

# Estimating Dynamic Panel Data Models: A Practical Guide for Macroeconomists

Ruth A. Judson  
rjudson@frb.gov

Ann L. Owen  
aowen@frb.gov

Federal Reserve Board of Governors\*  
January 1996

## Abstract

Previous research on dynamic panel estimation has focused on panels that, unlike a typical panel of macroeconomic data, have small time dimensions and large individual dimensions. We use a Monte Carlo approach to investigate the performance of several different methods designed to reduce the bias of the estimated coefficients for the longer, narrower panels commonly found for macro data. We find that the bias of the least squares dummy variable approach can be significant, even when the time dimension of the panel is as large as 30. For panels with small time dimensions, we find a corrected least squares dummy variable estimator to be the best choice. However, as the time dimension of the panel increases, the computationally simpler Anderson-Hsiao estimator performs equally well. We apply our recommendations to a panel of countries to show that increases in income growth precede increases in savings rates and increases in savings rates precede declines in income growth.

JEL Codes: C23, O11, E00

Keyword: Panel data, simulation, dynamic model, macroeconomics, growth

---

\*20th & C Sts., N.W., Washington, D.C. 20551. Phone: (202)736-5612 Fax: (202)452-2301. This paper represents the views of the authors and should not be interpreted as reflecting those of the Board of Governors of the Federal Reserve System or other members of its staff.

# Estimating Dynamic Panel Data Models: A Practical Guide for Macroeconomists

## 1 Introduction

The recent revitalization of interest in long-run growth and the availability of macroeconomic data for large panels of countries has generated interest among macroeconomists in estimating dynamic models with panel data. (See, e.g., Mankiw, Romer, and Weil (1992), Fischer (1993), and Levine and Renelt (1992).) Use of panel data in estimating common relationships across countries is particularly appropriate because it allows the identification of country-specific effects that control for missing or unobserved variables.<sup>1</sup> However, microeconomists have generally been more avid users of panel data, and, thus, existing panel data techniques have been devised and tested with the typical dimensions of a microeconomic panel dataset in mind.

Many macroeconomic panel datasets have a time dimension far greater and an individual (country) dimension far smaller than the typical microeconomic panel. This difference is important in choosing an estimation technique for two reasons. First, it is well known that using dummy variables to estimate individual effects in a model which includes a lagged value of the dependent variable results in biased estimates when the time dimension of the panel ( $T$ ) is small. Nickell (1981) derives a formula for the bias, showing that the bias approaches zero as  $T$  approaches infinity. Thus, for many macroeconomists, the question, "How big should  $T$  be before the bias can be ignored?", is a critical one. A second reason that typical macro panels may require different estimation techniques than those used on micro panels is that recent work investigating the appropriateness of competing estimators has generated conflicting results, showing that the characteristics of the data influence the performance of an estimator. Arellano and Bond (1991) run a Monte Carlo experiment to judge the performance of the Anderson-Hsiao estimator against various GMM estimators and find that the GMM procedures produce substantial efficiency gains. However, Kiviet (1995), using a slightly different experimental design, finds that the Anderson-Hsiao estimator compares favorably to GMM and concludes that no estimator has

---

<sup>1</sup>Many recent studies use panel data but do not use techniques that exploit the panel dimension of the data.

been found to be the appropriate choice in all circumstances. Our findings support this conclusion and suggest that the best technique changes with the size of the panel.

In this paper, we evaluate several different techniques for estimating dynamic models with panels characteristic of many macroeconomic panel datasets; our goal is to provide a guide to choosing appropriate techniques for panels of various dimensions. We build on previous work in this area that has examined dynamic panel data estimators both theoretically and with simulations. Anderson and Hsiao (1981) derive an instrumental variables approach. Holtz-Eakin, Newey and Rosen (1988) expand on the Anderson-Hsiao approach, showing how to implement it to estimate a vector autoregression with time-varying parameters. Arellano and Bond (1991) use Monte Carlo studies to evaluate a GMM estimator that is very similar to the Holtz-Eakin et. al. recommendation, and Kiviet (1995) uses simulations to compare these and several other techniques, including a corrected least squares dummy variable estimator he develops in his paper. Our work most closely follows Kiviet's, however, we focus our attention on data with the qualities normally encountered by macroeconomists while he focuses on the short (small T), wide (large N) panels typical of micro data.

We have three main conclusions. First, macroeconomists should not dismiss the least squares dummy variable bias as insignificant. Even with a time dimension as large as 30, we find that the bias may be equal to as much as 20% of the true value of the coefficient of interest. Second, a "restricted GMM" estimator that uses a subset of the available lagged values as instruments increases computational efficiency without significantly detracting from its effectiveness. Finally, the size of the panel influences the choice of estimator. For panels with a small time dimension, we find a corrected least squares dummy variable to be the best choice. However, as the time dimension of the panel increases, the computationally simpler Anderson-Hsiao estimator performs equally well.

In the final section of the paper, we demonstrate the importance of these findings by applying our recommendations to a panel of countries in order to learn more about the dynamic relationship between savings and growth. When using the appropriate technique for the size of the panel, we are able to establish that increases in income growth precede increases in savings and increases in savings precede declines in income growth.

We develop these results in the next five sections. Section 2 sets up the model we will estimate, and reviews the problems caused by adding a lagged dependent variable to a panel data

regression and discusses the proposed solutions, Section 3 describes our methodology, Section 4 provides results, Section 5 implements our recommendations in tests of Granger causality between savings and growth, and Section 6 concludes.

## 2 The Problem and Proposed Solutions

We consider a dynamic fixed effects model of the form

$$y_{i,t} = \gamma y_{i,t-1} + x'_{i,t} \beta + \eta_i + \epsilon_{i,t} \quad (1)$$

where  $\eta_i$  is a fixed-effect,  $x_{i,t}$  is a  $(K-1) \times 1$  vector of exogenous regressors and  $\epsilon_{i,t} \sim N(0, \sigma_\epsilon^2)$  is a random disturbance. We assume

$$\begin{aligned} \sigma_\epsilon^2 &\geq 0, \\ E(\epsilon_{i,t}, \epsilon_{j,s}) &= 0 \quad i \neq j \text{ or } t \neq s \\ E(\eta_i, \epsilon_{j,t}) &= 0 \quad \forall i, j, t \\ E(x_{i,t}, \epsilon_{j,s}) &= 0 \quad \forall i, j, t, s \end{aligned} \quad (2)$$

Equation 1 is a common specification for those wishing to estimate a VAR or test for Granger causality.

The fixed effects model we have chosen is a common choice for macroeconomists. It is generally more appropriate than a random effects model for many macro datasets for two reasons. First, if the individual effect represents omitted variables, it is highly likely that these country-specific characteristics are correlated with the other regressors. Second, it is also fairly likely that a typical macro panel will contain most of the countries of interest and, thus, will be less likely to be a random sample from a much larger universe of countries (e.g., an OECD panel is likely to contain all of the OECD countries and not just a random sample of them).

The model in Equation (1), however, includes as one of the regressors a lagged dependent variable. In this case, the usual approach to estimating a fixed-effects model -- the least squares dummy variable estimator (LSDV) -- generates a biased estimate of the coefficients. Nickell (1981) derives an expression for the bias of  $\gamma$  when there are no exogenous regressors, showing that the bias approaches zero as  $T$  approaches infinity. Thus, the LSDV estimator only performs well when the time dimension of the panel is large.

Several estimators have been proposed to estimate Equation (1) when T is not large.<sup>2</sup> Anderson and Hsiao (1981) propose two instrumental variable procedures. To remove the fixed effect, they first difference Equation (1) to obtain

$$(y_{i,t} - y_{i,t-1}) = \gamma(y_{i,t-1} - y_{i,t-2}) + (x_{i,t} - x_{i,t-1})'\beta + (\epsilon_{i,t} - \epsilon_{i,t-1}) \quad (3)$$

In the differenced equation, however, the errors  $(\epsilon_{i,t} - \epsilon_{i,t-1})$  are now correlated with the one of the independent variables  $(y_{i,t-1} - y_{i,t-2})$ , and they recommend instrumenting for  $(y_{i,t-1} - y_{i,t-2})$  with either  $y_{i,t-2}$  or  $(y_{i,t-2} - y_{i,t-3})$  which are uncorrelated with the disturbance in (3) but correlated with  $(y_{i,t-1} - y_{i,t-2})$ . Arellano (1989) shows that using the lagged difference as an instrument results in an estimator that has a very large variance. Arellano and Bond (1991) and Kiviet (1995) confirm the superiority of using the lagged level as an instrument with simulation results, and we, therefore, focus our attention only on this estimator (subsequently termed the AH estimator). Thus, when the dimension of the panel is  $N \times T$ , the Anderson-Hsiao estimator we employ is

$$\hat{\delta}_{AH} = (Z'X)^{-1}Z'Y \quad (4)$$

where Z is a  $K \times N(T-2)$  matrix of instruments, X is a  $K \times N(T-2)$  matrix of regressors and Y is an  $N(T-2) \times 1$  vector of dependent variables. Let  $\Delta y_{i,t} = y_{i,t} - y_{i,t-1}$ . Then,

$$Z_i = \begin{bmatrix} y_{i,1} & \Delta x_{i,3} \\ \cdot & \cdot \\ \cdot & \cdot \\ y_{i,T-2} & \Delta x_{i,T} \end{bmatrix} \quad X_i = \begin{bmatrix} \Delta y_{i,2} & \Delta x_{i,3} \\ \cdot & \cdot \\ \cdot & \cdot \\ \Delta y_{i,T-1} & \Delta x_{i,T} \end{bmatrix} \quad Y_i = \begin{bmatrix} \Delta y_{i,3} \\ \cdot \\ \cdot \\ \Delta y_{i,T} \end{bmatrix} \quad (5)$$

$$Z = \begin{bmatrix} Z_1 \\ \cdot \\ \cdot \\ Z_N \end{bmatrix} \quad X = \begin{bmatrix} X_1 \\ \cdot \\ \cdot \\ X_N \end{bmatrix} \quad Y = \begin{bmatrix} Y_1 \\ \cdot \\ \cdot \\ Y_N \end{bmatrix} \quad (6)$$

We also examine two GMM estimators suggested by Arellano and Bond (1991). The

---

<sup>2</sup>Further details on computational issues are discussed in the appendix. Gauss programs are available from the authors upon request.

Anderson-Hsiao estimator can be considered a special case of the GMM procedures, which also remove the individual effect by differencing Equation (1) to obtain Equation (3). The GMM procedures, however, gain efficiency by exploiting additional moment restrictions. They use all available lagged values of the dependent variables plus lagged values of the exogenous regressors as instruments.<sup>3</sup> The GMM estimators take the form

$$\hat{\delta}_{GMM} = (X'Z^*A_NZ^{*\prime}X)^{-1}X'Z^*A_NZ^{*\prime}Y \quad (7)$$

where X and Y are as defined above, but  $Z_i^*$  is a block diagonal matrix whose  $s$ th block is given by  $(y_{i1} \dots y_{is} x_{i1} \dots x_{i(s+1)})$  for  $s=1, \dots, T-2$ . Then  $Z^*=(Z_1^*, \dots, Z_N^*)'$ .

Two different choices for  $A_N$  result in two different GMM estimators. A one-step estimator, GMM1, can be found by using

$$A_N = \left( \frac{1}{N} \sum_i^N Z_i^{*\prime} H Z_i^* \right)^{-1} \quad (8)$$

where H is a T-2 square matrix with twos in the main diagonals, minus ones in the first subdiagonals, and zeros otherwise. A two-step estimator, GMM2, is found by letting

$$A_N = \left( \frac{1}{N} \sum_i^N Z_i^{*\prime} \Delta \hat{\epsilon}_i \Delta \hat{\epsilon}_i' Z_i^* \right)^{-1} \quad (9)$$

where  $\Delta \hat{\epsilon}_i = (\Delta \hat{\epsilon}_{i3}, \dots, \Delta \hat{\epsilon}_{iT})$  are the residuals from a consistent one-step estimator of  $\Delta y_i$  (we use GMM1).

An alternative estimation procedure is suggested by Kiviet (1995). He derives a formula for the bias of the LSDV estimator and recommends subtracting this from the estimated LSDV coefficients. While the formula for the LSDV correction is presented in the appendix, estimation of the LSDV correction also involves a two-step procedure in which the residuals from a first-step consistent estimator (we use AH) are used in the second-stage calculation of the bias. We call Kiviet's corrected LSDV estimator LSDVC.

### 3 Methodology

---

<sup>3</sup>When  $x_{it}$  are predetermined but not strictly exogenous, only lagged values of x are valid instruments. If  $x_{it}$  are strictly exogenous, then current and lagged values are valid instruments.

In this section, we describe the Monte Carlo analysis we used to investigate several questions relevant for the macroeconomist wishing to choose one estimator from the several different techniques to estimate a dynamic panel data model. First, we examine the bias of OLS and LSDV estimators to determine how their magnitudes vary with characteristics of the dataset. Next, we explore various GMM estimators to see which technique produces the most efficient and consistent estimate. Finally, we make an overall comparison between the LSDV, AH, GMM and LSDVC.

Our data generation process closely follows Kiviet (1995). The model for  $y_{it}$  is given in Equation (1);  $x_{it}$  was generated with

$$x_{i,t} = \rho x_{i,t-1} + \xi_{i,t} \quad \xi_{i,t} \sim N(0, \sigma_\xi^2) \quad (10)$$

Thus, in addition to  $\beta$ ,  $\rho$  and  $\sigma_\xi^2$  also determine the correlation between  $y_{it}$  and  $x_{it}$ . Kiviet defines a signal to noise ratio,  $\sigma_s^2$

$$\sigma_s^2 = \text{var}(v_{it} - \epsilon_{i,t}), \quad v_{i,t} \equiv y_{i,t} - \frac{1}{1-\gamma} \eta_i \quad (11)$$

and shows that it can be calculated from other parameters of the model as follows

$$\sigma_s^2 = \beta^2 \sigma_\xi^2 \left[ 1 + \frac{(\gamma + \rho)^2}{1 + \gamma \rho} [\gamma \rho - 1] - (\gamma \rho)^2 \right]^{-1} + \frac{\gamma^2}{1 - \gamma^2} \sigma_\epsilon^2 \quad (12)$$

The higher the signal-to-noise ratio, the more useful  $x_{it}$  is in explaining  $y_{it}$ . Kiviet (1995) finds that varying the signal-to-noise ratio significantly alters the relative performance of the bias of the various estimators.

We also choose  $\beta=1-\gamma$  so that a change in  $\gamma$  affects only the short-run dynamic relationship between  $x$  and  $y$  and not the steady-state relationship. Thus, given choices for  $\gamma$ ,  $\sigma_\epsilon^2$ ,  $\sigma_s^2$ , and  $\rho$ , all of the other parameters of the model are determined. Our parameter choices can be summarized as follows:  $\sigma_\epsilon^2$  is normalized to 1,  $\rho$  is set at the intermediate value of 0.5,  $\sigma_s^2$  alternates between a value of 2 and 8, and  $\gamma$  alternates between 0.2 and 0.8. For each combination of parameters we vary the size of our panel.  $N$ , the cross-sectional dimension, takes on values of 20 or 100, and  $T$ , the time dimension, is assigned values of 5, 10, 20 and 30. In total, we have 32 different parameter combinations.

We generate the data by choosing  $x_{i,0}$ ,  $y_{i,0} = 0$  and then discarding the first 50 observations

before selecting our sample. In general we performed 1,000 replications, although for the GMM comparisons for  $T=20$ ,  $N=100$  and  $T=30$ , it was only feasible to do 100 replications. We used a fixed set of seeds for the random number generator so that our results can be replicated.

## 4 Results

The first issue we addressed was the magnitude of the bias of the OLS and LSDV estimators for various panel sizes. Table 1 summarizes the results from this initial experiment for a subset of parameter values (results for the full set of parameter values are in an appendix available from the authors upon request.)

<b>Table 1: OLS and LSDV Bias Estimates</b>						
$\sigma_\epsilon=1, \sigma_\xi=2, \rho=0.5, N=100$						
<b>T</b>	<b><math>\gamma</math></b>	<b><math>\gamma</math> Bias</b>		<b><math>\beta</math></b>	<b><math>\beta</math> Bias</b>	
		<b>OLS (SE)</b>	<b>LSDV (SE)</b>		<b>OLS (SE)</b>	<b>LSDV (SE)</b>
5	0.2	0.225 (0.039)	-0.147 (0.040)	0.8	-0.098 (0.044)	0.006 (0.045)
	0.8	0.049 (0.026)	-0.504 (0.058)	0.2	-0.005 (0.055)	-0.027 (0.070)
10	0.2	0.225 (0.032)	-0.059 (0.023)	0.8	-0.099 (0.031)	0.015 (0.026)
	0.8	0.049 (0.017)	-0.232 (0.032)	0.2	-0.007 (0.037)	0.002 (0.045)
20	0.2	0.225 (0.028)	-0.027 (0.015)	0.8	-0.100 (0.023)	0.009 (0.017)
	0.8	0.049 (0.012)	-0.104 (0.019)	0.2	-0.008 (0.026)	0.006 (0.028)
30	0.2	0.226 (0.026)	-0.017 (0.012)	0.8	-0.100 (0.019)	0.006 (0.014)
	0.8	0.049 (0.011)	-0.066 (0.014)	0.2	-0.008 (0.020)	0.006 (0.022)

The results in Table 1 suggest that, with either method of estimation, the bias of the  $\gamma$  estimate is more severe than the bias of the estimate of  $\beta$ . They also confirm the well-documented result that using OLS to estimate a model with fixed effects generates a significant bias, even as  $T$  gets large. As Nickell (1981) predicts, the bias of the LSDV estimator increases with  $\gamma$  and decreases with  $T$ . For  $T=5$  and  $T=10$ , the magnitudes of the bias are quite large: for  $T=10$ , just under 30% and over 50% for  $T=5$ . Moreover, the bias of the LSDV estimate of  $\gamma$  is



not insignificant, even at  $T=20$ . An estimate that gives a bias within 2 standard deviations of our average bias would result in a bias that could be from 2% to 28% of the true value of  $\gamma$ . When  $T=30$ , the average bias becomes significantly smaller, although the LSDV does not become more efficient. Based on the results in Table 1, one could expect an LSDV estimate with a bias from 3% to 20% of the true value of the coefficient even when  $T=30$ . It is worthwhile to note, however, that errors of this magnitude would still result in an estimate with the correct sign, and that the bias of the estimates of  $\beta$  are relatively small. However, as discussed below, since easy to implement alternative techniques allow the estimation of these coefficients with greater precision, it is difficult to justify using LSDV, even when  $T>20$ .

Since in many circumstances, the LSDV estimator is not an appropriate choice, we explore the other estimators described above. Before making an overall comparison, however, we first narrow our selection by comparing various GMM procedures. Arellano and Bond (1991) discuss two variants of a GMM procedure that use all lagged values as instruments. When  $T$  gets large, however, computational requirements increase substantially. Therefore, we compare various "restricted GMM" estimators in which the number of lagged values used as instruments is reduced to 2, 3, 5, and 8 instruments.<sup>4</sup> These estimates were computed using 1000 draws for  $T=10$  and  $T=20$  and 100 draws for  $T=30$ ; the cases where  $T=5$  were omitted since the maximum number of instruments available in those cases is three. (Note that the Anderson-Hsiao estimator that uses only one lag of the dependent variable as an instrument can be thought of as a "fully restricted" GMM2 estimator.) Table 2 presents the results of these comparisons.

---

<sup>4</sup>In addition, we checked to see whether there were gains to iterating GMM or to using more than eight instruments for  $T=20$  (where 18 instruments would be available). There was no gain to either of these procedures.

Table 2: GMM Bias Estimates for $\gamma$										
$\sigma_\epsilon=1 \quad \rho=0.5 \quad \sigma_\xi=2$										
T	N	$\gamma$	GMM1				GMM2			
			2 Inst (SE)	3 Inst (SE)	5 Inst (SE)	8 Inst (SE)	2 Inst (SE)	3 Inst (SE)	5 Inst (SE)	8 Inst (SE)
10	20	0.2	-0.041 (0.064)	-0.045 (0.061)	-0.050 (0.057)	-0.052 (0.056)	-0.043 (0.081)	-0.053 (0.087)	-0.077 (0.102)	-0.091 (0.112)
		0.8	-0.222 (0.139)	-0.235 (0.131)	-0.241 (0.111)	-0.239 (0.099)	-0.249 (0.168)	-0.281 (0.174)	-0.336 (0.198)	-0.377 (0.220)
	100	0.2	-0.010 (0.032)	-0.011 (0.031)	-0.012 (0.029)	-0.012 (0.029)	-0.009 (0.036)	-0.010 (0.034)	-0.011 (0.032)	-0.012 (0.031)
		0.8	-0.056 (0.063)	-0.065 (0.060)	-0.079 (0.057)	-0.086 (0.055)	-0.058 (0.072)	-0.069 (0.069)	-0.081 (0.063)	-0.087 (0.059)
20	20	0.2	-0.032 (0.042)	-0.033 (0.039)	-0.037 (0.037)	-0.037 (0.035)	-0.084 (0.110)	-0.146 (0.141)	-0.263 (0.196)	-0.268 (2.703)
		0.8	-0.137 (0.077)	-0.140 (0.071)	-0.147 (0.064)	-0.137 (0.051)	-0.424 (0.225)	-0.576 (0.274)	-0.840 (0.487)	-0.986 (1.756)
	100	0.2	-0.005 (0.019)	-0.006 (0.018)	-0.007 (0.017)	-0.008 (0.016)	-0.005 (0.020)	-0.004 (0.020)	-0.008 (0.022)	-0.015 (0.027)
		0.8	-0.028 (0.037)	-0.031 (0.034)	-0.037 (0.032)	-0.045 (0.031)	-0.028 (0.037)	-0.029 (0.035)	-0.043 (0.040)	-0.069 (0.049)
30	20	0.2	-0.025 (0.036)	-0.025 (0.035)	-0.027 (0.034)	-0.027 (0.032)	-0.176 (0.143)	-0.292 (0.184)	-0.410 (0.323)	-0.093 (0.333)
		0.8	-0.113 (0.062)	-0.112 (0.052)	-0.112 (0.049)	-0.101 (0.042)	-0.825 (0.293)	-1.034 (0.306)	-1.139 (0.508)	-0.733 (1.504)
	100	0.2	-0.004 (0.015)	-0.005 (0.014)	-0.006 (0.014)	-0.007 (0.013)	-0.004 (0.017)	-0.006 (0.019)	-0.016 (0.023)	-0.043 (0.029)
		0.8	-0.026 (0.028)	-0.027 (0.026)	-0.028 (0.024)	-0.034 (0.022)	-0.030 (0.033)	-0.036 (0.034)	-0.080 (0.044)	-0.187 (0.073)

Note: 1000 draws were used for T=10 and T=20, N=20. For T=20, N=100 and T=30, it was only feasible to do 100 draws.

The results in Table 2 indicate that the one-step GMM1 procedure outperforms the two-step GMM2 estimator both in terms of producing a smaller bias and a lower standard deviation of

the estimates.<sup>5</sup> Interestingly, increasing the number of instruments used appears to create a tradeoff between the average bias and the efficiency of the estimator. Using a GMM1 estimator with only two instruments produces a smaller expected bias in most cases, but using the full set of instruments almost always increases the efficiency of the estimates.

In our final experiment, we make an overall comparison between OLS, LSDV, AH, GMM and LSDVC. Based on the GMM comparisons reported above, we focus only on two restricted GMM estimators--GMM12 is GMM1 using only 2 lagged values as instruments and GMM15 is GMM1 using only 5 lagged values as instruments. Results for a subset of parameter values are presented in Tables 3 and 4. Results for the full parameter set are qualitatively the same and are presented in the appendix.<sup>6</sup> Table 3 shows the average bias and standard deviations of our estimates of  $\gamma$ , while Table 4 shows these results for our estimates of  $\beta$ . The bias of the estimates of  $\beta$  shown in Table 4 are relatively small for all techniques except OLS, and, thus, cannot be used to distinguish between estimators. As a result, we will focus our attention on estimates of  $\gamma$ .

The results in Table 3 show that, as to be expected, all the estimators (with the exception of OLS) generally perform better with a larger N. However, the GMM estimators seem to show the largest improvement (largest percent reduction in average bias) as N increases, while the uncorrected LSDV estimator shows improvement in only a few cases. Even with the relatively large improvement in the GMM estimates, GMM methods do not produce a superior estimate in terms of either average bias or efficiency, and the relative performance of these estimators does not generally change as N gets larger.

---

<sup>5</sup>These results are consistent with the findings of Arellano and Bond (1991) who find that in two of the three parameterizations they discuss, GMM1 outperforms GMM2 in both of these dimensions. Kiviet (1995) also finds that GMM1 is more efficient than GMM2, but reports more cases in which GMM2 produces a slightly lower bias.

<sup>6</sup>Because the non-GMM methods are substantially less computationally intensive, we were able to confirm the non-GMM results with 10,000 draws. These results, as well as other diagnostics such as RMSEs are available upon request.

<b>Table 3: Bias Estimates for <math>\gamma</math> Using Various Estimators</b>								
$\sigma_\epsilon=1 \quad \rho=0.5 \quad \sigma_\xi=2$								
<b>N</b>	<b>T</b>	<b><math>\gamma</math></b>	<b>OLS (SE)</b>	<b>LSDV (SE)</b>	<b>LSDVc (SE)</b>	<b>AH (SE)</b>	<b>GMM12 (SE)</b>	<b>GMM15 (SE)</b>
5	20	0.2	0.207 (0.092)	-0.148 (0.089)	0.001 (0.110)	0.017 (0.194)	-0.062 (0.121)	---
		0.8	0.033 (0.062)	-0.507 (0.130)	-0.190 (0.161)	0.061 (1.862)	-0.394 (0.284)	---
	100	0.2	0.225 (0.039)	-0.147 (0.040)	-0.006 (0.046)	0.001 (0.077)	-0.014 (0.057)	---
		0.8	0.049 (0.026)	-0.504 (0.058)	-0.131 (0.080)	0.007 (0.202)	-0.126 (0.167)	---
10	20	0.2	0.210 (0.072)	-0.060 (0.049)	0.000 (0.052)	0.005 (0.098)	-0.041 (0.064)	-0.050 (0.057)
		0.8	0.038 (0.042)	-0.238 (0.072)	-0.049 (0.081)	0.013 (0.218)	-0.222 (0.139)	-0.241 (0.111)
	100	0.2	0.225 (0.032)	-0.059 (0.023)	0.000 (0.024)	0.000 (0.043)	-0.010 (0.032)	-0.012 (0.029)
		0.8	0.049 (0.017)	-0.232 (0.032)	-0.032 (0.041)	0.000 (0.088)	-0.056 (0.063)	-0.079 (0.057)
20	20	0.2	0.213 (0.061)	-0.028 (0.033)	-0.001 (0.034)	0.001 (0.063)	-0.032 (0.042)	-0.037 (0.037)
		0.8	0.041 (0.030)	-0.108 (0.040)	-0.007 (0.046)	0.003 (0.118)	-0.137 (0.077)	-0.147 (0.064)
	100	0.2	0.225 (0.028)	-0.027 (0.015)	0.000 (0.015)	0.001 (0.027)	-0.005 (0.019)	-0.007 (0.017)
		0.8	0.049 (0.012)	-0.104 (0.019)	-0.005 (0.023)	0.001 (0.050)	-0.028 (0.037)	-0.037 (0.032)
30	20	0.2	0.214 (0.057)	-0.018 (0.026)	0.000 (0.027)	-0.001 (0.049)	-0.025 (0.036)	-0.027 (0.034)
		0.8	0.043 (0.025)	-0.068 (0.030)	-0.001 (0.034)	0.003 (0.088)	-0.113 (0.062)	-0.112 (0.049)
	100	0.2	0.226 (0.026)	-0.017 (0.012)	0.000 (0.012)	0.000 (0.021)	-0.004 (0.015)	-0.006 (0.014)
		0.8	0.049 (0.011)	-0.066 (0.014)	-0.001 (0.015)	0.000 (0.037)	-0.026 (0.028)	-0.028 (0.024)

All the estimators, except OLS, also improve as T gets larger. Thus, for a sufficiently large N and T, the differences in efficiency and bias of the different techniques becomes quite small. Even so, the results in Table 3 do highlight two techniques that consistently outperform the others--LSDVC and AH. The AH estimator produces the lowest average bias and the LSDVC produces the most efficient estimates. While choice of either technique could be justified on the basis of our results, given the large size of the standard deviation of the AH estimates for  $T \leq 10$ , a "bad draw" would result in an estimate very far from the true value, perhaps even of the wrong sign. Therefore, a reasonable strategy to follow would be to use the LSDVC method for small panels, but for panels with a larger time dimension, use the AH estimator. In our opinion, when T is large, the increased efficiency of AH, in addition to its computational simplicity for a long panel, tip the scales in its favor.

Table 4: Estimates for $\beta$ Bias Using Various Estimators								
$\sigma_\epsilon=1 \quad \rho=0.5 \quad \sigma_\xi=2$								
N	T	$\beta$	OLS (SE)	LSDV (SE)	LSDVc (SE)	AH (SE)	GMM12 (SE)	GMM15 (SE)
5	20	0.8	-0.086 (0.104)	0.008 (0.110)	0.005 (0.112)	0.004 (0.132)	-0.017 (0.166)	---
		0.2	-0.001 (0.133)	-0.024 (0.171)	-0.007 (0.183)	0.003 (0.381)	-0.087 (0.265)	---
	100	0.8	-0.098 (0.044)	0.006 (0.045)	0.001 (0.045)	0.001 (0.054)	-0.005 (0.075)	---
		0.2	-0.005 (0.055)	-0.027 (0.070)	-0.005 (0.076)	0.002 (0.094)	-0.034 (0.127)	---
10	20	0.8	-0.091 (0.071)	0.016 (0.060)	0.001 (0.061)	0.000 (0.078)	0.002 (0.080)	0.007 (0.071)
		0.2	-0.005 (0.085)	0.003 (0.101)	0.003 (0.102)	-0.003 (0.137)	-0.029 (0.140)	-0.023 (0.122)
	100	0.8	-0.099 (0.031)	0.015 (0.026)	0.001 (0.027)	0.000 (0.033)	0.000 (0.035)	0.001 (0.033)
		0.2	-0.007 (0.037)	0.002 (0.045)	0.002 (0.045)	0.001 (0.057)	-0.010 (0.060)	-0.011 (0.057)
20	20	0.8	-0.094 (0.051)	0.010 (0.038)	0.001 (0.038)	0.000 (0.051)	0.003 (0.049)	0.009 (0.043)
		0.2	-0.006 (0.057)	0.008 (0.063)	0.003 (0.062)	-0.001 (0.089)	-0.008 (0.082)	-0.002 (0.073)
	100	0.8	-0.100 (0.023)	0.009 (0.017)	0.000 (0.017)	-0.001 (0.022)	0.000 (0.025)	0.001 (0.022)
		0.2	-0.008 (0.026)	0.006 (0.028)	0.000 (0.028)	-0.002 (0.039)	-0.002 (0.042)	-0.002 (0.037)
30	20	0.8	-0.095 (0.043)	0.007 (0.030)	0.000 (0.030)	-0.001 (0.040)	-0.003 (0.038)	0.003 (0.033)
		0.2	-0.007 (0.045)	0.006 (0.048)	0.001 (0.047)	-0.002 (0.070)	-0.013 (0.064)	-0.005 (0.054)
	100	0.8	-0.100 (0.019)	0.006 (0.014)	0.000 (0.014)	-0.001 (0.018)	0.000 (0.019)	0.001 (0.016)
		0.2	-0.008 (0.020)	0.006 (0.022)	0.000 (0.022)	-0.001 (0.032)	-0.003 (0.032)	-0.001 (0.027)

## 5 An Application: Savings and Growth

Carroll and Weil (1994) use macroeconomic panel data to investigate the long-run relationship between savings rates and growth of per capita income. Using the same data and

specification, we apply the recommendations above to reestimate these relationships and demonstrate the importance of technique selection. We find that in this relatively short panel, usage of the corrected LSDV estimator can make a difference in obtaining significant coefficient estimates for all variables.

Carroll and Weil estimate two equations to test for Granger causality between savings and growth:

$$g_{i,t} = \eta_i^g + \gamma^g g_{i,t-1} + \beta^g s_{i,t-1} + \eta_{i,t} \quad (13)$$

$$s_{i,t} = \eta_i^s + \gamma^s s_{i,t-1} + \beta^s g_{i,t-1} + \epsilon_{i,t} \quad (14)$$

The error terms in the two equations are assumed to be independently distributed and each equation is estimated separately.<sup>7</sup> Thus, in the estimation of the growth equation, lagged savings is equivalent to  $x_t$  in equation (2), and in the estimation of the savings equation, lagged growth plays a similar role.

They construct annual measures of the savings and growth rates from the Penn World Tables, Mark 5. Using only countries with a data quality grade of C- or better and excluding communist countries, countries whose economies are dominated by oil production, and countries with populations less than one million, leaves a panel of 64 countries over 39 years (1950 to 1988). However, to avoid identifying business cycle effects, five-year averages are taken, reducing the time dimension of the panel to six.<sup>8</sup> Not all countries have data for all six time periods--there are a total of 353 observations.

Table 5 presents initial results from Carroll and Weil from two estimators--the LSDV and

---

<sup>7</sup>Carroll and Weil also use an alternative specification with time-specific effects that generates the same qualitative results.

<sup>8</sup>David Weil provided the data for us. See the Data Appendix of Carroll and Weil for a more thorough description of the data.

an alternative Anderson-Hsiao estimator that uses the twice lagged difference as an instrument for the lagged difference. As mentioned previously, we did not include this alternative Anderson-Hsiao estimator in our simulations because the lagged level has been shown to provide more efficient estimates. Table 5 presents their initial results.

Results from the LSDV estimation are potentially quite interesting. Lagged growth positively Granger causes savings, but lagged savings negatively Granger causes growth. However, when the instrumental variables technique is used to correct for biased coefficients, the relationship between lagged savings and growth becomes insignificant. Since the coefficients of particular interest in these estimations are not the coefficients on the lagged dependent variable, it is disappointing that in order to correct for the bias in  $\gamma$ , the alternative Anderson-Hsiao estimator renders  $\beta$  insignificant.

Dependent Variable	$g_t$		$s_t$	
Independent Variable	LSDV	Alternative AH	LSDV	Alternative AH
$g_{t-1}$	-0.059 (0.068)	-0.071 (0.128)	0.318* (0.080)	0.459* (0.184)
$s_{t-1}$	-.117* (0.039)	-0.058 (0.068)	0.482* (0.046)	-1.08* (0.528)

\*Significant at the 5% level.

The simulations discussed above suggest that the corrected LSDV estimator may provide more efficient estimates and thus may allow us to retain significance of the coefficients of interest. In applying this technique, however, two practical implementation issues arise. First, the corrected LSDV technique cannot be applied to an unbalanced panel. Thus, one must make a choice between selecting a technique that increases efficiency but may not be able to use all available data and one that uses all available data but is less efficient. Of course, the sensible choice should depend on the severity of the necessary reduction in sample. If balancing a panel requires substantial reduction in coverage, the corrected LSDV technique will not be the optimal choice. This efficiency tradeoff is likely to be an issue for macroeconomists using large panels of



countries like the Penn World Tables because it is more likely that low income countries would have more missing values and, thus, a balanced panel may be less representative. In this particular case, the reduction in sample is notable but moderate. When balancing the panel, we lose 41 observations spread across 12 countries.

The second issue related to the estimation of the corrected LSDV estimator is that there is not an analytical form for the standard errors. Thus, it is necessary to use a bootstrap technique to estimate standard errors in practice. Since bootstrapping is an iterative procedure, this adds substantially to computational time, reinforcing the recommendation above for the Anderson-Hsiao estimator when the time dimension of the panel is large.

Results for the balanced panel estimation are in Table 6.

Table 6: Balanced Panel Estimation						
Dependent Variable						
	LSDV	AH	LSDVC	LSDV	AH	LSDVC
Independent Variable	$g_t$	$g_t$	$g_t$	$s_t$	$s_t$	$s_t$
$g_{t-1}$	.019 (.081)	.324* (.165)	.286* (.051)	.219* (.096)	.283 (.258)	.202* (.083)
$s_{t-1}$	-.173* (.055)	-.266* (.091)	-.179* (.051)	.435* (.065)	-.229 (.733)	.568* (.027)

\*Significant at the 5% level.

As predicted by our simulation results, the corrected LSDV estimator has smaller standard errors, and thus, gives a higher number of significant coefficients. It is interesting to note that, in the savings regression, the Anderson-Hsiao estimator seems to apply a cure that is worse than the disease. The attempt to correct for the  $\gamma$  bias results in both coefficients becoming insignificant. The relatively small correction to the LSDV estimates indicates that the estimated bias in this case was not particularly severe. In addition, the difference in the magnitudes of the standard errors for these two methods is fairly remarkable.

Finally, in Table 7 we present results for the AH estimator using an unbalanced panel. As can be seen in this table, the additional data does increase the number of significant coefficients obtained, however, by this measure, the corrected LSDV estimator is still superior.

Table 7: Unbalanced Panel Estimates AH Estimator		
Dependent Variable		
Independent Variable	$g_t$	$s_t$
$g_{t-1}$	.245 (.130)	.268 (.137)
$s_{t-1}$	-.159 (.059)	-.014 (.487)

\*Significant at the 5% level.

## 6 Conclusion

The characteristics of the data, in particular the size of the panel, influence the choice of an optimal estimator for panel data models. For a panel with a long time dimension, we find that the computationally simple Anderson-Hsiao estimator performs well. However, when the time dimension of the panel is less than or equal to 10 observations, and the panel does not suffer too severely from missing data, due to the inefficiency of the Anderson-Hsiao estimates, use of a corrected LSDV estimator is preferred. In an empirical example, we find that the proper choice of estimator can be critical.

## REFERENCES

- Anderson, T.W. and Cheng Hsiao, "Estimation of dynamic models with error components," *Journal of the American Statistical Association*, 1981, 589-606.
- Anderson, T.W., and Cheng Hsiao, 1982, "Formulation and Estimation of Dynamic Models Using Panel Data," *Journal of Econometrics*, 18, 47-82.
- Arellano, Manuel, 1989, "A Note on the Anderson-Hsiao Estimator for Panel Data," *Economic Letters*, 31, 337-341.
- 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, 277-297.
- Barro, Robert, J. "Inflation and economic growth," *Bank of England Quarterly Bulletin*, May 1995.
- Carroll, Christopher D. and David N. Weil, "Saving and Growth: A Reinterpretation," *Carnegie-Rochester Conference Series on Public Policy* 40 (1994) 133-192.
- Efron, B. and Tibshirani R., "Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy," *Statistical Science*, 1986, Vol. 1, No. 1, 54-77.
- Fischer, Stanley, "The Role of Macroeconomic Factors in Growth," *Journal of Monetary Economics*, December 1993, 485-512.
- Holtz-Eakin, Douglas, Whitney Newey, and Harvey S. Rosen, 1988, "Estimating Vector Autoregressions with Panel Data," *Econometrica*, 56(6), 1371-1395.
- Hsiao, Cheng, 1986, *Analysis of Panel Data*, (New York: Cambridge University Press).
- Kiviet, Jan F., 1995, "On Bias, Inconsistency, and Efficiency of Various Estimators in Dynamic Panel Data Models," *Journal of Econometrics*, 68, 53-78.
- Levine, Ross, and David Renelt, "A Sensitivity Analysis of Cross-Country Growth Regressions," *American Economic Review*, September 1992, 942-63.
- Mankiw, N. Gregory, David Romer and David N. Weil, "A Contribution to the Empirics of Economic Growth," *Quarterly Journal of Economics*, May 1992, 407-38.
- Nickell, S., "Biases in Dynamic Models with Fixed Effects," *Econometrica*, 49, 1981, 1417-1426.

## Appendix: Estimation and Computation Issues

### 1. GMM Estimation

In calculating GMM1 and GMM2, it is helpful to note that, when  $T=5$ ,

$$\mathbf{Z}_i^* \equiv \begin{bmatrix} y_{i1} & x_{i1} & x_{i2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & y_{i1} & y_{i2} & x_{i1} & x_{i2} & x_{i3} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & y_{i1} & y_{i2} & y_{i3} & x_{i1} & x_{i2} & x_{i3} & x_{i4} \end{bmatrix} \quad (15)$$

and

$$\mathbf{Z} \equiv \begin{bmatrix} \mathbf{Z}_1^* \\ \mathbf{Z}_2^* \\ \cdot \\ \cdot \\ \mathbf{Z}_N^* \end{bmatrix} \quad (16)$$

### 2. LSDV Correction

Kiviet (1995) derives the bias of the LSDV estimator:

$$\begin{aligned} E(\hat{\delta} - \delta) = & -\sigma_\epsilon^2 (\bar{\mathbf{D}})^{-1} \left( \frac{N}{T-1} (i'_{T-1} C i_{T-1}) [2q - \bar{\mathbf{W}}' A \bar{\mathbf{W}} (\bar{\mathbf{D}})^{-1} q] \right. \\ & + \text{tr} \{ \bar{\mathbf{W}}' (I_N \otimes A_{T-1} C A_{T-1}) \bar{\mathbf{W}} (\bar{\mathbf{D}})^{-1} \} q + \bar{\mathbf{W}}' (I_N \otimes A_{T-1} C A_{T-1}) \bar{\mathbf{W}} (\bar{\mathbf{D}})^{-1} q \\ & \left. + \sigma_\epsilon^2 N q' (\bar{\mathbf{D}})^{-1} q \times \left[ -\frac{N}{T-1} (i'_{T-1} C i_{T-1}) \text{tr} \{ C' A_{T-1} C \} + 2 \text{tr} \{ C' A_{T-1} C A_{T-1} C \} \right] q \right) \\ & + O(N^{-1} T^{-3/2}) \end{aligned} \quad (17)$$

where

$$\begin{aligned}
\bar{D} &\equiv \bar{W}'A\bar{W} + \sigma_{\epsilon}^2 Ntr(C'A_{T-1}C)qq' \\
A_{T-1} &\equiv I_{T-1} - \frac{1}{T-1} i_{T-1} i_{T-1}' \\
A &\equiv I_N \otimes A_{T-1} \\
AW &\equiv E(AW)
\end{aligned} \tag{18}$$

$$C \equiv \begin{bmatrix} 0 & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ 1 & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ \gamma & 1 & 0 & \cdot & \cdot & \cdot & 0 \\ \gamma^2 & \gamma & 1 & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \gamma^{T-3} & \cdot & \cdot & \cdot & \gamma & 1 & 0 \end{bmatrix}, (T-1) \times (T-1) \tag{19}$$

$$q \equiv [1 \ 0 \ \cdot \ \cdot \ 0]', K \times 1 \tag{20}$$

$i_{T-1}$  is a  $(T-1) \times 1$  vector of ones, and  $W$  is an  $N(T-1) \times K$  matrix of independent variables constructed as follows. Define  $y_i^{(-1)}$  and  $X_i$  as the independent variables for the  $i$ th individual. Then,

$$y_i^{(-1)} = \begin{bmatrix} y_{i,1} \\ \cdot \\ \cdot \\ y_{i,T-1} \end{bmatrix} \quad \text{and} \quad X_i = \begin{bmatrix} x_{i2} \\ \cdot \\ \cdot \\ x_{iT} \end{bmatrix} \tag{21}$$

and  $y^{(-1)}$  and  $X$  are  $N(T-1) \times 1$  and  $N(T-1) \times (K-1)$  matrices such that

$$y^{(-1)} = \begin{bmatrix} y_1^{(-1)} \\ \cdot \\ \cdot \\ y_N^{(-1)} \end{bmatrix} \quad \text{and} \quad X = \begin{bmatrix} X_1 \\ \cdot \\ \cdot \\ X_N \end{bmatrix} \tag{22}$$

then  $W = [y^{(-1)} \ X]$ .  $\bar{W}$  is the expected value of  $W$ . As mentioned in the text, in order to calculate the LSDV correction, one first needs to obtain a preliminary estimate of the coefficients in order to estimate  $\sigma_\epsilon^2$  and  $\bar{W}$ . Any consistent estimator can be used; we used AH.

While the LSDV correction can be calculated directly from the formula shown above, when  $N$  or  $T$  is large, the matrices involved in the calculations can become quite large and implementation of the LSDV correction can require a good deal of computing resources. However, the computational burden can be reduced by noting two relationships. First, calculation and inversion of the large  $A$  matrix can be avoided by recognizing that

$$A \bar{W} = W^* - A_{T-1} C \epsilon \quad (23)$$

where  $\epsilon$  is a vector of errors from the preliminary regression and  $W^*$  is a group-demeaned  $W$ . Second, since  $A$  is idempotent,

$$\bar{W} A \bar{W} = A \bar{W} A \bar{W} \quad (24)$$