# A joint serial correlation test for linear panel data models

Yamagata, Takashi

# Accepted Manuscript


A joint serial correlation test for linear panel data models

Takashi Yamagata

PII: S0304-4076(08)00081-X
DOI: 10.1016/j.jeconom.2008.08.005
Reference: ECONOM 3051

To appear in: *Journal of Econometrics*

Received date: 2 October 2006
Revised date: 11 July 2008
Accepted date: 7 August 2008

# A Joint Serial Correlation Test for Linear Panel Data Models

Takashi Yamagata*
Department of Economics and Related Studies, University of York, Heslington, York, UK

July 2008

**Abstract**

This paper proposes a joint error serial correlation test to be applied to linear panel data models after generalised method of moments estimation. This new test is as an alternative inferential tool to both the $m_2$ test of Arellano and Bond (1991) and the overidentifying restrictions test. The proposed test, called the $m^2_{(2,p)}$ test, involves an examination of the joint significance of estimates of second to $p^{th}$-order (first differenced) error serial correlations. The small sample properties of the $m^2_{(2,p)}$ test are investigated by means of Monte Carlo experiments. The evidence shows that the proposed test mostly outperforms the conventional $m_2$ test and has high power when the overidentifying restrictions test does not, under a variety of alternatives including slope heterogeneity and cross section dependence.

**Key Words:** method of moments; dynamic panel data; serial correlation test; slope heterogeneity; cross section dependence; $m_2$ test; overidentifying restrictions test.
**JEL Classification:** C12, C15, C23, C52, H71

*Corresponding author. I am grateful to two anonymous referees, Leslie Godfrey, Chris Orme, Vasilis Sarafidis, and participants of Econometric Society European Meeting 2007 for their helpful comments.

# 1 Introduction

The use of generalised method of moments (GMM) estimation for linear panel data models has gained popularity over last decade. This method has been widely used in economic analysis, such as labour participation, cross-country growth convergence, government behaviour, among many others. It is well-known that the validity of moment restrictions is essential for GMM estimation. Sargan (1958) and Hansen (1982) proposed a test for the validity of the overidentifying restrictions, called the overidentifying restrictions test hereafter, which serves as a general misspecification test. Also, in linear dynamic panel models, existence of error serial correlation will invalidate subsets of moment restrictions. A second order first differenced error serial correlation test, called $m_2$ test, and the Sargan's difference test for error serial correlations, proposed by Arellano and Bond (1991), have become standard diagnostic tools in applied research.

Recently, two major concerns over the use of GMM estimation have been raised in the literature. First, ignorance of the heterogeneity of slope coefficients in dynamic linear panel data models will result in persistent residual serial correlation, leading to inconsistency of GMM estimator (see Pesaran and Smith, 1995). Second, if the model is subject to (heterogeneous) cross section dependence arising from unobserved common factors, again the GMM estimator will be inconsistent (see Holtz-Eakin, Newey and Rosen, 1988, Ahn, Lee and Schmidt, 2001, Robertson and Sarafidis, 2006, Sarafidis, Yamagata and Robertson, 2006). Importantly, these scenarios generally imply non-trivial higher order error serial correlation, resulting in invalidity of *all* moment restrictions. In addition, under more general $q^{th}$-order moving average or autoregressive errors, the $m_s$ test of Arellano (2003), $s = 1, 2, ..., p$, may not be powerful enough, while the Sargan's difference test may not be appropriate.

Under such misspecifications, the overidentifying restrictions test is expected to reject the null hypothesis with probability tending to one as the sample size tends to infinity. However, recent literature contains reports on its poor finite sample behaviour, especially when there are relatively many overidentifying restrictions. The finite sample evidence of Bowsher (2002) and Windmeijer (2005) suggests that the overidentifying restrictions test tends to reject the null too infrequently, unless the time series dimension is very small. Accordingly, the power of the overidentifying restrictions test with finite sample can be very low. Furthermore, Windmeijer (2005) reports that the use of the infeasible weighting matrix (using the unknown true parameter) fails to improve the finite sample performance of the overidentifying restrictions test. Bond and Windmeijer (2005) illustrate that the bootstrap overidentifying restrictions tests, based on the bootstrap method proposed by Hall and Horowitz (1996) and Brown and Newey (2002), generally have inferior finite sample performance to the asymptotic tests.

In view of this, the current paper proposes a joint test for the second to $p^{th}$-order first-differenced error serial correlation, called the $m^2_{(2,p)}$ test, which can serve as an alternative misspecification test. This test has not been appeared in the existing literature.[1] The asymptotic local power of the $m^2_{(2,p)}$ test is investigated, which yields two main results. First, AR($q$) and MA($q$) errors are locally equivalent alternatives in Godfrey's (1981) sense. This implies that the rejection of the null hypothesis by the $m^2_{(2,p)}$ test may not help to indicate whether the errors are MA($q$) or AR($q$). Second, the asymptotic power of an overspecified $m^2_{(2,p)}$ test can be higher than that of the $m^2_{(2,q+1)}$ test. This implies

[1]See Inoue and Solon (2006) for a portmanteau test for serial correlation in a classical fixed effects model, in which the regressors are strictly exogenous.

that the power of the proposed joint serial correlation test can be higher than that of the conventional $m_2$ test, under the varieties of alternatives such as AR($q$) and MA($q$) errors, slope heterogeneity, and cross section dependence.

The small sample properties of the $m^2_{(2,p)}$ test with $p > 2$ will be compared to those of the $m_2$ test and the overidentifying restrictions test by means of Monte Carlo experiments. The evidence shows that the proposed test often outperforms the $m_2$ test under the varieties of alternatives, such as AR(1), AR(2) and MA(2) errors, slope heterogeneity, and error cross section dependence. In the case of MA(1) error, the joint test and the $m_2$ test have very similar power estimates. Importantly, the $m^2_{(2,p)}$ test with the maximum $p$ available has high power where the overidentifying restrictions test does not.

Section 2 contains a discussion of the model and the estimation method. The existing tests are reviewed in Section 3. Section 4 proposes the joint serial correlation test, $m^2_{(2,p)}$ test, then discusses its power properties under the various alternatives. The finite sample evidence is reported in Section 5, and Section 6 contains some concluding remarks.

## 2 Model and Estimation Method

Consider the following model

$$y_{it} = \alpha_i + \lambda y_{i,t-1} + \boldsymbol{\beta}'\mathbf{x}_{it} + u_{it},\ i = 1, 2, ..., N,\ t = 2, 3, ..., T, \tag{1}$$

where $\alpha_i$ is an individual effect with finite mean and finite variance, $|\lambda| < 1$, $\boldsymbol{\beta}$ is a $(K \times 1)$ parameter vector which is bounded, $\mathbf{x}_{it} = (x_{1it}, x_{2it}, ..., x_{Kit})'$ is a $(K \times 1)$ vector of predetermined regressors such that $E(\mathbf{x}_{is}u_{it}) \neq 0$ for $s > t$, zero otherwise. First differencing (1) gives

$$\Delta y_{it} = \lambda \Delta y_{i,t-1} + \boldsymbol{\beta}'\Delta\mathbf{x}_{it} + \Delta u_{it},\ i = 1, 2, ..., N,\ t = 3, 4, ..., T, \tag{2}$$

where $\Delta y_{it} = y_{it} - y_{it-1}$, $\Delta\mathbf{x}_{it} = \mathbf{x}_{it} - \mathbf{x}_{it-1}$, $\Delta u_{it} = u_{it} - u_{it-1}$. For further discussion, stacking (1) for each $i$ yields

$$\mathbf{y}_i = \alpha_i \boldsymbol{\iota}_{T-1} + \lambda \mathbf{y}_{i,-1} + \mathbf{X}_i\boldsymbol{\beta} + \mathbf{u}_i,\ i = 1, 2, ..., N, \tag{3}$$

where $\mathbf{y}_i = (y_{i2}, y_{i3}, ..., y_{iT})'$, $\boldsymbol{\iota}_g$ is a $(g \times 1)$ vector of unity with natural number $g$, $\mathbf{y}_{i,-1} = (y_{i1}, y_{i2}, ..., y_{iT-1})'$, $\mathbf{X}_i = (\mathbf{x}_{i2}, \mathbf{x}_{i3}, ..., \mathbf{x}_{iT})'$, $\mathbf{u}_i = (u_{i2}, u_{i3}, ..., u_{iT})'$. The matrix version of the first differenced equation is defined by

$$\Delta\mathbf{y}_i = \lambda \Delta\mathbf{y}_{i,-1} + \Delta\mathbf{X}_i\boldsymbol{\beta} + \Delta\mathbf{u}_i,\ i = 1, 2, ..., N, \tag{4}$$

where $\Delta\mathbf{y}_i = (\Delta y_{i3}, \Delta y_{i4}, ..., \Delta y_{iT})'$, $\Delta\mathbf{y}_{i,-1} = (\Delta y_{i2}, \Delta y_{i3}, ..., \Delta y_{iT-1})'$, $\Delta\mathbf{X}_i = (\Delta\mathbf{x}_{i3}, \Delta\mathbf{x}_{i4}, ..., \Delta\mathbf{x}_{iT})'$, $\Delta\mathbf{u}_i = (\Delta u_{i3}, \Delta u_{i4}, ..., \Delta u_{iT})'$, or

$$\Delta\mathbf{y}_i = \Delta\mathbf{W}_i\boldsymbol{\theta} + \Delta\mathbf{u}_i, \tag{5}$$

where $\Delta\mathbf{W}_i = (\Delta\mathbf{y}_{i,-1}, \Delta\mathbf{X}_i)$, $\boldsymbol{\theta} = (\lambda, \boldsymbol{\beta}')'$.

Define the matrix of instruments

$$\mathbf{Z}_i = \begin{bmatrix} \mathbf{Z}_{Yi} & \mathbf{Z}_{Xi} \end{bmatrix}\ (T - 2 \times h), \tag{6}$$

where $h = h_y + h_x$,

$$\mathbf{Z}_{Yi} = \begin{bmatrix} y_{i1} & 0 & 0 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & y_{i1} & y_{i2} & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & 0 & 0 & y_{i1} & \cdots & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & ... & ... & \vdots \\ 0 & 0 & 0 & 0 & \cdots & y_{i1} & \cdots & y_{iT-2} \end{bmatrix} (T-2 \times h_y), \tag{7}$$

where $h_y = (T-1)(T-2)/2$ and

$$\mathbf{Z}_{Xi} = \begin{bmatrix} \mathbf{x}'_{i1} & \mathbf{x}'_{i2} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \cdots & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{x}'_{i1} & \mathbf{x}'_{i2} & \mathbf{x}'_{i3} & \mathbf{0} & \cdots & \cdots & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{x}'_{i1} & \cdots & \cdots & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & ... & ... & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{x}'_{i1} & \cdots & \mathbf{x}'_{iT-1} \end{bmatrix} (T-2 \times h_x), \tag{8}$$

$h_x = K(T+1)(T-2)/2$. GMM estimation is based on the moment restrictions

$$E[\mathbf{Z}_i^{'} \Delta \mathbf{u}_i] = \mathbf{0}. \tag{9}$$

The Arellano-Bond two-step GMM estimator is defined as

$$\ddot{\boldsymbol{\theta}}_N = \left( \mathbf{A}'_N \ddot{\boldsymbol{\Omega}}_N^{-1} \mathbf{A}_N \right)^{-1} \mathbf{A}'_N \ddot{\boldsymbol{\Omega}}_N^{-1} \mathbf{b}_N,$$

where $\mathbf{A}_N = N^{-1} \sum_{i=1}^N \mathbf{Z}'_i \Delta \mathbf{W}_i$, $\mathbf{b}_N = N^{-1} \sum_{i=1}^N \mathbf{Z}'_i \Delta \mathbf{y}_i$, $\ddot{\boldsymbol{\Omega}}_N = N^{-1} \sum_{i=1}^N \mathbf{Z}'_i \Delta \dot{\mathbf{u}}_i \Delta \dot{\mathbf{u}}'_i \mathbf{Z}_i$ with $\Delta \dot{\mathbf{u}}_i = \Delta \mathbf{y}_i - \Delta \mathbf{W}_i \dot{\boldsymbol{\theta}}_N$, where $\dot{\boldsymbol{\theta}}_N$ is the one-step GMM estimator

$$\dot{\boldsymbol{\theta}}_N = \left( \mathbf{A}'_N \dot{\boldsymbol{\Omega}}_N^{-1} \mathbf{A}_N \right)^{-1} \mathbf{A}'_N \dot{\boldsymbol{\Omega}}_N^{-1} \mathbf{b}_N,$$

where $\dot{\boldsymbol{\Omega}}_N = N^{-1} \sum_{i=1}^N \mathbf{Z}'_i \mathbf{H} \mathbf{Z}_i$, $\mathbf{H}$ is a $(T-2 \times T-2)$ matrix, $(s,r)$ elements of which are $h_{s,r}$, where $h_{s,s} = 2$, $h_{s,s+1} = h_{s+1,s} = -1$, and $h_{sr} = 0$ for $|s-r| > 1$.

In order to proceed, the following assumptions are made:

**Assumption 1:** $\{\mathbf{y}_i^*, \mathbf{X}_i^*\}_{i=1}^N$ is a sequence of independently and identically distributed random matrices, where $\mathbf{y}_i^* = (y_{i1}, \mathbf{y}'_i)'$ and $\mathbf{X}_i^* = (\mathbf{x}_{i1}, \mathbf{X}'_i)'$.

**Assumption 2:**

**(i)** $u_{it}$ is independently and identically distributed, with mean zero and a strictly positive variance $\sigma^2$, and has a finite fourth order moment.

**(ii)** $E\left(u_{it} | y_{it-1}, y_{it-2}, ..., y_{i1}, \mathbf{x}_{it}, \mathbf{x}_{it-1}, ..., \mathbf{x}_{i1}, \alpha_i\right) = 0$, $t = 2, 3, ..., T$.

**(iii)** The coefficient on the lagged dependent variable satisfies $|\lambda| < 1$.

**Assumption 3:** $rank\left(E[\mathbf{Z}'_i \Delta \mathbf{W}_i]\right) = K + 1$.

**Assumption 4:** $\mathbf{M} = E\left(\boldsymbol{\kappa}_i \boldsymbol{\kappa}'_i\right)$ is a $(p+h-1 \times p+h-1)$ symmetric and positive definite matrix, where $\boldsymbol{\kappa}_i = \left(\boldsymbol{\eta}'_i, (\mathbf{Z}'_i \Delta \mathbf{u}_i)'\right)'$ with $\boldsymbol{\eta}_i = \left(\eta_{i2}, \eta_{i3}, ..., \eta_{ip}\right)'$, $\eta_{is} = \Delta u_{it} \Delta u_{it+s}$, $s = 2, 3, ..., p$.

Assumption 1 is required in order to apply the standard *iid* central limit theorem later. It can be relaxed to the "independently but not necessarily identically distributed" case. The stronger assumption is employed for the ease of the exposition. Assumption 2(i) excludes heteroskedastic time series. Assumption 2(ii) concerns sequential moment conditions, which imply $E(\alpha_i u_{it}) = 0$, $E(y_{i1}u_{it}) = 0$, $E(\mathbf{x}_{i1}u_{it}) = \mathbf{0}$, $t = 2, 3, ..., T$. Assumption 2(iii) assures the stability of $y_{it}$ process; see Arellano (2003), for example. Assumption 3 is an identification condition and Assumption 4 ensures that the test statistic proposed later has a chi-square distribution with $p - 1$ degrees of freedom, asymptotically.

# 3 Existing Tests

The standard serial correlation tests in dynamic linear GMM models are the $m_2$ test and Sargan's difference test, both of which are proposed in Arellano and Bond (1991).[2] Slightly more general versions of these test statistics are discussed below.

## 3.1 The $m_s$ Test

As a generalisation of the $m_2$ test, Arellano (2003; p.121-123) proposes the $m_s$ statistics, $s = 1, 2, .., p$, with $p \leq T - 3$, which are intended to detect particular orders of first differenced error serial correlation. The hypotheses of interest are $H_0 : E(\Delta u_{it}\Delta u_{it+s}) = 0$ against $H_1 : E(\Delta u_{it}\Delta u_{it+s}) \neq 0$. The $m_s$ test statistic is defined as

$$m_s = \frac{1}{\sqrt{N\ddot{v}_s^2}} \sum_{i=1}^{N} \ddot{\eta}_{is}, \tag{10}$$

where

$$\ddot{\eta}_{is} = \sum_{t=3}^{T-s} \Delta\ddot{u}_{it}\Delta\ddot{u}_{it+s}, \tag{11}$$

with $\Delta\ddot{u}_{it} = \Delta y_{it} - \Delta\mathbf{w}_{it}'\ddot{\boldsymbol{\theta}}_N$, and

$$\ddot{v}_s^2 = \left(N^{-1}\sum_{i=1}^{N}\ddot{\eta}_{is}^2\right) + \ddot{\boldsymbol{\omega}}_{Ns}'\ddot{\mathbf{Q}}_N^{-1}\ddot{\boldsymbol{\omega}}_{Ns} - 2\ddot{\boldsymbol{\omega}}_{Ns}'\ddot{\mathbf{Q}}_N^{-1}\mathbf{A}_N'\ddot{\boldsymbol{\Omega}}_N^{-1}\left(N^{-1}\sum_{i=1}^{N}\mathbf{Z}_i'\Delta\dot{\mathbf{u}}_i\ddot{\eta}_{is}\right),$$

where

$$\ddot{\boldsymbol{\omega}}_{Ns} = N^{-1}\sum_{i=1}^{N}\left(\sum_{t=3}^{T-s}\Delta\ddot{u}_{it}\Delta\mathbf{w}_{i,t+s}\right), \tag{12}$$

$$\ddot{\mathbf{Q}}_N = \mathbf{A}_N'\ddot{\boldsymbol{\Omega}}_N^{-1}\mathbf{A}_N. \tag{13}$$

Under the null hypothesis, $m_s \xrightarrow{d} N(0, 1)$, as $N \to \infty$ with $T$ fixed. The $m_s$ test is designed to be powerful against $s^{th}$ order first-differenced error serial correlation. Therefore, it may not have enough power against more general higher order serial correlations. In this paper, we focus on the $m_2$ test statistic, since it represents the properties of the $m_s$ statistics and is one of the most frequently reported test statistics in the empirical literature.

[2] We do not examine the Hausman (1978) test approach, which is considered by Arellano and Bond (1991), since it will be asymptotically equivalent to the Sargan's difference test; see Newey (1985) and Hayashi (2000).

### 3.2 Sargan's Difference Test

The Sargan's difference test is designed to check the validity of subsets of moment restrictions. Unlike the $m_s$ statistics, Sargan's difference test statistics can be applied as joint error serial correlation tests, such as $H_0 : E(\Delta u_{it}\Delta u_{it+s}) = 0$ for all $s = 2, 3, ..., p$.[3] To motivate the Sargan's difference test, suppose the alternative is the MA($q$) error, $u_{it} = \sum_{\ell=0}^{q} \psi_\ell \varepsilon_{it-\ell}$ with $\psi_0 = 1$. Then, decompose the matrix of instruments $\mathbf{Z}_i$ into two subsets

$$\mathbf{Z}_i = \left[ \underset{(T-2\times h_1)}{\mathbf{Z}_{1i}} , \underset{(T-2\times h_2)}{\mathbf{Z}_{2i}} \right], \tag{14}$$

such that, under the null hypothesis of no error serial correlation $E[\mathbf{Z}'_{1i}\Delta\mathbf{u}_i] = \mathbf{0}$ and $E[\mathbf{Z}'_{2i}\Delta\mathbf{u}_i] = \mathbf{0}$, but under the alternative hypothesis of MA($q$) error $E[\mathbf{Z}'_{1i}\Delta\mathbf{u}_i] \neq \mathbf{0}$ but $E[\mathbf{Z}'_{2i}\Delta\mathbf{u}_i] = \mathbf{0}$. For example, testing against the alternative of MA(1) errors, $\mathbf{Z}_{1i}$ consists of $y_{it-2}$, $\mathbf{x}_{it-1}$ and $\mathbf{x}_{it-2}$, and $\mathbf{Z}_{2i}$ consists of the additional lagged instruments. Then, the Sargan's difference test is defined as

$$SD = S(\ddot{\boldsymbol{\theta}}_N) - S(\ddot{\boldsymbol{\theta}}_{N2}) \tag{15}$$

where

$$S(\ddot{\boldsymbol{\theta}}_N) = \left( N^{-1/2} \sum_{i=1}^{N} \Delta\ddot{\mathbf{u}}'_i \mathbf{Z}_i \right) \ddot{\boldsymbol{\Omega}}_N^{-1} \left( N^{-1/2} \sum_{i=1}^{N} \mathbf{Z}'_i \Delta\ddot{\mathbf{u}}_i \right) \tag{16}$$

with $\Delta\ddot{\mathbf{u}}_i = \Delta\mathbf{y}_i - \Delta\mathbf{W}_i\ddot{\boldsymbol{\theta}}_N$, and

$$S(\ddot{\boldsymbol{\theta}}_{2N}) = \left( N^{-1/2} \sum_{i=1}^{N} \Delta\ddot{\mathbf{u}}'_{2i} \mathbf{Z}_{2i} \right) \ddot{\boldsymbol{\Omega}}_{2N}^{-1} \left( N^{-1/2} \sum_{i=1}^{N} \mathbf{Z}'_{2i} \Delta\ddot{\mathbf{u}}_{2i} \right) \tag{17}$$

with $\Delta\ddot{\mathbf{u}}_{2i} = \Delta\mathbf{y}_i - \Delta\mathbf{W}_i\ddot{\boldsymbol{\theta}}_{2N}$, where

$$\ddot{\boldsymbol{\theta}}_{2N} = \left( \mathbf{A}'_{2N} \ddot{\boldsymbol{\Omega}}_{2N}^{-1} \mathbf{A}_{2N} \right)^{-1} \mathbf{A}'_{2N} \ddot{\boldsymbol{\Omega}}_{2N}^{-1} \mathbf{b}_{2N},$$

$\mathbf{A}_{2N} = N^{-1}\sum_{i=1}^{N} \mathbf{Z}'_{2i}\Delta\mathbf{W}_i$, $\mathbf{b}_{2N} = N^{-1}\sum_{i=1}^{N} \mathbf{Z}'_{2i}\Delta\mathbf{y}_i$, $\ddot{\boldsymbol{\Omega}}_{2N}$ is based on the corresponding one-step GMM estimator based only on the instruments $\mathbf{Z}_{2i}$. Under the null hypothesis of $E[\mathbf{Z}'_{1i}\Delta\mathbf{u}_i] = \mathbf{0}$, $SD \xrightarrow{d} \chi^2(h_1)$. For later usage, $S(\ddot{\boldsymbol{\theta}}_N) \xrightarrow{d} \chi^2(h - K - 1)$ under the null.

The drawback of Sargan's difference test for testing general error serial correlations is the requirement of $h_2 \geq K + 1$ valid moment restrictions under the alternatives. Clearly, first order autoregressive error, for instance, does not allow this requirement to be satisfied.

In the next section, a joint error serial correlation test will be proposed, which is designed to detect higher order error serial correlations. Its power properties are analysed.

# 4 The $m^2_{(2,p)}$ Test for Error Serial Correlation

The hypotheses of our interest are

$$H_0 : E(\Delta u_{it}\Delta u_{it+s}) = 0 \text{ jointly for } s = 2, 3, ..., p(\leq T - 3) \tag{18}$$

[3] Arellano and Bond (1991) proposed the use of Sargan's difference test for testing against MA(1) error.

against

$$H_1 : E(\Delta u_{it}\Delta u_{it+s}) \neq 0, \text{ for some } s, \tag{19}$$

$t = 3, 4, ..., T-2$. We define a joint test statistic for second to $p^{th}$ order error serial correlation, for $T \geq 5$, as

$$m^2_{(2,p)} = \boldsymbol{\iota}'_N \ddot{\mathbf{H}} \left(\ddot{\mathbf{G}}'\ddot{\mathbf{G}}\right)^{-1} \ddot{\mathbf{H}}'\boldsymbol{\iota}_N, \tag{20}$$

where $\boldsymbol{\iota}_N$ is a $(N \times 1)$ vector of ones, $\ddot{\mathbf{H}} = (\ddot{\boldsymbol{\eta}}_1, \ddot{\boldsymbol{\eta}}_2, ..., \ddot{\boldsymbol{\eta}}_N)'$, $\ddot{\boldsymbol{\eta}}_i = (\ddot{\eta}_{i2}, ..., \ddot{\eta}_{ip})'$, $\ddot{\eta}_{is}$ is defined by (11), $\ddot{\mathbf{G}} = (\ddot{\mathbf{g}}_1, \ddot{\mathbf{g}}_2, ..., \ddot{\mathbf{g}}_N)'$, $\ddot{\mathbf{g}}_i = (\ddot{g}_{i2}, ..., \ddot{g}_{ip})'$, $\ddot{g}_{is} = \ddot{\eta}_{is} - \ddot{\boldsymbol{\omega}}'_{Ns}\ddot{\mathbf{Q}}_N^{-1}\mathbf{A}'_N\ddot{\boldsymbol{\Omega}}_N^{-1}\mathbf{Z}'_i\Delta\dot{\mathbf{u}}_i$, where $\ddot{\boldsymbol{\omega}}_{Ns}$ and $\ddot{\mathbf{Q}}_N$ are as defined by (12) and (13), respectively.[4] Now the theorem can be stated as:

**Theorem 1** *Consider the panel data model (1). Suppose Assumptions 1-4 hold. Then under the null hypothesis (18),*

$$m^2_{(2,p)} \xrightarrow{d} \chi^2(p-1), \tag{21}$$

*as $N \to \infty$ with fixed $T$, where $m^2_{(2,p)}$ is defined by (20).*[5]

See Appendix A for a proof.[6]

Observe that the Arellano's (2003) $m_s$ statistics for $s = 2, 3, ..., p$ is simply $\boldsymbol{\iota}'_N\ddot{\mathbf{h}}_{s-1}/\sqrt{N\ddot{v}_s^2}$, where $\ddot{\mathbf{h}}_{s-1}$ is the $(s-1)^{th}$ column of $\ddot{\mathbf{H}}$ and $N\ddot{v}_s^2$ is the $(s-1)^{th}$ diagonal element of $\ddot{\mathbf{G}}'\ddot{\mathbf{G}}$. Thus, the $m_2$ test and the $m^2_{(2,2)}$ test are equivalent.

### 4.1 Power Properties of the $m^2_{(2,p)}$ Test under Various Alternatives

As with any test for misspecification, the power properties of the $m^2_{(2,p)}$ test depends on the assumed alternative and the true data generation process; see, for example, Davidson and MacKinnon (1985). Initially the asymptotic power of the $m^2_{(2,p)}$ test under the local AR($q$) and MA($q$) errors are investigated. Discussions of the importance of the $m^2_{(2,p)}$ test under slope heterogeneity and error cross section dependence are then provided.

#### 4.1.1 Asymptotic Power Analysis under the Local AR($q$) and MA($q$) Errors

Without loss of generality, we focus on the panel AR(1) model specification, namely

$$y_{it} = \alpha_i + \lambda y_{i,t-1} + u_{it}, \; i = 1, 2, ..., N, \; t = 2, 3, ..., T, \tag{22}$$

[4] An alternative formulation, which is asymptotically equivalent to $m^2_{(2,p)}$, would be

$$\boldsymbol{\iota}'_N\ddot{\mathbf{G}}^*(\ddot{\mathbf{G}}^{*\prime}\ddot{\mathbf{G}}^*)^{-1}\ddot{\mathbf{G}}^{*\prime}\boldsymbol{\iota}_N$$

where the $(i, s)$ element of $\ddot{\mathbf{G}}^*$ is $\ddot{g}^*_{is} = \ddot{\eta}_{is} - \ddot{\boldsymbol{\omega}}'_{Ns}\ddot{\mathbf{Q}}_N^{-1}\mathbf{A}'_N\ddot{\boldsymbol{\Omega}}_N^{-1}\mathbf{Z}'_i\Delta\ddot{\mathbf{u}}_i$.

[5] The $m^2_{(2,p)}$ test based on Blundell and Bond (1998) GMM estimator could be easily constructed. However, to save the space, such a version is not considered in this paper.

[6] For unbalanced panel, $t = 1, 2, ..., T_i$, the joint test still can be computed, so long as $\min_{1\leq i\leq N} T_i \geq 5$. In this case, note that the cross section dimension depends on $s = 2, 3, ..., p$. Denoting this cross section dimension as $N_s$, it can be shown that $m^2_{(2,p)} \xrightarrow{d} \chi^2(p-1)$ as $\min_{2\leq s\leq p} N_s \to \infty$.

where $\alpha_i \sim iid(0, \sigma_\alpha^2)$, and, under no misspecification, $u_{it} \sim iid(0, \sigma^2)$. Also the $y_{it}$ process is assumed to be started long time ago. For simplicity, it is assumed that only the most recent lagged levels are used as instruments, namely, $\mathbf{Z}_i = diag(y_{i1}, y_{i2}, .., y_{iT-2})$.

An asymptotic expansion of $N^{-1/2} \sum_{i=1}^{N} \ddot{\boldsymbol{\eta}}_i$ around $\ddot{\lambda}_N = \lambda$, with all cross section averages replacing averages of expectations, yields

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \ddot{\boldsymbol{\eta}}_i = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \boldsymbol{\eta}_i - \bar{\boldsymbol{\omega}} \bar{Q}^{-1} \bar{\mathbf{a}}' \bar{\boldsymbol{\Omega}}^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \mathbf{Z}_i' \Delta \mathbf{u}_i + o_p(1), \tag{23}$$

where $\boldsymbol{\eta}_i = (\eta_{i2}, ..., \eta_{ip})'$, $\bar{\boldsymbol{\omega}} = (\bar{\omega}_2, ..., \bar{\omega}_p)'$, $\bar{\omega}_s = N^{-1} \sum_{i=1}^{N} \sum_{t=3}^{T-s} E(\Delta u_{it} \Delta y_{i,t+s})$, $\bar{Q} = \bar{\mathbf{a}}' \bar{\boldsymbol{\Omega}}^{-1} \bar{\mathbf{a}}$, $\bar{\mathbf{a}} = N^{-1} \sum_{i=1}^{N} E(\mathbf{Z}_i' \Delta \mathbf{y}_{i,-1})$, $\bar{\boldsymbol{\Omega}} = N^{-1} \sum_{i=1}^{N} E(\mathbf{Z}_i' \Delta \mathbf{u}_i \Delta \mathbf{u}_i' \mathbf{Z}_i)$.

**Asymptotic Local Equivalence** As with the results in Godfrey (1981) for the well-known joint Lagrange Multiplier (LM) serial correlation test, it turns out that AR($q$) and MA($q$) errors are asymptotically locally equivalent alternatives for the $m_{(2,p)}^2$ test, as below.

The natural alternatives which result in higher order error serial correlation may be MA($q$) errors

$$u_{it} = \sum_{\ell=1}^{q} \psi_\ell \varepsilon_{it-\ell} + \varepsilon_{it}, \tag{24}$$

where $|\psi_\ell| < \infty$, $\ell = 1, 2, ..., q$, as well as AR($q$) errors

$$u_{it} = \sum_{\ell=1}^{q} \rho_\ell u_{it-\ell} + \varepsilon_{it}, \tag{25}$$

$\varepsilon_{it} \sim iid(0, \sigma_\varepsilon^2)$, and it is assumed that the roots of $1 - \sum_{\ell=1}^{q} \rho_\ell z^\ell = 0$ lie strictly outside the unit circle. Observe that, as AR errors are persistent, a simple AR(1) errors results in higher order serial correlation, in a sense that $E\left(u_{it} u_{it+s}\right) \neq 0$, for $s > 1$.

Now consider local versions of MA($q$) and AR($q$) errors, namely $\psi_\ell = N^{-1/2} \delta_\ell$ in (24) and $\rho_\ell = N^{-1/2} \delta_\ell$ in (25), $\ell = 1, 2, ..., q$. It is assumed that $0 < |\delta_\ell| < \infty$, but satisfying stationarity condition of $u_{it}$ for given $N$, as above. Next define the $r^{th}$-order error autocovariance

$$\gamma_r = E(u_{it} u_{it+r}) = E(u_{it} u_{it-r}),\ r = 0, 1, ... \tag{26}$$

For both local AR($q$) and MA($q$) errors, $\gamma_r$ can be solved with respect to the parameters $\sigma_\varepsilon^2$ and $\delta_\ell$, $\ell = 1, 2, ..., q$,[7] and they are

$$\gamma_r = \begin{cases} \sigma^2 + o(N^{-1/2}), & \text{for } r = 0, \\ \sigma^2 \delta_r / \sqrt{N} + o(N^{-1/2}), & \text{for } r = 1, 2, ..., q, \\ o(N^{-1/2}), & \text{for } r > q. \end{cases}$$

Let the non-central chi-square distribution with $n$ degrees of freedom with non-centrality parameter $\zeta$ be denoted by $\chi^2(n, \zeta)$. Under these local alternatives,

$$m_{(2,p)}^2 \xrightarrow{d} \chi^2(p-1, \boldsymbol{\varphi}_p' \mathbf{V}_p^{-1} \boldsymbol{\varphi}_p), \tag{27}$$

[7] See, for example, Hamilton (1994).

where $\mathbf{V}_p$ is $\mathrm{plim}_{N\to\infty}(\ddot{\mathbf{G}}_p'\ddot{\mathbf{G}}_p/N)$, $\boldsymbol{\varphi}_p = \mathrm{plim}_{N\to\infty} N^{-1/2}\sum_{i=1}^N \ddot{\boldsymbol{\eta}}_i$, which is decomposed as

$$\boldsymbol{\varphi}_p = \mathbf{c}_p + \mathbf{d}_p, \tag{28}$$

where

$$\mathbf{c}_p = \sigma^2 \begin{pmatrix} (T-4)(2\delta_2 - \delta_3 - \delta_1) \\ \vdots \\ (T-q-1)(2\delta_{q-1} - \delta_q - \delta_{q-2}) \\ (T-q-2)(2\delta_q - \delta_{q-1}) \\ (T-q-3)(-\delta_q) \\ \mathbf{0}_{p-q-1} \end{pmatrix},$$

$$\mathbf{d}_p = \sigma^2 \bar{\boldsymbol{\varpi}} \bar{Q}^{-1} \bar{\mathbf{a}}' \bar{\boldsymbol{\Omega}}^{-1} \boldsymbol{\tau}_{T-2} \left( \sum_{\ell=2}^{q} \lambda^{\ell-2}\delta_\ell - \sum_{j=1}^{q} \lambda^{j-1}\delta_j \right),$$

with $\mathbf{0}_g$ being a $(g \times 1)$ vector of zeros.

This asymptotic local equivalence of the $m^2_{(2,p)}$ test statistic between AR($q$) and MA($q$) errors means that the $m^2_{(2,p)}$ test is powerful against both MA($q$) and AR($q$) alternatives, but that the rejection of the null hypothesis by the $m^2_{(2,p)}$ test may not help to indicate whether the errors are MA($q$) or AR($q$). Furthermore, the rejection of no error serial correlation merely means that the null hypothesis is not likely to be correct and does not necessarily mean that the test is in favour of particular alternatives (See Davidson and MacKinnon, 1993; p.364). An important implication of this result is that it might be a good idea to regard the $m^2_{(2,p)}$ test as a misspecification test, in a sense that the rejection of the null hypothesis by the test does not imply a particular model specification.[8] The same implication applies to the $m_2$ test, given its equivalence to the $m^2_{(2,2)}$ test.

Sargan's difference test for MA($q$) errors, defined by (15), uses precise information about the alternative, but the $m^2_{(2,p)}$ test does not. A question which may then arise is whether the Sargan's difference type test is more powerful than the $m^2_{(2,p)}$ test under MA($q$) errors. Arellano and Bond (1991) compare the power of the Sargan's difference test with that of the $m_2$ test under MA(1) error. They use a Monte Carlo experiment, the evidence from which shows that the $m_2$ test is more powerful than the Sargan's difference test.[9]

In equation (28), $\mathbf{c}_p$ reflects the asymptotic bias of $N^{-1/2}\sum_{i=1}^N \boldsymbol{\eta}_i$, and $\mathbf{d}_p$ is due to non-zero $\bar{\boldsymbol{\varpi}}$ and the asymptotic bias of $\sqrt{N}(\ddot{\lambda}_N - \lambda)$. Observe that only the first $q$ elements of $\mathbf{c}_p$ are non-zero and these elements affect power. In addition, all $p-1$ elements of the other component, $\mathbf{d}_p$, are non-zero, even though the magnitude of $\bar{\varpi}_s$ decays as $s$ increases. This finding might indicate that an overspecified joint test statistic, $m_{(2,p)}$ with $p > q+1$, may not lose much power comparing to just specified joint test statistic, $m_{(2,q+1)}$, under local AR($q$) or MA($q$) errors. This possibility will be investigated next.

**Some Local Power Comparison** Define the power function of the noncentral chi-square tests as

$$\pi_\alpha(n,\zeta) = \Pr\left(\chi^2(n,\zeta) > \chi^2_{n,\alpha}\right)$$

[8] See also Davidson and MacKinnon (1985) and Godfrey and Orme (1996).

[9] Godfrey (1981) examines the power of the LM test and the Likelihood Ratio (LR) test against the MA(1) alternative, and find that the LR test is more powerful, which uses precise information about the alternative. Note that the LR and LM test statistics in Godfrey (1981) are asymptotically equivalent, but the Sargan's difference test and the $m^2_{(2,2)}$ are not.

where $\alpha$ is the size of the test, such that $\Pr\left(\chi^2(n,0) > \chi^2_{n,\alpha}\right) = \alpha$. Das Gupta and Perlman (1974) claim that if we define

$$h(\zeta) = h(\zeta; n, \nu, \alpha) > 0,$$

for $\zeta > 0$, to be the unique value satisfying

$$\pi_\alpha(n,\zeta) = \pi_\alpha(n+\nu, \zeta + h(\zeta)), \tag{29}$$

where $\nu$ is a positive integer, then it is proved that $h(\zeta)$ is strictly increasing in $\zeta$. This result shows that the power of the noncentral chi-square test is strictly decreasing in the number of degrees of freedom.

We would like to compare the asymptotic power functions of $m^2_{(2,p)}$ and $m^2_{(2,q+1)}$ for $p > q+1$, using (29). By (27), the asymptotic power function of $m^2_{(2,p)}$ test statistic is $\pi_\alpha\left(p-1, \zeta\left(\delta_1^2, \delta_2^2, ..., \delta_q^2; T, \lambda, \sigma^2, \sigma_\alpha^2\right)\right)$. Since this is highly nonlinear in parameters, it seems impossible to obtain general results for local power comparison among the joint tests. Rather, we focus on the comparison of the local power functions of $m^2_{(2,2)}$ and $m^2_{(2,3)}$ test statistics under the AR(1) or MA(1) local alternatives, with the panel AR(1) model defined by (22) for $T = 6$, $\sigma^2 = 1$ and $\sigma^2_\alpha = 1$. In this case, it can be shown that

$$m^2_{(2,2)} \xrightarrow{d} \chi^2\left(1, \zeta_1\left(\delta_1^2; \lambda\right)\right), \tag{30}$$

$$m^2_{(2,3)} \xrightarrow{d} \chi^2\left(2, \zeta_2\left(\delta_1^2; \lambda\right)\right). \tag{31}$$

Given the value of $\lambda$, $\zeta_1\left(\delta_1^2;\lambda\right)$ and $\zeta_2\left(\delta_1^2;\lambda\right)$ become linear functions of $\delta_1^2$. As $\delta_1^2$ can take any finite non-negative value, what matters is the ratio $\xi(\lambda) = \zeta_2\left(\delta_1^2;\lambda\right)/\zeta_1\left(\delta_1^2;\lambda\right)$. Figure 1 reports the plot of the ratio $\xi(\lambda)$ for $-0.99 < \lambda < 0.99$, which is the range of interest.[10] The maximum of $\xi(\lambda)$ is 1.695 at $\lambda = 0.14$, and the minimum is 1.471 at $\lambda = -0.53$. Also, the local minimum where $\lambda$ is positive is 1.573 at $\lambda = 0.61$.

Table 1 provides the required value of $\delta_1^2$ to achieve the target power of $\pi_{0.05}\left(1,\delta^2\right) = \pi_{0.05}\left(2, \xi(\lambda)\delta^2\right) = 0.05$, 0.10, 0.20, 0.50, 0.90 and 0.95, at $\lambda = 0.14, 0.61, -0.53$. As can be seen, for all values of the target power, $m^2_{(2,3)}$ requires smaller values of $\delta^2$ than $m^2_{(2,2)}$, at the 5% significance level. This finding indicates that, at least with $T = 6, \sigma^2 = \sigma^2_\alpha = 1$, the proposed $m^2_{(2,3)}$ test achieves higher power than the $m^2_{(2,2)}$ test for $-1 < \lambda < 1$, under the local AR(1) or MA(1) errors. Moreover, since the $m^2_{(2,2)}$ test and the $m_2$ test are equivalent, the $m^2_{(2,3)}$ test is superior to the $m_2$ test, in terms of this asymptotic local power comparison, in this particular situation.[11]

Next we discuss power properties of the $m^2_{(2,p)}$ test under slope heterogeneity and cross section dependence.

#### 4.1.2 Slope Heterogeneity

The results of Pesaran and Smith (1995) imply that ignoring slope heterogeneity in the linear dynamic panel model may create persistent error serial correlation. Consider a slope heterogeneity version of the model (1)

$$y_{it} = \alpha_i + \lambda_i y_{i,t-1} + \boldsymbol{\beta}_i' \mathbf{x}_{it} + \varepsilon_{it},\ i = 1,2,...,N,\ t = 2,3,...,T, \tag{32}$$

[10] A note of derivation of this result is available from the author upon request.

[11] Small sample evidence on the power of these tests under non-local alternatives will be provided later.

where $\lambda_i = \lambda + \upsilon_{1i}$, $\upsilon_{1i} \sim iid(0, \sigma^2_{\upsilon 1})$, $\boldsymbol{\beta}_i = \boldsymbol{\beta} + \upsilon_{2i}$, $\upsilon_{2i} \sim iid(\mathbf{0}, \boldsymbol{\Sigma}_{\upsilon 2})$. Then the error of homogeneous model (1), $u_{it}$, can be written as

$$u_{it} = \upsilon_{1i} y_{i,t-1} + \boldsymbol{\upsilon}'_{2i} \mathbf{x}_{it} + \varepsilon_{it}. \tag{33}$$

It is clear that the error term is persistently serially correlated, and the regressors and the error term will be correlated. Thus, together with the local analysis in the case of AR($q$) and MA($q$) errors, the proposed joint serial correlation test, $m^2_{(2,p)}$ with $p > 2$, is likely to have higher power than the $m_2$ test. There does not seem to be a direct test of slope heterogeneity in dynamic linear panel models for large $N$ and fixed $T$ in the literature.[12] Therefore, the serial correlation test and the overidentifying restrictions test can be useful in playing the role of slope homogeneity test.

### 4.1.3 Cross Section Dependence

Ignorance of error cross section dependence also generate error serial correlation. Consider the multi-factor error structure of the model (1)

$$u_{it} = \boldsymbol{\phi}'_i \mathbf{f}_t + \varepsilon_{it}, \tag{34}$$

where $\boldsymbol{\phi}_i \sim iid(\mathbf{0}, \boldsymbol{\Sigma}_\phi)$, $E(\boldsymbol{\phi}_i \varepsilon_{jt}) = \mathbf{0}$ for all $i$, $j$, $t$, and $\mathbf{f}_t$ is a $(m \times 1)$ random vector which is distributed as $iid(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{f}})$. This type of error generates heterogeneous error cross section dependence, as discussed in Holtz-Eakin, Newey and Rosen (1988), Ahn, Lee and Schmidt (2001) and Sarafidis, Yamagata and Robertson (2006), among others. Taking expectations, after conditioning upon $\mathbf{f}_t$, yields

$$\begin{aligned} E(\Delta u_{it} \Delta u_{i,t+s}) &= E\left[(\Delta \mathbf{f}'_t \boldsymbol{\phi}_i + \Delta \varepsilon_{it})(\boldsymbol{\phi}'_i \Delta \mathbf{f}_{t+s} + \Delta \varepsilon_{i,t+s})\right] \\ &= \Delta \mathbf{f}'_t \boldsymbol{\Sigma}_\phi \Delta \mathbf{f}_{t+s}. \end{aligned} \tag{35}$$

Note that the magnitude of $E(\Delta u_{it} \Delta u_{it+s})$ does not necessarily decrease as $s$ increases with given $t$. Therefore, the power of the proposed joint serial correlation test is likely to increase as the value of $p$ for $m^2_{(2,p)}$ increases.[13]

## 4.2 Discussions

First, it is easily seen that the Sargan's difference test is not justified under the alternatives specified by (25), (33) and (34).[14] Even under the MA($q$) error model for which the Sargan's difference test is valid, the $m^2_{(2,q+1)}$ test will be recommended, given the finite sample evidence of Arellano and Bond (1991).[15]

A natural choice of the test against these and other misspecifications might be the overidentifying restrictions test proposed by Sargan (1958) and Hansen (1982), which is defined by (16). However, the evidence in the recent literature suggests that the finite sample behaviour of the overidentifying restrictions test can be very poor. Bowsher

[12] For large ($N$ and $T$) panels, see Pesaran, Smith and Im (1996) and Pesaran and Yamagata (2008), for related issues.

[13] Sarafidis, Yamagata and Robertson (2006) proposed Sargan's difference test for heterogeneous error cross section dependence.

[14] If $\mathbf{x}_{it}$ is strictly exogenous, a Sargan's difference test (and Hausman test) could be applicable, by utilising the instruments consists of subsets of $\mathbf{X}_i$; see Sarafidis, Yamagata and Robertson (2006).

[15] See the discussion in Section 4.1.

(2002) shows that the overidentifying restrictions test becomes severely undersized with an increasing number of overidentifying moment restrictions in pure autoregressive panel data models with normal errors. The most striking evidence is the finding of Windmeijer (2005). He considered the linear model with only predetermined regressors and heteroskedastic non-normal errors. He compared the size properties of the overidentifying restrictions test based on an infeasible weighting matrix, obtained treating true parameter as known, and the feasible one defined by (16). He found that the two statistics had almost exactly the same size properties, which deteriorates as the number of overidentifying restrictions rises. Given that the $m^2_{(2,p)}$ test has power against a broad range of model misspecifications, as shown above, it can serve as an alternative misspecification test to the overidentifying restrictions test.

In practice, there is no clear theoretical guidance about the best choice of $p$ for the $m^2_{(2,p)}$ test. The choice made partly depends on what kind of misspecifications one has in mind. If there is enough reason to doubt the usefulness of the $q^{th}$ order moving average or autoregressive error serial correlation alternative, it may be reasonable to choose $p$ to be slightly greater than $q+1$. If one uses the joint serial correlation test as an alternative to general misspecification tests, it may be desirable to set $p$ to be its maximum value or close to it, so long as $N$ is sufficiently large.[16]

When the proposed joint serial correlation test rejects the null hypothesis, it does not direct to a particular alternative model specification, as has been emphasized above. Thus, a researcher, who has faced by such a rejection, may have to proceed to identify the source of such misspecifications in separate analyses.[17]

# 5 Small Sample Properties of the Joint $m^2_{(2,p)}$ Test

In this section, the finite sample behaviour of proposed $m^2_{(2,p)}$ test with $p > 2$ is compared with that of the $m_2$ test of Arellano and Bond (1991) and the overidentifying restrictions test.[18] In order to see the effects of increasing $p$ in the $m^2_{(2,p)}$ test under a variety of alternatives, the performance of all $m^2_{(2,3)}$,..., $m^2_{(2,T-3)}$ tests is investigated. When the behaviour of the overidentifying restrictions test is discussed, particular attention is paid to the $m^2_{(2,T-3)}$ test. We consider six types of misspecifications which lead to error serial correlations: AR(1) errors; MA(1) errors; AR(2) errors; MA(2) errors; heterogeneous slopes; and heterogeneous error cross section dependence. The rejection frequencies based on the size-corrected critical value are also provided for the power comparison.[19]

[16] One could examine the $m^2_{(2,p)}$ tests for different values of $p$. However, in this case the test procedure would be subject to multiple testing problem and one cannot control the overall significance level, in general; see Savin (1983).

[17] For example, to sort out AR(1) errors, add one more further lagged dependent variable as a regressor; to cope with cross section dependence, adopt the estimation methods by Holtz-Eakin, Newey and Rosen (1988), Ahn, Lee and Schmidt (2001).

[18] The finite sample evidence reported in Bowsher (2002) suggests that to some limited extent, the size can be controlled by reducing the number of moment restrictions.

[19] As Horowitz and Savin (2000) point out, the tests based on the size-corrected critical values are of limited empirical relevance. We report the size-corrected power of the tests because the bootstrap test, which is the potential alternative to the size-corrected test, seems unreliable in this application, especially for the overidentifying restrictions test; see Bond and Windmeijer (2005).

## 5.1 Design

The first data generating process (DGP) considered is a panel ARDL(1,0) model

$$y_{it} = \alpha_i + \lambda y_{i,t-1} + \beta x_{it} + u_{it},\ i = 1, 2, ..., N; t = -48, -47, ..., T, \tag{36}$$

which may be of greater practical interest than the panel AR(1) model. $y_{i,-49} = 0$ and first 50 observations are discarded. Also, we set $\lambda = 0.5$, $\beta = 0.5$. $u_{it}$ is defined below but for the size of the test $u_{it} = \varepsilon_{it}$, where $\varepsilon_{it} \sim iidN(0, \sigma_\varepsilon^2)$.

The DGP of $x_{it}$ considered here is

$$x_{it} = \rho_x x_{i,t-1} + \pi u_{i,t-1} + v_{it},\ i = 1, 2, ..., N; t = -48, -47, ..., T, \tag{37}$$

where $\rho_x = 0.5$, $v_{it} \sim iidN(0, \sigma_v^2)$. $\pi$ is set to 0.5. $x_{i,-49} = 0$ and first 50 observations are discarded. Following Kiviet (1995) and Bun and Kiviet (2002), we control the signal-to-noise ratio under the null, $u_{it} = \varepsilon_{it}$ through $\sigma_v^2$. Define the signal as $\sigma_s^2 = var\,(y_{it}^* - \varepsilon_{it})$, where $y_{it}^* = y_{it} - \alpha_i/(1-\lambda)$. Then, denoting the variance of the error by $\sigma_\varepsilon^2 = var(\varepsilon_{it})$, we define the signal-to-noise ratio, $\varpi = \sigma_s^2/\sigma_\varepsilon^2$. Specifically

$$\sigma_v^2 = \beta^{-2}\left\{\left[\sigma_\varepsilon^2\left(1+\varpi\right)\right]/a_1 - b_1\right\}$$

where

$$\begin{aligned} a_1 &= \frac{(1+\lambda\rho_x)}{(1-\rho_x^2)(1-\lambda^2)(1-\lambda\rho_x)} \\ b_1 &= 1 + (\beta\pi - \rho_x)^2 + \frac{2(\beta\pi-\rho_x)(\lambda+\rho_x)}{1+\lambda\rho_x}. \end{aligned}$$

We set $\varpi = 3$. Also we choose $\sigma_\alpha^2$ such that the ratio of the impact on $var(y_{it})$ of the two variance components $\alpha_i$ and $\varepsilon_{it}$ is constant across designs. More precisely,

$$\sigma_\alpha^2 = (1-\lambda)^2\, a_1 b_1; \tag{38}$$

see Sarafidis, Yamagata and Robertson (2006) for detailed derivation.

Another DGP considered is derived from Windmeijer (2005) and can be written as

$$y_{it} = \alpha_i + x_{it}\beta + u_{it},\ i = 1, 2, ..., N; t = 1, ..., T, \tag{39}$$

where $\alpha_i \sim iidN(0,1)$, $\beta = 1$, $u_{it}$ is specified above, but $\varepsilon_{it} = \sigma_i\varphi_t\epsilon_{it}$, $\sigma_i \sim iidU[0.5, 1.5]$, $\varphi_t = 0.5$ for $t = -49, ..., 0$ and $\varphi_t = 0.5 + 0.1(t-1)$ for $t = 1, ..., T$, and $\epsilon_{it} \sim iid\chi^2(1) - 1$. The regressor $x_{it}$ is generated as (37), except that $\sigma_v^2 = 1$, and an extra term, $\alpha_i$, enters in the right hand side.

We consider seven different error specifications, denoted by (a)-(g). Constants $c$, $\psi$, $\sigma_\varepsilon$ are controlled so that $Var(u_{it}) = \gamma_0 = 1$ in the case of (36):

(a) First, there are no misspecifications:

$$u_{it} = \sigma_\varepsilon\varepsilon_{it}, \tag{40}$$

$\sigma_\varepsilon = 1$.

(b) The second specification is the AR(1) error model,

$$u_{it} = \rho_1 u_{it-1} + \sigma_\varepsilon\varepsilon_{it}, \tag{41}$$

where $\sigma_\varepsilon^2 = (1 - \rho_1^2)$. $\rho_1 = 0.2$ is considered; so that $\gamma_0 = 1$ and $\gamma_1 = 0.2$, with the DGP (36).

(c) The third specification is the MA(1) error scheme,

$$u_{it} = \sigma_\varepsilon \left( \varepsilon_{it} + \psi_1 \varepsilon_{it-1} \right), \tag{42}$$

where $\sigma_\varepsilon^2 = (1 + \psi_1^2)^{-1}$with $\psi_1 = 0.2$; so that $\gamma_0 = 1$ and $\gamma_1 = 0.2$ with the DGP (36).

(d) The fourth specification is the AR(2) error model,

$$u_{it} = \rho_1 u_{it-1} + \rho_2 u_{it-2} + \sigma_\varepsilon \varepsilon_{it}, \tag{43}$$

where $\sigma_\varepsilon^2 = (1 + \rho_2)[(1 - \rho_2)^2 - \rho_1^2]/(1 - \rho_2)$, with $\rho_1 = 0.2$ and $\rho_2 = 0.1$; so that $\gamma_0 = 1$, $\gamma_1 = 2/9$ and $\gamma_2 = 13/90$ with the DGP (36).

(e) The fifth specification is the MA(2) error process,

$$u_{it} = \sigma_\varepsilon \left( \varepsilon_{it} + \psi_1 \varepsilon_{it-1} + \psi_2 \varepsilon_{it-2} \right), \tag{44}$$

where $\sigma_\varepsilon^2 = (1 + \psi_1^2 + \psi_2^2)^{-1}$ with $\psi_1 = 20/103$, $\psi_2 = 13/90$ so that $\gamma_0 = 1$, $\gamma_1 = 2/9$ and $\gamma_2 = 13/90$ with the DGP (36).

Note that these particular designs of AR(2) and MA(2) errors are chosen to empathize the usefulness of joint serial correlation test relative to the $m_2$ test. Specifically, under these designs, $E\left(\Delta u_{it} \Delta u_{it+2}\right) = 2\gamma_2 - \gamma_1 = 0.07$ and $E\left(\Delta u_{it} \Delta u_{it+3}\right) = -\gamma_2 = -0.14$; so the latter is twice as large as the former in absolute value. This result implies that the $m_2$ test is likely to be less powerful than the $m_{(2,3)}^2$ test.

(f) The sixth specification allows for heterogeneous slopes. The term $\beta$ in (36) is replaced by $\beta_i \sim iidN(0.5, 1)$, and $\lambda$ is kept homogeneous. The constant $\beta$ in (39) is replaced with $\beta_i \sim iidN(1, 1)$.

(g) The final specification permits heterogeneous error cross section dependence, with

$$u_{it} = c^2 \left( \phi_i f_t + \sigma_\varepsilon^2 \varepsilon_{it} \right), \tag{45}$$

$\phi_i \sim iidU[-1, 1]$, $f_t \sim iidN(0, \sigma_f^2)$, $\sigma_f^2 = \sigma_\varepsilon^2 = 1$. We set $c^2 = 3/4$.

We consider all combinations of $N = 100, 200, 400$, $T = 7, 11$ for DGP (36) and $T = 6, 10$ for DGP (39). All experiments are based on 2000 replications. The rejection rates are based on an estimated 5% critical value, which is obtained as the 0.95 quantile of the test statistics under consideration over 10000 replications.

## 5.2 Results

Table 2 contains results for the case of a linear dynamic panel ARDL(1,0) model with predetermined regressors. The size results are reported in panel (a). The size of the $m_2$ test and the $m_{(2,p)}^2$ tests are satisfactory for all combinations of $N$ and $T$. On the other hand, the overidentifying restrictions test tends to reject the null too infrequently. The degree of under-rejection by the overidentifying restrictions test becomes worse when $T$ is increased to 11, due to a increase of the number of moment restrictions. This finding is consistent with the results of Bowsher (2002) and Windmeijer (2005).

Next, consider evidence about the power properties under varieties of alternatives, which is contained in panels (b)-(g) in Table 2. Given the size distortion of the overidentifying restrictions test, a size-adjusted power is reported in parentheses. In the case of AR(1) error specified by (41), the power of the $m_{(2,p)}^2$ tests with $p > 2$ dominates that of the $m_2$ test almost all of the cases, as predicted in section 4.1. Focusing on the choice of $p$ of the joint $m^2_{(2,p)}$ test, when $T = 11$ the power increases as $p$ rises from 2 to 5, then slightly decreases afterwards. Across designs, the overidentifying restrictions test has very low power, partly due to its size distortion towards below the significance level. Nevertheless, in terms of the size-adjusted power, the $m^2_{(2,T-3)}$ tests are also superior to the overidentifying restrictions test. Turning attention to MA(1) errors specified by (42), the power of the $m^2_{(2,p)}$ tests with $p > 2$ dominates that of the $m_2$ test most of the cases, as predicted in section 4.1. In the Monte Carlo design, MA(1) and AR(1) errors yield the same first order autocorrelation of $u_{it}$, though the power gained by increasing $p$ in the case of MA(1) error is not as much as in the case of AR(1) errors. This result may be explained as follows. Recall that (28) shows that the mean shift of the test statistic under the alternative is decomposed into non-zero $s^{th}$ order autocovariances of $\Delta u_{it}$, $E(\eta_{is})$, and the bias of the estimator of slope coefficient. In the case of non-local AR(1) error, all $E(\eta_{is})$, $s = 2, 3, ..., p$, are non-zero, whereas, in the case of MA(1) error only $E(\eta_{i2})$ is non-zero, which may lead to such a difference in power. Another property to point out is that the power of the $m^2_{(2,p)}$ tests is in general higher in the case of MA(1) error than in the case of AR(1) error,. Probably this property reflects the fact that the bias of the estimator of slope coefficient reduces the magnitude of the mean shift of the test statistics and such a bias is larger with AR(1) errors than with MA(1) errors.

In the case of AR(2) error specified by (43) the $m_2$ test has virtually no power, due to the choice of the parameters in autoregressive errors, as explained above. In contrast, the $m^2_{(2,p)}$ test increases its power substantially as $p$ rises. For $T = 7$, the overidentifying restrictions test seems more powerful than the $m^2_{(2,T-3)}$ test, but the reverse relationship is true for $T = 11$. In the case of MA(2) error specified by (44), similar properties of the behaviour of tests hold to those in the case of AR(2) error.[20]

In the case of slope heterogeneity, where $\beta$ in (36) is replaced with $\beta_i \sim iidN(0.5, 1)$ and $\lambda$ is kept constant, the $m^2_{(2,p)}$ tests with $p > 2$ dominate the $m_2$ test, except for $N = 100$. For $T = 7$, there is no clear ranking in terms of power between the $m^2_{(2,T-3)}$ test and the overidentifying restrictions test, but the $m^2_{(2,T-3)}$ test is superior for $T = 11$.

In the case of cross section dependence specified by (45), power estimates of the $m^2_{(2,p)}$ tests monotonically increase as $p$ increases, as discussed in Section 4.1. The power of the $m^2_{(2,T-3)}$ tests is exceeded by that of the overidentifying restrictions test when $N$ becomes larger.

Table 3 reports the results in the case of a linear panel model with predetermined regressors. The size results are reported in panel (a). The estimated size of the $m^2_{(2,p)}$ tests tend to lower than the significance level for small $N$ and large $p$. The overidentifying restrictions test tends to reject the null very infrequently. The power properties of the $m_2$ test and the $m^2_{(2,p)}$ tests are similar to those reported in Table 2. The evidences suggest that the power of the overidentifying restrictions test is extremely low across designs and dominated by the $m^2_{(2,T-2)}$ test.

Overall, the performance of the $m^2_{(2,p)}$ tests with $p > 2$ is at least as good as the conventional $m_2$ test, and is superior to the latter in the majority of cases. Also, the

[20]The power of the joint tests under the AR(2) or MA(2) errors is lower than that under the AR(1) or MA(1) errors. This is explained as follows. Under the AR(2) or MA(2) errors in the experiments, when $T = 7$, $|E(\eta_{i2})| = 0.21$ and $|E(\eta_{i3})| = 0.28$. On the other hand, under the AR(1) or MA(1) errors, $|E(\eta_{i2})| = 0.8$, which is much larger than the sum of $|E(\eta_{i2})|$ and $|E(\eta_{i3})|$ under the AR(2) or MA(2) error design.

$m^2_{(2,T-3)}$ tests are superior to the overidentifying restrictions test in many, but not all cases. The proposed joint serial correlation tests can serve as a general misspecification test as an alternative to the overidentifying restrictions test.

# 6 Concluding Remarks

This paper has proposed a joint error serial correlation test for linear panel data models estimated by the generalised method of moments (GMM) estimation. The proposed serial correlation test, called $m^2_{(2,p)}$ test, examines second to $p^{th}$-order (first differenced) error serial correlations jointly. The asymptotic local power analysis of the $m^2_{(2,p)}$ test reveals that (i) AR($q$) and MA($q$) errors are locally equivalent alternatives in Godfrey's (1981) sense; (ii) the asymptotic power of an overspecified $m^2_{(2,p)}$ test can be higher than that of the just specified test. This implies that the power of the proposed joint serial correlation test can be higher than that of the conventional $m_2$ test, under the varieties of alternatives such as AR($q$) and MA($q$) errors, slope heterogeneity, and cross section dependence.

The small sample properties of the $m^2_{(2,p)}$ tests with $p > 2$ has been compared with those of the $m_2$ test, which is equivalent to the $m^2_{(2,2)}$ test, and also with those of the overidentifying restrictions test by means of Monte Carlo experiments. The evidence shows that the $m^2_{(2,p)}$ tests with $p > 2$ mostly outperform the $m_2$ test under several alternatives, such as AR(1), AR(2) and MA(2) errors, slope heterogeneity and error cross section dependence. In the case of MA(1) errors, the $m^2_{(2,p)}$ tests with $p > 2$ and the $m_2$ test have very similar power. It is important to note that the $m^2_{(2,p)}$ test with the maximum $p$ available has high power when the overidentifying restrictions test does not.

In view of these results, it is concluded that the proposed joint serial correlation test may serve as a useful alternative to the conventional $m_2$ and the overidentifying restrictions tests.

It may be worth making two remarks. There is no clear theoretical guidance about how to choose $p$ for the $m^2_{(2,p)}$ test. The implications for power properties depend upon the nature of actual misspecification. The absence of prior information about the number of test indicators (i.e. $p$ here) is typical of the implementation of misspecification checks, e.g., the RESET test and the Lagrange multiplier test for serial correlation; see Godfrey (1988;p.79-80). Second, a rejection of the null of no error serial correlation by the proposed test does not necessarily imply the acceptance of any particular alternative model specification. Thus, a researcher, who has been faced by such a rejection, should proceed to identify the source of misspecifications without relying solely on the test outcome and estimation of the data-inconsistent model; see, for example, Davidson and MacKinnon (1985) and Godfrey and Orme (1996) for further discussion.

# A Proof of Theorem 1

Recall $\ddot{\boldsymbol{\eta}}_i = \left(\ddot{\eta}_{i2}, \ddot{\eta}_{i3}, ..., \ddot{\eta}_{ip}\right)'$ with $\ddot{\eta}_{is} = \sum_{t=3}^{T-s} \Delta\ddot{u}_{it}\Delta\ddot{u}_{it+s}$, $s = 2, 3, ..., p(\leq T-3)$. Replacing the averages in the right hand side of a Taylor series expansion of $N^{-1/2}\sum_{i=1}^{N} \ddot{\boldsymbol{\eta}}_i$ around $\ddot{\boldsymbol{\theta}}_N = \boldsymbol{\theta}$ with averages of expectations yields

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\ddot{\boldsymbol{\eta}}_i = \left[\mathbf{I}_{p-1}, -\bar{\mathbf{B}}\bar{\mathbf{Q}}^{-1}\bar{\mathbf{A}}'\bar{\boldsymbol{\Omega}}^{-1}\right]\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\begin{array}{c}\boldsymbol{\eta}_i\\ \mathbf{Z}_i'\Delta\mathbf{u}_i\end{array}\right) + o_p(1), \tag{A.1}$$

where $\bar{\mathbf{B}} = (\bar{\boldsymbol{\omega}}_2, \bar{\boldsymbol{\omega}}_3, ..., \bar{\boldsymbol{\omega}}_p)'$ with $\bar{\boldsymbol{\omega}}_s = N^{-1}\sum_{i=1}^{N} E(\sum_{t=3}^{T-s}\Delta u_{it}\Delta\mathbf{w}_{i,t+s})$, $\bar{\mathbf{Q}} = \bar{\mathbf{A}}'\bar{\boldsymbol{\Omega}}^{-1}\bar{\mathbf{A}}$, $\bar{\mathbf{A}} = N^{-1}\sum_{i=1}^{N} E(\mathbf{Z}_i'\Delta\mathbf{W}_i)$, $\bar{\boldsymbol{\Omega}} = N^{-1}\sum_{i=1}^{N} E(\mathbf{Z}_i'\Delta\mathbf{u}_i\Delta\mathbf{u}_i'\mathbf{Z}_i)$. As the first term of the right hand side have mean zero and the asymptotic variance-covariance matrix

$$\bar{\mathbf{V}}_p = \bar{\mathbf{D}}\bar{\mathbf{M}}\bar{\mathbf{D}}', \tag{A.2}$$

where $\bar{\mathbf{D}} = [\mathbf{I}_{p-1}, -\bar{\mathbf{B}}\bar{\mathbf{Q}}^{-1}\bar{\mathbf{A}}'\bar{\boldsymbol{\Omega}}^{-1}]$ and

$$\bar{\mathbf{M}} = \left(\begin{array}{cc}\bar{\mathbf{R}} & \bar{\mathbf{F}}\\ \bar{\mathbf{F}}' & \bar{\boldsymbol{\Omega}}\end{array}\right), \tag{A.3}$$

with $\bar{\mathbf{R}} = N^{-1}\sum_{i=1}^{N} E\left(\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\right)$ and $\bar{\mathbf{F}} = N^{-1}\sum_{i=1}^{N} E\left(\boldsymbol{\eta}_i\Delta\mathbf{u}_i'\mathbf{Z}_i\right)$. As $\bar{\mathbf{M}}$ is a $(p+h-1\times p+h-1)$ positive definite matrix by Assumption 4 and $\bar{\mathbf{D}}$ has full row rank**,** $\bar{\mathbf{V}}_p$ is positive definite**.** Under Assumptions 1-4,

$$N^{-1/2}\bar{\mathbf{V}}_p^{-1/2}\ddot{\mathbf{H}}'\boldsymbol{\iota}_N \xrightarrow{d} N(\mathbf{0}_{p-1}, \mathbf{I}_{p-1}), \tag{A.4}$$

then, since $\ddot{\mathbf{V}}_N - \bar{\mathbf{V}}_p = o_p(1)$ with $\ddot{\mathbf{V}}_N = \ddot{\mathbf{G}}'\ddot{\mathbf{G}}/N$,

$$\frac{\boldsymbol{\iota}_N'\ddot{\mathbf{H}}}{\sqrt{N}}\left(\frac{\ddot{\mathbf{G}}'\ddot{\mathbf{G}}}{N}\right)^{-1}\frac{\ddot{\mathbf{H}}'\boldsymbol{\iota}_N}{\sqrt{N}} \xrightarrow{d} \chi^2(p-1) \tag{A.5}$$

as $N \to \infty$, under the null hypothesis, as required.

# References


Ahn, S.C., Lee Y.H., Schmidt, P., (2001). GMM estimation of linear panel data models with time-varying individual effects. Journal of Econometrics 101, 219-255.

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

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

Blundell, R.W., Bond, S.R., (1998). Initial conditions and moment restrictions in dynamic panel data models. Journal of Econometrics 87, 115-143.

Bond, S., Windmeijer, F., (2005). Reliable inference for GMM estimators? Finite sample procedures in linear panel data models. Econometric Reviews 24, 1-37.

Bowsher, C.G., (2002). On testing overidentifying restrictions in dynamic panel data models. Economics Letters 77, 211-220.

Brown, B.W., Newey, W.K., (2002). Generalized method of moments, efficient bootstrapping, and improved inference. Journal of Business and Economic Statistics 20, 507-517.

Bun, M.J.G., Kiviet, J.F., (2006). The effects of dynamic feedbacks on LS and MM estimator accuracy in panel data models. Journal of Econometrics 132, 409-444.

Das Gupta, S., Perlman, M.D., (1974). Power of the noncentral F-test: effect of additional variates on Hotelling's $T^2$-test. Journal of the American Statistical Association 69, 174-180.

Davidson, R., MacKinnon, J.G., (1985). The interpretation of test statistics. Canadian Journal of Economics 69, 38-57.

Davidson, R., MacKinnon, J.G., (1993). Estimation and Inference in Econometrics. Oxford University Press, New York.

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.

Godfrey, L.G., (1988). Misspecification Tests in Econometrics. Cambridge University Press, Cambridge.

Godfrey, L.G., Orme, C.D., (1996). On the behaviour of conditional moment tests in the presence of unconsidered local alternatives. International Economic Review 37, 263-281.

Hall, P., Horowitz, J., (1996). Bootstrap critical values for tests based on generalized method of moments estimators. Econometrica 64, 891-916.

Hamilton, J.D., (1994). Time Series Analysis, Princeton University Press, Princeton.

Hansen, L., (1982). Large sample properties of generalized method of moments estimators. Econometrica 50, 1029-1054.

Hausman, J.A., (1978). Specification tests in econometrics. Econometrica 46, 1251-1272.

Hayashi, F., (2000). Econometrics, Princeton University Press, Princeton.

Holtz-Eakin, D., Newey, W., Rosen H.S., (1988). Estimating Vector Autoregressions with Panel Data. Econometrica 56, 1371–1395.

Horowitz, J., Savin, N.E., (2000). Empirically relevant critical values for hypothesis tests: A bootstrap approach. Journal of Econometrics 95, 375-389.

Inoue, A., Solon, G., (2006). A portmanteau test for serially correlated errors in fixed effects models. Econometric Theory 22, 835-851.

Kiviet, J.F., (1995). On bias, inconsistency, and efficiency of various estimators in dynamic panel data models. Journal of Econometrics 68, 53-78.

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

Pesaran, M.H., Smith, Ron, (1995). Estimating long-run relationships from dynamic heterogeneous panels. Journal of Econometrics 68, 79-113.

Pesaran, M.H., Yamagata, T., (2008). Testing slope homogeneity in large panels. Journal of Econometrics 142, 50-93.

Pesaran, H., Smith, R., Im, K.S., (1996). Dynamic linear models for heterogenous panels. In: Matyas, L., Sevestre, P., eds, The Econometrics of Panel Data: A Handbook of the Theory with Applications, second revised edition, pp. 145-195.

Robertson, D., Sarafidis, V., (2006). The impact of cross-sectional dependence in short dynamic panel data estimation. Unpublished manuscript.

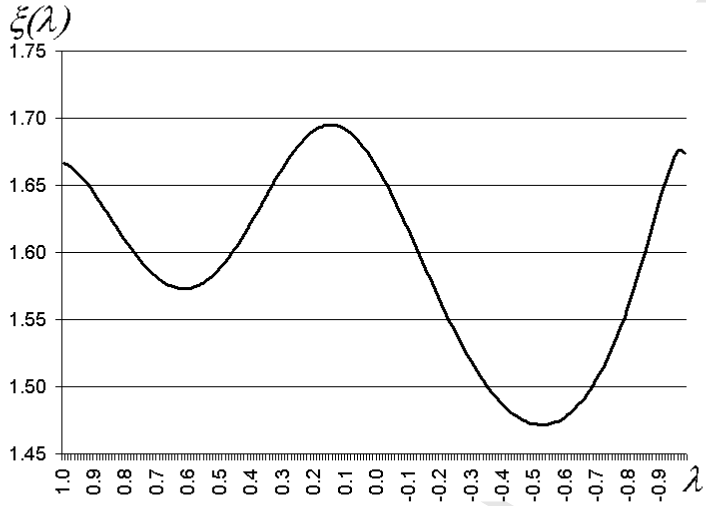
Sarafidis, V., Yamagata, T., Robertson, D., (2006). A test of cross section dependence for a linear dynamic panel model with regressors. Unpublished manuscript.

Sargan, D., (1958). The estimation of economic relationships using instrumental variables. Econometrica 26, 393-415.

Savin, N.E., (1983). Multiple hypothesis testing, in: Z. Griliches and M.D. Intriligator, (Eds.), Handbook of Econometrics, Vol. 2. Elsevier, Amsterdam, pp. 827-879.

Windmeijer, F., (2005). A finite sample correction for the variance of linear efficient two-step GMM estimators. Journal of Econometrics 126, 25-51.

**Figure 1: The plot of the ratio of noncentrality parameters of the asymptotic distributions of $m^2_{(2,3)}$ and $m^2_{(2,2)}$ statistics under local MA(1)/AR(1) errors, in the case of a panel AR(1) model**




Notes: $\xi(\lambda) = \zeta_2(\delta^2;\lambda)/\zeta_1(\delta^2;\lambda)$ is the ratio of noncentral parameters, where $\zeta_2(\delta^2;\lambda)$ and $\zeta_1(\delta^2;\lambda)$ are noncentral parameter of the asymptotic distribution of $m^2_{(2,3)}$ and $m^2_{(2,2)}$ statistics under local MA(1)/AR(1) errors, respectively, given the panel AR(1) model $y_{it} = \alpha_i + \lambda y_{it-1} + u_{it}$, $i = 1, 2, ..., N$, $t = 1, 2, ..., T$, $\alpha_i \sim iid(0, \sigma^2_\alpha)$, $|\lambda| < 1$, $u_{it} \sim iid(0, \sigma^2)$, with $T = 6$, $\sigma^2 = \sigma^2_\alpha = 1$.

**Table 1: Value of $\delta^2$ in the noncentral chi-square distributions $\chi^2(1, \delta^2)$ and $\chi^2(2, \xi(\lambda)\delta^2)$ to achieve the power $\pi_{0.05}(n, \zeta)$**

| | $\chi^2(1,\delta^2)$ | $\chi^2(2,\xi(\lambda)\delta^2)$ | | |
|---|---|---|---|---|
| $\pi_{0.05}(n,\zeta)$ | – | $\xi(0.14) = 1.695$ | $\xi(0.61) = 1.573$ | $\xi(-0.53) = 1.471$ |
| 0.05 | 0.00 | 0.00 | 0.00 | 0.00 |
| 0.10 | 0.65 | 0.47 | 0.50 | 0.54 |
| 0.20 | 1.11 | 0.78 | 0.84 | 0.89 |
| 0.50 | 1.96 | 1.31 | 1.42 | 1.51 |
| 0.90 | 3.24 | 2.10 | 2.26 | 2.42 |
| 0.95 | 3.60 | 2.32 | 2.50 | 2.67 |

Note: The required $\delta^2$ with which the power functions $\pi_{0.05}(1, \delta^2)$ and $\pi_{0.05}(2, \xi(\lambda)\delta^2)$ achieve the target value, 0.05, 0.10,...,0.95, are obtained, where $\xi(\lambda)$ is the ratio of noncentral parameters of the asymptotic distribution of $m^2_{(2,3)}$ and $m^2_{(2,2)}$ statistics under local MA(1)/AR(1) errors, given the panel AR(1) model $y_{it} = \alpha_i + \lambda y_{it-1} + u_{it}$, $i = 1, 2, ..., N$, $t = 1, 2, ..., T$, $\alpha_i \sim iid(0, \sigma^2_\alpha)$, $|\lambda| < 1$, $u_{it} \sim iid(0, \sigma^2)$, with $T = 6$, $\sigma^2 = \sigma^2_\alpha = 1$. As shown in Figure 1, at $\lambda = 0.14$ $(-0.53)$ $\xi(\lambda)$ is at the maximum (minimum) for $-0.99 \le \lambda \le 0.99$, and at $\lambda = 0.61$ $\xi(\lambda)$ is at the local maximum for $0 \le \lambda \le 0.99$.

**Table 2: Size and Power of the Tests: A Dynamic Panel ARDL(1,0) Model with Predetermined Regressors**

| | (a) Size | | | (b) AR(1) Error | | |
|---|---|---|---|---|---|---|
| **Test,N** | **100** | **200** | **400** | **100** | **200** | **400** |
| | $T=7$ | | | $T=7$ | | |
| $m_2$ | 5.40 | 5.05 | 5.50 | 24.10 (22.75) | 42.45 (43.45) | 69.50 (69.15) |
| $m^2_{(2,3)}$ | 4.95 | 4.80 | 4.95 | 25.80 (27.35) | 49.80 (51.25) | 83.40 (83.50) |
| $m^2_{(2,4)}$ | 4.90 | 4.90 | 4.80 | 26.20 (27.80) | 52.55 (53.10) | 86.30 (87.50) |
| $S(\ddot{\boldsymbol{\theta}}_N)$ | 2.90 | 4.40 | 4.80 | 8.05 (14.15) | 32.85 (37.25) | 81.75 (79.75) |
| | $T=11$ | | | $T=11$ | | |
| $m_2$ | 4.55 | 4.95 | 5.30 | 43.80 (45.50) | 72.80 (74.15) | 95.35 (95.80) |
| $m^2_{(2,3)}$ | 5.10 | 4.80 | 4.90 | 55.10 (57.20) | 88.10 (87.10) | 99.55 (99.55) |
| $m^2_{(2,4)}$ | 4.65 | 5.40 | 5.35 | 62.15 (63.85) | 92.60 (92.20) | 99.90 (99.98) |
| $m^2_{(2,5)}$ | 4.80 | 5.35 | 5.10 | 62.85 (64.25) | 94.00 (93.85) | 100.00 (100.00) |
| $m^2_{(2,6)}$ | 4.55 | 5.05 | 5.15 | 61.90 (63.40) | 94.50 (93.90) | 100.00 (100.00) |
| $m^2_{(2,7)}$ | 4.05 | 5.25 | 4.90 | 59.25 (62.90) | 94.40 (94.35) | 100.00 (100.00) |
| $m^2_{(2,8)}$ | 4.05 | 5.65 | 4.70 | 57.00 (62.05) | 94.35 (93.85) | 100.00 (100.00) |
| $S(\ddot{\boldsymbol{\theta}}_N)$ | 0.00 | 1.45 | 3.90 | 0.00 (5.85) | 18.50 (34.50) | 90.85 (89.75) |
| | (e) MA(1) Error | | | (d) AR(2) Errors | | |
| **Test,N** | **100** | **200** | **400** | **100** | **200** | **400** |
| | $T=7$ | | | $T=7$ | | |
| $m_2$ | 44.15 (43.20) | 73.05 (73.80) | 95.45 (95.50) | 5.15 (4.80) | 5.30 (5.70) | 5.40 (5.60) |
| $m^2_{(2,3)}$ | 44.00 (44.00) | 75.55 (76.00) | 97.30 (97.30) | 7.45 (7.45) | 9.60 (9.90) | 15.45 (15.75) |
| $m^2_{(2,4)}$ | 43.20 (44.05) | 77.30 (77.35) | 97.90 (98.10) | 8.20 (8.98) | 12.80 (12.90) | 24.20 (25.55) |
| $S(\ddot{\boldsymbol{\theta}})$ | 10.20 (18.10) | 48.75 (51.95) | 92.85 (93.00) | 4.15 (9.10) | 15.60 (17.80) | 45.25 (45.95) |
| | $T=11$ | | | $T=11$ | | |
| $m_2$ | 75.45 (76.30) | 97.25 (97.50) | 99.90 (99.90) | 5.15 (5.25) | 5.10 (5.45) | 6.35 (6.70) |
| $m^2_{(2,3)}$ | 80.80 (81.60) | 99.00 (99.05) | 100.00 (100.00) | 12.30 (13.10) | 21.00 (21.35) | 36.30 (36.35) |
| $m^2_{(2,4)}$ | 83.60 (84.55) | 99.45 (99.55) | 100.00 (100.00) | 20.55 (21.50) | 36.60 (37.40) | 66.35 (67.25) |
| $m^2_{(2,5)}$ | 83.00 (83.65) | 99.45 (99.45) | 100.00 (100.00) | 23.55 (24.25) | 45.60 (46.10) | 80.15 (79.95) |
| $m^2_{(2,6)}$ | 81.80 (82.00) | 99.50 (99.50) | 100.00 (100.00) | 25.40 (25.70) | 50.45 (50.90) | 84.45 (84.40) |
| $m^2_{(2,7)}$ | 80.95 (82.05) | 99.45 (99.45) | 100.00 (100.00) | 25.10 (27.25) | 52.55 (52.60) | 87.25 (87.50) |
| $m^2_{(2,8)}$ | 78.35 (80.20) | 99.25 (99.30) | 100.00 (100.00) | 24.75 (27.00) | 53.15 (53.85) | 87.85 (87.95) |
| $S(\ddot{\boldsymbol{\theta}}_N)$ | 0.00 (5.85) | 27.85 (45.55) | 97.25 (97.45) | 0.00 (4.95) | 8.50 (20.25) | 63.75 (65.65) |

(Table 2 continued)

| | **(e) MA(2) Errors** | | | **(f) Slope Heterogeneity:** | | |
|---|---|---|---|---|---|---|
| **Test,N** | **100** | **200** | **400** | **100** | **200** | **400** |
| | $T=7$ | | | $T=7$ | | |
| $m_2$ | 8.15 (7.70) | 13.60 (14.40) | 21.80 (22.25) | 19.25 (18.60) | 31.00 (32.15) | 50.30 (51.15) |
| $m^2_{(2,3)}$ | 14.35 (14.35) | 22.55 (23.10) | 45.70 (46.15) | 17.95 (17.95) | 37.90 (38.65) | 65.10 (65.40) |
| $m^2_{(2,4)}$ | 16.20 (16.70) | 30.65 (30.80) | 57.05 (59.00) | 17.15 (17.60) | 37.75 (37.90) | 70.90 (72.55) |
| $S(\ddot{\boldsymbol{\theta}}_N)$ | 6.25 (12.10) | 25.10 (28.25) | 67.10 (67.70) | 3.90 (8.35) | 27.70 (30.40) | 80.15 (80.95) |
| | $T=11$ | | | $T=11$ | | |
| $m_2$ | 12.20 (13.00) | 23.10 (23.55) | 43.70 (44.85) | 33.65 (34.65) | 53.90 (54.70) | 83.05 (83.60) |
| $m^2_{(2,3)}$ | 31.20 (23.30) | 57.95 (58.45) | 89.05 (89.05) | 40.45 (41.15) | 72.60 (73.30) | 96.55 (96.55) |
| $m^2_{(2,4)}$ | 47.05 (48.55) | 80.10 (80.60) | 98.25 (98.40) | 44.95 (46.80) | 80.70 (81.20) | 98.95 (99.00) |
| $m^2_{(2,5)}$ | 48.90 (49.75) | 82.50 (83.20) | 99.15 (99.15) | 43.30 (43.95) | 81.15 (81.70) | 99.35 (99.35) |
| $m^2_{(2,6)}$ | 48.45 (48.90) | 84.10 (84.35) | 99.40 (99.40) | 41.50 (41.95) | 82.00 (82.25) | 99.45 (99.45) |
| $m^2_{(2,7)}$ | 46.50 (48.70) | 84.80 (84.85) | 99.45 (99.45) | 40.30 (42.85) | 81.25 (81.35) | 99.45 (99.45) |
| $m^2_{(2,8)}$ | 47.05 (48.60) | 85.40 (85.85) | 99.55 (99.55) | 37.40 (40.15) | 80.20 (80.55) | 99.25 (99.25) |
| $S(\ddot{\boldsymbol{\theta}}_N)$ | 0.00 (5.15) | 14.45 (29.95) | 85.65 (86.25) | 0.00 (4.55) | 7.05 (19.40) | 81.45 (82.60) |

| | **(g) Cross Section Dependence** | | |
|---|---|---|---|
| **Test,N** | **100** | **200** | **400** |
| | $T=7$ | | |
| $m_2$ | 37.00 (34.90) | 49.20 (47.90) | 57.65 (60.80) |
| $m^2_{(2,3)}$ | 50.60 (47.40) | 65.70 (64.10) | 75.25 (77.45) |
| $m^2_{(2,4)}$ | 56.30 (54.50) | 73.15 (71.90) | 82.40 (85.00) |
| $S(\ddot{\boldsymbol{\theta}})$ | 39.75 (49.85) | 80.75 (80.70) | 93.20 (93.85) |
| | $T=11$ | | |
| $m_2$ | 38.60 (36.75) | 54.40 (51.35) | 63.90 (63.50) |
| $m^2_{(2,3)}$ | 55.00 (50.30) | 71.00 (68.30) | 80.50 (80.60) |
| $m^2_{(2,4)}$ | 63.45 (61.40) | 80.55 (79.80) | 88.85 (88.60) |
| $m^2_{(2,5)}$ | 68.05 (66.65) | 85.50 (84.30) | 92.65 (92.95) |
| $m^2_{(2,6)}$ | 72.35 (71.15) | 89.25 (88.15) | 95.65 (95.40) |
| $m^2_{(2,7)}$ | 75.10 (75.75) | 91.30 (90.00) | 96.50 (96.70) |
| $m^2_{(2,8)}$ | 76.40 (77.20) | 92.10 (90.85) | 97.15 (97.45) |
| $S(\ddot{\boldsymbol{\theta}}_N)$ | 0.00 (8.10) | 72.05 (84.20) | 99.15 (99.30) |

Notes: The data is generated as $y_{it} = \alpha_i + 0.5y_{i,t-1} + \beta x_{it} + u_{it}$, where $\alpha_i \sim iidN(0, \sigma^2_\alpha)$ with $\sigma^2_\alpha$ defined by (38); $\beta = 0.5$ except panel (f), where $\beta_i \sim iidN(0.5, 1)$; $u_{it}$ is specified by (40)-(45); $x_{it} = 0.5x_{i,t-1} + 0.5u_{i,t-1} + v_{it}$, $v_{it} \sim iidN(0, \sigma^2_v)$, $i = 1, 2, ..., N; t = -48, -47, ..., T$, with $y_{i,-49} = 0$ and $x_{i,-49} = 0$. The first 50 observations are discarded. The signal-to-noise ratio is fixed 3 through $\sigma^2_v$ under the null. $m_2$ signifies the Arellano and Bond (1991) test, $m^2_{(2,p)}$ signifies the proposed joint test for second to $p^{th}$ order first differenced error serial correlation, $S(\ddot{\boldsymbol{\theta}}_N)$ signifies the overidentifying restrictions test. All tests are based on optimal two-step Arellano and Bond (1991) GMM estimator. The $m_2$ test results is based on the $m^2_{(2,2)}$ test, and the $m^2_{(2,p)}$ statistics are compared to $\chi^2(p-1)$ distributions. The $S(\ddot{\boldsymbol{\theta}}_N)$ statistic is compared to $\chi^2(h-2)$ distributions, where $h$ is defined by (6). The figures in parenthesis are size-adjusted power, which are based on the simulated distributions of test statistics with 10000 replications. All tests are conducted at 5% significance level. All experiments are based on 2000 replications.

**Table 3: Size and Power of the Tests: A Linear Model with Predetermined Regressors**

| | **(a) Size** | | | **(b) AR(1) Errors** | | |
|---|---|---|---|---|---|---|
| **Test,N** | **100** | **200** | **400** | **100** | **200** | **400** |
| | $T=6$ | | | $T=6$ | | |
| $m_2$ | 4.65 | 4.75 | 4.90 | 19.35 (18.60) | 29.20 (31.65) | 50.15 (46.70) |
| $m^2_{(2,3)}$ | 2.75 | 3.95 | 5.10 | 16.60 (22.02) | 32.35 (37.60) | 59.20 (59.55) |
| $m^2_{(2,4)}$ | 3.15 | 3.45 | 4.80 | 14.70 (20.00) | 31.75 (36.30) | 61.70 (60.80) |
| $S(\ddot{\boldsymbol{\theta}}_N)$ | 2.00 | 3.00 | 4.00 | 2.10 (5.65) | 3.15 (5.95) | 4.70 (7.15) |
| | $T=10$ | | | $T=10$ | | |
| $m_2$ | 5.15 | 4.50 | 4.60 | 31.10 (32.80) | 49.40 (52.05) | 75.25 (76.15) |
| $m^2_{(2,3)}$ | 3.65 | 4.00 | 4.35 | 34.30 (37.45) | 59.30 (63.20) | 89.55 (88.85) |
| $m^2_{(2,4)}$ | 3.50 | 4.00 | 4.80 | 36.60 (40.25) | 63.55 (69.25) | 92.50 (92.95) |
| $m^2_{(2,5)}$ | 3.30 | 3.50 | 4.95 | 34.75 (41.75) | 65.15 (69.45) | 94.10 (94.60) |
| $m^2_{(2,6)}$ | 3.10 | 3.60 | 4.85 | 34.50 (42.35) | 64.90 (69.80) | 94.50 (94.55) |
| $m^2_{(2,7)}$ | 2.90 | 2.85 | 5.00 | 31.70 (40.75) | 65.15 (70.60) | 94.55 (94.80) |
| $m^2_{(2,8)}$ | 2.95 | 3.50 | 5.35 | 28.00 (38.85) | 62.30 (71.05) | 94.35 (94.50) |
| $S(\ddot{\boldsymbol{\theta}}_N)$ | 0.45 | 1.65 | 2.30 | 0.45 (5.05) | 1.95 (6.60) | 6.15 (9.40) |
| | **(c) MA(1) Errors** | | | **(d) AR(2) Errors** | | |
| **Test,N** | **100** | **200** | **400** | **100** | **200** | **400** |
| | $T=6$ | | | $T=6$ | | |
| $m_2$ | 30.55 (31.00) | 52.90 (52.75) | 75.25 (73.95) | 5.65 (5.65) | 4.50 (4.45) | 4.15 (4.00) |
| $m^2_{(2,3)}$ | 25.95 (30.90) | 53.05 (56.50) | 80.25 (80.85) | 6.10 (8.30) | 8.85 (10.55) | 13.25 (13.55) |
| $m^2_{(2,4)}$ | 23.50 (29.00) | 50.00 (53.10) | 81.15 (81.55) | 6.40 (8.40) | 9.80 (11.40) | 16.30 (16.75) |
| $S(\ddot{\boldsymbol{\theta}})$ | 2.40 (5.75) | 4.20 (6.60) | 6.75 (8.60) | 1.65 (5.10) | 3.80 (5.45) | 4.50 (6.35) |
| | $T=10$ | | | $T=10$ | | |
| $m_2$ | 53.15 (53.90) | 78.95 (79.80) | 95.25 (95.25) | 3.90 (3.95) | 4.60 (5.05) | 4.25 (4.35) |
| $m^2_{(2,3)}$ | 53.90 (56.65) | 83.50 (84.80) | 98.15 (98.15) | 9.00 (9.95) | 14.50 (16.50) | 23.00 (23.00) |
| $m^2_{(2,4)}$ | 52.50 (57.10) | 85.55 (87.35) | 98.85 (98.85) | 12.45 (14.20) | 23.50 (26.30) | 40.90 (41.75) |
| $m^2_{(2,5)}$ | 50.05 (57.15) | 85.40 (87.10) | 98.95 (98.95) | 13.45 (18.10) | 27.25 (30.80) | 51.20 (51.70) |
| $m^2_{(2,6)}$ | 47.90 (56.30) | 83.45 (86.25) | 98.75 (98.80) | 13.40 (19.40) | 30.20 (34.95) | 57.55 (59.90) |
| $m^2_{(2,7)}$ | 45.30 (54.30) | 81.75 (85.65) | 98.70 (98.85) | 13.85 (20.25) | 32.00 (37.55) | 59.85 (62.45) |
| $m^2_{(2,8)}$ | 41.75 (52.55) | 80.90 (85.00) | 98.60 (98.85) | 12.80 (19.15) | 30.85 (38.80) | 61.80 (63.45) |
| $S(\ddot{\boldsymbol{\theta}}_N)$ | 0.15 (5.10) | 3.25 (8.35) | 8.70 (12.70) | 0.10 (4.40) | 2.00 (5.00) | 3.90 (6.75) |

(Table 3 continued)

| | (e) MA(2) Errors | | | (f) Slope Heterogeneity | | |
|---|---|---|---|---|---|---|
| **Test,N** | **100** | **200** | **400** | **100** | **200** | **400** |
| | $T=6$ | | | $T=6$ | | |
| $m_2$ | 3.80 (4.00) | 5.05 (4.85) | 8.85 (8.45) | 11.90 0.(11.45) | 15.95 (17.05) | 26.45 (25.85) |
| $m^2_{(2,3)}$ | 8.45 (10.90) | 13.50 (16.00) | 22.90 (23.60) | 10.45 (14.70) | 18.40 (23.00) | 36.40 (38.20) |
| $m^2_{(2,4)}$ | 8.60 (11.30) | 13.90 (16.20) | 30.15 (30.55) | 10.00 (14.25) | 19.80 (23.40) | 40.25 (42.05) |
| $S(\ddot{\boldsymbol{\theta}}_N)$ | 2.50 (5.40) | 3.85 (5.60) | 4.15 (6.10) | 2.00 (5.95) | 3.20 (5.45) | 4.30 (5.60) |
| | $T=10$ | | | $T=10$ | | |
| $m_2$ | 5.65 (5.90) | 9.00 (9.60) | 19.90 (20.35) | 14.45 (12.95) | 20.95 (22.60) | 34.45 (34.70) |
| $m^2_{(2,3)}$ | 17.30 (19.05) | 28.75 (30.20) | 54.00 (53.95) | 17.10 (18.70) | 31.00 (33.70) | 56.60 (54.15) |
| $m^2_{(2,4)}$ | 23.90 (26.75) | 44.05 (47.70) | 75.30 (76.05) | 21.45 (22.80) | 38.00 (43.45) | 68.65 (68.15) |
| $m^2_{(2,5)}$ | 23.30 (28.85) | 46.55 (50.25) | 78.90 (79.50) | 21.30 (26.05) | 41.45 (46.30) | 77.10 (75.00) |
| $m^2_{(2,6)}$ | 23.40 (30.65) | 47.90 (52.10) | 83.50 (84.25) | 21.70 (27.35) | 42.70 (48.60) | 80.05 (79.10) |
| $m^2_{(2,7)}$ | 23.20 (30.30) | 48.80 (54.45) | 84.05 (85.45) | 19.85 (26.55) | 42.55 (50.05) | 80.20 (80.40) |
| $m^2_{(2,8)}$ | 20.65 (29.95) | 47.55 (54.90) | 83.50 (85.65) | 18.05 (25.80) | 41.30 (50.45) | 81.15 (80.65) |
| $S(\ddot{\boldsymbol{\theta}}_N)$ | 0.20 (4.65) | 1.40 (6.60) | 4.75 (7.50) | 0.15 (6.00) | 2.80 (5.20) | 3.20 (5.30) |

| | (g) Cross Section Dependence | | |
|---|---|---|---|
| **Test,N** | **100** | **200** | **400** |
| | $T=6$ | | |
| $m_2$ | 29.30 (29.90) | 41.50 (40.30) | 48.20 (51.75) |
| $m^2_{(2,3)}$ | 43.00 (46.15) | 57.05 (57.75) | 68.20 (72.50) |
| $m^2_{(2,4)}$ | 49.70 (54.20) | 65.50 (65.25) | 76.05 (78.20) |
| $S(\ddot{\boldsymbol{\theta}})$ | 20.95 (29.90) | 42.60 (46.75) | 64.55 (69.25) |
| | $T=10$ | | |
| $m_2$ | 20.45 (19.60) | 27.65 (26.35) | 37.75 (40.25) |
| $m^2_{(2,3)}$ | 29.30 (27.60) | 38.20 (40.90) | 53.90 (55.35) |
| $m^2_{(2,4)}$ | 34.60 (34.75) | 47.35 (50.55) | 65.20 (66.40) |
| $m^2_{(2,5)}$ | 39.50 (43.05) | 55.00 (57.55) | 72.25 (74.05) |
| $m^2_{(2,6)}$ | 46.20 (48.05) | 61.65 (64.55) | 78.50 (79.60) |
| $m^2_{(2,7)}$ | 48.80 (53.85) | 65.70 (69.15) | 81.90 (83.95) |
| $m^2_{(2,8)}$ | 51.80 (57.90) | 69.70 (72.95) | 83.70 (85.80) |
| $S(\ddot{\boldsymbol{\theta}}_N)$ | 6.25 (27.20) | 42.05 (56.85) | 76.55 (82.35) |

Notes: The data is generated as $y_{it} = \alpha_i + \beta x_{it} + u_{it}$, where $\alpha_i \sim iidN(0,1)$; $\beta = 1$ except panel (f), where $\beta_i \sim iidN(1,1)$; $u_{it}$ is specified by (40)-(45), but under the null $u_{it} = \varepsilon_{it}$, $\varepsilon_{it} = \sigma_i \varphi_t \epsilon_{it}$, $\sigma_i \sim iidU[0.5, 1.5]$, $\varphi_t = 0.5$ for $t = -49, ..., 0$ and $\varphi_t = 0.5 + 0.1(t-1)$ for $t = 1, ..., T$, and $\epsilon_{it} \sim iid\chi^2(1) - 1$; $x_{it} = \alpha_i + 0.5x_{i,t-1} + 0.5u_{i,t-1} + v_{it}$, $v_{it} \sim iidN(0,1)$, $i = 1, 2, ..., N; t = -48, -47, ..., T$, with $y_{i,-49} = 0$ and $x_{i,-49} = 0$. $m_2$ signifies the Arellano and Bond (1991) test, $m^2_{(2,p)}$ signifies the proposed joint test for second to $p^{th}$ order first differenced error serial correlation, $S(\ddot{\boldsymbol{\theta}}_N)$ signifies the overidentifying restrictions test. All tests are based on optimal two-step Arellano and Bond (1991) GMM estimator. The $m_2$ test results is based on the $m^2_{(2,2)}$ test, and the $m^2_{(2,p)}$ statistics are compared to $\chi^2(p-1)$ distributions. The $S(\ddot{\boldsymbol{\theta}}_N)$ statistic is compared to $\chi^2(h_x - 1)$ distributions, where $h_x$ is defined by (8). The figures in parenthesis are size-adjusted power, which are based on the simulated distributions of test statistics with 10000 replications. All tests are conducted at 5% significance level. All experiments are based on 2000 replications.