Accepted Manuscript

Testing for serial correlation in hierarchical linear models

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

Please cite this article as: J. Alejo, G. Montes-Rojas, W. Sosa-Escudero, Testing for serial correlation in hierarchical linear models, *Journal of Multivariate Analysis* (2017), https://doi.org/10.1016/j.jmva.2017.11.007

Testing for serial correlation in hierarchical linear models

Javier Alejo[a], Gabriel Montes-Rojas[b], Walter Sosa-Escudero[c,*]

[a] Universidad Nacional de La Plata and CONICET
[b] CONICET-IIEP-BAIRES-Universidad de Buenos Aires and Universitat Autònoma de Barcelona
[c] Universidad de San Andrés and CONICET

Abstract

This paper proposes a simple hierarchical model and a testing strategy to identify intra-cluster correlations, in the form of nested random effects and serially correlated error components. We focus on intra-cluster serial correlation at different nested levels, a topic that has not been studied in the literature before. A Neyman's $C(\alpha)$ framework is used to derive LM-type tests that allow researchers to identify the appropriate level of clustering as well as the type of intra-group correlation. An extensive Monte Carlo exercise shows that the proposed tests perform well in finite samples and under non-Gaussian distributions.

Keywords: Clusters, random effects, serial correlation.
2010 MSC: 62P20, 62H15, 62J10

1. Introduction

Intra-group correlation has received considerable interest. When the data can be grouped in clusters, observations within each group are typically dependent. As highlighted by Bertrand et al. [13], failure to take such dependence into account can lead to misleading statistical inferences; this concern dates back to Moulton's seminal paper [27]. Consequently, the problem of whether 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.

Practitioners typically rely on "cluster robust methods", e.g., on estimates of standard errors that explicitly allow for correlations among observations within a group. However, the consistency of this approach depends on the number of independent groups growing large. This is problematic when grouping obeys a nested structure, as would be the case of students in a given class, school, etc. In such a scenario a safer strategy that allows for arbitrary correlations at a higher level (e.g., at the school instead of the class level) comes at the price of leaving fewer independent groups, making asymptotic approximations less reliable. In their recent survey, Cameron and Miller [15] point 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. 321].

We are thus concerned with the appropriate level of clustering in a hierarchical linear model. Proper identification of the source of intra-group correlation is important to decide how to estimate the parameters of interest and their standard deviation. For example, when only random effects cause intra-cluster correlation, feasible generalized least-squares (GLS) strategies as in [6] might offer a simple and convenient alternative over cluster robust methods in the few groups scenario. The most obvious source 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. Tests for nested random effects have been studied in [8]. Another source of intra-cluster correlation that has received particular consideration in [13] is time, i.e., cluster correlation is induced when

---

*Corresponding author. Universidad de San Andrés, Vito Dumas 284, B1644BID Victoria, Buenos Aires, Argentina.
Email addresses: javieralejo@gmail.com (Javier Alejo), gabriel.montes@fce.uba.ar (Gabriel Montes-Rojas), wsosa@udesa.edu.ar (Walter Sosa-Escudero)

observations are sorted chronologically, i.e., serial correlation. Tests for nested random effects allowing for serial correlation at the "finest" level only (students, in our example) were proposed in [7].

Our paper considers intra-group correlations as a combination of random effects and serially correlated error components in a nested, hierarchical structure. It focuses only on the issue of different levels of serial correlation in a hierarchical model, assuming the presence of nested random effects. This topic has not been analyzed in the literature. We argue that these tests are important to understand the nature of intra-cluster correlation, because controlling for random effects only tends to underestimate standard errors in the presence of serial correlation [25]. These tests complement the results in the literature, notably [7, 8]. A comprehensive testing framework for both random effects and/or serial correlation, at different nested levels, could thus be developed based on our results and those in [7, 8].

In particular, our testing strategy allows for serial correlation at *both* hierarchical levels, jointly or conditional on the presence of the other. Our tests are based on the Lagrange Multiplier (LM) principle, constructed under Gaussian error components. Our simulation experiments show that the tests work under both normality and non-normality, in line with the results in Honda [20], who shows that the classical Breusch–Pagan test is robust to alternative distributional assumptions. Consistent estimators of the parameters under the null hypothesis can be obtained using an ANOVA-type analysis (see, e.g., [4, 6]), which are easier to obtain than full maximum likelihood estimators. Hence we propose Neyman's $C(\alpha)$ tests, which are asymptotically equivalent to likelihood-based LM tests under $\sqrt{n}$-consistent non-maximum likelihood estimation of the nuisance parameters; see [10] for a discussion.

The paper is organized as follows. Section 2 discusses a simple model for grouped data and the relevant hypotheses for intra-cluster correlations. Section 3 derives tests for all possible combinations of cluster effects. The reliability of the asymptotic results in the 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 6 concludes.

2. Nested intra-group and serial correlation

Consider a hierarchical linear model with two nested cluster groups, so that, for all $i \in \{1, \ldots, M\}$, $j \in \{1, \ldots, N\}$ and $t \in \{1, \ldots, T\}$,

$$y_{ijt} = x_{ijt}^\top \beta + u_{ijt}, \quad u_{ijt} = \phi_i + \delta_{it} + \mu_{ij} + \nu_{ijt}.$$

To simplify notation and derivations we will assume a balanced panel data. Here, y_{ijt} is the outcome of interest and as in [6], each observation (i, j, t) will be referred to as corresponding to individual j in group i and period t. Furthermore, $x_{ijt}^\top$ and β are $1 \times K$ and $K \times 1$ vectors with the observable covariates and unknown parameters, respectively.

The error structure allows for unobserved heterogeneity at the i, it, ij and ijt levels in the form of unobserved random effects and autocorrelation that determine the error structure u_{ijt}. The quantities ϕ_i and μ_{ij} are nested random effects at the i and ij levels, respectively. The presence of two hierarchical levels leads to two autocorrelation patterns. Consider two nested stationary AR(1) processes, viz.

$$\delta_{it} = \lambda \delta_{it-1} + \eta_{it}, \quad \nu_{ijt} = \rho \nu_{ijt-1} + \epsilon_{ijt},$$

with $|\lambda| < 1$ and $|\rho| < 1$. A canonical example for this model may be the following. Consider M classrooms each with N students each observed during T periods, where each student belongs to only one classroom. Let y_{ijt} 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 dependence may arise due to the time dependence of shocks at the student or classroom level, modeled as AR(1) processes in our case.

The full null hypothesis of no cluster effect is the joint null of no random effect 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", i.e., 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 higher level leaves fewer groups of independent observations, harming the reliability of cluster robust standard errors. Second, 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

2

null hypothesis of no serial correlation, only random effects cause intra-cluster correlation, in which case minimum norm quadratic unbiased estimates of variances can be simply derived as in [6] or [24, p. 61], which may have a considerable advantage over cluster robust methods in the few groups scenario, especially in terms of bias.

Consequently, in this setup, testing for cluster correlations amounts to checking for random effects and serial correlation at different hierarchical levels. When there is only one hierarchical level (e.g., students in different periods) the setup is a standard panel data structure, hence the problem reduces to learning the source of intra-group correlation in the form of random effects or serial correlation. The classic test of Breusch and Pagan [14] checks for random effects in a simple error components model. Baltagi and Li [4] propose a test for first order serial correlation in the same framework. Bera et al. [12] point out that both tests reject their nulls incorrectly when the unwanted effect is present, i.e., the Breusch–Pagan test rejects under serial correlation even when no random effects are present and a similar symmetric concern affects the test by Baltagi and Li [4]. Consequently, both tests might detect intra-group correlation but are unable to identify its source. Bera et al. [12] propose a modification that can identify each effect separately. Finally, Inoue and Solon [21] propose a test for first order serial correlation after fixed effect estimation.

When more than one hierarchical level is allowed for, Baltagi et al. [8] develop LM tests for random effects in a nested error components model, but with no serial correlation. Baltagi et al. [7] 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, i.e., 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 [22], who proposes an omnibus test based on the comparison of variance estimates with or without allowing for cluster correlation, in the spirit of the classic White test for heteroscedasticity. King and Roberts [23] 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 $x_i = (x_{i11}^\top, \ldots, x_{i1T}^\top, \ldots, x_{iN1}^\top, \ldots, x_{iNT}^\top)^\top$. We will make the following assumptions.

Assumption 1:

The set $\{(y_{i\cdot}, x_i, \phi_i, \eta_{i\cdot}, \mu_{i\cdot}, \epsilon_{i\cdot}) : i \in \{1, \ldots, M\}\}$ forms an independent and identically distributed random sample.

Assumption 2:

Correct mean specification: $\forall_{i,j,t} \; \mathrm{E}(\phi_i \mid x_i) = \mathrm{E}(\eta_{it} \mid x_i) = \mathrm{E}(\mu_{ij} \mid x_i) = \mathrm{E}(\epsilon_{ijt} \mid x_i) = 0$.

Assumption 3:

Variance: $\mathrm{var}(\phi_i \mid x_i) = \sigma_\phi^2$, $\mathrm{var}(\delta_{it} \mid x_i) = \sigma_\delta^2 = \sigma_\eta^2/(1-\lambda^2)$, $\mathrm{var}(\mu_{ij} \mid x_i) = \sigma_\mu^2$, $\mathrm{var}(\nu_{ijt} \mid x_i) = \sigma_\nu^2 = \sigma_\epsilon^2/(1-\rho^2)$. $\sigma_\phi^2 > 0$, $\sigma_\delta^2 > 0$, $\sigma_\eta^2 > 0$, $\sigma_\mu^2 > 0$, $\sigma_\nu^2 > 0$, $\sigma_\epsilon^2 > 0$, $|\lambda| < 1$, $|\rho| < 1$.

Assumption 4:

Nested autocovariance structure: $\forall_{i,j,t,h} \; \mathrm{cov}(\delta_{it}, \delta_{ih} \mid x_i) = \lambda^{|t-h|}\sigma_\eta^2$, $\mathrm{cov}(\nu_{ijt}, \nu_{ijh} \mid x_i) = \rho^{|t-h|}\sigma_\epsilon^2$.

Assumption 5:

Normality of iid random samples: $\phi_i \sim \mathcal{N}(0, \sigma_\phi^2)$, $\mu_{ij} \sim \mathcal{N}(0, \sigma_\mu^2)$, $\eta_{it} \sim \mathcal{N}(0, \sigma_\eta^2)$, $\epsilon_{ijt} \sim \mathcal{N}(0, \sigma_\epsilon^2)$.

In matrix form the model can be written as $y = X\beta + u$, where y and u 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. Under Assumptions 1–5, the covariance matrix is

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

where $I_\cdot$ is a $\cdot$-dimensional identity matrix, $J_\cdot$ is a $\cdot$-dimensional matrix of 1s, V_λ is a $T \times T$ matrix with (i, j)th element equal to $\lambda^{|i-j|}$ for all $i \in \{1, \ldots, T\}$ and $j \in \{1, \ldots, T\}$, V_ρ is the same with ρ replacing λ, and $\otimes$ is the Kronecker product.

For future reference define the idempotent matrices $\bar{J}. \equiv \frac{1}{\cdot} J.$ and $\bar{E}. \equiv I. - \bar{J}.$, 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

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

with $\theta = (\sigma_\phi^2, \sigma_\eta^2, \sigma_\mu^2, \sigma_\epsilon^2, \rho, \lambda)$, $u = y - X\beta$, and Ω is given by Eq. (1). In this model we have that the Fisher information matrix is block diagonal in terms of β and θ. This feature also applies to non-Gaussian error components, where in fact ordinary least-squares (OLS) estimators for β are consistent, a consequence of Assumption 2. In turn, this simplifies the subsequent algebra where we only consider θ for constructing our LM tests.

Baltagi et al. [8] develop LM tests for random effects in a nested error components model, assuming $\sigma_\eta^2 = \lambda = \rho = 0$. They derive tests for the joint null hypothesis

$$\mathcal{H}_0^{\sigma_\phi^2, \sigma_\mu^2} : \sigma_\phi^2 = \sigma_\mu^2 = 0$$

and for the conditional hypotheses $\mathcal{H}_0^{\sigma_\mu^2} : \sigma_\mu^2 = 0$, assuming $\sigma_\phi^2 > 0$, and $\mathcal{H}_0^{\sigma_\phi^2} : \sigma_\phi^2 = 0$, assuming $\sigma_\mu^2 > 0$.

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 et al. [7]. i.e., their null hypothesis is

$$\mathcal{H}_0^{\sigma_\phi^2, \sigma_\mu^2, \rho} : \sigma_\phi^2 = \sigma_\mu^2 = \rho = 0,$$

assuming $\sigma_\eta^2 = \lambda = 0$.

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 ($\mathcal{H}_0^{\rho, \lambda} : \rho = 0, \lambda = 0$, assuming $\sigma_\phi^2 > 0, \sigma_\eta^2 > 0, \sigma_\mu^2 > 0$) and conditional tests for one type of serial correlation given that the other is present, i.e., $\mathcal{H}_0^\rho : \rho = 0$, assuming $\sigma_\phi^2 > 0, \sigma_\eta^2 > 0, \sigma_\mu^2 > 0, |\lambda| < 1$ and $\mathcal{H}_0^\lambda : \lambda = 0$, assuming $\sigma_\phi^2 > 0, \sigma_\eta^2 > 0, \sigma_\mu^2 > 0, |\rho| < 1$.

The combination of the proposed tests with those previously proposed by Baltagi et al. [7, 8] allows researchers to fully identify the levels and the sources of intra-cluster correlation, and to decide on an appropriate strategy to handle it. For example, and as mentioned in the Introduction, under the joint null of no serial correlation, a hierarchical feasible generalized least-squares strategy produces asymptotically efficient estimates of the parameters of interest and consistent estimates of their variances, which may have considerable advantages over cluster robust methods that unnecessarily allow intra-class 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 $\theta \in \Theta \subseteq \mathbb{R}^p$, where p is the dimension of θ. Using the formulas on p. 326 of Harville [19] (see also [3]) the score functions can be expressed, for all $r \in \{1, \ldots, p\}$, as

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

The information matrix $\mathcal{J}$ can be obtained, for all $r, k \in \{1, \ldots, p\}$, as

$$\frac{\partial^2}{\partial \theta_r \partial \theta_k} L = \frac{1}{2} \operatorname{tr} \left\{ \Omega^{-1} \left(\frac{\partial^2 \Omega}{\partial \theta_r \partial \theta_k} - \frac{\partial \Omega}{\partial \theta_r} \Omega^{-1} \frac{\partial \Omega}{\partial \theta_k} \right) \right\} + \frac{1}{2} u^\top \Omega^{-1} \left(\frac{\partial^2 \Omega}{\partial \theta_r \partial \theta_k} - 2 \frac{\partial \Omega}{\partial \theta_r} \Omega^{-1} \frac{\partial \Omega}{\partial \theta_r} \right) \Omega^{-1} u,$$

and

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

Note that

4

$$\partial\Omega/\partial\sigma_\phi^2 = I_M \otimes J_N \otimes J_T, \tag{2}$$

$$\partial\Omega/\partial\sigma_\eta^2 = \frac{1}{1-\lambda^2}\,(I_M \otimes J_N \otimes V_\lambda), \tag{3}$$

$$\partial\Omega/\partial\sigma_\mu^2 = I_M \otimes I_N \otimes J_T, \tag{4}$$

$$\partial\Omega/\partial\sigma_\epsilon^2 = \frac{1}{1-\rho^2}\,(I_M \otimes I_N \otimes V_\rho), \tag{5}$$

$$\partial\Omega/\partial\rho = -\frac{2\rho\sigma_\epsilon^2}{(1-\rho^2)^2}\,(I_M \otimes I_N \otimes V_\rho) + \frac{\sigma_\epsilon^2}{1-\rho^2}\,(I_M \otimes I_N \otimes W_\rho), \tag{6}$$

and

$$\partial\Omega/\partial\lambda = -\frac{2\lambda\sigma_\eta^2}{(1-\lambda^2)^2}(I_M \otimes J_N \otimes V_\lambda) + \frac{\sigma_\eta^2}{1-\lambda^2}\,(I_M \otimes J_N \otimes W_\lambda), \tag{7}$$

where $W_\rho = \partial V_\rho/\partial\rho$ is a derivative $T \times T$ matrix with (i, j)th entry equal to $|i - j|\rho^{|i-j|-1}$. The matrix $W_\lambda = \partial V_\lambda/\partial\lambda$ is the same with λ replacing ρ.

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 [4, 6] and 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 $\sqrt{n}$-consistent non-maximum likelihood estimation of the nuisance parameters.

Consider a partition of $\theta = (\theta_1^\top, \theta_2^\top)^\top$, where θ_2 contains the parameters under the corresponding null hypothesis $\mathcal{H}_0^2 : \theta_2 = 0$, and θ_1 the nuisance parameters that need to be estimated. In our particular case, θ will be partitioned into either $\theta_1 = (\sigma_\phi^2, \sigma_\eta^2, \sigma_\mu^2, \sigma_\epsilon^2), \theta_2 = (\rho, \lambda)$ (Section 3.1), $\theta_1 = (\sigma_\phi^2, \sigma_\eta^2, \sigma_\mu^2, \sigma_\epsilon^2, \rho), \theta_2 = \lambda$ (Section 3.2) or $\theta_1 = (\sigma_\phi^2, \sigma_\eta^2, \sigma_\mu^2, \sigma_\epsilon^2, \lambda), \theta_2 = \rho$ (Section 3.3). Correspondingly, the score will be partitioned as $s(\theta) = (s_1(\theta)^\top, s_2(\theta)^\top)^\top$, and the information matrix as

$$\mathcal{J}(\theta) = \begin{pmatrix} \mathcal{J}_{11}(\theta) & \mathcal{J}_{12}(\theta) \\ \mathcal{J}_{21}(\theta) & \mathcal{J}_{22}(\theta) \end{pmatrix}.$$

Conditional LM statistics for $\mathcal{H}_0^2$ under maximum likelihood estimation are defined as

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

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

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

A well known result is that $LM_{2\cdot 1}(\hat{\theta}^*) \rightsquigarrow \chi^2_{\dim(\theta_2)}$ as $n \to \infty$, where $\hat{\theta}^*$ is a $\sqrt{n}$-consistent estimator under the corresponding null hypothesis.

3.1. LM test for serial correlation under random effects: $\mathcal{H}_0^{\rho,\lambda} : \rho = 0, \lambda = 0$, assuming $\sigma_\phi^2 > 0, \sigma_\eta^2 > 0, \sigma_\mu^2 > 0$

Consider first a test for no autocorrelation at both hierarchical levels but allowing for a nested error components random effects structure. In this case the covariance matrix under the null hypothesis Ω_0 is given by Eq. (1) with $\sigma_\delta^2 = \sigma_\eta^2$ and $\sigma_\nu^2 = \sigma_\epsilon^2$. Then

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

Each partial derivative of Ω under the null hypothesis is constructed by replacing ρ and λ by 0 in Eqs. (2)–(7). Note that Eqs. (6)–(7) are simplified to

5

$$\partial\Omega/\partial\rho|_{\mathcal{H}_0} = \sigma_\epsilon^2 (I_M \otimes I_N \otimes B_T), \quad \partial\Omega/\partial\lambda|_{\mathcal{H}_0} = \sigma_\eta^2 (I_M \otimes J_N \otimes B_T),$$

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

For this case define $\theta_1 = (\sigma_\phi^2, \sigma_\eta^2, \sigma_\mu^2, \sigma_\epsilon^2)$ and $\theta_2 = (\rho, \lambda)$. Appendix A1 provides consistent estimates for all the elements of θ under the null hypothesis, $\hat{\theta} = (\hat{\sigma}_\phi^2, \hat{\sigma}_\eta^2, \hat{\sigma}_\mu^2, \hat{\sigma}_\epsilon^2, 0, 0)$. Then a test for absence of autocorrelation at any level is constructed by replacing in $LM_{2\cdot 1}$ all the unknown parameters by its consistent estimates, using the matrix derivative formulas above and replacing the unobserved u by OLS residuals $\hat{u}$. The resulting test statistic will be labeled $LM_{(\rho,\lambda)\cdot\sigma}$. Computer routines to implement all the proposed tests are available from the authors upon request.

3.2. Test for serial correlation at the group level: $\mathcal{H}_0^\lambda : \lambda = 0$, assuming $\sigma_\phi^2 > 0, \sigma_\eta^2 > 0, \sigma_\mu^2 > 0, |\rho| < 1$

This is a test for autocorrelation at the most aggregate level. In this case we compute Ω_0 replacing $\lambda = 0$ in Eq. (1) and in Eqs. (2)–(7) for the partial derivatives of Ω under the null hypothesis. The matrix B_T is used in Eq. (7).

Two tests will be proposed for this case. First a test for $\mathcal{H}_0^\lambda$ that imposes $\rho = 0$, i.e., 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. The second test is based on consistent estimates of ρ as well as other variance parameters as detailed in Appendix A2. For this case $\theta_1 = (\sigma_\phi^2, \sigma_\eta^2, \sigma_\mu^2, \sigma_\epsilon^2, \rho)$ and $\theta_2 = \lambda$. Replacing these estimates in the previous formula, we obtain the conditional test $LM_{\lambda\cdot(\sigma,\rho)}$.

3.3. LM test for autocorrelation at the individual level: $\mathcal{H}_0^\rho : \rho = 0$, assuming $\sigma_\phi^2 > 0, \sigma_\eta^2 > 0, \sigma_\mu^2 > 0, |\lambda| < 1$

This is a test for autocorrelation at the finest level. In this case, ρ is replaced by zero in Eq. (1) to compute Ω_0 and in Eqs. (2)–(7) for the partial derivatives under $\mathcal{H}_0$. The matrix B_T is used in Eq. (6).

Once again, two different tests will be derived. The first one is a test for $\mathcal{H}_0^\rho$ that imposes $\lambda = 0$, i.e., 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, $\check{\theta} = (\check{\sigma}_\phi^2, \check{\sigma}_\eta^2, \check{\sigma}_\mu^2, \check{\sigma}_\epsilon^2, 0, \check{\lambda})$. For this case let $\theta_1 = (\sigma_\phi^2, \sigma_\eta^2, \sigma_\mu^2, \sigma_\epsilon^2, \lambda)$ and $\theta_2 = \rho$. The test derived by replacing these estimates in the formula will be labeled $LM_{\rho\cdot(\sigma,\lambda)}$.

4. Monte Carlo experiments

This section explores the small-sample performance of the proposed tests through a Monte Carlo experiment. We will consider the following simple hierarchical model:

$$y_{ijt} = \beta_1 x_{1,i} + \beta_2 x_{2,it} + \beta_3 x_{3,ij} + \beta_4 x_{4,ijt} + u_{ijt},$$

where $u_{ijt} = \phi_i + \delta_{it} + \mu_{ij} + v_{ijt}$ for all $i \in \{1, \dots, M\}$, $j \in \{1, \dots, N\}$, $t \in \{1, \dots, T\}$, 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 $\mathcal{N}(0, 1)$ random variables, and μ_{ij} be $\mathcal{N}(0, 0.1)$. 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,

$$
\begin{aligned}
\delta_{it} &= \lambda\delta_{it-1} + \eta_{it}, \quad v_{ijt} = \rho v_{ijt-1} + \epsilon_{ijt}, \\
x_{2,it} &= x_{1,i} + \lambda_x x_{2,it-1} + v_{2,it}, \quad x_{2,i1} = x_{1,i} + v_{2,i1}, \\
x_{4,ijt} &= x_{1,i} + x_{2,it} + x_{3,ij} + \rho_x x_{4,ijt-1} + v_{ijt}, \quad x_{4,ij1} = x_{1,i} + x_{2,i1} + x_{3,ij} + v_{4,ij1}.
\end{aligned}
$$

with $|\lambda| < 1$ and $|\rho| < 1$.

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 each district) and $T = \{5, 10\}$ (i.e., number of repeated observations of the same school). Alternative sample sizes only reinforce the results, and are not shown to save space. We evaluate the tests using a nominal size of 0.05 and 1,000 replications. For all data generating processes we consider the performance of the LM statistics

constructed in the previous section: (i) joint test for $\mathcal{H}_0 : \lambda = \rho = 0$, $LM_{(\rho,\lambda)\cdot\sigma}$, (ii) tests for $\mathcal{H}_0 : \lambda = 0$, $LM_{\lambda\cdot\sigma}$, $LM_{\lambda\cdot(\sigma,\rho)}$, and (iii) for $\mathcal{H}_0 : \rho = 0$, $LM_{\rho\cdot\sigma}$, $LM_{\rho\cdot(\sigma,\lambda)}$.

Table 1 focuses on size performance under the joint null of no serial correlation at both hierarchical levels ($\mathcal{H}_0 : \rho = \lambda = 0$), and under serial correlation at each hierarchical level separately ($\rho = 0.2, \lambda = 0$ and $\rho = 0, \lambda = 0.2$). Rows correspond to rejections rates of alternative tests for different sample sizes of the temporal dimension (T). Columns consider alternative values for M and N, and for alternative configurations of the serial correlation parameters.

[INSERT TABLE 1 HERE]

The simulations show that all tests have approximately correct empirical size for all panel size dimensions, i.e., close to 5% in all cases when $\mathcal{H}_0 : \rho = \lambda = 0$ is true. Moreover, when one correlation parameter is increased while keeping the other constant, all tests properly aimed at detecting it correctly increase their rejection rates.

In order to explore the power performance of the tests with more detail results are shown graphically to avoid cluttering information. Figure 1 studies the performance of all tests when λ takes values in $\{0, 0.1, 0.2, \ldots, 0.9\}$. Graphs (a) and (b) present results when $\rho = 0$. Graph (a) presents results for tests checking for $\lambda = 0$ (LM_λ and $LM_{\lambda\rho}$) while graph (b) considers tests for $\rho = 0$ (LM_ρ and $LM_{\rho\cdot\lambda}$). The joint test $LM_{\rho\lambda}$ is reproduced in both graphs for easy comparison. Graphs (c) and (d) present the same information, but when $\rho = 0.2$. Finally, Figure 2 presents the same information but now altering ρ in $\{0, 0.1, 0.2, \ldots, 0.9\}$, and λ in $0, 0.2$.

[INSERT FIGURE 1 HERE]

[INSERT FIGURE 2 HERE]

The experiments suggest the following results. First, and as expected, all tests (joint, marginal and conditional) have larger power when the alternative hypothesis they are designed to test for is activated. Second, when only one pattern of serial correlation is present, the power ranking always favors the marginal test, followed by the conditional and the joint test. Third, conditional tests perform very similarly to marginal tests. Fourth, marginal and conditional tests in one direction are not affected by the direction not tested for, i.e., for example, the presence of serial correlation at the "fine" level does not affect tests for serial correlation at the "coarse" level.

A topic not explored in previous work [7, 8] is the relevance of the normality assumption. We repeated the Monte Carlo exercises assuming errors from the centered and standardized chi-square distribution with 1 degree of freedom, and for the standardized Student's t distribution with 5 degrees of freedom. Results are presented in Table 2, Figures 3–4 for the chi-square distribution and Table 3, and Figures 5–6 for Student's t distribution, which are organized as the ones corresponding to the normal case. All results are virtually unaltered, suggesting that the Gaussian assumption is not restrictive. This is in line with the analytic result in Honda [20], who shows that the classical Breusch–Pagan test is indeed robust to non-normality in spite of being derived from a Gaussian likelihood. Even though an analog analytical result for serial correlation is, to the best of our knowledge, not available, the empirical results in Evans [16] suggest that unlike tests for heteroscedasticity, standard LM based tests for serial correlation are consistent and size robust to non-normalities. In the strict time series framework Furno [17] suggests that power improvements may arise by replacing OLS by least absolute deviations residuals, but her empirical results do not find conclusive improvements when testing for serial correlation.

[INSERT TABLE 2 HERE]

[INSERT FIGURES 3 AND 4 HERE]

[INSERT TABLE 3 HERE]

[INSERT FIGURES 5 AND 6 HERE]

In summary, the Monte Carlo results suggest that a proper combination of joint and marginal tests is able to identify the right pattern of serial correlation, i.e., a joint test can be used to check if serial correlation is present, and the marginal tests to check which one is active. A "multiple testing" strategy (as in [11]) can be implemented using a Bonferroni approach, by rejecting the joint null if at least one of the marginal test lies in its rejection region, where the significance level for the marginal tests is halved to preserve the asymptotic size of the resulting joint test.

We highlight the fact that both marginal tests for serial correlation are needed to correctly identify the relevant serial correlation pattern, which is the main contribution of this paper. Conditional tests do not seem to offer any practical gain over marginal ones. This is practically convenient since the former require previous estimation of parameters of the serial correlation process to be controlled for, i.e., marginal test require simple GLS estimation of variances of random effects only. Finally, the Gaussian assumption does not seem restrictive.

5. Empirical application: Educational performance

As an empirical illustration we apply the proposed tests to study the dynamics of educational performance. The Program for International Student Assessment (PISA) tests are administered every three years in the OECD and a group of partner countries. The program collects harmonized information about students and schools using a single questionnaire, thus being comparable across countries.

Understanding the channels behind the dynamics of educational performance is a relevant issue for policy making purposes. Following [18], consider a stylized educational production function model

$$\text{score}_{ijt} = \alpha \, \text{stratio}_{ijt} + \beta \, \text{grade}_{ijt} + \gamma \, \text{pcgirls}_{ijt} + \delta \, \text{hisei}_{ijt} + u_{ijt},$$

where the outcome variable (score) is the mean score in a standardized international reading test. In this case i corresponds to the country, j to type of school, and t is the year in which the survey information was collected. Covariates include some of the usual inputs proposed in the literature: the average students-teacher ratio (stratio), the school year that students attend (grade), the proportion of girls at school (pcgirls) and an index of socio-economic level (hisei). All variables are averages at the school level. The first covariate is a proxy of the educational resources of the school, the second is a measure of students' experience, and the last two variables capture differences in educational performance related to demographic and economic factors. Finally, the error term u_{ijt} is assumed to have the nested structure in Section 2. 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. In each country there are three types of schools ($N = 3$): Private independent, Private government-dependent and Public. The information was collected for 2000, 2003, 2006, 2009 and 2012 ($T = 5$).

Table 4 shows the results of applying the tests proposed in this paper. At a 5% significance level, the joint null hypothesis of no autocorrelation in both cluster groups is rejected. However, further analysis reveals that both tests for $\lambda = 0$ rejected their corresponding null hypotheses, at the 1% of significance level for $LM_{\lambda \cdot \sigma}$ and 5% for $LM_{\lambda \cdot (\sigma, \rho)}$. Interestingly, the tests for $\rho = 0$ do not provide enough evidence to reject their null hypothesis. Therefore, the persistence of temporal exogenous shocks that affect educational performance seems to be related to those affecting the country, and not to the type of school.

[INSERT TABLE 4 HERE]

Clearly a detailed study of this subject exceeds the scope of this illustration. Nevertheless, the key point of the exercise is clear, which is to highlight the usefulness of the proposed tests to isolate the relevant source of intra-cluster correlation. The joint tests suggest correlated shocks in terms of serially correlated errors, but the proposed testing strategy indicates that the temporal persistence of shocks occurs only at the coarse (country) level, but not at the finer one (school). As noted by an anonymous referee, this could be due to the fact that at the school level one can expect less persistence, may be due to fresh students every year and some others graduating regularly, but more heterogeneity. Moreover, the country level reflects structural conditions that induce persistence. In particular, community and (extended) family conditions are more stable across time. The fact that autocorrelation is significant also shows that aggregate shocks are more persistent than individual shocks.

6. Conclusion and extensions

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

Several extensions could be considered. First, the simulation exercises reveal that the estimates of the nuisance parameter is demanding for moderate to large panel sizes (i.e., $M = 10, N = 10, T > 10$). The main problem relates to

the Wallace and Hussain [30] transformation of OLS residuals to obtain consistent estimators of the σ parameters. In particular, the computation of the inverse of the corresponding matrices is slow in both Stata and R, the two standard platforms used for the implementation. Thus, alternatives as those analyzed in [6] could be explored to speed up the process.

Second, the quest for the adequate level of clustering should also be analyzed in terms of heteroscedasticity. As argued by Wooldridge [31] both serial correlation and heteroscedasticity concerns call for cluster robust standard errors, even after GLS random effects estimation. Then, an important extension would be to adapt the results of [26] on testing for heteroscedasticity 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 should consider random effects, serial correlation and heteroscedasticity jointly.

Third, the model can be easily extended to the unbalanced case following [6–8] by considering that each i group is of size N_i, and each j_i intra-group cluster has T_{j_i} observations. We remark that in the unbalanced case the Kronecker product needs to be changed to allow for different intra-group sizes and all matrices are to be indexed by the corresponding group they belong to.

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

We consider best quadratic unbiased estimators for $(\sigma_\phi^2, \sigma_\eta^2, \sigma_\mu^2, \sigma_\epsilon^2)$ as a simple extension of the spectral decomposition given in Wallace and Hussain [30] and Baltagi ([3], pp. 38–39). Rewriting the variance covariance matrix under the null we have

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

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

$$\begin{aligned}
\Omega_0 &= NT\sigma_\phi^2(I_M \otimes \bar{J}_N \otimes \bar{J}_T) + N\sigma_\eta^2(I_M \otimes \bar{J}_N \otimes \bar{J}_T) + N\sigma_\eta^2(I_M \otimes \bar{J}_N \otimes \bar{E}_T) \\
&\quad + T\sigma_\mu^2(I_M \otimes \bar{J}_N \otimes \bar{J}_T) + T\sigma_\mu^2(I_M \otimes \bar{E}_N \otimes \bar{J}_T) + \sigma_\epsilon^2(I_M \otimes \bar{J}_N \otimes \bar{J}_T) \\
&\quad + \sigma_\epsilon^2(I_M \otimes \bar{J}_N \otimes \bar{E}_T) + \sigma_\epsilon^2(I_M \otimes \bar{E}_N \otimes \bar{J}_T) + \sigma_\epsilon^2(I_M \otimes \bar{E}_N \otimes \bar{E}_T) \\
&= \sigma_1^2(I_M \otimes \bar{E}_N \otimes \bar{E}_T) + \sigma_2^2(I_M \otimes \bar{E}_N \otimes \bar{J}_T) + \sigma_3^2(I_M \otimes \bar{J}_N \otimes \bar{E}_T) + \sigma_4^2(I_M \otimes \bar{J}_N \otimes \bar{J}_T) \\
&= \sigma_1^2 Q_1 + \sigma_2^2 Q_2 + \sigma_3^2 Q_3 + \sigma_4^2 Q_4,
\end{aligned}$$

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

Thus, asymptotically unbiased and consistent estimates can be obtained as

$$\hat{\sigma}_\epsilon^2 = \frac{u^\top Q_1 u}{M(N-1)(T-1)}, \quad \hat{\sigma}_2^2 = \frac{u^\top Q_2 u}{M(N-1)}, \quad \hat{\sigma}_3^2 = \frac{u^\top Q_3 u}{M(T-1)}, \quad \hat{\sigma}_4^2 = \frac{u^\top Q_4 u}{M},$$

and

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

Given that u is not observed, using $\hat{u} = Q_X u$, the OLS residuals where $Q_X = I_{MNT} - X(X^\top X)^{-1} X^\top$ is the residual matrix projection, produces an asymptotic bias. We follow [6] in adapting the estimator from [30] to our case.

Note that if a is an n-dimensional normal random vector and $a \sim \mathcal{N}(0, \Sigma)$, then if A is an $n \times n$ constant symmetric matrix, and $\mathrm{E}(a^\top A a) = \mathrm{tr}(A\Sigma)$. Now, $\hat{u} \sim \mathcal{N}(0, Q_X \Omega_0 Q_X)$ and then

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

This generates a 4×4 system of equations from which estimates of $\sigma_1^2, \ldots, \sigma_4^2$ 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 the strategy described in [9]. First, we construct the within residuals from a least-squares dummy variables fixed effects model. Consider the regression model

$$y_{ijt} = x_{ijt}^\top \beta + \sum_{i=1}^{M} \sum_{t=1}^{T} \eta_{it} d_{it} + \sum_{i=1}^{M} \sum_{j=1}^{N} \mu_{ij} d_{ij} + u_{ijt},$$

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. 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

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

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 [8, pp. 256–257], we use $\bar{J}_T^\rho = \iota_{\rho T} \iota_{\rho T}^\top / d^2$, $\bar{E}_T^\rho = I_T - \bar{J}_T^\rho$, where $\iota_{\rho T} = (\alpha_\rho, 1, 1, \ldots, 1)^\top$, $\alpha_\rho = \sqrt{(1+\rho)/(1-\rho)}$, $d^2 = \iota_{\rho T}^\top \iota_{\rho T} = \alpha_\rho^2 + T - 1$, 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 $\text{tr}(C_\rho C_\rho^\top) = T_\rho/T \text{tr}(I_T) = T_\rho/T \text{tr}(\bar{J}_T^\rho + \bar{E}_T^\rho)$, a factor that applies to the σ_η^2 term. These are used to construct Q_1, Q_2, Q_3, Q_4. Then,

$$\begin{aligned} \text{tr}(C_\rho \Omega_0 C_\rho^\top) &= NT_\rho \sigma_\phi^2 \text{tr}(I_M \otimes \bar{J}_N \otimes \bar{J}_T^\rho) + NT_\rho/T \sigma_\eta^2 \text{tr}(I_M \otimes \bar{J}_N \otimes \bar{J}_T^\rho) + NT_\rho/T \sigma_\eta^2 \text{tr}(I_M \otimes \bar{J}_N \otimes \bar{E}_T^\rho) \\ &\quad + T_\rho \sigma_\mu^2 \text{tr}(I_M \otimes \bar{J}_N \otimes \bar{J}_T^\rho) + T_\rho \sigma_\mu^2 \text{tr}(I_M \otimes \bar{E}_N \otimes \bar{J}_T^\rho) + \sigma_\epsilon^2 \text{tr}(I_M \otimes \bar{J}_N \otimes \bar{J}_T^\rho) \\ &\quad + \sigma_\epsilon^2 (I_M \otimes \bar{J}_N \otimes \bar{E}_T^\rho) + \sigma_\epsilon^2 \text{tr}(I_M \otimes \bar{E}_N \otimes \bar{J}_T \rho) + \sigma_\epsilon^2 \text{tr}(I_M \otimes \bar{E}_N \otimes \bar{E}_T^\rho) \\ &= \sigma_1^2 \text{tr}(I_M \otimes \bar{E}_N \otimes \bar{E}_T^\rho) + \sigma_2^2 \text{tr}(I_M \otimes \bar{E}_N \otimes \bar{J}_T^\rho) + \sigma_3^2 \text{tr}(I_M \otimes \bar{J}_N \otimes \bar{E}_T^\rho) + \sigma_4^2 \text{tr}(I_M \otimes \bar{J}_N \otimes \bar{J}_T^\rho) \\ &= \sigma_1^2 \text{tr}(Q_1) + \sigma_2^2 \text{tr}(Q_2) + \sigma_3^2 \text{tr}(Q_3) + \sigma_4^2 \text{tr}(Q_4), \end{aligned}$$

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

$$\tilde{\sigma}_\epsilon^2 = \frac{u^\top Q_1 u}{M(N-1)(T_\rho-1)}, \quad \tilde{\sigma}_2^2 = \frac{u^\top Q_2 u}{M(N-1)}, \quad \tilde{\sigma}_3^2 = \frac{u^\top Q_3 u}{M(T_\rho-1)}, \quad \tilde{\sigma}_4^2 = \frac{u^\top Q_4 u}{M},$$

and

$$\tilde{\sigma}_\mu^2 = \frac{\tilde{\sigma}_2^2 - \tilde{\sigma}_\epsilon^2}{T_\rho}, \quad \tilde{\sigma}_\eta^2 = \frac{\tilde{\sigma}_3^2 - \tilde{\sigma}_\epsilon^2}{NT_\rho/T}, \quad \tilde{\sigma}_\phi^2 = \frac{\tilde{\sigma}_4^2 - NT_\rho/T \tilde{\sigma}_\eta^2 - T_\rho \sigma_\mu^2 - \sigma_\epsilon^2}{NT_\rho}.$$

Given that u is not observed, we use the same adaptation procedure for the estimator from [30], 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 the strategy from [9] adapted to this case. First, we consider the regression model

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

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

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

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

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

This is equivalent to the transformation

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

Then follow Appendix A2, where λ replaces ρ and all the corresponding matrices and factors 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:

$$\begin{aligned} \mathrm{tr}(C_\lambda \Omega_0 C_\lambda^\top) &= NT_\lambda \sigma_\phi^2 \mathrm{tr}(I_M \otimes \bar{J}_N \otimes \bar{J}_T^\lambda) + N\sigma_\eta^2 \mathrm{tr}(I_M \otimes \bar{J}_N \otimes \bar{J}_T^\lambda) + N\sigma_\eta^2 \mathrm{tr}(I_M \otimes \bar{J}_N \otimes \bar{E}_T^\lambda) + T_\lambda \sigma_\mu^2 \mathrm{tr}(I_M \otimes \bar{J}_N \otimes \bar{J}_T^\lambda) \\ &\quad + T_\lambda \sigma_\mu^2 \mathrm{tr}(I_M \otimes \bar{E}_N \otimes \bar{J}_T^\lambda) + T_\lambda/T \sigma_\epsilon^2 \mathrm{tr}(I_M \otimes \bar{J}_N \otimes \bar{J}_T^\lambda) + T_\lambda/T \sigma_\epsilon^2 (I_M \otimes \bar{J}_N \otimes \bar{E}_T^\lambda) \\ &\quad + T_\lambda/T \sigma_\epsilon^2 \mathrm{tr}(I_M \otimes \bar{E}_N \otimes \bar{J}_T \lambda) + T_\lambda/T \sigma_\epsilon^2 \mathrm{tr}(I_M \otimes \bar{E}_N \otimes \bar{E}_T^\lambda) \\ &= \sigma_1^2 \mathrm{tr}(I_M \otimes \bar{E}_N \otimes \bar{E}_T^\lambda) + \sigma_2^2 \mathrm{tr}(I_M \otimes \bar{E}_N \otimes \bar{J}_T^\lambda) + \sigma_3^2 \mathrm{tr}(I_M \otimes \bar{J}_N \otimes \bar{E}_T^\lambda) + \sigma_4^2 \mathrm{tr}(I_M \otimes \bar{J}_N \otimes \bar{J}_T^\lambda) \\ &= \sigma_1^2 tr(Q_1) + \sigma_2^2 \mathrm{tr}(Q_2) + \sigma_3^2 \mathrm{tr}(Q_3) + \sigma_4^2 \mathrm{tr}(Q_4), \end{aligned}$$

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

$$\check{\sigma}_\epsilon^2 = T/T_\lambda \frac{u^\top Q_1 u}{M(N-1)(T_\lambda-1)}, \quad \check{\sigma}_2^2 = \frac{u^\top Q_2 u}{M(N-1)}, \quad \check{\sigma}_3^2 = \frac{u^\top Q_3 u}{M(T_\lambda-1)}, \quad \check{\sigma}_4^2 = \frac{u^\top Q_4 u}{M},$$

and

$$\check{\sigma}_\mu^2 = \frac{\check{\sigma}_2^2 - T_\lambda/T \check{\sigma}_\epsilon^2}{T_\lambda}, \quad \check{\sigma}_\eta^2 = \frac{\check{\sigma}_3^2 - T_\lambda/T \check{\sigma}_\epsilon^2}{N}, \quad \check{\sigma}_\phi^2 = \frac{\check{\sigma}_4^2 - N\check{\sigma}_\eta^2 - T_\lambda \sigma_\mu^2 - T_\lambda/T \sigma_\epsilon^2}{NT_\lambda}.$$

However, since u is not observed, we use the same adaptation procedure as in Appendix A1.

References

[1] J. Angrist, J.-S. Pischke, Mostly harmless econometrics: An empiricist's companion, Princeton University Press, Princeton 2009.

[2] T. Amemiya, The estimation of the variances in a variance-components mode, International Economic Review 12 (1971) 1–13.

[3] B.H. Baltagi, Econometric Analysis of Panel Data, Wiley, NJ, 2013.

[4] B.H. Baltagi, Q. Li, A transformation that will circumvent the problem of autocorrelation in an error-component model, J. Econometrics 48 (1991) 385–393.

[5] B.H. Baltagi, Q. Li, Testing AR(l) against MA(l) disturbances in an error component model, J. Econometrics 68 (1995) 133–151.

[6] B.H. Baltagi, S.H. Song, B.C. Jung, The unbalanced nested error component regression model, J. Econometrics 101 (2001) 357–381.

[7] B.H. Baltagi, S.H. Song, B.C. Jung, LM tests for the unbalanced nested panel data regression model with serially correlated errors, Annales d'économie et de statistique 65 (2002) 219–268.

[8] B.H. Baltagi, S.H. Song, B.C. Jung, B.C., Simple LM tests for the unbalanced nested error component regression model, Econometric Reviews 21 (2002) 167–187.

[9] B.H. Baltagi, P.X. Wu, Unequally spaced panel data regressions with AR(1) disturbances, Econometric Theory 15 (1999) 814–823.

[10] A. Bera, Y. Bilias, Rao's score, Neyman's $C(\alpha)$ and Silvey's LM tests: An essay on historical developments and some new results, J. Statist. Plann. Inf. 97 (2001) 9–44.

[11] A. Bera, C. Jarque, Model specification tests: A simultaneous approach, J. Econometrics 20 (1982) 59–82.

[12] A. Bera, W. Sosa-Escudero, M. Yoon, Tests for the error component model in the presence of local misspecification, J. Econometrics 101 (2001) 1–23.

[13] M. Bertrand, E. Duflo, S. Mullainathan, How much should we trust differences-in-differences estimates? Quarterly J. Economics 119 (2004) 249–275.

[14] T.S. Breusch, A.R. Pagan, The Lagrange multiplier test and its applications to model specification in econometrics, Review of Economic Studies 47 (1980) 239–253.

[15] C. Cameron, D.L. Miller, A practitioner's guide to cluster-robust inference, J. Human Resources 50 (2015) 317–372.

[16] M. Evans, Robustness of size of tests of autocorrelation and heteroscedasticity to nonnormality, J. Econometrics 51 (1992) 7–24.

[17] M. Furno, LM tests in the presence of non-normal error distributions, Econometric Theory 16 (2000) 249–261.

[18] E.A. Hanushek, L. Woessmann, The economics of international differences in educational achievement, In: E.A. Hanushek, S. Machin, L. Woessmann (Eds.), Handbook of the Economics of Education 3, North Holland, Amsterdam, 2011, pp. 89–200.

[19] D.A. Harville, Maximum likelihood approaches to variance component estimation and to related problems, J. Amer. Statist. Assoc. 72 (1977) 320–340.

[20] Y. Honda, Testing the error components model with non-normal disturbances, Rev. Economic Studies 52 (1985) 681–690.

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

[22] G. Kézdi, Robust standard error estimation in fixed-effects panel models, Hungarian Statistical Review Special Number 9 (2004) 95–116.

[23] G. King, M.E. Roberts, How robust standard errors expose methodological problems they do not fix, and what to do about it, Political Analysis 23 (2015) 159–179.

[24] L. Matyas, Error components models, In: L. Matyas, P. Sevestre (Eds.), The Econometrics of Panel Data, 2nd edition, Kluwer Academic Publishers, Dordrecht, 2012, pp. 50–76.

[25] G. Montes-Rojas, An equicorrelation Moulton factor in the presence of arbitrary intra-cluster correlation, Economics Letters 45 (2016) 221–224.

[26] G. Montes-Rojas, W. Sosa-Escudero, Robust tests for heteroskedasticity in the one-way error components model, J. Econometrics 160 (2011) 300–310.

[27] B.R. Moulton, Random group effects and the precision of regression estimates, J. Econometrics 32 (1986) 385–397.

[28] B.R. Moulton, Diagnostics for group effects in regression analysis, J. Bus. Econom. Statist. 5 (1987) 275–282.

[29] B.R. Moulton, An illustration of a pitfall in estimating the effects of aggregate variables in micro units, Review of Economics and Statistics 72 (1990) 334–338.

[30] T.D. Wallace, A. Hussain, The use of error components models in combining cross section with time series data, Econometrica 37 (1969) 55–72.

[31] J.M. Wooldridge, Econometric analysis of cross section and panel data, 2nd edition, MIT Press, Boston, MA, 2010.

Table 1: Empirical size and power for normally distributed errors

	T	ρ=0, λ=0 M=5 N=5	ρ=0, λ=0 M=5 N=10	ρ=0, λ=0 M=10 N=5	ρ=0, λ=0 M=10 N=10	ρ=0.2, λ=0 M=5 N=5	ρ=0.2, λ=0 M=5 N=10	ρ=0.2, λ=0 M=10 N=5	ρ=0.2, λ=0 M=10 N=10	ρ=0, λ=0.2 M=5 N=5	ρ=0, λ=0.2 M=5 N=10	ρ=0, λ=0.2 M=10 N=5	ρ=0, λ=0.2 M=10 N=10
$LM_{(\rho,\lambda)\tau}$	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%
$LM_{\lambda\tau}$	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%
$LM_{\lambda(\sigma\varphi)}$	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%
$LM_{\rho\tau}$	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%
$LM_{\rho\cdot(\sigma,\lambda)}$	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 χ_1^2 distributed errors

	T	ρ=0, λ=0 M=5 N=5	ρ=0, λ=0 M=5 N=10	ρ=0, λ=0 M=10 N=5	ρ=0, λ=0 M=10 N=10	ρ=0.2, λ=0 M=5 N=5	ρ=0.2, λ=0 M=5 N=10	ρ=0.2, λ=0 M=10 N=5	ρ=0.2, λ=0 M=10 N=10	ρ=0, λ=0.2 M=5 N=5	ρ=0, λ=0.2 M=5 N=10	ρ=0, λ=0.2 M=10 N=5	ρ=0, λ=0.2 M=10 N=10
$LM_{(\rho,\lambda)\tau}$	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%
$LM_{\lambda\tau}$	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%
$LM_{\lambda(\sigma\varphi)}$	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%
$LM_{\rho\tau}$	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%
$LM_{\rho\cdot(\sigma,\lambda)}$	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's t_5 distributed errors
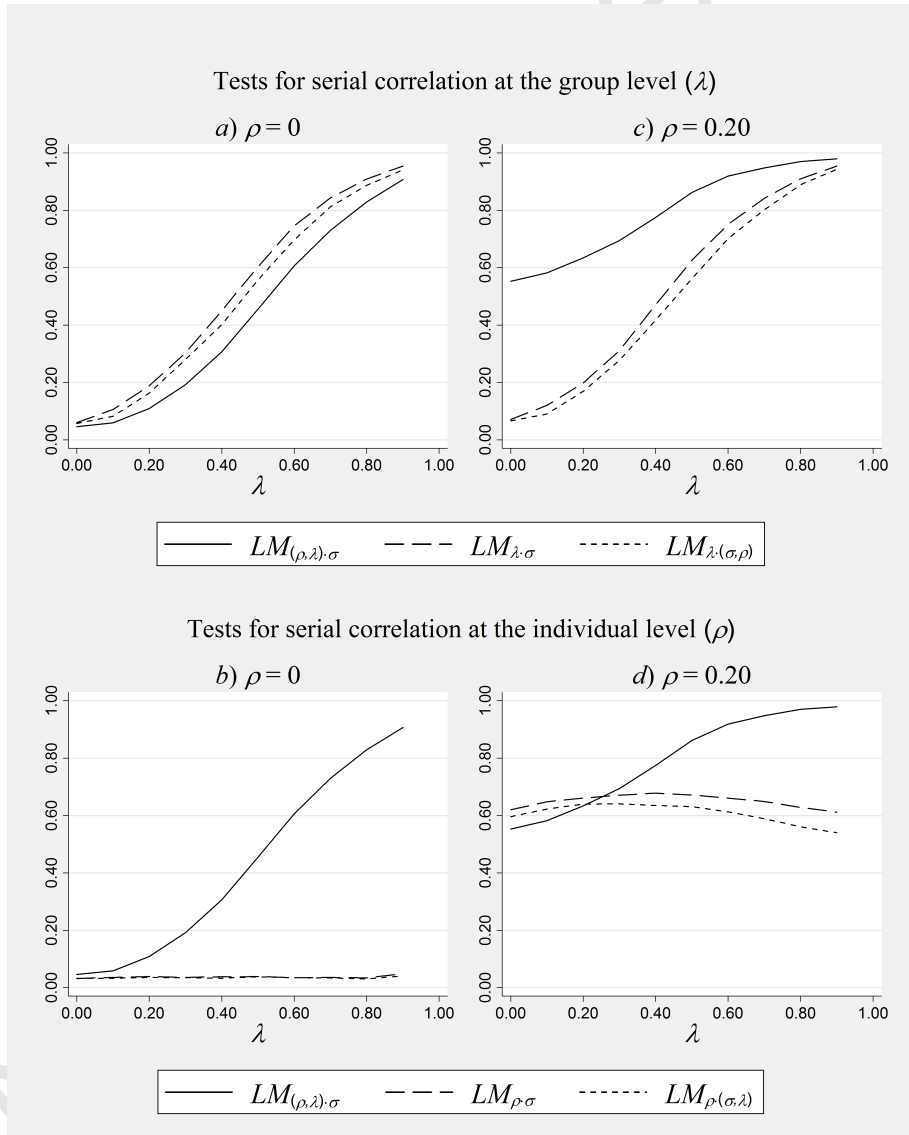
	T	$\rho=0,\lambda=0$ $M=5$ $N=5$	$N=10$	$M=10$ $N=5$	$N=10$	$\rho=0.2,\lambda=0$ $M=5$ $N=5$	$N=10$	$M=10$ $N=5$	$N=10$	$\rho=0,\lambda=0.2$ $M=5$ $N=5$	$N=10$	$M=10$ $N=5$	$N=10$
$LM_{(\rho,\lambda)\sigma}$	5	5.0%	5.2%	3.4%	4.9%	18.2%	35.3%	36.0%	67.7%	6.5%	8.6%	4.9%	8.8%
	10	3.8%	4.4%	4.7%	4.5%	52.8%	85.9%	90.6%	99.8%	9.2%	15.7%	13.6%	21.2%
$LM_{\lambda\sigma}$	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%
$LM_{\lambda(\sigma\rho)}$	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%
$LM_{\rho\sigma}$	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%
$LM_{\rho(\sigma,\lambda)}$	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%

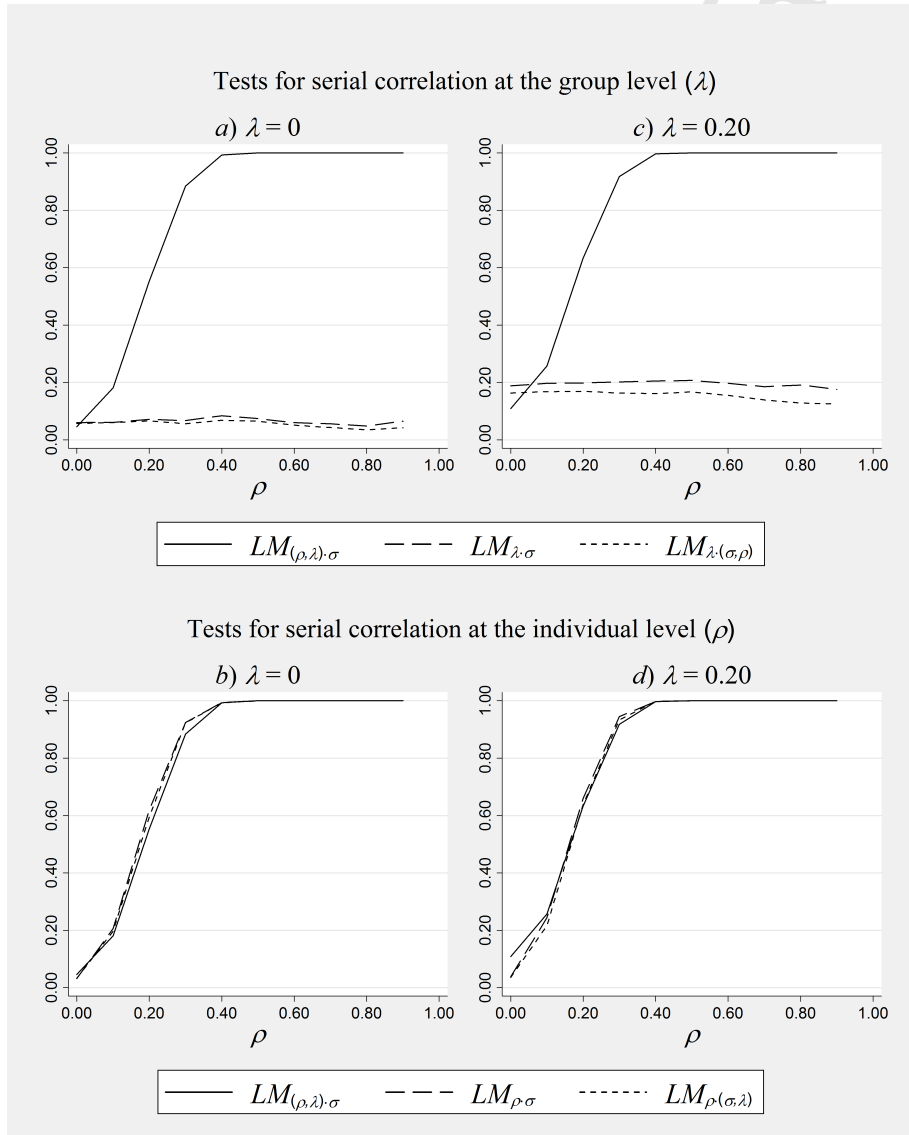Notes: Monte Carlo experiments based on 1000 replications and a 5% nominal size.

14

Table 4: PISA nested autocorrelation analysis

|	Test statistic	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.

Figure 1: Empirical size and power, normally distributed errors and $\lambda \in \{0, 0.1, 0.2, \ldots, 0.9\}$

Tests for serial correlation at the group level (λ)

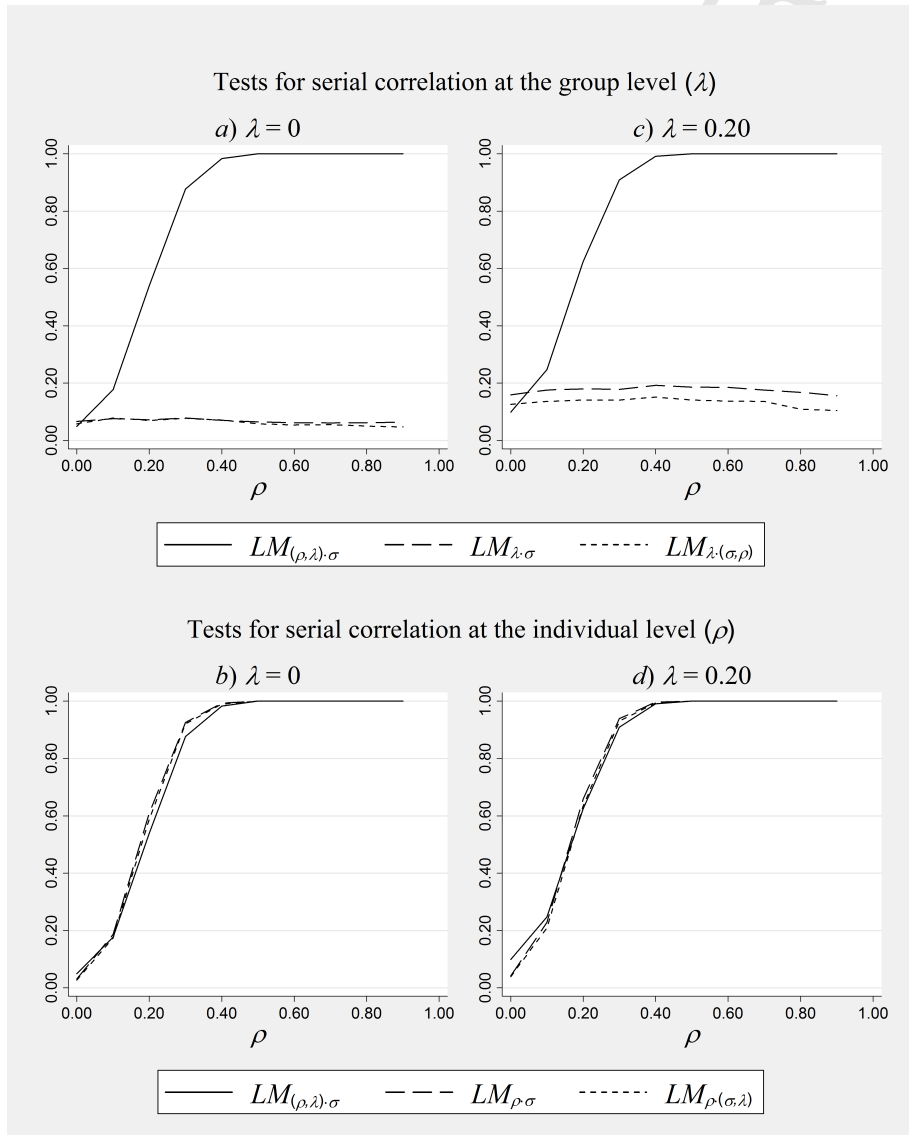Tests for serial correlation at the individual level (ρ)

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 and $\rho \in \{0, 0.1, 0.2, \ldots, 0.9\}$
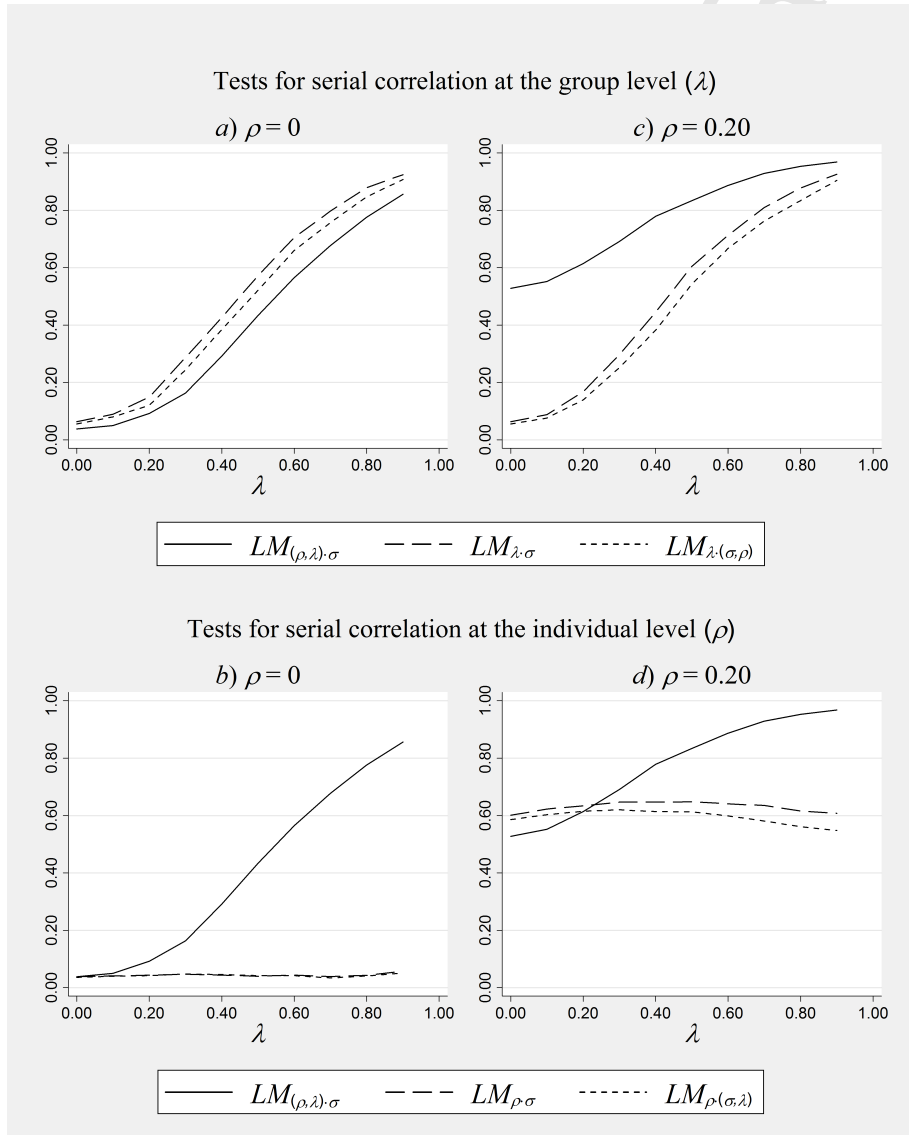

Notes: Monte Carlo experiments based on 1000 replications and a 5% nominal size. Panel size $M = 5, N = 5, T = 10$.

Figure 3: Empirical size and power, χ_1^2 distributed errors and $\lambda \in \{0, 0.1, 0.2, \ldots, 0.9\}$
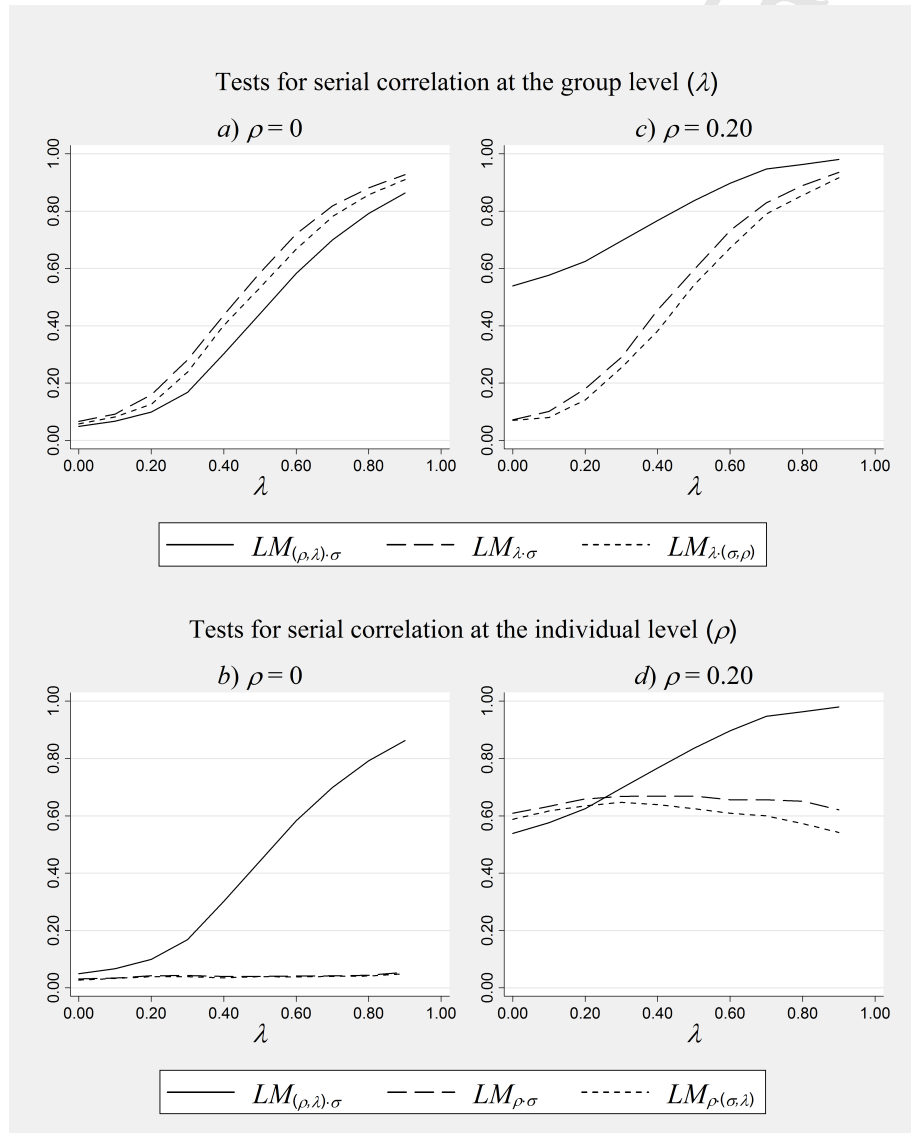
Notes: Monte Carlo experiments based on 1000 replications and a 5% nominal size. Panel size $M = 5, N = 5, T = 10$.

Figure 4: Empirical size and power, χ_1^2 distributed errors and $\rho \in \{0, 0.1, 0.2, \ldots, 0.9\}$
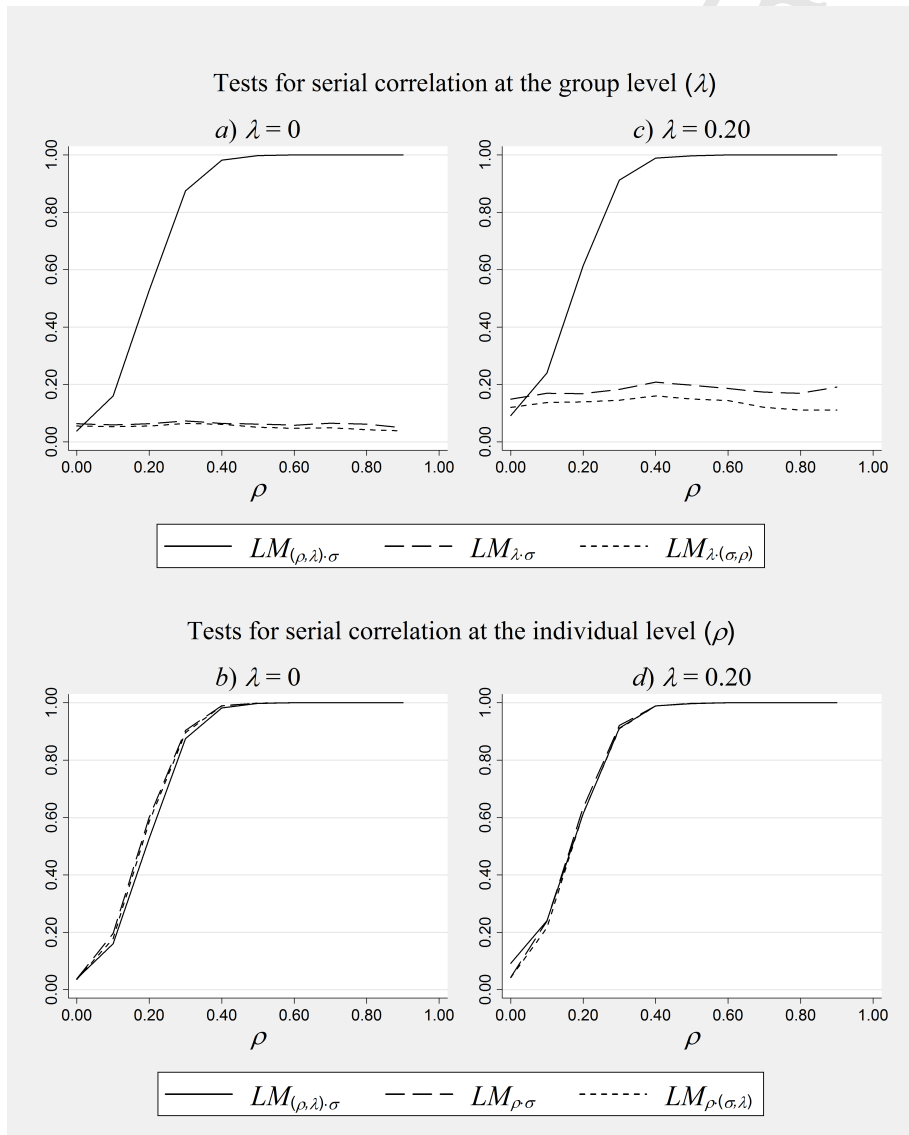

Notes: Monte Carlo experiments based on 1000 replications and a 5% nominal size. Panel size $M = 5, N = 5, T = 10$.

Figure 5: Empirical size and power, Student's t_5 distributed errors and $\lambda \in \{0, 0.1, 0.2, \ldots, 0.9\}$

Tests for serial correlation at the group level (λ)

a) $\rho = 0$

c) $\rho = 0.20$

$LM_{(\rho,\lambda)\cdot\sigma}$ $LM_{\lambda\cdot\sigma}$ $LM_{\lambda\cdot(\sigma,\rho)}$

Tests for serial correlation at the individual level (ρ)

b) $\rho = 0$

d) $\rho = 0.20$

$LM_{(\rho,\lambda)\cdot\sigma}$ $LM_{\rho\cdot\sigma}$ $LM_{\rho\cdot(\sigma,\lambda)}$

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

19

Figure 6: Empirical size and power, Student's t_5 distributed errors and $\rho \in \{0, 0.1, 0.2, \ldots, 0.9\}$

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