A Nonlinear Least Squares Approach to Estimating Fixed Effects Panel Data Models with Lagged Dependent Variables, with Applications to the Incidental Parameters Problem*
PRELIMINARY – PLEASE DO NOT CITE OR REPRODUCE

Jerry A. Hausman
MIT Department of Economics, E17-238A
50 Ames Street (packages)
77 Massachusetts Ave (mail)
Cambridge MA 02142-1347

Maxim L. Pinkovskiy
Research Department, Federal Reserve Bank of New York
33 Liberty Street
New York, NY 10046

September 2013

Abstract

We propose a novel estimator for the dynamic panel model, which solves the failure of strict exogeneity by the use of nonlinear least squares and an incidental parameters correction. We show that this estimator performs well as compared with approaches in current use. We extend our incidental parameters correction to propose a general solution to the incidental parameters problem in a class of panel data models, as well as a general approach whose bias rapidly diminishes with the number of observations per unit.

1 Introduction

Our paper attempts to contribute to solving two major problems in nonlinear panel data econometrics: failure of strict exogeneity in dynamic panel estimation and the general incidental parameters problem. The idea that estimating the dynamic panel equation by OLS will produce biased and inconsistent estimates has been explored in the literature since 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 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

*We would like to thank Guido Kuersteiner, Joseph Shapiro, and seminar participants at Cambridge University and MIT for very helpful comments. We would like to thank Yuqi Song for outstanding research assistance. Pinkovskiy would like to thank the Paul and Daisy Soros Foundation for New Americans for intellectual stimulation, and the NSF GRFP and the Institute of Humane Studies for funding. This paper solely represents the views of the authors and not of the organizations listed above or of the Federal Reserve Bank of New York. All errors are our own.

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.

The incidental parameters problem has been known in econometrics since Neyman and Scott (1948), and no general methods of its solution exist. Rasch (1960) showed that conditional logistic estimation does not suffer from this problem, and Anderson (1970) developed a methodology for resolving the problem when the incidental parameters constitute a "sufficient statistic" of a likelihood function. Honoré (1992) and Honoré and Kyriazidou (1999) created an ingenious approach for solving the incidental parameter problem in models with censoring and truncation. A literature (e.g. Hahn and Newey (2004), Bai (2009)) has developed general analytical and numerical methods for addressing the incidental parameters problem if the number of panel units (N) and the number of time periods (T) both go to infinity at specified rates. However, we are not aware of general methods for the case in which T is fixed.

We propose a new estimator for the dynamic panel model, which is based on back-substituting the dynamic equation into itself until the right-hand side consists only of exogenous regressors and the initial value of the dependent variable, which should be uncorrelated with the error terms if they do not have serial correlation. We estimate the resulting equation by nonlinear least squares, with a correction to solve the incidental parameter problem that arises in this case. We obtain the correction by computing the probability limit of the first-order condition of the nonlinear least squares estimator and creating an estimator for this probability limit. We expand our estimator to accomodate endogenous regressors and serial correlation in the error terms. Simulations of the performance of this dynamic panel nonlinear least squares estimator (DPNLS) against that of previous GMM-based estimators suggests that DPNLS nearly always has lower bias and variance in its estimates of the coefficient on the lagged dependent variable, and that it is considerably more efficient in the estimation of the coefficients on the covariates, which often tend to be of primary interest in applications. A similar estimator has been proposed by Wooldridge (2013) in the context of dynamic panel estimation with random effects. However, it is the presence of fixed effects that requires the incidental parameters correction, and applications frequently involve the use of fixed effects to help identify the parameters of interest.

We further adapt our technique for solving the incidental parameters problem in the dynamic panel model to other nonlinear panel data models. We show that in models that are affine functions of the incidental parameters (of which the dynamic panel data model is a subset), the incidental parameter problem may be removed completely. We also provide a method for estimating arbitrary nonlinear panel data models that have linearly independent derivatives, which has very rapidly vanishing bias as the number of time periods T goes to infinity.

The rest of the paper is organized as follows. Section 2 presents a simple version of our dynamic panel estimator and its performance in simulations. Section 3 expands the estimator to accomodate weaker assumptions on the data. Section 4 presents a general method for eliminating the incidental parameter problem in models that are affine in these parameters. Section 5 presents our general approach for estimating nonlinear panel data models. Section 6 concludes.

2 Dynamic Panel Estimator

We are interested in estimating the equation

$$y_{i,t} = \alpha_0 y_{i,t-1} + X_{i,t}'\gamma_0 + \mu_{i,0} + \varepsilon_{i,t} \tag{1}$$

where $X_{i,t}$ is a $K \times 1$ vector, $i = 1, ..., N$ and $t = 1, .., T$.

We make the following assumptions:

$$\begin{aligned} \operatorname{rank}\left(E\left(X_{i,t}X_{i,t}'\right)\right) &= K && \text{(A1)} \\ E\left(\varepsilon_{i,t}|X_{\iota,\tau}\right) &= 0 \text{ for all } (i,t,\iota,\,\tau) && \text{(A2)} \\ E\left(\varepsilon_{i,t}\varepsilon_{\iota,\tau}|X\right) &= 0 \text{ for } (i,t) \neq (\iota,\tau) && \text{(A3)} \end{aligned}$$

Assumption A1 is the traditional full rank assumption in regression analysis. Assumption A2 ensures that the regressors are strictly exogenous, so the only identification problem is the one arising from the presence of a lagged dependent variable. Assumption A3 rules out serial correlation in the error terms. This is a traditional assumption in the analysis of panel data models with lagged dependent variables, as the lagged variable tends to capture the serial correlation in the error. We will show how this assumption can be relaxed in Section 2.

The traditional approach has been to estimate this model by differencing equation 1 and using GMM (Arellano and Bond 1991, Blundell and Bond 1998). We proceed differently by back-substituting for $y_{i,t-1}$ into equation 1 so as to express the right hand-side in terms of the current and lagged values of the regressors $X_{i,t}$ and the strictly exogenous initial value $y_{i,0}$

$$y_{i,t}=\alpha_0^t y_{si,0}+\left(\sum_{\tau=0}^{\tau=t-1}\alpha_0^\tau X_{i,t}^\tau\right)'\gamma_0+\mu_{i,0}\sum_{\tau=0}^{\tau=t-1}\alpha_0^\tau+\sum_{\tau=0}^{\tau=t-1}\alpha_0^\tau\varepsilon_{i,t-\tau} \tag{3}$$

(I use the notation $X_{i,t}^\tau$ to designate $X_{i,t-\tau}$.) Since $y_{i,0}$ is determined before $\varepsilon_{i,1},...,\varepsilon_{i,T}$, we have

$$E\left(\varepsilon_{i,t}|y_{\hat{\sigma},0},X_{\sigma,\tau}\right)=0 \text{ for all } (i,t,\sigma,\hat{\sigma},\,\tau)$$

and we no longer have the endogeneity problem. Here, we rely on Assumption A3 to ensure that there is no correlation between $\varepsilon_{i,0}$ and $y_{i,0}$ because of possible serial correlation between $\varepsilon_{i,0}$ and $\varepsilon_{i,-1}$.

The most straightforward way to estimate 3 is by nonlinear least squares. This approach has been used by Wooldridge (2013) in a context in which the individual effects $\mu_{i,0}$ could be viewed as random effects. However, if the individual effects are treated as fixed effects, and if $S\to\infty$ while T remains fixed (as is plausible for much microeconometric data), we have an incidental parameters problem, because the fixed effects are inconsistently estimated. Specifically, the first order conditions of the nonlinear least squares minimization problem imply:

$$\hat{\mu}_i=\left[\frac{1}{T}\sum_{t=1}^{T}\left(\sum_{\tau=0}^{\tau=t-1}\hat{\alpha}^\tau\right)^2\right]^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}\left(y_{i,t}-\hat{\alpha}^t y_{i,0}-\left(\sum_{\tau=0}^{\tau=t-1}\hat{\alpha}^\tau X_{i,t}^\tau\right)'\hat{\gamma}\right)\sum_{\tau=0}^{\tau=t-1}\hat{\alpha}^\tau\right] \tag{4a}$$

$$\hat{\gamma}=\left[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\sum_{\tau=0}^{\tau=t-1}\hat{\alpha}^\tau X_{i,t}^\tau\right)\left(\sum_{\tau=0}^{\tau=t-1}\hat{\alpha}^\tau X_{i,t}^\tau\right)'\right]^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{i,t}-\hat{\alpha}^t y_{i,0}-\hat{\mu}_i\sum_{\tau=0}^{\tau=t-1}\hat{\delta}^\tau\right)\left(\sum_{\tau=0}^{\tau=t-1}\hat{\alpha}^\tau X_{i,t}^\tau\right) \tag{4b}$$

$$0=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\begin{array}{l}\left(y_{i,t}-\left(\hat{\alpha}^t y_{i,0}+\left(\sum_{\tau=0}^{\tau=t-1}\hat{\alpha}^\tau X_{i,t}^\tau\right)'\hat{\gamma}+\hat{\mu}_i\sum_{\tau=0}^{\tau=t-1}\hat{\alpha}^\tau\right)\right)\times\\ \left(t\hat{\alpha}^{t-1}y_{i,0}+\left(\sum_{\tau=0}^{\tau=t-1}\tau\hat{\alpha}^{\tau-1}X_{i,t}^\tau\right)'\hat{\gamma}+\hat{\mu}_i\sum_{\tau=0}^{\tau=t-1}\tau\hat{\alpha}^{\tau-1}\right)\end{array}\right] \tag{4c}$$

Hence, if $\hat{\alpha}=\alpha_0$, the true value of α, it is clear that $\hat{\gamma}$ is a consistent estimator of γ, and $\hat{\mu}_i$ is an unbiased, though inconsistent, estimator of $\mu_{i,0}$ if T is fixed.

Define $\gamma\left(\hat{\alpha}\right)$ and $\left\{\mu_i\left(\hat{\alpha}\right)\right\}_{i=1}^N$ as the solutions to the system of equations (4b) and (4a) for all i. Then, we can express equation 4c as a function of $\hat{\alpha}$ alone:

$$\phi_{N,0}\left(\hat{\alpha}\right):=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\begin{array}{c}\left(y_{i,t}-\left(\hat{\alpha}^t y_{i,0}+\left(\sum_{\tau=0}^{\tau=t-1}\hat{\alpha}^\tau X_{i,t}^\tau\right)'\gamma\left(\hat{\alpha}\right)+\mu_i\left(\hat{\alpha}\right)\sum_{\tau=0}^{\tau=t-1}\hat{\alpha}^\tau\right)\right)\\ \times\left(t\hat{\alpha}^{t-1}y_{i,0}+\left(\sum_{\tau=0}^{\tau=t-1}\tau\hat{\alpha}^{\tau-1}X_{i,t}^\tau\right)'\hat{\gamma}\left(\hat{\alpha}\right)+\mu_i\left(\hat{\alpha}\right)\sum_{\tau=0}^{\tau=t-1}\tau\hat{\alpha}^{\tau-1}\right)\end{array}\right]=0 \tag{4e}$$

From equation (4e) it is clear that

$$\operatorname*{plim}_{N\to\infty}\phi_{N,0}\left(\alpha_0\right)\neq 0$$

and hence, $\hat{\alpha}$ is not a consistent estimator of α_0. This is just a restatement of the standard result that nonlinear least squares is typically inconsistent in the presence of incidental parameters. Our approach is to exploit the linear structure of the model in equation (3) to estimate the value of $E\left(\phi_0\left(\alpha_0\right)\right)$ and subtract it from $\phi_0\left(\hat{\alpha}\right)$ to obtain an equation that is annihilated by α_0 in expectation.

We see that

$$\begin{aligned}
\operatorname*{plim}_{N\to\infty} \phi_{N,0}\left(\alpha_0\right) &= \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} E\left[\frac{1}{T}\sum_{t=1}^{T}\left[\left(\left(\mu_{i,0}-\hat{\mu}_i\left(\alpha_0\right)\right)\sum_{\tau=0}^{\tau=t-1}\alpha_0^{\tau}+\eta_{i,t}\left(\alpha_0\right)\right)\left(\left(\mu_{i,0}-\hat{\mu}_i\right)\sum_{\tau=0}^{\tau=t-1}\tau\alpha_0^{\tau-1}\right)\right]\right] \\
&= \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N} S_i\left(\alpha\right) =: S\left(\alpha\right)
\end{aligned}$$

where

$$S_i\left(\alpha\right) := \frac{1}{\frac{1}{T}\sum_{t=1}^{T}\left(D\left(\alpha.t\right)\right)^2}\frac{1}{T^2}\sum_{t=1}^{T}\sum_{t'=1}^{T}\left\{D\left(\alpha.t\right)\left(D\left(\alpha.t'\right)K\left(\alpha\right)-\frac{\partial}{\partial\alpha}D\left(\alpha,t'\right)\right)\left(\alpha^{t+t'}\left(\sum_{q=1}^{q=\min\left(t,t'\right)}\alpha^{-(2q)}\left[E\left(\varepsilon_{i,q}^2\right)\right]\right)\right)\right\}$$

$$D\left(\alpha,t\right)=\left(\sum_{\tau=0}^{\tau=t-1}\alpha^{\tau}\right)$$

$$K\left(\alpha\right)=\frac{\frac{1}{T}\sum_{t=1}^{T}D\left(\alpha,t\right)\frac{\partial}{\partial\alpha}D\left(\alpha,t\right)}{\frac{1}{T}\sum_{t=1}^{T}\left(D\left(\alpha,t\right)\right)^2}$$

To correct the incidental parameter problem, we need a feasible estimator of $S\left(\alpha\right)$, and specifically of $E\left(\varepsilon_{i,t}^2\right)$ for any i and t. We can obtain such an estimator by using the de-meaned residuals

$$\hat{\varepsilon}_{i,t}\left(\alpha\right)=\left(y_{i,t}-\frac{1}{T}\sum_{t=1}^{T}y_{i,t}\right)-\alpha\left(y_{i,t-1}-\frac{1}{T}\sum_{t=1}^{T}y_{i,t-1}\right)-\left(X_{i,t}-\frac{1}{T}\sum_{t=1}^{T}X_{i,t}\right)'\gamma\left(\alpha\right)$$

and by defining the moment

$$\hat{\varepsilon}_{i,t}^{(2)}\left(\alpha\right) := \frac{1}{T-2}\left[T\hat{\varepsilon}_{i,t}^2\left(\alpha\right)-\frac{1}{T-1}\left(\sum_{t=1}^{T}\hat{\varepsilon}_{i,t}^2\left(\alpha\right)\right)\right]$$

Then,

$$\operatorname*{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\hat{\varepsilon}_{i,t}^{(2)}\left(\alpha\right)=\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}E\left(\varepsilon_{i,t}^2\right)$$

and a feasible, consistent estimator of $S\left(\alpha\right)$ is

$$\hat{S}_N\left(\alpha\right) := \frac{1}{\frac{1}{T}\sum_{t=1}^{T}\left(D\left(\alpha.t\right)\right)^2}\frac{1}{T^2}\sum_{t=1}^{T}\sum_{t'=1}^{T}\left\{D\left(\alpha.t\right)\left(D\left(\alpha.t'\right)K\left(\alpha\right)-\frac{\partial}{\partial\alpha}D\left(\alpha,t'\right)\right)\left(\alpha^{t+t'}\left(\sum_{q=1}^{q=\min\left(t,t'\right)}\alpha^{-(2q)}\left(\frac{1}{N}\sum_{i=1}^{N}\hat{\varepsilon}_{i,t}^{(2)}\left(\alpha\right)\right)\right)\right)\right\}$$

Define the estimator $\hat{\alpha}_{DPNLS}$ by

$$\phi_N\left(\hat{\alpha}_{DPNLS}\right) := \phi_{N,0}\left(\hat{\alpha}_{DPNLS}\right)-\hat{S}_N\left(\hat{\alpha}_{DPNLS}\right)=0 \tag{6}$$

Then, assuming that a unique value for $\hat{\alpha}_{DPNLS}$ exists, we have the following proposition

Proposition 1 *The estimator $\hat{\alpha}_{DPNLS}$ is a consistent estimator of α_0. Furthermore, if $E\left(\varepsilon_{i,t}^4\right)\leq M<\infty$ for all i and t, this estimator is $\sqrt{N}$-consistent for and asymptotically normal estimator of α_0*

Proof. Outlined in text. The consistency of $\hat{\alpha}_{DPNLS}$ follows from the fact that

$$\operatorname*{plim}_{N\to\infty} \phi_N(\alpha_0) = 0$$

The asymptotic normality follows from writing $\phi_N(\hat{\alpha})$ as

$$\phi_N(\alpha_0) = \frac{1}{N}\sum_{i=1}^{N}\left[\begin{array}{c} \frac{1}{T}\sum_{t=1}^{T}\left[\begin{array}{c}\left(y_{i,t} - \left(\alpha_0^t y_{i,0} + \left(\sum_{\tau=0}^{\tau=t-1}\alpha_0^\tau X_{i,t}^\tau\right)'\gamma(\alpha_0) + \mu_i(\alpha_0)\sum_{\tau=0}^{\tau=t-1}\alpha_0^\tau\right)\right) \\ \times\left(t\alpha_0^{t-1}y_{i,0} + \left(\sum_{\tau=0}^{\tau=t-1}\tau\alpha_0^{\tau-1}X_{i,t}^\tau\right)'\hat{\gamma}(\alpha_0) + \mu_i(\alpha_0)\sum_{\tau=0}^{\tau=t-1}\tau\alpha_0^{\tau-1}\right)\end{array}\right] \\ -\frac{1}{\frac{1}{T}\sum_{t=1}^{T}(D(\alpha_0,t))^2}\frac{1}{T^2}\sum_{t=1}^{T}\sum_{t'=1}^{T}\Bigl\{D(\alpha_0,t)\left(D(\alpha_0,t')K(\alpha) - \frac{\partial}{\partial\alpha}D(\alpha_0,t')\right)\Bigl(\alpha_0^{t+t'}\Bigl(\sum_{q=1}^{q=\min(t,t')}\alpha_0^{-(2q)}\hat{\varepsilon}_{i,t}^{(2)}(\alpha_0)\end{array}\right.$$

Hence, $\phi_N(\alpha_0)$ is an average of independent random variables with bounded variance, and therefore satisfies a central limit theorem. The asymptotic normality of $\hat{\alpha}_{DPNLS}$ follows by the delta method, since

$$\sqrt{N}\left(\hat{\alpha}_{DPNLS} - \alpha_0\right) = \frac{\sqrt{N}\phi_N(\alpha_0)}{\phi_N'(\alpha_0)}$$

■

Corollary 2 *Under the conditions of Proposition 1, $\gamma(\hat{\alpha}_{DPNLS})$ is a consistent, asymptotically normal estimator of γ_0.*

Solving equation 6 in finite samples may be impossible because the equation may fail to have a zero. For data realizations for which a zero fails to exist, we take $\hat{\alpha}_{DPNLS}$ to be the local minimum of $\hat{\phi}(\hat{\alpha}_0)^2$ in the region in which $\hat{\phi}(\hat{\alpha}_0) > 0$. This correction vanishes asymptotically, but is numerically important even for samples with $N = 200$, as will be shown.

2.1 Monte Carlo Simulation

We test the performance of the nonlinear least squares estimator against that of the traditional Arellano-Bond estimator and the Blundell-Bond estimator. We set the number of panel units equal to 200, and the number of time periods to 5 (not counting an initial observation of the dependent variable). We create a correlated, lognormally distributed covariate $x_{i,t}$, which enters the equation for $y_{i,t}$ with a coefficient of 0.1 and we allow for time fixed effects. Specifically, we generate the data as follows:

$$\begin{aligned} x_{i,t} &= \rho x_{i,t-1} + (1-\rho)\eta_{i,t-1},\ E\left(\ln\left(\eta_{i,t}\right)\right) = 0,\ \operatorname{var}\left(\ln\left(\eta_{i,t}\right)\right) = \sigma_X^2,\ \rho = 0.7,\ \sigma_X^2 = 1 \\ & x_{i,0}\tilde{\ }N\left(\exp\left(\frac{1}{2(1-\rho)}\sigma_X^2\right), \left(\exp\left(\sigma_X^2\right)-1\right)\exp\left(\sigma_X^2\right)\frac{(1-\rho)^2}{1-\rho^2}\right) \end{aligned}$$

$$y_{i,t} = \alpha y_{i,t-1} + \gamma x_{i,t} + \mu_i + \lambda_t + \varepsilon_{i,t},\ \mu_i\tilde{\ }N\left(0,\sigma_\alpha^2\right),\ \varepsilon_{i,t}\tilde{\ }N\left(0,\sigma_\varepsilon^2\right),\ \sigma_\alpha^2 = \sigma_\varepsilon^2 = \sigma_\lambda^2 = 1$$

We consider two different distributions for the initial condition $y_{i,0}$. The first is a normal distribution with the mean and variance of the stationary distribution for $y_{i,t}$.

$$y_{i,0}\tilde{\ }N\left(\frac{\gamma}{1-\alpha}\exp\left(\frac{1}{2(1-\rho)}\sigma_X^2\right) + \frac{\mu_i}{1-\alpha}, \frac{1}{1-\alpha^2}\left[\sigma_\varepsilon^2 + \sigma_\lambda^2 + \gamma^2\left(\exp\left(\sigma_X^2\right)-1\right)\exp\left(\sigma_X^2\right)\frac{(1-\rho)^2}{1-\rho^2}\right]\right)$$

The second distribution is a nonstationary distribution used by Blundell and Bond (1998):

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

First, consider the estimation results for the autoregressive parameter α. Table **1** presents a distribution of the estimates from the Monte Carlo simulation when the starting values are drawn from the stationary distribution. As is well-known, the Arellano-Bond estimator performs poorly because the correlation between the instruments and the lagged difference approaches zero as α approaches unity. The Blundell-Bond estimator does not suffer from this problem, but instead, slightly overestimates the autoregressive parameter. Our estimator, $\hat{\alpha}_{DPNLS}$, underestimates the parameter α on average, but only slightly. Its mean squared error is comparable to (and occasionally, lower than) that of the minimum of the mean squared errors of the Arellano-Bond and Blundell-Bond estimators. When we consider Monte Carlo simulations of α when the initial condition is nonstationary, we see that the nonlinear least squares estimator not only compares well with the existing estimators in terms of mean squared error, but performs consistently across a variety of conditions in which the existing estimators perform poorly. With a nonstationary initial condition, it is the Arellano-Bond estimator performs very well (because during the transition to a stationary distribution, the autoregressive process is substantially different from a random walk) and the Blundell-Bond estimator performs very poorly, typically overestimating the true parameter on average. The nonlinear least squares estimator performs similarly to Arellano-Bond in this setting (slightly worse for high values of α). Hence, the nonlinear least squares estimator performs well regardless of the initial condition, while the two existing GMM estimators are sensitive to the choice of the initial condition.

Second, consider the estimation of the coefficient γ in Table 2. Often, the purpose of panel data analysis is not to explore the dynamic behavior of $y_{i,t}$ but to better understand the association between $y_{i,t}$ and $x_{i,t}$. Regardless of whether the initial condition is stationary or nonstationary, and regardless of the value of the autoregressive coefficient δ, the nonlinear least squares estimator regularly provides a much lower-MSE estimate of γ than does either Arellano-Bond or Blundell-Bond. The bias is much smaller and the standard deviation is at most half of the standard deviation of either of the GMM estimators. Such a result is not counterintuitive; we have replaced GMM estimation with a regression technique, which tends to reduce the variance of our estimators. However, the improvement in our estimation of γ is striking.[1]

To better understand the working of the nonlinear least squares estimator, I present histograms of the distributions of the estimator for various values of α when we draw the initial values from the stationary distribution. We see that the distributions are asymmetric, and that they have pronounced peaks slightly to the right of the true estimator value. These peaks reflect the cases in which the function $\phi(\alpha)$ does not have a zero, and the estimate of α is based on its local minimum. We see that this procedure might slightly overestimate α, but that it does not do violence to the estimation of α.

Finally, Table 3 assesses the importance of the incidental parameters correction. We see that failing to correct for incidental parameters generates substantial bias under the stationary initial condition if the parameter α is close to unity. For the nonstationary initial condition, the incidental parameters bias is smaller but still present.

3 Instrumental Variables and Optimal Instruments-GMM

We now consider the estimation of the dynamic panel data model in the case that the regressors are endogenous. Suppose there exists a matrix Z that is $NT \times J$, where $J \geq K+1$. We make the following assumptions:

$$\begin{aligned} \operatorname{rank}\left(E\left(X_{i,t}X_{i,t}'\right)\right) &= K && \text{(AIV1a))} \\ \operatorname{rank}\left(E\left(X_{i,t}Z_{i,t}'\right)\right) &= K && \text{(AIV1b))} \\ E\left(\varepsilon_{i,t}|Z_{\iota,\tau}\right) &= 0 \text{ for all } (i,t,\iota,\tau) && \text{(AIV2)} \\ E\left(\varepsilon_{i,t}\varepsilon_{\iota,\tau}|Z\right) &= 0 \text{ for } (i,t)\neq(\iota,\tau) && \text{(AIV3)} \end{aligned}$$

In general, we can estimate the dynamic panel model under these assumptions using a GMM approach . Define the residual function of the model to be

$$\eta_{i,t}(\Gamma) = y_{i,t} - \alpha y_{i,0} - \sum_{\tau=0}^{\tau=t-1} \alpha^{\tau} X_{i,t}^{\tau\prime}\gamma - \mu_i \sum_{\tau=0}^{t-1} \alpha^{\tau}$$

[1] Unfortunately, when we try to estimate the long-run effect of a unit increase in $x_{s,t}$, which is defined by $\frac{\gamma}{1-\delta}$, we do not get substantial improvement from using the nonlinear least squares estimator.

where

$$\Gamma = \left(\alpha, \gamma, \{\mu_i\}_{i=1}^{n}\right)$$

Assumption AIV2 implies that

$$E\left(\eta_{i,t} F\left(Z_{i,t}\right)\right) = 0$$

for any (vector) function F. Newey (1995) shows that the optimal (minimum-variance) GMM estimator to use is given by

$$F^{*}\left(Z_{i,t}\right) = \frac{1}{E\left(\eta_{i,t}^{2} | Z_{i,t}\right)} E\left(\frac{\partial \eta_{i,t}}{\partial \Gamma} | Z_{i,t}\right)$$

Under homoskedasticity, this function is just

$$F^{*}\left(Z_{i,t}\right) = E\left(\frac{\partial \eta_{i,t}}{\partial \Gamma} | Z_{i,t}\right)$$

Now, in our context,

$$E\left(\frac{\partial \eta_{i,t}}{\partial \Gamma} | Z_{i,t}\right) = -\begin{pmatrix} t\alpha^{t-1} E\left(y_{i,0} | Z_{i,t}\right) + \sum_{\tau=0}^{\tau=t-1} \tau \alpha^{\tau-1} E\left(X_{i,t}^{\tau} | Z_{i,t}\right) \gamma_0 + \mu_i \sum_{\tau=0}^{\tau=t-1} \tau \alpha^{\tau-1} \\ \sum_{\tau=0}^{\tau=t-1} \alpha^{\tau} E\left(X_{i,t}^{\tau} | Z_{i,t}\right) \\ \left(\sum_{\tau=0}^{\tau=t-1} \alpha_0^{\tau}\right)_{i=1}^{n} \end{pmatrix}$$

so all we need are expectations of $y_{i,0}$ and the regressors, conditional on the instrument set $Z_{i,t}$.

We can compute estimators of $E\left(X_{i,t}^{\tau} | Z_{i,t}\right)$ and $E\left(y_{i,0} | Z_{i,t}\right)$ via regression. Let $\hat{X}_{i,t}^{\tau}$ and $\hat{y}_{i,0}$ denote the predicted values. Define

$$\hat{\mu}_i\left(\hat{\alpha}\right) = \left[\frac{1}{T} \sum_{t=1}^{T} \left(\sum_{\tau=0}^{\tau=t-1} \hat{\alpha}^{\tau}\right)^{2}\right]^{-1} \left[\frac{1}{T} \sum_{t=1}^{T} \left(y_{i,t} - \hat{\alpha}^{t} y_{i,0} - \left(\sum_{\tau=0}^{\tau=t-1} \hat{\alpha}^{\tau} X_{i,t}^{\tau}\right)' \hat{\gamma}\left(\hat{\alpha}\right)\right) \sum_{\tau=0}^{\tau=t-1} \hat{\alpha}^{\tau}\right]$$

$$\hat{\gamma}\left(\hat{\alpha}\right) = \left[\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(\sum_{\tau=0}^{\tau=t-1} \hat{\alpha}^{\tau} X_{i,t}^{\tau}\right) \left(\sum_{\tau=0}^{\tau=t-1} \hat{\alpha}^{\tau} \hat{X}_{i,t}^{\tau}\right)'\right]^{-1} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(y_{i,t} - \hat{\alpha}^{t} y_{i,0} - \hat{\mu}_i\left(\hat{\alpha}\right) \sum_{\tau=0}^{\tau=t-1} \hat{\delta}^{\tau}\right) \left(\sum_{\tau=0}^{\tau=t-1} \hat{\alpha}^{\tau} \hat{X}_{i,t}^{\tau}\right)$$

and

$$\phi_N^{GMM}\left(\alpha_0\right) = \frac{1}{N} \sum_{i=1}^{N} \left[\begin{array}{c} \frac{1}{T} \sum_{t=1}^{T} \left[\begin{array}{c} \left(y_{i,t} - \left(\alpha_0^{t} y_{i,0} + \left(\sum_{\tau=0}^{\tau=t-1} \alpha_0^{\tau} X_{i,t}^{\tau}\right)' \hat{\gamma}\left(\alpha_0\right) + \hat{\mu}_i\left(\alpha_0\right) \sum_{\tau=0}^{\tau=t-1} \alpha_0^{\tau}\right)\right) \\ \times \left(t\alpha_0^{t-1} \hat{y}_{i,0} + \left(\sum_{\tau=0}^{\tau=t-1} \tau \alpha_0^{\tau-1} \hat{X}_{i,t}^{\tau}\right)' \hat{\gamma}\left(\alpha_0\right) + \hat{\mu}_i\left(\alpha_0\right) \sum_{\tau=0}^{\tau=t-1} \tau \alpha_0^{\tau-1}\right) \end{array}\right] \\ -\frac{1}{\frac{1}{T} \sum_{t=1}^{T} \left(D\left(\alpha_0, t\right)\right)^2} \frac{1}{T^2} \sum_{t=1}^{T} \sum_{t'=1}^{T} \left\{D\left(\alpha_0, t\right) \left(D\left(\alpha_0, t'\right) K\left(\alpha\right) - \frac{\partial}{\partial \alpha} D\left(\alpha_0, t'\right)\right) \left(\alpha_0^{t+t'} \left(\sum_{q=1}^{q=\min\left(t,t'\right)} \alpha_0^{-(2q)} \hat{\varepsilon}_{i,t}^{(2)} \left(\alpha \right.\right.\right.\right. \end{array}\right.$$

Define $\hat{\alpha}_{OI-GMM}$ as the solution to the equation $\phi_N^{GMM}\left(\hat{\alpha}\right) = 0$ (assuming it is unique). Then, the following proposition holds:

Proposition 3 *The estimator $\hat{\alpha}_{OI-GMM}$ is a consistent estimator of α_0. Furthermore, if $E\left(\varepsilon_{i,t}^{4}\right) \leq M < \infty$ for all i and t, this estimator is $\sqrt{N}$-consistent for and asymptotically normal estimator of α_0*

The proof is analogous to that in Section 2.

Note that DPNLS and OI-GMM are equivalent if the instrument set is

$$Z_{i,t} = \left[y_{i,0}, X_{i,t}\right]$$

3.1 Serial Correlation in the Error Terms

Suppose that Assumption (A2) holds, so the regressors act as their own instruments, but Assumption (AIV3) fails. Then, DPNLS fails, but we may still perform OI-GMM by setting

$$Z_{i,t} = \left[X_{i,t}, X_{i,t-1}, ..., X_{i,1}\right]$$

However, we then need a different formula for the correction term $S\left(\hat{\alpha}\right)$.

3.2 Predetermined Regressors

Suppose that instead of Assumption (A2) we only have

$$E\left(\varepsilon_{i,t}|X_{\iota,\tau}\right)=0 \text{ for all } (i,\iota) \text{ and } t\geq\tau$$

Then, we can perform OI-GMM by setting

$$Z_{i,t}=\left[X_{i,1},X_{i,0}\right]$$

3.3 Monte Carlo Simulation

We consider a Monte Carlo simulation with the same parameters as in Section 1, but with the regressors now predetermined rather than exogenous. Specifically,

$$x_{i,t}=\rho x_{i,t-1}+(1-\rho)\,\eta_{i,t-1}+\psi\varepsilon_{i,t-1}$$

We consider only the stationary initial condition and compare the DPNLS and OI-GMM estimators for $N=1000$. Tables 4 and 5 present results for α and γ respectively. We see that OI-GMM does exactly what was ordered: it is unbiased when there is no endogeneity problem ($\psi=0$), though it has much higher MSE than does DPNLS, while it has much lower bias (and MSE) than does DPNLS in the case when an endogeneity problem is present ($\psi=1$).

4 Incidental Parameters Problem in an Arbitrary Affine Model

Suppose that we wish to estimate a model with a vector of fixed effects that affects $y_{i,t}$ affinely. Specifically,

$$y_{i,t}=f\left(x_{i,t},\gamma\right)+g\left(x_{i,t},\gamma\right)'\mu_i+\varepsilon_{i,t}$$

where γ is a $Q\times1$ parameter vector, μ_i is an $L\times1$ vector, and $g\left(x_{i,t},\gamma\right)$ is an $L\times1$ vector. We also assume

$$\begin{aligned}
\operatorname{rank}\left(g\left(x_{i,t},\gamma\right)g\left(x_{i,t},\gamma\right)'\right) &= L && \text{(AA1)}\\
E\left(\varepsilon_{i,t}|x_{i,t}\right) &= 0 && \text{(AA2)}\\
E\left(\varepsilon_{i,t}\varepsilon_{j,s}|X_{i,\cdot}\right) &= \sigma^2\left((i,t)=(j,s)\right) && \text{(AA3)}
\end{aligned}$$

The variance assumption (AA3) may be weakened considerably, and is made for expositional convenience.

We can use analogous methods to Section 1 to estimate γ consistently. Once again, define the fixed effect as a function of the parameter vector γ to be

$$\mu_i\left(\gamma\right)=\left(\frac{1}{T}\sum_{t=1}^{T}g\left(x_{i,t},\gamma_0\right)g\left(x_{i,t},\gamma_0\right)'\right)^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}\left(y_{i,t}-f\left(x_{i,t},\gamma\right)\right)g\left(x_{i,t},\gamma\right)\right]$$

Estimating the error variance is now more complicated because μ_i is a vector and the components of μ_i are multiplied by functions of $x_{i,t}$. To do this, we consider stacks of L observations, so that it is possible to isolate μ_i by premultiplying each stack by the inverse of the matrix multiplying the vector μ_i. Then, we construct de-meaned residuals and proceed as before. The full (and admittedly, complicated) procedure is given below.

1. (a) For each i, define M as the **least common multiple** of T and L. Then construct the $M\times1$ vector $\varepsilon_i^{(1)}\left(\gamma\right)$ with

$$\left(\varepsilon_i^{(1)}\left(\gamma\right)\right)_m=y_{i,m'}-f\left(x_{i,m'},\gamma\right)$$

where

$$m'=m\bmod T$$

and mod denotes the modulo function.

(b) Break up this vector into M/L components, each of size $L \times 1$, and call each component $\varepsilon_i^{(1,k)}$, where $k = 1, ..., M/L = K$. Denote $m^{(k)}(j)$ as the value of m corresponding to the jth element of component k: for example, $m^{(2)}(2) = L + 2$.

(c) For each $\varepsilon_{i,P_k}^{(1,k)}$, construct

$$\varepsilon_i^{(2,k)}(\gamma) = \left(F_i^{(k)}(\gamma)\right)^{-1} \varepsilon_i^{(1,k)}(\gamma)$$

where F_{i,P_k} is a $L \times L$ matrix with element

$$\left[F_i^{(k)}(\gamma)\right]_{p,q} = \left[g\left(x_{i,m^{(k)}(q)}, \gamma\right)\right]_p$$

(d) Next, compute

$$\varepsilon_i^{(3,k)}(\gamma_1) = \varepsilon_i^{(2,k)}(\gamma) - \frac{1}{K}\sum_{k=1}^{K} \varepsilon_i^{(2,k)}(\gamma)$$

(e) Finally, define the estimator of the true error variance as

$$\hat{\sigma}^2(\gamma) = \frac{1}{N}\sum_{i=1}^{N} \frac{\operatorname{tr}\left(\varepsilon_i^{(3,1)}(\gamma)\left(\varepsilon_i^{(3,1)}(\gamma)\right)'\right)}{\operatorname{tr}(D_i)}$$

where

$$\begin{aligned} D_i &= \left(F_i^{(1)}(\gamma)\right)^{-1} U_{1,1}\left(\left(F_i^{(1)}(\gamma)\right)'\right)^{-1} \\ &\quad - \frac{2}{K}\sum_{k=1}^{K}\left(F_i^{(k)}(\gamma)\right)^{-1} U_{1,k}\left(\left(F_i^{(1)}(\gamma)\right)'\right)^{-1} \\ &\quad + \frac{1}{K^2}\sum_{k=1}^{K}\sum_{k'=1}^{K}\left(F_i^{(k)}(\gamma)\right)^{-1} U_{k,k'}\left(\left(F_i^{(k)}(\gamma)\right)'\right)^{-1} \end{aligned}$$

where $\operatorname{tr}()$ denotes the trace function, and where $U_{k,l}$ is a $L \times L$ matrix defined as follows:

$$[U_{k,l}]_{p,q} = 1\left(\left(m^{(k)}(p) \bmod T\right) = \left(m^{l}(q) \bmod T\right)\right)$$

In other words, $U_{k,l}$ is a matrix of indicators for whether the pth error term in component k corresponds to the error term from the same observation as the qth error term in component l.

The following proposition holds:

Proposition 4 *Let γ_{IP-NLS} solve*

$$\begin{aligned} 0 &= \phi(\gamma) := \frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T}\left(y_{i,t} - f(x_{i,t},\gamma) - g(x_{i,t},\gamma)'\mu_i(\gamma)\right)\left(f_\gamma(x_{i,t},\gamma) + g_\gamma(x_{i,t},\gamma)'\mu_i(\gamma)\right) \\ &\quad - \frac{\hat{\sigma}^2(\gamma)}{T}\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T}\left[\begin{array}{c} \left(\frac{1}{T}\sum_{t=1}^{T}\left[g(x_{i,t},\gamma)'\left(\frac{1}{T}\sum_{t=1}^{T} g(x_{i,t},\gamma)g(x_{i,t},\gamma)'\right)^{-1} g(x_{i,\tau},\gamma)\right]\right) \times \\ \left(\left[g_\gamma(x_{i,t},\gamma)'\left(\frac{1}{T}\sum_{t=1}^{T} g(x_{i,t},\gamma)g(x_{i,t},\gamma)'\right)^{-1} g(x_{i,\tau},\gamma)\right]\right) \end{array}\right] \\ &\quad + \frac{\hat{\sigma}^2(\gamma)}{T}\frac{1}{N}\sum_{i=1}^{N}\left[\frac{1}{T}\sum_{t=1}^{T}\left(g_\gamma(x_{i,t},\gamma)'\left(\frac{1}{T}\sum_{t=1}^{T} g(x_{i,t},\gamma)g(x_{i,t},\gamma)'\right)^{-1} g(x_{i,t},\gamma)\right)\right] \end{aligned}$$

where $\hat{\sigma}^2(\gamma)$ and $\mu_i(\gamma)$ are defined above. Then γ_{IP-NLS} is consistent and, if $E\left(\varepsilon_{i,t}^4\right) \le M < \infty$ for all i and t, is $\sqrt{N}$-consistent and asymptotically normal.

5 General Approach to Incidental Parameters Problem in Nonlinear Models

Suppose that we wish to estimate the arbitrarily nonlinear model

$$y_{i,t} = f\left(x_{i,t}, \gamma, \mu_i\right) + \varepsilon_{i,t}$$

under the assumptions

$$\begin{aligned} E\left(\varepsilon_{i,t}|x_{i,t}\right) &= 0 \\ E\left(\varepsilon_{i,t}\varepsilon_{j,s}|X_{i,\cdot}\right) &= \sigma^2\left((i,t)=(j,s)\right) \end{aligned}$$

The variance assumption is made for convenience and can be relaxed. Once again, T is assumed to be fixed. The complication is that the model is now nonlinear in the fixed effect.

We propose an estimator whose asymptotic bias goes to zero much faster than (to our knowledge) has been observed in the literature. First, we obtain an estimator for μ_i, denoted by β_i, by solving the first-order condition. Then, we consider a Taylor expansion of f around ν_i to $T-1$ terms, and allow the coefficients on the derivative terms to be unrestricted. Hence, we transform the nonlinear model into a model with a vector of $T-1$ fixed effects that enters linearly. We can solve the transformed model by setting $f\left(x_{i,t}, \gamma\right) := f\left(x_{i,t}, \gamma, \beta_i\right)$, and $g\left(x_{i,t}, \gamma\right) = \left[\frac{1}{j!} f^{(j)}\left(x_{i,t}, \gamma, \beta_i\right)\right]_{j=1}^{T-1}$. In particular, Assumption (AA3) implies that we need the Wronskian of the first $T-1$ derivatives of f to be nonsingular for generic $x_{i,t}$, γ, and β_i.

However, the remainder from the Taylor expansion gives rise to the inconsistency. In particular, since β_i will depend on the average of the error terms for panel unit i, we will have

$$\beta_i = \mu_i + O_P\left(\frac{1}{\sqrt{T}}\right)$$

By Taylor's Theorem, the remainder term for each panel unit i will satisfy

$$\left|R_{T-1,i,t}\right| \leq \frac{K}{T!}\left(\beta_i - \alpha_i\right)^T = \frac{K}{T!}\left(O_P\left(\frac{1}{\sqrt{T}}\right)\right)^T$$

so

$$(T!)\, T^{\frac{T}{2}} R_{T-1,i,t} = O_P(1)$$

Hence, the bias will be of order $1/\left((T!)\, T^{\frac{T}{2}}\right)$, which goes to zero extremely quickly, as shown in the table below.

T	$T!T^{\frac{T-1}{2}}$
2	4
3	31.177
4	384
5	6708.2
6	1.5552×10^5
7	4.5738×10^6
8	1.6515×10^8

6 Conclusion

We have presented a new method of estimating the basic dynamic panel model with covariates. We have shown through simulation evidence and empirical examples that this method appears to outperform GMM-based methods in terms of both bias and variance of the estimates, and that it appears to make a difference in applications. We have also expanded this method to wider classes of dynamic panel models, as well as relaxed some of the more stringent assumptions underpinning it. Furthermore, we have adapted the methodology behind the derivation of our dynamic panel estimator to solve the incidental parameters problem for a large class of nonlinear panel data models, and we have proposed a general approach of estimating nonlinear panel data models with fixed-T asymptotics, whose bias is rapidly diminishing in T.

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, E.** 1970. "Asymptotic Properties of Conditional Maximum Likelihood Estimators," *Journal of the Royal Statistical Society*, Series B, 32, 283-301.

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

[4] **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.

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

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

[7] **Griliches, Zvi, and Jerry A. Hausman**. 1986. "Errors in Variables in Panel Data." *Journal of Econometrics* 31, no. 1: 93-118.

[8] **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.

[9] **Hahn, Jinyong, and Whitney Newey**. 2004. "Jackknife and Analytical Bias Reduction for Nonlinear Panel Models." *Econometrica* 72, no. 4: 1295-1319.

[10] **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.

[11] **Honore, Bo E.** 1992. "Trimmed LAD and Least Squares Estimation of Truncated and Censored Regression Models with Fixed Effects." *Econometrica* 60, no. 3: 533-565.

[12] **Honore, Bo E., Ekaterini Kyriazidou, and Christopher Udry**. 1997. "Estimation of Type 3 Tobit Models Using Symmetric Trimming and Pairwise Comparisons." *Journal of Econometrics* 76, no. 1-2: 107-128.

[13] **Neyman, J., and E. L. Scott.** 1948. "Consistent estimates based on partially consistent observations." *Econometrica* 16, 1-32.

[14] **Semykina, Anastasia, and Jeffrey M. Wooldridge**. 2013. "Estimation of Dynamic Panel Data Models with Sample Selection." *Journal of Applied Econometrics* 28, no. 1: 47-61.

7 Tables and Figures

Table 1

Simulations of α for DPNLS and GMM-Style Estimators								
T=5, N=200								
Method	Mean	SD	RMSE	Median	1st pctile	25th pctile	75th pctile	99th pctile
Stationary Initial Condition								
Arellano-Bond, α=.25	.235	.054	.055	.230	.125	.202	.267	.366
Arellano-Bond, α=.50	.472	.072	.077	.468	.329	.419	.529	.639
Arellano-Bond, α=.75	.669	.126	.150	.674	.369	.571	.752	.973
Arellano-Bond, α=.90	.601	.278	.408	.651	-.223	.426	.774	1.155
Arellano-Bond, α=.95	.474	.315	.570	.489	-.549	.286	.714	1.101
Blundell-Bond, α=.25	.280	.057	.065	.273	.174	.242	.323	.400
Blundell-Bond, α=.50	.549	.069	.085	.543	.407	.503	.593	.709
Blundell-Bond, α=.75	.839	.071	.114	.848	.665	.788	.888	.989
Blundell-Bond, α=.90	.979	.042	.090	.975	.886	.951	1.004	1.127
Blundell-Bond, α=.95	.995	.020	.049	.993	.956	.982	1.006	1.059
DPNLS, α=.25	.248	.060	.060	.245	.108	.209	.280	.397
DPNLS, α=.50	.499	.068	.068	.491	.340	.451	.536	.664
DPNLS, α=.75	.759	.096	.096	.750	.558	.695	.801	.962
DPNLS, α=.90	.888	.092	.093	.895	.676	.828	.976	1.024
DPNLS, α=.95	.913	.087	.094	.923	.701	.847	.991	1.033
Nonstationary Initial Condition								
Arellano-Bond, α=.25	.237	.048	.049	.235	.141	.206	.270	.348
Arellano-Bond, α=.50	.468	.080	.086	.470	.301	.409	.525	.650
Arellano-Bond, α=.75	.719	.102	.107	.718	.470	.647	.799	.944
Arellano-Bond, α=.90	.897	.039	.039	.897	.819	.869	.929	.980
Arellano-Bond, α=.95	.948	.030	.030	.948	.887	.926	.972	1.013
Blundell-Bond, α=.25	.452	.066	.213	.452	.296	.414	.494	.599
Blundell-Bond, α=.50	.540	.069	.080	.546	.382	.493	.590	.683
Blundell-Bond, α=.75	.980	.014	.230	.981	.939	.971	.991	1.006
Blundell-Bond, α=.90	1.102	.009	.202	1.102	1.069	1.098	1.109	1.118
Blundell-Bond, α=.95	1.139	.008	.189	1.139	1.105	1.135	1.144	1.153
DPNLS, α=.25	.249	.059	.059	.247	.128	.212	.285	.405
DPNLS, α=.50	.503	.093	.093	.494	.297	.444	.544	.773
DPNLS, α=.75	.752	.085	.085	.750	.575	.688	.807	.959
DPNLS, α=.90	.896	.040	.040	.894	.799	.868	.923	.988
DPNLS, α=.95	.947	.032	.032	.946	.869	.924	.968	1.021

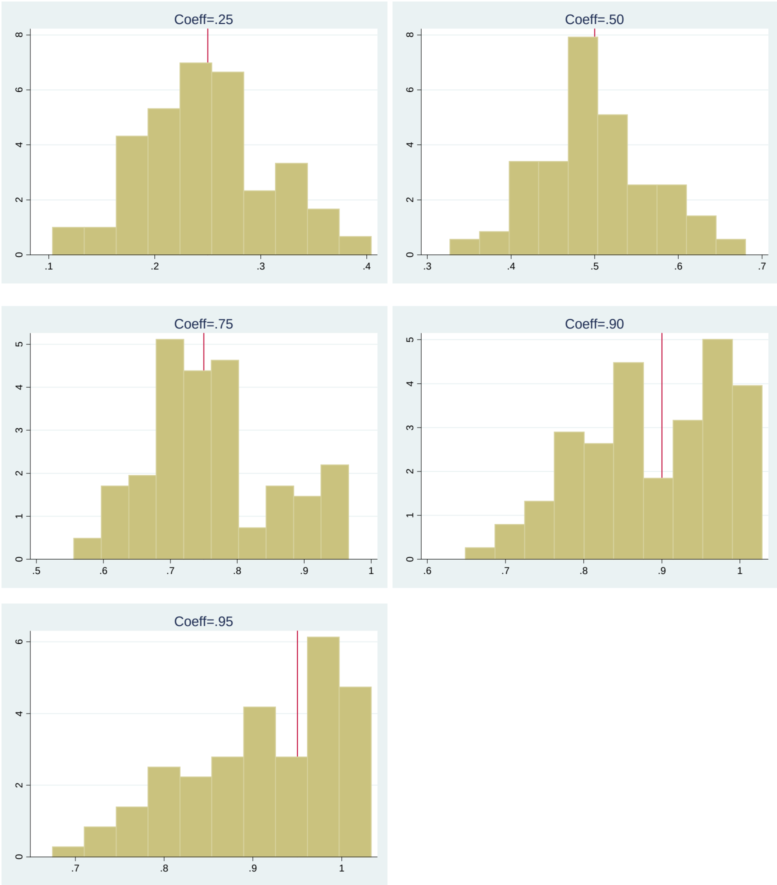
(1)

Table 2

Simulations of γ for DPNLS and GMM-Style Estimators								
True γ=0.10 For All Specifications, T=5, N=200								
Method	Mean	SD	RMSE	Median	1st pctile	25th pctile	75th pctile	99th pctile
Stationary Initial Condition								
Arellano-Bond, α=.25	.066	.158	.161	.069	-.262	-.028	.148	.548
Arellano-Bond, α=.50	.069	.156	.159	.069	-.248	-.028	.152	.551
Arellano-Bond, α=.75	.084	.156	.156	.083	-.241	-.019	.164	.546
Arellano-Bond, α=.90	.139	.160	.165	.137	-.200	.023	.235	.615
Arellano-Bond, α=.95	.186	.165	.186	.169	-.141	.076	.284	.702
Blundell-Bond, α=.25	.064	.145	.150	.056	-.274	-.024	.151	.437
Blundell-Bond, α=.50	.055	.146	.153	.045	-.278	-.036	.141	.426
Blundell-Bond, α=.75	.037	.152	.165	.027	-.302	-.063	.128	.401
Blundell-Bond, α=.90	.043	.146	.157	.045	-.275	-.058	.126	.428
Blundell-Bond, α=.95	.056	.143	.150	.056	-.249	-.043	.142	.440
DPNLS, α=.25	.115	.060	.062	.107	-.013	.078	.154	.291
DPNLS, α=.50	.115	.062	.064	.108	-.020	.075	.146	.297
DPNLS, α=.75	.116	.066	.068	.114	-.027	.071	.153	.302
DPNLS, α=.90	.117	.068	.070	.114	-.033	.071	.153	.303
DPNLS, α=.95	.116	.068	.069	.113	-.036	.070	.153	.300
Nonstationary Initial Condition								
Arellano-Bond, α=.25	.064	.162	.166	.067	-.360	-.034	.155	.546
Arellano-Bond, α=.50	.070	.162	.164	.074	-.320	-.032	.164	.547
Arellano-Bond, α=.75	.074	.154	.156	.073	-.204	-.028	.160	.543
Arellano-Bond, α=.90	.069	.157	.160	.070	-.219	-.035	.147	.547
Arellano-Bond, α=.95	.069	.159	.162	.070	-.221	-.040	.149	.556
Blundell-Bond, α=.25	-.132	.353	.422	-.176	-.981	-.314	.047	1.114
Blundell-Bond, α=.50	.048	.149	.157	.041	-.304	-.046	.138	.421
Blundell-Bond, α=.75	.035	.177	.188	.042	-.475	-.085	.133	.560
Blundell-Bond, α=.90	.071	.218	.219	.064	-.540	-.043	.174	.967
Blundell-Bond, α=.95	.087	.236	.237	.073	-.563	-.034	.183	1.106
DPNLS, α=.25	.115	.060	.062	.106	-.010	.077	.154	.291
DPNLS, α=.50	.115	.062	.064	.108	-.019	.074	.146	.298
DPNLS, α=.75	.115	.065	.067	.107	-.038	.075	.149	.297
DPNLS, α=.90	.116	.067	.068	.108	-.034	.072	.148	.298
DPNLS, α=.95	.116	.067	.069	.109	-.034	.071	.150	.298

(2)

Figure 1



(1)

Table 3

Simulations of α for DPNLS – IP Correction					
T=5, N=200					
Method	Mean	RMSE	Median	25th pctile	75th pctile
Stationary Initial Condition					
DPNLS, Corrected, α=.25	.244	.072	.244	.187	.298
DPNLS, Corrected, α=.50	.497	.080	.491	.435	.544
DPNLS, Corrected, α=.75	.756	.101	.736	.674	.820
DPNLS, Corrected, α=.90	.883	.094	.873	.802	.972
DPNLS, Corrected, α=.95	.911	.095	.908	.837	.988
DPNLS, Uncorrected, α=.25	.252	.070	.255	.197	.305
DPNLS, Uncorrected, α=.50	.483	.067	.481	.432	.522
DPNLS, Uncorrected, α=.75	.678	.091	.673	.630	.710
DPNLS, Uncorrected, α=.90	.764	.144	.761	.730	.793
DPNLS, Uncorrected, α=.95	.782	.174	.779	.751	.808

(3)

Table 4

Simulations of α for Optimal GMM Estimator						
T=5, N=1000, Stationary Init. Cond., α=0.9						
Model	Method	Mean	RMSE	Median	25th pctile	75th pctile
ψ=0	Optimal GMM, DPNLS Instruments.	.899	.055	.894	.857	.951
ψ=0	Optimal GMM, Init. Value. Instruments	.921	.069	.921	.870	.959
ψ=1	Optimal GMM, DPNLS Instruments.	.809	.093	.810	.796	.824
ψ=1	Optimal GMM, Init. Value. Instruments	.919	.075	.922	.865	.955

(4)

Table 5

Simulations of γ for Optimal GMM Estimator						
T=5, N=1000, Stationary Init. Cond., γ=0.1						
Model	Method	Mean	RMSE	Median	25th pctile	75th pctile
ψ=0	Optimal GMM, DPNLS Instruments.	.099	.029	.099	.080	.119
ψ=0	Optimal GMM, Init. Value. Instruments	.104	.060	.103	.065	.144
ψ=1	Optimal GMM, DPNLS Instruments.	.040	.062	.040	.028	.052
ψ=1	Optimal GMM, Init. Value. Instruments	.089	.042	.092	.063	.118

(5)