cemmap

centre for microdata methods and practice

Estimating dynamic panel models: backing out the Nickell Bias

Jerry A. Hausman
Maxim L. Pinkovskiy

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP53/17

E·S·R·C
ECONOMIC & SOCIAL RESEARCH COUNCIL

An ESRC Research Centre

Estimating Dynamic Panel Models: Backing out the Nickell Bias

Jerry A. Hausman and Maxim L. Pinkovskiy*

November 2017

Abstract

We propose a new estimator for the dynamic panel model, which solves the failure of strict exogeneity by calculating the bias in the first-order conditions as a function of the autoregressive parameter and solving the resulting equation. The estimator does well in a wide variety of situations where other estimators do not perform well: stationary initial condition, predetermined but not strictly exogenous regressors, and the presence of correlation between the error terms and the fixed effects. We also propose a general method for including predetermined variables in fixed-effects panel regressions.

JEL Codes: C01, C22, C23

*We are very grateful to Jacob Conway for superb research assistance. The views expressed in this article are those of the authors and do not necessarily reflect the opinions of the Federal Reserve Bank of New York or of the Federal Reserve System

1 Introduction

Our paper contributes to the literature on estimating linear dynamic panel data models with lagged dependent variables. The idea that estimating the dynamic panel equation by OLS will produce biased and inconsistent estimates has been explored in the literature since Nickell (1981) and Anderson and Hsiao (1982), with Arellano and Bond (1991) proposing an optimal GMM estimator. The Arellano-Bond estimator exhibits substantial downward bias when the coefficient on the lagged dependent variable is close to unity, as then the dependent variable follows a near random walk and lagged levels correlate poorly with lagged differences, thus creating a weak instrument problem. A strand of the literature (Ahn and Schmidt (1995) , Blundell and Bond (1998), Hahn (1999)) solves this problem by imposing further restrictions on the dependent variable process and exploiting the resulting moment conditions; however, these restrictions may not hold in practice. Hahn, Hausman and Kuersteiner (2007) follow Griliches and Hausman (1986) and take long differences of the data to improve the correlation between levels and differences; however, this approach does not make use of all the data available. Hence the estimation of dynamic panel models is still an open problem.

We propose a new estimator for the dynamic panel model, which is based on computing the bias terms in the first-order condition for the autoregressive coefficient that result from the failure of strict exogeneity. The main assumption that we must maintain for this approach is the lack of serial correlation between the model errors, as in Arellano and Bond (1991). We find a modified version of this first-order condition, one of whose roots is a consistent estimator of the true autoregressive parameter. We can expand our estimator to accomodate all predetermined variables, and we develop a general method for predetermined variables in a panel regression context that is also based on the idea of correcting the first-order conditions to make them unbiased estimators of zero at the true parameter values.

Simulations of the performance of our estimator against that of previous GMM-based estimators suggests that our estimator nearly always produces unbiased estimates of the coefficients on the lagged dependent variable and on the covariate (which often tends to be of primary interest in applications) in finite samples. In particular, we present evidence that, unlike many instrumental-variables based estimators, our technique delivers consistent estimates regardless of the distribution of the initial values of the dependent variable. Our estimator also produces estimates of both coefficients with variances that are somewhat smaller than those of Arellano-Bond for multiple data generating processes, and tends to produce estimates with variances that are close to the variances of Arellano-Bond in other settings. Our estimator also matches the performance of existing estimators in terms of allowing other regressors to be predetermined but not exogenous. We also compare our estimator with the factor-based approach recently proposed by Bai (2013) and find that our estimator can accommodate the case in which fixed effects and model errors are correlated (also matching the performance of Arellano and Bond 1991), while the Bai (2013) estimator delivers consistent estimates only on the assumption that the two are uncorrelated.[1]

The rest of the paper is organized as follows. Section 2 presents a simple version of our dynamic panel estimator. Section 3 expands the estimator to accomodate weaker assumptions on the data. Section 4 presents simulation evidence on the properties of our estimator. Section 5 concludes.

2 The Estimator

We consider the problem of estimating the model

$$y_{i,t} = \alpha_0 y_{i,t-1} + x'_{i,t}\beta_0 + \mu_{i,0} + \varepsilon_{i,t} \tag{1}$$

where $y_{i,t}$ is the dependent variable, $x_{i,t}$ is a vector of regressors, $\mu_{i,0}$ is a fixed effect and $\varepsilon_{i,t}$ is the error term. There are N panel units i, with N thought of as large, and T time units t, with T treated as a fixed parameter. We consider combinations of the following assumptions:

[1] Hsiao, Pesaran and Tahmiscioglu (2002) also propose an estimator under additional assumptions on the covariates $x_{i,t}$. Our approach does not require any assumptions on $x_{i,t}$.

$$
\begin{aligned}
E\left(\varepsilon_{i,t}\varepsilon_{j,t'}\right) &= 0 \text{ if } i \neq j \text{ or } t \neq t' \text{ (NSC)} \\
E\left(x_{i,t}\varepsilon_{j,t'}\right) &= 0 \text{ if } i \neq j \text{ or } t \neq t' \text{ (GM)} \\
E\left(x_{i,t}\varepsilon_{j,t'}\right) &= 0 \text{ if } i \neq j \text{ or } t' \geq t \text{ (PR)}
\end{aligned}
$$

Assumption NSC is the no-serial correlation assumption used by much of the literature following Arellano and Bond (1991), and it will be maintained for this estimator.[2] Assumption GM states that the regressors $x_{i,t}$ are strictly exogenous, and assumption PR states that they are predetermined, but not necessarily exogenous. We will see that assumption GM can be weakened to assumption PR. We will also consider an additional assumption, which we will not require to hold in our approach.

$$E\left(\mu_{i,0}\varepsilon_{i,t}\right) = 0 \text{ (ECF)}$$

Assumption (ECF) states that the errors are uncorrelated with the true fixed effects $\mu_{i,0}$. This assumption will be important later on, as the estimator of Bai (2013) implicitly relies on it. [3]

2.1 Notation

First, we define the empirical fixed effects as functions of estimators of α and β:

$$\hat{\mu}_i\left(\alpha, \beta\right) = \frac{1}{T}\sum_{t=1}^{T}\left(y_{i,t} - \alpha y_{i,t-1} - x'_{i,t}\beta\right)$$

[2] In principle, failure of assumption (NSC) can be seen as evidence for model misspecification and lack of inclusion of an adequate number of lags of the dependent variable into the model. Our methodology can easily be extended in principle to multiple lags of the dependent variable, with the key step becoming the solution of a system of polynomial equations in the coefficients on the multiple lags.

[3] Strictly speaking, we make an additional assumption that the errors are uncorrelated with the initial values of the process, $y_{i,0}$.

$$E\left(y_{i,0}\varepsilon_{i,t}\right) = 0 \; (ECI)$$

We can relax this assumption in our estimator, while failure of this assumption would be problematic for the Arellano-Bond approach as $y_{i,0}$ is typically part of the Arellano-Bond moment conditions. However, given the assumption (NSC) that errors are not serially correlated, it is hard to envision a scenario in which assumption (ECI) fails while (NSC) holds. One such scenario would be if the processes $y_{i,t}$ were started at very nonrandom values that had to do with the subsequent error processes $\varepsilon_{i,t}$ (which were, otherwise, serially correlated), but such a situation appears contrived since practitioners usually choose the starting period $t = 0$ for reasons of data availability rather than because the data generating process changes at that date.

The assumption can be accommodated in our estimator by including an additional term $E\left(\hat{\varepsilon}_{i,t}\left(\beta\left(\alpha\right)\right)y_{i,0}\right)$, which is easily computable using the data, in the bias correction. It is, once again, a polynomial in α, so nothing changes qualitatively.

Suppose that we know β_0 and α_0. Then,

$$\hat{\mu}_i(\alpha_0, \beta_0) = \mu_{i,0} + \frac{1}{T}\sum_{t=1}^{T} \varepsilon_{i,t} = \mu_{i,0} + O_p\left(\frac{1}{T}\right)$$

under any combination of the assumptions above. Hence, the empirical fixed effects are unbiased (but not consistent) for the true fixed effects $\mu_{i,0}$.

Now, for any variable $r_{i,t}$, define

$$\hat{r}_{i,t} = r_{i,t} - \frac{1}{T}\sum_{\tau=1}^{T} r_{i,\tau}$$

the "demeaned" version of the variable $r_{i,t}$.

In particular, we have

$$\hat{y}_{i,t} = \alpha_0 \hat{y}_{i,t-1} + \hat{x}'_{i,t}\beta_0 + \hat{\varepsilon}_{i,t}$$

2.2 Coefficients as Functions of Autoregressive Parameter

2.2.1 Case 1: Strictly Exogenous Covariates

In this subsection, we assume that assumption (GM) – that the covariates $x_{i,t}$ are strictly exogenous, and hence uncorrelated with $\varepsilon_{i,\tau}$ in all periods τ, holds. Then, for a given value of α, the best estimator of β_0 in a least-squares sense is

$$\hat{\beta}_{GM}(\alpha) = \left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} \hat{x}_{i,t}\hat{x}'_{i,t}\right)^{-1}\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} \hat{x}_{i,t}\left(\hat{y}_{i,t} - \alpha\hat{y}_{i,t-1}\right)\right)$$

the OLS estimate of the coefficient on the regressors. The estimator $\hat{\beta}_{GM}(\alpha_0)$ is consistent for β_0 because assumption (GM) implies that

$$E\left(\hat{x}_{i,t}\hat{\varepsilon}_{i,t}\right) = 0 \tag{2}$$

(since $x_{i,t}$ is uncorrelated with the leads and lags of $\varepsilon_{i,t}$ as well as with its current value) so

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{x}_{i,t}\hat{\varepsilon}_{i,t}\rightarrow E\left(\hat{x}_{i,t}\hat{\varepsilon}_{i,t}\right)=0$$

and

$$\hat{\beta}_{GM}\left(\alpha_0\right)=\beta_0+\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{x}_{i,t}\hat{x}'_{i,t}\right)^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{x}_{i,t}\hat{\varepsilon}_{i,t}\rightarrow\beta_0$$

Hence, if the true value of the autoregressive parameter were known, the OLS estimate for the coefficient on the regressors in equation (1) would be consistent for β_0. The inconsistency in this estimate is entirely a result of having an inconsistent estimate of α_0.

2.2.2 Case 2: Predetermined Covariates

It is interesting to relax strict exogeneity (assumption (GM)) to the weaker assumption that the covariates are uncorrelated with current and future errors, but may be correlated with past errors (assumption (PR)). Assuming predetermined, instead of strictly exogenous, regressors is natural in the dynamic panel data context, as the lagged dependent variable itself can be thought of as a predetermined regressor. Many dynamic panel data estimators, including those of Arellano and Bond (1991) can deliver consistent estimates of the autoregressive parameter and the coefficient vector in the presence of predetermined regressors.

If we relax assumption (GM) to assumption (PR), equation (2) is no longer true. However, we can instead compute the following estimator for β_0 given a value for α:

$$\hat{\beta}_{PR}\left(\alpha\right)=\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{z}_{i,t}\hat{x}'_{i,t}\right)^{-1}\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{z}_{i,t}\left(\hat{y}_{i,t}-\alpha\hat{y}_{i,t-1}\right)\right)$$

where

$$\begin{aligned}z_{i,t} &= x_{i,t}+\sum_{\tau=t+1}^{T}\left(\frac{1}{T-\tau-1}\right)x_{i,\tau},\ t<T\\ &= x_{i,T},\ t=T\end{aligned}$$

It is straightforward to show that

$$\hat{\beta}_{PR}\left(\alpha_0\right)\rightarrow\beta_0$$

because

$$E\left(x_{i,t}\varepsilon_{i,t'}\right) = E\left(x_{i,t}\hat{\varepsilon}_{i,t'}\right) + \frac{1}{T-(t-1)}\sum_{\tau=1}^{t-1} E\left(x_{i,t}\hat{\varepsilon}_{i,\tau}\right)$$

where $\hat{\varepsilon}_{i,t}$ are the empirical residuals evaluated at $\alpha = \alpha_0$ and $\beta = \beta_0$. The complete derivation of the form of the variable $z_{i,t}$ is presented in the Appendix. It is worth noting that although $\hat{\beta}_{PR}(\alpha)$ is numerically identical to an instrumental variables estimator with $\hat{z}_{i,t}$ as an instrument for $\hat{x}_{i,t}$, the exclusion restriction plainly need not hold, since, for example $E\left(z_{i,T}\varepsilon_{i,1}\right) = E\left(x_{i,T}\varepsilon_{i,1}\right)$ need not be equal to zero. Once again, however, it is clear that even if the regressors are predetermined but not exogenous, the fundamental source of the inconsistency of their estimates lies with having an incorrect value for α.

2.3 Modified FOC for α

Consider the first-order condition for α derived from OLS. We have

$$F_{\alpha}(\alpha) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{i,t} - \alpha y_{i,t-1} - x'_{i,t}\beta(\alpha) - \hat{\mu}_i(\alpha)\right) y_{i,t-1}$$

where $\beta(\alpha)$ and $\hat{\mu}_i(\alpha) = \hat{\mu}_i(\alpha, \beta(\alpha))$ have been definued in the previous subsection. For consistent estimation, we need

$$F_{\alpha}(\alpha_0) = 0$$

However,

$$
\begin{aligned}
F_{\alpha}(\alpha_0) &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{i,t}-\alpha_0 y_{i,t-1}-x_{i,t}'\beta(\alpha_0)-\hat{\mu}_i(\alpha_0)\right)y_{i,t-1} \\
&= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\varepsilon_{i,t}-\frac{1}{T}\sum_{\tau=1}^{T}\varepsilon_{i,\tau}\right)y_{i,t-1}+o_p(1) \\
&= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}\left(\varepsilon_{i,t}-\frac{1}{T}\sum_{\tau=1}^{T}\varepsilon_{i,\tau}\right)\left(\alpha_0^{t-1}y_{i,0}+\sum_{\tau=1}^{t-1}\alpha_0^{\tau-1}x_{i,t-\tau}'\beta_0+\left(\sum_{\tau=1}^{t-1}\alpha_0^{\tau-1}\right)\mu_{i,0}+\sum_{\tau=1}^{t-1}\alpha_0^{\tau-1}\varepsilon_{i,t-\tau}\right)+o_p(1) \\
&\to \frac{1}{T}\sum_{t=2}^{T}E\left[\left(\varepsilon_{i,t}-\frac{1}{T}\sum_{\tau=1}^{T}\varepsilon_{i,\tau}\right)\sum_{\tau=1}^{t-1}\alpha_0^{\tau-1}\varepsilon_{i,t-\tau}\right]\ (I) \\
&\quad +\frac{1}{T}\sum_{t=2}^{T}\left(\sum_{\tau=1}^{t-1}\alpha_0^{\tau-1}\right)E\left[\left(\varepsilon_{i,t}-\frac{1}{T}\sum_{\tau=1}^{T}\varepsilon_{i,\tau}\right)\mu_{i,0}\right]\ (II) \\
&\quad +\frac{1}{T}\sum_{t=2}^{T}E\left(\left(\varepsilon_{i,t}-\frac{1}{T}\sum_{\tau=1}^{T}\varepsilon_{i,\tau}\right)\sum_{\tau=1}^{t-1}\alpha_0^{\tau-1}x_{i,t-\tau}'\beta_0\right)\ (III)
\end{aligned}
$$

So the FOC evaluated at the true value of $\alpha = \alpha_0$ approaches in probability a sum of three terms, not necessarily zero.[4] These terms are all $O\left(\frac{1}{T}\right)$ and comprise the Nickell bias of OLS in the presence of a lagged dependent variable (Nickell 1981). Term I will be nonzero (specifically, negative) under any combination of assumptions discussed earlier in this section. Term II will be zero iff the errors are uncorrelated with the fixed effects (assumption (ECF) holds). Lastly, Term III will be zero under assumption (GM) – strict exogeneity of the regressors – but not under assumption (PR), which allows regressors that are predetermined, but not exogenous.

The term that is always nonzero is (I). It is straightforward to see that as long as the errors are not serially correlated (assumption (NSC) holds)

$$
(I) = -\frac{1}{T^2}\sum_{t=2}^{T}\sum_{\tau=1}^{t-1}\alpha_0^{t-1-\tau}E\left(\varepsilon_{i,\tau}^2\right)
$$

Define the empirical residual as

$$
\begin{aligned}
\hat{\varepsilon}_{i,t}(\alpha) &= \hat{y}_{i,t}-\alpha\hat{y}_{i,t-1}-\hat{x}_{i,t}'\beta(\alpha) \\
&= \hat{y}_{i,t}-\hat{x}_{i,t}'\hat{\beta}_0-\alpha\left(\hat{y}_{i,t-1}-\hat{x}_{i,t}'\hat{\beta}_1\right)
\end{aligned}
$$

[4] Recall that we are assuming $E(y_{i,0}\varepsilon_{i,t})=0$ as an extension of Assumption (NSC). If this assumption is violated, we have an additional term, whose properties are straightforward, and similar to the properties of the other three terms.

where

$$
\begin{aligned}
\hat{\beta}_0 &= \left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{x}_{i,t}\hat{x}_{i,t}'\right)^{-1}\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{x}_{i,t}\hat{y}_{i,t}\right) \\
\hat{\beta}_1 &= \left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{x}_{i,t}\hat{x}_{i,t}'\right)^{-1}\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{x}_{i,t}\hat{y}_{i,t-1}\right)
\end{aligned}
$$

or the equivalent of these terms under assumption PR (predetermined regressors), and note that

$$
\hat{\varepsilon}_{i,t}(\alpha_0) = \varepsilon_{i,t} - \frac{1}{T}\sum_{\tau=1}^{T}\varepsilon_{i,\tau} + o_p(1)
$$

Then, we can estimate the quantities $E\left(\varepsilon_{i,\tau}^2\right)$ as a function of α as follows:

$$
E\left(\varepsilon_{i,t}^2\right) \leftarrow \frac{1}{N}\sum_{i=1}^{N}\left(\frac{T}{T-2}\left(\hat{\varepsilon}_{i,t}^2(\alpha_0) - \frac{1}{T-1}\frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_{i,t}^2(\alpha_0)\right)\right)
$$

This approach is similar to the result of Stock and Watson (2008) for the estimation of standard errors in a panel setting with fixed T.

In particular, rather than being a potentially more complicated function, term (I) is a polynomial in α_0 of order T. We note that

$$
\hat{\varepsilon}_{i,t}(\alpha) = r_{i,t}^0 - \alpha r_{i,t}^1
$$

where $r_{i,t}^0$ and $r_{i,t}^1$ are residuals from regressions of $\hat{y}_{i,t}$ on $\hat{x}_{i,t}$ and $\hat{y}_{i,t-1}$ on $\hat{x}_{i,t}$, respectively (instrumented by $\hat{z}_{i,t}$ when we assume predetermined, rather than strictly exogenous, regressors, or relax assumption GM to assumption PR).

Then, we define the following moments of residuals:

$$
\begin{aligned}
R_t^1 &= \frac{T}{T-2}\frac{1}{N}\sum_{i=1}^{N}\left[\left(r_{i,t}^1\right)^2 - \frac{1}{T-1}\frac{1}{T}\sum_{t=1}^{T}\left(r_{i,t}^1\right)^2\right] \\
R_t^\rho &= \frac{T}{T-2}\frac{1}{N}\sum_{i=1}^{N}\left[r_{i,t}^0 r_{i,t}^1 - \frac{1}{T-1}\frac{1}{T}\sum_{t=1}^{T} r_{i,t}^0 r_{i,t}^1\right] \\
R_t^0 &= \frac{T}{T-2}\frac{1}{N}\sum_{i=1}^{N}\left[\left(r_{i,t}^0\right)^2 - \frac{1}{T-1}\frac{1}{T}\sum_{t=1}^{T}\left(r_{i,t}^0\right)^2\right]
\end{aligned}
$$

It follows straightforwardly that

$$
P_1(\alpha_0) := -\frac{1}{T^2}\left(\sum_{t=2}^{T}\left(\sum_{\tau=1}^{T-t+1} R_\tau^1\right)\alpha_0^t - 2\sum_{t=1}^{T-1}\left(\sum_{\tau=1}^{T-t} R_\tau^\rho\right)\alpha_0^t + \sum_{t=0}^{T-2}\left(\sum_{\tau=1}^{T-t-1} R_\tau^0\right)\alpha_0^t\right) \to (I)
$$

and we can rewrite the modified first order condition as

$$
P_1(\alpha_0) - \alpha_0 + \alpha_{OLS} = 0
$$

or

$$
\tilde{P}_1(\alpha_0) = 0
$$

The fact that the modified first-order condition in α takes the form of a polynomial makes our estimator tractable, as it does not involve numerically solving an equation or maximizing a criterion function, where the existence and uniqueness of roots, as well as the convergence properties of most root-finding algorithms are not generally known. Instead, we obtain exactly T roots, some imaginary and some real.

As N goes to infinity, $\tilde{P}_1(\alpha)$ should have at least one real root – at α_0. However, in finite samples, $\tilde{P}_1(\alpha)$ may not have any real roots. Therefore, we also consider values of α that are local minima of $(P_1(\alpha))^2$, or, equivalently, solve $P_1'(\alpha) = 0$ subject to $P_1''(\alpha) > 0$.

We then face the problem of finding which member of our solution set to select as our estimate.[5] One straightforward approach is to select the root that is closest to another, consistent estimator of α_0. We will present simulations using an infeasible version of the

[5] While we do not have a criterion function, as in maximum likelihood, to select the root that attains the global maximum, we can exhaustively catalogue the candidate roots, while this is generally not possible to do with a likelihood function that is not globally concave.

estimator in which we select the root that is closest to the true value of the autoregressive parameter used to construct the simulation, as well as using the root that is closest to an estimator based on instrumenting the lagged dependent variable with lags of strictly exogenous regressors.

3 Extensions of the Basic Estimator

As alluded to in the previous subsection, we can easily relax all of the assumptions under which Terms II and III are nonzero by approximating them in ways that are similar to the approximation of Term I. All of the approximations are polynomials in α, because consistent estimators of the error variances, of moments of fixed effects and of interactions between predetermined covariates and errors are linear or quadratic in α. We discuss the construction of these approximations as polynomials of α below:

3.1 Term II

Term (II) is nonzero iff we have $E\left(\mu_{i,0}\varepsilon_{i,t}\right) \neq 0$ (failure of assumption (ECF), which states that errors and fixed effects are uncorrelated). It can also be estimated straightforwardly, since

$$\frac{1}{N}\sum_{i=1}^{N}\hat{\varepsilon}_{i,t}\left(\alpha_0\right)\hat{\mu}_{i,0}\left(\alpha_0\right) = \frac{1}{N}\sum_{i=1}^{N}\left(\varepsilon_{i,t} - \frac{1}{T}\sum_{\tau=1}^{T}\varepsilon_{i,\tau}\right)\left(\mu_{i,0} + \frac{1}{T}\sum_{t=1}^{T}\varepsilon_{i,t}\right)$$

where we recall that

$$\hat{\mu}_{i,0}\left(\alpha_0\right) = \hat{\mu}_i\left(\alpha_0, \beta\left(\alpha_0\right)\right) = \frac{1}{T}\sum_{t=1}^{T}\left(y_{i,t} - \alpha_0 y_{i,t-1} - x'_{i,t}\beta_0\right)$$

The second term can be further analyzed as

$$\frac{1}{T}\sum_{\tau=1}^{T}\frac{1}{N}\sum_{i=1}^{N}\left(\varepsilon_{i,t} - \frac{1}{T}\sum_{t'=1}^{T}\varepsilon_{i,t'}\right)\varepsilon_{i,\tau} \rightarrow \frac{1}{T}E\left(\varepsilon_{i,t}^2\right) - \frac{1}{T^2}\sum_{\tau=1}^{T}E\left(\varepsilon_{i,\tau}^2\right)$$

and then the entire sum of second terms becomes

$$\frac{1}{N}\sum_{i=1}^{N}\left(\varepsilon_{i,t}-\frac{1}{T}\sum_{\tau=1}^{T}\varepsilon_{i,\tau}\right)\mu_{i,0}=\frac{1}{N}\sum_{i=1}^{N}\hat{\varepsilon}_{i,t}\hat{\mu}_{i}\left(\alpha_{0}\right)-\frac{1}{T}\sum_{t=2}^{T}\left(\sum_{\tau=1}^{t-1}\alpha_{0}^{\tau-1}\right)\left[\frac{1}{T}E\left(\varepsilon_{i,t}^{2}\right)-\frac{1}{T^{2}}\sum_{\tau=1}^{T}E\left(\varepsilon_{i,\tau}^{2}\right)\right]$$

where $E\left(\varepsilon_{i,t}^{2}\right)$ is estimated as in term (I), so everything on the right hand-side is estimable.

We can consequently express term (II) as another polynomial in α of order T. First, we define

$$\begin{aligned}\tilde{R}_{t}^{1} &= \frac{1}{N}\sum_{i=1}^{N}r_{i,t}^{1}\tilde{r}_{i}^{1}\\ \tilde{R}_{t}^{1/2} &= \frac{1}{N}\sum_{i=1}^{N}(1/2)\left(r_{i,t}^{0}\tilde{r}_{i}^{1}+r_{i,t}^{1}\tilde{r}_{i}^{0}\right)\\ \tilde{R}_{t}^{0} &= \frac{1}{N}\sum_{i=1}^{N}r_{i,t}^{0}\tilde{r}_{i}^{0}\end{aligned}$$

where $\tilde{r}_{i}^{0}$ and $\tilde{r}_{i}^{1}$ are the panel unit fixed effects from the regressions generating $\hat{\beta}_{0}$ and $\hat{\beta}_{1}$ respectively. Then, we define

$$\bar{R}_{t}^{k}=\frac{1}{N}\sum_{i=1}^{N}\left(\left(r_{i,t}^{0}\right)^{2-2k}\left(r_{i,t}^{1}\right)^{2k}-\frac{1}{T}\sum_{\tau=1}^{T}\left(r_{i,t}^{0}\right)^{2-2k}\left(r_{i,t}^{1}\right)^{2k}\right),\ k=0,1/2,1$$

and

$$Z_{t}^{k}=\tilde{R}_{t}^{k}-\frac{1}{T}\bar{R}_{t}^{k}$$

It is then easy to see that

$$P_{2}\left(\alpha_{0}\right):=\frac{1}{T}\sum_{t=2}^{T}\left(\sum_{\tau=t}^{T}Z_{\tau}^{1}\right)\alpha_{0}^{t}-2\frac{1}{T}\sum_{t=1}^{T-1}\left(\sum_{\tau=t+1}^{T}Z_{\tau}^{\rho}\right)\alpha_{0}^{t}+\frac{1}{T}\sum_{t=0}^{T-2}\left(\sum_{\tau=t+2}^{T}Z_{\tau}^{0}\right)\alpha_{0}^{t}\rightarrow(II)$$

where $P_{2}\left(\alpha_{0}\right)$ is a polynomial in α_{0} of order T.

3.2 Term III

Next, we may need to estimate term (III) if assumption (GM) – the strict exogeneity of the regressors – does not hold, but assumption (PR) does (regressors are not strictly exogenous, but are predetermined). Then

$$(IV) \rightarrow -\frac{1}{T}\sum_{t=2}^{T}\left(\frac{1}{T}\sum_{\tau'=1}^{t-1}\alpha_0^{t-\tau'-1}\left(\sum_{\tau=1}^{\tau'-1}E\left(\varepsilon_{i,\tau}x'_{i,\tau'}\right)\right)\right)\beta_0$$

and we can estimate $E\left(\varepsilon_{i,\tau}x'_{i,\tau'}\right)$ for $\tau' > \tau$ by the formula

$$E\left(x_{i,t}\varepsilon_{i,t'}\right) = E\left(x_{i,t}\hat{\varepsilon}_{i,t'}\left(\alpha_0\right)\right) + \frac{1}{T-(t-1)}\sum_{\tau=1}^{t-1}E\left(x_{i,t}\hat{\varepsilon}_{i,\tau}\left(\alpha_0\right)\right)$$

which is derived in Appendix I as part of the general estimator for predetermined variables. If we define

$$f_{i,t}^{k} = x_{i,t}\hat{\beta}_k$$

where $k \in \{0,1\}$ as before, and we define

$$\begin{aligned}
\tilde{X}_t^0 &= \left(\frac{T}{T-(t-1)}\right)\frac{1}{N}\sum_{i=1}^{N}\sum_{t'=1}^{t-1}f_{i,t}^0 r_{i,t'}^0 \\
\tilde{X}_t^{1/2} &= \left(\frac{T}{T-(t-1)}\right)\frac{1}{N}\sum_{i=1}^{N}\sum_{t'=1}^{t-1}(1/2)\left(f_{i,t}^0 r_{i,t'}^1 + f_{i,t}^1 r_{i,t'}^0\right) \\
\tilde{X}_t^1 &= \left(\frac{T}{T-(t-1)}\right)\frac{1}{N}\sum_{i=1}^{N}\sum_{t'=1}^{t-1}f_{i,t}^1 r_{i,t'}^1
\end{aligned}$$

we can easily show that

$$P_3\left(\alpha_0\right) := -\left(\frac{1}{T^2}\sum_{t=2}^{T-1}\left(\sum_{\tau=2}^{T-t+1}\tilde{X}_\tau^1\right)\alpha_0^t - 2\frac{1}{T^2}\sum_{t=1}^{T-2}\left(\sum_{\tau=2}^{T-t}\tilde{X}_\tau^{1/2}\right)\alpha_0^t + \frac{1}{T^2}\sum_{t=0}^{T-3}\left(\sum_{\tau=2}^{T-t-1}\tilde{X}_\tau^1\right)\alpha_0^t\right) \rightarrow (III)$$

another polynomial of order T.

4 Simulations

We run simulations to illustrate the properties of our new estimator. All of these simulations involved the model

$$y_{i,t} = \alpha_0 y_{i,t-1} + \beta_0 x_{i,t} + \mu_{i,0} + \varepsilon_{i,t}$$

with various assumptions. We use the version of our estimator that includes Terms I and II, and we consider a version of our estimator that also includes term III when we investigate the estimates under predetermined variables. We typically compute two versions of our estimator: an infeasible estimator, where we select the root that is closest to the true value of α_0 to be our estimate; and a feasible estimator, where we select the root that is closest to the estimate of α_0 provided by instrumenting the lagged dependent variable with lags of $x_{i,t}$[6]. The second approach requires that $E\left(x_{i,t}\varepsilon_{i,s}\right) = 0$ for all s, t, which is equivalent to $x_{i,t}$ being strictly exogenous. Hence, in specifications involving predetermined regressors, we select the root that is closest to the estimate of α_0 provided by instrumenting with $x_{i,0}$, the only lag of $x_{i,t}$ that is uncorrelated with all the error terms.

4.1 Stationary Initial Condition

We assume that

$$\mu_i \,\tilde{}\, N(0,1), \text{ iid}$$
$$\varepsilon_{i,t} \,\tilde{}\, N(0,1), \text{ iid}$$
$$x_{i,t} \,\tilde{}\, N(\mu_i,1), \text{ iid}$$

We set $\beta_0 = 1$ and allow α_0 to take values from the set $\{0.25, 0.5, 0.75, 0.9, 0.95, 0.99\}$. This set enables us to see the performance of our estimator for a wide variety of autoregressive parameters,

Table I presents simulation results in which we draw $y_{i,0}$ from the stationary distribution of this process, specifically

$$y_{i,0} \,\tilde{}\, N\left(\frac{1+\beta_0}{1-\alpha_0}\mu_i, \frac{1+\beta_0^2}{1-\alpha_0^2}\right) \text{ iid}$$

Each row of Table I presents summary statistics of the distribution of Monte Carlo estimates for the given method (in the panel) and the given value of α_0 (in the row). The first panel presents results that use the Arellano-Bond estimator. We see from the first row, where $\alpha_0 = 0.25$, that the mean of the estimates in the simulations is 0.246 with a standard deviation of 0.025. Hence, for $\alpha_0 = 0.25$, a low value of α_0, Arellano-Bond performs well,

[6] Specifically, we construct our instruments as

$$z_{i,t}^j = x_{i,t-j} \cdot (t \geq j), \text{ for } j = 1, ..., T$$

since the process is far from a random walk in $y_{i,t}$ and the change in the dependent variable is correlated with its level. However, if we look at the row with $\alpha_0 = 0.99$, we see that the mean of the estimates in the simulations is 0.58 with a standard deviation of 0.308, very far from the actual value of α_0 and with large variations across the data samples. It is clear that as α_0 becomes larger, Arellano-Bond delivers downward biased estimates with large standard errors. For large values of α_0, the bias in α affects the measurement of β, causing it to be biased away from β_0. For example, when $\alpha_0 = 0.99$, the mean of the estimates of β is 0.795, given that $\beta_0 = 1$.

Proceeding similarly, we see that the method of Blundell and Bond (1998) in the second panel and the method of Bai (2013) in the fifth panel deliver consistent estimates of α_0 with fairly low MSE, for low as well as high values of α_0. For example, when $\alpha_0 = 0.99$, the mean of the Blundell-Bond estimates is 0.994 and the mean of the Bai estimates is 0.991. The standard deviation of the Bai estimator is 0.023, while the standard deviation of the Blundell-Bond estimator is a much smaller 0.002. In the third panel of the table, we use the infeasible version of our estimator (in which the closest root to the true value is picked as the estimator). First, we see that the resulting estimates are consistent for α_0, with the mean of the estimates for each value of α_0 differing from the true value by no more than 0.005. The standard deviations of the estimates are heterogeneous, and vary with α_0. For low and high values of α_0 ($\alpha_0 = 0.25, 0.5$ and 0.99), the standard deviations produced by the infeasible estimator are close to or below those of the Blundell-Bond estimator (for $\alpha_0 = 0.25$ and 0.5) or the Bai estimator ($\alpha_0 = 0.99$). However, the standard deviations of the estimates for $\alpha_0 = 0.75$, 0.9 or 0.95 are twice or three times as large as the standard deviations produced by the Blundell and Bond estimator or the Bai estimator for these values of α_0. This can be entirely explained by the inclusion of Term II (which accounts for correlation between errors and fixed effects, which is not present here), and if Term II were removed, the standard deviations produced by the infeasible estimator would be virtually identical to those of the Bai estimator. In any case, the standard deviations produced by the infeasible estimator are considerably smaller than the standard deviations of the Arellano-Bond estimator. If we select the closest root to the "simple IV" estimator, the simulation results are almost identical to the infeasible case. Hence, for data simulated from this stationary distribution, our estimator yields consistent estimates unlike the Arellano-Bond estimator, and performs comparably to the Bai (2013) estimator and the Blundell-Bond (1998) estimator in terms of mean squared error.

4.2 Nonstationary Initial Condition

Table II presents simulation results in which we draw $y_{i,0}$ from the nonstationary distribution

$$y_{i,0} \tilde{} N(2\mu_i, 4/3) \text{ iid}$$

following Blundell and Bond (1998). Here, the Arellano-Bond estimator delivers consistent estimates with low standard errors, as does the Bai (2013) estimator. The MSEs range from 0.025 to 0.006 for the Arellano-Bond estimator (declining with α_0) and from 0.012 to 0.004 for the Bai estimator (similarly). On the other hand, the Blundell-Bond (1998) estimator performs poorly, generating upward-biased estimates (though with low standard errors). For example, for $\alpha_0 = 0.25$, the mean of the Blundell-Bond estimates is 0.4 (with a standard error of 0.012) and for $\alpha_0 = 0.99$, it is 1.086 (with a standard error of 0.002). The infeasible version of our approach yields consistent estimates for all values of α_0 (the mean of the estimates is within 0.001 of the true α_0) and has better MSE than does the Arellano-Bond procedure (though slightly worse MSE than the Bai (2013) estimator, although the differences are small enough to be simulation error). The feasible version (starting from the "simple-IV" estimates) is identical to the infeasible version. It is also notable that, while both the Arellano-Bond estimator and our estimator produce consistent estimates of the covariate coefficient β, our estimator yields lower RMSE than does Arellano-Bond.

The virtue of our approach (which, up to this point, it shares with the Bai (2013) estimator) is that it delivers consistent estimates of α_0 and β_0 regardless of whether the initial condition of the dynamic process is stationary or nonstationary. We have seen in Table I that Arellano-Bond is inconsistent with large MSE when the initial distribution is stationary, and we have seen here that Blundell-Bond is inconsistent with large MSE when it is nonstationary. Our approach allows for consistent estimation without having to pre-test for the stationarity of the distribution, which introduces additional variance into the procedure.

4.3 Correlated Fixed Effects and Errors

While the estimator of Bai (2013) performs as well (or slightly better) than our estimator in the two settings considered above, it relies on the assumption that the errors of the dynamic process are uncorrelated with the fixed effects (assumption ECF in Section 1). In this simulation, we consider data that does not satisfy this assumption. We use the nonstationary distribution from the nonstationary simulation exercise, but also define the fixed effect as

$$\mu_i = \tilde{\mu}_i + \varepsilon_{i,1}$$

and

$$\tilde{\mu}_i \tilde{} N(0,1) \text{ iid}$$

while drawing

$$x_{i,t} \tilde{} N(\tilde{\mu}_i, 1), \text{ iid}$$

$$y_{i,0} \tilde{} N(2\tilde{\mu}_i, 4/3), \text{ iid}$$

to avoid making the regressors be predetermined. Here, we no longer consider the Blundell-Bond method as we know that it does not work well when the initial distribution is nonstationary. We present the simulation results in Table III. We see that Arellano-Bond delivers consistent estimates with low RMSE that are very similar to the ones in Table II. On the other hand, the Bai (2013) estimator delivers estimates that are biased downwards, with the bias being particularly severe for low values of α_0. For example, for $\alpha_0 = 0.25$, the mean of the Bai (2013) estimates is 0.11 with a standard error of 0.012. If we do not include Term (II) in our estimator (but include only Term I), our estimator yields estimates that are also biased downwards in a very similar way to the Bai (2013) estimator. However, once we include the correction (Term II), our estimates become consistent, with somewhat smaller variance than the Arellano-Bond estimates (for example, if $\alpha_0 = 0.25$, the MSE of our estimator that uses the "simple IV" estimate to select the root is 0.015, while the MSE of the Arellano-Bond estimator is 0.025). Therefore, our estimator delivers consistent estimates with low MSE in cases where the Bai (2013) estimator is biased and inconsistent. We can also see that the standard deviations of the Bai estimator and of the version of our estimator that includes only term (I) are nearly identical, suggesting that including Term II when assumption (ECF) holds, and the errors are uncorrelated with the fixed effects (such as in Tables (I) and (II)) increases the variance of our estimator.[7]

4.4 Predetermined Regressors

Lastly, we investigate how our approach performs when the regressors are predetermined, but not exogenous. The fourth table also starts with the nonstationary distribution simulation, but makes $x_{i,t}$ be predetermined. Specifically, we define

$$\tilde{x}_{i,t} \tilde{} N(\mu_i, 1), \text{ iid}$$

and

$$x_{i,t} = \tilde{x}_{i,t} + \varepsilon_{i,t-1}$$

The Bai (2013) estimator is equipped to handle only regressors that are autoregressive, which is not the case in this simulation, but the Arellano-Bond estimator can accommodate arbitrary predetermined regressors if the econometrician appropriately specifies them. Therefore, we consider how well our estimator performs relative to the Arellano-Bond estimator. Since $x_{i,t}$ is predetermined and $\tilde{x}_{i,t}$ is unobservable, we base our feasible estimator on a modified version of the the "simple IV" estimator, which just includes the initial values of $x_{i,t}$ interacted with year fixed effects as instruments for $y_{i,t-1}$. It is clear that these instruments satisfy the exclusion restriction, since $x_{i,0}$ is uncorrelated with any $\varepsilon_{i,t}$ for $t \geq 1$.

We present the simulation results in Table IV. The version of our estimator (feasible or infeasible) that includes Term III deliver consistent estimates with acceptably low RMSE,

[7] An extension of our estimator could involve pre-testing for whether assumption (ECF) holds or fails, and including Term (II) in our estimator only when the test fails to reject the null that assumption (ECF) holds.

although for middling values of α_0, such as $\alpha_0 = 0.75$, the RMSE is 0.069 for the infeasible estimator and 0.078 for the feasible one. The Arellano-Bond estimator also delivers consistent estimates with lower MSE than does our estimator. On the other hand, the version of our estimator that includes only Terms I and II delivers estimates that are downward biased, especially for low values of α_0. In particular, when $\alpha_0 = 0.25$, the mean of the estimates is 0.15. The general predetermined correction described in the Appendix works very well in obtaining a consistent estimate of β, with the mean of the estimates being essentially at the true value of 1; the uncorrected estimator yields estimates of β below 0.91 on average.

5 Conclusion

A persistent problem in dynamic panel data analysis is finding an estimator that performs well in many different data settings under minimal assumptions. For example, the Arellano-Bond (1991) estimator tends to perform poorly when the underlying process comes from a stationary distribution and the dependent variable follows a near random walk. Attempts to improve the Arellano-Bond estimator, such as Blundell and Bond (1991), tend to perform poorly in situations when Arellano-Bond performs well.

We propose a new estimator for linear dynamic panel data models with serially uncorrelated errors that is less sensitive to the distribution of initial values than are the popular Arellano and Bond (1991) and Blundell and Bond (1998) estimators, and that does not rely on any additional assumptions about the canonical model. The approach behind this estimator is not to use linear moment restrictions to find instruments for the lagged dependent variable, but to compute the bias in the first-order condition for the autoregressive parameter α, and use it to obtain a modified first-order condition that is an unbiased estimator of zero when $\alpha = \alpha_0$. Our estimator performs well in simulations regardless of the initial distribution of the outcome variable. We also compare our estimator with the recent estimator by Bai (2013), which also delivers consistent estimates in the fixed-T setting regardless of initial conditions. We find that our estimator performs just about as well as the Bai (2013) estimator in terms of MSE when model errors and panel unit fixed effects are uncorrelated. However, when they are correlated, our estimator continues to deliver consistent estimates with low MSE, while the Bai estimator is biased and inconsistent.

While our methodology is presented for the case with only one lag of the dependent variable in the true model, it is straightforward to extend our approach to generate a system of polynomial equations that would be a consistent estimator of zero at the true values of the lag coefficients. We leave this extension and its implementation to future research.

References

[1] **Ahn, Seung C.**, and **Peter Schmidt**. 1995. "Efficient Estimation of Models for Dynamic Panel Data." *Journal of Econometrics* 68, no. 1: 5-27.

[2] **Anderson, T. W., and Cheng Hsiao**. 1982. "Formulation and Estimation of Dynamic Models Using Panel Data." *Journal of Econometrics* 18, no. 1: 47-82.

[3] **Arellano, Manuel, and Stephen Bond.** 1991. "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations." *Review of Economic Studies* 58, no. 2: 277-297.

[4] **Bai, Jushan**. 2013. "Fixed-Effects Dynamic Panel Models: A Factor Analytical Method." *Econometrica* 81, no. 1: 285-314.

[5] **Blundell, Richard, and Stephen Bond**. 1998. "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models." *Journal of Econometrics* 87, no. 1: 115-143.

[6] **Hahn, Jinyong**. 1999. "How Informative Is the Initial Condition in the Dynamic Panel Model with Fixed Effects?." *Journal of Econometrics* 93, no. 2: 309-326.

[7] **Hahn, Jinyong, Jerry Hausman, and Guido Kuersteiner**. 2007. "Long Difference Instrumental Variables Estimation for Dynamic Panel Models with Fixed Effects." *Journal of Econometrics* 140, no. 2: 574-617.

[8] **Hsiao, Cheng, M. Hashem Pesaran, and A. Kamil Tahmiscioglu**. 2002. "Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods." *Journal of Econometrics* 109, no. 1: 107-150.

[9] **Stock, James H., and Mark W. Watson.** 2008. "Heteroskedasticity-robust standard errors for fixed effects panel data regression." *Econometrica* 76, no. 1: 155-174.

6 Tables

Table I (I)

Simulations of α and β: Stationary Initial Condition

T=5, N=1000, Distribution of y_0 is stationary. There are 1000 replications.

α_0	Mean, α	SD, α	RMSE, α	Median, α	IDR, α	Mean, β	SD, β	RMSE, β	Median, β	IDR, β
Arellano-Bond (1991)										
0.25	.247	.021	.021	.247	.057	.998	.019	.019	.998	.050
0.50	.498	.032	.032	.498	.083	.998	.023	.023	.997	.060
0.75	.739	.051	.052	.741	.128	.994	.028	.029	.995	.072
0.90	.833	.126	.143	.833	.316	.968	.064	.071	.968	.161
0.95	.765	.210	.280	.783	.518	.909	.106	.139	.917	.259
0.99	.580	.308	.512	.617	.784	.795	.155	.256	.814	.395
Blundell-Bond (1998)										
0.25	.254	.020	.020	.252	.049	1.003	.019	.019	1.003	.048
0.50	.506	.023	.024	.506	.058	1.002	.021	.021	1.002	.055
0.75	.758	.024	.025	.761	.062	1.002	.019	.019	1.002	.048
0.90	.918	.018	.026	.919	.040	1.005	.018	.019	1.005	.047
0.95	.968	.011	.021	.968	.027	1.005	.017	.018	1.005	.045
0.99	.994	.002	.005	.994	.005	1.001	.017	.017	1.001	.044
Hausman-Pinkovskiy (2017) Infeasible										
0.25	.250	.014	.014	.251	.036	.999	.015	.015	.998	.038
0.50	.499	.017	.017	.501	.043	.999	.016	.016	1.000	.043
0.75	.750	.023	.023	.751	.058	1.000	.018	.018	1.000	.046
0.90	.894	.041	.041	.902	.087	.997	.024	.024	.998	.059
0.95	.955	.037	.037	.960	.086	1.002	.023	.023	1.003	.057
0.99	.994	.010	.011	.994	.017	1.001	.016	.016	1.002	.040
Hausman-Pinkovskiy (2017) Simple IV										
0.25	.250	.014	.014	.251	.036	.999	.015	.015	.998	.038
0.50	.499	.017	.017	.501	.043	.999	.016	.016	1.000	.043
0.75	.750	.023	.023	.751	.058	1.000	.018	.018	1.000	.046
0.90	.900	.044	.044	.904	.082	.999	.024	.024	1.000	.057
0.95	.957	.038	.039	.961	.085	1.003	.024	.024	1.004	.058
0.99	.993	.013	.013	.994	.017	1.000	.016	.016	1.001	.040
Bai (2013)										
0.25	.249	.013	.013	.248	.033	.999	.015	.016	.998	.040
0.50	.500	.015	.015	.500	.038	.999	.017	.017	.999	.042
0.75	.751	.016	.016	.750	.041	1.000	.017	.017	.999	.044
0.90	.901	.019	.019	.900	.051	1.000	.017	.017	1.000	.045
0.95	.950	.021	.021	.950	.056	1.000	.018	.018	1.000	.049
0.99	.991	.023	.023	.990	.061	1.001	.020	.020	1.001	.051

This table presents simulation results for the model described in Section 4.1. IDR refers to the difference between the 90th and the 10th percentiles of the coefficient in question.

Simulations of α and β: Nonstationary Initial Condition *T=5, N=1000. Distribution of y_0 is $N(2\mu_i, 4/3)$. There are 1000 replications.*										
α_0	Mean, α	SD, α	RMSE, α	Median, α	IDR, α	Mean, β	SD, β	RMSE, β	Median, β	IDR, β
Arellano-Bond (1991)										
0.25	.246	.025	.025	.247	.067	.998	.020	.020	.998	.052
0.50	.495	.035	.035	.497	.091	.997	.025	.025	.997	.064
0.75	.749	.012	.012	.749	.033	.999	.019	.019	.998	.050
0.90	.899	.007	.007	.899	.020	1.000	.018	.018	.999	.047
0.95	.949	.006	.006	.949	.017	.999	.018	.018	1.000	.049
0.99	.989	.006	.006	.989	.015	1.000	.018	.018	1.000	.047
Blundell-Bond (1998)										
0.25	.402	.012	.152	.402	.032	1.083	.018	.085	1.084	.046
0.50	.683	.007	.183	.683	.018	1.094	.019	.096	1.095	.051
0.75	.894	.004	.144	.894	.010	1.120	.017	.121	1.120	.044
0.90	1.014	.003	.114	1.014	.009	1.155	.018	.156	1.155	.046
0.95	1.054	.003	.104	1.054	.008	1.169	.018	.170	1.169	.045
0.99	1.086	.002	.096	1.086	.007	1.181	.017	.182	1.181	.044
Hausman-Pinkovskiy (2017) Infeasible										
0.25	.250	.014	.014	.251	.037	.999	.015	.015	.998	.038
0.50	.499	.014	.014	.500	.037	.999	.016	.016	.999	.041
0.75	.750	.010	.010	.750	.025	1.000	.016	.016	1.001	.043
0.90	.900	.006	.006	.900	.017	.999	.016	.016	.999	.040
0.95	.950	.005	.005	.950	.014	1.000	.016	.016	1.000	.042
0.99	.990	.005	.005	.989	.012	.999	.015	.015	.999	.039
Hausman-Pinkovskiy (2017) Simple IV										
0.25	.250	.014	.014	.251	.037	.999	.015	.015	.998	.038
0.50	.499	.014	.014	.500	.037	.999	.016	.016	.999	.041
0.75	.750	.010	.010	.750	.025	1.000	.016	.016	1.001	.043
0.90	.900	.006	.006	.900	.017	.999	.016	.016	.999	.040
0.95	.950	.005	.005	.950	.014	1.000	.016	.016	1.000	.042
0.99	.990	.005	.005	.989	.012	.999	.015	.015	.999	.039
Bai (2013)										
0.25	.249	.012	.012	.249	.034	.999	.015	.015	.998	.039
0.50	.500	.012	.012	.500	.032	.999	.016	.016	.999	.043
0.75	.750	.008	.008	.750	.021	1.000	.016	.016	.999	.041
0.90	.900	.006	.006	.900	.015	1.000	.015	.015	1.000	.041
0.95	.949	.005	.005	.949	.013	1.000	.016	.016	1.000	.041
0.99	.989	.004	.004	.989	.013	1.000	.016	.016	1.000	.042

This table presents simulation results for the model described in Section 4.2. IDR refers to the difference between the 90th and the 10th percentiles of the coefficient in question.

Table III (III)

Simulations of α and β: Correlation Between Errors and Fixed Effects *T=5, N=1000. NS distribution is* $N(2\tilde{\mu}_i, 4/3)$. *Fixed effect is* $\mu_i = \tilde{\mu}_i + \varepsilon_{i,1}$. There are 1000 reps										
α_0	Mean, α	SD, α	RMSE, α	Median, α	IDR, α	Mean, β	SD, β	RMSE, β	Median, β	IDR, β
Arellano-Bond (1991)										
0.25	.247	.025	.025	.246	.064	.998	.020	.020	.998	.052
0.50	.498	.034	.034	.497	.089	.998	.024	.024	.998	.064
0.75	.749	.012	.012	.748	.032	.999	.018	.018	.998	.048
0.90	.899	.007	.007	.899	.018	.999	.018	.018	.999	.046
0.95	.950	.006	.006	.949	.015	.999	.018	.018	.999	.047
0.99	.990	.005	.005	.990	.013	1.000	.018	.018	1.000	.047
Hausman-Pinkovskiy (2017): No Correction, Infeasible										
0.25	.114	.010	.135	.114	.026	.966	.015	.036	.966	.039
0.50	.366	.009	.133	.366	.026	.959	.014	.043	.959	.037
0.75	.667	.006	.082	.667	.016	.967	.015	.036	.967	.039
0.90	.848	.005	.051	.848	.012	.976	.015	.027	.976	.038
0.95	.907	.004	.043	.907	.012	.979	.015	.025	.979	.040
0.99	.952	.004	.038	.952	.011	.981	.014	.023	.980	.037
Hausman-Pinkovskiy (2017): Correlation Correction, Infeasible										
0.25	.249	.015	.015	.248	.038	.999	.016	.016	.999	.042
0.50	.500	.018	.018	.500	.046	1.000	.015	.016	1.001	.040
0.75	.749	.011	.011	.749	.028	.999	.016	.016	1.000	.042
0.90	.900	.006	.006	.899	.017	.999	.015	.015	.999	.039
0.95	.950	.005	.005	.950	.015	1.000	.016	.016	1.000	.042
0.99	.990	.005	.005	.990	.013	1.000	.015	.015	.999	.038
Hausman-Pinkovskiy (2017): Correlation Correction, Simple IV										
0.25	.249	.015	.015	.248	.038	.999	.016	.016	.999	.042
0.50	.500	.018	.018	.500	.046	1.000	.015	.016	1.001	.040
0.75	.749	.011	.011	.749	.028	.999	.016	.016	1.000	.042
0.90	.900	.006	.006	.899	.017	.999	.015	.015	.999	.039
0.95	.950	.005	.005	.950	.015	1.000	.016	.016	1.000	.042
0.99	.990	.005	.005	.990	.013	1.000	.015	.015	.999	.038
Bai (2013)										
0.25	.111	.012	.138	.111	.032	.965	.015	.037	.965	.040
0.50	.365	.010	.135	.365	.026	.958	.015	.044	.958	.037
0.75	.672	.007	.077	.672	.019	.969	.014	.033	.969	.036
0.90	.855	.005	.044	.855	.015	.979	.015	.025	.980	.039
0.95	.913	.005	.036	.913	.013	.982	.015	.023	.982	.040
0.99	.959	.004	.031	.959	.012	.985	.015	.021	.985	.040

This table presents simulation results for the model described in Section 4.3. IDR refers to the difference between the 90th and the 10th percentiles of the coefficient in question.

Table IV (IV)

Simulations of α and β: Predetermined Variables

T=5, N=1000. y_0 distribution is $N(2\mu_i, 4/3)$. Covariate is $x_{i,t} = \tilde{x}_{i,t} + \varepsilon_{i,t-1}$. There are 1000 reps.

α_0	Mean, α	SD, α	RMSE, α	Median, α	IDR, α	Mean, β	SD, β	RMSE, β	Median, β	IDR, β
Arellano-Bond (1991)										
0.25	.244	.017	.018	.244	.043	.996	.017	.017	.996	.044
0.50	.493	.022	.023	.492	.060	.994	.023	.024	.995	.062
0.75	.746	.012	.013	.746	.032	.995	.021	.021	.996	.054
0.90	.898	.007	.007	.898	.019	.997	.017	.018	.996	.044
0.95	.949	.006	.006	.948	.016	.997	.018	.018	.997	.046
0.99	.989	.005	.006	.989	.015	.997	.017	.017	.997	.044
Hausman-Pinkovskiy (2017): No Correction, Infeasible										
0.25	.148	.027	.104	.156	.081	.907	.012	.093	.907	.031
0.50	.334	.008	.166	.333	.022	.905	.011	.095	.906	.030
0.75	.630	.007	.119	.630	.017	.886	.011	.114	.886	.029
0.90	.818	.005	.082	.818	.014	.880	.011	.120	.880	.028
0.95	.879	.005	.070	.879	.013	.879	.011	.121	.879	.029
0.99	.927	.004	.063	.927	.012	.878	.010	.122	.877	.026
Hausman-Pinkovskiy (2017): Predetermined Correction, Infeasible										
0.25	.249	.024	.024	.247	.065	.999	.023	.023	.999	.058
0.50	.490	.069	.069	.499	.173	.992	.066	.067	.999	.173
0.75	.747	.018	.018	.748	.046	.997	.030	.030	.997	.077
0.90	.898	.009	.009	.899	.024	.997	.024	.024	.997	.061
0.95	.949	.010	.010	.948	.026	.998	.031	.031	.996	.078
0.99	.986	.010	.010	.986	.025	.989	.036	.038	.989	.089
Hausman-Pinkovskiy (2017): Predetermined Correction, Simple IV										
0.25	.249	.024	.024	.247	.065	.999	.023	.023	.999	.058
0.50	.510	.078	.079	.514	.193	1.010	.074	.074	1.016	.189
0.75	.747	.018	.018	.748	.046	.997	.030	.030	.997	.077
0.90	.898	.009	.009	.899	.024	.997	.024	.024	.997	.061
0.95	.949	.010	.010	.948	.026	.998	.031	.031	.996	.078
0.99	.988	.012	.012	.989	.033	.995	.039	.040	.998	.102

This table presents simulation results for the model described in Section 4.4. IDR refers to the difference between the 90th and the 10th percentiles of the coefficient in question.

7 Addendum: Proof of method of calculating β for general predetermined variables in fixed effect setting

Suppose that we seek to estimate the model

$$y_{i,t} = x'_{i,t}\beta_0 + \mu_i + \varepsilon_{i,t}$$

Suppose errors are uncorrelated with each other (but heteroskedastic) but regressors are predetermined. So we assume that

$$E\left(x_{i,t}\varepsilon_{i,\tau}\right) = 0,\ \tau \geq t$$

but

$$E\left(x_{i,t}\varepsilon_{i,\tau}\right) \neq 0,\ \tau < t$$

The objective function is

$$\min_{\beta} \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{i,t} - x'_{i,t}\beta - \mu_i\right)^2$$

$$\begin{aligned} F_{\mu_i}(\beta) &= -2\frac{1}{T}\sum_{t=1}^{T}\left(y_{i,t} - x'_{i,t}\beta - \mu_i\right) = 0 \\ &\Rightarrow \mu_i^*(\beta) = \frac{1}{T}\sum_{t=1t}^{T}\left(y_{i,t} - x'_{i,t}\beta\right) \end{aligned}$$

$$\begin{aligned} F_{\beta}(\beta) &= -2\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}x_{i,t}\left(y_{i,t} - x'_{i,t}\beta - \mu_i^*\right) \\ F_{\beta}(\beta_0) &= -2\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\varepsilon_{i,t} - \frac{1}{T}\sum_{\tau=1}^{T}\varepsilon_{i,\tau}\right)x_{i,t} \\ &\rightarrow -2\frac{1}{T}\sum_{t=1}^{T}E\left[\left(\varepsilon_{i,t} - \frac{1}{T}\sum_{\tau=1}^{T}\varepsilon_{i,\tau}\right)x_{i,t}\right] \\ &= 2\frac{1}{T^2}\sum_{t=2}^{T}\sum_{\tau=1}^{t-1}E\left(x_{i,t}\varepsilon_{i,\tau}\right) \neq 0 \end{aligned}$$

Now, let $\hat{y}_{i,t}$ and $\hat{x}_{i,t}$ be de-meaned $y_{i,t}$ and $x_{i,t}$ by panel unit.

$$\hat{y}_{i,t} = y_{i,t} - \frac{1}{T}\sum_{\tau=1}^{T}y_{i,\tau}$$

and

$$\hat{\varepsilon}_{i,t}(\beta) = \hat{y}_{i,t} - \hat{x}'_{i,t}\beta$$

the detrended residuals.
Then,

$$\hat{\varepsilon}_{i,t}(\beta_0) = \varepsilon_{i,t} - \frac{1}{T}\sum_{\tau=1}^{T}\varepsilon_{i,\tau}$$

and, for any $t>1$,

$$E\left(x_{i,t}\hat{\varepsilon}_{i,t'}\left(\beta_0\right)\right)=\frac{1}{N}\sum_{i=1}^{N}x_{i,t}\left(\hat{y}_{i,t'}-\hat{x}_{i,t'}'\beta_0\right)=E\left(x_{i,t}\left(\varepsilon_{i,t'}-\frac{1}{T}\sum_{\tau=1}^{T}\varepsilon_{i,\tau}\right)\right)=E\left(x_{i,t}\varepsilon_{i,t'}\right)-\frac{1}{T}\sum_{\tau=1}^{t-1}E\left(x_{i,t}\varepsilon_{i,\tau}\right)$$

so

$$\begin{aligned}\frac{1}{t-1}\sum_{t'=1}^{t-1}E\left(x_{i,t}\hat{\varepsilon}_{i,t'}\left(\beta_0\right)\right) &= \frac{1}{t-1}\sum_{t'=1}^{t-1}E\left(x_{i,t}\varepsilon_{i,t'}\right)-\frac{1}{T}\sum_{\tau=1}^{t-1}E\left(x_{i,t}\varepsilon_{i,\tau}\right)\\ &= \frac{1}{t-1}\left(1-\frac{t-1}{T}\right)\sum_{\tau=1}^{t-1}E\left(x_{i,t}\varepsilon_{i,\tau}\right)\end{aligned}$$

Then, for any $t>1$,

$$\sum_{\tau=1}^{t-1}E\left(x_{i,t}\varepsilon_{i,\tau}\right)=\frac{T}{T-(t-1)}\sum_{\tau=1}^{t-1}E\left(x_{i,t}\hat{\varepsilon}_{i,\tau}\left(\beta_0\right)\right)$$

and for $t'<t$

$$E\left(x_{i,t}\varepsilon_{i,t'}\right)=E\left(x_{i,t}\hat{\varepsilon}_{i,t'}\left(\beta_0\right)\right)+\frac{1}{T-(t-1)}\sum_{\tau=1}^{t-1}E\left(x_{i,t}\hat{\varepsilon}_{i,\tau}\left(\beta_0\right)\right)$$

Then, the limit of the FOC is

$$F_\beta\left(\beta_0\right)\rightarrow 2\frac{1}{T}\sum_{t=1}^{T-1}\sum_{\tau=t+1}^{T}\left(\frac{1}{T-(\tau-1)}\right)E\left(x_{i,\tau}\hat{\varepsilon}_{i,t}\left(\beta_0\right)\right)$$

Hence, we look for a $\hat{\beta}$ satisfying

$$F_\beta\left(\hat{\beta}\right)=-2\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}x_{i,t}\left(\hat{y}_{i,t}-\hat{x}_{i,t}'\hat{\beta}\right)=2\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T-1}\sum_{\tau=t+1}^{T}\left(\frac{1}{T-(\tau-1)}\right)x_{i,\tau}\left(\hat{y}_{i,t}-\hat{x}_{i,t}'\hat{\beta}\right)$$

or

$$\frac{1}{N}\sum_{i=1}^{N}\left[\frac{1}{T}\sum_{t=1}^{T-1}\sum_{\tau=t+1}^{T}\left(\frac{1}{T-(\tau-1)}\right)x_{i,\tau}\hat{y}_{i,t}+\frac{1}{T}\sum_{t=1}^{T}x_{i,t}\hat{y}_{i,t}-\left[\frac{1}{T}\sum_{t=1}^{T-1}\sum_{\tau=t+1}^{T}\left(\frac{1}{T-(\tau-1)}\right)x_{i,\tau}\hat{x}_{i,t}'+\frac{1}{T}\sum_{t=1}^{T}x_{i,t}\hat{x}_{i,t}'\right]\hat{\beta}\right]=0$$

Let

$$\begin{aligned}z_{i,t} &= x_{i,t}+\sum_{\tau=t+1}^{T}\left(\frac{1}{T-(\tau-1)}\right)x_{i,\tau},\ t<T\\ &= x_{i,T},\ t=T\end{aligned}$$

Define the matrices

$$
\begin{aligned}
W_{XX} &= \frac{1}{N}\sum_{i=1}^{N}\left[\frac{1}{T}\sum_{t=1}^{T-1}\sum_{\tau=t+1}^{T}\left(\frac{1}{T-(\tau-1)}\right)x_{i,\tau}\hat{x}'_{i,t}+\frac{1}{T}\sum_{t=1}^{T}x_{i,t}\hat{x}'_{i,t}\right] \\
&= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}z_{i,t}\hat{x}'_{i,t}=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{z}_{i,t}\hat{x}'_{i,t} \\
W_{XY} &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}z_{i,t}\hat{y}_{i,t}=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{z}_{i,t}\hat{y}_{i,t}
\end{aligned}
$$

(the last equalities following mechanically because of idempotence of residual maker matrix)
Then,

$$\hat{\beta} = W_{XX}^{-1}W_{XY}$$

So, β_0 can be estimated by IV with $\hat{z}_{i,t}$ as the "instrument".