On the Use of Fixed Effects Estimators for Time-Series Cross-Section Data

August 10, 2011

Gregory J. Wawro*

Cyrus Samii

Ida Pagter Kristensen

Under review. Please do not cite without permission.

A previous version of this paper was presented at the 2003 Summer Methods Conference. The authors gratefully acknowledge Neal Beck and Jonathan Katz for sharing code and comments. Other helpful comments were provided by Genie Baker, Robert Erikson, Robert Franzese, Jennifer Hill, Charles Himmelberg, Walter Mebane, Michael K. Miller, Helen Milner, and Michael Sobel. Lucas Leeman provided valuable research assistance. Computing support was provided by the Institute for Social and Economic Policy and Research at Columbia University. Wawro acknowledges the generous support of a John M. Olin Faculty Fellowship awarded by the National Association of Scholars. The authors alone bear responsibility for errors or omissions in the paper.

*Corresponding author, reachable at gjw10@columbia.edu.

On the Use of Fixed Effects Estimators for Time-Series Cross-Section Data

August 10, 2011

Abstract

This paper examines issues related to the use of fixed effects estimators for analysis of time-series cross-section (TSCS) data. We focus on issues concerning serial correlation and the use of dynamic specifications to deal with it, as well as potential pitfalls related to time invariant explanatory variables. We discuss a general framework for thinking about robust standard error estimators for TSCS data and evaluate alternative approaches for various sample sizes and conditions. We conduct Monte Carlo experiments to evaluate the performance of "robust FE" approaches relative to other popular estimators and provide guidance to those who confront the various estimation issues presented by TSCS data.

1 Introduction

This paper examines issues related to the use of fixed effects estimators for analysis of time-series cross-section (TSCS) data. Fixed effects (FE) models can be a useful tool because they exploit the existence of multiple observations on cross-sectional units over time to account for unobserved, unit-specific heterogeneity while obtaining estimates on observable, substantive variables of interest.[1] While this is an attractive feature of these models, two key issues raise concerns about their performance for TSCS analysis. The first issue was pointed out by Beck and Katz (1995; 1996) in their seminal and widely influential articles on methods for TSCS data where they discuss how researchers are likely to encounter non-spherical errors that can compromise the performance of standard estimators. In particular, TSCS data are often characterized by serial correlation, which leads to incorrect standard errors if not accounted for. However, the popular correction of including a lagged dependent variable (LDV) may not work well, because the inclusion of the lag can cause bias and inconsistency in the parameter estimates in FE models.[2] We present analytical results and simulations to show that while biases from combining LDVs with FE are typically insubstantial in moderately sized samples, robust standard error estimators are available that perform quite well relative to other options to address non-spherical errors and obviate the need to include the LDV to correct for serial correlation. Including an LDV in a model specification is a choice that should never be taken lightly, and we show that researchers concerned about serial correlation in TSCS data should not feel compelled to include them when estimating FE models. When the LDV is included in a model for substantive theoretical reasons, we show that with enough time periods in the data, reliable inferences can be made if the LDV is of substantive interest, but only if no serial correlation remains after conditioning on the LDV.

A second issue concerns the loss of the ability to make inferences about time invariant or slow-moving variables. Such variables are either perfectly or highly collinear with fixed

[1] By "fixed effects" models here we mean the textbook within-group estimator (e.g., see Hsiao (2003).

[2] Note that this is also potentially problematic for FE models if a lag is included for substantive theoretical reasons.

unit-specific effects and thus coefficients on them are either not identified or are difficult to estimate with any precision. Attempting to account for unit heterogeneity may come at the cost of restricting us from reaching substantive conclusions about theoretically important variables.[3] Thus, researchers face a difficult trade-off when contemplating the use of FE models. However, we will show that choosing to include time invariant variables (TINVs) in the place of unit effects can lead to incorrect inferences on such variables, such that little of substantive value can be learned about them.

We address these issues in detail and provide practical guidance on estimation approaches. We discuss a general framework for thinking about robust standard error estimators for TSCS data and evaluate alternative approaches for various sizes of N and T. We conduct Monte Carlo experiments to assess the performance of FE approaches relative to one of the most popular approaches in the political science—using ordinary least squares point estimates with panel corrected standard errors (PCSEs) and a LDV to correct for serial correlation—as well as other plausible candidates.

2 Debates about fixed effects models for TSCS data in political science

TSCS analyses are prevalent in political science research, despite the fact that methodological work on TSCS data still leaves open many questions. For example, while methodological work on FE models for panel contexts (large N, short T) is voluminous, work on FE models

[3]This is not a problem with random effects estimators, since they impose the restriction that the unit effects and the explanatory variables are independent of each other. If that restriction is violated, however, we will get inconsistent estimates. We seek to avoid entering the philosophical debate of whether it is more appropriate to consider unit effects as fixed or random, since it is impossible to resolve here. It is more of a practical question in our view, since in our experience it is rarely the case that the assumption that unit effects and the explanatory variables are completely orthogonal is met. As such, we will not investigate the performance of random effects estimators (including random coefficient and hierarchical model estimators) that make these kinds of orthogonality assumptions. Another potential drawback of using FE models that we will not discuss here concerns problems of panel attrition (see Hsiao 2003, ch. 8–9).

for TSCS data is comparatively thin. In the article that profoundly changed the way political scientists analyze TSCS data, Beck and Katz (1995) in endnote 4 (p. 645) mention that unit-specific effects may be included among the set of explanatory variables in model specifications, contending that, "fixed effects present no special problems for TSCS models, because the number of unit-specific dummy variables required is not large." Yet certain features of TSCS data—particularly serial correlation and the possibility that dynamics will play a more important role than they do in short T settings—are not clearly addressed in the panel data literature on FE estimators.

Beck (2001, 282–287) addresses the issue of unit heterogeneity in TSCS data, but does not link this discussion with the issue of serial correlation or dynamics. He does recommend that researchers test for unit heterogeneity in their data, however. The dangers of unmodeled unit-specific effects in TSCS data are explicitly addressed by Green, Kim and Yoon (2001) who drive home the point that when unmodeled unit effects are correlated with explanatory variables, OLS slope coefficients are biased and inconsistent.[4] The source of the problem is that unit-specific effects are relegated to the disturbance term if they are not modeled explicitly, which results in omitted variable bias. Green et al. argue for the application of standard FE estimators to eliminate this problem.

In response to Green, Kim and Yoon (2001), Beck and Katz (2001) argue that including FE for models with continuous dependent variables to account for unobserved heterogeneity can be worse than leaving them out. The bias may be minimal in certain situations—namely, when the explanatory power of the unit effects is low. Since FE are perfectly collinear with time invariant variables and highly collinear with variables that move slowly, it is generally the case that the former must be dropped if FE are included in the model, while the latter will have imprecisely estimated coefficients. Thus, the loss in terms of inference on important substantive variables that are time invariant or move slowly can outweigh the gains of using FE. Beck and Katz (2001, 493) also discuss the issue of dynamics, arguing that including lags of the dependent variable can make FE less relevant (e.g., FE are similar to including a lag with a coefficient of one). This is argument requires further investigation. In the literature

[4] If they are not correlated with explanatory variables, then we should simply get bias in the intercept.

on dynamic panel models, unit heterogeneity and dynamics are treated as separate features to be modeled, which is consistent with Beck's (2001) position that these aspects of the data generating process (DGP) should be treated as substantive issues and not mere nuisance.[5] Our concern is that this claim might lead researchers to believe that including unit effects is not necessary in TSCS data in order to produce reliable inferences.

Wilson and Butler's (2007) review of 195 articles that use TSCS data indicates that there is insufficient attention paid to the use of FE estimators for TSCS data, with the lion's share of these articles neither reporting tests for unit effects nor attempting to account for their presence. It is our sense that a good deal of confusion remains among researchers as to the implications of using FE with TSCS data. While researchers must consider a variety of potential estimation problems that we cannot address in one paper, we hope to resolve some important questions involving how to address temporal dynamics and TINVs when researchers believe that FE are present in their data.

3 Models and Estimators for TSCS Data

3.1 Fixed effects models for TSCS data

We consider a general model for TSCS data that enables us to incorporate very basic dynamics, as well as various types of non-spherecity in the error term.[6] The following outcome model motivates our discussion of estimators and also defines a data generating process for the Monte Carlo simulations later in the paper:

$$y_{it} = \gamma y_{i,t-1} + \beta x_{it} + \theta x_{i,t-1} + \lambda z_i + \alpha_i + \epsilon_{it}, \tag{1}$$

for $i = 1, ..., N$ and $t = 1, ..., T$. As a TSCS model, as distinct from a "panel" model, we presume N is small relative to T. The coefficients γ and θ capture dynamic effects, which

[5] For a review of the literature on dynamic panel models see Arellano and Honoré (2001) and Baltagi (2008, Ch. 8).

[6] We do not consider more complicated dynamics or lags because the goal here is to focus attention on the main problem at hand, which is the possible presence of unmeasured unit effects.

in the present model are limited to effects that carry over from a single period.[7] The α_i's are unmeasured "unit effects" that reduce to unit-specific intercepts for each i. The variable x_{it} is a time-varying covariate and z_i is a time-invariant covariate. These covariates may or may not be correlated with the unit effect, α_i. For simplicity, assume that the underlying data generating process is such that no other measured covariates serve to confound the relationship between these two covariates and the outcome. The variable u_{it} is an error term assumed to have zero mean conditional on any of the other terms on the right hand side. We allow for the error term to exhibit any of the following:

- Contemporaneous correlation: the errors across cross-sectional units are correlated due to common shocks in a given time period.
- Panel heteroskedasticity: the error variance differs across cross-sectional units due to characteristics unique to the units.
- Serial correlation: the errors within units are temporally correlated. Following the convention in the literature, we will focus on autoregressive processes of order 1 (AR(1)).

Following Beck and Katz (1995), we also allow for the time varying covariate, x_{it}, to exhibit such dependencies, although in a manner that is uncorrelated with u_{it}.

3.2 Coefficient estimates

We assume that the primary goal is to estimate the effect (or partial association) of x_{it} on y_{it}. In our estimation equation, this effect is modeled in terms of a time- and unit-invariant instantaneous effect, β, and lagged effect, θ.[8] The estimation method needs to attend to possible confounding that α_i may cause. To the extent that the α_i's are correlated with the x_{it}'s, pooled OLS estimates of β and θ, which relegate the α_i's to the error, will result

[7]See DeBoef and Keele (2008) on interpreting coefficients from this type of autoregressive distributed lag model.

[8]Even if in the "real world" the relationship is not time- and unit-invariant, a model that assumes fixed coefficients may still provide a useful summary. See Angrist and Pischke (2009, Ch. 4) for a relevant discussion.

in omitted variable bias. The FE estimator applies the "within" transformation, which subtracts off unit-specific means for y_{it}, x_{it}, and $x_{i,t-1}$. This transformation purges the data of estimated α_i values, thereby removing the potential for confounding due to such fixed, additively separable variation.[9] The time invariant z_i is also swept away, given that the unit specific mean of z_i is just z_i. Therefore, our ability to estimate λ is purposefully sacrificed due to non-separability between z_i and α_i.

If $\gamma \neq 0$, certain problems arise. Of course, if serial correlation is present in u_{it}, then identification of γ becomes problematic due to the correlation between $y_{i,t-1}$ and the error. But even without such serial correlation in the error, the within transformation induces correlation between the transformed LDV and the transformed error term. Both the transformed error term and the transformed LDV are produced by subtracting off the mean of error terms across periods. In finite samples, this may induce non-trivial dependence, although this dependence weakens as the number of time periods grows. Thus, FE estimates of γ suffer from the bias discussed by Nickell (1981), and such bias propagates to estimates of β if x_{it} and $y_{i,t-1}$ are correlated. The situation is something of a catch-22, because omitting $y_{i,t-1}$ may result in omitted variable bias. Importantly, however, the bias on γ is of order $1/T$, in which case estimates of γ are consistent for TSCS data.[10] Therefore, this sort of "Nickell bias" may not present much of a problem in TSCS data with large T. For panel data situations, instrumental-variables (IV)-based "dynamic panel models" have been proposed to surmount the Nickell bias problem, but it is doubtful that they are appropriate for TSCS data. These are generalized methods of moments (GMM) estimators, and their consistency depends crucially on N, which is typically not large in TSCS data. Adolph, Butler and Wilson (N.d.) find that IV/GMM estimators do not perform well with data of the dimensions that we examine in this paper. Beck and Katz (2004) also show that with typical TSCS data, the performance and simplicity of LSDV trumps both IV/GMM estimators as well as the approach advocated by Kiviet (1995).[11] Thus, it would seem that with large enough T,

[9]Using unit-specific dummy variables to account for unit effects is algebraically equivalent to the within transformation for estimating β. This approach is known as least squares dummy variable or LSDV.

[10]Nickell's original discussion concerned panel data which are presumed to go to infinity in N with fixed T, in which case the γ estimates would be biased and inconsistent.

[11]Simulations that we have carried out have found that IV-based dynamic panel estimators require that N

there is justification for recommending including $y_{i,t-1}$ in the specification of the FE model to address serial correlation of the autoregressive variety, even though the LDV is not explicitly part of the DGP. We assess this recommendation in our Monte Carlo experiments below.

As a final consideration we might ask might there be ways that the judicious use of lags could allow us to account for unobserved, correlated unit effects while also permitting us to include time invariant variables in the specification? Unfortunately not. If equation (1) represents the true data generating process, then even conditional on the LDV, $y_{i,t-1}$, the regressor, x_{it}, is correlated with unobserved α_i and so OLS will fail to estimate β without bias. If $\gamma = 0$, then introducing y_{it} into specification may reduce some of the bias, but it will not remove it completely. In addition, doing so will induce new serial dependencies in the error term.[12] If $x_{i,t-1}$ is added to the specification, it may reduce the bias on β a bit more, but at a cost of obscuring our ability to estimate λ in such a way as to add no appeal relative to FE.[13] The Monte Carlo simulations below illustrate this point, clarifying how lag specifications and FE are not substitutes.

be in the many hundreds to obtain reasonable results. Results are available from the authors upon request.

[12]To see this, $\gamma = 0$ implies that introducing $y_{i,t-1}$ into the specification is analogous to attempting to estimate,

$$y_{it} = \gamma y_{i,t-1} + \beta x_{it} + \lambda z_i + \alpha_i + \epsilon_{it} - \gamma y_{i,t-1} = \gamma y_{i,t-1} + \beta x_{it} + \lambda z_i + \underbrace{(1-\gamma)\alpha_i + \epsilon_{it} - \gamma\epsilon_{i,t-1} - \gamma\beta x_{i,t-1}}_{\text{compound error term}},$$

in which case x_{it} is still endogenous and therefore OLS will not be consistent for β. The bias will be less due to the estimated $(1-\hat{\gamma})$ term's attenuation of the correlation between α_i and x_{it}. Furthermore, $y_{i,t-1}$ is also endogenous to the last component of the compound error, and the OLS residual may exhibit positive or negative serial correlation, due to the correlations between α_i and the $x_{i,t-1}$'s.

[13]Consider the first-differenced version of equation (1),

$$y_{it} - y_{i,t-1} = \beta(x_{it} - x_{i,t-1}) + \epsilon_{it} - \epsilon_{i,t-1}.$$

Including both $y_{i,t-1}$ and $x_{i,t-1}$ in the specification amounts to estimating a version of the first-differenced equation without appropriate restrictions on the coefficients for $y_{i,t-1}$ and $x_{i,t-1}$ necessary to obtain unbiased β estimates (the restricted coefficient values would be 1 and β, respectively). To the extent that the estimates depart from these restricted values, we may obtain a coefficient estimate on z_i when it is clear that we should not. The estimate for λ is a product of the biased coefficient estimates on $y_{i,t-1}$ and x_{it}.

3.3 Cluster robust standard errors

Beck and Katz (1995, 1996) and Beck (2001) have shown convincingly that feasible generalized least squares (FGLS) approaches to addressing issues of contemporaneous correlation, panel heteroskedasticity, and serial correlation in TSCS data are unreliable. The central problem with FGLS is that it requires the estimation of a large number of paramaters with relatively few data points. A popular alternative approach for addressing such dependencies is through the use of various "cluster robust" standard errors. The basic idea is as follows. Suppose that one stacks the $T \times K$ cross-section-specific matrices of regressors (including a constant) into a single matrix $NT \times K$ matrix, $\mathbf{X}$, and the same is done for the outcome variable to produce an $NT \times 1$ vector, $\mathbf{y}$. Least squares regression estimates are given by $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$, and the covariance matrix for these estimate is,

$$\mathrm{Cov}\,(\mathbf{b}) = \mathrm{E}\,\left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\epsilon}\boldsymbol{\epsilon}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\right]. \tag{2}$$

The inner term, $\boldsymbol{\epsilon}\boldsymbol{\epsilon}'$, is a matrix of error term cross-products since $\boldsymbol{\epsilon}$ is the $NT \times 1$ vector containing the stacked $\epsilon_{i,t}$ for all i and t. Holding the regressors fixed enables us to focus attention on $\mathrm{E}\,(\boldsymbol{\epsilon}\boldsymbol{\epsilon}') = \mathrm{Cov}\,(\boldsymbol{\epsilon})$, the $NT \times NT$ covariance matrix for the error terms. Assumptions about error term dependencies are conveyed by this covariance matrix.

One way to think about robust estimation strategies for accounting for error dependencies among cross-sectional units and time periods is to write $\mathrm{Cov}\,(\boldsymbol{\epsilon})$ as a partitioned matrix:

$$\mathrm{Cov}\,(\boldsymbol{\epsilon}) = \begin{bmatrix} \mathrm{Cov}\,(\boldsymbol{\epsilon}_1) & \mathrm{Cov}\,(\boldsymbol{\epsilon}_1, \boldsymbol{\epsilon}_2) & \cdots & \mathrm{Cov}\,(\boldsymbol{\epsilon}_1, \boldsymbol{\epsilon}_N) \\ \mathrm{Cov}\,(\boldsymbol{\epsilon}_1, \boldsymbol{\epsilon}_2) & \mathrm{Cov}\,(\boldsymbol{\epsilon}_2) & \cdots & \mathrm{Cov}\,(\boldsymbol{\epsilon}_2, \boldsymbol{\epsilon}_N) \\ \vdots & \vdots & \ddots & \vdots \\ & & & \mathrm{Cov}\,(\boldsymbol{\epsilon}_{N-1}, \boldsymbol{\epsilon}_N) \\ \mathrm{Cov}\,(\boldsymbol{\epsilon}_1, \boldsymbol{\epsilon}_N) & \mathrm{Cov}\,(\boldsymbol{\epsilon}_2, \boldsymbol{\epsilon}_N) & \cdots & \mathrm{Cov}\,(\boldsymbol{\epsilon}_N) \end{bmatrix}.$$

where $\boldsymbol{\epsilon}_i$ is a $T \times 1$ vector of error terms for unit i for all time periods (and therefore each block of the matrix is $T \times T$). The matrix $\mathrm{Cov}\,(\boldsymbol{\epsilon}_i)$ contains the covariances for pairs of time periods for unit i, while the matrix $\mathrm{Cov}\,(\boldsymbol{\epsilon}_i, \boldsymbol{\epsilon}_j)$ contains the covariances for units i and j within and across time periods. There are potentially $NT(NT+1)/2$ unknown variance and covariance parameters to estimate in $\mathrm{Cov}\,(\boldsymbol{\epsilon})$—far more than the NT data points available to estimate them with. The advantage of robust standard error approaches is that, rather than estimating $\mathrm{Cov}\,(\boldsymbol{\epsilon})$ directly and then plugging it into eq. (2), we can instead focus on computing $\mathbf{X}'\text{Cov}(\boldsymbol{\epsilon})\mathbf{X}$ which reduces the dimensionality of the estimation problem by considering the variance and covariance terms in conjunction with the explanatory variables (note that this matrix is only $K \times K$). This effectively addresses the parameter to data ratio problem, and given certain conditions, produces consistent estimates of $\text{Cov}(\mathbf{b})$, and therefore, reliable estimates of standard errors.[14]

Alternative robust estimators propose different ways of computing $\mathbf{X}'\widehat{\text{Cov}(\boldsymbol{\epsilon})}\mathbf{X}$, in part by imposing various restrictions on the elements of $\text{Cov}(\boldsymbol{\epsilon})$ that represent particular correlation patterns ($\widehat{\text{Cov}(\boldsymbol{\epsilon})}$ denotes that we have replaced the $\epsilon_{i,t}$ with consistent residual estimates $\hat{\epsilon}_{i,t}$). In other words, a "cluster robust" estimator for $\text{Cov}(\mathbf{b})$ is given by substituting products of residuals (that is, $\hat{\boldsymbol{\epsilon}}$'s), or averages of such products in the case of constant correlations, for the non-zero products of error terms that appear in $\text{Cov}(\boldsymbol{\epsilon})$. For example, PCSEs assume $\text{Cov}(\boldsymbol{\epsilon}_i)$ and $\text{Cov}(\boldsymbol{\epsilon}_{i,j})$ are diagonal matrices with the variance terms σ_i along the diagonal in the former and the covariance terms $\sigma_{i,j}$ along the diagonal in the latter. Zeros in the off-diagonal elements in these matrices imply no serial correlation in the errors, hence the need to do something else to correct for it if present. To achieve this structure for the covariance matrix, OLS residuals are used to estimate

$$\widehat{\text{Cov}(\boldsymbol{\epsilon})} = \frac{\mathbf{E}'\mathbf{E}}{T} \otimes \mathbf{I}_T, \tag{3}$$

where $\mathbf{E}$ is a $T \times N$ matrix of the re-shaped $NT \times 1$ vector of OLS residuals, such that the columns contain the $T \times 1$ vectors of residuals for each cross-sectional unit. Thus, the non-zero elements in the covariance matrix are filled with cross-products that are averaged over cross-sections and time. Note that we could apply the within-group transformation to the data first and use the resulting residuals in computing eq. (3). This estimator, which we refer to as FE-PCSEs, would enable us to account explicitly for unit effects in addition to the kinds of non-sphericity that PCSEs address.

[14]The convergence of the resulting matrix to $\text{Cov}(\mathbf{b})$ is based on the condition of diminishing influence of constituent components as the sample size increases, as demonstrated by Huber (1967) and, specifically for the linear model, by White (1980). The generalization of Huber's and White's results to clustered dependencies is not attributable to any single author. Nonetheless, foundational work on the topic includes Liang and Zeger (1986), as well as Rogers (1993), who implemented the method into Stata in the early 1990s.

Within-cross-section dependence (e.g., serial correlation) may be accounted for by clustering units by cross section, with consistency derived as a condition of an ever increasing number of (conditionally) independent cross-sections (in other words, fixed T and growing N). Thus, one may estimate,

$$\frac{N}{N-1}\frac{NT-1}{NT-K}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\left(\sum_{i=1}^{N}\mathbf{X}_i{}'\hat{\boldsymbol{\epsilon}}_i\hat{\boldsymbol{\epsilon}}_i'\mathbf{X}_i\right)\left(\mathbf{X}'\mathbf{X}\right)^{-1}, \tag{4}$$

where $\mathbf{X}'\widehat{\mathrm{Cov}\,(\boldsymbol{\epsilon})}\mathbf{X} = \sum_{i=1}^{N}\mathbf{X}_i{}'\hat{\boldsymbol{\epsilon}}_i\hat{\boldsymbol{\epsilon}}_i'\mathbf{X}_i$, which captures the matrix multiplication over a block diagonal matrix. This formulation allows one to see more intuitively how asymptotics in N apply, as $\text{plim}_{N\to\infty} N^{-1}\sum_{i=1}^{N}\mathbf{X}_i{}'\hat{\boldsymbol{\epsilon}}_i\hat{\boldsymbol{\epsilon}}_i'\mathbf{X}_i = \mathbf{X}'\mathrm{Cov}\,(\boldsymbol{\epsilon})\mathbf{X}$ by the law of large numbers. The $[N/(N-1)][(NT-1)/(NT-K)]$ normalization has been added as a finite sample correction to ensure appropriate t statistics for finite N (Rogers, 1993; Hansen, 2007, 608).[15] For FE estimation, the $\mathbf{X}$ terms would have been subjected to the within transformation and the residuals would be produced from FE estimates on the transformed data. Since Arellano (1987) originally derived this robust covariance estimator (minus the finite sample correction), we refer to the standard errors produced by it as Arellano standard errors (ASEs). ASEs account for any arbitrary dependence between time-periods within independently drawn cross-sections that remains after application of the within-transformation—namely, serial correlation. Thus, using ASEs may obviate the need for including a LDV to correct for serial correlation. However, ASEs do not account for contemporaneous correlation. The intuition here is that the summation in the computation of $\mathbf{X}'\widehat{\mathrm{Cov}\,(\boldsymbol{\epsilon})}\mathbf{X}$ in eq. (4) is over i, whereas addressing contemporaneous correlation would require summation over the t dimension as well.

It is important to keep in mind that the performance of robust standard errors that compute covariance matrices in this way (i.e., computing $\mathrm{Cov}\,(\boldsymbol{\epsilon})$ in conjunction with the $\mathbf{X}$) will depend in part on whether the variables in $\mathbf{X}$ display correlation structures that mirror those of the disturbance term. For the linear model, correlation in the error terms

[15] Finite sample corrections may be employed generally to match t-statistics to limiting distributions for finite samples under assumptions of independent and identically distributed observations. Such corrections have been shown to improve small sample performance (Rogers, 1993; Hansen, 2007; Angrist and Pischke, 2009).

is propagated into variability in coefficient estimates via corresponding dependence in the regressors. If correlations in the disturbances are not mirrored by correlations in a regressor, then the correlation in the disturbances is inconsequential for the standard error for the coefficient on that regressor.[16] Cluster dependencies in the regressors can be checked in sample data, e.g. by checking the intra-class correlations of regressors. This could allow the analyst to make informed decisions about what kinds of standard error estimates are likely to be the least biased among a set of imperfect alternatives.

Note that the assumption of fixed T and increasing N is exactly antithetical to the structure of TSCS data. Nevertheless, Hansen (2007) has demonstrated that, withassumptions that are reasonable for TSCS data (i.e., N fixed and T increasing and "strong mixing" over time), the cross-section cluster covariance estimator represented in expression (4) is well-behaved (by "strong mixing," we mean that two observations rapidly approach statistical independence as the time between them increases). A key result from his analysis, however, is that the standard t-statistics should be tested against the t_{N-1} distribution.[17]

Finally, we might consider dependency in the errors that accounts for clustering within and across cross-sections, as well as across time-periods within cross-sections. In that case, $\text{Cov}(\boldsymbol{\epsilon})$ would be a block matrix: blocks along the diagonal would contain within-cross-section covariances, as in expression (4); blocks off the diagonal would be diagonal matrices containing within-time-period covariances—similar to eq. (3) but without the restriction of equal covariance for all cross-section pairs. If we allow the assumption that data grow in both N and T, a covariance matrix estimator based on the corresponding residual cross-products would be consistent (Cameron, Miller and Gelbach, 2011). For TSCS data, we typically want to assume fixed N while allowing that T may grow infinitely. Presumably, one could derive a limiting distribution for test statistics holding one or another dimension fixed under a strong-mixing assumption, but to our knowledge this has not been attempted.[18] In any

[16] Kezdi (2003) develops this point for fixed effects models, while Moulton (1986) makes this point more generally.

[17] Hansen notes that the default for Stata's "cluster" command is to use the normalized cluster robust estimator (equation (4)) and perform a t test with the appropriate degrees of freedom, a happy situation based on Rogers (1993)'s derivation of an optimal finite sample adjustment.

[18] Conley (1999) comes close by exploiting strong mixing over an arbitrary distance metric to generalize the

case, following Cameron, Miller and Gelbach (2011) we can construct this robust estimator as

$$\begin{aligned}
&\tfrac{N}{N-1}\tfrac{NT-1}{NT-K}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\left(\sum_{i=1}^{N}\mathbf{X}_i{}'\hat{\boldsymbol{\epsilon}}_i\hat{\boldsymbol{\epsilon}}_i'\mathbf{X}_i\right)\left(\mathbf{X}'\mathbf{X}\right)^{-1}\\
&+\tfrac{T}{T-1}\tfrac{NT-1}{NT-K}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\left(\sum_{t=1}^{T}\mathbf{X}_t{}'\hat{\boldsymbol{\epsilon}}_t\hat{\boldsymbol{\epsilon}}_t'\mathbf{X}_t\right)\left(\mathbf{X}'\mathbf{X}\right)^{-1}\\
&-\tfrac{NT-1}{NT-K}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{x}_{it}'\mathbf{x}_{it}\hat{\boldsymbol{\epsilon}}_{it}^2\right)\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1},
\end{aligned}$$

where $\mathbf{x}_{it}$ is a $1 \times K$ vector (that is, a row from $\mathbf{X}$), $\mathbf{X}_i$ is the matrix of regressors for unit i, and $\mathbf{X}_t$ is the matrix of regressors for period t. This formulation demonstrates how the estimator combines the cross-section and time-period cluster robust estimators (first two lines, respectively), taking care to ensure no "double counting" of the diagonal (third line). For heuristic sake, we refer to standard errors derived from this estimator as APCSEs, since it is essentially a hybrid of ASEs and PCSEs. It is not clear how demanding this estimator is in terms of asymptotics, and given that it involves more parameters than PCSEs and ASEs, it is essential to assess its performance in conditions that we encounter with actual TSCS data.

To conclude this section, we note that the validity of the robust standard error estimators discussed here depends on the consistency of the residual estimates, which in turn depends on the consistency of the coefficient estimates. Thus, concerns about the consistency of point estimates discussed in Section 3.2 dovetail with concerns about the performance of cluster robust estimators. Since PCSEs do not address serial correlation, additional steps must be taken to produce reliable inferences with them. Beck and Katz (1996) conducted simulations which showed that the lag correction with OLS generally outperforms an FGLS estimator that uses the Prais-Winsten transformation, and therefore recommend that researchers begin with the former in TSCS data. Introducing a lag in this way, can be problematic, however, because if serial correlation is not adequately addressed by the lag, the resulting dependence between the LDV and error term may render OLS inconsistent. Beck (2001, 279–280) advocates a Lagrange multiplier test to make sure that including the lag (or additional lags) adequately addresses temporal dependence. Interestingly, such a test can indicate the presence of unit effects, since they are form of temporal dependence that can show up in the

proposals of Newey and West (1987) and develop spatial heteroskedasticity and autocorrelation consistent estimators.

errors if they are not modeled. Although ASEs and APCSEs address serial correlation without the need for including an LDV, if the LDV is in the DGP (i.e., γ in eq. (1)), Nickell bias may propogate to the computation of these robust estimators and adversely affect their performance. Our Monte Carlo analysis, which we turn to now, enables us to assess concerns about point estimates as well as standard error estimates.

4 Comparison of Alternative Estimators

We conducted Monte Carlo experiments to investigate the issues developed in section 3. The details of the design of the experiments are discussed in the appendix. In reporting results, we devote attention to the following:

1. We examine the performance the estimators using sample sizes that resemble what researchers encounter in applied work. These include samples with rather small N or T.

2. We study how standard error estimates are affected by the variance of unit-effects or omitted variable bias due to unit effects that are correlated with regressors.

3. We assess the extent to which Nickell bias affects FE estimates of β when a LDV is included to handle potential dynamics—both when the LDV is including for substantive reasons as well as when it is to address the nuisance of serial correlation. Because the size of T is likely to be important here, we examine various values of T.

4.1 Measures of performance

We are primarily interested in the performance of different approaches for estimating β and its standard error. To gauge the accuracy of the point estimates for β, we computed the bias and root mean squared error (RMSE) of the estimated βs across simulations. We compute the bias as,

$$\text{Bias} \equiv L^{-1} \sum_{l=1}^{L} (\hat{\beta}_l - \beta),$$

where l refers to a given simulation and L refers to the total number of simulations. To assess the performance of the standard errors, we computed the measure of "over/under confidence" advocated by Beck and Katz:

$$\text{Confidence} \equiv 100 \times \frac{\sum_{l=1}^{L} \left(\hat{\beta}_l - \bar{\hat{\beta}}\right)^2}{\sum_{l=1}^{L} \left[\text{SE}\left(\hat{\beta}_l\right)\right]^2}, \tag{5}$$

where $\hat{\beta}_l$ denotes estimates, $\bar{\hat{\beta}}$ denotes the mean of the estimates across all simulations, and $\text{SE}\left(\hat{\beta}_l\right)$ denotes the estimated standard error for $\hat{\beta}_l$. Values above 100 indicate that true sampling variability is greater than the reported estimate of that variability, while values less than 100 indicate that the estimate overstates true variability. Understating variability (being overconfident) means that we might reject the null of a zero coefficient when the null is true (i.e., commit a Type I error). Overstating variability (being underconfident) implies that we might not reject the null of a zero coefficient when that null is false, leading us to conclude that a variable does not have effects when in fact it does (i.e., increases the risk of Type II errors). While confidence provides a measure of the mean performance of a standard error estimator, we are sometimes also interested in the stability or volatility of a standard error estimator. Such volatility affects the reliability of tests based on the standard error. As a measure of volatility, we report in the text below coefficients of variation, which are computed by taking the standard deviation of the standard error estimates over simulation runs and dividing by the mean of the estimates. A standard error estimator with accurate confidence but a high coefficient of variation would provide less reliable tests or confidence intervals than an estimator with accurate confidence and a lower coefficient of variation. As per convention, we express the coefficient of variation as a percentage. In each simulation, we also performed a Lagrange Multiplier (LM) test to determine the existence of serial correlation, even after the inclusion of the LDV to remove it.

Our core simulation results are presented in tables in this paper, with extended results available in an online supplement.[19] All experiments were based on 500 simulation runs.

[19]We acknowledge that the tables that we report contain a great deal of information and can be taxing on the eye. Yet this gives readers the best opportunity to evaluate the performance of the estimators and make the optimal choice for their particular research situations. For those who are more graphically inclined, the

All estimates in the table include the LDV in the specification, except for those demarcated with an asterisk (*). The information reported in the tables is organized is a straightforward way. The tables are set up so that horizontal blocks pertain to progressively higher degrees of serial correlation, captured by values of 0, .3, .6, and .9 for ρ_ϵ, the parameter in an AR(1) process (see the column marked with "ρ_ϵ").[20] Within each of these blocks, we cycle through three values each for contemporaneous correlation (0, .25, .5) and panel heteroskedasticity (0, .3, .5), indicated in columns marked "Cor." and "Het.", respectively.[21]

4.2 Results for finite samples of varying dimensions

We begin with experiments where $N = 16$ and $T = 20$, since this sample size is broadly representative of what political science research using TSCS data typically encounters. This is a rather small number of units and a moderate number of time periods. We will discuss how changes in N or T affect conclusions. In order to isolate the effects of sample size, we keep the DGP simple, and so the unit effects have rather small variance ($\sigma_\alpha = 5$, while $\beta = 10$) and there is no confounding due to either the unit effects or the LDV (i.e., x_{it} is uncorrelated with α_i and the LDV is not part of the DGP).[22] That is, the DGP is a special case of (1):

$$y_{it} = \beta x_{it} + \alpha_i + \epsilon_{it}, \tag{6}$$

where $\epsilon_{it} = \rho_\epsilon \epsilon_{i,t-1} + \varepsilon_{it}$ (i.e., the errors follow an AR(1) process). Importantly, the contemporaneous correlation and panel heteroskedasticity in the errors is mirrored by such correlation and heteroskedasticity in x_{it} (except where noted), although serial correlation in the errors

web supplement to this paper will include figures that display relevant results.

[20]This is the same ρ_ϵ from the Monte Carlo DGP described in eq. (11) in the appendix.

[21]"Cor." is the value of the $\rho_{cc,u}$ parameter in the Monte Carlo DGP described in the appendix. "Het." is equal to the heteroskedasticity measure defined by Beck and Katz (1995), which equals "the standard deviation of the normalized weights that would be used in panel-weighted least squares" (fn. 21). It is a concave, monotonic transformation of v_{u2}. A Het. coefficient of 0 corresponds to $v_{u2} = 0$; 0.3 corresponds to $v_{u2} = 3.45$; and 0.5 corresponds to $v_{u2} = 10$.

[22]This choice for σ_α produced model fits such that the evidence in favor of models that accounted for unit effects relative to models that did not was "weak" to "positive", according to Raftery's (1995) criteria for BIC comparisons.

is not mirrored by serial correlation in x_{it}. This follows Beck and Katz (1995), allowing for comparison.

The initial set of experiments is reported in Table 1. The first thing that stands out is the column labeled "% reject $\rho_\epsilon = 0$", which reports the percentage of times that the LM test rejects the null of no serial correlation. Under these experimental conditions, we always would conclude that serial correlation is present, even though we have included a LDV to address it, because the unit effects are relegated to the disturbance term. Even when there is no serial correlation in the disturbances (i.e., $\rho_\epsilon = 0$) in the DGP, relegating the unit effects to the disturbance in the estimation model tends to trip the alarm of the serial correlation test.[23] This suggests we should not proceed to make inferences without doing more to address the problem. Yet, Wilson and Butler's (2007) finding that studies often fail to check for serial correlation suggests many researchers might proceed to make inferences at this point.

Interestingly, doing so does not appear to be problematic under the conditions of this experiment. The bias in OLS point estimates is minimal, as expected since there is no confounding. The PCSEs from OLS with the LDV perform about as well or better than the other standard error estimators that we investigated, hewing close to the value of 100 across the various experimental conditions.[24] The very slight overconservatism (evident in confidence values below 100) is due to the PCSEs ignoring the slight negative serial correlation that the LDV induces in the error (refer to fn. 12). This issue disappears when the lag $x_{i,t-1}$ is included in the model (refer to fn. 13), as shown in Table A-1 in the Web appendix . The RMSE of OLS point estimates is somewhat higher than that for the FE estimates of β, and there is essentially no differnece between the bias and RMSE values for FE with and without the LDV. With non-zero values of contemporaneous correlation, ASEs perform very poorly as they ignore this dependence across units.[25] FE-PCSEs and

[23]It is important to keep in mind that with observational data, we will not know what the nature of the disturbance is and can only try to divine it through appropriate diagnostic tests.

[24]Note that we stay true to Beck and Katz's original derivation and do not add a finite sample correction to the PCSEs. This results in some overconfidence in smaller samples.

[25]In experiments not reported here, we investigated the performance of standard fixed effects standard errors (FESEs) and an estimator derived by Arellano (1989) inspired by Kiefer (1980). Both of these

APCSEs are on balance a bit more over-confident than the PCSEs, but only slightly so.[26] Indeed, despite the rather small N and moderate T, it is striking just how well the APCSEs perform. There is only a slight cost in terms of the volatility of estimates: the coefficient of variation for the APCSEs was about 18% across these simulations whereas for PCSEs with the LDV (and no FE) specification, it was about 16%. The fact that the difference is slight is due in part to the efficiency gains from FE estimation. The coefficient of variation for the more restrictive FE-PCSE estimates ranged from 6% to 17%, with higher values for the scenarios with large amounts of contemporaneous correlation and panel heteroskedasticity. Thus, among FE standard error estimators, one trades off robustness with estimator stability in this setting.

Generally, OLS with PCSEs, FE with PCSEs (both with an LDV to correct for serial correlation), and FE with APCSEs (no LDV necessary) perform about equally here. If researchers are uncomfortable with adding an LDV to their model, there really is no reason to under these conditions, since FE with APCSEs appear to be quite reliable. As we report in our web appendix, the performance of ASEs improves to the levels of the other FE robust estimators if no contemporaneous correlation exists in the x_{it} (see Table A-3). This pattern held throughout the range of experiments that we performed and confirms that researchers should test for correlation patterns in their explanatory variables when making the choice of which robust standard error estimator to use.

The importance of the size of T can be seen if we decrease the number of periods to 5, as shown in Table 2.[27] OLS point estimates are noisier than with larger T, and thus RMSE for β is worse, as is the case for FE with and without the LDV. The PCSEs hold up reasonably well despite the small number of periods. Again, we see overconservatism when there is no contemporaneous correlation or panel heteroskedasticity due to PCSEs ignoring the negative serial correlation induced by the LDV. As contemporaneous correlation and panel heteroskedasticity increase, the PCSEs move toward overconfidence. This bias reflects

---

performed so poorly that we do not include them in the discussion here. However, we do note that Wilson and Butler's (2007) survey of work on TSCS data suggests that practitioners regularly use FESEs.

[26]Note that FE-PCSEs perform worse if no LDV is included to address serial correlation in the errors.

[27]For samples of publications that conduct TSCS analysis with T in this range, see Iversen and Soskice (2006), Meguid (2005), and Treisman (2000).

the fact that the increasing dependence reduces the effective sample size. The nature of this effect is made more clear when we include the lagged $x_{i,t-1}$ (in Table A-2), which removes the overconservatism result and lays bare that it is due to increasing dependence and variability across units. Interestingly, the performance of FE-PCSEs degrades much more substantially with $T = 5$, in contrast to their generally good performance with $T = 20$.[28] ASEs continue to perform badly when contemporaneous correlation is present. APCSEs are, on average, accurate only at low levels of contemporaneous correlation, and become more unreliable as such correlation grows, again reflecting the loss in effective sample size.[29] Still, FE always has lower RMSE than OLS, so researchers should consider the this tradeoff. Estimation volatility, measured with the coefficients of variation, was high for all estimators, meaning that testing would be unreliable. The coefficient of variation for PCSEs was around 22% across simulations, FE-PCSEs between 23% and 42%, ASEs between 23% and 35%, and APCSEs being 24% and 40%, with higher values again for the scenarios with large amounts of contemporaneous correlation and panel heteroskedasticity. Thus, when the sample is shrunk to $T = 5$ and with only $N = 16$ units, inference becomes challenging, especially for FE estimators and when there exists substantial contemporaneous correlation. The more restrictive OLS with LDV model with PCSEs was most reliable here, although OLS displays more bias and higher RMSE that FE. Also, consistency of OLS depends crucially on the fact that the unit effect is not correlated with $x_{i,t}$—an issue that we return to later.

Increasing T beyond 20 brings the performance of the estimators more closely in line. As can be seen in Table 3, for example, when we double the size of T to 40, OLS, FE, and FE* are essentially equivalent in terms of bias and RMSE. PCSEs, FE-PCSEs, and APCSEs perform equally, although ASEs still do very poorly as contemporaneous correlation increases. For example, with $T = 40$ and $N = 16$, we see the same basic patterns with respect to confidence as in the $T = 20$ case. Coefficients of variation for PCSEs, FE-PCSEs, ASEs, and APCSEs

[28]A finite sample adjustment for the standard error along the lines of those discussed in the previous previous section would be fixed for all FE-PCSE estimates in these simulations to be $\sqrt{\frac{5}{4}\frac{79}{78}} = 1.125$. This would account for less than half of the over-confidence that we observe.

[29]In a few instances among the 500 simulation runs, APCSEs failed to produce valid estimates due to negative variance values.

were 12% to 14%, 4% to 11%, 19% to 24%, and 11% to 18%, respectively.

Changing the size of N while holding T fixed, however, does result in variation in performance across the esimators in question. Table 4 reports results when conditions are the same as for the results reported in Table 1, except that N is increased to 100.[30] The gap between RMSE for OLS and RMSE for FE is much smaller overall, as we would expect with more data. We again see that PCSEs tend to overstate variability, but now this has increased to as much as 30 percentage points when there is no contemporaneous correlation and/or heteroskedasticity in the errors.[31] Using either FE-PCSEs (or introducing the lag $x_{i,t-1}$) to the specification resolves this particular problem. The large N greatly magnifies the overconfidence of the ASEs when contemporaneous correlation is present. The APCSEs performed well in this setting—about the same or slightly better than FE-PCSEs. They produce confidence values very close to 100, and the coefficients of variation were between 8% to 17%. Coefficients of variation of PCSEs, FE-PCSEs, and ASEs were 8% to 9%, 4% to 17%, and 8% to 16%, respectively—slightly less volatile, but at the cost of potential bias.

To summarize the results to this point, we find that with at least moderately sized samples and under the conditions explored with the above sets of experiments (most notably, no correlation between unit effects and explanatory variables), FE with PCSEs and APCSEs perform well in terms of removing bias due to unit effects, producing reliable standard errors, and improving estimation efficiency. With T in the range of 20, the bias induced by including LDVs is tolerable, but the results for APCSEs suggest that an LDV is not necessary to address serial correlation in the errors. In situations with very small sample dimensions (e.g., $N = 16, T = 5$), researchers appear to be better off using PCSEs without FE than with FE. While APCSEs tended to be more confident than PCSEs with the highest values of contemporaneous correlation that we looked at, the point estimates associated with APCSEs had smaller RMSE than the OLS estimates. Researchers need to contemplate this tradeoff when deciding among estimators and be appropriately humble about what exactly

[30]In the past decade, numerous articles have appeared in top journals that analyze samples with dimensions equal to or approximately equal to this size.

[31]PCSEs overstate variability even more if contemporaneous correlation does note exist in the explanatory variables.

can be learned from data of these limited dimensions.

4.3 Results for unit effect variance and correlation with x_{it}

For the experiments reported in Table 5, we increased the influence of the unit effects such that $\sigma_\alpha = 20$.[32] OLS becomes substantially noisier, given that the now-larger unit effect is relegated to the error term, and comparison of RMSE reveals that FE provides clearer gains in precision. The PCSEs ignore correlation induced by the unit effect. We might expect this to be less of an issue because there is no corresponding serial correlation in x_{it}. However, as discussed above, negative serial correlation in the errors induced by the LDV causes the PCSEs to be underconfident, with coefficient variability now overstated by between 3 and 23 percentage points. The most underconfidence comes when there is little to no contemporaneous correlation or heteroskedasticity in the disturbances.[33] FE-PCSEs and APCSEs perform about the same as with the $\sigma_\alpha = 5$ case, only slightly understating variability. The only situation where PCSEs do better than the these robust estimators is at the highest values for contemporaneous correlation, panel heteroskedasticity, and serial correlation examined in our experiments. This is particular to the Monte Carlo set-up, as the lack of serial correlation in x_{it} means that the residual serial correlation in ϵ_{it} does not undermine the performance of FE-PCSEs.[34] As before, the ASEs suffer tremendously by ignoring contemporaneous correlation.

We expect bias in OLS point estimates to increase in both the correlation between α_i and $x_{i,t}$ and the unit effect variance, although it is unclear to what extent this would affect the performance of standard errors that rely on such estimates. Tables 6 and 7 report results with the correlation between the unit effect and the explanatory variable set at approximately .3 (labeled as $\rho_{\alpha,x}$ in the table and in the DGP description above). The bias in β_{OLS} follows our expectations. Yet this variation in the severity of omitted variable bias

[32] For this value of σ_α, comparisons of BIC values indicated that evidence in favor of models that accounted for unit effects relative to models that did not was "positive" to "strong."

[33] As before, this problem disappears when the lagged regressor $x_{i,t-1}$ is included in the specification.

[34] If x_{it} displayed the same kind of serial correlation as ϵ_{it}, we would expect the performence of FE-PCSEs to decline.

does not appreciably affect the relative performance of the standard error estimators.[35] The general patterns noted above still appear where FE-PCSEs and APCSEs perform better than PCSEs (in absolute terms) except at the highest levels of contemporaneous correlation and panel heteroskedasticity. The FE approach, though, clearly dominates, when bias and RMSE are taken into account.

This is especially the case when N and the size of the unit effects increase. The decline in the performance of PCSEs that we saw in Tables 4 and 5 due to increases in the size of the unit effects and in sample size on the cross-sectional dimension is greatly exacerbated by correlation between $x_{i,t}$ and α_i. As reported in Table 8, PCSEs are underconfident by over 50 percentage points when there is no contemporaneous correlaton in the errors. This improves as such correlation increases, but FE-PCSEs and APCSEs clearly dominate here across the range of experimental conditions, with the latter doing slightly better. The poor performance of PCSEs, combined with the bias and higher RMSE associated with OLS point estimates, suggest that researchers should be wary about using them with Ns of this size or greater if there is positive to strong evidence of the presence of unit effects that are correlated with explanatory variables.[36]

4.4 Problems when the DGP includes a LDV and FE

In the simulations discussed thus far, the only correlation between $y_{i,t-1}$ and x_{it} is due to their both being correlated with α_i. Because FE purges the data of variation associated with α_i, this dependence between $y_{i,t-1}$ and x_{it} is broken. Therefore, there is no risk that Nickell bias on the coefficient for $y_{i,t-1}$ induces bias on the estimates of β. This is confirmed by our main simulation results, as there are no discernible differences for FE estimates when $y_{i,t-1}$ is included in the specification across the main simulations, even with small T.

We now look into the case where $y_{i,t-1}$ must be included in the specification in order to

[35]This general result holds for the various sample dimensions as well.

[36]In results available from our web supplement, when $x_{i,t-1}$ is included in the specification, the magnitude of the omitted variable bias decreases because of the pseudo-first-differencing that this specification implies (refer to footnote 13). However, as discussed above, there are no advantages to this strategy over FE. Indeed, FE is more efficient, as indicated by smaller RMSE.

remove omitted variable bias in estimating β. Such is the case when correlation between $y_{i,t-1}$ and x_{it} exists even after purging variation due to α_i. For example, it might be that the x_{it} is a function of $y_{i,t-1}$. We examined this scenario in a set of auxiliary simulations. We used the same basic settings as in the "central" experiments described above, except to make patterns associated with bias pronounced, we set $\sigma_\alpha = 20$, the coefficient on the LDV $\gamma = 0.8$, the correlation between the unit effect and regressor $\rho_{\alpha,x} = 0.8$, and the correlation between the LDV and x_{it} to 0.67. We varied T from 2 to 20 to explore the rate at which Nickell bias disappeared. The graphs in the left column of Figure 1 displays the results from these simulations. We omit results for the OLS estimates that exclude both the LDV and unit effects, as they were horribly biased (ADL in Figure 1 refers to "autoregressive distributed lag"). Rather, we show results for five specifications based on various combinations of the LDV and FE. The convergence toward unbiased estimates in the two estimates at the top demonstrate the rate of decay in the Nickell bias. The key lesson from these simulations is that for $T > 4$, the Nickell bias that propagates to β is of much less concern than bias resulting from omitting either the LDV or FE. Such specification bias is show in the bottom three graphs. We take this to be a general lesson, and propose that for moderately deep TSCS datasets, analysts not worry about issues associated with Nickell bias. When the data are at least moderately deep in the time dimension—in our case $T \geq 8$—the possibility of omitted variable bias associated with the LDV or FE should be the primary concern.

The right column adds an additional wrinkle to the estimation problem: in addition to the confounding due to the FE and LDV, the error term is generated with AR(1) serial correlation equal to $\rho_\epsilon = 0.5$. In this situation, there exists no FE or LDV conditioning strategy that allows for unbiased estimation of β. The fact the FE/LDV and FE/ADL estimates hover around the unbiasedness line is an artifact of the particular correlation values settings—it is possible to come up with values such that the estimates steadily drift away. At the same time, the fact that the bias is generally smaller than the bias due to omitting either the LDV or FE is general. This demonstrates again that when the data are moderately deep in the time dimension, issues associated with Nickell biases need not be a concern. With respect to standard errors, the results in section 4.3 demonstrated that omitted variable biases do not change the conclusions that we draw in section 4.2 on the relative performance of the

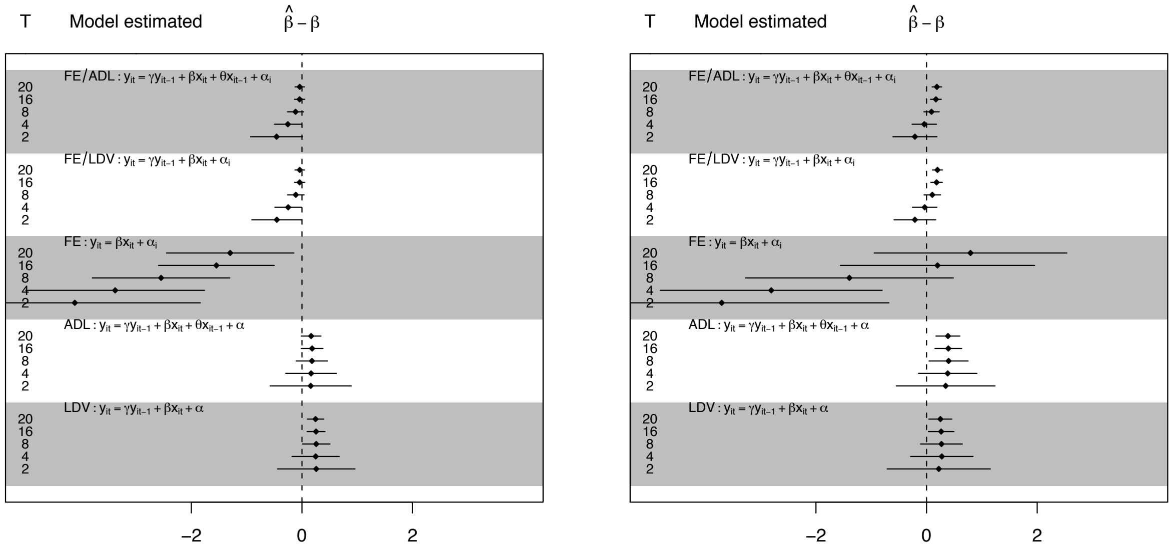

Figure 1: LDV and FE specifications and Nickell bias.

estimators. The same logic applies for the FE-LDV simulations, with the results on standard errors reported in 4.2 holding up in these new simulations as well.

5 Time invariant regressors

In the previous section, we demonstrated that in certain situations—situations which we believe are quite common—using a FE estimator with cluster robust standard errors is preferable to trying to use lags to handle unit heterogeneity and serial correlation. Lags may be included, though, if there is a presumed need to account for dynamics in the DGP, with little concern for Nickell bias for data that is at least moderately deep in the time dimension. Thus far, we have sidestepped the issue of what to do if a model contains substantively important variables that vary little, if at all, over time. Yet, time invariant or slow moving variables are quite common in TSCS data. Cross-national data sets often have country-level variables, such as institutional measures, that are time invariant or near-invariant.

Using FE for TSCS data then often involves a trade-off as coefficients on time invariant variables are not identified when the standard FE estimator is employed.[37] FE uses only within-group variation for identification, whereas OLS uses both within- and between-group variation. In some cases, researchers choose to use OLS over FE because of their interest in making inferences about time invariant variables.[38] The argument that including a time invariant variable that we care about substantively instead of a generic time invariant unit effect of vague or ambiguous meaning is intuitively appealing. Technically, however, adopting this approach can be highly problematic. In essence, if the true specification contains a unit effect instead of the time invariant variable, then we are estimating a model with a disturbance term that includes components that are not accounted for by the standard least squares estimator. While it is well known that this can lead to bias in point estimates, we show here that even under conditions where there is no such bias, the least squares estimator for the variance-covariance matrix—as well as PCSEs—produces misleading standard errors. In particular, we show that ignoring unit effects for the sake of estimating coefficients on time invariant variables can lead to the incorrect conclusion that these coefficients are statistically significant when they are not.

Suppose the DGP follows eq. 6—i.e., γ, θ, and λ all equal zero and (for now) we assume $\epsilon_{i,t}$ is spherical and uncorrelated with α_i.

If we ignore the unit effects, then we are estimating the above model with the disturbance $v_{i,t} = \alpha_i + \epsilon_{i,t}$. If α_i is correlated with $x_{i,t}$, then this leads to bias and inconsistency in OLS estimates of *beta*. But even if there is no such correlation, there can still be problems with inferences. The problem, again, is that relegating α_i to the disturbance term in essence induces correlation in the errors—an issue that became apparent in the Monte Carlo experiments discussed in Section 4. In this case, we can rewrite the covariance matrix in eq. 2 as

$$\text{Cov}\,[\mathbf{b}] = E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{v}\mathbf{v}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}], \tag{7}$$

[37]Unbiased coefficient estimates can be recovered for time invariant variables using the FE estimator, but only if time invariant variables are available that are uncorrelated with the unit-effects (Hausman and Taylor, 1981).

[38]For recent published papers where authors explicitly chose not to use FE because of collinearity with time invariant or slow-moving variables, see Gerring, Thacker and Moreno (2005), Long and Leeds (2006), Hecock (2006), and Mosley and Uno (2007).

where $\mathbf{X}$ and $\mathbf{v}$ represent the $x_{i,t}$ and $v_{i,t}$ stacked over all i and t. Let $\mathbf{\Sigma} = E[\mathbf{X}'\mathbf{v}\mathbf{v}'\mathbf{X}]$. For the case of repeated cross-sections, this can be rewritten as

$$\mathbf{\Sigma} = E\left[\sum_i \sum_j \sum_t \sum_s \mathbf{x}_{it} v_{it} v_{js} \mathbf{x}'_{js}\right] \tag{8}$$

If $v_{i,t} = \alpha_i + \epsilon_{i,t}$, then eq. 8 can be rewritten as

$$\mathbf{\Sigma} = E\left[\sum_i \sum_t \mathbf{x}_{it}\mathbf{x}'_{it} v_{it}^2\right] + 2E\left[\sum_i \sum_{t>s} \mathbf{x}_{it}\mathbf{x}'_{is}\alpha_i^2\right]. \tag{9}$$

Even if the ϵ_{it} are spherical, the standard OLS estimator for the variance-covariance matrix will be wrong, since the second term in eq. 9 will be ignored if the α_i are not accounted for in the model specification. PCSEs will also ignore this term, since they do not account for correlation across time periods.

This presents a problem particularly for time invariant variables because the standard errors for the coefficients on such variables will generally be too small, possibly leading to type I errors.[39] To see this, consider the second term of eq. 9 when $\mathbf{x}_{i,t} = (x_{i,t}, z_i)'$:

$$2E\left[\sum_i \sum_{t>s} \begin{bmatrix} x_{i,t}x_{i,s} & x_{i,t}z_i \\ x_{i,s}z_i & z_i^2 \end{bmatrix} \alpha_i^2\right] \tag{10}$$

It is possible that the xs are uncorrelated across cross-sectional units and time periods and are also uncorrelated with the zs, which would give zeros in the first diagonal element and the off-diagonals and hence will not contribute anything to the standard errors in expectation. However, the term z_i^2, which pertains to the standard error for the time invariant variable, is guaranteed to be positive (assuming the zs have positive variance), and it will remain so after doing the summation, multiplying by α_i^2, and taking its expectation. OLS and other estimators that are not robust to longitudinal correlation in the errors (like PCSEs) ignore this term, and therefore produce standard error estimates that are too small.[40] Thus,

[39]The results here essentially generalize Moulton (1986)'s results for random effects to the fixed effects context.

[40]The degree to which the standard errors are understated is proportional to the variance of the zs and the square of their expectation. A similar downward bias will be seen with slow moving variables, with the degree of bias being proportional to the covariance between the zs across time periods and the product of the means of the zs across periods.

including a time invariant variable in a regression instead of estimating FE could lead to the inference that the variable has a statistically significant effect when in reality it does not.

To confirm this analytical result and get a sense of the magnitude of the problem, we conducted Monte Carlo analysis where we generated the data with an explanatory variable, $x_{i,t}$, and a unit effect, but then estimated a model that replaces the unit effect with a randomly generated z_i, which is independent of $x_{i,t}$ and the unit effect. This set-up mimics a scenario where a researcher forsakes the FE approach in order to include a time invariant variable, even though—unbeknown to the researcher—the time invariant variable actually has no relationship with the dependent variable.

We estimated the model via standard OLS as well as with PCSEs. To verify that the problem is a kind of serial correlation induced by omission of the unit effects, we computed cluster robust standard errors along the lines of expression (4) above, but using OLS residuals instead of FE residuals. We also compute the between group estimator, which can also provide a useful check on the impact of the time-invariant variable while accounting for unit effects (and other time varying explanatory variables). We assess performance by looking at bias in terms of the mean of the point estimates across simulations, as well as look at the proportion of samples that produce estimates of λ (the coefficient on z_i) that are bounded away from zero (i.e., the coverage of the confidence intervals). In other words, we calculated the proportion of samples where we would draw the inference that the time invariant variable has a statistically significant effect (at the .05 level) when in reality it has no effect on the dependent variable.

Table 9 reports the results. The column marked "No lag" indicates that, while on average the OLS point estimates of λ are zero, in about 70% of the simulations we would incorrectly conclude that the time invariant variable has a statistically significant coefficient. PCSEs do even worse, producing statistically significant λs in over 90% of the simulations. Both the robust standard error estimator from eq. 5 and the standard error from the between estimator do much better, producing very few significant λs. The one silver lining here is that if the Lagrange Multiplier test for serial correlation is conducted, it will almost always indicate that serial correlation is present and something must be done to correct for it. These results again indicate how crucial it is to test for serial correlation and reemphasize how troubling it is that Wilson and Butler (2007) find that a large proportion of studies that use TSCS data did not check for serial correlation. The between group estimator produces statistically significant coefficients in only about 6% of the simulations, which is in line with what we would expect when computing 95% confidence intervals. Thus, these results suggest that a useful check on whether we are drawing incorrect conclusions about the effect of a time invariant variable is to compute the between group estimator to see if it produces results that are different from estimators that use more than between variation for identification. Computing standard errors from the variance-covariance matrix given in eq. 5 also gives us clues that something is amiss with a time invariant variable. It produces estimates of λ that are bounded away from zero in 10% of the simulations, which is a little high, but far better than the other standard error estimators we examined.

We checked to see whether the results changed if lags were included in the specification, since including lags would be a logical next step if serial correlation was found. The columns marked "One lag" and "Two lags" in Table 9 report the results. When one lag is included, the performance of OLS and PCSEs improves some, but we still would conclude λ was statistically significant, respectively, in 65 and 80 percent of the simulations. Fortunately, we still reject the null of no serial correlation in every sample. Thus, it is important that researchers re-test for serial correlation after including a lag to make sure the problem has been addressed. However, including two lags to address serial correlation does not promise to make the situation much better. With two lags the performance of the standard errors for OLS and PCSEs remains abysmal, but the test for serial correlation indicates that it is still present. If we did not check for remaining serial correlation after including an additional lag and proceeded to conduct inferences, we would be likely to conclude—incorrectly—that the time invariant variable is an important predictor of the dependent variable.

These experiments confirm the analytical result that choosing OLS—or OLS with PCSEs—instead of FE can lead to seriously misleading inferences on time invariant variables. But they also suggest one way that researchers can avoid these pitfalls. The standard F test for unit effects cannot be performed when a time invariant variable is included in the specification, since that variable will be perfectly collinear with unit-specific dummies. Yet if we estimate the model both by OLS with PCSEs and the between group estimator and find a statistically significant coefficient on the time invariant variable in the former but not the latter, this raises a red flag that we need to be especially cautious about drawing conclusions on the effects of this variable. Computing the robust standard error estimator from eq. 5 also provides some help in determining whether a time invariant variable belongs in a specification.

The experiments just discussed set aside the issue of whether the disturbance term $\epsilon_{i,t}$ was itself non-spherical. But if serial correlation exists and we use a LDV to correct for it without accounting for unit effects, additional inferential problems arise for time-invariant variables. As we stated in section 3.2, the bias that results from the correlation between a LDV and a unit effect that is relegated to the disturbance term can lead to bias in coefficients on independent variables. The extent of the bias depends on the degree of correlation between the independent variables and the dependent variable, and, by extension, the LDV. When unit effects account for a relatively small part of the variation in the dependent variable, we see an acceptably small amount of bias in the coefficients on time-varying explanatory variables. However, we are likely to see a greater amount of bias with time-invariant variables because they will be more highly correlated with the LDV than are time-varying variables. The intuition here is that—while a time-varying variable in a given period may have little relation to its value in the previous period—a time-invariant variable has exactly the same value in every period, and is thus perfectly correlated with itself as a component of the LDV. This correlation can lead to serious bias in coefficients on time-invariant variables.[41]

To illustrate this point, we extended our experiments by including a time-invariant variable (z_i) in addition to a time-varying variable ($x_{i,t}$) in the DGP. This variable was generated such that its correlation with both the unit-specific effect α_i and the time varying explanatory variable was essentially zero—the best case scenario for finding good performance for the pooled approach. The remaining aspects of the simulations follow those discussed in section 4.

Table 10 reports the results from these experiments for $N = 16$ and $T = 20$. We report the average correlations between the LDV, the unit effect, and both the time-varying and

[41]For slow-moving variables, the degree of bias would be a function of of just how slowly the variables change over time, with the limit being no change at all.

time-invariant variables to demonstrate that any bias that we see is due to the relationship between the LDV, the unit effect, and the time-invariant variable. The column labeled $\rho_{\alpha,z}$ indicates that the average correlation between the unit-specific effects and the time-invariant variable is essentially zero, while the columns labeled $\rho_{\alpha,y_{t-1}}$ and $\rho_{y_{t-1},z}$ indicate the degree of correlation between the unit effect and the LDV and the LDV and the time-invariant variable, respectively. The medium to high correlation between the LDV and the time-invariant variable that we see results in the poor performance of OLS in estimating the effects of the time-invariant variable. OLS underestimates the coefficient across the range of experimental conditions. The serial correlation induced by ignoring the unit effects is seriously detrimental for the performance of PCSEs, which drastically understate coefficient variability. If we rely on the test for serial correlation (after including a lag) to decide whether it is appropriate to use OLS with PCSEs to make inferences on the time-invariant variable, we would almost certainly conclude that a problem existed that requires more attention before we proceeded to make inferences. If we did not properly test for serial correlation or did nothing more to correct for it and used OLS with PCSEs, we risk making grossly inaccurate inferences, even under the best case scenario of orthogonality between the unit effects and the independent variables.

While we are sympathetic to researchers who desire to make substantive inferences on time-invariant variables in TSCS data, these experiments show that using OLS with PCSEs and ignoring unit effects is potentially highly problematic. If a LDV is included in the specification and unit effects are present in the data, ignoring them for the sake of including a time invariant variable in the specification appears to buy us little in terms of substantive inferences, since the results that we obtain from pooling the data can be remarkably misleading. If researchers are primarily concerned with the effects of time-invariant variables, a better option might be to use the between estimator, and rely on between-group rather than within-group variation for identification.[42]

[42]Plümper and Troeger (2007) advocate the use of a method they call *xtfevd* to estimate models with time invariant variables and unit effects on TSCS data. The logic and performance of the method has been called into serious question recently by Breusch, Ward, Nguyen and Kompas (2011*b*,*a*) and Greene (2011*a*,*b*). We conducted our own Monte Carlo experiments with the conditions investigated throughout this paper and

In addition to the statistical concerns demonstrated in these simulations, we wish to highlight other concerns that aries when including TINVs in TSCS models. First, suppose the goal is to estimate parameters in expression (1) in order to forecast outcomes for some units that, for whatever reason, were not part of the original dataset. Such a goal violates the presumptions on which *fixed* effects models were originally developed, and prompts the need for a random effects or hierarchical modeling approach that imposes constraints on the unit effects to establish a distribution for out of sample predictions. Second, suppose that what we really care about is the effect of z_i, and we feel the need to "control" for fluctuations of the sort measured by x_{it}. The rationale for doing so may be questioned. If the x_{it} values are causal descendants of the z_i's, then "controlling" for them creates post-treatment bias for the coefficient on z_i. It is difficult to imagine a situation in which a time-varying variable could be causally prior to a time invariant variable, in which case it is difficult to imagine a situation where we would need x_{it} to reduce bias in the estimate of λ. All of these issues associated with estimating coefficients on TINVs along with TVCs are pertinent even when using methods such as the "fixed effects vector decomposition" method of Plümper and Troeger (2007).

6 Discussion

This paper has investigated the performance of cluster robust FE estimators relative to other estimators under the kinds of conditions that researchers are likely to encounter when analyzing TSCS data. A key focus has been how to deal with dynamics and the presence of unobserved fixed effects. We have demonstrated that in some cases where theory raises concers about the performance of certain estimators, those estimators appear to perform adequately, given the limited amount of data that we have to work with. Although theory tells us that bias in OLS and FE point estimates results when unit effects are present in the data and a lag is introduced into the specification, in data sets with T above and N below particular thresholds, the bias is acceptably low and robust standard errors (which

our results are consistent with the concerns raised by Breusch et al. and Greene. Given space constraints, we do not report these results here but they are available upon request.

are computed using possibly biased estimates) are not too adversely affected. $T > 10$ appears to be a good rule of thumb when the LDV is merely a control variable; $T > 20$ is reasonable when the coefficient on the LDV is of primary interest and the disturbance is free from serial correlation.[43] Neverthless, one of the key take away points of this paper is that researchers who are rightly concerned about how introducing a LDV into a specificiation might dramatically change the nature of a model can simply leave it out and use APCSEs to address serial correlation.

The pooling approach that employs OLS with a LDV and PCSEs proved to be reasonably robust in the presence of serial correlation and unit effects under a limited set of conditions, although the performance of this approach declines—not surprisingly—as the correlation between the unit effects and explanatory variables increases, the influence of the unit effects increases, the time dimension of the sample decreases or the cross-sectional dimension of the sample increases. We would recommend against the pooled approach for Ns greater than 100, if BIC comparisons indicate more than a weak preference for fixed effects, and if there is evidence that unit effects are correlated with explanatory variables at .3 or higher. The FE approach which uses FE point estimates and residuals to compute PCSEs generally dominates the pooling approach in terms of bias, RMSE, and standard error estimates, except when T is very small—around 5 periods.[44] With T of 20 or greater, FE point estimates appear to suffer from little bias from the inclusion of a LDV and what bias exists does not really propogate to FE-PCSEs. Nevertheless, since APCSEs performed as well as or better than FE-PCSEs without altering the specification by including a LDV, we think researchers should find them to be a particularly attractive option.[45]

[43]This point is consistent with the findings of Beck and Katz (2004).

[44]It is doubtful that Beck and Katz intended PCSEs to be used for [such small] Ts [or such large Ns] of this size, but over the past decade several articles have been publishd in top journals that do just that.

[45]In Stata, general cluster robust standard errors (including ASEs) are implemented using the `cluster` option. Douglas Miller provides an `ado` file for multiway cluster robust standard errors (such as APCSEs) at `http://www.econ.ucdavis.edu/faculty/dlmiller/statfiles/`. In `R`, the `xtpcse` package implements PCSEs, the `plm` package implements a variety of robust standard errors estimators, although it is not clear that they include the small sample correction. Our code, which includes the correction, is available for researchers who want to employ one-way and multi-way cluster robust standard errors (such as ASEs and

We have also shown that the strategy of pooling the data rather than accounting for unit effects so that time invariant variables can be included in the specification can lead to incorrect inferences on such variables. While time invariant (or near-invariant) variables are of interest for many TSCS researchers, we have shown mathematically and through simulations that particular caution is necessary when making inferences about these variables. We advise that researchers check any results they obtain from pooling against results from the between-group estimator. Including both time invariant variables and LDVs in specifications can lead to grossly incorrect inferences on the former, and in the end we may simply be unable to distinguish the effects of unobserved unit heterogeneity from the effects of observed time invariant variables. Quantitative methods are limited in what they can tell us in these situations and it is not clear that fundamental issues of identification can be overcome.

We have discussed the performance of some standard estimators under a variety of conditions but we obviously have not covered every possible scenario that TSCS researchers might face.[46] As such, we can offer no "silver bullet" approach that will always be the best option. However, we do emphasize that it is crucial for researchers to conduct and report a battery of specification tests, including tests of unit effects, tests for correlation between unit effects and explanatory variables, and tests for serial correlation (both before and after attempts to correct for it), as well as other sensitivity analysis when time-invariant variables are involved. Although it is not a novel or controversial recommendation, we close by reiterating that extensive diagnostic tests and the computation of alternative estimators can help us sort through these issues, and should be *de rigeur* in any TSCS analysis.

7 Appendix: Data generating process for experiments

The data generating process (DGP) is based on the outcome equation given in expression (1). The α_i's are drawn first. We use a uniform distribution centered on zero and scaled

APCSEs).

[46]However, we will provide our Monte Carlo code through a Web supplement that will enable researchers to explore the performance of various estimators for a range of conditions that they may encounter for their samples of interest.

to have standard deviation equal to σ_α, which is a parameter that we are free to manipulate. This amounts to selecting N α_i's as independent and identically distributed draws from Unif$[-\sigma_\alpha\sqrt{3}, \sigma_\alpha\sqrt{3}]$. A uniform distribution is used here simply to demonstrate the generality of the DGP. Other distributions could be chosen.[47]

We want to manipulate contemporaneous correlation (denoted $\rho_{cc,x}$) as well as "panel heteroskedasticity" in the x_{it}'s.[48] The panel heteroskedasticity is such that for units with indices $1, ..., N/2$, the variance of x_{it} is 1, and for units with indices $N/2+1, ..., N$, the variance is v_{x2}.[49] We also want the x_{it}'s to exhibit correlation with $\alpha_i \otimes \iota_T$ equal to $\rho_{\alpha,x}$, where ι_T is a T-length vector of ones.[50]

In order to achieve all of this, we specify,

$$x_{it} = \gamma_i \alpha_i + \eta_{it},$$

[47] There are consequences for estimated and actual variability in parameter estimates depending on whether the DGP holds the α_i's fixed or draws them anew for each sample. If the α_i's are held fixed but FE is not used to purge the variation that they induce, standard error estimators will treat such residual variation as a component of random error. The result will be an overestimate of the variability of parameter estimates. This issue has some deeper implications that get to the heart of sources of variability in real world data; refer to Mosteller and Tukey (1977, Ch. 7). But we decided not to let this issue complicate our analysis, and therefore draw the α_i's anew for each sample. This captures a data-generating process whereby unit-effects, which affect both outcome and regressor values, are originally the result of stochastic forces prior to the first period of observation.

[48] This follows Beck and Katz (1995, 1996).

[49] This implies that there be an even number of time periods, something that we have set for programming simplicity. If for some reason one wanted an odd number, one could simply drop one of the cross-sections without disturbing the essential correlational structure of the panel.

[50] We want the correlation to be specified in terms of x_{it} and the long-form $\alpha_i \otimes \iota_T$, rather than relative to the short-form α_i. With pooled OLS, the long-form $\alpha_i \otimes \iota_T$ is the actual omitted variable. Thus, omitted variable bias is most easily written in terms of the correlation between x_{it} and $\alpha_i \otimes \iota_T$.

where for a given time period, t, we have,

$$\begin{bmatrix} \eta_{1t} \\ \vdots \\ \eta_{Nt} \end{bmatrix} \sim MVN \left(\begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & & & & & & & \\ \rho_{cc,x} & 1 & & & & & & \\ \vdots & & \ddots & & & & & \\ \rho_{cc,x} & \cdots & \cdots & 1 & & & & \\ \sqrt{v_{x2}}\rho_{cc,x} & \cdots & \cdots & \sqrt{v_{x2}}\rho_{cc,x} & v_{x2} & & & \\ \vdots & \ddots & & \vdots & v_{x2}\rho_{cc,x} & \ddots & & \\ \vdots & & \ddots & \vdots & \vdots & \ddots & \ddots & \\ \sqrt{v_{x2}}\rho_{cc,x} & \cdots & \cdots & \sqrt{v_{x2}}\rho_{cc,x} & v_{x2}\rho_{cc,x} & \cdots & v_{x2}\rho_{cc,x} & v_{x2} \end{bmatrix} \right),$$

and $\gamma_i = \frac{\rho_{\alpha,x}}{\sqrt{1-\rho^2_{\alpha,x}}} \frac{\sqrt{1+\mathbf{1}(i>\frac{N}{2})(v_{x2}-1)}}{\sigma_\alpha}$, where $\mathbf{1}()$ is the indicator function. The result is that the x_{it}'s exhibit the desired levels of contemporaneous correlation, panel heteroskedasticity, and correlation with α_i. In addition, correlation with α_i implies that the x_{it}'s exhibit some serial correlation.

The ϵ_{it}'s are constructed in a manner that allows us to control contemporaneous correlation, panel heteroskedasticity, and serial correlation. To achieve this, we decompose the error as follows,

$$\epsilon_{it} = \sqrt{v_\epsilon}[\rho_\epsilon \epsilon_{i,t-1} + (\sqrt{\rho_{cc,u}})\tau_t + (\sqrt{1-\rho_{cc,u}})\epsilon_{it}], \tag{11}$$

where panel heteroskedasticity is determined by $v_u = 1$ for $i \leq N/2$ and $v_u = v_{u2}$ for $i > N/2$; ρ_ϵ is the degree of serial correlation; $\tau_t \sim \mathrm{N}(\rho_\tau \tau_{t-1}, 1)$ is a time shock for period t that may also exhibit serial correlation (if $\rho_\tau \neq 0$); $\rho_{cc,u}$ gives the amount of contemporaneous correlation in the u_{it}'s when $\rho_\tau = 0$; and $\epsilon_{it} \sim \mathrm{N}(0,1)$ is an idiosyncratic error.

For the purposes of our simulations, we only need exogenous z_i. So they are drawn as $z_i \sim N(0,1)$ and then expanded to conform to the full panel dataset as $z_i \otimes \iota_T$.

A null period ($t = 0$) is generated to allow for the generation of observations for $t = 1$, given the first order autoregressive nature of the DGP. The $y_{i,0}$'s are random draws from $\mathrm{N}(\alpha_i, 1)$. The parameter of interest, β, is fixed to equal 10 throughout. The DGP runs for a set of 50 burn-in periods that are then discarded. The subsequent T periods are used for our analysis.

References

Adolph, Christopher, Daniel M. Butler and Sven E. Wilson. N.d. "Like Shoes and Shirt, One Size Does Not Fit All: Evidence on Time Series Cross-Section Estimators and Specifications from Monte Carlo Experiments." Unpublished paper.

Angrist, Joshua D. and Jorn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton: Princeton University Press.

Arellano, Manuel. 1987. "Computing Robust Standard Errors for Within-Group Estimators." *Oxford Bulletin of Economics and Statistics* 49:431–434.

Arellano, Manuel. 1989. "A Note on the Anderson-Hsiao Estimator for Panel Data." *Economics Letters* 31:337–41.

Arellano, Manuel and Bo Honoré. 2001. Panel Data Models: Some Recent Developments. In *Handbook of Econometrics*, ed. J. J. Heckman and E. Leamer. Vol. 5 North-Holland chapter 53, pp. 3229–3296.

Baltagi, Badi H. 2008. *Econometric Analysis of Panel Data.* 3rd ed. Hoboken, NJ: Wiley.

Beck, Nathaniel. 2001. "Time-Series Cross-Section Data: What Have We Learned in the Past Few Years?" *Annual Review of Political Science* 4:271–93.

Beck, Nathaniel and Jonathan N. Katz. 1995. "What To Do (and Not To Do) with Time-Series Cross-Section Data in Comparative Politics." *American Political Science Review* 89:634–647.

Beck, Nathaniel and Jonathan N. Katz. 1996. "Nuisance vs. Substance: Specifying and Estimating Time-Series-Cross-Section Models." *Political Analysis* 6:1–36.

Beck, Nathaniel and Jonathan N. Katz. 2001. "Throwing Out the Baby with the Bath Water: A Comment on Green, Kim, and Yoon." *International Organization* 55(2):487–495.

Beck, Nathaniel and Jonathan N. Katz. 2004. "Time-Series Cross-Section Issues: Dynamics 2004." Paper presented at the Annual Meeting of the Political Methodology Society, Stanford University.

Breusch, Trevor, Michael B. Ward, Hoa Thi Minh Nguyen and Tom Kompas. 2011*a*. "FEVD: Just IV or Just Mistaken?" *Political Analysis* 19(2):165–169.

Breusch, Trevor, Michael B. Ward, Hoa Thi Minh Nguyen and Tom Kompas. 2011*b*. "On the Fixed-Effects Vector Decomposition." *Political Analysis* 19(2):123–134.

Cameron, A. Colin, Dougals Miller and Jonah Gelbach. 2011. "Robust inference with multiway clustering." *Journal of Business and Economic Statistics* 29(2):238–249.

Conley, Timothy G. 1999. "GMM Estimation with Cross Sectional Dependence." *Journal of Econometrics* 92(1):1–45.

DeBoef, Suzanna and Luke Keele. 2008. "Taking time seriously." *American Journal of Political Science* 52(1):184–200.

Gerring, John, Strom Thacker and Carola Moreno. 2005. "Centripetal Democratic Governance: A Theory and Global Inquiry." *American Political Science Review* 99(4):567–81.

Green, Donald P., Soo Yeon Kim and David Yoon. 2001. "Dirty Pool." *International Organization* 55:441–68.

Greene, William. 2011*a*. "Fixed Effects Vector Decomposition: A Magical Solution to the Problem of Time-Invariant Variables in Fixed Effects Models?" *Political Analysis* 19(2):135–146.

Greene, William. 2011*b*. "Reply to Rejoinder by Pluemper and Troeger." *Political Analysis* 19(2):170–172.

Hansen, Christian B. 2007. "Asymptotic Properties of a Robust Variance Matrix Estimator for Panel Data when T is Large." *Journal of Econometrics* 141(2):597–620.

Hausman, Jerry A. and W. Taylor. 1981. "Panel Data and Unobservable Individual Effects." *Econometrica* 49:1377–1398.

Hecock, R. Douglas. 2006. "Electoral Competition, Globalization, and Subnational Education Spending in Mexico, 1999–2004." *American Journal of Political Science* 50(4):950–961.

Hsiao, Cheng. 2003. *Analysis of Panel Data*. 2nd ed. Cambridge: Cambridge University Press.

Huber, Peter J. 1967. "The behavior of maximum likelihood estimates under non-standard conditions." *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* 1:221–233.

Iversen, Torben and David Soskice. 2006. "Electoral Institutions and the Politics of Coalitions: Why Some Democracies Redistribute Morethan Others." *American Political Science Review* 100(2):165–181.

Kezdi, Gabor. 2003. "Robust Standard Error Estimation in Fixed-Effects Panel Models." Budapest University and Central European University.

Kiefer, N. M. 1980. "Estimation of Fixed Effects Models for Time Series of Cross Sections with Arbitrary Intertemporal Covariance." *Journal of Econometrics* 26:646–79.

Kiviet, Jan F. 1995. "On Bias, Inconsistency, and Effciency of Various Estimators in Dynamic Panel Data Models." *Journal of Econometrics* 68:53–78.

Liang, Kung-Yee and Scott L. Zeger. 1986. "Longitudinal Data Analysis Using Generalized Linear Models." *Biometrika* 73:13–22.

Long, Andrew G. and Brett Ashley Leeds. 2006. "Trading For Security: Military Alliances and Economic Agreements." *Journal of Peace Research* 43(4):433–451.

Meguid, Bonnie M. 2005. "Competition between Unequals: The Role of Mainstream Party Strategy in Niche Party Success." *American Political Science Review* 99(3):347–359.

Mosley, Layna and Saika Uno. 2007. "Racing to the Bottom or Climbing to the Top." *Comparative Political Studies* 40(8):923–948.

Mosteller, Frederick and John W. Tukey. 1977. *Data Analysis and Regression*. Reading, MA: Addison-Wesley.

Moulton, Brent R. 1986. "Random group effects and the precision of regression estimates." *Journal of Econometrics* 32:385–397.

Newey, Whitney K and Kenneth D. West. 1987. "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica* 55(3):703–708.

Nickell, Stephen. 1981. "Biases in Dynamic Models with Fixed Effects." *Econometrica* 49:1417–1426.

Plümper, Thomas and Vera E. Troeger. 2007. "Efficient Estimation of Time-Invariant and Rarely Changing Variables in Finite Sample Panel Analyses with Unit Fixed Effects." *Political Analysis* 15(2):124–139.

Raftery, Adrian E. 1995. "Bayesian Model Selection in Social Research." *Sociological Methodology* 25:111–163.

Rogers, William. 1993. "Regression standard errors in clustered samples." *Stata Technical Bulletin* 13:19–23.

Treisman, Daniel. 2000. "Decentralization and Inflation: Commitment, Collective Action, or Continuity." *American Political Science Review* 94:837–857.

White, Halbert. 1980. "A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity." *Econometrica* 48:817–830.

Wilson, Sven E. and Daniel M. Butler. 2007. "A Lot More to Do: The Sensitivity of Time-Series Cross-Section Analyses to Simple Alternative Specifications." *Political Analysis* 15(2):101–123.

Table 1: Monte Carlo Analysis with $N = 16$ and $T = 20$; "Weak to Positive" Evidence in Favor of Unit Effects and Approximately Zero Correlation Between x_{it} and α_i

Cor.	Het.	ρ_ϵ	% reject $\rho_\epsilon = 0$	$\rho_{\alpha,x}$	Bias β_{OLS}	RMSE OLS	Bias β_{FE}	RMSE FE	Bias β^*_{FE}	RMSE FE*	Confidence PCSE	FE-PCSE	ASE	APCSE
0.00	0.0	0.0	100	−0.00	−0.009	0.26	−0.002	0.06	−0.002	0.06	96	100	100	100
0.00	0.3	0.0	100	−0.00	0.000	0.19	−0.000	0.07	−0.001	0.07	98	109	108	108
0.00	0.5	0.0	100	0.00	0.000	0.15	−0.007	0.08	−0.006	0.08	105	110	111	112
0.25	0.0	0.0	100	−0.00	0.011	0.27	0.000	0.09	−0.000	0.08	95	111	150	109
0.25	0.3	0.0	100	0.00	0.019	0.19	0.001	0.09	0.001	0.09	95	110	144	111
0.25	0.5	0.0	100	0.00	0.012	0.15	0.001	0.10	0.002	0.10	102	107	138	113
0.50	0.0	0.0	100	0.00	0.089	0.32	0.004	0.13	0.005	0.13	93	108	250	110
0.50	0.3	0.0	100	0.00	0.027	0.23	−0.009	0.13	−0.009	0.13	102	109	228	112
0.50	0.5	0.0	100	0.00	0.013	0.18	0.003	0.14	0.003	0.14	104	109	218	115
0.00	0.0	0.3	100	0.00	0.018	0.26	0.000	0.06	−0.001	0.06	95	106	101	101
0.00	0.3	0.3	100	−0.00	−0.013	0.20	−0.003	0.07	−0.004	0.07	99	105	103	103
0.00	0.5	0.3	100	0.00	0.002	0.14	0.001	0.08	−0.001	0.07	97	106	104	105
0.25	0.0	0.3	100	−0.00	0.022	0.27	−0.003	0.08	−0.005	0.08	97	103	142	103
0.25	0.3	0.3	100	−0.00	0.016	0.20	0.007	0.09	0.007	0.09	101	111	142	110
0.25	0.5	0.3	100	0.00	0.010	0.15	0.002	0.10	0.001	0.09	100	109	134	109
0.50	0.0	0.3	100	0.00	0.046	0.30	−0.001	0.13	−0.003	0.13	92	110	250	112
0.50	0.3	0.3	100	−0.00	0.022	0.22	0.006	0.13	0.004	0.12	97	105	220	107
0.50	0.5	0.3	100	0.00	0.006	0.18	−0.013	0.13	−0.013	0.13	108	108	205	108
0.00	0.0	0.6	100	0.00	0.005	0.24	0.001	0.06	−0.001	0.06	88	109	104	103
0.00	0.3	0.6	100	0.00	0.002	0.20	0.005	0.07	0.004	0.06	98	104	100	99
0.00	0.5	0.6	100	−0.01	−0.015	0.14	0.003	0.08	0.001	0.07	98	111	108	107
0.25	0.0	0.6	100	0.00	0.021	0.26	0.007	0.08	0.005	0.08	95	108	144	106
0.25	0.3	0.6	100	−0.00	0.014	0.19	0.006	0.09	0.005	0.08	96	107	140	108
0.25	0.5	0.6	100	−0.00	0.002	0.15	−0.000	0.09	−0.001	0.09	103	105	130	105
0.50	0.0	0.6	100	0.00	0.071	0.31	−0.002	0.12	−0.004	0.11	93	106	236	105
0.50	0.3	0.6	100	−0.00	0.024	0.21	0.004	0.13	0.001	0.12	96	109	225	108
0.50	0.5	0.6	100	0.00	0.009	0.17	−0.006	0.13	−0.005	0.12	101	105	203	106
0.00	0.0	0.9	100	−0.00	0.004	0.27	0.002	0.05	−0.000	0.05	97	111	104	105
0.00	0.3	0.9	100	−0.00	−0.006	0.21	−0.000	0.06	−0.002	0.05	104	103	101	101
0.00	0.5	0.9	100	0.00	0.002	0.15	0.001	0.07	−0.002	0.06	101	111	105	105
0.25	0.0	0.9	100	0.00	0.040	0.25	0.005	0.07	0.003	0.07	90	109	142	106
0.25	0.3	0.9	100	−0.00	0.006	0.19	0.004	0.07	0.001	0.07	96	107	136	105
0.25	0.5	0.9	100	0.01	0.021	0.15	0.007	0.09	0.004	0.08	100	116	141	116
0.50	0.0	0.9	100	−0.00	0.055	0.31	0.001	0.11	−0.002	0.11	94	113	248	114
0.50	0.3	0.9	100	0.00	0.030	0.22	−0.006	0.10	−0.009	0.10	105	108	211	110
0.50	0.5	0.9	100	0.00	0.017	0.17	0.006	0.12	0.001	0.12	106	117	221	119

Notes: Results from 500 simulations.

Cor. = contemporaneous correlation; Het.= panel heteroskedasticity; ρ_ϵ = autocorrelation parameter.

% reject $\rho_\epsilon = 0$ is the proportion of samples where the null of no serial correlation is rejected.

$\rho_{\alpha,x}$ = degree of correlation between the unit effects and the explanatory variable.

Bias β^*_{FE} and RMSE FE* refer to FE results when the LDV is not included in the specification.

Table 2: Monte Carlo Analysis with $N = 16$ and $T = 5$; "Weak to Positive" Evidence in Favor of Unit Effects and Approximately Zero Correlation Between x_{it} and α_i

Cor.	Het.	ρ_ϵ	% reject $\rho_\epsilon = 0$	$\rho_{\alpha,x}$	Bias β_{OLS}	RMSE OLS	Bias β_{FE}	RMSE FE	Bias β^*_{FE}	RMSE FE*	Confidence PCSE	Confidence FE-PCSE	Confidence ASE	Confidence APCSE
0.00	0.0	0.0	100	−0.00	−0.006	0.55	−0.016	0.16	−0.009	0.13	79	133	101	103
0.00	0.3	0.0	100	0.00	0.017	0.42	0.001	0.17	0.006	0.15	92	131	108	108
0.00	0.5	0.0	100	−0.00	−0.009	0.31	−0.012	0.20	−0.001	0.16	96	135	105	107
0.25	0.0	0.0	100	−0.00	0.107	0.59	−0.002	0.21	−0.001	0.18	85	146	146	121
0.25	0.3	0.0	100	−0.01	0.018	0.45	−0.007	0.22	−0.002	0.19	97	139	137	115
0.25	0.5	0.0	100	−0.00	0.000	0.36	−0.001	0.24	0.006	0.21	109	137	133	117
0.50	0.0	0.0	100	−0.00	0.250	0.76	0.005	0.32	0.010	0.28	95	160	236	132
0.50	0.3	0.0	100	−0.01	0.041	0.51	−0.008	0.33	−0.000	0.28	99	160	216	134
0.50	0.5	0.0	99	−0.00	0.044	0.41	0.003	0.32	0.009	0.27	114	150	189	126
0.00	0.0	0.3	100	−0.01	−0.066	0.62	0.007	0.15	0.005	0.12	87	143	100	101
0.00	0.3	0.3	100	0.00	0.019	0.45	−0.003	0.17	0.002	0.13	96	140	103	106
0.00	0.5	0.3	100	0.00	0.024	0.36	0.016	0.19	0.016	0.16	108	142	111	114
0.25	0.0	0.3	100	0.01	0.161	0.64	0.010	0.19	0.017	0.16	89	148	144	119
0.25	0.3	0.3	100	0.00	0.026	0.45	−0.008	0.20	0.003	0.17	95	143	137	118
0.25	0.5	0.3	100	−0.00	−0.001	0.36	0.002	0.22	0.006	0.19	111	143	131	113
0.50	0.0	0.3	99	0.00	0.259	0.79	0.004	0.29	−0.004	0.25	95	165	227	129
0.50	0.3	0.3	100	0.00	0.113	0.54	0.020	0.30	0.019	0.26	101	161	220	138
0.50	0.5	0.3	100	−0.00	0.041	0.40	0.015	0.30	0.016	0.26	115	156	206	133
0.00	0.0	0.6	100	0.01	0.057	0.56	0.002	0.12	0.001	0.10	83	138	105	108
0.00	0.3	0.6	100	−0.00	−0.023	0.44	−0.003	0.14	−0.003	0.12	94	144	103	105
0.00	0.5	0.6	100	−0.01	−0.004	0.34	0.009	0.16	0.010	0.14	103	151	108	112
0.25	0.0	0.6	100	−0.01	0.025	0.60	−0.001	0.16	−0.002	0.14	86	151	142	118
0.25	0.3	0.6	100	−0.00	0.030	0.43	0.005	0.18	0.007	0.15	94	162	136	118
0.25	0.5	0.6	100	0.00	0.014	0.34	−0.015	0.20	−0.014	0.17	105	151	136	117
0.50	0.0	0.6	100	0.00	0.261	0.79	−0.004	0.24	−0.009	0.21	96	167	225	127
0.50	0.3	0.6	100	0.00	0.104	0.52	0.002	0.26	−0.004	0.23	100	167	217	137
0.50	0.5	0.6	100	−0.00	0.035	0.40	0.005	0.26	−0.002	0.23	114	159	197	131
0.00	0.0	0.9	100	−0.00	0.008	0.58	0.010	0.09	0.005	0.07	85	145	100	103
0.00	0.3	0.9	100	−0.01	−0.023	0.43	−0.004	0.09	−0.005	0.08	91	138	103	105
0.00	0.5	0.9	100	0.00	−0.022	0.34	0.002	0.11	−0.003	0.10	103	148	112	115
0.25	0.0	0.9	100	−0.00	0.126	0.60	−0.004	0.11	−0.002	0.10	85	151	143	115
0.25	0.3	0.9	100	−0.00	0.041	0.46	0.009	0.12	0.008	0.11	97	150	132	112
0.25	0.5	0.9	100	−0.01	0.013	0.31	0.004	0.13	0.002	0.12	98	153	135	117
0.50	0.0	0.9	100	−0.00	0.209	0.68	0.000	0.17	0.000	0.15	85	157	211	121
0.50	0.3	0.9	100	−0.00	0.074	0.50	0.016	0.17	0.007	0.16	98	163	218	134
0.50	0.5	0.9	100	−0.01	0.024	0.36	−0.003	0.17	−0.002	0.15	105	159	180	120

Notes: Results from 500 simulations.

Cor. = contemporaneous correlation; Het.= panel heteroskedasticity; ρ_ϵ = autocorrelation parameter.

% reject $\rho_\epsilon = 0$ is the proportion of samples where the null of no serial correlation is rejected.

$\rho_{\alpha,x}$ = degree of correlation between the unit effects and the explanatory variable.

Bias β^*_{FE} and RMSE FE* refer to FE results when the LDV is not included in the specification.

Table 3: Monte Carlo Analysis with $N = 16$ and $T = 40$; "Weak to Positive" Evidence in Favor of Unit Effects and Approximately Zero Correlation Between x_{it} and α_i

Cor.	Het.	ρ_ϵ	% reject $\rho_\epsilon = 0$	$\rho_{\alpha,x}$	Bias β_{OLS}	RMSE OLS	Bias β_{FE}	RMSE FE	Bias β^*_{FE}	RMSE FE*	Confidence PCSE	Confidence FE-PCSE	Confidence ASE	Confidence APCSE
0.00	0.0	0.0	100	0.00	0.002	0.19	−0.001	0.04	−0.001	0.04	100	104	103	103
0.00	0.3	0.0	100	−0.00	−0.008	0.13	0.001	0.05	0.001	0.05	98	105	108	107
0.00	0.5	0.0	100	0.00	0.003	0.10	−0.000	0.05	−0.000	0.05	98	101	103	102
0.25	0.0	0.0	100	−0.00	0.014	0.18	−0.001	0.06	−0.001	0.06	95	106	147	108
0.25	0.3	0.0	100	−0.00	0.010	0.13	0.000	0.06	−0.000	0.06	98	99	133	101
0.25	0.5	0.0	100	−0.00	−0.002	0.11	0.001	0.07	0.001	0.06	105	104	134	105
0.50	0.0	0.0	100	0.00	0.022	0.23	−0.000	0.09	0.000	0.09	103	102	244	104
0.50	0.3	0.0	100	−0.00	0.004	0.15	−0.002	0.09	−0.002	0.09	99	107	238	109
0.50	0.5	0.0	100	−0.00	0.000	0.12	−0.003	0.09	−0.001	0.09	98	101	207	105
0.00	0.0	0.3	100	−0.00	−0.007	0.18	0.002	0.04	0.001	0.04	95	101	98	98
0.00	0.3	0.3	100	0.00	0.006	0.14	0.002	0.05	0.001	0.05	103	105	101	100
0.00	0.5	0.3	100	0.00	−0.005	0.10	−0.002	0.05	−0.003	0.05	103	108	109	109
0.25	0.0	0.3	100	0.00	0.022	0.19	−0.002	0.06	−0.002	0.06	100	104	141	103
0.25	0.3	0.3	100	0.00	0.012	0.13	0.004	0.06	0.002	0.06	99	104	138	105
0.25	0.5	0.3	100	0.00	0.005	0.11	−0.000	0.07	−0.001	0.06	104	107	135	108
0.50	0.0	0.3	100	−0.00	0.021	0.23	0.005	0.09	0.004	0.09	104	106	251	109
0.50	0.3	0.3	100	0.00	0.009	0.14	−0.000	0.09	−0.000	0.09	94	99	222	104
0.50	0.5	0.3	100	−0.00	0.001	0.12	0.004	0.09	0.004	0.09	102	104	208	107
0.00	0.0	0.6	100	−0.00	−0.006	0.18	−0.002	0.04	−0.004	0.04	95	101	97	97
0.00	0.3	0.6	100	−0.00	−0.004	0.13	−0.000	0.05	−0.002	0.04	99	102	100	100
0.00	0.5	0.6	100	0.00	−0.003	0.10	0.001	0.05	0.000	0.05	100	107	106	105
0.25	0.0	0.6	100	0.00	0.014	0.18	0.002	0.06	0.001	0.06	98	104	146	105
0.25	0.3	0.6	100	−0.00	0.001	0.13	0.000	0.06	−0.000	0.06	99	102	133	102
0.25	0.5	0.6	100	0.00	0.014	0.10	0.003	0.06	0.002	0.06	100	103	132	104
0.50	0.0	0.6	100	−0.00	0.028	0.22	−0.001	0.09	−0.003	0.09	98	112	266	114
0.50	0.3	0.6	100	0.00	0.011	0.15	0.004	0.09	0.005	0.09	101	103	218	105
0.50	0.5	0.6	100	−0.00	0.005	0.12	0.002	0.09	0.001	0.09	98	101	204	104
0.00	0.0	0.9	100	0.00	0.010	0.17	0.002	0.04	0.000	0.04	95	99	97	97
0.00	0.3	0.9	100	0.00	0.009	0.13	0.001	0.04	−0.002	0.04	100	102	101	100
0.00	0.5	0.9	100	−0.00	−0.004	0.10	−0.000	0.05	−0.002	0.05	104	101	101	101
0.25	0.0	0.9	100	0.00	0.019	0.18	0.004	0.05	0.002	0.05	94	102	141	103
0.25	0.3	0.9	100	−0.00	−0.001	0.13	0.003	0.05	0.001	0.05	101	101	133	101
0.25	0.5	0.9	100	−0.00	−0.006	0.10	0.003	0.06	0.001	0.06	99	102	133	106
0.50	0.0	0.9	100	0.00	0.041	0.21	−0.003	0.08	−0.004	0.08	96	106	249	105
0.50	0.3	0.9	100	0.00	0.010	0.14	0.000	0.08	−0.001	0.08	96	101	215	104
0.50	0.5	0.9	100	−0.00	0.000	0.12	−0.002	0.09	−0.003	0.08	100	102	213	107

Notes: Results from 500 simulations.

Cor. = contemporaneous correlation; Het.= panel heteroskedasticity; ρ_ϵ = autocorrelation parameter.

% reject $\rho_\epsilon = 0$ is the proportion of samples where the null of no serial correlation is rejected.

$\rho_{\alpha,x}$ = degree of correlation between the unit effects and the explanatory variable.

Bias β^*_{FE} and RMSE FE* refer to FE results when the LDV is not included in the specification.

Table 4: Monte Carlo Analysis with $N = 100$ and $T = 20$; "Weak to Positive" Evidence in Favor of Unit Effects and Approximately Zero Correlation Between x_{it} and α_i

Cor.	Het.	ρ_ϵ	% reject $\rho_\epsilon = 0$	$\rho_{\alpha,x}$	Bias β_{OLS}	RMSE OLS	Bias β_{FE}	RMSE FE	Bias β^*_{FE}	RMSE FE*	Confidence PCSE	Confidence FE-PCSE	Confidence ASE	Confidence APCSE
0.00	0.0	0.0	100	−0.00	0.001	0.11	−0.000	0.02	−0.000	0.02	73	105	102	104
0.00	0.3	0.0	100	−0.00	−0.003	0.08	0.001	0.03	0.001	0.03	87	106	104	104
0.00	0.5	0.0	100	0.00	0.003	0.06	0.002	0.03	0.002	0.03	97	105	101	101
0.25	0.0	0.0	100	0.00	0.033	0.15	0.001	0.06	0.001	0.06	80	108	278	108
0.25	0.3	0.0	100	−0.00	0.005	0.11	−0.001	0.07	−0.001	0.06	95	110	253	107
0.25	0.5	0.0	100	0.00	0.004	0.08	−0.003	0.06	−0.002	0.06	97	103	221	104
0.50	0.0	0.0	100	0.00	0.051	0.27	−0.004	0.11	−0.006	0.11	97	102	551	103
0.50	0.3	0.0	100	0.00	0.033	0.17	0.005	0.12	0.003	0.11	98	106	488	103
0.50	0.5	0.0	100	0.00	0.020	0.13	0.010	0.12	0.008	0.12	101	106	445	106
0.00	0.0	0.3	100	−0.00	0.000	0.11	0.002	0.02	0.001	0.02	72	108	104	104
0.00	0.3	0.3	100	0.00	0.001	0.08	0.001	0.03	0.000	0.03	87	107	103	103
0.00	0.5	0.3	100	0.00	0.005	0.06	0.001	0.03	−0.000	0.03	90	105	102	104
0.25	0.0	0.3	100	−0.00	0.013	0.15	0.002	0.06	0.002	0.06	81	105	269	103
0.25	0.3	0.3	100	0.00	0.011	0.11	0.002	0.06	0.002	0.06	91	105	241	105
0.25	0.5	0.3	100	−0.00	−0.002	0.08	−0.002	0.06	−0.003	0.06	95	105	223	106
0.50	0.0	0.3	100	0.00	0.045	0.26	0.000	0.12	−0.002	0.11	94	107	557	105
0.50	0.3	0.3	100	−0.00	0.019	0.17	0.004	0.12	0.004	0.12	99	110	514	109
0.50	0.5	0.3	100	−0.00	0.007	0.13	−0.001	0.12	0.001	0.12	105	111	464	111
0.00	0.0	0.6	100	0.00	−0.002	0.11	0.002	0.02	−0.001	0.02	74	100	95	96
0.00	0.3	0.6	100	0.00	0.001	0.07	0.003	0.03	0.000	0.02	79	102	97	96
0.00	0.5	0.6	100	0.00	−0.001	0.06	0.002	0.03	−0.000	0.03	91	110	104	105
0.25	0.0	0.6	100	0.00	0.039	0.16	0.001	0.06	−0.002	0.06	83	109	275	104
0.25	0.3	0.6	100	0.00	0.017	0.10	0.006	0.06	0.004	0.06	88	110	244	107
0.25	0.5	0.6	100	0.00	0.002	0.08	0.000	0.07	−0.003	0.06	98	113	236	111
0.50	0.0	0.6	100	0.00	0.056	0.26	0.001	0.11	−0.000	0.11	93	108	561	104
0.50	0.3	0.6	100	−0.00	0.018	0.17	−0.006	0.11	−0.008	0.11	97	107	485	105
0.50	0.5	0.6	100	0.00	0.007	0.13	−0.000	0.12	−0.000	0.11	104	110	439	106
0.00	0.0	0.9	100	−0.00	−0.006	0.10	0.004	0.02	0.002	0.02	68	108	103	103
0.00	0.3	0.9	100	−0.00	−0.004	0.08	0.003	0.02	−0.000	0.02	80	105	101	99
0.00	0.5	0.9	100	0.00	0.004	0.06	0.004	0.03	0.002	0.02	90	105	99	98
0.25	0.0	0.9	100	−0.00	0.018	0.14	0.003	0.05	0.000	0.05	75	109	263	105
0.25	0.3	0.9	100	−0.00	0.012	0.10	0.005	0.05	0.003	0.05	86	110	240	105
0.25	0.5	0.9	100	0.00	0.006	0.08	−0.000	0.05	−0.004	0.05	96	108	213	103
0.50	0.0	0.9	100	−0.00	0.043	0.25	0.002	0.10	0.002	0.10	94	115	571	112
0.50	0.3	0.9	100	−0.00	0.024	0.17	0.002	0.10	−0.000	0.09	102	107	478	104
0.50	0.5	0.9	100	0.00	0.011	0.11	0.005	0.09	0.000	0.09	97	102	417	102

Notes: Results from 500 simulations.

Cor. = contemporaneous correlation; Het.= panel heteroskedasticity; ρ_ϵ = autocorrelation parameter.

% reject $\rho_\epsilon = 0$ is the proportion of samples where the null of no serial correlation is rejected.

$\rho_{\alpha,x}$ = degree of correlation between the unit effects and the explanatory variable.

Bias β^*_{FE} and RMSE FE* refer to FE results when the LDV is not included in the specification.

Table 5: Monte Carlo Analysis with $N = 16$ and $T = 20$; "Positive to Strong" Evidence in Favor of Unit Effects and Approximately Zero Correlation Between x_{it} and α_i

Cor.	Het.	ρ_ϵ	% reject $\rho_\epsilon = 0$	$\rho_{\alpha,x}$	Bias β_{OLS}	RMSE OLS	Bias β_{FE}	RMSE FE	Bias β^*_{FE}	RMSE FE*	Confidence PCSE	FE-PCSE	ASE	APCSE
0.00	0.0	0.0	98	−0.00	0.036	0.53	−0.002	0.06	−0.002	0.06	80	100	100	100
0.00	0.3	0.0	100	−0.00	0.039	0.48	−0.000	0.07	−0.001	0.07	81	109	108	108
0.00	0.5	0.0	100	0.00	0.028	0.42	−0.007	0.08	−0.006	0.08	87	110	111	112
0.25	0.0	0.0	97	−0.00	0.110	0.67	0.000	0.09	−0.000	0.08	87	111	150	109
0.25	0.3	0.0	100	0.00	0.096	0.55	0.001	0.09	0.001	0.09	82	110	144	111
0.25	0.5	0.0	100	0.00	0.072	0.47	0.001	0.10	0.002	0.10	90	107	138	113
0.50	0.0	0.0	91	0.00	0.301	1.00	0.004	0.13	0.005	0.13	93	108	250	110
0.50	0.3	0.0	100	0.00	0.206	0.78	−0.009	0.13	−0.009	0.13	88	109	228	112
0.50	0.5	0.0	100	0.00	0.088	0.59	0.003	0.14	0.003	0.14	93	109	218	115
0.00	0.0	0.3	99	0.00	0.077	0.51	0.000	0.06	−0.001	0.06	77	106	101	101
0.00	0.3	0.3	100	−0.00	−0.011	0.52	−0.003	0.07	−0.004	0.07	87	105	103	103
0.00	0.5	0.3	100	0.00	0.022	0.44	0.001	0.08	−0.001	0.07	90	106	104	105
0.25	0.0	0.3	98	−0.00	0.147	0.68	−0.003	0.08	−0.005	0.08	88	103	142	103
0.25	0.3	0.3	100	−0.00	0.108	0.58	0.007	0.09	0.007	0.09	87	111	142	110
0.25	0.5	0.3	100	0.00	0.059	0.45	0.002	0.10	0.001	0.09	87	109	134	109
0.50	0.0	0.3	93	0.00	0.201	0.94	−0.001	0.13	−0.003	0.13	90	110	250	112
0.50	0.3	0.3	100	−0.00	0.171	0.82	0.006	0.13	0.004	0.12	94	105	220	107
0.50	0.5	0.3	100	0.00	0.131	0.60	−0.013	0.13	−0.013	0.13	94	108	205	108
0.00	0.0	0.6	99	0.00	0.051	0.49	0.001	0.06	−0.001	0.06	76	109	104	103
0.00	0.3	0.6	100	0.00	0.016	0.50	0.005	0.07	0.004	0.06	84	104	100	99
0.00	0.5	0.6	100	−0.01	−0.041	0.43	0.003	0.08	0.001	0.07	89	111	108	107
0.25	0.0	0.6	99	0.00	0.088	0.64	0.007	0.08	0.005	0.08	84	108	144	106
0.25	0.3	0.6	100	−0.00	0.130	0.55	0.006	0.09	0.005	0.08	82	107	140	108
0.25	0.5	0.6	100	−0.00	0.061	0.45	−0.000	0.09	−0.001	0.09	88	105	130	105
0.50	0.0	0.6	95	0.00	0.278	0.95	−0.002	0.12	−0.004	0.11	89	106	236	105
0.50	0.3	0.6	100	−0.00	0.145	0.78	0.004	0.13	0.001	0.12	91	109	225	108
0.50	0.5	0.6	100	0.00	0.154	0.61	−0.006	0.13	−0.005	0.12	95	105	203	106
0.00	0.0	0.9	100	−0.00	0.055	0.53	0.002	0.05	−0.000	0.05	80	111	104	105
0.00	0.3	0.9	100	−0.00	−0.006	0.54	−0.000	0.06	−0.002	0.05	91	103	101	101
0.00	0.5	0.9	100	0.00	0.053	0.43	0.001	0.07	−0.002	0.06	88	111	105	105
0.25	0.0	0.9	99	0.00	0.147	0.65	0.005	0.07	0.003	0.07	83	109	142	106
0.25	0.3	0.9	100	−0.00	0.083	0.61	0.004	0.07	0.001	0.07	91	107	136	105
0.25	0.5	0.9	100	0.01	0.086	0.46	0.007	0.09	0.004	0.08	88	116	141	116
0.50	0.0	0.9	95	−0.00	0.264	0.92	0.001	0.11	−0.002	0.11	87	113	248	114
0.50	0.3	0.9	100	0.00	0.225	0.83	−0.006	0.10	−0.009	0.10	96	108	211	110
0.50	0.5	0.9	100	0.00	0.116	0.61	0.006	0.12	0.001	0.12	97	117	221	119

Notes: Results from 500 simulations.

Cor. = contemporaneous correlation; Het.= panel heteroskedasticity; ρ_ϵ = autocorrelation parameter.

% reject $\rho_\epsilon = 0$ is the proportion of samples where the null of no serial correlation is rejected.

$\rho_{\alpha,x}$ = degree of correlation between the unit effects and the explanatory variable.

Bias β^*_{FE} and RMSE FE* refer to FE results when the LDV is not included in the specification.

Table 6: Monte Carlo Analysis with $N = 16$ and $T = 20$; "Weak to Positive" Evidence in Favor of Unit Effects and Correlation Between x_{it} and α_i

Cor.	Het.	ρ_ϵ	% reject $\rho_\epsilon = 0$	$\rho_{\alpha,x}$	Bias β_{OLS}	RMSE OLS	Bias β_{FE}	RMSE FE	Bias β^*_{FE}	RMSE FE*	Confidence PCSE	Confidence FE-PCSE	Confidence ASE	Confidence APCSE
0.00	0.0	0.0	100	0.29	0.851	0.88	−0.004	0.06	−0.004	0.06	89	104	103	102
0.00	0.3	0.0	100	0.28	0.635	0.66	0.002	0.07	0.004	0.07	96	106	106	106
0.00	0.5	0.0	100	0.26	0.423	0.44	−0.001	0.08	0.000	0.08	95	109	108	109
0.25	0.0	0.0	100	0.29	0.873	0.90	−0.005	0.08	−0.005	0.08	84	103	133	99
0.25	0.3	0.0	100	0.28	0.661	0.68	0.006	0.09	0.007	0.09	85	105	138	107
0.25	0.5	0.0	100	0.26	0.439	0.46	−0.000	0.09	0.002	0.09	94	101	128	104
0.50	0.0	0.0	100	0.29	0.905	0.97	−0.002	0.13	−0.002	0.13	101	109	246	107
0.50	0.3	0.0	100	0.28	0.672	0.71	−0.003	0.13	−0.002	0.13	95	106	224	108
0.50	0.5	0.0	100	0.26	0.452	0.49	0.000	0.15	0.001	0.14	102	115	222	116
0.00	0.0	0.3	100	0.29	0.853	0.88	0.001	0.06	0.000	0.06	83	104	98	97
0.00	0.3	0.3	100	0.27	0.622	0.65	0.003	0.07	0.002	0.07	91	106	106	107
0.00	0.5	0.3	100	0.26	0.428	0.45	0.004	0.07	0.003	0.07	94	103	103	103
0.25	0.0	0.3	100	0.29	0.880	0.91	−0.002	0.08	−0.003	0.08	86	106	143	104
0.25	0.3	0.3	100	0.28	0.668	0.69	−0.000	0.08	−0.001	0.08	92	102	131	103
0.25	0.5	0.3	100	0.26	0.433	0.46	0.001	0.09	−0.000	0.09	97	102	130	104
0.50	0.0	0.3	100	0.30	0.931	0.98	0.005	0.13	0.004	0.13	91	113	253	112
0.50	0.3	0.3	100	0.28	0.673	0.71	−0.001	0.13	−0.000	0.13	95	107	223	108
0.50	0.5	0.3	100	0.26	0.445	0.48	−0.003	0.13	−0.006	0.13	101	105	204	106
0.00	0.0	0.6	100	0.29	0.849	0.87	0.001	0.06	−0.001	0.05	81	105	100	99
0.00	0.3	0.6	100	0.28	0.617	0.64	0.001	0.06	−0.001	0.06	88	100	98	98
0.00	0.5	0.6	100	0.26	0.423	0.44	0.003	0.07	0.001	0.07	95	100	99	100
0.25	0.0	0.6	100	0.29	0.896	0.93	−0.003	0.08	−0.004	0.08	87	109	146	108
0.25	0.3	0.6	100	0.28	0.654	0.68	−0.003	0.08	−0.005	0.08	91	105	136	105
0.25	0.5	0.6	100	0.26	0.437	0.46	0.006	0.09	0.002	0.09	97	110	136	110
0.50	0.0	0.6	100	0.29	0.921	0.97	0.008	0.13	0.005	0.12	88	110	246	108
0.50	0.3	0.6	100	0.29	0.665	0.70	−0.003	0.12	−0.003	0.12	94	106	218	107
0.50	0.5	0.6	100	0.27	0.457	0.49	0.005	0.13	0.005	0.13	97	110	212	112
0.00	0.0	0.9	100	0.29	0.841	0.86	0.001	0.05	−0.001	0.05	78	107	100	99
0.00	0.3	0.9	100	0.28	0.629	0.65	0.000	0.06	−0.003	0.05	87	103	100	100
0.00	0.5	0.9	100	0.27	0.427	0.45	−0.001	0.07	−0.003	0.07	90	111	108	109
0.25	0.0	0.9	100	0.29	0.854	0.88	0.007	0.07	0.004	0.06	86	106	133	98
0.25	0.3	0.9	100	0.28	0.646	0.67	0.005	0.07	0.001	0.07	96	105	131	103
0.25	0.5	0.9	100	0.26	0.426	0.45	−0.003	0.08	−0.003	0.08	97	114	137	111
0.50	0.0	0.9	100	0.30	0.932	0.98	0.005	0.11	−0.000	0.10	90	108	242	109
0.50	0.3	0.9	100	0.28	0.663	0.70	0.001	0.11	−0.001	0.11	95	110	227	110
0.50	0.5	0.9	100	0.27	0.444	0.48	−0.003	0.11	−0.004	0.11	101	109	200	109

Notes: Results from 500 simulations.

Cor. = contemporaneous correlation; Het.= panel heteroskedasticity; ρ_ϵ = autocorrelation parameter.

% reject $\rho_\epsilon = 0$ is the proportion of samples where the null of no serial correlation is rejected.

$\rho_{\alpha,x}$ = degree of correlation between the unit effects and the explanatory variable.

Bias β^*_{FE} and RMSE FE* refer to FE results when the LDV is not included in the specification.

Table 7: Monte Carlo Analysis with $N = 16$ and $T = 20$; "Positive to Strong" Evidence in Favor of Unit Effects and Correlation Between x_{it} and α_i

Cor.	Het.	ρ_ϵ	% reject $\rho_\epsilon = 0$	$\rho_{\alpha,x}$	Bias β_{OLS}	RMSE OLS	Bias β_{FE}	RMSE FE	Bias β^*_{FE}	RMSE FE*	Confidence PCSE	FE-PCSE	ASE	APCSE
0.00	0.0	0.0	100	0.29	1.006	1.10	−0.004	0.06	−0.004	0.06	79	104	103	102
0.00	0.3	0.0	100	0.28	1.023	1.10	0.002	0.07	0.004	0.07	82	106	106	106
0.00	0.5	0.0	100	0.26	0.912	0.97	−0.001	0.08	0.000	0.08	82	109	108	109
0.25	0.0	0.0	100	0.29	1.040	1.18	−0.005	0.08	−0.005	0.08	82	103	133	99
0.25	0.3	0.0	100	0.28	1.081	1.18	0.006	0.09	0.007	0.09	82	105	138	107
0.25	0.5	0.0	100	0.26	0.958	1.03	−0.000	0.09	0.002	0.09	82	101	128	104
0.50	0.0	0.0	99	0.29	1.099	1.41	−0.002	0.13	−0.002	0.13	98	109	246	107
0.50	0.3	0.0	100	0.28	1.117	1.30	−0.003	0.13	−0.002	0.13	87	106	224	108
0.50	0.5	0.0	100	0.26	0.979	1.10	0.000	0.15	0.001	0.14	88	115	222	116
0.00	0.0	0.3	100	0.29	0.988	1.08	0.001	0.06	0.000	0.06	75	104	98	97
0.00	0.3	0.3	100	0.27	0.998	1.07	0.003	0.07	0.002	0.07	78	106	106	107
0.00	0.5	0.3	100	0.26	0.921	0.98	0.004	0.07	0.003	0.07	82	103	103	103
0.25	0.0	0.3	100	0.29	1.068	1.20	−0.002	0.08	−0.003	0.08	81	106	143	104
0.25	0.3	0.3	100	0.28	1.120	1.22	−0.000	0.08	−0.001	0.08	84	102	131	103
0.25	0.5	0.3	100	0.26	0.929	1.01	0.001	0.09	−0.000	0.09	84	102	130	104
0.50	0.0	0.3	99	0.30	1.156	1.41	0.005	0.13	0.004	0.13	88	113	253	112
0.50	0.3	0.3	100	0.28	1.111	1.30	−0.001	0.13	−0.000	0.13	88	107	223	108
0.50	0.5	0.3	100	0.26	0.980	1.13	−0.003	0.13	−0.006	0.13	98	105	204	106
0.00	0.0	0.6	100	0.29	0.982	1.06	0.001	0.06	−0.001	0.05	70	105	100	99
0.00	0.3	0.6	100	0.28	0.990	1.06	0.001	0.06	−0.001	0.06	77	100	98	98
0.00	0.5	0.6	100	0.26	0.911	0.97	0.003	0.07	0.001	0.07	80	100	99	100
0.25	0.0	0.6	100	0.29	1.097	1.22	−0.003	0.08	−0.004	0.08	80	109	146	108
0.25	0.3	0.6	100	0.28	1.090	1.19	−0.003	0.08	−0.005	0.08	81	105	136	105
0.25	0.5	0.6	100	0.26	0.969	1.04	0.006	0.09	0.002	0.09	86	110	136	110
0.50	0.0	0.6	98	0.29	1.137	1.38	0.008	0.13	0.005	0.12	87	110	246	108
0.50	0.3	0.6	100	0.29	1.091	1.27	−0.003	0.12	−0.003	0.12	87	106	218	107
0.50	0.5	0.6	100	0.27	0.983	1.09	0.005	0.13	0.005	0.13	84	110	212	112
0.00	0.0	0.9	100	0.29	0.989	1.07	0.001	0.05	−0.001	0.05	73	107	100	99
0.00	0.3	0.9	100	0.28	1.000	1.08	0.000	0.06	−0.003	0.05	78	103	100	100
0.00	0.5	0.9	100	0.27	0.902	0.96	−0.001	0.07	−0.003	0.07	77	111	108	109
0.25	0.0	0.9	100	0.29	1.009	1.14	0.007	0.07	0.004	0.06	78	106	133	98
0.25	0.3	0.9	100	0.28	1.090	1.18	0.005	0.07	0.001	0.07	80	105	131	103
0.25	0.5	0.9	100	0.26	0.925	1.00	−0.003	0.08	−0.003	0.08	85	114	137	111
0.50	0.0	0.9	100	0.30	1.155	1.41	0.005	0.11	−0.000	0.10	90	108	242	109
0.50	0.3	0.9	100	0.28	1.120	1.30	0.001	0.11	−0.001	0.11	88	110	227	110
0.50	0.5	0.9	100	0.27	0.974	1.11	−0.003	0.11	−0.004	0.11	93	109	200	109

Notes: Results from 500 simulations.

Cor. = contemporaneous correlation; Het.= panel heteroskedasticity; ρ_ϵ = autocorrelation parameter.

% reject $\rho_\epsilon = 0$ is the proportion of samples where the null of no serial correlation is rejected.

$\rho_{\alpha,x}$ = degree of correlation between the unit effects and the explanatory variable.

Bias β^*_{FE} and RMSE FE* refer to FE results when the LDV is not included in the specification.

Table 8: Monte Carlo Analysis with $N = 100$ and $T = 20$; "Positive to Strong" Evidence in Favor of Unit Effects and Correlation Between x_{it} and α_i

Cor.	Het.	ρ_ϵ	% reject $\rho_\epsilon = 0$	$\rho_{\alpha,x}$	Bias β_{OLS}	RMSE OLS	Bias β_{FE}	RMSE FE	Bias β^*_{FE}	RMSE FE*	Confidence PCSE	Confidence FE-PCSE	Confidence ASE	Confidence APCSE
0.00	0.0	0.0	100	0.30	0.996	1.01	0.001	0.02	0.002	0.02	44	104	99	97
0.00	0.3	0.0	100	0.29	1.035	1.05	0.002	0.03	0.002	0.03	48	105	104	104
0.00	0.5	0.0	100	0.27	0.936	0.95	−0.000	0.03	0.000	0.03	52	105	104	104
0.25	0.0	0.0	100	0.30	1.083	1.16	0.001	0.06	0.002	0.06	74	104	273	106
0.25	0.3	0.0	100	0.29	1.068	1.12	−0.002	0.07	−0.001	0.06	74	109	252	109
0.25	0.5	0.0	100	0.27	0.984	1.02	0.003	0.06	0.003	0.06	76	103	213	102
0.50	0.0	0.0	100	0.31	1.225	1.42	0.011	0.12	0.012	0.12	83	112	594	112
0.50	0.3	0.0	100	0.29	1.131	1.28	−0.005	0.12	−0.004	0.12	86	111	512	108
0.50	0.5	0.0	100	0.27	1.005	1.10	−0.002	0.11	−0.002	0.11	88	101	433	103
0.00	0.0	0.3	100	0.30	0.989	1.01	0.001	0.02	−0.000	0.02	45	105	102	102
0.00	0.3	0.3	100	0.29	1.032	1.05	0.002	0.03	0.001	0.03	51	106	104	104
0.00	0.5	0.3	100	0.27	0.936	0.95	0.004	0.03	0.003	0.03	54	107	103	103
0.25	0.0	0.3	100	0.30	1.077	1.15	0.002	0.06	−0.000	0.06	76	108	278	105
0.25	0.3	0.3	100	0.29	1.080	1.13	0.001	0.07	−0.001	0.06	72	112	250	109
0.25	0.5	0.3	100	0.27	0.976	1.01	−0.000	0.06	−0.002	0.06	73	103	217	103
0.50	0.0	0.3	100	0.31	1.120	1.33	0.000	0.12	−0.002	0.12	84	110	577	110
0.50	0.3	0.3	100	0.29	1.148	1.32	0.012	0.12	0.009	0.11	93	108	508	105
0.50	0.5	0.3	100	0.27	0.990	1.09	0.008	0.12	0.004	0.12	90	113	464	109
0.00	0.0	0.6	100	0.30	0.983	1.00	0.003	0.02	0.000	0.02	42	106	100	100
0.00	0.3	0.6	100	0.29	1.020	1.03	0.003	0.03	0.001	0.03	46	108	104	104
0.00	0.5	0.6	100	0.27	0.944	0.95	0.002	0.03	−0.001	0.03	55	106	101	100
0.25	0.0	0.6	100	0.30	1.049	1.13	0.002	0.06	−0.000	0.06	76	114	292	110
0.25	0.3	0.6	100	0.29	1.117	1.16	0.000	0.06	−0.002	0.06	71	102	233	98
0.25	0.5	0.6	100	0.27	0.949	0.98	0.004	0.07	0.001	0.06	73	116	235	111
0.50	0.0	0.6	100	0.31	1.110	1.32	0.004	0.11	0.001	0.11	85	109	563	108
0.50	0.3	0.6	100	0.29	1.127	1.28	0.001	0.11	0.001	0.10	90	100	453	95
0.50	0.5	0.6	100	0.27	1.006	1.10	0.001	0.12	−0.003	0.12	90	112	472	111
0.00	0.0	0.9	100	0.30	0.998	1.01	0.003	0.02	0.000	0.02	42	107	100	101
0.00	0.3	0.9	100	0.29	1.020	1.03	0.002	0.02	−0.001	0.02	48	110	102	102
0.00	0.5	0.9	100	0.27	0.928	0.94	0.003	0.02	0.000	0.02	52	101	97	97
0.25	0.0	0.9	100	0.30	1.046	1.12	0.000	0.06	−0.001	0.05	73	116	276	107
0.25	0.3	0.9	100	0.29	1.071	1.12	0.003	0.05	0.000	0.05	71	113	255	112
0.25	0.5	0.9	100	0.27	0.984	1.02	0.005	0.05	0.003	0.05	78	107	221	106
0.50	0.0	0.9	100	0.30	1.171	1.39	0.001	0.10	0.000	0.10	89	109	578	107
0.50	0.3	0.9	100	0.29	1.105	1.25	−0.011	0.10	−0.012	0.10	87	111	500	106
0.50	0.5	0.9	100	0.27	1.010	1.10	−0.002	0.10	−0.004	0.10	88	115	463	112

Notes: Results from 500 simulations.

Cor. = contemporaneous correlation; Het.= panel heteroskedasticity; ρ_ϵ = autocorrelation parameter.

% reject $\rho_\epsilon = 0$ is the proportion of samples where the null of no serial correlation is rejected.

$\rho_{\alpha,x}$ = degree of correlation between the unit effects and the explanatory variable.

Bias β^*_{FE} and RMSE FE* refer to FE results when the LDV is not included in the specification.

Table 9: Monte Carlo Analysis for Time Invariant Variables; $N = 16$ and $T = 20$; "Weak to Positive" Evidence in Favor of Unit Effects and Approximately Zero Correlation Between x_{it} and α_i

	No lag	One lag	Two lags
Mean OLS γ	-0.03	-0.03	-0.03
Mean Between γ	-0.02	-0.01	-0.01
Proportion of significant OLS γs	0.68	0.65	0.62
Proportion of significant OLS γs, PCSEs	0.93	0.80	0.75
Proportion of significant OLS γs, Robust	0.10	0.11	0.11
Proportion of significant Between γs	0.06	0.05	0.07
% reject H_0 of no autocorrelation	100.00	100.00	100.00

Notes: $N = 15; T = 20$. 1000 simulations for each model. The DGP does not include a time invariant variable, but the estimation equation does.

Table 10: Monte Carlo Analysis for Time Invariant Variables; $N = 16$ and $T = 20$; "Weak to Positive" Evidence in Favor of Unit Effects and Approximately Zero Correlation Between x_{it}, z_i, and α_i

Cor.	Het.	ρ_ϵ	% reject $\rho_\epsilon = 0$	$\rho_{\alpha,y_{t-1}}$	$\rho_{\alpha,x}$	$\rho_{\alpha,z}$	$\rho_{y_{t-1},x}$	$\rho_{y_{t-1},z}$	Bias λ_{OLS}	RMSE OLS	Confidence PCSE
0.00	0.0	0.00	100	0.32	0.00	−0.00	−0.00	0.63	−1.821	2.18	290
0.00	0.3	0.00	100	0.25	0.00	−0.01	−0.00	0.51	−0.945	1.57	441
0.00	0.5	0.00	100	0.20	0.00	0.02	−0.01	0.38	−0.355	1.42	610
0.25	0.0	0.00	100	0.30	−0.00	−0.03	−0.01	0.64	−1.920	2.22	269
0.25	0.3	0.00	100	0.26	0.00	0.00	−0.01	0.52	−0.929	1.57	445
0.25	0.5	0.00	100	0.19	−0.01	0.01	−0.02	0.38	−0.360	1.34	597
0.50	0.0	0.00	100	0.33	−0.00	0.01	−0.02	0.65	−1.808	2.16	272
0.50	0.3	0.00	100	0.26	0.00	0.01	−0.02	0.52	−0.873	1.61	461
0.50	0.5	0.00	100	0.19	−0.00	−0.00	−0.01	0.39	−0.445	1.38	593
0.00	0.0	0.30	100	0.31	0.00	−0.02	−0.00	0.64	−1.907	2.31	316
0.00	0.3	0.30	100	0.26	0.00	0.00	−0.00	0.52	−0.924	1.59	450
0.00	0.5	0.30	100	0.19	−0.00	−0.00	−0.00	0.38	−0.471	1.47	623
0.25	0.0	0.30	100	0.32	−0.00	0.00	−0.01	0.64	−1.804	2.20	304
0.25	0.3	0.30	100	0.26	0.00	0.01	−0.00	0.52	−0.928	1.59	456
0.25	0.5	0.30	100	0.20	0.00	0.01	−0.01	0.38	−0.407	1.38	609
0.50	0.0	0.30	100	0.33	0.01	0.01	−0.02	0.64	−1.829	2.25	299
0.50	0.3	0.30	100	0.23	−0.00	−0.04	−0.02	0.51	−1.082	1.67	455
0.50	0.5	0.30	100	0.20	0.00	0.02	−0.03	0.38	−0.414	1.40	597
0.00	0.0	0.60	100	0.33	−0.00	0.01	−0.00	0.64	−1.755	2.15	300
0.00	0.3	0.60	100	0.25	−0.01	−0.01	−0.01	0.51	−1.010	1.65	466
0.00	0.5	0.60	100	0.20	−0.00	0.02	−0.01	0.38	−0.363	1.44	631
0.25	0.0	0.60	100	0.32	−0.00	−0.01	−0.01	0.65	−1.856	2.24	302
0.25	0.3	0.60	100	0.25	−0.00	−0.01	−0.01	0.51	−0.965	1.58	452
0.25	0.5	0.60	100	0.20	−0.00	0.01	−0.02	0.38	−0.448	1.42	626
0.50	0.0	0.60	100	0.33	0.00	0.02	−0.01	0.65	−1.815	2.19	280
0.50	0.3	0.60	100	0.25	0.00	−0.01	−0.02	0.52	−1.042	1.66	454
0.50	0.5	0.60	100	0.20	0.00	0.00	−0.02	0.39	−0.489	1.39	587
0.00	0.0	0.90	100	0.33	0.01	0.01	−0.00	0.64	−1.807	2.19	300
0.00	0.3	0.90	100	0.26	0.00	0.00	−0.00	0.51	−0.972	1.63	462
0.00	0.5	0.90	100	0.19	−0.00	0.01	−0.00	0.38	−0.442	1.43	636
0.25	0.0	0.90	100	0.33	0.00	0.00	−0.01	0.63	−1.874	2.29	324
0.25	0.3	0.90	100	0.26	0.00	0.01	−0.01	0.51	−0.934	1.68	510
0.25	0.5	0.90	100	0.20	0.00	0.01	−0.02	0.39	−0.459	1.43	654
0.50	0.0	0.90	100	0.33	0.00	0.00	−0.02	0.64	−1.931	2.29	284
0.50	0.3	0.90	100	0.26	0.00	−0.00	−0.02	0.52	−1.039	1.68	461
0.50	0.5	0.90	100	0.19	−0.00	0.01	−0.02	0.39	−0.447	1.38	637

Notes: Cor. = contemporaneous correlation; Het.= panel heteroskedasticity; ρ_ϵ = autocorrelation parameter.

% reject $\rho_\epsilon = 0$ is the proportion of samples where the null of no serial correlation is rejected.

$\rho_{\alpha,\cdot}$ = correlation between the unit effects and RHS components.

The DGP includes the time invariant variable z_i.

Ancillary Analysis For Web Appendix

While we have done an extensive amount of ancillary analysis that is not discussed in detail in the paper, we include in this appendix some key results that are referred to as being available from the authors. We have additional results should reviewers like to see them, but the tables included here give results for what we see as the most relevant experiments referenced but not reported in detail in the paper. The results included here will be made available through a web appendix.

- Table A-1 reports the results of experiments with $N = 16$ and $T = 20$ and a lag of x_{it} included in the specification. Table A-2 reports results for this specification when $T = 5$. Tables A-5 and A-6 report results for the inclusion of $x_{i,t-1}$ but with correlation between the explanatory variables and the unit effect.

- Table A-3 reports the results of experiments with $N = 16$ and $T = 20$ and no contemporaneous correlation in the x_{it}. Table A-4 has the same design for the x_{it} but increases N to 100.

Table A-1: Monte Carlo Analysis with $N = 16$ and $T = 20$; "Weak to Positive" Evidence in Favor of Unit Effects and Approximately Zero Correlation Between x_{it} and α_i; $x_{i,t-1}$ Included in Specification

Cor.	Het.	ρ_ϵ	% reject $\rho_\epsilon = 0$	$\rho_{\alpha,x}$	Bias β_{OLS}	RMSE OLS	Bias β_{FE}	RMSE FE	Bias β^*_{FE}	RMSE FE*	Confidence PCSE	Confidence FE-PCSE	Confidence ASE	Confidence APCSE
0.00	0.0	0.0	100	−0.00	−0.000	0.08	−0.002	0.06	−0.002	0.06	96	100	100	100
0.00	0.3	0.0	100	−0.00	−0.002	0.09	0.000	0.07	−0.001	0.07	104	109	108	108
0.00	0.5	0.0	100	0.00	−0.008	0.10	−0.007	0.08	−0.006	0.08	107	111	111	112
0.25	0.0	0.0	100	−0.00	−0.003	0.11	0.000	0.09	−0.000	0.08	102	113	150	109
0.25	0.3	0.0	100	0.00	−0.001	0.12	0.001	0.09	0.001	0.09	106	111	144	111
0.25	0.5	0.0	100	0.00	−0.002	0.12	0.002	0.10	0.002	0.10	101	109	138	113
0.50	0.0	0.0	100	0.00	0.017	0.18	0.004	0.13	0.005	0.13	106	112	250	110
0.50	0.3	0.0	100	0.00	−0.016	0.17	−0.007	0.13	−0.009	0.13	106	112	228	112
0.50	0.5	0.0	100	0.00	0.004	0.17	0.004	0.14	0.003	0.14	106	112	218	115
0.00	0.0	0.3	100	0.00	−0.000	0.07	−0.001	0.06	−0.001	0.06	98	105	101	101
0.00	0.3	0.3	100	−0.00	−0.000	0.08	−0.003	0.07	−0.004	0.07	102	105	103	103
0.00	0.5	0.3	100	0.00	−0.001	0.08	−0.000	0.07	−0.001	0.07	101	105	104	105
0.25	0.0	0.3	100	−0.00	−0.004	0.10	−0.004	0.08	−0.005	0.08	105	104	142	103
0.25	0.3	0.3	100	−0.00	0.007	0.10	0.005	0.09	0.007	0.09	107	111	142	110
0.25	0.5	0.3	98	0.00	0.002	0.11	0.001	0.09	0.001	0.09	108	111	134	109
0.50	0.0	0.3	99	0.00	0.006	0.15	0.000	0.12	−0.003	0.13	106	111	250	112
0.50	0.3	0.3	98	−0.00	−0.008	0.14	0.002	0.12	0.004	0.12	103	106	220	107
0.50	0.5	0.3	96	0.00	−0.010	0.15	−0.014	0.13	−0.013	0.13	105	109	205	108
0.00	0.0	0.6	94	0.00	0.001	0.06	0.000	0.05	−0.001	0.06	105	110	104	103
0.00	0.3	0.6	87	0.00	−0.003	0.06	−0.000	0.06	0.004	0.06	99	102	100	99
0.00	0.5	0.6	74	−0.01	0.000	0.07	0.001	0.06	0.001	0.07	100	107	108	107
0.25	0.0	0.6	87	0.00	0.005	0.08	0.005	0.07	0.005	0.08	109	112	144	106
0.25	0.3	0.6	81	−0.00	−0.002	0.08	0.001	0.07	0.005	0.08	100	103	140	108
0.25	0.5	0.6	71	−0.00	−0.001	0.08	−0.001	0.08	−0.001	0.09	101	104	130	105
0.50	0.0	0.6	75	0.00	−0.003	0.11	−0.003	0.10	−0.004	0.11	103	105	236	105
0.50	0.3	0.6	76	−0.00	0.008	0.12	0.005	0.11	0.001	0.12	108	108	225	108
0.50	0.5	0.6	67	0.00	0.003	0.12	−0.002	0.11	−0.005	0.12	109	109	203	106
0.00	0.0	0.9	17	−0.00	−0.001	0.03	−0.001	0.03	−0.000	0.05	102	107	104	105
0.00	0.3	0.9	25	−0.00	−0.003	0.04	−0.003	0.04	−0.002	0.05	99	102	101	101
0.00	0.5	0.9	23	0.00	0.001	0.04	0.000	0.04	−0.002	0.06	104	110	105	105
0.25	0.0	0.9	28	0.00	0.002	0.04	0.002	0.04	0.003	0.07	102	105	142	106
0.25	0.3	0.9	27	−0.00	−0.000	0.05	−0.000	0.05	0.001	0.07	107	111	136	105
0.25	0.5	0.9	28	0.01	0.002	0.05	0.002	0.05	0.004	0.08	102	106	141	116
0.50	0.0	0.9	41	−0.00	0.001	0.06	−0.000	0.06	−0.002	0.11	95	101	248	114
0.50	0.3	0.9	39	0.00	−0.006	0.07	−0.007	0.07	−0.009	0.10	105	106	211	110
0.50	0.5	0.9	39	0.00	0.004	0.08	0.004	0.08	0.001	0.12	110	115	221	119

Notes: Results from 500 simulations.

Cor. = contemporaneous correlation; Het.= panel heteroskedasticity; ρ_ϵ = autocorrelation parameter.

% reject $\rho_\epsilon = 0$ is the proportion of samples where the null of no serial correlation is rejected.

$\rho_{\alpha,x}$ = degree of correlation between the unit effects and the explanatory variable.

Bias β^*_{FE} and RMSE FE* refer to FE results when the LDV is not included in the specification.

Table A-2: Monte Carlo Analysis with $N = 16$ and $T = 5$; "Weak to Positive" Evidence in Favor of Unit Effects and Approximately Zero Correlation Between x_{it} and α_i; $x_{i,t-1}$ Included in Specification

Cor.	Het.	ρ_ϵ	% reject $\rho_\epsilon = 0$	$\rho_{\alpha,x}$	Bias β_{OLS}	RMSE OLS	Bias β_{FE}	RMSE FE	Bias β^*_{FE}	RMSE FE*	Confidence PCSE	Confidence FE-PCSE	Confidence ASE	Confidence APCSE
0.00	0.0	0.0	87	−0.00	−0.012	0.18	−0.011	0.16	−0.009	0.13	106	147	101	103
0.00	0.3	0.0	81	0.00	0.007	0.20	0.006	0.17	0.006	0.15	111	144	108	108
0.00	0.5	0.0	71	−0.00	0.000	0.22	−0.007	0.20	−0.001	0.16	110	148	105	107
0.25	0.0	0.0	84	−0.00	0.008	0.25	0.003	0.20	−0.001	0.18	126	164	146	121
0.25	0.3	0.0	80	−0.01	−0.002	0.24	0.001	0.22	−0.002	0.19	117	162	137	115
0.25	0.5	0.0	65	−0.00	−0.004	0.27	0.002	0.24	0.006	0.21	119	153	133	117
0.50	0.0	0.0	78	−0.00	0.015	0.39	0.006	0.32	0.010	0.28	146	198	236	132
0.50	0.3	0.0	74	−0.01	−0.009	0.39	−0.001	0.33	−0.000	0.28	141	196	216	134
0.50	0.5	0.0	67	−0.00	0.012	0.38	0.011	0.31	0.009	0.27	141	181	189	126
0.00	0.0	0.3	55	−0.01	0.005	0.16	0.010	0.15	0.005	0.12	112	152	100	101
0.00	0.3	0.3	54	0.00	0.001	0.18	−0.002	0.17	0.002	0.13	110	149	103	106
0.00	0.5	0.3	42	0.00	0.011	0.20	0.016	0.20	0.016	0.16	116	154	111	114
0.25	0.0	0.3	60	0.01	0.007	0.20	0.012	0.19	0.017	0.16	123	162	144	119
0.25	0.3	0.3	55	0.00	0.001	0.22	−0.005	0.20	0.003	0.17	122	155	137	118
0.25	0.5	0.3	48	−0.00	0.009	0.23	0.001	0.22	0.006	0.19	120	157	131	113
0.50	0.0	0.3	59	0.00	0.013	0.33	0.004	0.28	−0.004	0.25	143	190	227	129
0.50	0.3	0.3	54	0.00	0.010	0.33	0.021	0.30	0.019	0.26	142	193	220	138
0.50	0.5	0.3	49	−0.00	−0.001	0.32	0.015	0.30	0.016	0.26	138	182	206	133
0.00	0.0	0.6	27	0.01	0.004	0.12	0.002	0.12	0.001	0.10	109	142	105	108
0.00	0.3	0.6	25	−0.00	−0.002	0.14	−0.007	0.14	−0.003	0.12	113	149	103	105
0.00	0.5	0.6	24	−0.01	0.006	0.16	0.006	0.16	0.010	0.14	119	156	108	112
0.25	0.0	0.6	30	−0.01	−0.002	0.16	−0.002	0.16	−0.002	0.14	123	159	142	118
0.25	0.3	0.6	31	−0.00	0.001	0.17	0.001	0.18	0.007	0.15	124	172	136	118
0.25	0.5	0.6	29	0.00	−0.011	0.19	−0.015	0.20	−0.014	0.17	127	162	136	117
0.50	0.0	0.6	40	0.00	−0.013	0.27	−0.011	0.23	−0.009	0.21	153	193	225	127
0.50	0.3	0.6	37	0.00	0.002	0.26	−0.001	0.25	−0.004	0.23	147	189	217	137
0.50	0.5	0.6	30	−0.00	−0.001	0.26	0.001	0.25	−0.002	0.23	143	183	197	131
0.00	0.0	0.9	9	−0.00	0.004	0.08	0.007	0.08	0.005	0.07	109	143	100	103
0.00	0.3	0.9	12	−0.01	−0.008	0.09	−0.006	0.09	−0.005	0.08	107	140	103	105
0.00	0.5	0.9	14	0.00	−0.001	0.10	0.002	0.10	−0.003	0.10	112	147	112	115
0.25	0.0	0.9	13	−0.00	−0.009	0.10	−0.007	0.10	−0.002	0.10	128	159	143	115
0.25	0.3	0.9	17	−0.00	0.007	0.11	0.007	0.11	0.008	0.11	129	162	132	112
0.25	0.5	0.9	20	−0.01	0.006	0.12	0.003	0.12	0.002	0.12	124	157	135	117
0.50	0.0	0.9	25	−0.00	−0.013	0.16	−0.008	0.15	0.000	0.15	147	182	211	121
0.50	0.3	0.9	21	−0.00	0.004	0.16	0.014	0.16	0.007	0.16	141	187	218	134
0.50	0.5	0.9	22	−0.01	−0.008	0.16	−0.004	0.17	−0.002	0.15	139	182	180	120

Notes: Results from 500 simulations.

Cor. = contemporaneous correlation; Het.= panel heteroskedasticity; ρ_ϵ = autocorrelation parameter.

% reject $\rho_\epsilon = 0$ is the proportion of samples where the null of no serial correlation is rejected.

$\rho_{\alpha,x}$ = degree of correlation between the unit effects and the explanatory variable.

Bias β^*_{FE} and RMSE FE* refer to FE results when the LDV is not included in the specification.

Table A-3: Monte Carlo Analysis with $N = 16$ and $T = 20$; "Weak to Positive" Evidence in Favor of Unit Effects and Approximately Zero Correlation Between x_{it} and α_i; No Contemporaneous Correlation in the x_{it}

Cor.	Het.	ρ_ϵ	% reject $\rho_\epsilon = 0$	$\rho_{\alpha,x}$	Bias β_{OLS}	RMSE OLS	Bias β_{FE}	RMSE FE	Bias β^*_{FE}	RMSE FE*	Confidence PCSE	Confidence FE-PCSE	Confidence ASE	Confidence APCSE
0.00	0.0	0.0	100	−0.00	−0.009	0.26	−0.002	0.06	−0.002	0.06	96	100	100	100
0.00	0.3	0.0	100	−0.00	0.000	0.19	−0.000	0.07	−0.001	0.07	98	109	108	108
0.00	0.5	0.0	100	0.00	0.000	0.15	−0.007	0.08	−0.006	0.08	105	110	111	112
0.25	0.0	0.0	100	−0.00	−0.006	0.26	−0.000	0.06	−0.001	0.06	94	109	105	105
0.25	0.3	0.0	100	0.00	0.010	0.19	0.001	0.07	0.002	0.07	93	106	105	104
0.25	0.5	0.0	100	0.00	0.007	0.15	0.001	0.08	0.001	0.08	101	104	107	107
0.50	0.0	0.0	100	0.00	0.036	0.25	0.002	0.06	0.002	0.06	90	108	103	102
0.50	0.3	0.0	100	0.00	−0.000	0.19	−0.005	0.07	−0.005	0.07	96	111	108	108
0.50	0.5	0.0	100	0.00	0.002	0.14	−0.000	0.08	−0.000	0.08	94	105	108	107
0.00	0.0	0.3	100	0.00	0.018	0.26	0.000	0.06	−0.001	0.06	95	106	101	101
0.00	0.3	0.3	100	−0.00	−0.013	0.20	−0.003	0.07	−0.004	0.07	99	105	103	103
0.00	0.5	0.3	100	0.00	0.002	0.14	0.001	0.08	−0.001	0.07	97	106	104	105
0.25	0.0	0.3	100	−0.00	−0.002	0.26	−0.001	0.06	−0.002	0.06	95	102	99	99
0.25	0.3	0.3	100	−0.00	0.006	0.20	0.005	0.07	0.005	0.07	99	107	105	105
0.25	0.5	0.3	100	0.00	0.004	0.15	−0.000	0.08	−0.001	0.08	101	110	107	107
0.50	0.0	0.3	100	0.00	0.004	0.24	−0.001	0.06	−0.002	0.06	86	105	101	101
0.50	0.3	0.3	100	−0.00	−0.008	0.18	0.003	0.07	0.001	0.06	93	101	99	99
0.50	0.5	0.3	100	0.00	−0.000	0.15	−0.007	0.08	−0.008	0.07	102	104	100	100
0.00	0.0	0.6	100	0.00	0.005	0.24	0.001	0.06	−0.001	0.06	88	109	104	103
0.00	0.3	0.6	100	0.00	0.002	0.20	0.005	0.07	0.004	0.06	98	104	100	99
0.00	0.5	0.6	100	−0.01	−0.015	0.14	0.003	0.08	0.001	0.07	98	111	108	107
0.25	0.0	0.6	100	0.00	−0.001	0.25	0.004	0.06	0.002	0.05	91	106	100	101
0.25	0.3	0.6	100	−0.00	0.000	0.19	0.006	0.07	0.003	0.07	97	108	107	107
0.25	0.5	0.6	100	−0.00	0.000	0.15	0.003	0.07	0.001	0.07	103	106	103	102
0.50	0.0	0.6	100	0.00	0.019	0.27	0.000	0.06	−0.002	0.05	97	105	100	100
0.50	0.3	0.6	100	−0.00	0.004	0.18	0.004	0.06	0.001	0.06	93	104	100	99
0.50	0.5	0.6	100	0.00	−0.000	0.14	−0.003	0.07	−0.004	0.07	96	107	104	104
0.00	0.0	0.9	100	−0.00	0.004	0.27	0.002	0.05	−0.000	0.05	97	111	104	105
0.00	0.3	0.9	100	−0.00	−0.006	0.21	−0.000	0.06	−0.002	0.05	104	103	101	101
0.00	0.5	0.9	100	0.00	0.002	0.15	0.001	0.07	−0.002	0.06	101	111	105	105
0.25	0.0	0.9	100	0.00	0.015	0.25	0.002	0.05	−0.001	0.05	91	107	100	101
0.25	0.3	0.9	100	−0.00	−0.008	0.18	0.004	0.06	0.001	0.05	92	107	102	103
0.25	0.5	0.9	100	0.01	0.016	0.14	0.005	0.07	0.003	0.07	98	114	110	111
0.50	0.0	0.9	100	−0.00	−0.012	0.26	0.001	0.05	−0.001	0.05	93	106	97	98
0.50	0.3	0.9	100	0.00	0.002	0.20	−0.002	0.05	−0.005	0.05	102	105	100	102
0.50	0.5	0.9	100	0.00	0.007	0.14	0.006	0.07	0.002	0.06	99	110	106	104

Notes: Results from 500 simulations.

Cor. = contemporaneous correlation; Het.= panel heteroskedasticity; ρ_ϵ = autocorrelation parameter.

% reject $\rho_\epsilon = 0$ is the proportion of samples where the null of no serial correlation is rejected.

$\rho_{\alpha,x}$ = degree of correlation between the unit effects and the explanatory variable.

Bias β^*_{FE} and RMSE FE* refer to FE results when the LDV is not included in the specification.

Table A-4: Monte Carlo Analysis with $N = 100$ and $T = 20$; "Weak to Positive" Evidence in Favor of Unit Effects and Approximately Zero Correlation Between x_{it} and α_i; No Contemporaneous Correlation in the x_{it}

Cor.	Het.	ρ_ϵ	% reject $\rho_\epsilon = 0$	$\rho_{\alpha,x}$	Bias β_{OLS}	RMSE OLS	Bias β_{FE}	RMSE FE	Bias β^*_{FE}	RMSE FE*	Confidence PCSE	Confidence FE-PCSE	Confidence ASE	Confidence APCSE
0.00	0.0	0.0	100	−0.00	0.001	0.11	−0.000	0.02	−0.000	0.02	73	105	102	104
0.00	0.3	0.0	100	−0.00	−0.003	0.08	0.001	0.03	0.001	0.03	87	106	104	104
0.00	0.5	0.0	100	0.00	0.003	0.06	0.002	0.03	0.002	0.03	97	105	101	101
0.25	0.0	0.0	100	0.00	0.010	0.10	0.000	0.03	0.001	0.02	69	109	105	103
0.25	0.3	0.0	100	−0.00	−0.006	0.09	−0.001	0.03	−0.001	0.03	92	105	103	103
0.25	0.5	0.0	100	0.00	0.000	0.06	−0.002	0.03	−0.002	0.03	94	103	100	101
0.50	0.0	0.0	100	0.00	0.009	0.11	−0.002	0.02	−0.001	0.02	75	101	98	97
0.50	0.3	0.0	100	0.00	0.005	0.08	0.001	0.03	0.001	0.03	84	103	100	99
0.50	0.5	0.0	100	0.00	0.005	0.06	0.002	0.03	0.002	0.03	88	101	98	98
0.00	0.0	0.3	100	−0.00	0.000	0.11	0.002	0.02	0.001	0.02	72	108	104	104
0.00	0.3	0.3	100	0.00	0.001	0.08	0.001	0.03	0.000	0.03	87	107	103	103
0.00	0.5	0.3	100	0.00	0.005	0.06	0.001	0.03	−0.000	0.03	90	105	102	104
0.25	0.0	0.3	100	−0.00	−0.005	0.10	0.001	0.02	0.001	0.02	72	104	100	99
0.25	0.3	0.3	100	0.00	−0.000	0.08	0.002	0.03	0.001	0.03	82	104	102	102
0.25	0.5	0.3	100	−0.00	−0.005	0.05	0.001	0.03	−0.000	0.03	87	99	96	97
0.50	0.0	0.3	100	−0.00	0.000	0.10	0.001	0.02	−0.000	0.02	70	104	99	99
0.50	0.3	0.3	100	−0.00	−0.001	0.08	0.003	0.03	0.001	0.03	84	104	101	102
0.50	0.5	0.3	100	0.00	0.001	0.06	0.001	0.03	0.001	0.03	93	105	102	102
0.00	0.0	0.6	100	0.00	−0.002	0.11	0.002	0.02	−0.001	0.02	74	100	95	96
0.00	0.3	0.6	100	0.00	0.001	0.07	0.003	0.03	0.000	0.02	79	102	97	96
0.00	0.5	0.6	100	0.00	−0.001	0.06	0.002	0.03	−0.000	0.03	91	110	104	105
0.25	0.0	0.6	100	0.00	0.004	0.10	0.002	0.02	−0.001	0.02	70	105	97	96
0.25	0.3	0.6	100	0.00	0.002	0.08	0.004	0.03	0.002	0.03	81	108	102	102
0.25	0.5	0.6	100	0.00	0.000	0.06	0.002	0.03	−0.000	0.03	91	109	104	104
0.50	0.0	0.6	100	0.00	0.007	0.11	0.001	0.02	−0.001	0.02	72	101	96	96
0.50	0.3	0.6	100	−0.00	−0.003	0.08	0.000	0.03	−0.002	0.03	83	106	101	100
0.50	0.5	0.6	100	0.00	0.000	0.06	0.002	0.03	0.000	0.03	88	109	101	100
0.00	0.0	0.9	100	−0.00	−0.006	0.10	0.004	0.02	0.002	0.02	68	108	103	103
0.00	0.3	0.9	100	−0.00	−0.004	0.08	0.003	0.02	−0.000	0.02	80	105	101	99
0.00	0.5	0.9	100	0.00	0.004	0.06	0.004	0.03	0.002	0.02	90	105	99	98
0.25	0.0	0.9	100	−0.00	−0.006	0.10	0.002	0.02	−0.000	0.02	68	107	98	99
0.25	0.3	0.9	100	−0.00	0.000	0.08	0.003	0.02	0.001	0.02	82	107	100	101
0.25	0.5	0.9	100	0.00	0.001	0.06	0.001	0.03	−0.002	0.02	91	109	101	102
0.50	0.0	0.9	100	−0.00	−0.002	0.10	0.003	0.02	0.001	0.02	70	113	106	106
0.50	0.3	0.9	100	−0.00	0.000	0.08	0.003	0.02	0.000	0.02	84	105	99	100
0.50	0.5	0.9	100	0.00	0.001	0.06	0.003	0.03	0.000	0.02	88	104	100	101

Notes: Results from 500 simulations.

Cor. = contemporaneous correlation; Het.= panel heteroskedasticity; ρ_ϵ = autocorrelation parameter.

% reject $\rho_\epsilon = 0$ is the proportion of samples where the null of no serial correlation is rejected.

$\rho_{\alpha,x}$ = degree of correlation between the unit effects and the explanatory variable.

Bias β^*_{FE} and RMSE FE* refer to FE results when the LDV is not included in the specification.

Table A-5: Monte Carlo Analysis with $N = 16$ and $T = 20$; "Weak to Positive" Evidence in Favor of Unit Effects and Correlation Between x_{it} and α_i; $x_{i,t-1}$ Included in Specification

Cor.	Het.	ρ_ϵ	% reject $\rho_\epsilon = 0$	$\rho_{\alpha,x}$	Bias β_{OLS}	RMSE OLS	Bias β_{FE}	RMSE FE	Bias β^*_{FE}	RMSE FE*	Confidence PCSE	Confidence FE-PCSE	Confidence ASE	Confidence APCSE
0.00	0.0	0.0	100	0.29	0.057	0.10	−0.004	0.06	−0.004	0.06	100	104	103	102
0.00	0.3	0.0	100	0.28	0.080	0.12	0.002	0.07	0.004	0.07	105	106	106	106
0.00	0.5	0.0	100	0.26	0.096	0.13	−0.001	0.08	0.000	0.08	100	110	108	109
0.25	0.0	0.0	100	0.29	0.063	0.13	−0.005	0.08	−0.005	0.08	106	104	133	99
0.25	0.3	0.0	100	0.28	0.091	0.15	0.006	0.09	0.007	0.09	102	106	138	107
0.25	0.5	0.0	100	0.26	0.102	0.15	−0.000	0.09	0.002	0.09	101	103	128	104
0.50	0.0	0.0	100	0.29	0.059	0.18	−0.002	0.13	−0.002	0.13	103	113	246	107
0.50	0.3	0.0	100	0.28	0.076	0.19	−0.003	0.13	−0.002	0.13	103	108	224	108
0.50	0.5	0.0	100	0.26	0.099	0.20	0.001	0.14	0.001	0.14	113	118	222	116
0.00	0.0	0.3	100	0.29	0.045	0.08	0.001	0.06	0.000	0.06	102	104	98	97
0.00	0.3	0.3	100	0.27	0.061	0.10	0.003	0.07	0.002	0.07	109	108	106	107
0.00	0.5	0.3	99	0.26	0.073	0.11	0.002	0.07	0.003	0.07	96	101	103	103
0.25	0.0	0.3	100	0.29	0.044	0.10	−0.002	0.08	−0.003	0.08	104	106	143	104
0.25	0.3	0.3	100	0.28	0.063	0.12	−0.000	0.08	−0.001	0.08	102	102	131	103
0.25	0.5	0.3	98	0.26	0.065	0.12	−0.002	0.09	−0.000	0.09	99	101	130	104
0.50	0.0	0.3	98	0.30	0.051	0.17	0.003	0.13	0.004	0.13	112	115	253	112
0.50	0.3	0.3	97	0.28	0.059	0.15	−0.001	0.12	−0.000	0.13	99	107	223	108
0.50	0.5	0.3	93	0.26	0.057	0.16	−0.007	0.13	−0.006	0.13	108	106	204	106
0.00	0.0	0.6	93	0.29	0.026	0.06	−0.001	0.05	−0.001	0.05	97	102	100	99
0.00	0.3	0.6	91	0.28	0.036	0.07	0.000	0.05	−0.001	0.06	99	100	98	98
0.00	0.5	0.6	81	0.26	0.045	0.08	0.001	0.06	0.001	0.07	103	102	99	100
0.25	0.0	0.6	86	0.29	0.021	0.08	−0.005	0.07	−0.004	0.08	104	106	146	108
0.25	0.3	0.6	82	0.28	0.033	0.08	−0.005	0.07	−0.005	0.08	102	104	136	105
0.25	0.5	0.6	71	0.26	0.047	0.10	0.004	0.08	0.002	0.09	105	108	136	110
0.50	0.0	0.6	77	0.29	0.033	0.13	0.005	0.11	0.005	0.12	113	116	246	108
0.50	0.3	0.6	69	0.29	0.039	0.12	−0.002	0.11	−0.003	0.12	107	108	218	107
0.50	0.5	0.6	64	0.27	0.043	0.12	−0.001	0.11	0.005	0.13	102	106	212	112
0.00	0.0	0.9	19	0.29	0.007	0.03	−0.002	0.03	−0.001	0.05	101	104	100	99
0.00	0.3	0.9	24	0.28	0.011	0.04	−0.002	0.03	−0.003	0.05	97	98	100	100
0.00	0.5	0.9	20	0.27	0.016	0.05	−0.000	0.04	−0.003	0.07	102	107	108	109
0.25	0.0	0.9	27	0.29	0.011	0.05	0.002	0.05	0.004	0.06	109	110	133	98
0.25	0.3	0.9	29	0.28	0.010	0.05	−0.002	0.05	0.001	0.07	105	107	131	103
0.25	0.5	0.9	24	0.26	0.011	0.05	−0.004	0.05	−0.003	0.08	106	111	137	111
0.50	0.0	0.9	39	0.30	0.012	0.07	0.002	0.07	−0.000	0.10	110	113	242	109
0.50	0.3	0.9	40	0.28	0.012	0.07	−0.001	0.07	−0.001	0.11	105	106	227	110
0.50	0.5	0.9	41	0.27	0.014	0.08	−0.003	0.07	−0.004	0.11	108	110	200	109

Notes: Results from 500 simulations.

Cor. = contemporaneous correlation; Het.= panel heteroskedasticity; ρ_ϵ = autocorrelation parameter.

% reject $\rho_\epsilon = 0$ is the proportion of samples where the null of no serial correlation is rejected.

$\rho_{\alpha,x}$ = degree of correlation between the unit effects and the explanatory variable.

Bias β^*_{FE} and RMSE FE* refer to FE results when the LDV is not included in the specification.

Table A-6: Monte Carlo Analysis with $N = 16$ and $T = 20$; "Positive to Strong" Evidence in Favor of Unit Effects and Correlation Between x_{it} and α_i; $x_{i,t-1}$ Included in Specification

Cor.	Het.	ρ_ϵ	% reject $\rho_\epsilon = 0$	$\rho_{\alpha,x}$	Bias β_{OLS}	RMSE OLS	Bias β_{FE}	RMSE FE	Bias β^*_{FE}	RMSE FE*	Confidence PCSE	FE-PCSE	ASE	APCSE
0.00	0.0	0.0	100	0.29	0.013	0.08	−0.004	0.06	−0.004	0.06	99	104	103	102
0.00	0.3	0.0	100	0.28	0.021	0.10	0.002	0.07	0.004	0.07	104	106	106	106
0.00	0.5	0.0	100	0.26	0.030	0.11	−0.001	0.08	0.000	0.08	100	110	108	109
0.25	0.0	0.0	100	0.29	0.018	0.12	−0.005	0.08	−0.005	0.08	105	104	133	99
0.25	0.3	0.0	100	0.28	0.031	0.12	0.006	0.09	0.007	0.09	101	106	138	107
0.25	0.5	0.0	100	0.26	0.035	0.13	−0.000	0.09	0.002	0.09	100	103	128	104
0.50	0.0	0.0	100	0.29	0.012	0.17	−0.002	0.13	−0.002	0.13	102	113	246	107
0.50	0.3	0.0	100	0.28	0.016	0.18	−0.003	0.13	−0.002	0.13	102	108	224	108
0.50	0.5	0.0	100	0.26	0.028	0.20	0.001	0.14	0.001	0.14	112	118	222	116
0.00	0.0	0.3	100	0.29	0.014	0.07	0.001	0.06	0.000	0.06	103	104	98	97
0.00	0.3	0.3	100	0.27	0.020	0.09	0.003	0.07	0.002	0.07	110	108	106	107
0.00	0.5	0.3	100	0.26	0.023	0.09	0.002	0.07	0.003	0.07	96	101	103	103
0.25	0.0	0.3	100	0.29	0.011	0.10	−0.002	0.08	−0.003	0.08	104	106	143	104
0.25	0.3	0.3	100	0.28	0.020	0.10	−0.000	0.08	−0.001	0.08	101	102	131	103
0.25	0.5	0.3	99	0.26	0.015	0.10	−0.002	0.09	−0.000	0.09	98	101	130	104
0.50	0.0	0.3	99	0.30	0.017	0.16	0.003	0.13	0.004	0.13	111	115	253	112
0.50	0.3	0.3	99	0.28	0.016	0.15	−0.001	0.12	−0.000	0.13	99	107	223	108
0.50	0.5	0.3	97	0.26	0.005	0.16	−0.007	0.13	−0.006	0.13	108	106	204	106
0.00	0.0	0.6	95	0.29	0.007	0.05	−0.001	0.05	−0.001	0.05	97	102	100	99
0.00	0.3	0.6	95	0.28	0.011	0.06	0.000	0.05	−0.001	0.06	100	100	98	98
0.00	0.5	0.6	91	0.26	0.015	0.07	0.001	0.06	0.001	0.07	103	102	99	100
0.25	0.0	0.6	88	0.29	0.001	0.08	−0.005	0.07	−0.004	0.08	104	106	146	108
0.25	0.3	0.6	89	0.28	0.007	0.08	−0.005	0.07	−0.005	0.08	101	104	136	105
0.25	0.5	0.6	83	0.26	0.017	0.09	0.004	0.08	0.002	0.09	105	108	136	110
0.50	0.0	0.6	79	0.29	0.012	0.13	0.005	0.11	0.005	0.12	112	116	246	108
0.50	0.3	0.6	76	0.29	0.012	0.12	−0.002	0.11	−0.003	0.12	106	108	218	107
0.50	0.5	0.6	74	0.27	0.011	0.12	−0.001	0.11	0.005	0.13	102	106	212	112
0.00	0.0	0.9	20	0.29	0.000	0.03	−0.002	0.03	−0.001	0.05	100	104	100	99
0.00	0.3	0.9	27	0.28	0.002	0.04	−0.002	0.03	−0.003	0.05	97	98	100	100
0.00	0.5	0.9	26	0.27	0.005	0.04	−0.000	0.04	−0.003	0.07	103	107	108	109
0.25	0.0	0.9	28	0.29	0.004	0.05	0.002	0.05	0.004	0.06	109	110	133	98
0.25	0.3	0.9	32	0.28	0.001	0.05	−0.002	0.05	0.001	0.07	105	107	131	103
0.25	0.5	0.9	26	0.26	0.001	0.05	−0.004	0.05	−0.003	0.08	106	111	137	111
0.50	0.0	0.9	40	0.30	0.005	0.07	0.002	0.07	−0.000	0.10	110	113	242	109
0.50	0.3	0.9	40	0.28	0.002	0.07	−0.001	0.07	−0.001	0.11	105	106	227	110
0.50	0.5	0.9	43	0.27	0.003	0.07	−0.003	0.07	−0.004	0.11	108	110	200	109

Notes: Results from 500 simulations.

Cor. = contemporaneous correlation; Het.= panel heteroskedasticity; ρ_ϵ = autocorrelation parameter.

% reject $\rho_\epsilon = 0$ is the proportion of samples where the null of no serial correlation is rejected.

$\rho_{\alpha,x}$ = degree of correlation between the unit effects and the explanatory variable.

Bias β^*_{FE} and RMSE FE* refer to FE results when the LDV is not included in the specification.