ASOCIACION ARGENTINA
DE ECONOMIA POLITICA

ANALES | ASOCIACION ARGENTINA DE ECONOMIA POLITICA

L Reunión Anual

Noviembre de 2015

ISSN 1852-0022
ISBN 978-987-28590-3-9

Intra-cluster correlation as serial vs. equicorrelation in a hierarchical linear model.

Alejo, Javier
Montes-Rojas, Gabriel
Sosa Escudero, Walter

Intra-cluster correlation as serial vs. equicorrelation in a hierarchical linear model *

Javier Alejo

Universidad Nacional de La Plata and CONICET

Gabriel Montes-Rojas

CONICET and Universidad de San Andres

Walter Sosa-Escudero

Universidad de San Andrés and CONICET

August 27, 2015

Abstract

This paper proposes a simple hierarchical model and a testing strategy to identify intra-cluster correlations. Intra-group correlations are modeled as a combination of nested random effects and serially correlated error components in a hierarchical model. A Neyman $C(\alpha)$ framework is used to derive LM-type tests to identify the appropriate level of clustering and the type of intra-group correlation.

Keywords: Clusters, random effects, serial correlation.

JEL Classification: I14, I18, I19

*Corresponding author: Gabriel Montes-Rojas. CONICET - Universidad de San Andrés, Vito Dumas 285, Victoria, Pcia. de Buenos Aires, Argentina. Email: gmontesrojas@udesa.edu.ar.

1 Introduction

Intra-group correlation has received considerable interest in the applied and theoretical literature. When the data can be grouped in clusters it is the rule rather than the exception that observations within a group are not independent. Failure to accomodate these interactions can lead to misleading statistical inferences, as highlighted by the influential article by Bertrand, Duflo and Mullainathan (2004); a concern that dates back to Moulton's (1986) seminal paper.

The empirical practice relies on 'cluster robust methods', that is, for example, on estimates of standard errors that explicitly allow for correlations among observations within a group. The reliability of such strategy comes at a cost, since its consistency depends on the number of independent groups growing large. This is problematic in the case where grouping obeys a nested structure, as would be the case of students in a given class, in a particular school, etc. In this scenario a safer strategy that allows for arbitrary correlations at a larger group (say, at the school instead of the class level) comes at the price of leaving fewer independent groups, rendering asymptotic approximations less reliable. The recent exhaustive survey by Cameron and Miller (2015) points out that 'there is no general solution to this trade-off, and there is no formal test of the level at which to cluster. The consensus is to be conservative and avoid bias and use bigger and more aggregated clusters when possible, up to and including the point at which there is concern about having too few clusters." (p.21).

This paper proposes a testing framework for the appropriate level of clustering in a hierarchical linear model. The most obvious type of intra-group correlation arises when all observations within a group share an unobserved common factor, hence all observations in a group are 'equicorrelated' in the sense that all pairwise correlations are the same. Beyond equicorrelation little can be said if observations within a group do not follow a relevant ordering, like, for example, students sorted alphabetically in a class. Hence more complex patterns requires imposing a sensible ordering within the cluster. A natural one that has received particular consideration in Bertrand et al.'s (2004) influential article is time, that is, cluster correlation induced when observations are sorted chronologically. We will model intra-group correlations as a combination of random effects and serially correlated error components, in a nested, hierarchical structure. Consequently, the problem of what and how to cluster observations is related to identifying a) the 'finest' grouping structure that leaves out more independent groups and, b) the type of intra-cluster correlation, in the form of either random effects, serial correlation or both.

Baltagi, Song and Jung (2002a, 2002b) propose tests for nested random effects allowing for serial correlation at the 'finest' level (students, in our example). Our testing strategy allows for serial correlation at *both* hierarchical levels, jointly or conditional on the presence of the other. Consequently, the proposed testing procedure allows researchers to detect whether intra group correlation operates at a finer or coarser level and also if it is due to equicorrelation, serially correlation or both.

Tests are derived in Neyman's $C(\alpha)$ likelihood framework under normality, based on simple consistent estimators for nuisance parameters. An extensive Monte Carlo exercise is implemented to explore the performance of the proposed tests in finite samples and under non-Gaussian distributions.

The paper is organized as follows. The next section discusses a simple model for grouped data and the relevant hypotheses for intra-cluster correlations. Section 3 derives tests for all possible combination of cluster effects. The reliability of the asymptotic results in a small sample context is evaluated in a comprehensive Monte Carlo experiment in Section 4. Section 5 presents an empirical case that illustrates how to implement the proposed testing strategy in practice. Section 5 concludes.

2 Nested intra-group and serial correlation

Consider a hierarchical linear model with two nested cluster groups,

$$y_{ijt} = x'_{ijt}\beta + u_{ijt},$$

$$u_{ijt} = \phi_i + \delta_{it} + \mu_{ij} + \nu_{ijt},$$

$i = 1, 2, ...M, \quad j = 1, 2, ..., N, \quad t = 1, 2, ..., T.$. As in Baltagi et al. (2001), each observation (i, j, t) will be referred to as corresponding to individual j in group i and period t. To simplify notation and derivations we will assume a balanced panel data.

The error structure allows for unobserved heterogeneity at the i, it, ij and ijt levels, respectively, in the form of unobserved random effects. The presence of two hierarchical levels leads to two autocorrelation patterns. Consider two nested stationary AR(1) processess:

$$\delta_{it} = \lambda\delta_{it-1} + \eta_{it}, \quad 0 \leq |\lambda| < 1,$$

$$\nu_{ijt} = \rho\nu_{ijt-1} + \epsilon_{ijt}, \quad 0 \leq |\rho| < 1.$$

A canonical example for this model may be the following. Consider M classrooms each with N students observed during T periods, where each student belongs to only one classroom. Let y denote a learning outcome such as GPA. Intra cluster correlation in the unobservables may occur due to the presence of an unobserved time invariant term that is student specific (μ_{ij}, i.e. ability, family background) or classroom specific (ϕ_i, i.e. teachers' effect). Alternatively, intra-group dependences may arise due to the time dependence of shocks at the student or classroom levels, modelled as AR(1) processes in our case.

The full null hypothesis of no cluster effects is the joint null of no random effects nor serial correlation at both levels. Departures away from this joint null are informative about two practical issues. The first one is the decision about 'what to cluster over', that is, choosing the appropriate hierarchical level up to which to allow for possible intra-group correlations. As mentioned in the Introduction, this is a crucial question since allowing for correlations at a bigger level leaves fewer groups of independent observations, harming the reliability of cluster robust standard errors. Secondly, it is relevant to know not only the level at which to cluster but also the source of intra-group correlation, as a previous step in deciding how to handle correlations to estimate standard errors consistently. For example, under the null of no serial correlation only random effects cause intracluster correlation, in which case minimum norm quadratic unbiased estimates of the parameters needed to construct estimates of the variances of GLS estimators can be simply derived (as in Matyas (1996, pp. 61)), which may have a considerable advantage over cluster robust methods in the few groups scenario, specially in terms of bias.

Consequently, in this setup testing for cluster correlations amount to checking for random effects and serial correlation at different hierarchical levels. When there is only one hierarchical level (students in different periods, for example), the problem reduces to learning the source of intra-group correlation in the form of random effects or serial correlation. The classic Breusch and Pagan (1980) test for random effects assumes no serial correlation and, symmetrically, the test by Baltagi and Li (1991) checks for first order serial correlation assuming no random effects. Bera, Sosa-Escudero and Yoon (2001) point out that both tests reject their nulls incorrectly when the unwanted effect is present, that is, the Breusch-Pagan test rejects under serial correlation even when no random effects are present and a similar concern affects the test by Baltagi and Li (1991). Consequently, both tests might detect intra-group correlation but are unable to identify its source. Bera, Sosa-Escudero and Yoon (2001) propose a modification that can identify each effect separately. Inoue and Solon (2006) propose a test for first order serial correlation after fixed effect estimation.

When more than one hierarchical level is allowed for, Baltagi, Song and

Jung (2002b) develop LM tests for random effects in a nested error components model, but with no serial correlation. Baltagi, Song and Jung (2002a) allow for serial correlation although at the finest level only (i.e. ijt). By allowing a full nested autocorrelation structure, the testing strategy proposed in this paper can correctly identify the level at which cluster effects take place and their sources, that is, whether they are caused by unobserved random effects and/or serial correlation and, more importantly, at which hierarchical level each of them operates.

Related strategies include Kézdi (2004), who proposes a based on the comparisson of variance estimates with or without allowing for cluster correlation, in the spirit of the classic White test for heteroskedasticity. King and Roberts (2015) propose a similar procedure using the generalized information matrix. These two procedures do not detect the appropriate level of clusters since they only check for differences with respect to the joint null of absence of cluster correlation.

3 Tests for cluster effects

Let $x_i = [x_{i11}, ..., x_{i1T}, ..., x_{iN1}, ..., x_{iNT}]$. We will make the following assumptions:

Assumption 1:

$\{y_{i\cdot}, x_i, \phi_i, \eta_{i\cdot}, \mu_{i\cdot}, \epsilon_{i\cdot}\}_{i=1}^{M}$ is an independent and identically distributed random sample;

Assumption 2:

Correct mean specification: $E[\phi_i|x_i] = E[\eta_{it}|x_i] = E[\mu_{ij}|x_i] = E[\epsilon_{ijt}|x_i] = 0, \forall i, j, t$;

Assumption 3:

Variance: $Var[\phi_i|x_i] = \sigma^2_\phi$, $Var[\delta_{it}|x_i] = \sigma^2_\delta = \sigma^2_\eta/(1-\lambda^2)$, $Var[\mu_{ij}|x_i] = \sigma^2_\mu$, $Var[\nu_{ijt}|x_i] = \sigma^2_\nu = \sigma^2_\epsilon/(1-\rho^2)$;

Assumption 4:

Nested autocovariance structure: $Cov[\delta_{it}, \delta_{ih}|x_i] = \lambda^{|t-h|}\sigma^2_\eta, \forall i, h, t,\ 0 \leq |\lambda| < 1$, $Cov[\nu_{ijt}, \nu_{ijh}|x_i] = \rho^{|t-h|}\sigma^2_\epsilon, \forall i, j, t, h,\ 0 \leq |\rho| < 1$.

Assumption 5:

Normality: $\phi_i \sim i.i.d.\ N(0, \sigma^2_\phi), \forall i$, $\mu_{ij} \sim i.i.d.\ N(0, \sigma^2_\mu), \forall i, j$, $\eta_{it} \sim i.i.d.\ N(0, \sigma^2_\eta), \forall i, j, t$, $\epsilon_{ijt} \sim i.i.d.\ N(0, \sigma^2_\epsilon)$.

In matrix form the model can be written as

$$y = X\beta + u,$$

$$\Omega \equiv E(uu^\top) = \sigma^2_\phi(I_M \otimes J_N \otimes J_T) + \sigma^2_\delta(I_M \otimes J_N \otimes V_\lambda) + \sigma^2_\mu(I_M \otimes I_N \otimes J_T) + \sigma^2_\nu(I_M \otimes I_N \otimes V_\rho), \tag{1}$$

where y and e are the $MNT \times 1$ column vectors with all the dependent variable and residual observations, and X is the $MNT \times K$ matrix with the observable covariates. $I_\cdot$ is a $\cdot$-dimensional identity matrix, $J_\cdot$ is a $\cdot$-dimensional matrix of ones,

$$V_\lambda = \begin{bmatrix} 1 & \lambda & \lambda^2 & \cdots & \lambda^{T-1} \\ \lambda & 1 & \lambda & \cdots & \lambda^{T-2} \\ \lambda^2 & \lambda & 1 & \cdots & \lambda^{T-3} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \lambda^{T-1} & \lambda^{T-2} & \lambda^{T-3} & \cdots & 1 \end{bmatrix},$$

$$V_\rho = \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{T-1} \\ \rho & 1 & \rho & \cdots & \rho^{T-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{T-3} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \rho^{T-3} & \cdots & 1 \end{bmatrix},$$

and $\otimes$ is the Kronecker product. For future reference define the idempotent matrices $\bar{J}_\cdot \equiv \frac{1}{\cdot}J_\cdot$ and $\bar{E}_\cdot \equiv I_\cdot - \bar{J}_\cdot$, which correspond to the projection and residual projection matrices, respectively, on a set of dummy variables for the $\cdot$ level.

The log likelihood function for this problem is given by

$$L(\beta,\theta) \propto -\frac{1}{2}\ln|\Omega| - \frac{1}{2}u^{\top}\Omega^{-1}u,$$

with $\theta = (\sigma^2_\phi, \sigma^2_\eta, \sigma^2_\mu, \sigma^2_\epsilon, \rho, \lambda)$ and Ω is given by equation (1).

Baltagi, Song and Jung (2002a) develop LM tests for random effects in a nested error components model, assuming $\sigma^2_\eta = \lambda = \rho = 0$. They derive tests for the joint null $H_0^{\sigma^2_\phi,\sigma^2_\mu} : \sigma^2_\phi = \sigma^2_\mu = 0$ and for the conditional hypotheses $H_0^{\sigma^2_\mu} : \sigma^2_\mu = 0$, assuming $\sigma^2_\phi \geq 0$, and $H_0^{\sigma^2_\phi} : \sigma^2_\phi = 0$, assuming $\sigma^2_\mu \geq 0$.

A joint test for no cluster effects in a nested random effects model and no serial correlation at the finest level was studied by Baltagi, Song and Jung (2002b). That is, their null hypotesis is $H_0^{\sigma^2_\phi,\sigma^2_\mu,\rho} : \sigma^2_\phi = \sigma^2_\mu = \rho = 0$, assuming $\sigma^2_\eta = \lambda = 0$.

In this paper we develop tests for detecting the appropriate level of autocorrelation in a nested random effects structure. We propose tests for the join null of no serial correlation at any hierarchical level ($H_0^{\rho,\lambda} : \rho = 0, \lambda = 0$, assuming $\sigma^2_\phi \geq 0, \sigma^2_\eta \geq 0, \sigma^2_\mu \geq 0$) and conditional tests for one type of serial correlation given that the other is present, that is, $H_0^{\rho} : \rho = 0$, assuming $\sigma^2_\phi \geq 0, \sigma^2_\eta \geq 0, \sigma^2_\mu \geq 0, 0 \leq |\lambda| < 1$ and $H_0^{\lambda} : \lambda = 0$, assuming $\sigma^2_\phi \geq 0, \sigma^2_\eta \geq 0, \sigma^2_\mu \geq 0, 0 \leq |\rho| < 1$. The combination of the proposed tests with those previously proposed by Baltagi et al (2002b) allows researchers to fully identify the levels and the sources of intra-cluster correlation, and decide on a appropriate strategy to handle it. For example, and as mentioned in the Introduction under the joint null of no serial correlation, a hierarchical FGLS strategy produces asymptotically efficient estimates of the parameters of interest and consistent estimates of their variances, which may have considerable advantages over cluster robuts methods who unnecesarily allow intraclass correlations to vary. Also, in the case of serial correlation, the tests identify whether it takes place at the fine or coarse level, indicating at which level to cluster observations, which, as stressed previously, is crucial to maximize the number of independent groups in order to make asymptotic approximations more reliable if clustering occurs at the finer level.

Let $\theta \in \Theta \subseteq \mathbb{R}^p$. Using the formulas in Harville (1977, p.326) (see also Baltagi, 2003, p.63) the score functions can be expressed as

$$s_r(\theta) = \partial L/\partial \theta_r = -\frac{1}{2} tr[\Omega^{-1} \partial \Omega / \partial \theta_r] + \frac{1}{2}[u^\top \Omega^{-1} (\partial \Omega / \partial \theta_r) \Omega^{-1} u],$$

for $1 \leq r \leq p$. The information matrix $\mathcal{J}$ can be obtained as

$$\begin{aligned} \partial^2 L/\partial \theta_r \partial \theta_k &= \frac{1}{2} tr \left[\Omega^{-1} \left\{ \frac{\partial^2 \Omega}{\partial \theta_r \partial \theta_k} - \frac{\partial \Omega}{\partial \theta_r} \Omega^{-1} \frac{\partial \Omega}{\partial \theta_k} \right\} \right] \\ &+ \frac{1}{2} u^\top \Omega^{-1} \left[\frac{\partial \Omega}{\partial \theta_r \partial \theta_k} - 2 \frac{\partial \Omega}{\partial \theta_r} \Omega^{-1} \frac{\partial \Omega}{\partial \theta_r} \right] \Omega^{-1} u, \end{aligned}$$

and

$$\mathcal{J}_{rk}(\theta) \equiv -E[\partial^2 L / \partial \theta_r \partial \theta_k] = \frac{1}{2} tr \left[\Omega^{-1} \frac{\partial \Omega}{\partial \theta_r} \Omega^{-1} \frac{\partial \Omega}{\partial \theta_k} \right].$$

Note that

$$\begin{aligned} \partial \Omega / \partial \sigma_\phi^2 &= (I_M \otimes J_N \otimes J_T), \\ \partial \Omega / \partial \sigma_\eta^2 &= \frac{1}{1-\lambda^2} (I_M \otimes J_N \otimes V_\lambda), \\ \partial \Omega / \partial \sigma_\mu^2 &= (I_M \otimes I_N \otimes J_T), \\ \partial \Omega / \partial \sigma_\epsilon^2 &= \frac{1}{1-\rho^2} (I_M \otimes I_N \otimes V_\rho), \\ \partial \Omega / \partial \rho &= -\frac{2\rho\sigma_\epsilon^2}{(1-\rho^2)^2} (I_M \otimes I_N \otimes V_\rho) + \frac{\sigma_\epsilon^2}{(1-\rho^2)} (I_M \otimes I_N \otimes V_\rho'), \\ \partial \Omega / \partial \lambda &= -\frac{2\lambda\sigma_\eta^2}{(1-\lambda^2)^2} (I_M \otimes J_N \otimes V_\lambda) + \frac{\sigma_\eta^2}{(1-\lambda^2)} (I_M \otimes J_N \otimes V_\lambda'), \end{aligned}$$

where

$$V'_\rho = \begin{bmatrix} 0 & 1 & 2\rho & \cdots & (T-1)\rho^{T-2} \\ 1 & 0 & 1 & \cdots & (T-2)\rho^{T-3} \\ 2\rho & 1 & 0 & \cdots & (T-3)\rho^{T-4} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ (T-1)\rho^{T-2} & (T-2)\rho^{T-3} & (T-3)\rho^{T-4} & \cdots & 0 \end{bmatrix},$$

and

$$V'_\lambda = \begin{bmatrix} 0 & 1 & 2\lambda & \cdots & (T-1)\lambda^{T-2} \\ 1 & 0 & 1 & \cdots & (T-2)\lambda^{T-3} \\ 2\lambda & 1 & 0 & \cdots & (T-3)\lambda^{T-4} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ (T-1)\lambda^{T-2} & (T-2)\lambda^{T-3} & (T-3)\lambda^{T-4} & \cdots & 0 \end{bmatrix}.$$

In order to construct LM tests, first note that the block diagonality between β and θ allow us to focus on the scores corresponding to θ only. Second, consistent estimators of θ under the null can be obtained using an ANOVA-type analysis (in particular see Baltagi and Li, 1991, Baltagi, Jung and Song, 2001; see also the Appendices). Hence our tests will be based on Neyman's $C(\alpha)$ principle, which produces tests that are asymptotically equivalent to likelihood based LM tests under any initial $\sqrt{N}$-consistent non-ML estimation of the nuisance parameters.

Consider a partition of $\theta = (\theta_1', \theta_2')'$, where θ_2 contains the parameters under the corresponding null hypothesis $H_0^2 : \theta_2 = 0$, and θ_1 the nuisance paramaters that need to be estimated. In our particular case, θ will be partitioned into either $\theta_1 = (\sigma_\phi^2, \sigma_\eta^2, \sigma_\mu^2, \sigma_\epsilon^2)', \theta_2 = (\rho, \lambda)'$ (subsection 3.1), $\theta_1 = (\sigma_\phi^2, \sigma_\eta^2, \sigma_\mu^2, \sigma_\epsilon^2, \rho)', \theta_2 = \lambda$ (subsection 3.2) or $\theta_1 = (\sigma_\phi^2, \sigma_\eta^2, \sigma_\mu^2, \sigma_\epsilon^2, \lambda)', \theta_2 = \rho$ (subsection 3.3). Correspondingly, the score will be partitioned as $s(\theta) = (s_1(\theta), s_2(\theta))'$, and the information matrix as $\mathcal{J}(\theta) = \begin{bmatrix} \mathcal{J}_{11}(\theta) & \mathcal{J}_{12}(\theta) \\ \mathcal{J}_{22}(\theta) & \mathcal{J}_{22}(\theta) \end{bmatrix}$.

Conditional LM statistics for H_0^2 under ML estimation are defined as

$$LM_2(\theta) = s_2(\theta)'[\mathcal{J}_{22}(\theta) - \mathcal{J}_{21}(\theta)\mathcal{J}_{11}^{-1}(\theta)\mathcal{J}_{12}(\theta)]^{-1}s_2(\theta).$$

Neyman's $C(\alpha)$ adjusted scores are defined as

$$s_{2\cdot 1}(\theta) \equiv s_2(\theta) - \mathcal{J}_{21}(\theta)\mathcal{J}_{11}^{-1}(\theta)\mathcal{J}_{12}(\theta)s_1(\theta).$$

Then, the Neyman's $C(\alpha)$ LM statistic is

$$LM_{2\cdot 1}(\theta) = s_{2\cdot 1}(\theta)'[\mathcal{J}_{22}(\theta) - \mathcal{J}_{21}(\theta)\mathcal{J}_{11}^{-1}(\theta)\mathcal{J}_{12}(\theta)]^{-1}s_{2\cdot 1}(\theta).$$

A well known result is that $LM_{2\cdot 1}(\hat{\theta}^*) \xrightarrow{d} \chi^2_{dim(\theta_2)}$, where $\hat{\theta}^*$ is any consistent estimator under the corresponding null hypothesis.

3.1 LM test for serial correlation under random effects: $H_0^{\rho,\lambda} : \rho = 0, \lambda = 0$, assuming $\sigma^2_\phi \geq 0, \sigma^2_\eta \geq 0, \sigma^2_\mu \geq 0$

Consider first a test for no autocorrelation at both hierarchical levels but allowing for a nested error component random effects structure. In this case

$$\Omega_0 = \sigma^2_\phi(I_M \otimes J_N \otimes J_T) + \sigma^2_\eta(I_M \otimes J_N \otimes I_T) + \sigma^2_\mu(I_M \otimes I_N \otimes J_T) + \sigma^2_\epsilon(I_M \otimes I_N \otimes I_T),$$

and

$$\begin{aligned}
\partial\Omega/\partial\sigma^2_\phi|_{H_0} &= (I_M \otimes J_N \otimes J_T),\\
\partial\Omega/\partial\sigma^2_\eta|_{H_0} &= (I_M \otimes J_N \otimes I_T),\\
\partial\Omega/\partial\sigma^2_\mu|_{H_0} &= (I_M \otimes I_N \otimes J_T),\\
\partial\Omega/\partial\sigma^2_\epsilon|_{H_0} &= (I_M \otimes I_N \otimes I_T),\\
\partial\Omega/\partial\rho|_{H_0} &= \sigma^2_\epsilon(I_M \otimes I_N \otimes B_T),\\
\partial\Omega/\partial\lambda|_{H_0} &= \sigma^2_\eta(I_M \otimes J_N \otimes B_T),
\end{aligned}$$

where B_T is a $T \times T$ bi-diagonal matrix, that is, with zeros in all its elements except $b_{t,t-1} = b_{t,t+1} = 1$ for $t = 1, 2, 3, ..., T$.

For this case define $\theta_1 = (\sigma^2_\phi, \sigma^2_\eta, \sigma^2_\mu, \sigma^2_\epsilon)$ and $\theta_2 = (\rho, \lambda)$. Appendix A1 provides consistent estimates for all the elements of θ under the null hypothesis, $\hat{\theta} = (\hat{\sigma}^2_\phi, \hat{\sigma}^2_\eta, \hat{\sigma}^2_\mu, \hat{\sigma}^2_\epsilon, 0, 0)$.

Then a test for absence of autocorrelation at any level can be constructed by replacing in $LM_{2\cdot 1}$ all unknown quantities by its consistent estimates, using the matrix derivative formulae above and replacing the unobserved u by OLS residuals $\hat{u}$. The resulting test statistic will be labeled as $LM_{(\rho,\lambda)\cdot\sigma}$. Computer routines to implement all the proposed tests are available upon request.

3.2 Test for serial correlation at the group level: H_0^λ : $\lambda = 0$, assuming $\sigma^2_\phi \geq 0, \sigma^2_\eta \geq 0, \sigma^2_\mu \geq 0, 0 \leq |\rho| < 1$

This is a test for autocorrelation at the most aggregate level. In this case we have

$$\Omega_0 = \sigma^2_\phi(I_M \otimes J_N \otimes J_T) + \sigma^2_\eta(I_M \otimes J_N \otimes I_T) + \sigma^2_\mu(I_M \otimes I_N \otimes J_T) + \frac{\sigma^2_\epsilon}{1-\rho^2}(I_M \otimes I_N \otimes V_\rho)$$

Note that:

$$\partial\Omega/\partial\sigma^2_\phi|_{H_0} = (I_M \otimes J_N \otimes J_T),$$
$$\partial\Omega/\partial\sigma^2_\eta|_{H_0} = (I_M \otimes J_N \otimes I_T),$$
$$\partial\Omega/\partial\sigma^2_\mu|_{H_0} = (I_M \otimes I_N \otimes J_T),$$
$$\partial\Omega/\partial\sigma^2_\epsilon|_{H_0} = \frac{1}{1-\rho^2}(I_M \otimes I_N \otimes V_\rho),$$
$$\partial\Omega/\partial\rho|_{H_0} = -\frac{2\rho\sigma^2_\epsilon}{(1-\rho^2)^2}(I_M \otimes I_N \otimes V_\rho) + \frac{\sigma^2_\epsilon}{(1-\rho^2)}(I_M \otimes I_N \otimes V'_\rho),$$
$$\partial\Omega/\partial\lambda|_{H_0} = \sigma^2_\eta(I_M \otimes J_N \otimes B_T).$$

Two tests will be proposed for this case. First a test for H_0^λ that imposes $\rho = 0$, that is, assuming that that there is no autocorrelation at the aggregate level while testing for autocorrelation at the individual level. It implicitly defines $\theta_1 = (\sigma_\phi^2, \sigma_\eta^2, \sigma_\mu^2, \sigma_\epsilon^2)$ and $\theta_2 = \lambda$. This is based on $LM_{2\cdot 1}$ and will be defined as $LM_{\lambda\cdot\sigma}$, a marginal LM statistic. Second, consider an estimator of ρ. ρ is estimated jointly with the rest of the σs. Appendix A2 provides consistent estimates of θ under the null hypothesis, $\tilde{\theta} = (\tilde{\sigma}_\phi^2, \tilde{\sigma}_\eta^2, \tilde{\sigma}_\mu^2, \tilde{\sigma}_\epsilon^2, \tilde{\rho}, 0)$. For this case let $\theta_1 = (\sigma_\phi^2, \sigma_\eta^2, \sigma_\mu^2, \sigma_\epsilon^2, \rho)$ and $\theta_2 = \lambda$ and consider a $LM_{2\cdot 1}$ test defined as $LM_{\lambda\cdot(\sigma,\rho)}$, a conditional LM statistic.

3.3 LM test for autocorrelation at the individual level: $H_0^\rho : \rho = 0$, assuming $\sigma_\phi^2 \geq 0, \sigma_\eta^2 \geq 0, \sigma_\mu^2 \geq 0, 0 \leq |\lambda| < 1$

This is a test for autocorrelation at the individual level.

$$\Omega_0 = \sigma_\phi^2(I_M \otimes J_N \otimes J_T) + \frac{\sigma_\eta^2}{1-\lambda^2}(I_M \otimes J_N \otimes V_\lambda) + \sigma_\mu^2(I_M \otimes I_N \otimes J_T) + \sigma_\epsilon^2(I_M \otimes I_N \otimes I_T)$$

Note that

$$\begin{aligned}
\partial\Omega/\partial\sigma_\phi^2|_{H_0} &= (I_M \otimes J_N \otimes J_T), \\
\partial\Omega/\partial\sigma_\eta^2|_{H_0} &= \frac{1}{1-\lambda^2}(I_M \otimes J_N \otimes V_\lambda), \\
\partial\Omega/\partial\sigma_\mu^2|_{H_0} &= (I_M \otimes I_N \otimes J_T), \\
\partial\Omega/\partial\sigma_\epsilon^2|_{H_0} &= (I_M \otimes I_N \otimes I_T), \\
\partial\Omega/\partial\rho|_{H_0} &= \sigma_\epsilon^2(I_M \otimes I_N \otimes B_T), \\
\partial\Omega/\partial\lambda|_{H_0} &= -\frac{2\lambda\sigma_\eta^2}{(1-\lambda^2)^2}(I_M \otimes J_N \otimes V_\lambda) + \frac{\sigma_\eta^2}{1-\lambda^2}(I_M \otimes J_N \otimes V_\lambda'),
\end{aligned}$$

Once again, two different tests will be derived. The first one is a test for H_0^ρ that imposes $\lambda = 0$, that is, it assumes that there is not autocorrelation at the aggregate level while it tests for autocorrelation at the individual level. It implicitly defines $\theta_1 = (\sigma_\phi^2, \sigma_\eta^2, \sigma_\mu^2, \sigma_\epsilon^2)$ and $\theta_2 = \rho$. This is based on $LM_{2\cdot 1}$ and will be defined as $LM_{\rho\cdot\sigma}$, a marginal LM statistic. The second test checks for serial correlation after having estimated λ. Appendix A3 provides consistent estimates of θ under the null hypothesis, $\check{\theta} = (\check{\sigma}^2_\phi, \check{\sigma}^2_\eta, \check{\sigma}^2_\mu, \check{\sigma}^2_\epsilon, 0, \check{\lambda})$. For this case let $\theta_1 = (\sigma^2_\phi, \sigma^2_\eta, \sigma^2_\mu, \sigma^2_\epsilon, \lambda)$ and $\theta_2 = \rho$ and consider a $LM_{2\cdot 1}$ test defined as $LM_{\rho\cdot(\sigma,\lambda)}$, a conditional LM statistic.

4 Monte Carlo experiments

This section explores the small sample performance of the proposed tests through and extensive Monte Carlo experiment.

The hierarchical model is:

$$y_{ijt} = \beta_1 x_{1,i} + \beta_2 x_{2,it} + \beta_3 x_{3,ij} + \beta_4 x_{4,ijt} + u_{ijt}, \qquad (2)$$
$$u_{ijt} = \phi_i + \delta_{it} + \mu_{ij} + \nu_{ijt},$$
$$i = 1, ...M,\ j = 1, 2, ..., N,\ t = 1, 2, ..., T,$$

with $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 1$ and $\rho_x = \lambda_x = 0.5$. Let $[v_{1,i}, v_{2,it}, v_{3,ij}, v_{4,ijt}]$ and $[\phi_i, \eta_{it}, \epsilon_{ijt}]$ be independent and identically distributed vectors of $N(0, 1)$ random variables and μ_{ij} is N(0,0.1). We set $x_{1,i} = v_{1,i}$ and $x_{3,ij} = x_{1,i} + v_{3,ij}$. We consider two AR(1) structures for both the covariates and error terms,

$$\begin{aligned} \delta_{it} &= \lambda\delta_{it-1} + \eta_{it}, |\lambda| < 1, \\ \nu_{ijt} &= \rho\nu_{ijt-1} + \epsilon_{ijt}, |\rho| < 1, \\ x_{2,it} &= x_{1,i} + \lambda_x x_{2,it-1} + v_{2,it}, |\lambda_x| < 1, x_{2,i1} = x_{1,i} + v_{2,i1}, \\ x_{4,ijt} &= x_{1,i} + x_{2,it} + x_{3,ij} + \rho_x x_{4,ijt-1} + v_{ijt}, |\rho_x| < 1, \end{aligned}$$

$x_{2,ij1} = x_{1,i} + x_{2,i1} + x_{3,ij} + v_{4,ij1}$.

We consider different panel sizes with $M \in \{5, 10\}$ (i.e. number of school districts), $N \in \{5, 10\}$ (i.e. number of schools within district) and $T = \{3, 5, 10\}$ (i.e. number of repeated observations of the same school). We evaluate the tests using a nominal size of 0.05 and 1000 replications. For all data generating processes we consider the performance of the LM statistics constructed above: (i) joint test for $H_0 : \lambda = \rho = 0$, $LM_{(\rho,\lambda)\cdot\sigma}$, (ii) tests for $H_0 : \lambda = 0$, $LM_{\lambda\cdot\sigma}$, $LM_{\lambda\cdot(\sigma,\rho)}$, and (iii) for $H_0 : \rho = 0$, $LM_{\rho\cdot\sigma}$, $LM_{\rho\cdot(\sigma,\lambda)}$.

The Monte Carlo studies can be divided into two strands. First, we analyze the tests size performance under the joint null hypothesis $H_0 : \rho = \lambda = 0$. and different values of the autocorrelation parameters: $\rho \in \{0, 0.2\}$ and $\lambda \in \{0, 0.2\}$. Second, we analyze the tests' relative power performance for $\lambda \in \{0, 0.2\} \times \rho \in \{0, 0.1, 0.2, ..., 0.9\}$ and $\rho \in \{0, 0.2\} \times \lambda \in \{0, 0.1, 0.2, ..., 0.9\}$.

The first set of rows in Table 1 presents Monte Carlo results for the joint test for $H_0 : \rho = \lambda = 0$, $LM_{(\rho,\lambda)\cdot\sigma}$, when σ is estimated as in Appendix A1. The simulations show that the test has approximate correct empirical size for all panel size dimensions, i.e. close to 5% in all cases when $H_0 : \rho = \lambda = 0$ is true. Moreover, when one correlation parameter is increased, keeping the other constant, the test correctly increases its rejection rates. Power increases in all dimensions, that is, as either, M, N or T increases, keeping the other dimensions constant. This implies that the test is consistent in every possible direction. Note that rejection rates for $\rho \neq 0$ are higher than those for $\lambda \neq 0$. This is because a potential estimator of λ uses less information than that of ρ.

The second and third set of rows in Table 1 report the simulation results for testing $H_0 : \lambda = 0$, marginal $LM_{\lambda\cdot\sigma}$, with σ estimated as in Appendix A1, and conditional $LM_{\lambda\cdot(\sigma,\rho)}$ when (σ, ρ) are estimated as in Appendix A2. The tests have correct empirical size when the data is simulated with $\rho = \lambda = 0$. Rejection rates are marginally larger when $\rho \neq 0$ but the size distortion is small. Note that there are marginal improvements when ρ is estimated, i.e. when comparing $LM_{\lambda\cdot\sigma}$ with $LM_{\lambda\cdot(\sigma,\rho)}$. Regarding empirical power, the tests correctly detects $\lambda > 0$. As expected, the marginal tests have better power than the conditional when in fact $\rho = 0$. The simulations show that power increases as either M, N or T increases, keeping the other dimensions constant. However, the best improvement is observed for increments in T.

Finally, the fourth and fifth set of rows in Table 1 show the simulation results for testing $H_0 : \rho = 0$, marginal $LM_{\rho\cdot\sigma}$, with σ estimated as in Appendix A1, and conditional $LM_{\rho\cdot(\sigma,\lambda)}$ when (σ, λ) are estimated as in Appendix A3. The tests have correct empirical size when the data is simulated with $\rho = \lambda = 0$ for $T = 10$ but are undersized with low rejection rates for $T = 3, 5$. Rejection rates do not increase when $\lambda \neq 0$. Note that there are marginal improvements when λ is estimated, i.e. when comparing $LM_{\rho\cdot\sigma}$ with $LM_{\rho\cdot(\sigma,\lambda)}$. Regarding empirical power, the tests correctly detects $\rho \neq 0$. As expected, the marginal tests have slightly better power than the conditional when in fact $\lambda = 0$. The simulations show that power increases as either M, N or T increases, keeping the other dimensions constant.

In order to analyze the test statistics' relative power performance we compute the rejection rates for different parameter values. Figure 1 shows the five tests performance for the cases $\lambda = 0 \times \rho \in \{0, 0.1, 0.2, ..., 0.9\}$ (left) and $\rho = 0 \times \lambda \in \{0, 0.1, 0.2, ..., 0.9\}$ (right). Figure 2 repeats the same exercise for the five tests performance but for the cases $\lambda = 0.2 \times \rho \in \{0.1, 0.2, ..., 0.9\}$ (left) $\rho = 0.2 \times \lambda \in \{0.1, 0.2, ..., 0.9\}$ (right). This is done for the $M = 5, N = 5, T = 10$ panel size and for the same DGP of Table 1.

Figure 1 clearly shows that tests in a particular direction are not contaminated for the presence of autocorrelation in the other nested level. Moreover, the figure shows a clear ordering in terms of empirical power, i.e. *joint* $\leq$ *conditional* $\leq$ *marginal*. Figure 2 shows that the tests correctly identify the presence of autocorrelation in one cluster level, even in the presence of autocorrelation in the other level (set at 0.2).

A particular motivation for the construction of these tests was to disentangle the presence of nested autocorrelation structures, in the presence of nested unobserved effects in the form of random effects. As pointed out by Bera, Sosa-Escudero and Yoon (2001), tests for random effects also reject the null hypothesis in the presence of serial correlation. For our particular nested structure, Baltagi, Song and Jung (2002b) LM tests for nested random effects should also be able reject the null hypothesis pointing out to the presence of serial correlation given by either ρ or λ. We argue that while this is the case, our proposed tests have higher power. To illustrate this point, we compute two alternative tests for random effects only.

1. A test for $H_0 : \sigma^2_\phi = 0$ based on Baltagi et al. (2002a), ignoring serial correlation at any level but estimating σ^2_η, σ^2_μ and σ^2_ϵ. This is defined as $LM_{V(\phi)\cdot\sigma}$ and it is included as an additional test (the sixth LM test in the figures).

2. A test for $H_0 : \sigma^2_\mu = 0$ based on Baltagi et al. (2002a), ignoring serial correlation at any level but estimating σ^2_ϕ, σ^2_η and σ^2_ϵ. This is defined as $LM_{V(\mu)\cdot\sigma}$ and it is included as an additional test (the seventh LM test in the figures).

As expected for the DGP generated with $\sigma^2_\phi = 1$, $LM_{V(\phi)\cdot\sigma}$ has rejection rates above the 5% theoretical level. The test also shows that its rejection rates does not increase with respect to ρ while it does when λ increases, thus providing some evidence on the appropriate level of clustering. However, it is clear that while it might be able to detect serial correlation at the $M-$level, it does so with lower power than the constructed test for λ, such as $LM_{\lambda\cdot\sigma}$. Similar results are obtained for a test for $H_0 : \sigma^2_\mu = 0$, that is, it detects departures from $\rho = 0$ but not on $\lambda = 0$ but with a lower performance than the corresponding tests for ρ. As a result, while we corroborate the fact that tests for nested random-effects are affected by serial correlation (an application of the Bera, Sosa-Escudero and Yoon, 2001, principle to the nested hierarchical model), these tests are not the most appropriate to identify which level to cluster on. These results also complement the findings of Baltagi, Song and Jung (2002a), in which serial correlation in ρ only is condidered.

Tables 2 and 3 repeat the simulations above but with error components that are not normally distributed. In the first case, Table 2, each component follows a centered and standardized chi-squared distribution with 1 degree of freedom ($\frac{\chi_1^2-1}{\sqrt{2}}$). In the second case, Table 2, each component follows a $Student - t$ distribution with 5 degrees of freedom (standardized to have variance of 1, $\sqrt{\frac{(d-2)}{d}}t_d, d = 5$). The tables show that our constructed tests have correct empirical size and similar power performance to those observed in Table 1 for Gaussian errors. This determines that the robustness of 'covariance' tests to non-normalities observed by Honda (1985) in the one-level error-components model also applies to the nested model used in this paper. Thus, our LM tests can be applied to different distributions of the error-components, not just Gaussian. In fact, the tests also have a similar power performance to that of the normal counterparts, as shown by Figures 3, 4, 5 and 6.

5 Empirical application: educational performance

As an empirical illustration we apply the proposed tests to study the dynamics of educational performance. The PISA tests are held every three years in the OECD member countries and a group of partner countries whose number has been growing over time. The program collects harmonized information about students and schools using a single questionnaire in all countries, thus being comparable across countries. Clearly, understanding the channels behind the dynamics of educational performance is a relevant issue for policy making purposes. Therefore, this application shows the importance of having a test that allows to identify the underlying sources of autocorrelation in a reliable way.

The usual approach is to estimate an educational production function. Following Hanushek and Woessmann (2011), we estimate a simplified model given by the linear equation,

$$score_{ijt} = \alpha\, stratio_{ijt} + \beta\, grade_{ijt} + \gamma\, pcgirls_{ijt} + \delta\, hisei_{ijt} + u_{ijt},$$

where the outcome variable is the mean score in a standardized international reading test (*score*), and the covariates include some of the usual inputs proposed in the literature: the average students-teacher ratio (*stratio*), the average school year that students attend (*grade*), the average proportion of girls at school (*pcgirls*) and an average index of socio-economic level (*hisei*). The first covariate is a proxy of the educational resources of the school, the second is a measure of students' experience, while the last two capture differences in educational performance related to demographic and economic factors. Finally, the error term u_{ijt} is assumed the nested structure in Section 2. In this case we have a nested structure given by country and school type. In our panel, i correspond to the country, j to type of school, and t is the year in which the survey information was collected. Since our testing strategy was developed for balanced panel data, we use the sub-sample of the eight ($M = 8$) countries with complete data in the five existing surveys: Austria, Belgium, Switzerland, Spain, Hong Kong, Ireland, Republic of Korea, Portugal and Thailand. Within those countries there are three types of schools ($N = 3$): Private independent, Private government-dependent and Public. This information was collected for 2000, 2003, 2006, 2009 and 2012 ($T = 5$).

Table 4 shows the results of applying our tests to the model of educational performance using this data. At 5% of significance level, the joint hypothesis of no autocorrelation in both cluster groups is rejected. However, further analysis reveals that both tests for $\lambda = 0$ are rejected at the 1% of significance level for $LM_{\lambda\cdot\sigma}$ and 5% for $LM_{\lambda\cdot(\sigma,\rho)}$. The tests for $\rho = 0$ does not provide enough evidence to reject this null hypothesis. Therefore, the temporal persistence in educational performance seems to be related to the country rather than the type of school.

This result has an interesting interpretation. Given the inputs levels, the fact that a country is systematically good (or bad) does not appear to be related to the type of school, i.e. in relation to the public vs. private discussion. Consequently, this dynamic is associated to the general situation of the country as a whole and not to the performance of different types of schools. This component is likely to be associated with the educational culture and the design of the educational system, among others.

6 Discussion and conclusion

The proposed testing framework allows for a comprehensive analysis of the appropriate level of clustering in a multi-level nested longitudinal panel data structure.

The model proposed in this paper suggests several extensions. First, the simulation exercises reveal that the estimates of the nuisance parameter is quite demanding for moderate to large panel sizes (i.e. $M = 10, N = 10, T \geq 10$). The main problem arises from the Wallace and Hussain (1969) transformation of OLS residuals to obtain consistent estimators of the σ parameters. In particular, the inverse of the corresponding matrices is slow in both `STATA` and `R`, the two softwares that were used for implementation. Thus, different alternatives analyzed on Baltagi, Song and Jung (2001) could be explored to speed-up the process.

Second, the quest for the adequate level of clustering should also be analyzed in terms of heteroskedasticity. As argued by Wooldridge (2012) both serial correlation and heteroskedasticity concerns call for cluster robust standard errors, even after GLS random effects estimation. Then, an important extension would be the adpatation of Montes-Rojas and Sosa-Escudero (2011) heteroskedasticity tests for the error-components model to the nested structure combined here. A general testing framework to identify the appropriate level and type of clustering to be used requires nested random effects, serial correlation and heteroskedasticity jointly.

Appendix 1: Estimates of $(\sigma^2_\phi, \sigma^2_\eta, \sigma^2_\mu, \sigma^2_\epsilon)$, assuming $\rho = \lambda = 0$ using invariant quadratic forms

We consider best quadratic unbiased estimators $(\sigma^2_\phi, \sigma^2_\eta, \sigma^2_\mu, \sigma^2_\epsilon)$ as a simple extension of the spectral decomposition given in Wallace and Hussain (1969) and Baltagi (2003, pp.38-39).

Rewriting the variance covariance matrix under the null we have

$$\Omega_0 = \sigma^2_\phi(I_M \otimes J_N \otimes J_T) + \sigma^2_\eta(I_M \otimes J_N \otimes I_T) + \sigma^2_\mu(I_M \otimes I_N \otimes J_T) + \sigma^2_\epsilon(I_M \otimes I_N \otimes I_T).$$

Then replacing $J.$ by its idempotent counterpart $\bar{J}.$ and using the fact that $I. = \bar{J}. + \bar{E}.$, we obtain

$$\begin{aligned}
\Omega_0 &= NT\sigma^2_\phi(I_M \otimes \bar{J}_N \otimes \bar{J}_T) + N\sigma^2_\eta(I_M \otimes \bar{J}_N \otimes \bar{J}_T) + N\sigma^2_\eta(I_M \otimes \bar{J}_N \otimes \bar{E}_T) \\
&\quad + T\sigma^2_\mu(I_M \otimes \bar{J}_N \otimes \bar{J}_T) + T\sigma^2_\mu(I_M \otimes \bar{E}_N \otimes \bar{J}_T) + \sigma^2_\epsilon(I_M \otimes \bar{J}_N \otimes \bar{J}_T) \\
&\quad + \sigma^2_\epsilon(I_M \otimes \bar{J}_N \otimes \bar{E}_T) + \sigma^2_\epsilon(I_M \otimes \bar{E}_N \otimes \bar{J}_T) + \sigma^2_\epsilon(I_M \otimes \bar{E}_N \otimes \bar{E}_T) \\
&= \sigma^2_1(I_M \otimes \bar{E}_N \otimes \bar{E}_T) + \sigma^2_2(I_M \otimes \bar{E}_N \otimes \bar{J}_T) + \sigma^2_3(I_M \otimes \bar{J}_N \otimes \bar{E}_T) \\
&\quad + \sigma^2_4(I_M \otimes \bar{J}_N \otimes \bar{J}_T) \\
&= \sigma^2_1 Q_1 + \sigma^2_2 Q_2 + \sigma^2_3 Q_3 + \sigma^2_4 Q_4,
\end{aligned}$$

where $\sigma^2_1 = \sigma^2_\epsilon$, $\sigma^2_2 = T\sigma^2_\mu + \sigma^2_\epsilon$, $\sigma^2_3 = N\sigma^2_\eta + \sigma^2_\epsilon$, $\sigma^2_4 = NT\sigma^2_\phi + N\sigma^2_\eta + T\sigma^2_\mu + \sigma^2_\epsilon$, $Q_1 = (I_M \otimes \bar{E}_N \otimes \bar{E}_T)$, $Q_2 = (I_M \otimes \bar{E}_N \otimes \bar{J}_T)$, $Q_3 = (I_M \otimes \bar{J}_N \otimes \bar{E}_T)$ and $Q_4 = (I_M \otimes \bar{J}_N \otimes \bar{J}_T)$.

Thus, asymptotically unbiased and consistent estimates can be obtained as

$$\hat{\sigma}^2_\epsilon = \frac{u'Q_1u}{M(N-1)(T-1)},$$

$$\hat{\sigma}^2_2 = \frac{u'Q_2u}{M(N-1)},$$

$$\hat{\sigma}_3^2 = \frac{u'Q_3u}{M(T-1)},$$
$$\hat{\sigma}_4^2 = \frac{u'Q_4u}{M},$$

and

$$\hat{\sigma}_\mu^2 = \frac{\hat{\sigma}_2^2 - \hat{\sigma}_\epsilon^2}{T},$$
$$\hat{\sigma}_\eta^2 = \frac{\hat{\sigma}_3^2 - \hat{\sigma}_\epsilon^2}{N},$$
$$\hat{\sigma}_\phi^2 = \frac{\hat{\sigma}_4^2 - N\hat{\sigma}_\eta^2 - T\sigma_\mu^2 - \sigma_\epsilon^2}{NT}.$$

However, since u is not observed, using $\hat{u} = Q_X u$, the OLS residuals where $Q_X = I_{MNT} - X(X'X)^{-1}X'$ is the residual matrix projection, produces an asymptotic bias. We follow Baltagi, Song and Jung (2001) adaptation for the Wallace and Hussain (1969) estimator to our particular case. Note that if a is a n-dimensional normal random vector and $a \sim N(0, \Sigma)$, then if A is a $n \times n$ constant symmetric matrix, $E(a'Aa) = tr(A\Sigma)$. Now, $\hat{u} \sim N(0, Q_X\Omega_0 Q_X)$ and then

$$E(\hat{u}'Q.\hat{u}) = tr(Q.Q_X\Omega_0 Q_X) = \sum_{h=1}^{4} \sigma_h^2 tr(Q.Q_X Q_h Q_X).$$

This generates a 4×4 system of equations from which estimates of σ_h^2, $h = 1, 2, 3, 4$ can be obtained and the variance component estimates follow.

Appendix 2: Estimates of $(\sigma_\phi^2, \sigma_\eta^2, \sigma_\mu^2, \sigma_\epsilon^2, \rho)$, assuming $\lambda = 0$

We follow Baltagi and Wu (1999) strategy. First, we construct the within residuals from a least-squares dummy variables fixed effects model. Consider the regression model

$$y_{ijt} = x'_{ijt}\beta + \sum_{i=1}^{M}\sum_{t=1}^{T}\eta_{it}d_{it} + \sum_{i=1}^{M}\sum_{j=1}^{N}\mu_{ij}d_{ij} + u_{ijt},$$

where $\{d_{ij}\}$ is a set of dummies for the NM clusters, $\{d_{it}\}$ is another set for the MT interactions of time and M-group, and let $\{\tilde{u}_{ijt}\}$ be the residuals. Second we estimate ρ using the estimator

$$\tilde{\rho} = \frac{MNT}{MN(T-1)} \frac{\sum_{i=1}^{M}\sum_{j=1}^{N}\sum_{t=2}^{T}\tilde{u}_{ijt}\tilde{u}_{ijt-1}}{\sum_{i=1}^{M}\sum_{j=1}^{N}\sum_{t=1}^{T}\tilde{u}_{ijt}^2}.$$

Third, we transform the data to eliminate the AR(1) structure. In particular, this is done for all variables $\{y_{ijt}, x_{ijt}\}_{i=1,j=1,t=1}^{M,N,T}$ with the transformation obtained from the pre-multiplication of the matrix

$$C_\rho = \begin{bmatrix} (1-\rho^2)^{1/2} & 0 & 0 & \cdots & 0 & 0 & 0 \\ -\rho & 1 & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -\rho & 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & -\rho & 1 \end{bmatrix}.$$

This is equivalent to the transformation

$$\tilde{a}_{ijt} = \begin{cases} (1-\tilde{\rho}^2)^{1/2} a_{ijt} & \text{if } t=1 \\ (1-\tilde{\rho}^2)^{1/2}\left[\left(\frac{1}{1-\tilde{\rho}^2}\right)^{1/2} a_{ijt} - \left(\frac{\tilde{\rho}^2}{1-\tilde{\rho}^2}\right)^{1/2} a_{ijt-1}\right] & \text{if } t>1 \end{cases}.$$

Then consider the spectral decomposition and solution given in Appendix A1 for the residuals $\{\tilde{u}\}$ transformed by C_ρ. In this case, following Baltagi, Song and Jung (2002b, p256-257), we use $\bar{J}_T^\rho = \iota_{\rho T}\iota_{\rho T}'/d^2$, $\bar{E}_T^\rho = I_T - \bar{J}_T^\rho$, where $\iota_{\rho T} = (\alpha_\rho, 1, 1, ..., 1)'$, $\alpha_\rho = \sqrt{\frac{1+\rho}{1-\rho}}$, $d^2 = \iota_{\rho T}'\iota_{\rho T} = \alpha_\rho^2 + T - 1$, instead of $\bar{J}_T$ and $\bar{E}_T$ and T needs to be replaced by $T_\rho = (1-\rho^2)(\alpha_\rho^2 + T - 1)$. In particular, note that $tr(C_\rho C_\rho') = T_\rho/Ttr(I_T) = T_\rho/Ttr(\bar{J}_T^\rho + \bar{E}_T^\rho)$, a factor that applies to the σ_η^2 term. These are used to construct Q_h, $h = 1,2,3,4$.

Then,

$$
\begin{aligned}
tr(C_\rho \Omega_0 C_\rho') &= NT_\rho \sigma_\phi^2 tr(I_M \otimes \bar{J}_N \otimes \bar{J}_T^\rho) \\
&\quad + NT_\rho/T\sigma_\eta^2 tr(I_M \otimes \bar{J}_N \otimes \bar{J}_T^\rho) + NT_\rho/T\sigma_\eta^2 tr(I_M \otimes \bar{J}_N \otimes \bar{E}_T^\rho) \\
&\quad + T_\rho \sigma_\mu^2 tr(I_M \otimes \bar{J}_N \otimes \bar{J}_T^\rho) + T_\rho \sigma_\mu^2 tr(I_M \otimes \bar{E}_N \otimes \bar{J}_T^\rho) \\
&\quad + \sigma_\epsilon^2 tr(I_M \otimes \bar{J}_N \otimes \bar{J}_T^\rho) + \sigma_\epsilon^2 (I_M \otimes \bar{J}_N \otimes \bar{E}_T^\rho) \\
&\quad + \sigma_\epsilon^2 tr(I_M \otimes \bar{E}_N \otimes \bar{J}_T \rho) + \sigma_\epsilon^2 tr(I_M \otimes \bar{E}_N \otimes \bar{E}_T^\rho) \\
&= \sigma_1^2 tr(I_M \otimes \bar{E}_N \otimes \bar{E}_T^\rho) + \sigma_2^2 tr(I_M \otimes \bar{E}_N \otimes \bar{J}_T^\rho) + \sigma_3^2 tr(I_M \otimes \bar{J}_N \otimes \bar{E}_T^\rho) \\
&\quad + \sigma_4^2 tr(I_M \otimes \bar{J}_N \otimes \bar{J}_T^\rho) \\
&= \sigma_1^2 tr(Q_1) + \sigma_2^2 tr(Q_2) + \sigma_3^2 tr(Q_3) + \sigma_4^2 tr(Q_4),
\end{aligned}
$$

where $\sigma_1^2 = \sigma_\epsilon^2$, $\sigma_2^2 = T_\rho \sigma_\mu^2 + \sigma_\epsilon^2$, $\sigma_3^2 = NT_\rho/T\sigma_\eta^2 + \sigma_\epsilon^2$, $\sigma_4^2 = NT_\rho \sigma_\phi^2 + NT_\rho/T\sigma_\eta^2 + T_\rho \sigma_\mu^2 + \sigma_\epsilon^2$, $Q_1 = (I_M \otimes \bar{E}_N \otimes \bar{E}_T^\rho)$, $Q_2 = (I_M \otimes \bar{E}_N \otimes \bar{J}_T^\rho)$, $Q_3 = (I_M \otimes \bar{J}_N \otimes \bar{E}_T^\rho)$ and $Q_4 = (I_M \otimes \bar{J}_N \otimes \bar{J}_T^\rho)$.

Thus, asymptotically unbiased and consistent estimates can be obtained as

$$
\tilde{\sigma}_\epsilon^2 = \frac{u'Q_1 u}{M(N-1)(T_\rho - 1)},
$$

$$
\tilde{\sigma}_2^2 = \frac{u'Q_2 u}{M(N-1)},
$$

$$
\tilde{\sigma}_3^2 = \frac{u'Q_3 u}{M(T_\rho - 1)},
$$

$$
\tilde{\sigma}_4^2 = \frac{u'Q_4 u}{M},
$$

and

$$
\tilde{\sigma}_\mu^2 = \frac{\tilde{\sigma}_2^2 - \tilde{\sigma}_\epsilon^2}{T_\rho},
$$

$$
\tilde{\sigma}_\eta^2 = \frac{\tilde{\sigma}_3^2 - \tilde{\sigma}_\epsilon^2}{NT_\rho/T},
$$

$$
\tilde{\sigma}_\phi^2 = \frac{\tilde{\sigma}_4^2 - NT_\rho/T\tilde{\sigma}_\eta^2 - T_\rho \sigma_\mu^2 - \sigma_\epsilon^2}{NT_\rho}.
$$

However, since u is not observed, we use the same procedure of Baltagi, Song and Jung (2001) adaptation for the Wallace and Hussain (1969) as in Appendix A1.

Appendix 3: Estimates of $(\sigma_\phi^2, \sigma_\eta^2, \sigma_\mu^2, \sigma_\epsilon^2, \lambda)$, assuming $\rho = 0$

We follow Baltagi and Wu (1999) strategy adapted to this case. First, we consider the regression model

$$y_{ijt} = x'_{ijt}\beta + \sum_{i=1}^{M}\sum_{j=1}^{N}\mu_{ij}d_{ij} + u_{ijt},$$

where $\{d_{ij}\}$ is a set of dummies for the NM clusters and let $\{\breve{u}_{ijt}\}$ be the corresponding residual estimates. Second we estimate λ using the estimator

$$\breve{\lambda} = \frac{MT}{M(T-1)} \frac{\sum_{i=1}^{M}\sum_{t=2}^{T}\left(\frac{1}{N}\sum_{j=1}^{N}\breve{u}_{ijt}\right)\left(\frac{1}{N}\sum_{j=1}^{N}\breve{u}_{ijt-1}\right)}{\sum_{i=1}^{M}\sum_{t=1}^{T}\left(\frac{1}{N}\sum_{j=1}^{N}\breve{u}_{ijt}\right)^2}.$$

Third, we transform the data to eliminate the AR(1) structure. In particular, this is done for all variables $\{y_{ijt}, x_{ijt}\}_{i=1,j=1,t=1}^{M,N,T}$ with the transformation

$$C_\lambda = \begin{bmatrix} (1-\lambda^2)^{1/2} & 0 & 0 & \cdots & 0 & 0 & 0 \\ -\lambda & 1 & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -\lambda & 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & -\lambda & 1 \end{bmatrix}.$$

This is equivalent to the transformation

$$\breve{a}_{ijt} = \begin{cases} (1-\breve{\lambda}^2)^{1/2} a_{ijt} & \text{if } t = 1 \\ (1-\breve{\lambda}^2)^{1/2}\left[\left(\frac{1}{1-\breve{\lambda}^2}\right)^{1/2} a_{ijt} - \left(\frac{\breve{\lambda}^2}{1-\breve{\lambda}^2}\right)^{1/2} a_{ijt-1}\right] & \text{if } t > 1 \end{cases}.$$

Then follow Appendix A2, where λ replaces ρ and all the corresponding matrices and factores are defined accordingly, i.e. $\bar{J}_T^\lambda$, $\bar{E}_T^\lambda$ and T_λ. Note that in this case multiplying by C_λ produces the following:

$$\begin{aligned}
tr(C_\lambda \Omega_0 C_\lambda') &= NT_\lambda \sigma_\phi^2 tr(I_M \otimes \bar{J}_N \otimes \bar{J}_T^\lambda) \\
&+ N\sigma_\eta^2 tr(I_M \otimes \bar{J}_N \otimes \bar{J}_T^\lambda) + N\sigma_\eta^2 tr(I_M \otimes \bar{J}_N \otimes \bar{E}_T^\lambda) \\
&+ T_\lambda \sigma_\mu^2 tr(I_M \otimes \bar{J}_N \otimes \bar{J}_T^\lambda) + T_\lambda \sigma_\mu^2 tr(I_M \otimes \bar{E}_N \otimes \bar{J}_T^\lambda) \\
&+ T_\lambda/T\sigma_\epsilon^2 tr(I_M \otimes \bar{J}_N \otimes \bar{J}_T^\lambda) + T_\lambda/T\sigma_\epsilon^2 (I_M \otimes \bar{J}_N \otimes \bar{E}_T^\lambda) \\
&+ T_\lambda/T\sigma_\epsilon^2 tr(I_M \otimes \bar{E}_N \otimes \bar{J}_T \lambda) + T_\lambda/T\sigma_\epsilon^2 tr(I_M \otimes \bar{E}_N \otimes \bar{E}_T^\lambda) \\
&= \sigma_1^2 tr(I_M \otimes \bar{E}_N \otimes \bar{E}_T^\lambda) + \sigma_2^2 tr(I_M \otimes \bar{E}_N \otimes \bar{J}_T^\lambda) + \sigma_3^2 tr(I_M \otimes \bar{J}_N \otimes \bar{E}_T^\lambda) \\
&+ \sigma_4^2 tr(I_M \otimes \bar{J}_N \otimes \bar{J}_T^\lambda) \\
&= \sigma_1^2 tr(Q_1) + \sigma_2^2 tr(Q_2) + \sigma_3^2 tr(Q_3) + \sigma_4^2 tr(Q_4),
\end{aligned}$$

where $\sigma_1^2 = T_\lambda/T\sigma_\epsilon^2$, $\sigma_2^2 = T_\lambda\sigma_\mu^2 + T_\lambda/T\sigma_\epsilon^2$, $\sigma_3^2 = N\sigma_\eta^2 + T_\lambda/T\sigma_\epsilon^2$, $\sigma_4^2 = NT_\lambda\sigma_\phi^2 + N\sigma_\eta^2 + T_\lambda\sigma_\mu^2 + T_\lambda/T\sigma_\epsilon^2$, $Q_1 = (I_M \otimes \bar{E}_N \otimes \bar{E}_T^\lambda)$, $Q_2 = (I_M \otimes \bar{E}_N \otimes \bar{J}_T^\lambda)$, $Q_3 = (I_M \otimes \bar{J}_N \otimes \bar{E}_T^\lambda)$ and $Q_4 = (I_M \otimes \bar{J}_N \otimes \bar{J}_T^\lambda)$.

Thus, asymptotically unbiased and consistent estimates can be obtained as

$$\check{\sigma}_\epsilon^2 = T/T_\lambda \frac{u'Q_1 u}{M(N-1)(T_\lambda - 1)},$$

$$\check{\sigma}_2^2 = \frac{u'Q_2 u}{M(N-1)},$$

$$\check{\sigma}_3^2 = \frac{u'Q_3 u}{M(T_\lambda - 1)},$$

$$\check{\sigma}_4^2 = \frac{u'Q_4 u}{M},$$

and

$$\check{\sigma}_\mu^2 = \frac{\check{\sigma}_2^2 - T_\lambda/T\check{\sigma}_\epsilon^2}{T_\lambda},$$

$$\check{\sigma}_\eta^2 = \frac{\check{\sigma}_3^2 - T_\lambda/T\check{\sigma}_\epsilon^2}{N},$$
$$\check{\sigma}_\phi^2 = \frac{\check{\sigma}_4^2 - N\check{\sigma}_\eta^2 - T_\lambda\sigma_\mu^2 - T_\lambda/T\sigma_\epsilon^2}{NT_\lambda}.$$

However, since u is not observed, we use the same procedure of Baltagi, Song and Jung (2001) adaptation for the Wallace and Hussain (1969) as in Appendix A1.

References

Angrist, J. and Pischke, J-S (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press, Princeton.

Amemiya, T. (1971). "The estimation of the variances in a variance-components model," *International Economic Review* 12(1), 1-13.

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

Baltagi, B. H. (2003) *Econometric Analysis of Panel Data* (third ed.) John Wiley & Sons.

Baltagi, B. H., and P. X. Wu. (1999). "Unequally spaced panel data regressions with AR(1) disturbances." *Econometric Theory* 15, 814-823.

Baltagi, B. H., and Li. Q. (1991). "A transformation that will circumvent the problem of autocorrelation in an error-component model ." *Journal of Econometrics* 48, 385–393.

Baltagi, B. H., and Li., Q. (1995). "Testing AR(l) against MA(l) disturbances in an error component model." *Journal of Econometrics* 68, 133-151.

Baltagi, B. H., Song, S.H., and Jung, B.C. (2001). "The unbalanced nested error component regression model." *Journal of Econometrics* 101, 357-381.

Baltagi, B. H., Song, S.H., and Jung, B.C. (2002a). "LM Tests for the

Unbalanced Nested Panel Data Regression Model with Serially Correlated Errors." *Annales d'conomie et de Statistique* 65, 219–268.

Baltagi, B. H., Song, S.H., and Jung, B.C. (2002b). "Simple LM tests for the unbalanced nested error component regression model." *Econometric Reviews* 21(2), 167-187.

Bera, A., Sosa-Escudero, W., and Yoon, M. (2001). "Tests for the error component model in the presence of local misspecification." *Journal of Econometrics* 101, 1–23.

Bertrand, M., Duflo, E. and Mullainathan, S. (2004). "How much should we trust differences-in-differences estimates?" *Quarterly Journal of Economics* 119(1), 249-275.

Cameron, C., and Miller, D.L. (2015). "A practicioner's guide to cluster-robust inference," forthcoming *Journal of Human Resources.*

Hanushek, E.A., and Woessmann, L. (2011). *The Economics of International Differences in Educational Achievement* in In E.A. Hanushek, S. Machin and L. Woessmann (Eds.), *Handbook of the Economics of Education*, Vol. 3, Amsterdam: North Holland, pp 89-200.

Harville, D.A. (1977). "Maximum likelihood approaches to variance component estimation and to related problems," *Journal of the American Statistical Association* 72, 320–340.

Honda, Y. (1985). "Testing the Error Components Model with Non-Normal Disturbances" *Review of Economic Studies* 52(4), 681–690.

Matyas, L. (1996). "Error Components Models", chapter 4 in Matyas, L. and Sevestre, P. (eds), *The Econometrics of Panel Data, 2nd edition*, Kluver Academic Publishers, Dordrecht.

Montes-Rojas, G., and Sosa-Escudero, W. (2011). "Robust tests for heteroskedasticity in the one-way error components model," *Journal of Econo-*

metrics 160(2), 300–310.

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

Moulton, B.R. (1987). “Diagnostics for group effects in regression analysis,” *Journal of Business & Economic Sataistics* 5(2), 275-282.

Moulton, B.R. (1990). “An illustration of a pitfall in sstimating the effects of aggregate variables in micro units,” *Review of Economics and Statistics* 72, 334-338.

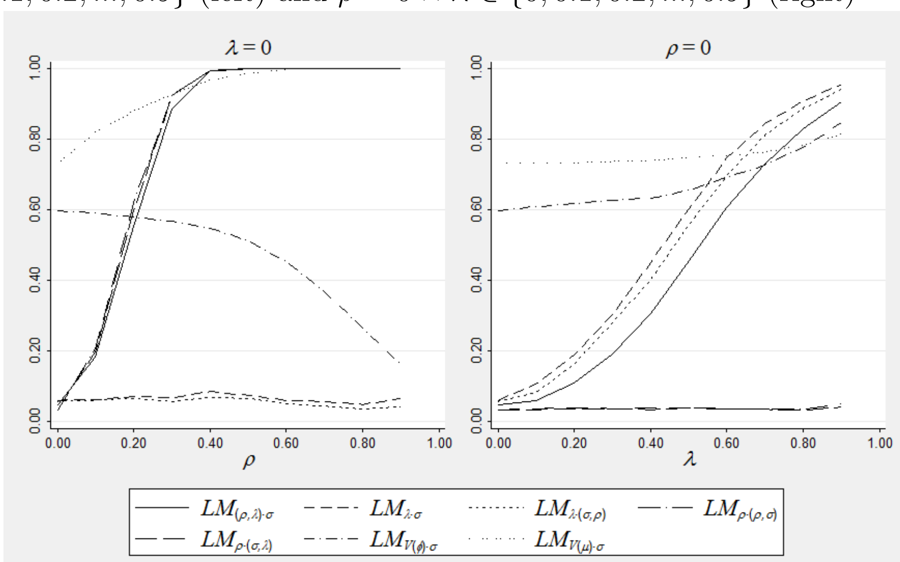
Kézdi, G. (2004). “Robust standard error estimation in fixed-effects panel models,” *Hungarian Statistical Review* Special Number 9, 95-116.

King, G., and Roberts, M.E. (2015). How robust standard errors expose methodological problems they do not fix, and what to do about it,” *Political Analysis* 23(2), 159179.

Wallace, T.D., and Hussain, A. (1969). “The Use of Error Components Models in Combining Cross Section with Time Series Data, ” *Econometrica* 37(1), 55-72.

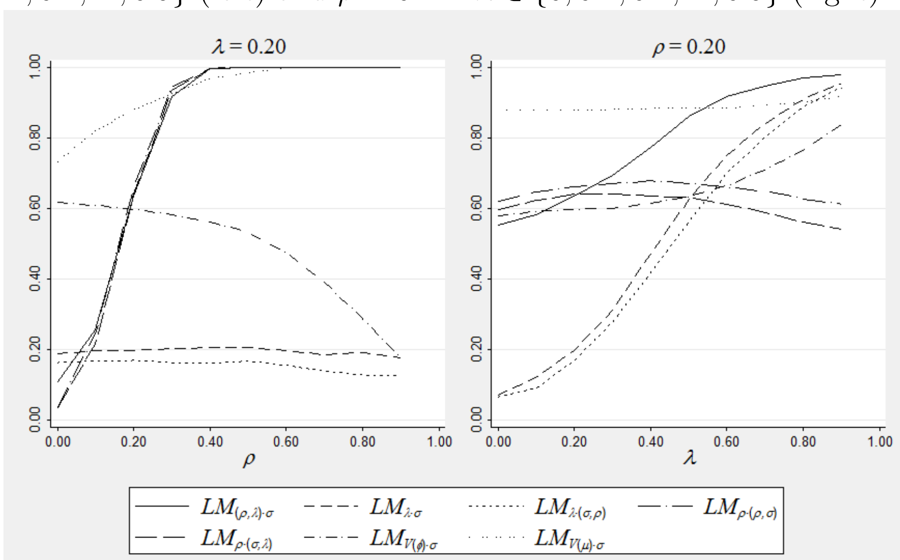
Wooldridge, J.M. (2007). *Econometric Analysis of Cross Section and Panel Data* (second ed.) MIT Press.

Figure 1: Empirical size and power, normally distributed errors, $\lambda = 0 \times \rho \in \{0, 0.1, 0.2, ..., 0.9\}$ (left) and $\rho = 0 \times \lambda \in \{0, 0.1, 0.2, ..., 0.9\}$ (right)



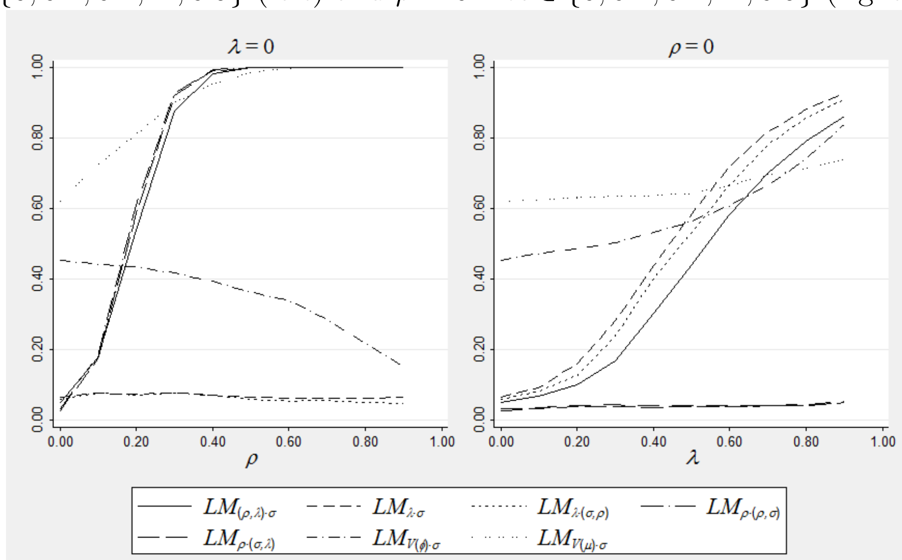
Notes: Monte Carlo experiments based on 1000 replications and a 5% nominal size. Panel size $M = 5, N = 5, T = 10$.

Figure 2: Empirical size and power, normally distributed errors, $\lambda = 0.2 \times \rho \in \{0, 0.1, 0.2, ..., 0.9\}$ (left) and $\rho = 0.2 \times \lambda \in \{0, 0.1, 0.2, ..., 0.9\}$ (right)



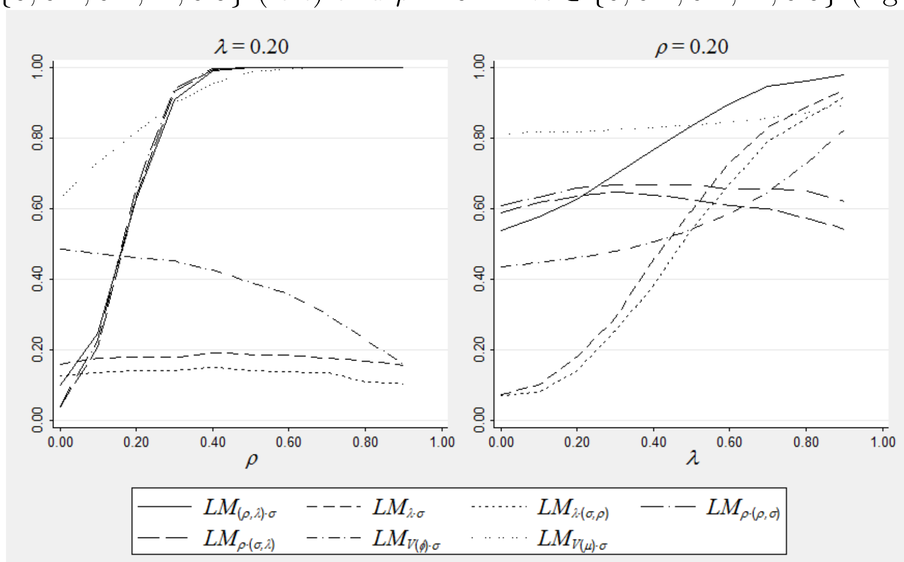
Notes: Monte Carlo experiments based on 1000 replications and a 5% nominal size. Panel size $M = 5, N = 5, T = 10$.

Figure 3: Empirical size and power, chi-squared distributed errors, $\lambda = 0 \times \rho \in \{0, 0.1, 0.2, ..., 0.9\}$ (left) and $\rho = 0 \times \lambda \in \{0, 0.1, 0.2, ..., 0.9\}$ (right)



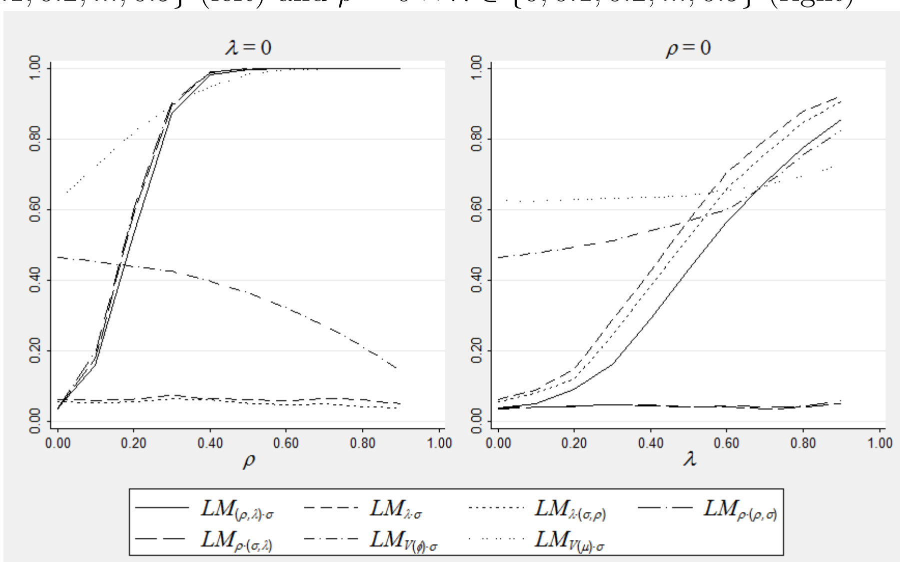
Notes: Monte Carlo experiments based on 1000 replications and a 5% nominal size. Panel size $M = 5, N = 5, T = 10$.

Figure 4: Empirical size and power, chi-squared distributed errors, $\lambda = 0.2 \times \rho \in \{0, 0.1, 0.2, ..., 0.9\}$ (left) and $\rho = 0.2 \times \lambda \in \{0, 0.1, 0.2, ..., 0.9\}$ (right)



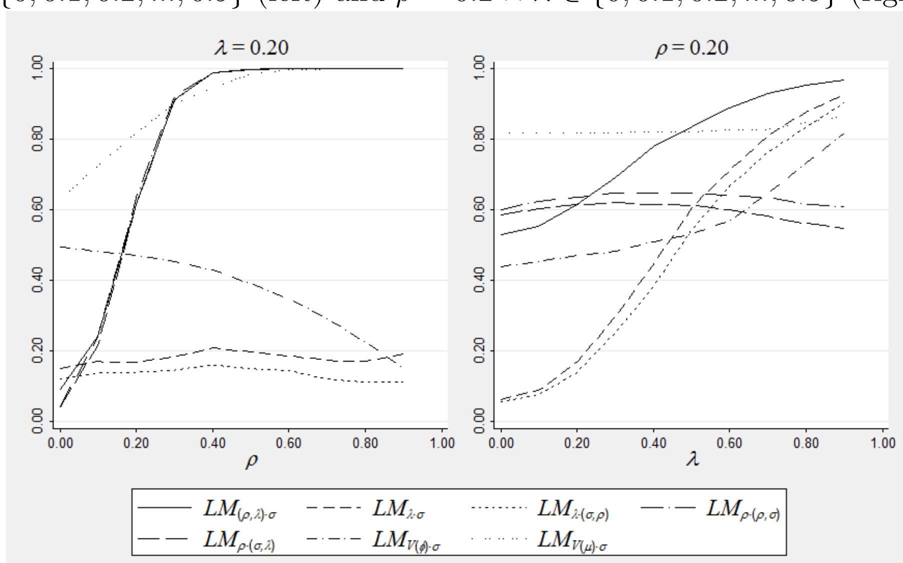
Notes: Monte Carlo experiments based on 1000 replications and a 5% nominal size. Panel size $M = 5, N = 5, T = 10$.

Figure 5: Empirical size and power, Student-t distributed errors, $\lambda = 0 \times \rho \in \{0, 0.1, 0.2, ..., 0.9\}$ (left) and $\rho = 0 \times \lambda \in \{0, 0.1, 0.2, ..., 0.9\}$ (right)



Notes: Monte Carlo experiments based on 1000 replications and a 5% nominal size. Panel size $M = 5, N = 5, T = 10$.

Figure 6: Empirical size and power, Student-t distributed errors, $\lambda = 0.2 \times \rho \in \{0, 0.1, 0.2, ..., 0.9\}$ (left) and $\rho = 0.2 \times \lambda \in \{0, 0.1, 0.2, ..., 0.9\}$ (right)



Notes: Monte Carlo experiments based on 1000 replications and a 5% nominal size. Panel size $M = 5, N = 5, T = 10$.

Table 1: Empirical size and power for normally distributed errors

		$\rho = 0$, $\lambda = 0$				$\rho = 0.2$, $\lambda = 0$				$\rho = 0$, $\lambda = 0.2$			
		M = 5		**M = 10**		**M = 5**		**M = 10**		**M = 5**		**M = 10**	
		N = 5	**N = 10**	**N = 5**	**N = 10**	**N = 5**	**N = 10**	**N = 5**	**N = 10**	**N = 5**	**N = 10**	**N = 5**	**N = 10**
$LM_{(\rho,\lambda)\cdot\sigma}$	**T**												
	3	2.8%	3.9%	4.0%	3.5%	4.3%	8.5%	9.8%	16.2%	2.8%	3.7%	4.7%	3.6%
	5	4.2%	3.6%	4.0%	5.9%	17.6%	35.5%	40.9%	68.8%	4.9%	6.4%	5.1%	9.2%
	10	4.6%	4.6%	5.0%	3.5%	55.3%	85.4%	89.7%	99.7%	10.9%	20.2%	12.1%	21.4%
$LM_{\lambda\cdot\sigma}$	**T**												
	3	5.6%	6.8%	4.0%	7.9%	4.9%	9.2%	6.4%	9.1%	5.1%	8.4%	5.0%	9.2%
	5	4.9%	8.3%	6.2%	9.9%	7.9%	12.5%	6.8%	13.4%	9.9%	15.0%	10.7%	16.7%
	10	6.0%	7.1%	8.4%	7.2%	7.1%	8.0%	7.8%	7.7%	18.8%	30.3%	18.8%	32.0%
$LM_{\lambda\cdot(\sigma,\rho)}$	**T**												
	3	5.8%	7.3%	3.6%	7.5%	5.7%	9.4%	6.6%	9.0%	5.4%	8.6%	4.8%	8.9%
	5	4.8%	8.9%	6.4%	9.9%	7.3%	11.7%	7.2%	12.9%	8.6%	13.0%	10.3%	16.0%
	10	5.7%	6.8%	8.2%	6.7%	6.6%	8.5%	7.1%	7.1%	16.3%	25.8%	17.1%	29.8%
$LM_{\rho\cdot\sigma}$	**T**												
	3	0.9%	0.7%	0.3%	0.2%	1.4%	0.8%	1.4%	1.4%	0.9%	0.9%	0.7%	0.3%
	5	3.1%	1.0%	2.1%	2.6%	15.1%	30.9%	29.8%	57.4%	3.3%	1.5%	2.5%	3.9%
	10	3.2%	4.0%	3.2%	3.2%	62.0%	89.7%	92.4%	99.8%	3.9%	5.6%	4.5%	4.3%
$LM_{\rho\cdot(\sigma,\lambda)}$	**T**												
	3	0.4%	0.1%	0.2%	0.1%	0.6%	0.3%	1.2%	1.1%	0.3%	0.1%	0.3%	0.2%
	5	2.9%	0.7%	2.1%	2.8%	13.4%	29.1%	28.8%	56.2%	3.0%	1.2%	2.5%	3.9%
	10	3.3%	3.6%	3.0%	3.2%	59.6%	88.9%	92.3%	99.8%	3.6%	5.2%	4.4%	4.3%

Notes: Monte Carlo experiments based on 1000 replications and a 5% nominal size.

Table 2: Empirical size and power for chi-squared distributed errors

		$\rho = 0$, $\lambda = 0$				$\rho = 0.2$, $\lambda = 0$				$\rho = 0$, $\lambda = 0.2$			
		M = 5		**M = 10**		**M = 5**		**M = 10**		**M = 5**		**M = 10**	
		N = 5	**N = 10**	**N = 5**	**N = 10**	**N = 5**	**N = 10**	**N = 5**	**N = 10**	**N = 5**	**N = 10**	**N = 5**	**N = 10**
$LM_{(\rho,\lambda)\cdot\sigma}$	**T**												
	3	5.3%	5.2%	5.5%	6.2%	6.9%	10.1%	12.5%	22.5%	5.5%	5.2%	6.6%	6.2%
	5	5.6%	5.8%	5.1%	4.5%	18.1%	36.5%	34.9%	69.2%	7.2%	8.0%	5.9%	8.5%
	10	4.9%	5.4%	4.4%	4.4%	53.9%	86.4%	89.1%	99.7%	9.9%	20.4%	12.0%	21.0%
$LM_{\lambda\cdot\sigma}$	**T**												
	3	4.1%	9.0%	5.1%	7.3%	5.1%	9.8%	5.1%	8.2%	4.2%	9.5%	4.6%	7.8%
	5	6.8%	9.1%	4.9%	6.4%	7.3%	9.9%	7.0%	11.2%	9.7%	14.9%	8.4%	14.0%
	10	6.6%	5.7%	4.6%	5.1%	7.2%	6.6%	5.9%	7.1%	15.9%	29.8%	16.8%	30.8%
$LM_{\lambda\cdot(\sigma,\rho)}$	**T**												
	3	4.9%	9.8%	5.1%	7.6%	6.0%	11.4%	7.4%	9.3%	5.7%	10.3%	4.9%	8.3%
	5	7.2%	8.2%	4.9%	6.5%	5.8%	8.8%	6.5%	10.1%	9.4%	12.6%	8.0%	12.5%
	10	5.8%	5.4%	4.1%	4.3%	7.0%	6.2%	6.0%	5.9%	12.6%	24.7%	14.4%	27.5%
$LM_{\rho\cdot\sigma}$	**T**												
	3	0.2%	0.5%	0.7%	0.2%	0.7%	0.8%	2.1%	1.5%	0.6%	1.0%	0.8%	0.5%
	5	2.9%	2.5%	2.1%	1.2%	15.8%	30.9%	27.8%	58.1%	3.6%	3.0%	2.8%	1.8%
	10	3.1%	4.8%	2.9%	2.8%	60.9%	90.8%	91.5%	99.7%	4.2%	5.6%	3.7%	3.8%
$LM_{\rho\cdot(\sigma,\lambda)}$	**T**												
	3	0.1%	0.0%	0.5%	0.1%	0.3%	0.7%	1.2%	1.3%	0.2%	0.3%	0.6%	0.1%
	5	2.5%	1.6%	2.2%	1.3%	13.6%	28.4%	26.8%	56.6%	3.1%	2.4%	2.6%	1.7%
	10	2.7%	4.4%	2.8%	2.7%	58.8%	89.6%	91.2%	99.7%	3.9%	5.1%	3.8%	3.8%

Notes: Monte Carlo experiments based on 1000 replications and a 5% nominal size.

Table 3: Empirical size and power for $Student - t_4$ distributed errors

		$\rho = 0$, $\lambda = 0$				$\rho = 0.2$, $\lambda = 0$				$\rho = 0$, $\lambda = 0.2$			
		M = 5		**M = 10**		**M = 5**		**M = 10**		**M = 5**		**M = 10**	
		N = 5	**N = 10**	**N = 5**	**N = 10**	**N = 5**	**N = 10**	**N = 5**	**N = 10**	**N = 5**	**N = 10**	**N = 5**	**N = 10**
$LM_{(\rho,\lambda)\cdot\sigma}$	**T**												
	3	3.2%	2.7%	4.4%	3.4%	5.0%	9.7%	12.1%	24.2%	4.0%	2.8%	4.8%	4.0%
	5	5.0%	5.2%	3.4%	4.9%	18.2%	35.3%	36.0%	67.7%	6.5%	8.6%	4.9%	8.8%
	10	3.8%	4.4%	4.7%	4.5%	52.8%	85.9%	90.6%	99.8%	9.2%	15.7%	13.6%	21.2%
$LM_{\lambda\cdot\sigma}$	**T**												
	3	4.2%	8.8%	5.9%	8.9%	5.3%	10.4%	5.4%	9.5%	4.3%	9.1%	6.5%	9.6%
	5	5.7%	8.3%	5.5%	6.0%	7.0%	10.9%	5.9%	11.0%	11.1%	15.5%	7.5%	14.7%
	10	6.3%	5.7%	6.0%	6.4%	6.3%	7.1%	6.9%	6.6%	14.9%	27.7%	18.5%	32.3%
$LM_{\lambda\cdot(\sigma,\rho)}$	**T**												
	3	4.7%	10.0%	6.8%	9.6%	7.3%	11.8%	6.3%	10.7%	5.6%	9.1%	7.2%	10.0%
	5	6.0%	9.0%	5.5%	6.0%	6.5%	10.2%	5.5%	9.9%	9.9%	12.6%	7.3%	13.7%
	10	5.5%	5.3%	5.8%	5.8%	5.5%	7.0%	6.0%	6.5%	12.0%	23.0%	17.2%	29.7%
$LM_{\rho\cdot\sigma}$	**T**												
	3	0.6%	0.9%	0.5%	0.7%	1.1%	1.1%	1.4%	2.8%	0.7%	0.2%	0.8%	0.5%
	5	2.9%	1.9%	1.4%	1.4%	16.3%	31.4%	29.2%	60.2%	4.4%	2.8%	2.4%	2.8%
	10	3.8%	2.8%	4.3%	3.0%	60.1%	91.0%	92.9%	99.7%	4.4%	3.9%	4.6%	4.2%
$LM_{\rho\cdot(\sigma,\lambda)}$	**T**												
	3	0.3%	0.0%	0.4%	0.3%	0.6%	0.5%	0.9%	2.0%	0.2%	0.0%	0.6%	0.3%
	5	2.7%	1.7%	1.3%	1.3%	14.1%	28.5%	27.4%	59.1%	3.8%	2.8%	2.3%	2.5%
	10	3.6%	3.2%	4.0%	3.2%	58.6%	90.5%	92.6%	99.6%	4.2%	4.0%	4.4%	4.4%

Notes: Monte Carlo experiments based on 1000 replications and a 5% nominal size.

Table 4: PISA nested autocorrelation analysis

	Estimation	**p-value**
$LM_{(\rho,\lambda)\cdot\sigma}$	6.90	0.0318
$LM_{\lambda\cdot\sigma}$	7.56	0.0060
$LM_{\lambda\cdot(\sigma,\rho)}$	5.02	0.0251
$LM_{\rho\cdot\sigma}$	0.64	0.4252
$LM_{\rho\cdot(\sigma,\lambda)}$	0.10	0.7491

Notes: computations with data from PISA survey.