Properties of Fixed Effects Dynamic Panel Data Estimators for a Typical Growth Dataset

Arya B. Gaduh

May 2002

Economics Working Paper Series
http://www.csis.or.id/papers/wpe062

Properties of Fixed Effects Dynamic Panel Data Estimators for a Typical Growth Dataset
Arya B. Gaduh
CSIS Working Paper Series
WPE 062
May 2002

ABSTRACT

This paper investigates the properties of dynamic panel data (DPD) estimators in the context of a typical growth dataset. Using Monte Carlo simulations, we compare the performance of various DPD estimators, namely the Anderson-Hsiao (AH) and Arellano-Bond's General Method of Moment (GMM) one-step and two-step estimators, using the least-square dummy variable (LSDV) as a benchmark. We arrive at three conclusions. First, LSDV produces biased estimates and the biases are significant even for a moderate-sized time dimension. Second, there is no immediately obvious choice to replace LSDV among the estimators considered here. For one, there is the bias-efficiency trade-off. In addition, differences in the characteristics of data influence the performances of the various estimators. Finally, serial correlations in the error terms, even at a low degree, can introduce significant biases to the estimations.

Keywords: *dynamic panel data estimators, economic growth, Anderson-Hsiao (AH), General Methods of Moment (GMM), least-square dummy variable (LSDV), Monte Carlo simulation*.

Arya B. Gaduh
abgaduh@csis.or.id
Department of Economics
CSIS Jakarta

Properties of Fixed Effects Dynamic Panel Data Estimators for a Typical Growth Dataset: Monte Carlo Evidence

Arya B. Gaduh

I. Introduction

Among most existing empirical studies on growth, single cross-section methodology is the dominant tool of analysis. Its simple implementation is appealing, yet this comes at a cost. Single cross-section specification puts implicit restrictions that production functions and technologies across countries are identical. Hence, if we believe that production functions and technologies differ across countries, single cross-section regression suffers from omitted variable bias. Furthermore, cross-section fails to capture the dynamic aspects of growth. As such, recent years saw macroeconomists trying to accommodate these features using dynamic panel data (DPD) specification (Islam, 1995; Caselli, Esquivel and Lefort, 1996).

There are many studies on properties of DPD estimators; most, however, are oriented towards typical microeconomic datasets with large cross-section but small time dimensions.[1] In contrast, little is written for the typical dimension of macroeconomic datasets with sizable cross-section dimensions and moderate size time dimensions. This difference is important for two reasons.[2] First, for small time dimensions (T), least square dummy variable (LSDV) estimates are biased. This bias becomes less significant as T increases. Those working with macroeconomic panels need to know how large must T be before this bias becomes negligible. Second*,* the characteristics of the data influence the robustness of the estimation techniques used. Hence, when working with macro panel datasets, there is no hard-and-fast rule as to which estimation technique is best.[3]

The paper by Judson and Owen (1999) seems to be the only one that specifically examines DPD estimations for macro datasets. The present paper is similar to theirs in the basic designs of the experiments. However, there are three differences. First, we modify the parameter values. We use parameter values that are calibrated (using the Heston and Summer's Penn World Table) to mimic typical growth regressions. Also, we only take a subset of estimation techniques considered in their paper. Finally, this study extends their experiment by incorporating different degrees of serial correlation of the disturbances to observe their effects on the properties of the estimators.

We draw three conclusions. First, LSDV produces biased estimates and the biases are significant even for a moderate T, especially when γ takes a low value. For γ close to one, this problem is much less pronounced. Second, there is no immediately obvious choice to replace LSDV among the estimators considered here. For one, there is the bias-efficiency trade-off. In addition, differences in the characteristics of data influence the performances of the various estimators. Finally, serial correlations in the error terms, even at a low degree, can introduce significant biases to the estimations.

We begin in the next section with a discussion of several estimation techniques whose properties are examined in this paper. In Section III, we discuss the methodology. Section IV considers the results, both in the absence and presence of autocorrelations in the disturbance terms. Section V presents our conclusions.

II. Inconsistency of LSDV and proposed solutions

For growth regressions, it is usually preferable to use fixed effects instead of random effects models. Since we use panel data to capture the omitted variables ignored by the single cross-section, these individual (country) specific effects are likely to be correlated with the regressors (Judson and Owen, 1999; Islam, 1995).

The canonical dynamic fixed effects model is

$$y_{it} = \gamma y_{it-1} + x'_{it}.\beta + \mu_i + \varepsilon_{it} \qquad i = 1, \ldots, N;\ t = 1, \ldots T \tag{1}$$

where γ is a scalar, x'_{it} is 1 x (K-1) matrix, and β is (K-1) x 1, μ_i is a fixed effect and $\varepsilon_{it} \sim$ IID (0, σ_ε^2).

We also assume

$$\begin{aligned} &\sigma_\varepsilon^2 > 0, \\ &E(\varepsilon_{it}, \varepsilon_{js}) = 0 \text{ where } i \neq j,\ t \neq s, \\ &E(x_{it}, \varepsilon_{it}) = 0\ \forall i, j, t, s \end{aligned} \tag{2}$$

Fixed effects model is usually estimated using LSDV. However, with the lagged dependent variable as a regressor, LSDV generates biased estimates. Subtracting the time mean of (1) from itself gives us

$$y_{it} - y_{i.} = \gamma(y_{it-1} - y_{it.-1}) + (x'_{it} - x'_{i.})\ \beta + (\varepsilon_{it} - \varepsilon_{i.}) \tag{3}$$

where for any variable z_{it}, $z_{i.} = (1/T)\sum_{t=1}^{T} z_{it}$ and $z_{i.-1} = (1/T)\sum_{t=0}^{T-1} z_{it}$. LSDV estimates the parameters by performing ordinary least square (OLS) on Eq. (3). However, since the time mean of ε_{it} includes the disturbance of $y_{i,t-1}$, the transformed regressor $(y_{i,t-1} - y_{i,t.-1})$ is not orthogonal to the transformed disturbance term $(\varepsilon_{it} - \varepsilon_{i.})$. Even when N goes to infinity, this will give biased estimates. Nickell (1981) calculates this bias to be of $O(1/T)$; as T grows, the bias becomes smaller. However, it is not clear at what critical value of T will this bias become negligible.

Anderson and Hsiao (1981) suggest using instrumental variable (IV) estimation as an alternative to LSDV. Taking the first difference of Eq. (1), we remove the fixed effects and arrive at

$$y_{it} - y_{it-1} = \gamma(y_{it-1} - y_{it-2}) + (x'_{it} - x'_{it-1}).\beta + (\varepsilon_{it} - \varepsilon_{it-1}) \tag{4}$$

With this specification, both lagged difference Δy_{it-2} and lagged level y_{it-2} are potential instruments. These instruments are highly correlated with the regressor Δy_{it-1} and uncorrelated with the disturbance terms, as long as the disturbances are not serially correlated. Arellano (1989), however, argues that for a wide range of parameters, estimations using lagged differences Δy_{it-1} has a singularity point and very large variances and therefore, suggests the use of lagged level instead.[4]

Meanwhile, Arellano and Bond (1991) suggest using the orthogonality conditions between lagged values of y_{it} and the disturbance terms to obtain additional instruments. Observing Eq. (4), they notice that for t = 3, for example, y_{i1} would be a valid instrument for Δy_{it-1}. For t = 4, both y_{i1} and y_{i2} would be appropriate instruments for Δy_{it-1} since neither is correlated with the disturbance terms $(\varepsilon_{it} - \varepsilon_{it-1})$. By applying this kind of argument, they obtain an additional instrument for each subsequent t, such that for $t = T$, Δy_{iT-1} has a set of valid instruments ($y_{i1}, y_{i2}, \dots y_{iT-2}$). Stacking these sets of instruments, we get

$$W_i = \begin{bmatrix} [y_{i1}] & & & 0 \\ & [y_{i1}, y_{i2}] & & \\ & & \ddots & \\ 0 & & & [y_{i1}, y_{i2}, \dots, y_{iT-2}] \end{bmatrix} \tag{5}$$

Given $\Delta\varepsilon'_i = (\varepsilon_{i3} - \varepsilon_{i2,\ldots,}\ \varepsilon_{iT} - \varepsilon_{iT-1})$ and observing that $E(W'_i.\Delta\varepsilon'_i) = 0$, they derive moment conditions which are then used to perform *generalised method of moments* (GMM) estimations.

Arellano and Bond propose a two-step estimation process. In the first step, we need an approximation of $\Delta\varepsilon'_{i.}$ By constructing

$$G = \begin{bmatrix} 2 & -1 & 0 & \cdots & 0 & 0 & 0 \\ -1 & 2 & -1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 2 & -1 \\ 0 & 0 & 0 & \cdots & 0 & -1 & 2 \end{bmatrix} \qquad (6)$$

which is *(T-2)* x *(T-2)*, the Kronecker product ($I_N \otimes G$) gives a preliminary approximation of $\Delta\varepsilon'_i$ in an MA(1) process. With both Wi's and $\Delta\varepsilon'_i$ available, we can do the one-step consistent GMM estimation. From this first step, we obtain estimates of $\Delta\varepsilon'_{i.}$ This can immediately be utilised to perform the second step of the two-step GMM estimation. They claim that GMM gives more efficient estimates than IV.

In this paper, using a Monte Carlo study, we consider LSDV, Anderson-Hsiao IV (AH) using lagged levels as instruments, and Arellano-Bond's one-step (GMM1) and two-step (GMM2) estimations. In the first part, we consider the case where all of the assumptions listed in Eq. (2) are satisfied. It is important to note that all of the results derived above rest on Eq. (2) being satisfied, particularly about the disturbances not being serially correlated. However, we rarely find serially uncorrelated disturbances in macroeconomic data. Therefore, we simulate different degrees of serial correlations here. This constitutes the second part of our results.

III. Methodology

The data generation process follows that of previous studies (Judson and Owen, 1999; Kiviet, 1995). x_{it} was generated with

$$x_{it} = \delta x_{it-1} + \xi_{it} \qquad \xi_{it} \sim N(0, \sigma_\xi^2) \qquad (7)$$

With growth regressions in mind when designing the experiment, we use the log of investment share in GDP in 1960 from the Penn World Table (PWT) as the initial values of x_{i0}. To find values for δ and σ_ξ^2, we took the x_{it} data for all countries between 1960 and 1990, and did a fixed effects regression of x_{it} on $x_{it-1.}$ This exercise gave a value of 0.77 for δ, and values between –2.5 and 3.3 for the estimated residuals. The latter is used to approximate the volatility of σ_ξ^2.

Meanwhile, y_{it} was generated using Eq. (1). We took the log of GDP per worker in 1960 as the initial values of y_{i0}. For the parameters in Eq. (1), Islam (1995) suggests a range of 0.59 to 0.79 for γ, and 0.12 to 0.15 for β.[5] He also estimates the μ_i's to be between 1.11 and 1.81.[6] Similar to the case of x_{it}, the volatility of σ_ε^2 was estimated by fixed effects regression of y_{it} on y_{it-1} and x_{it}. From this exercise, the disturbances were in the range of -0.56 to 0.95.

Hence, our parameter choices can be summarised as follows. We use a value 0.8 for δ and set the volatility of ξ_{it} to 3. Meanwhile, for the values of γ, we consider 0.6 and 0.8 as suggested by Islam (1995), as well as 0.2 to investigate whether a low γ give a qualitatively different result. β's are 0.15 and 0.75 and the volatility of ε_{it} is set at 0.75. μ_i's are between 1.0 and 2.0, generated by a random number generator. Disturbance terms were generated with a fixed seed for replicability and we threw away the first 100 randomly generated numbers. We then run simulations for $N = 50$ and $T = 5$, 10, 20, and 30. In this first set of experiments, we run a total of 24 simulations.

IV. Results

A. In a world with no serially correlated errors

An immediate issue facing our simulation was the choice of instrumental variables for the Arellano and Bond estimation. Taking a full set of instruments might seem sensible for small T's; however, for $T \geq 20$, for instance, computational complexity increases significantly. Preliminary simulations suggested that it could take almost a day to complete estimations of LSDV, AH, GMM1 and GMM2 with complete sets of instruments for $N = 50$, $T = 20$ and 1000 iterations. Fortunately, restricting the maximum number of instruments used incurs minimal loss on the robustness of the estimations. In fact, Arellano and Bond (1998) advise against using the full history of instruments because, in the presence of an exogenous regressor, this might introduce small sample overfitting biases. This is confirmed by our preliminary exercise.

We also find that GMM2 estimation does not provide a significant improvement over GMM1. In most cases, we find GMM1 outperforms its GMM2 counterpart in terms of root mean squared errors (RMSE). This result confirms that of Judson and Owen (1999), who then abandoned GMM2 and instead, just chose a few well performing GMM1 estimators. We follow their strategy here. Early experiments suggest that GMM103, and GMM110, which are GMM1 with 3, and 10 instruments, perform rather well.

Table 1 shows biases in LSDV estimations. Clearly, the bias in γ is much more severe than that in β. It also confirms previous studies (e.g., Kiviet, 1995; Judson and Owen, 1999; Nickell, 1981) that the bias decreases in *T*. For a small *T*, the bias increases in γ; however this relationship does not hold for moderate and large *T*'s. For low γ's, the bias can be quite significant in relative terms: even for $T = 20$, the bias can be as large as 20 percent of the true value for $\gamma = 0.2$.[7]

Table 1. Biases of LSDV estimations

T	γ	γ–bias (S.E.)	β	β–bias (S.E.)
5	0.2	-0.1664(0.0590)	0.15	0.0071(0.0223)
	0.6	-0.1740(0.0557)	0.15	0.0056(0.0233)
	0.8	-0.2874(0.0671)	0.15	-0.0038(0.0226)
10	0.2	-0.0820(0.0392)	0.15	0.0071(0.0119)
	0.6	-0.0755(0.0313)	0.15	0.0064(0.0121)
	0.8	-0.1043(0.0303)	0.15	0.0055(0.0115)
20	0.2	-0.0398(0.0247)	0.15	0.0050(0.0069)
	0.6	-0.0364(0.0183)	0.15	0.0055(0.0071)
	0.8	-0.0400(0.0162)	0.15	0.0054(0.0067)
30	0.2	-0.0259(0.0204)	0.15	0.0035(0.0054)
	0.6	-0.0238(0.0143)	0.15	0.0040(0.0053)
	0.8	-0.0229(0.0112)	0.15	0.0045(0.0050)

$N = 50$, $\delta = 0.8$, $\sigma_{\xi}=3$, $\sigma_{\varepsilon}=0.75$.

The full result for γ is presented in Table 2. Here, we see a case of the cure being worse than the disease. Only in two simulations, to wit, with $T = 5$, $\beta = 0.15$ and γ being either 0.2 or 0.6, do we have a strong case for using either AH or GMM1 instead of LSDV. Even when $T = 5$ – the case where the biases of LSDV are expected to be large enough to merit alternative estimators – the RMSE's of LSDV in the other four simulations are better than those of AH and GMM1.

For low values of γ, AH estimates tend to be unbiased but less efficient than LSDV. However, AH is somewhat unreliable: it frequently gives highly biased estimates for certain parameters.[8] On the other hand, as discussed above, GMM103 estimates are subject to overfitting bias which, while not as large as that of LSDV, can be rather significant. In terms of efficiency, GMM103 performs quite well for high values of γ while AH tends to perform poorly. As T increases, the performance of GMM1 worsens. In almost all cases for $T \geq 10$, GMM103 and GMM110 give significantly large biases, especially for low βs. For $T \geq 10$, AH gives unbiased but inefficient estimates with standard errors that are increasing in γ. At their worst, these standard errors can be more than five times the true value of the coefficients.

Table 2. Biases, Standard Errors and RMSE of Various Estimators

T	β	γ	*LSDV-bias* *(S.E.)* *[R.M.S.E]*	*AH-bias* *(S.E.)* *[R.M.S.E]*	*GMM103-bias* *(S.E.)* *[R.M.S.E]*	*GMM110-bias* *(S.E.)* *[R.M.S.E]*
5	*0.15*	*0.2*	-0.1664	0.0008	-0.0894	-
			(0.0590)	(0.0959)	(0.132)	-
			[0.1766]	[0.0959]	[0.1594]	-
		0.6	-0.1740	0.0021	-0.0529	-
			(0.0557)	(0.0921)	(0.1011)	-
			[0.1827]	[0.0921]	[0.1141]	-
		0.8	-0.2874	-0.5940	-0.3618	-
			(0.0671)	(15.5292)	(0.2626)	-
			[0.2951]	[15.5406]	[0.4471]	-
	0.75	*0.2*	-0.0316	0.0027	-0.0367	-
			(0.0273)	(0.0967)	(0.0838)	-
			[0.0418]	[0.0967]	[0.0915]	-
		0.6	-0.0318	-22.4095	-0.0395	-
			(0.0241)	(716.1240)	(0.0858)	-
			[0.0399]	[716.4745]	[0.0945]	-
		0.8	-0.0262	-0.0421	-0.0315	-
			(0.0200)	(1.6729)	(0.0763)	-
			[0.0330]	[1.6734]	[0.0825]	-
10	*0.15*	*0.2*	-0.0820	0.0012	-0.1868	-
			(0.0392)	(0.0643)	(0.0952)	-
			[0.0909]	[0.0643]	[0.2096]	-
		0.6	-0.0755	0.0042	-0.0830	-
			(0.0313)	(0.0615)	(0.0699)	-
			[0.0817]	[0.0617]	[0.1085]	-
		0.8	-0.1043	-0.1063	-0.2649	-
			(0.0303)	(3.9120)	(0.1308)	-
			[0.1086]	[3.9135]	[0.2955]	-
	0.75	*0.2*	-0.0118	0.0020	-0.0412	-
			(0.0140)	(0.0410)	(0.0468)	-
			[0.0184]	[0.0410]	[0.0624]	-
		0.6	-0.0093	-0.0021	-0.0258	-
			(0.0102)	(0.1330)	(0.0351)	-
			[0.0138]	[0.1330]	[0.0436]	-
		0.8	-0.0075	0.0200	-0.0204	-
			(0.0073)	(0.4714)	(0.0322)	-
			[0.0105]	[0.4718]	[0.0381]	-
20	*0.15*	*0.2*	-0.0398	0.0016	-0.2774	-0.2800
			(0.0247)	(0.0483)	(0.0790)	(0.0523)
			[0.0468]	[0.0483]	[0.2885]	[0.2848]
		0.6	-0.0364	-0.0015	-0.1493	-0.2012
			(0.0183)	(0.0467)	(0.0559)	(0.0431)
			[0.0408]	[0.0467]	[0.1594]	[0.2057]
		0.8	-0.0400	-0.2267	-0.2450	-0.2783
			(0.0162)	(4.3662)	(0.0804)	(0.0570)
			[0.0431]	[4.3721]	[0.2578]	[0.2840]
	0.75	*0.2*	-0.0048	0.0014	-0.0498	-0.0455
			(0.0093)	(0.0240)	(0.0328)	(0.0203)
			[0.0104]	[0.0240]	[0.0596]	[0.0498]
		0.6	-0.0032	0.0013	-0.0267	-0.0301
			(0.0056)	(0.0308)	(0.0239)	(0.0167)
			[0.0065]	[0.0308]	[0.0358]	[0.0345]
		0.8	-0.0024	0.0178	-0.0172	-0.0213
			(0.0036)	(0.7941)	(0.0192)	(0.0135)
			[0.0044]	[0.7943]	[0.0257]	[0.0253]

Table 2. continued …

T	*β*	*γ*	*LSDV-bias* (*S.E.*) *[R.M.S.E]*	*AH-bias* (*S.E.*) *[R.M.S.E]*	*GMM103-bias* (*S.E.*) *[R.M.S.E]*	*GMM110-bias* (*S.E.*) *[R.M.S.E]*
30	*0.15*	*0.2*	-0.0259	0.0012	-0.3212	-0.2952
			(0.0204)	(0.0365)	(0.0651)	(0.0418)
			[0.0330]	[0.0365]	[0.3278]	[0.2981]
		0.6	-0.0238	0.0001	-0.1967	-0.2431
			(0.0143)	(0.0398)	(0.0538)	(0.0393)
			[0.0278]	[0.0398]	[0.2039]	[0.2462]
		0.8	-0.0229	0.0507	-0.2370	-0.2661
			(0.0112)	(1.3205)	(0.0612)	(0.0433)
			[0.0255]	[1.3214]	[0.2447]	[0.2696]
	0.75	*0.2*	-0.0031	-0.0002	-0.0551	-0.0466
			(0.0071)	(0.0177)	(0.0268)	(0.0158)
			[0.0078]	[0.0177]	[0.0613]	[0.0492]
		0.6	-0.0021	0.0001	-0.0299	-0.0311
			(0.0041)	(0.0212)	(0.0194)	(0.0127)
			[0.0046]	[0.0212]	[0.0356]	[0.0336]
		0.8	-0.0014	0.0002	-0.0176	-0.0205
			(0.0025)	(0.0428)	(0.0153)	(0.0101)
			[0.0029]	[0.0428]	[0.0233]	[0.0228]

N = 50, $\delta = 0.8$, $\sigma_\xi=3$, $\sigma_\varepsilon=0.75$.

B. In a world with serially correlated errors

Serially correlated errors introduce an upward bias. In this experiment, the serial correlation is modelled as an AR(1) process:

$$\varepsilon_{it} = \alpha\, \varepsilon_{it-1} + \eta_{it} \qquad\qquad \eta_{it} \sim N(0,\ \sigma_\eta^{\ 2}) \qquad (8)$$

We run the same simulations as above, but this time, we introduce serially correlated errors by setting $\alpha = 0.25$ and 0.95. In this case, the bias of LSDV estimations will be a combination of the downward Nickell bias as well as the upward bias from the correlation between y_{it-1} and ε_{it}. The bias of γ, given autocorrelated errors, increases with α and T.[9] Moreover, autocorrelated errors also introduce a downward bias to β, whose size increases with α and T. Table 3 below presents this result. For conciseness, the table only includes a subset of results whose parameters approximate that of typical growth regressions.

Table 3. Bias and standard errors of LSDV estimates given serially correlated errors

T	γ	*Serial Correlation of* ε			β	*Serial Correlation of* ε		
		0 *Bias (S.E.)*	*0.25* *Bias (S.E.)*	*0.95* *Bias (S.E.)*		*0* *Bias (S.E.)*	*0.25* *Bias (S.E.)*	*0.95* *Bias (S.E.)*
5	*0.6*	-0.1740(0.0557)	-0.1171(0.0570)	0.0811(0.0520)	*0.15*	0.0056(0.0233)	0.0051(0.0237)	-0.0035(0.0238)
	0.8	-0.2874(0.0671)	-0.2128(0.0663)	0.0703(0.0518)		-0.0038(0.0226)	-0.0031(0.0229)	0.0015(0.0243)
10	*0.6*	-0.0755(0.0313)	-0.0176(0.0299)	0.1938(0.0319)	*0.15*	0.0064(0.0121)	0.0021(0.0126)	-0.0186(0.0176)
	0.8	-0.1043(0.0303)	-0.0606(0.0300)	0.1226(0.0268)		0.0055(0.0115)	0.0021(0.0127)	-0.0051(0.0192)
20	*0.6*	-0.0364(0.0183)	0.0285(0.0179)	0.2796(0.0184)	*0.15*	0.0055(0.0071)	-0.0044(0.0080)	-0.0429(0.0128)
	0.8	-0.0400(0.0162)	-0.0040(0.0154)	0.1478(0.0151)		0.0054(0.0067)	0.0008(0.0076)	-0.0193(0.0144)
30	*0.6*	-0.0238(0.0143)	0.0429(0.0145)	0.3141(0.0127)	*0.15*	0.0040(0.0053)	-0.0078(0.0060)	-0.0565(0.0104)
	0.8	-0.0229(0.0112)	0.0098(0.0108)	0.1608(0.0104)		0.0045(0.0050)	-0.0019(0.0059)	-0.0299(0.0137)

N = 50, $\delta = 0.8$, $\sigma_\xi=3$, $\sigma_\varepsilon=0.75$.

Other estimators experience similar, albeit stronger, biases as a result of serially correlated errors. As shown in Table 4 below, these biases are negligible for low T's. However, as *T* increases, these biases become significant. A mild serial correlation, i.e., with α =0.25, induces a bias of more than 10 percent of the true value of the coefficients in AH estimation for $T \geq 10$. A similar bias is introduced in GMM estimations; however, this bias is less apparent for a low α as it tends to cancel the overfitting bias which has the opposite sign.

Table 4. Biases and standard errors of γ for various estimators given serially correlated errors

T	β	γ	*Estimation*	*Serial Correlation of* ε *0* *Bias (S.E.)*	*0.25* *Bias (S.E.)*	*0.95* *Bias (S.E.)*
5	*0.15*	*0.6*	*LSDV*	-0.1740(0.0557)	-0.1171(0.0570)	0.0811(0.0520)
			AH	0.0021(0.0921)	0.0349(0.0936)	0.0176(0.1385)
			GMM103	-0.0529(0.1011)	-0.0415(0.0978)	0.0077(0.1038)
		0.8	*LSDV*	-0.2874(0.0671)	-0.2128(0.0663)	0.0703(0.0518)
			AH	-0.5940(15.5292)	-0.0865(3.1085)	-0.1298(2.7771)
			GMM103	-0.3618(0.2626)	-0.2788(0.2388)	0.0330(0.1771)
10	*0.15*	*0.6*	*LSDV*	-0.0755(0.0313)	-0.0176(0.0299)	0.1938(0.0319)
			AH	0.0042(0.0615)	0.0699(0.0613)	0.1433(0.5723)
			GMM103	-0.0830(0.0699)	-0.0534(0.0670)	0.0636(0.0782)
	0.15	*0.8*	*LSDV*	-0.1043(0.0303)	-0.0606(0.0300)	0.1226(0.0268)
			AH	-0.1063(3.9120)	-0.3095(10.4990)	-0.0846(3.3137)
			GMM103	-0.2649(0.1308)	-0.1843(0.1196)	0.1112(0.0800)

Table 4. continued …

T	*β*	*γ*	*Estimation*	*Serial Correlation of ε* 0 Bias (S.E.)	0.25 Bias (S.E.)	0.95 Bias (S.E.)
20	0.15	0.6	LSDV	-0.0364(0.0183)	0.0285(0.0179)	0.2796(0.0184)
			AH	-0.0015(0.0467)	0.1144(0.0468)	0.8027(2.0128)
			GMM103	-0.1493(0.0559)	-0.0965(0.0582)	0.1544(0.0577)
			GMM110	-0.2012(0.0431)	-0.1320(0.0441)	0.1392(0.0465)
		0.8	LSDV	-0.0400(0.0162)	-0.0040(0.0154)	0.1478(0.0151)
			AH	-0.2267(4.3662)	0.9040(18.0152)	-0.8140(15.4561)
			GMM103	-0.2450(0.0804)	-0.1561(0.0741)	0.1360(0.0413)
			GMM110	-0.2783(0.0570)	-0.1874(0.0543)	0.1115(0.0344)
30	0.15	0.6	LSDV	-0.0238(0.0143)	0.0429(0.0145)	0.3141(0.0127)
			AH	0.0001(0.0398)	0.1420(0.0419)	0.8426(1.9379)
			GMM103	-0.1967(0.0538)	-0.1266(0.0507)	0.1915(0.0451)
			GMM110	-0.2431(0.0393)	-0.1560(0.0381)	0.1680(0.0364)
		0.8	LSDV	-0.0229(0.0112)	0.0098(0.0108)	0.1608(0.0104)
			AH	0.0507(1.3205)	0.0393(10.6171)	-1.0028(6.5026)
			GMM103	-0.2370(0.0612)	-0.1549(0.0572)	0.1426(0.0322)
			GMM110	-0.2661(0.0433)	-0.1826(0.0393)	0.1185(0.0281)

N = 50, $\delta = 0.8$, $\beta = 0.15$, $\sigma_\xi=3$, $\sigma_\varepsilon=0.75$.

V. Conclusions

We begin with the problem of LSDV bias and proposed solutions to it. After six dozen simulations, we have bad news and good news. The bad news is that for small T’s, LSDV produces significant biases and neither of the two proposed alternatives discussed in this paper provides a satisfactory solution to this problem. The good news is that with the improved quality and quantity of data for growth regression, very few macroeconomists need to worry about having panel datasets with $T < 20$. In such cases, LSDV estimations around the parameters relevant to growth regressions (high γ, low β) have negligible biases and are efficient.

With serially correlated disturbances, however, the bias can become unpredictable. Even if the serial correlations are so mild that they cannot be detected by existing tests, they still might cause significant biases in the estimates. The problem is since we do not know the degree of serial correlations, we cannot estimate the induced upward bias. Based on our simulations, we know that LSDV is the still the preferred technique for $T \geq 10$ when we suspect the presence of a mild degree of serial correlation. As for more severe serial correlations, there exist tests that can detect them.

With no serial correlation and $T \leq 10$, we cannot draw definite conclusions about preferred estimation techniques. There is clearly a bias-efficiency trade-off between LSDV and AH; yet in either case, neither one is satisfactory. Therefore, one must go beyond this study to find an appropriate estimator. Several alternatives are available. Kiviet (1995) suggests corrected LSDV, which Judson and Owen (1999) found to be the preferred estimator for balanced panels with small *T*'s. Hsiao et. al. (2001) argue for maximum likelihood and minimum distance estimators, while some others propose limited information maximum likelihood (LIML) and symmetrically normalised GMM (SNM) (Baltagi, 2001).

Notes

[1] See, Nickell (1981), Arellano and Bond (1991), Kiviet (1995) and Hsiao et. al. (2001).

[2] Judson and Owen (1999, p.9-10)

[3] For instance, Arellano and Bond (1991) find GMM to be more efficient than the Anderson-Hsiao estimator. However, Kiviet (1995) shows that, for different experimental designs, Anderson-Hsiao can be as efficient as an Arellano-Bond GMM estimator.

[4] Using lagged level does not make AH immune to the possibility of having singularity points; it just narrows the range of parameters that can generate such problems.

[5] Islam (1995), Table IV.

[6] *Ibid,* Appendix 1.

[7] This makes sense if we consider Kiviet's (1995) derivation of the bias being of $O(N^{-1}T^{-3/2})$, where the order of the bias is not dependent on the value of γ. Therefore, a lower γ implies a higher relative bias.

[8] For example, for T=5, β =0.75 and γ = 0.6, AH gives a highly biased and inefficient estimate. It is possible that this was an isolated case – subsequent experiments with identical parameters were unable to replicate this result with such a high bias and standard errors. However, we suspect it is an instance of the 'singularity point' problem pointed out by Arellano (1989).

[9] For an analytical derivation of this bias, see Appendix 1.

Reference

Anderson, T. and C. Hsiao. 1981. "Estimation of dynamic models with error components." *Journal of the American Statistical Association* **76:** 598-606.

Arellano, M. 1989. "A note on the Anderson-Hsiao estimator for panel data." *Economic Letters* **31:** 337-41.

Arellano, M. and S. Bond. 1991. "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equation." *Review of Economic Studies* **58:** 277-97.

_______ and _______ 1998. "Using DPD98 for Gauss: A Guide for Users." Available from ftp://ftp.cemfi.es/pdf/papers/ma/dpd98.pdf

Baltagi, B. 2001. *Econometric Analysis of Panel Data.* Chichester: John Wiley and Sons.

Caselli, F., Esquivel, G. and Lefort, F. 1996. "Reopening the convergence debate: a new look at cross-country growth empirics." *Journal of Economic Growth* **1**(3): 363-90.

Hsiao, C., M. Pesaran and A. Tahmiscioglu. 2001. "Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods." Available from http://www.econ.cam.ac.uk/faculty/pesaran/Panmaxr1.pdf

Islam, N. 1995. "Growth empirics: A panel data approach." *Quarterly Journal of Economics* **110:** 1127-1170.

Judson, R. and Owen, A. 1999. "Estimating dynamic panel data models: a guide for macroeconomists." *Economic Letters* **65:** 9-15.

Kiviet, J. 1995. "On bias, inconsistency, and efficiency of various estimators in dynamic panel data models." *Journal of Econometrics* **68:** 53-78.

Nickell, S. 1981. "Biases in dynamic models with fixed effects." *Econometrica* **49**(6)**:** 1417-26.

Appendix 1. Bias of LSDV estimation in dynamic models with autocorrelated errors

In the following, we derive the total bias of LSDV estimates given AR(1) disturbances as in Eq. (8). The derivation extends Eqs.(7) to (14) of Nickell (1981, p.1419-21). We have the dynamic model

$$y_{it} = \gamma y_{it-1} + \mu_i + \varepsilon_{it} \qquad i = 1, \ldots, N;\ t = 1, \ldots T \qquad \text{(A1)}$$

with no exogenous regressor. We then do the usual step to eliminate the individual effects and arrive at

$$y_{it} - y_{i.} = \gamma(y_{it-1} - y_{it.-1}) + (\varepsilon_{it} - \varepsilon_{i.}) \qquad \text{(A2)}$$

which is analogous to Eq. (3) above. The asymptotic bias of this specification is

$$\underset{N\to\infty}{plim}(\hat{\gamma} - \gamma) = \frac{plim_{N\to\infty} 1/N \sum_{i=1}^{N}(y_{it-1} - y_{i.-1})(\varepsilon_{it} - \varepsilon_{i.})}{plim_{N\to\infty} 1/N \sum_{i=1}^{N}(y_{it-1} - y_{i.-1})^2} \equiv \frac{A_t}{B_t} \qquad \text{(A3)}$$

Replacing plims as N→∞ with expectations across i, E_i, we obtain

$$A_t = E_i\, y_{it-1}.\varepsilon_{it} - E_i\, y_{it-1}.\varepsilon_{i.} - E_i\, y_{i.-1}.\varepsilon_{it} + E_i\, y_{i.-1}.\varepsilon_{i.} \qquad \text{(A4)}$$

while Eq. (A1) implies

$$y_{it} = \mu_i/(1-\gamma) + \sum_{j=0}^{\infty}\gamma^i \varepsilon_{it-j} \qquad \text{(A5)}$$

Assuming no autocorrelated errors, Nickell [1981, p. 1420, Eq. (11)] derives the last three terms on the right hand side of Eq. (A4) to be

$$A_t^N = -\frac{\sigma_\varepsilon^2}{T(1-\gamma)}\left\{1 - \gamma^{t-1} - \gamma^{T-t} + \frac{1}{T}\frac{(1-\gamma^T)}{(1-\gamma)}\right\}. \qquad \text{(A6)}$$

Meanwhile, the first term on the right hand side of (A4) simplifies to

$$
\begin{aligned}
Ei\ y_{it-1}.\varepsilon_{it} &= E_i \sum_{j=0}^{\infty} \gamma^j .\varepsilon_{t-j-1}.\varepsilon_{it} \\
&= E_i \left[\varepsilon_{t-1}{}^2 + \gamma.\varepsilon_{t-2}{}^2 + \ldots + \varepsilon_{t-1}.\gamma\varepsilon_{t-1} + \varepsilon_{t-1}.\gamma^2\varepsilon_{t-1}\right] \\
&= \sigma_\varepsilon^2\left[1+\gamma^2+\gamma^4+\ldots+2\alpha\gamma+2\alpha\gamma^3+\ldots+2\alpha^2\gamma^2+2\alpha^2\gamma^4+\ldots\right] \\
&= \sigma_\varepsilon^2\left\{\frac{1}{1-\gamma^2}+\frac{2\alpha\gamma}{(1-\gamma^2)}\left(1+\alpha\gamma+\alpha^2\gamma^2+\ldots\right)\right\} \\
&= \sigma_\varepsilon^2\left\{\frac{1}{1-\gamma^2}+\frac{2\alpha\gamma}{(1-\gamma^2)(1-\alpha.\gamma)}\right\} \\
&= \sigma_\varepsilon^2\left\{\frac{1+\alpha\gamma}{(1-\gamma^2)(1-\alpha.\gamma)}\right\} \qquad \text{(A7)}
\end{aligned}
$$

Hence under serial correlation, the numerator of the bias becomes

$$
A_t = \sigma_\varepsilon^2\left\{\frac{1+\alpha\gamma}{(1-\gamma^2)(1-\alpha.\gamma)}\right\} - \frac{\sigma_\varepsilon^2}{T(1-\gamma)}\left\{1-\gamma^{t-1}-\gamma^{T-t}+\frac{1}{T}\frac{(1-\gamma^T)}{(1-\gamma)}\right\} \qquad \text{(A8)}
$$

Meanwhile, taking t = 1 to simplify calculation, the denominator of the bias is

$$
\begin{aligned}
B_t &= E_i\left(\sum_{j=0}^{\infty}\gamma^j\varepsilon_{it-j-1} - \frac{1}{T}\sum_{s=1}^{T}\sum_{j=0}^{\infty}\gamma^j\varepsilon_{is-j-1}\right)^2 \\
&= E_i\left(\sum_{j=0}^{\infty}\gamma^j\varepsilon_{it-j-1}\right)^2 - \frac{2}{T}E_i\left(\sum_{j=0}^{\infty}\gamma^j\varepsilon_{it-j-1}\right)\left(\frac{1}{T}\sum_{s=1}^{T}\sum_{j=0}^{\infty}\gamma^j\varepsilon_{is-j-1}\right) \\
&\quad + \frac{1}{T^2}E_i\left(\sum_{s=1}^{T}\sum_{j=0}^{\infty}\gamma^j\varepsilon_{is-j-1}\right)^2 \\
&= \sigma_\varepsilon^2\left\{\frac{1+\alpha\gamma}{(1-\gamma^2)(1-\alpha\gamma)}\right\} - \frac{2\sigma_\varepsilon^2}{T}\left\{\frac{1+\alpha\gamma}{(1-\gamma^2)(1-\alpha\gamma)}\right\}\left\{\frac{1-\alpha^{T-1}}{1-\alpha}\right\} \\
&\quad + \frac{2\sigma_\varepsilon^2}{T^2}\left\{\frac{1+\alpha\gamma}{(1-\gamma^2)(1-\alpha\gamma)}\right\}\left\{\frac{T\alpha-\dfrac{1-\alpha^{T-1}}{1-\alpha}}{1-\alpha}\right\} \qquad \text{(A9)}
\end{aligned}
$$

Without writing out A_t/B_t explicitly, we notice two differences between this and the Nickell bias. First*,* the bias of LSDV in the presence of AR(1) autocorrelated errors is not of *O(1/T)*. In fact, when the first term of A_t in Eq. (A8) dominates, the bias increases in T. Second*,* the sign of the bias is ambiguous: it is positive when the first right hand side term in Eq. (A8) is greater than the Nickell bias. Finally, the bias increases in α.