2/(1+\theta_2)$. Indeed, here, $t = -\frac{1}{2}$ for any value of θ_2. The portmanteau test of Inoue and Solon (2006) can be seen as a generalization of the Wooldridge-Drukker test.

Figure 2: Theoretical power comparison with the Wooldridge-Drukker statistic

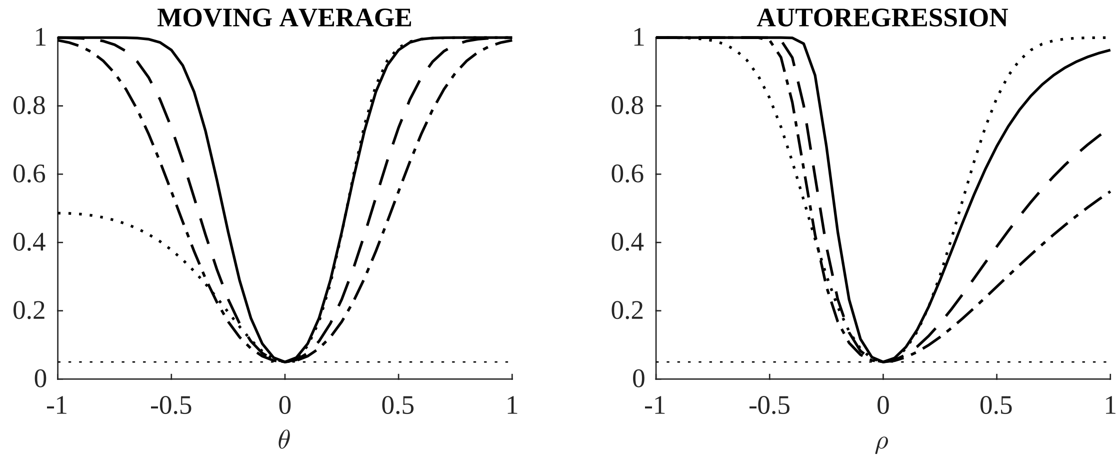

Figure notes. Our test for $\gamma^2 = 0$ (solid line), $\gamma^2 = 1$ (dashed line), and $\gamma^2 = 2$ (dashed-dotted line), and the Wooldridge-Drukker test (dotted line). The size of the tests (.05) is indicated by a horizontal dotted line.

It evaluates whether the correlations between $\varepsilon_{g,i_1} - \overline{\varepsilon}_g$ and $\varepsilon_{g,i_2} - \overline{\varepsilon}_g$ are different from $-1/(m-1)$, where $\overline{\varepsilon}_g$ is the within-group average of the errors. The Inoue and Solon (2006) test depends on a regularization parameter that serves to handle the fact that the demeaned errors sum to zero within each group. With three waves their procedure yields three possible tests statistics, say $t_{(1,2)}$, $t_{(1,3)}$, and $t_{(2,3)}$. The statistic $t_{(i_1,i_2)}$ tests the null that

$$\operatorname{corr}((\varepsilon_{g,i_1} - \overline{\varepsilon}_g), (\varepsilon_{g,i_2} - \overline{\varepsilon}_g)) = -1/(m-1).$$

In stationary designs $t_{(1,2)}$ and $t_{(2,3)}$ will have the same limit behavior. Calculations reveal that

$$t_{(1,2)} \xrightarrow{L} \chi^2\left(1, \left(\frac{1}{3}d\right)^2\right), \qquad t_{(1,3)} \xrightarrow{L} \chi^2\left(1, \left(\frac{2}{3}d\right)^2\right),$$

where $d = \theta$ for moving average alternatives and $d = \varrho$ for autoregressive alternatives. In both cases, $t_{(1,3)}$ is revealed to be much more powerful than $t_{(1,2)}$ and $t_{(2,3)}$. Figure 3 provides a power comparison between our test and the Inoue and Solon (2006) test. The power curves for our test coincide with those in Figure 2. The power curves for the Inoue and Solon (2006) tests (dotted lines) carry a marker to indicate which test statistic is used;

Figure 3: Theoretical power comparison with the Inoue-Solon statistic

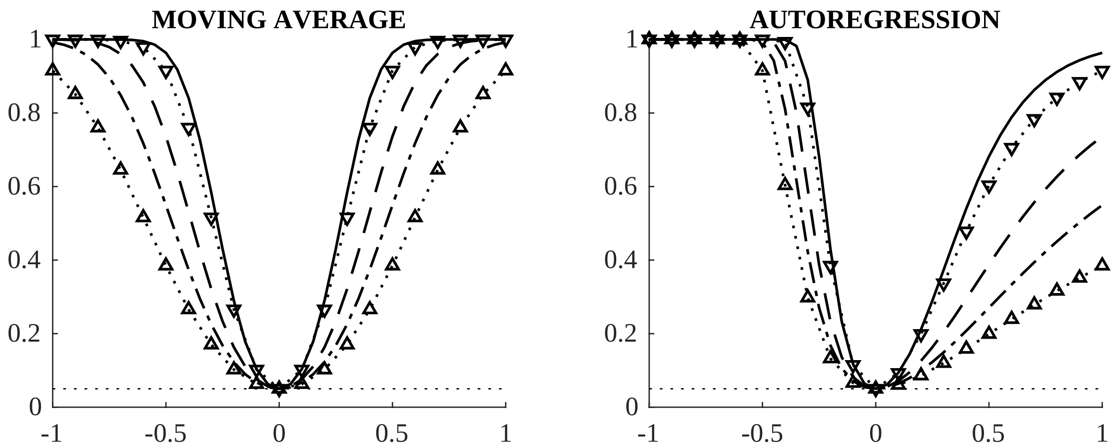

Figure notes. Our test for $\gamma^2 = 0$ (solid line), $\gamma^2 = 1$ (dashed line), and $\gamma^2 = 2$ (dashed-dotted line), and the Inoue-Solon tests (dotted line with triangular markers). The size of the tests (.05) is indicated by a horizontal dotted line.

upward triangles are for $t_{(1,2)}$ and $t_{(2,3)}$ while downward triangles are for $t_{(1,3)}$.

For a four-wave panel we can compare the power of our test to the m_2-statistic of Arellano and Bond (1991) (which, here, coincides with the test of Yamagata 2008). It tests the single moment condition $\mathbb{E}(\varDelta\eta_{g,2}\varDelta\eta_{g,4}) = 0$. The (square of the) m_2-statistic is equal to

$$\left(\frac{\sum_{g=1}^{n} \varDelta\eta_{g,2}\varDelta\eta_{g,4}}{\sum_{g=1}^{n} \varDelta\eta_{g,2}^2\varDelta\eta_{g,4}^2}\right)^2$$

and has a $\chi^2(1,0)$ limit distribution under the null. Under local moving-average alternatives of the same type as before the non-centrality parameter is

$$\theta^2 \frac{\sigma_2^4}{(\sigma_1^2 + \sigma_2^2)(\sigma_3^2 + \sigma_4^2)},$$

which reduces to $\theta^2/4$ under homoskedasticity. In contrast, our procedure tests five moment conditions, and the non-centrality parameter of its limit distribution equals

$$\frac{3}{2} \frac{\theta^2}{1 + \gamma^2/\sigma^2}.$$

The left plot in Figure 4 provides the power curves for our test statistic (solid line) and for the m_2-statistic (dashed line), again for a sample of size $n = 100$ and $\gamma^2 = 0$. Our approach

Figure 4: Theoretical power comparison with the m_2-statistic

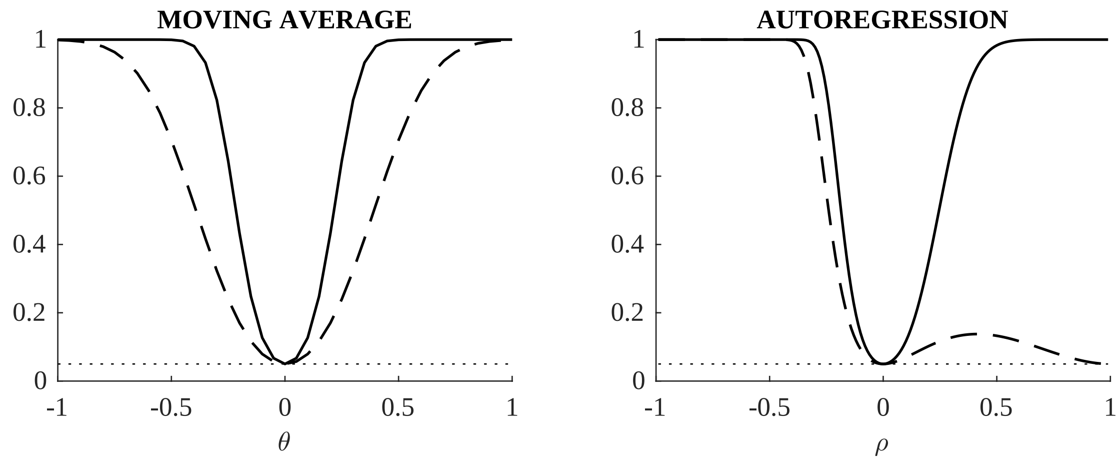

Figure notes. Our test (solid line) and the Arellano-Bond test (dashed line). The size of the tests (.05) is indicated by a horizontal dotted line.

is seen to be uniformly more powerful. As γ^2/σ^2 increases relative power decreases, in line with the left plot in Figure 2.

Under stationary autoregressive alternatives,

$$\mathbb{E}(\mathit{\Delta}\eta_{g,2}\mathit{\Delta}\eta_{g,4}) = -\varrho\,(1-\rho)\sigma^2,$$

and the non-centrality parameter for violations of the moment condition is

$$\frac{1}{4}(\varrho(1-\rho))^2.$$

This expression is non-monotone in ρ. Furthermore, it approaches zero as $\rho \to 1$ and so the m_2-statistic will have low power against near unit-root alternatives. The non-centrality parameter for our test statistic in this context equals

$$\frac{3}{2}\frac{\varrho^2}{1+\gamma^2/\sigma^2}\left(1+\frac{2}{3}\rho+\frac{5}{9}\rho^2\right).$$

This is strictly larger than the non-centrality parameter for three-wave data for all ρ. The right plot in Figure 4 —again for $n = 100$ and $\gamma^2 = 0$—confirms the poor power of the m_2-statistic against positive autocorrelation patterns and shows favorable power for our test.

Simulations. We next provide results from a simulation experiment on a heteroskedastic regression model. We generated outcomes with $\alpha_g \sim \mathrm{N}(0,1)$, a bivariate regressor $\boldsymbol{x}_{g,i}$ containing a standard-normal regressor and a binary regressor (whose success probability is equal to one half), $\boldsymbol{\beta} = (1,1)'$, and the idiosyncratic errors generated either through (i) the moving-average process

$$\varepsilon_{g,i} = u_{g,i} + \theta\, u_{g,i-1}, \qquad u_{g,i} \sim \mathrm{N}(0, \sigma_i^2),$$

or (ii) the first-order autoregressive process

$$\varepsilon_{g,i} = \rho\, \varepsilon_{g,i-1} + u_{g,i}, \qquad u_{g,i} \sim \mathrm{N}(0, \sigma_i^2);$$

here, $\sigma_i^2 = 1 + \log(i)$ for $i \geq 1$ and the error processes are initialized at zero in each case. We estimate $\boldsymbol{\beta}$ by the within-group estimator and use the test statistic from Theorem 2 in the simulations below. Results are presented in Figures 5–8 for $n = 250$ and (clockwise) $m = 3, 6, 9, 12$.

Figure 5 contains power curves (as obtained over 10,000 Monte Carlo replications) for our test (solid line), the test of Yamagata (2008) (dashed line) and the robust version of the Wooldridge-Drukker test due to Born and Breitung (2016). The horizontal dotted line in each plot indicates the (theoretical) size of the tests. Because both the test of Yamagata (2008) and the test of Born and Breitung (2016) require $m \geq 4$, a power curve for these tests is absent from the upper-left plot. Each of the three tests is size-correct under heteroskedasticity and this is borne out of our results. Our test and the test of Yamagata (when it exists) show comparable power, with the latter being slightly more powerful for the larger values of m. This can be explained by the fact that we test $r = m(m-1)/2 - 1$ restrictions while his test involves $m-3$ restrictions. The test of Born and Breitung (2016) has much lower power against all alternatives and for all sample sizes. Figure 6 compares the performance of our test with the test of Inoue and Solon (2006) (dashed line) implemented with the regularization parameter set as in Inoue and Solon (2006, p. 841) (which is also how it is implemented in Stata), and the test of Wooldridge (2002). These two approaches are not designed to handle heteroskedasticity and so are not size correct. The results are indicative of how ignoring heteroskedasticity can lead to incorrect inference. Indeed, the portmanteau test severely overrejects under the null while the test against first-order correlation is biased toward positive autocorrelation.

Figures 7 and 8 provide corresponding power results against autoregressive alternatives. The power curves for our test are essentially the same compared to the moving-average alternatives. The test of Yamagata has difficulties detecting alternatives that are close to a unit root. The test of Born and Breitung again has low power against a wide range of values for the autoregressive parameter. Similarly to before, Figure 8 shows that using a test statistic that is not robust to heteroskedasticity leads to tests that are not size correct and biased.

4 Empirical illustration

Infering the effect of smoking on birth outcomes is complicated by latent characteristics that are likely to be correlated with the mother's decision to smoke. The early literature has aimed to tackle this problem via instrumental variables; see Permutt and Hebel (1989), and Currie and Gruber (1996) and Reichman and Florio (1996) for related work. An alternative approach, taken by Abrevaya (2006), is to fit a fixed-effect model to data on mothers with multiple children. Exploiting repeated measurements is a powerful device in this context as it allows to control for unobserved characteristics of the mother that are constant across births.

As mothers with two children are uninformative for our purposes we restrict the sample of Abrevaya (2006) to mothers with three children (a larger number was not observed in these data). This yields three observations on $12,360$ mothers. Such a dimension fits our asymptotic approximation very well. Table 1 provides a summary of the variables in the data. `weight` is the newborn's weight (in grams). `smokes` is a binary variable indicating whether or not the mother smokes and `n_cigarettes` is the number of cigarettes smoked (per day). The control variables that vary across births for a given mother are the age of the mother (in years) (`age`), the newborn's gender (`male`) together with several variables

Figure 5: Simulated power against moving-average alternatives (robust tests)

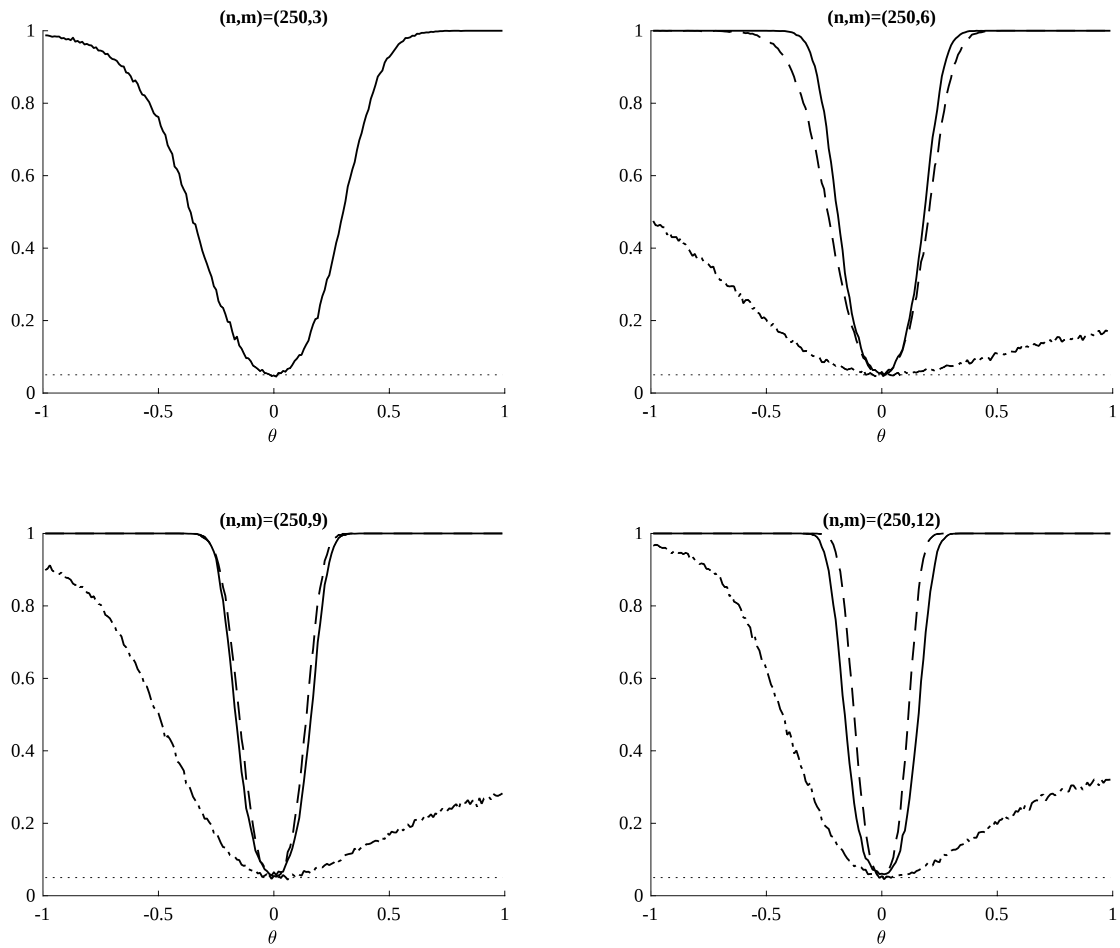

Figure notes. Our test (solid line), the Arellano-Bond/Yamagata test (dashed line), and the Born-Breitung test (dashed-dotted line). The size of the tests (.05) is indicated by a horizontal dotted line.

Figure 6: Simulated power against moving-average alternatives (non-robust tests)

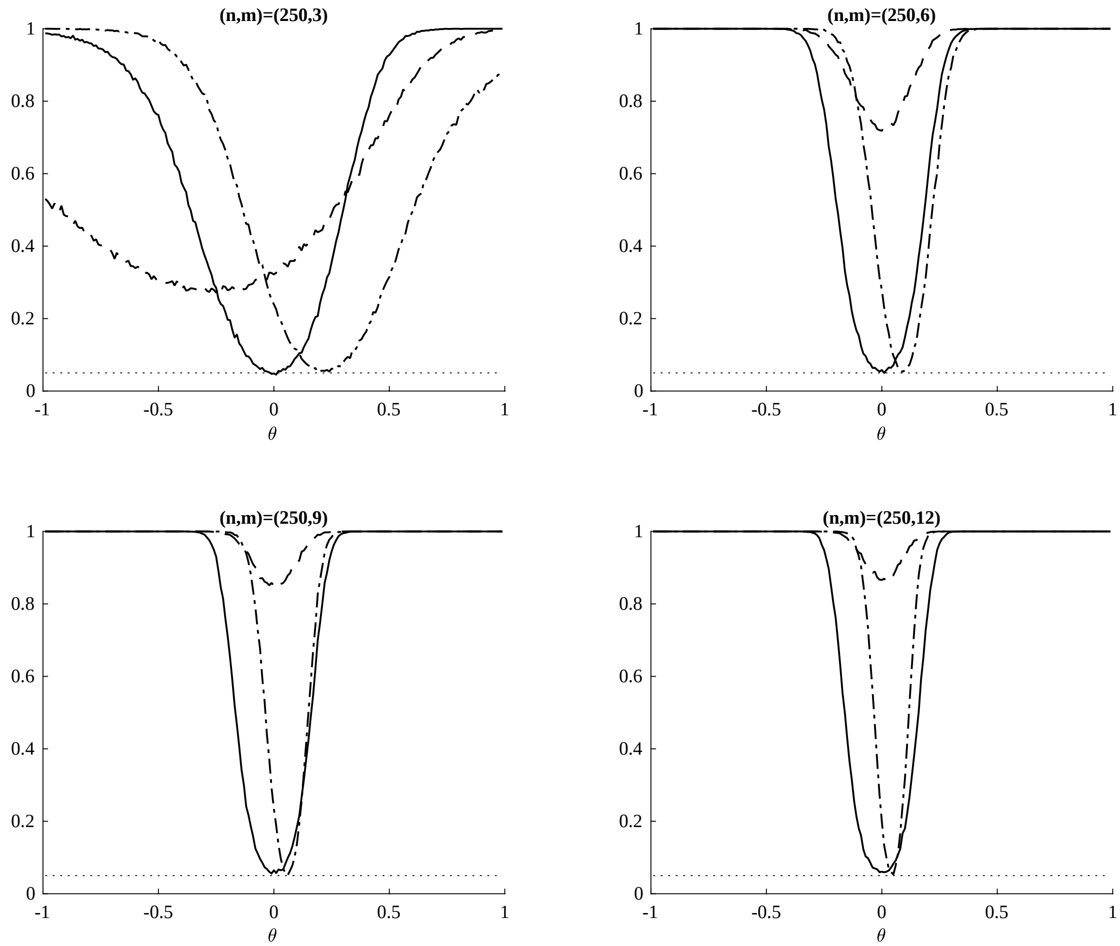

Figure notes. Our test (solid line), the Inoue-Solon test (dashed line), and the Wooldridge-Drukker test (dashed-dotted line). The size of the tests (.05) is indicated by a horizontal dotted line.

Figure 7: Simulated power against autoregressive alternatives (robust tests)

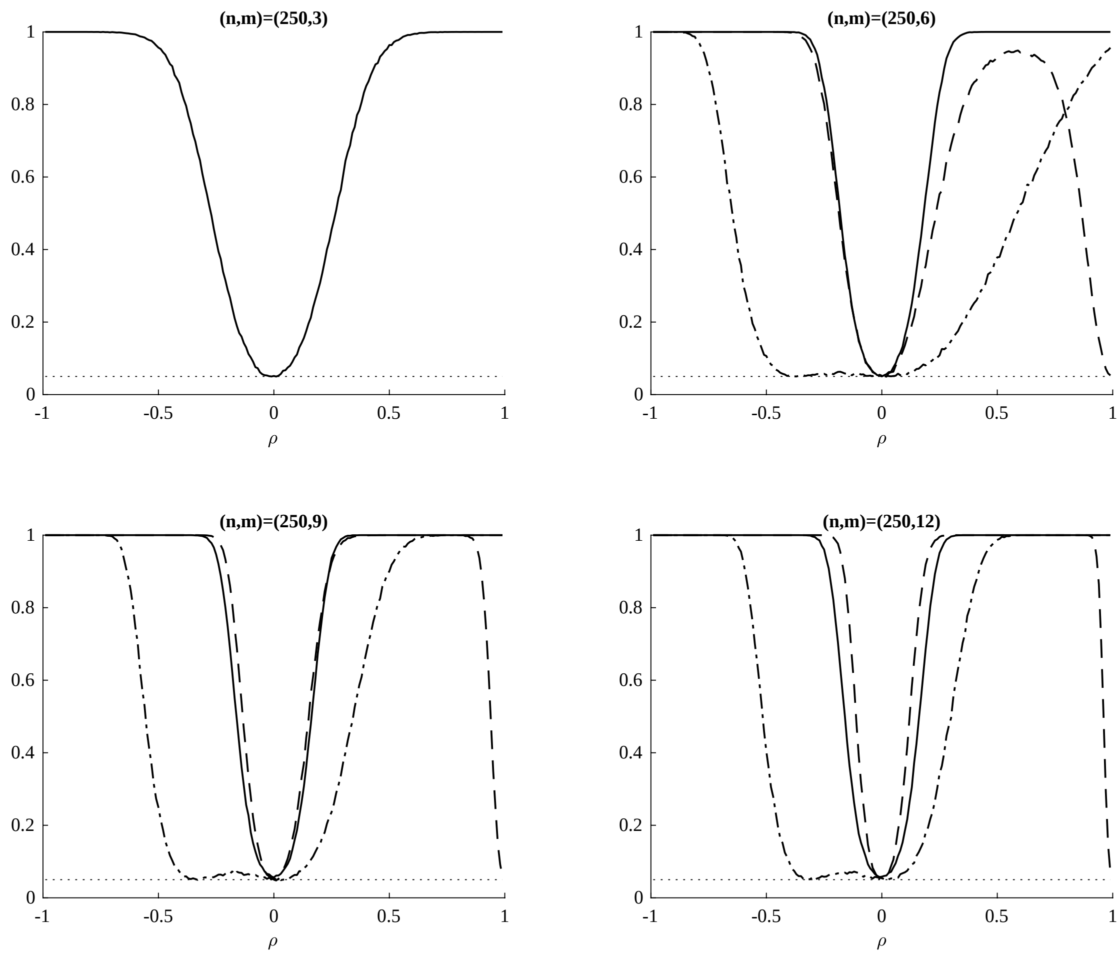

Figure notes. Our test (solid line), the Arellano-Bond/Yamagata test (dashed line), and the Born-Breitung test (dashed-dotted line). The size of the tests (.05) is indicated by a horizontal dotted line.

Figure 8: Simulated power against autoregressive alternatives (non-robust tests)

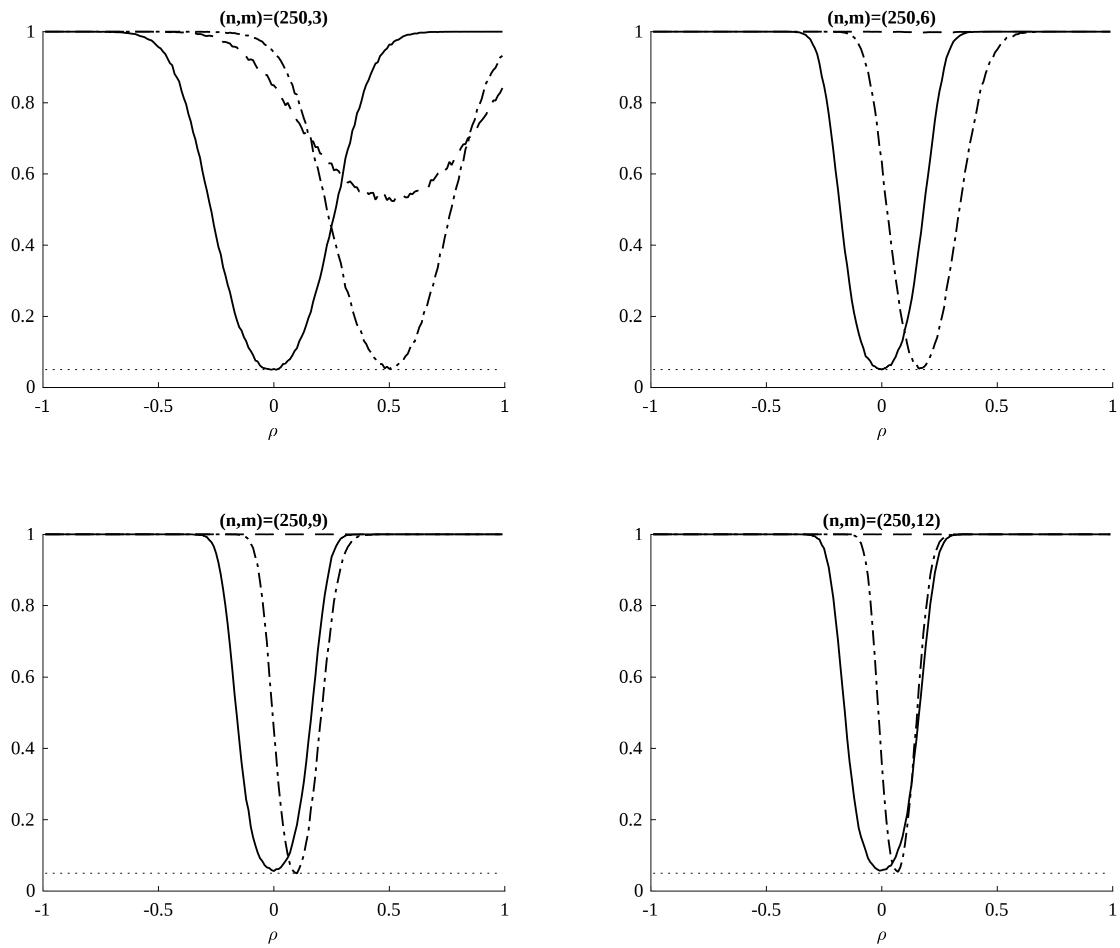

Figure notes. Our test (solid line), the Inoue-Solon test (dashed line), and the Wooldridge-Drukker test (dashed-dotted line). The size of the tests (.05) is indicated by a horizontal dotted line.

Table 1: Descriptive statistics

Variable	mean	sd	min	max
weight	3,489	536.7	290	5,925
smokes	0.112	0.316	0	1
n_cigarettes	1.837	8.237	0	99
age	28.67	5.349	14	46
male	0.514	0.500	0	1
novisit	0.0114	0.106	0	1
pretri2	0.134	0.340	0	1
pretri3	0.0258	0.159	0	1
adeqcode2	0.192	0.394	0	1
adeqcode3	0.0541	0.226	0	1

that aim to measure the extent to which the mother took adequate prenatal care. These are `novisit`, a dummy that switches on if no prenatal visit occured, `pretri2` and `pretri3`, which state that prenatal visits occured in the second and third trimester, respectively, and `adeqcode2` and `adeqcode3`, which indicate whether the so-called Kessner index equals two or three, respectively. The Kessner index is a categorical measure for adequacy of prenatal care which is based upon length of gestation, number of prenatal visits and date of initial prenatal visit. The three categories of the Kessner index are 'adequate' (a value of one), 'intermediate' (a value of two) and 'inadequate' (a value of three).

Table 2 summarizes the results of fitting a standard random-effect and a fixed-effect model to these data. The sign on the coefficients is as expected. Most notably, smoking has a substantially negative impact on birthweight. The table also provides p-values for the null that there is no unobserved mother heterogeneity. Under both specifications this null is strongly rejected. Furthermore, both the random-effect and the fixed-effect model associate a large fraction of the variation in birthweight to latent heterogeneity at the level of the mother, 39% and 49%, respectively.

In these data our test statistic for the null of no within-mother correlation tests two moment restrictions. The statistic (based on the fixed-effect estimator) takes the value

Table 2: Regression results

Dependent variable:	birthweight	
Independent variable	RE	FE
smokes	-241.3	-153.3
	(11.23)	(18.43)
n_cigarettes	0.00565	-0.471
	(0.381)	(0.493)
male	134.8	135.8
	(4.741)	(5.158)
age	36.73	23.32
	(4.773)	(7.305)
agesq	-0.373	-0.103
	(0.0822)	(0.122)
novisit	12.56	15.93
	(30.73)	(39.27)
adeqcode2	-79.85	-69.28
	(9.435)	(10.97)
adeqcode3	-168.8	-116.1
	(20.10)	(24.05)
pretri2	72.01	52.09
	(11.06)	(12.53)
pretri3	144.6	109.3
	(24.72)	(29.21)
no heterogeneity	.0000	.0000
between contr.	.3922	.4996

Table notes. Standard errors are in parentheses. for the fixed-effect estimator these are clustered at the level of the mother. The entries for 'no heterogeneity' are p-values for the null of no mother-specific unobserved heterogeneity. The entries for 'between contr.' give the estimated fraction of the variance that is due to between heterogeneity.

112.64 whicj, with a p-value that is zero up to 24 decimal digits, gives very strong evidence against our null. (The test of Inoue and Solon (2006) gives a p-value of .059.) Consequently, additional unobserved heterogeneity that is persistent across births but is not captured by the inclusion of mother-specific effects seems present in these data. One implication of this is that the attribution of half of the variation in birthweight to mother heterogeneity is not theoretically justified. Another consequence is that the random-effect model is misspecified. More importantly, however, the use of a fixed-effect strategy to identify the causal effect of smoking on infant health need not be sufficient to solve the omitted-variable problem described by Abrevaya (2006).

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 $\sum_{g=1}^{n} \boldsymbol{v}_g/\sqrt{n} \overset{L}{\to} \mathrm{N}(\mathbf{0}, \boldsymbol{V})$ and that $n^{-1}\sum_{g=1}^{n} \boldsymbol{v}_g \boldsymbol{v}_g' \overset{P}{\to} \boldsymbol{V}$. Hence, $\boldsymbol{z} := \left(\sum_{g=1}^{n} \boldsymbol{v}_g \boldsymbol{v}_g'\right)^{-1/2} \sum_{g=1}^{n} \boldsymbol{v}_g \overset{L}{\to} \mathrm{N}(\mathbf{0}, \boldsymbol{I}_r)$, and so $s_n = \boldsymbol{z}'\boldsymbol{z} \overset{L}{\to} \chi^2(r, 0)$. This is Part (i) of the theorem. Under the sequence of local alternatives $\mathbb{E}(\boldsymbol{v}_g) = \boldsymbol{\delta}/\sqrt{n}$ the limit distribution of $\boldsymbol{z}$ contains an asymptotic bias term. Moreover, now, $\boldsymbol{z} \overset{L}{\to} \mathrm{N}(\boldsymbol{V}^{-1/2}\boldsymbol{\delta}, \boldsymbol{I}_r)$, and so $s_n = \boldsymbol{z}'\boldsymbol{z} \overset{L}{\to} \chi^2(r, \boldsymbol{\delta}'\boldsymbol{V}^{-1}\boldsymbol{\delta})$. This is Part (ii) of the theorem. $\square$

Proof of Theorem 2. The main difference with the proof of Theorem 1 is accounting for the estimation noise in $\hat{\boldsymbol{\beta}}$. We first prove (2.7). Because

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

$\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|^2 = O_P(n^{-1})$ by (2.6), and the covariates have finite second moments, the expansion

$$\sum_{g=1}^{n} \hat{\boldsymbol{v}}_g = \sum_{g=1}^{n} \boldsymbol{v}_g + \sum_{g=1}^{n} (\partial \boldsymbol{v}_g/\partial \boldsymbol{\beta}')\,(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + O_P(1)$$

holds. Further, as $\partial \boldsymbol{v}_g / \partial \boldsymbol{\beta}' = \dot{\boldsymbol{H}}_g'(\boldsymbol{I}_p \otimes \boldsymbol{D}\, \boldsymbol{\eta}_g) - \boldsymbol{H}_g' \boldsymbol{D}\, \boldsymbol{X}_g$, the moment conditions postulated in the theorem allow the application of a law of large numbers to establish that

$$\sum_{g=1}^{n} \frac{\partial \boldsymbol{v}_g / \partial \boldsymbol{\beta}'}{n} \xrightarrow{P} \boldsymbol{\Omega};$$

(2.7) then follows. Next, again by the moment requirements on $\eta_{g,i}$ and $\boldsymbol{\omega}_g$, the law of large numbers yields

$$\sum_{g=1}^{n} \frac{(\boldsymbol{v}_g + \boldsymbol{\Omega}\boldsymbol{\omega}_g)(\boldsymbol{v}_g + \boldsymbol{\Omega}\boldsymbol{\omega}_g)'}{n} \xrightarrow{P} \tilde{\boldsymbol{V}}.$$

Consequently, with

$$\boldsymbol{z} := \left(\sum_{g=1}^{n} (\boldsymbol{v}_g + \boldsymbol{\Omega}\boldsymbol{\omega}_g)(\boldsymbol{v}_g + \boldsymbol{\Omega}\boldsymbol{\omega}_g)' \right)^{-1/2} \sum_{g=1}^{n} \hat{\boldsymbol{v}}_g$$

$\boldsymbol{z}'\boldsymbol{z}$ will follow a (non-central) χ^2-distribution with r degrees of freedom. The non-centrality parameter is zero under the null of no within-group correlation and $\tilde{\boldsymbol{\delta}}' \tilde{\boldsymbol{V}}^{-1} \tilde{\boldsymbol{\delta}}$ under local alternatives. It remains only to show that

$$\sum_{g=1}^{n} \frac{(\hat{\boldsymbol{v}}_g + \boldsymbol{O}\hat{\boldsymbol{\omega}}_g)(\hat{\boldsymbol{v}}_g + \boldsymbol{O}\hat{\boldsymbol{\omega}}_g)'}{n} - \frac{(\boldsymbol{v}_g + \boldsymbol{\Omega}\boldsymbol{\omega}_g)(\boldsymbol{v}_g + \boldsymbol{\Omega}\boldsymbol{\omega}_g)'}{n} \xrightarrow{P} 0,$$

so that the distributional approximations carry over to the feasible statistic $\hat{s}_n$. Given that $n^{-1} \sum_{g=1}^{n} \|\hat{\boldsymbol{\omega}}_g - \boldsymbol{\omega}_g\|^2 \xrightarrow{P} 0$ by assumption it suffices to prove that $\boldsymbol{O} \xrightarrow{P} \boldsymbol{\Omega}$ and that $n^{-1} \sum_{g=1}^{n} \|\hat{\boldsymbol{v}}_g - \boldsymbol{v}_g\|^2 \xrightarrow{P} 0$. First,

$$\boldsymbol{O} = \sum_{g=1}^{n} \frac{\dot{\boldsymbol{H}}_g'(\boldsymbol{I}_p \otimes \boldsymbol{D}\, \boldsymbol{e}_g)}{n} - \sum_{g=1}^{n} \frac{\boldsymbol{E}_g' \boldsymbol{D}\, \boldsymbol{X}_g}{n}.$$

From above, $\boldsymbol{e}_g = \boldsymbol{\eta}_g - \boldsymbol{X}_g\, (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$, from which, up to $O_P(n^{-1})$,

$$\boldsymbol{O} = \sum_{g=1}^{n} \frac{\partial \boldsymbol{v}_g / \partial \boldsymbol{\beta}'}{n} + \sum_{g=1}^{n} \frac{\dot{\boldsymbol{H}}_g'(\boldsymbol{I}_p \otimes \boldsymbol{D}\, \boldsymbol{X}_g)}{n} \left(\boldsymbol{I}_p \otimes (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})\right) - \left((\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \otimes \boldsymbol{I}_p\right)' \sum_{g=1}^{n} \frac{\dot{\boldsymbol{H}}_g' \boldsymbol{D}\, \boldsymbol{X}_g}{n}$$

follows by re-arrangement. Then

$$\boldsymbol{O} = \sum_{g=1}^{n} \frac{\partial \boldsymbol{v}_g / \partial \boldsymbol{\beta}'}{n} + o_P(1) \xrightarrow{P} \boldsymbol{\Omega}$$

is obtained on noting that the sample averages on the righ-hand side of the expression all converge in probability to finite quantities and that $\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\| \xrightarrow{P} 0$. Next,

$$\sum_{g=1}^{n} \frac{\|\hat{\boldsymbol{v}}_g - \boldsymbol{v}_g\|^2}{n} = \sum_{g=1}^{n} \frac{\|(\partial \boldsymbol{v}_g/\partial \boldsymbol{\beta}')\,(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})\|^2}{n} + o_P(1) \leq O_P(n^{-1}) \sum_{g=1}^{n} \frac{\|(\partial \boldsymbol{v}_g/\partial \boldsymbol{\beta}')\|^2}{n} + o_P(1).$$

With $\boldsymbol{i}_q$ denoting the p-dimensional unit vector in direction $q = 1, \ldots, p$ we can conveniently write

$$\|(\partial \boldsymbol{v}_g/\partial \boldsymbol{\beta}')\|^2 = \operatorname{trace}\big((\partial \boldsymbol{v}_g/\partial \boldsymbol{\beta}')'(\partial \boldsymbol{v}_g/\partial \boldsymbol{\beta}')\big) = \sum_{q=1}^{p} \|(\partial \boldsymbol{H}_g/\partial \boldsymbol{\beta}_q)' \boldsymbol{D} \boldsymbol{\eta}_g - \boldsymbol{H}_g' \boldsymbol{D} \boldsymbol{X}_g \boldsymbol{i}_q\|^2.$$

For all $q = 1, \ldots, p$ we have

$$\mathbb{E}(\boldsymbol{\eta}_g' \boldsymbol{D}' (\partial \boldsymbol{H}_g/\partial \boldsymbol{\beta}_q)(\partial \boldsymbol{H}_g/\partial \boldsymbol{\beta}_q)' \boldsymbol{D} \boldsymbol{\eta}_g) \leq \max_{i_1, i_2} \mathbb{E}((\Delta \eta_{g,i_1})^2\,(\boldsymbol{i}_q' \boldsymbol{x}_{g,i_2})^2) = O(1)$$

and

$$\mathbb{E}(\boldsymbol{i}_q' \boldsymbol{X}_g' \boldsymbol{D}' \boldsymbol{H}_g \boldsymbol{H}_g' \boldsymbol{D} \boldsymbol{X}_g \boldsymbol{i}_q) \leq \max_{i_1, i_2} \mathbb{E}((\eta_{g,i_1})^2\,(\boldsymbol{i}_q' \Delta \boldsymbol{x}_{g,i_2})^2) = O(1)$$

by Cauchy-Schwarz because $\varepsilon_{g,i}$, α_g and $\boldsymbol{x}_{g,i}$ all have finite fourth-order moments. We then have that

$$\sum_{g=1}^{n} \frac{\|(\partial \boldsymbol{v}_g/\partial \boldsymbol{\beta}')\|^2}{n} = O_P(1)$$

from which $\sum_{g=1}^{n} \|\hat{\boldsymbol{v}}_g - \boldsymbol{v}_g\|^2/n \xrightarrow{P} 0$ follows. The proof is thus complete. □

References

Abrevaya, J. (2006). Estimating the effect of smoking on birth outcomes using a matched panel data analysis. *Journal of Applied Econometrics 21*, 489–519.

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.

Cox, D. R. and P. J. Solomon (1988). On testing for serial correlation in large numbers of small samples. *Biometrika 75*, 145–148.

Currie, J. and J. Gruber (1996). Saving babies: The efficacy and cost of recent changes in the Medicaid eligibility of pregnant women. *Journal of Political Economy 104*, 457–470.

Drukker, D. M. (2003). Testing for serial correlation in linear panel-data models. *Stata Journal 3*, 168–177.

Godfrey, L. G. (1981). On the invariance of the Lagrange multiplier test with respect to certain changes in the alternative hypothesis. *Econometrica 49*, 1443–1455.

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.

Kline, P., R. Saggio, and M. Sølvsten (2018). Leave-out estimation of variance components. Mimeo.

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

Ljung, G. M. and G. E. P. Box (1978). On a measure of a lack of fit in time series models. *Biometrika 65*, 297–303.

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.

Newey, W. K. (1985). Generalized method of moments specification testing. *Journal of Econometrics 29*, 229–256.

Neyman, J. and E. L. Scott (1948). Consistent estimates based on partially consistent observations. *Econometrica 16*, 1–32.

Orme, C. D. and T. Yamagata (2006). The asymptotic distribution of the F-test statistic for individual effects. *Econometrics Journal 9*, 404–422.

Orme, C. D. and T. Yamagata (2013). A heteroskedasticity-robust f-test statistic for individual effects.

Econometric Reviews 33, 431–471.

Permutt, T. and J. R. Hebel (1989). Simultaneous-equation estimation in a clinical trial of the effect of smoking on birth weight. *Biometrics 45*, 619–622.

Reichman, N. E. and M. J. Florio (1996). The effects of enriched prenatal care services on Medicaid birth outcomes in New Jersey. *Journal of Health Economics 15*, 455–476.

Stock, J. H. and M. W. Watson (2008). Heteroskedasticity-robust standard errors for fixed effects panel data regression. *Econometrica 76*, 155–174.

Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*. MIT Press.

Wooldridge, J. M. (2003). Cluster-sample methods in applied econometrics. *American Economic Review: Papers & Proceedings 93*, 133–138.

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