

IZA DP No. 3487

Fixed Effects Bias in Panel Data Estimators

Hielke Buddelmeyer
Paul H. Jensen
Umut Oguzoglu
Elizabeth Webster

May 2008

Fixed Effects Bias in Panel Data Estimators

Hielke Buddelmeyer
University of Melbourne and IZA

Paul H. Jensen
University of Melbourne

Umut Oguzoglu
University of Melbourne and IZA

Elizabeth Webster
University of Melbourne

Discussion Paper No. 3487
May 2008

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post World Net. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ABSTRACT

Fixed Effects Bias in Panel Data Estimators^{*}

Since little is known about the degree of bias in estimated fixed effects in panel data models, we run Monte Carlo simulations on a range of different estimators. We find that Anderson-Hsiao IV, Kiviet's bias-corrected LSDV and GMM estimators all perform well in both short and long panels. However, OLS outperforms the other estimators when the following holds: the cross-section is small ($N = 20$), the time dimension is short ($T = 5$) and the coefficient on the lagged dependent variable is large ($\gamma = 0.8$).

JEL Classification: C23, O11, E00

Keywords: panel data, LSDV, dynamic model, fixed effects

Corresponding author:

Umut Oguzoglu
Melbourne Institute of Applied Economic and Social Research
Alan Gilbert Building
The University of Melbourne
Parkville, Victoria 3010
Australia
E-mail: umuto@unimelb.edu.au

^{*} We thank Anne Owen for making her Gauss code available online.

1. Introduction

In the past, researchers have regarded estimated fixed effects in panel data models as “nuisance” or “ancillary” parameters. However, these parameters often convey useful information in industrial, labour, environmental and health economics (for example, see McClellan and Staiger 2000; Murdock 2007). Since these applications are often characterised by small panel data sets (either due to small N ; or small T ; or both) – which is precisely where the biases are most likely to be most acute – it is vital that researchers are informed about the properties of different estimators with respect to the estimated fixed effects. The aim of this paper is to provide researchers with a guide to the extent of fixed effects bias in panel data estimators across a range of different panel sizes.

There are at least two types of application where estimated fixed-effects are important. First, when the unit of analysis is, for example, the country, industry or occupation, researchers are interested in estimates of the relative magnitude of the fixed effects. In these cases, the cardinal ranking of countries, industries or occupations is crucial for the interpretation of the results and the policy implications drawn. For example, policy makers need to know *which* industries have above average knowledge spillovers if they want to most efficiently target industry assistance.

Secondly, fixed effects soak up much of the explanatory power of relatively time-invariant explanatory variables. This can occur when linking panel data with cross-sectional data and can render the relatively time-invariant explanatory variables statistically insignificant even when they are economically significant.¹ In these situations, a two-stage analysis using the fixed effects as the dependent variable in the second stage regression may be required to disentangle the relevant explanatory factors. Jensen and Webster (2008) provide an example of this approach by estimating company-level fixed effects for a large stock market panel data set of firms and for a much smaller subset of firms regressing these on high quality survey-based data. Alternatively, extracted fixed effects can be introduced as regressors in a second stage model (Bulkley et al (2004)).

Despite the varied uses of estimated fixed effects, little is known about the performance of commonly-used panel data estimators with respect to fixed effects. It has been argued that the least squares dummy variable (LSDV) estimator produces estimated fixed effects which are unbiased but inconsistent in short panels. However, there are few practical guides as to the definition of a ‘short panel’ (Cameron and Trivedi 2005, p.727). Even less is known *a priori* about the properties of fixed effects in dynamic panels. While several simulation studies have provided sensitivity tests for the bias

¹ Alternatively, the researcher may employ a random-effects estimator. However, this approach may be contra-indicated if the individual effects represent omitted variables which are likely to be correlated with the other regressors.

on slope coefficients, they are universally silent on the bias inherent in the estimated fixed effects (Judson and Owen 1999).

The paper proceeds as follows. In Section 2, we present the dynamic fixed effects model and outline our framework for the Monte Carlo simulations. In Section 3, we introduce the performance indicators used to evaluate six panel data estimators and present the results from the simulations. Section 4 concludes.

2. The model

Following Judson and Owen (1999), we consider the dynamic fixed effects model:

$$y_{i,t} = \gamma y_{i,t-1} + x'_{i,t} \beta + \eta_i + \varepsilon_{i,t}; \quad |\gamma| < 1; \quad \beta = 1 - \gamma; \quad \text{for } i = 1, \dots, N; t = 1, \dots, T \quad (1)$$

where η_i is a fixed effect, $x_{i,t}$ is a $(K-1) \times 1$ vector of strictly exogenous regressors, and $\varepsilon_{i,t} \sim N(0, \sigma_\varepsilon^2)$ is the serially uncorrelated random disturbance. $x_{i,t}$ is generated using a simple AR(1) model:

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

We depart from the existing simulation studies by focusing primarily on the fixed effect, η_i . We perform the Monte Carlo simulations using 2500 replications, where new $x_{i,t}$ and $\varepsilon_{i,t}$ vectors are generated for each draw. Fixed effects are treated the same way as γ and β . Thus, we hold the value of the fixed-effects constant across all draws.²

Several estimators are compared: a cross-sectional OLS, the LSDV, an instrumental variables estimator (A-HIV) proposed by Anderson and Hsiao (1981), a corrected LSDV estimator (LSDVC) derived in Kiviet (1995), and one-step and two-step GMM estimators (GMM13 and GMM23)³ as in Arellano and Bond (1991). We examine how the fixed-effect bias associated with each of these estimators varies across: three levels of γ ($=0$, $=0.4$ and $=0.8$); three values of the time dimension, T ($=5$, $=10$, $=20$); and three values of the cross sectional dimension, N ($=20$, $=500$, $=2000$).⁴

Unlike other Monte Carlo results which generally report the properties of a single estimated coefficient, we need to report indicators that summarise the properties of a vector of fixed effect

² Fixed effects are assumed to vary across individuals following a normal distribution with mean 0 and variance σ_η^2 .

³ We report the performance of the ‘restricted’ GMM estimators where $t-3$ is the earliest period that is used to extract instruments.

⁴ We fixed $\rho = 0.5$, and the ratio $\sigma_\eta / \sigma_\varepsilon = (1 - \gamma)$ and the signal to noise ratio at 2. See Judson & Owen (1999) for a detailed discussion on these parameters.

coefficients.⁵ Accordingly, the performance statistics we use to assess the magnitude of the bias are slightly more complicated than in the single coefficient case. We first proceed by computing the average of estimated individual fixed effects, across all simulation draws:

$$\hat{\eta}_i = \frac{1}{D} \sum_{d=1}^D (\hat{\eta}_i^d) \quad \text{for } i=1, \dots, N \quad (3)$$

where D is number of replications and $\hat{\eta}_i^d$ is the estimated fixed-effect for unit i from draw d . Then we calculate two performance statistics – the mean absolute bias (MAB) and the root mean squared error (RMSE). While the MAB measures the extent of the bias, the RMSE is used to reflect both bias and dispersion of the estimators⁶.

1. The mean absolute bias: $MAB = \frac{1}{N} \sum_{i=1}^N (|\eta_i - \hat{\eta}_i|)$ (4)

2. The root mean squared error: $RMSE = \frac{1}{N} \sqrt{\sum_{i=1}^N (\eta_i - \hat{\eta}_i)^2}$ (5)

3. Results

Table 1 summarises the results from the six estimators. We have highlighted the estimator which has the smallest MAB (which coincides with the smallest RMSE) in each different case. Note that we have included the special case $\gamma = 0$ so that we can compare the results across both static and dynamic panel models. Overall, our results clearly show that the LSDVC, A-HIV, GMM13 and GMM23 estimators are the best-performing estimators, with the smallest MAB across a wide spectrum of cases. Accordingly, some general rules of thumb can be drawn:

- i) as T and N increase, the bias in the LSDVC, A-HIV, GMM13 and GMM23 estimators decreases;
- ii) when N is small ($= 20$) and $\gamma \leq 0.4$, the LSDVC estimator outperforms all other estimators;
- iii) when N is large (≥ 500) and $\gamma = 0.8$, the A-HIV, and GMM estimators outperform the LSDVC;
- iv) OLS is less biased than LSDV when $\gamma = 0.8$; and
- v) the LSDV estimator is never the most efficient.

⁵ Fixed effects are extracted by plugging estimated slope coefficients into (1) and taking averages over T .

⁶ Our approach is closely related to the nonparametric econometrics literature where the accuracy of estimated fitted values is measured using mean square error (see Pagan & Ullah (1999) for an extensive review of this literature).

Given the fact that recent developments allow for straightforward extraction of the fixed-effects from the LSDVC estimations (using the STATA ado command `xtlsdvc`), this suggests that the LSDVC estimator would rank high as the preferred choice for applied researchers— unless γ is large.

4. Conclusions

Rather than regarding estimated fixed-effects in panels as “nuisance” parameters, we argue that these parameters convey useful information in a number of microeconomic applications. In order to identify biases in estimated fixed-effects in panel data models, we run Monte Carlo simulations on a range of different estimators. Our results are intended to provide researchers with a general guide to the best choice of panel estimator given the dimensions of the panel they are working with. The summary of results presented in Table 2 suggest that LSDVC, A-HIV and GMM estimators perform well across a range of different panel dimensions. The only exception to this rule of thumb occurs when T and N are small and $\gamma = 0.8$. In this case, OLS is the best-performing estimator.

Table 2
Summary of results

Parameters	$T = 5, N = 20$	$T = 10, N = 500$	$T = 20, N = 2000$
$\gamma = 0$	LSDVC	LSDVC, A-HIV, GMM	LSDVC, A-HIV, GMM
$\gamma = 0.4$	LSDVC	LSDVC, A-HIV, GMM	LSDVC, A-HIV, GMM
$\gamma = 0.8$	OLS	A-HIV	A-HIV, GMM

Table 1

Bias estimates for the fixed effects (η) using various estimators^a

T	N	Gamma	Statistic	OLS	LSDV	LSDVC	A-HIV	GMM13	GMM23
5	20	0	MAB	0.343	0.098	0.012	0.020	0.032	0.034
			RMSE	0.442	0.126	0.014	0.025	0.039	0.042
5	20	0.4	MAB	0.201	0.186	0.011	0.028	0.067	0.069
			RMSE	0.261	0.238	0.013	0.038	0.083	0.085
5	20	0.8	MAB	0.059	0.462	0.187	0.693	0.141	0.146
			RMSE	0.077	0.597	0.242	0.854	0.177	0.183
5	500	0	MAB	0.243	0.091	0.009	0.008	0.008	0.008
			RMSE	0.301	0.113	0.011	0.010	0.010	0.010
5	500	0.4	MAB	0.142	0.167	0.014	0.008	0.008	0.008
			RMSE	0.176	0.206	0.017	0.011	0.011	0.010
5	500	0.8	MAB	0.043	0.415	0.101	0.009	0.012	0.012
			RMSE	0.054	0.514	0.125	0.011	0.015	0.014
5	2000	0	MAB	0.229	0.087	0.008	0.008	0.008	0.008
			RMSE	0.287	0.110	0.010	0.010	0.010	0.010
5	2000	0.4	MAB	0.133	0.160	0.014	0.008	0.008	0.008
			RMSE	0.168	0.202	0.017	0.010	0.010	0.010
5	2000	0.8	MAB	0.041	0.399	0.097	0.008	0.008	0.008
			RMSE	0.052	0.504	0.122	0.010	0.010	0.010

10	20	0	MAB	0.334	0.042	0.005	0.006	0.011	0.011
			RMSE	0.430	0.055	0.006	0.008	0.013	0.013
10	20	0.4	MAB	0.195	0.074	0.005	0.006	0.024	0.024
			RMSE	0.252	0.096	0.007	0.008	0.029	0.029
10	20	0.8	MAB	0.058	0.215	0.047	0.013	0.064	0.064
			RMSE	0.074	0.279	0.061	0.016	0.078	0.078
10	500	0	MAB	0.244	0.038	0.005	0.005	0.005	0.005
			RMSE	0.301	0.047	0.006	0.006	0.006	0.007
10	500	0.4	MAB	0.142	0.065	0.005	0.005	0.005	0.005
			RMSE	0.176	0.080	0.006	0.006	0.007	0.007
10	500	0.8	MAB	0.043	0.190	0.026	0.005	0.006	0.006
			RMSE	0.053	0.235	0.032	0.007	0.008	0.008
10	2000	0	MAB	0.228	0.036	0.006	0.006	0.006	0.006
			RMSE	0.287	0.046	0.007	0.007	0.007	0.007
10	2000	0.4	MAB	0.133	0.063	0.006	0.006	0.006	0.006
			RMSE	0.167	0.079	0.007	0.007	0.007	0.007
10	2000	0.8	MAB	0.041	0.183	0.026	0.006	0.006	0.006
			RMSE	0.051	0.231	0.032	0.007	0.007	0.007

20	20	0	MAB	0.329	0.020	0.003	0.004	0.005	0.005
			RMSE	0.424	0.025	0.004	0.005	0.006	0.006
20	20	0.4	MAB	0.192	0.034	0.003	0.004	0.013	0.013
			RMSE	0.248	0.043	0.004	0.005	0.017	0.017
20	20	0.8	MAB	0.057	0.099	0.007	0.006	0.026	0.026
			RMSE	0.074	0.127	0.009	0.007	0.033	0.033
20	500	0	MAB	0.243	0.018	0.004	0.004	0.004	0.004
			RMSE	0.301	0.022	0.005	0.005	0.005	0.005
20	500	0.4	MAB	0.142	0.029	0.004	0.004	0.004	0.004
			RMSE	0.176	0.036	0.005	0.005	0.005	0.005
20	500	0.8	MAB	0.043	0.086	0.006	0.004	0.004	0.004
			RMSE	0.053	0.106	0.007	0.005	0.005	0.005
20	2000	0	MAB	0.228	0.017	0.004	0.004	0.004	0.004
			RMSE	0.287	0.022	0.005	0.005	0.005	0.005
20	2000	0.4	MAB	0.133	0.028	0.004	0.004	0.004	0.004
			RMSE	0.167	0.035	0.005	0.005	0.005	0.005
20	2000	0.8	MAB	0.040	0.082	0.006	0.004	0.004	0.004
			RMSE	0.051	0.104	0.007	0.005	0.005	0.005

^a Estimates based on 2500 draws; $\rho=0.5$; and $\sigma_{\eta} / \sigma_{\varepsilon} = 1 - \gamma$

References

- Arellano, M. and Bond, S. 1991. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economics Studies* 58, 277-297.
- Bulkley, G., Harrif R.D.F. and Herrerias, R. (2004) Why does book-to-market value of equity forecast cross-section stock returns? *International Review of Financial Analysis* 13,153-160.
- Cameron, A. and Trivedi, P. 2005. *Microeconometrics. Methods and Applications*. Cambridge University Press, Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, Sao Paulo.
- Jensen, P.H. and Webster, E. 2008. What creates abnormal profits: Collusion, efficiency or strategy? Unpublished mimeo, University of Melbourne.
- Judson, R. and Owen, A. 1999. Estimating dynamic panel data models: a guide for macroeconomists. *Economics 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.
- McClellan, M. and Staiger, D. Comparing Hospital Quality at For-Profit and Not-for-Profit Hospitals in *The Changing Hospital Industry: Comparing Not-for-Profit and For-Profit Institutions*, (eds.) Cutler, David M. pp.93-112, The University of Chicago Press, 2000.
- Murdock, J. 2006. Handling unobserved site characteristics in random utility models of recreation demand. *Journal of Environmental Economics and Management*,51, 1-25.
- Pagan, A.R. and Ullah, A. *Nonparamateric Econometrics*. Cambridge University Press, 1999.