

Mastering Panel Metrics: Causal Impact of Democracy on Growth

Shuowen Chen
Victor Chernozhukov
Iván Fernández-Val

The Institute for Fiscal Studies
Department of Economics,
UCL

cemmap working paper CWP33/19

Mastering Panel Metrics: Causal Impact of Democracy on Growth

By SHUOWEN CHEN AND VICTOR CHERNOZHUKOV AND IVÁN FERNÁNDEZ-VAL*

The relationship between democracy and economic growth is of long standing interest. We revisit the panel data analysis of this relationship by Acemoglu et al. (forthcoming) using state of the art econometric methods. We argue that this and lots of other panel data settings in economics are in fact high-dimensional, resulting in principal estimators – the fixed effects (FE) and Arellano-Bond (AB) estimators – to be *biased* to the degree that *invalidates* statistical inference. We can however remove these biases by using simple analytical and sample-splitting methods, and thereby restore valid statistical inference. We find that the debiased FE and AB estimators produce substantially higher estimates of the long-run effect of democracy on growth, providing even stronger support for the key hypothesis in Acemoglu et al. (forthcoming). Given the ubiquitous nature of panel data, we conclude that the use of debiased panel data estimators should substantially improve the quality of empirical inference in economics.

I. Mastering Panel Metrics

A. The Setting

We consider the dynamic linear panel data model

$$(1) \quad Y_{it} = a_i + b_t + D'_{it}\alpha + W'_{it}\beta + \epsilon_{it},$$

* Chen: Department of Economics, Boston University, swchen@bu.edu. Chernozhukov: Department of Economics, MIT, vchern@mit.edu. Fernández-Val: Department of Economics, Boston University, ivanf@bu.edu. We gratefully acknowledge research support from the British Academy's visiting fellowships, National Science Foundation, and Spanish State Research Agency MDM-2016-0684 under the María de Maeztu Unit of Excellence Program. Fernández-Val was visiting UCL and Cemfi while working on this paper, he is grateful for their hospitality.

where $i = 1, \dots, N$ and $t = 1, \dots, T$. Here Y_{it} is the outcome for an observational unit i at time t , D_{it} is a vector of variables of interest or treatments, whose predictive effect α we would like to estimate, W_{it} is a vector of covariates or controls including a constant and lags of Y_{it} , a_i and b_t are unobserved unit and time effects that can be correlated to D_{it} , and ϵ_{it} is an error term normalized to have zero mean for each unit and time that satisfies the weak exogeneity condition

$$(2) \quad \epsilon_{it} \perp I_{it}, \quad I_{it} := \{(D_{is}, W_{is}, b_s)_{s=1}^t, a_i\}.$$

We assume that the vectors $Z_i := \{(Y_{it}, D'_{it}, W'_{it})'_{t=1}^T\}$, which collect these variables for the observational unit i , are i.i.d. across i , and make other conventional regularity assumptions. The main challenge in the estimation of panel data models is how to deal with the unobserved effects. We review two approaches.

B. The Fixed Effects Approach

This approach treats the unit and time effects as parameters to be estimated by applying OLS in the model:

$$Y_{it} = D'_{it}\alpha + X'_{it}\gamma + \epsilon_{it},$$

where $X_{it} := (W'_{it}, Q'_i, Q'_t)'$, Q_i is an N -dimensional vector of indicators for observational units with a 1 in the i -th position and 0's otherwise, and Q_t is a T -dimensional vector of indicators for time periods with a 1 in the t -th position and 0's otherwise. The elements of γ appearing in front of Q_i and Q_t are called unit fixed effects and time fixed effects, respectively. The resulting estimator is the fixed effect (FE) estimator. For our purposes, it can be seen as an exactly identified GMM

estimator with the score function

$$g(Z_i, \alpha, \gamma) = \{(Y_{it} - D'_{it}\alpha - X'_{it}\gamma)M_{it}\}_{t=1}^T,$$

where $M_{it} := (D'_{it}, X'_{it})'$.

The FE estimator is biased with bias of order $N/(NT) = 1/T$, due to estimation of many (N) nuisance parameters with NT observations, and the bias decreases as T becomes large. The estimator approaches the true value α as both NT and T become large, but unfortunately the bias of the estimator is too big relative to the order $1/\sqrt{NT}$ of the stochastic error, resulting in invalid assessment of statistic significance of the estimates. This necessitates the use of bias correction to restore the validity of the statistical inference.

C. The AB Approach

This approach eliminates the unit effects a_i by taking differences across time and uses moment conditions for the variables in differences. Specifically, define the differencing operator Δ acting on doubly indexed random variables V_{it} by creating the difference $\Delta V_{it} = V_{it} - V_{it-1}$. Apply this operator to both sides of (1) to obtain:

$$(3) \quad \Delta Y_{it} = \Delta D'_{it}\alpha + \Delta X'_{it}\gamma + \Delta \epsilon_{it},$$

where $X_{it} = (W'_{it}, Q'_t)'$. Note that by (2),

$$\Delta \epsilon_{it} \perp (D_{is}, W_{is})_{s=1}^{t-1}, \quad t = 2, \dots, T.$$

This means that estimation and inference can be done using an overidentified GMM with score function

$$g(Z_i, \alpha, \gamma) = \{(\Delta Y_{it} - \Delta D'_{it}\alpha - \Delta X'_{it}\gamma)M_{it}\}_{t=2}^T,$$

where $M_{it} = [(D'_{is}, W'_{is})_{s=1}^{t-1}, Q'_t]$. This is the Arellano and Bond (1991) estimator.

The AB estimator enjoys good properties when T is very small, but when T is even modestly large, it uses many ($m = O(T^2)$) moment conditions, which results in a bias of order $m/NT = O(T/N)$, which can be too large relative to the size of the stochastic error $1/\sqrt{NT}$ of the estimator. In the latter case statistical inference becomes invalid, and we need to employ bias correction

methods to restore its validity.

D. GMM under High Dimensionality and Need to Bias Corrections

Both FE and AB are GMM estimators in a high-dimensional regime – with either the number of nuisance parameters or the number of moment equations being large.

In the FE approach, the dimension of α is low, but the dimension p of the nuisance parameter γ is high. We can approximate this situation as $p = \dim(\gamma) \rightarrow \infty$ when $n \rightarrow \infty$, while $d_\alpha := \dim(\alpha)$ is held fixed. In the AB approach, the number of moment conditions, $m = \dim(g(Z_i, \alpha, \gamma))$, could be high, so we can approximate this situation as $m \rightarrow \infty$ when $n \rightarrow \infty$.

In either regime, there exist regularity conditions such that if $(p \vee m)^2$ is small compared to n :¹

$$(4) \quad (p \vee m)^2/n \rightarrow 0 \text{ as } n \rightarrow \infty,$$

then the standard approximate normality and consistency results of the GMM estimator continue to hold, namely

$$(5) \quad \sqrt{n}(\hat{\alpha} - \alpha) \overset{a}{\rightsquigarrow} N(0, V_{11}),$$

where V_{11} is the $d_\alpha \times d_\alpha$ upper-left block of the asymptotic variance of the GMM estimator corresponding to $\hat{\alpha}$.²

The key rate condition (4) can be interpreted as the *small bias condition*. This condition fails to hold in the FE approach where $p^2 = O(N^2 + T^2)$ and $n = NT$, and in the AB approach when T is large because $m^2 = O(T^4)$ and $n = NT$. Both of these failures apply to our empirical setting.

To understand where (4) comes from, let us focus on the exactly identified case where $p = m$. An asymptotic second order expansion of $\hat{\alpha}$ around α gives

$$\hat{\alpha} - \alpha = Z_n/\sqrt{n} + b/n + r_n,$$

¹For $a, b \in \mathbb{R}$, $a \vee b := \max(a, b)$.

²Sufficient conditions are given, for example, by Newey and Windmeijer (2009) for GMM problems with $m \rightarrow \infty$ and p fixed; and by Hahn and Newey (2004), Hahn and Kuersteiner (2011) and Fernández-Val and Weidner (2018) for nonlinear panel data models where $m \propto p \rightarrow \infty$.

where $Z_n \stackrel{a}{\sim} N(0, V_{11})$, $b = O(p)$ is a first order bias term coming from the quadratic term of the expansion, and r_n is a higher order remainder such as $r_n = O_p((p/n)^{3/2} + p^{1/2}/n)$. Then, (5) holds if both

$$\sqrt{nb}/n \rightarrow 0, \quad \text{i.e. } p^2/n \rightarrow 0,$$

and $\sqrt{nr}_n \rightarrow_P 0$, i.e. $p^{3/2}/n \rightarrow 0$.

The sketch above illustrates that the bias is the bottleneck. If we remove the bias somehow, then we can improve the rate requirement (4) to a weaker condition listed below.

There are several ways of removing the bias:

- a) *Analytical bias correction*, where we estimate b/n using analytical expressions for the bias and set

$$\tilde{\alpha} = \hat{\alpha} - \hat{b}/n.$$

- b) *Split-sample bias correction*, where we split the sample into two parts, compute the estimator on the two parts $\hat{\alpha}_{(1)}$ and $\hat{\alpha}_{(2)}$ to obtain $\bar{\alpha} = (\hat{\alpha}_{(1)} + \hat{\alpha}_{(2)})/2$, and then set

$$\tilde{\alpha} = \hat{\alpha} - (\bar{\alpha} - \hat{\alpha}) = 2\hat{\alpha} - \bar{\alpha}.$$

In some cases we can average over many splits to reduce variability.³

Why does the sample-splitting method work? Assuming that we estimate the same number of nuisance parameters and use the same number of moment conditions in all the parts of the sample, and that these parts are homogenous, then the first order biases of $\hat{\alpha}$, $\hat{\alpha}_{(1)}$, and $\hat{\alpha}_{(2)}$ are

$$\frac{b}{n}, \quad \frac{b}{n/2}, \quad \frac{b}{n/2},$$

so that the first order bias of $\tilde{\alpha}$ is

$$2\frac{b}{n} - \left(\frac{1}{2} \left[\frac{b}{n/2} \right] + \frac{1}{2} \left[\frac{b}{n/2} \right] \right) = 0.$$

After debiasing, the resulting rate condi-

³ In some cases it is also possible to use the bootstrap and leave-one-out methods for bias correction.

tions are weaker. In particular, there exist regularity conditions such that if the dimensionality is not overly high:

$$(p \vee m)^{3/2}/n \rightarrow 0 \text{ as } n \rightarrow \infty,$$

then the approximate normality and consistency results for the bias-corrected GMM estimator continue to hold:⁴

$$\sqrt{n}(\tilde{\alpha} - \alpha) \stackrel{a}{\sim} N(0, V_{11}).$$

E. The Debiased FE and AB Estimators

To construct analytical debiased FE estimator (DFE-A), we need to characterize the first order bias. An analysis similar to Nickell (1981) yields that first order bias b/n obeys:

$$Hb = -\frac{1}{T} \sum_{i=1}^N \sum_{t=1}^{T-1} \sum_{s=t+1}^T E[D_{is}\epsilon_{it}],$$

for

$$H = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T E[\tilde{D}_{it}\tilde{D}'_{it}],$$

where \tilde{D}_{it} is the residual of the sample linear projection of D_{it} on X_{it} . Note that $b = O(N)$ because the source of the bias is the estimation of the N unit fixed effects and the order of the bias is $b/n = O(T^{-1})$ because there are only T observations that are informative about each unit fixed effect.⁵ An estimator of the bias can be formed as

$$\hat{H}\hat{b} = -\sum_{i=1}^N \sum_{t=1}^{T-1} \sum_{s=t+1}^{(t+M)\wedge T} \frac{D_{is}\hat{\epsilon}_{it}}{T-s+t},$$

⁴Sufficient conditions are given in Kiviet (1995), Hahn and Kuersteiner (2002) and Chudik, Pesaran and Yang (2018) for dynamic linear panel data models and Fernández-Val and Weidner (2016) and Fernández-Val and Weidner (2018) for nonlinear panel data models.

⁵There is no bias coming from the estimation of the time fixed effects because the model is linear and we assume independence across i .

where $\hat{\epsilon}_{it}$ is the fixed effect residual,

$$\hat{H} = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \tilde{D}_{it} \tilde{D}'_{it},$$

and M is a trimming parameter such that $M/T \rightarrow 0$ and $M \rightarrow \infty$ as $T \rightarrow \infty$ (Hahn and Kuersteiner, 2011).

To implement debiasing by sample splitting, we need to determine the partition of the data. For the debiased FE estimator via sample splitting (DFE-SS), we split the panel along the time series dimension because the source of the bias is the estimation of the unit fixed effects. Thus, following Dhaene and Jochmans (2015), the parts contain the observations $\{i = 1, \dots, N; t = 1, \dots, \lceil T/2 \rceil\}$ and $\{i = 1, \dots, N; t = \lfloor T/2 \rfloor, \dots, T\}$, where $\lceil \cdot \rceil$ and $\lfloor \cdot \rfloor$ are the ceiling and floor functions. This partition preserves the time series structure and delivers two panels with the same number of unit fixed effects, where there are $T/2$ observations that are informative about each unit fixed effect. For the debiased AB estimator via sample splitting (DAB-SS), we split the panel along the cross section dimension because the source of the bias is the number of moment conditions relative to the sample size. Thus, the parts contain the observations $\{i = 1, \dots, \lceil N/2 \rceil; t = 1, \dots, T\}$ and $\{i = \lfloor N/2 \rfloor, \dots, N; t = 1, \dots, T\}$. This partition delivers two panels where the number of observations relative to the number of moment conditions is half of the original panel. Note that there are multiple possible partitions because the ordering of the observations along the cross section dimension is arbitrary. We can therefore average across multiple splits to reduce variability.

II. Democracy and Growth

We revisit the application to the causal effect of democracy on economic growth of Acemoglu et al. (forthcoming) using the econometric methods described in Section I. To keep the analysis simple, we use a balanced sub-panel of 147 countries over the period from 1987 through 2009 extracted from the data set used in Acemoglu et al.

(forthcoming). The outcome variable Y_{it} is the logarithm of GDP per capita in 2000 USD as measured by the World Bank for country i at year t . The treatment variable of interest D_{it} is a democracy indicator constructed in Acemoglu et al. (forthcoming), which combines information from several sources including Freedom House and Polity IV. It characterizes whether countries have free and competitive elections, checks on executive power, and an inclusive political process. We report some descriptive statistics of the variables used in the analysis in the online supplemental Appendix.

We control for unobserved country effects, time effects and rich dynamics of GDP using the linear panel model (1), where W_{it} includes four lags of Y_{it} . The weak exogeneity condition (2) implies that democracy and past GDP are orthogonal to contemporaneous and future GDP shocks ϵ_{it} and that these shocks are serially uncorrelated (since W_{it} includes the lagged values of Y_{it}).

In addition to the instantaneous or short-run effect of a transition to democracy to economic growth measured by the coefficient α , we are interested in a permanent or long-run dynamic effect. This effect in the dynamic linear panel model (1) is

$$(6) \quad \alpha / (1 - \sum_{j=1}^4 \beta_j),$$

where β_1, \dots, β_4 are the coefficients corresponding to the lags of Y_{it} .

We consider the FE and a one-step AB estimators as well as their debiased versions (DFE and DAB). Indeed, the raw AB and FE fail to satisfy the small bias condition: the AB approach relies on $m = 632$ moment conditions to estimate $p = 169$ parameters with $n = 147 \times 18 = 2,646$ observations, after using the first five periods as initial conditions, so that $(m \vee p)^2/n \approx 150$, which is not close to zero; the FE approach estimates $p = 170$ parameters with $n = 147 \times 19 = 2,793$ observations, after using the first four periods as initial conditions, yielding $(m \vee p)^2/n \approx 10$, which is

not close to zero.

To debias the estimators, we consider both analytical and split sample bias corrections. For the fixed effect approach, DFE-A implements the analytical debiasing with $M = 4$, whereas DFE-SS implements debiasing by sample splitting. We consider two versions of the debiasing via sample-splitting for AB, where DAB-SS1 uses one random split and DAB-SS5 uses the average of five random splits.

For each estimator, we report analytical standard errors clustered at the country level and bootstrap standard errors based on resampling countries with replacement. The estimates of the long-run effect are obtained by plugging-in estimates of the coefficients in the expression (6). We use the delta method to construct analytical standard errors clustered at the country level, and resample countries with replacement to construct bootstrap standard errors. There is no need to recompute the analytical standard errors for debiased estimators, because the ones obtained for the uncorrected estimators remain valid for the bias corrections. We also report bootstrap standard errors for the debiased estimators.

Table 1 presents the empirical results.⁶ FE finds that a transition to democracy increases economic growth by almost 1.9% in the first year and 16% in the long run, while AB finds larger impacts of 4% and 21% but less precisely estimated. We find that the debiasing changes the estimates by a significant amount in both statistical sense (relative to the standard error) and economic sense (relative to the uncorrected estimates). The debiased estimators, DFE and DAB, find that a transition to democracy increases economic growth by about 2.3-5.2% in the first year, and about 25-26% in the long run. Interestingly, the two debiased approaches produce very similar estimates. Moreover, the results coincide with the results obtained using the method of Hahn, Hausman and Kuersteiner (2005), as reported in Acemoglu et al.

⁶We obtained the estimates with the commands `p1m` and `pgmm` of the package `p1m` in R.

(forthcoming). We believe that the estimates reported here as well as the later estimates reported in Acemoglu et al. (forthcoming) represent an adequate, state of the art analysis. Of course, it would be interesting to continue to explore other modern, perhaps even more refined, econometric approaches to thoroughly examine the empirical question.

We conclude with comments on the standard errors. The analytical standard errors are smaller than the bootstrap standard errors for the split-sample bias corrections. These differences might indicate that the analytical standard errors miss the additional sampling error introduced by the estimation in smaller panels. The analytical correction produces more precise estimates than the split-sample correction.

REFERENCES

- Acemoglu, Daron, Suresh Naidu, Pascual Restrepo, and James A. Robinson.** forthcoming. "Democracy Does Cause Growth." *Journal of Political Economy*.
- Arellano, Manuel, and Stephen Bond.** 1991. "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations." *The Review of Economic Studies*, 58(2): 277–297.
- Chudik, Alexander, M. Hashem Pesaran, and Jui-Chung Yang.** 2018. "Half-panel jackknife fixed-effects estimation of linear panels with weakly exogenous regressors." *Journal of Applied Econometrics*, 33(6): 816–836.
- Dhaene, Geert, and Koen Jochmans.** 2015. "Split-panel Jackknife Estimation of Fixed-effect Models." *The Review of Economic Studies*, 82(3): 991–1030.
- Fernández-Val, Iván, and Martin Weidner.** 2016. "Individual and time effects in nonlinear panel models with large N, T." *Journal of Econometrics*, 192(1): 291–312.

TABLE 1—EFFECT OF DEMOCRACY ON ECONOMIC GROWTH

	Initial and Debiased FE			Initial and Debiased AB		
	FE	DFE-A	DFE-SS	AB	DAB-SS1	DAB-SS5
Short Run Effect of Democracy ($\times 100$)	1.89 (0.65) [0.64]	2.27 [0.64]	2.44 [0.96]	3.94 (1.50) [1.52]	5.22 [1.83]	4.53 [1.91]
1st lag of log GDP	1.15 (0.05) [0.05]	1.23 [0.05]	1.30 [0.08]	1.00 (0.06) [0.06]	0.98 [0.07]	1.03 [0.08]
2nd lag of log GDP	-0.12 (0.06) [0.05]	-0.14 [0.05]	-0.13 [0.08]	-0.06 (0.06) [0.06]	-0.05 [0.07]	-0.07 [0.07]
3rd lag of log GDP	-0.07 (0.04) [0.04]	-0.09 [0.04]	-0.13 [0.06]	-0.04 (0.04) [0.04]	-0.04 [0.04]	-0.06 [0.04]
4th lag of log GDP	-0.08 (0.02) [0.02]	-0.08 [0.03]	-0.08 [0.04]	-0.08 (0.03) [0.03]	-0.08 [0.03]	-0.08 [0.03]
Long-run effect of democracy ($\times 100$)	16.05 (6.67) [6.63]	25.91 [9.31]	25.69 [12.12]	20.97 (9.51) [9.38]	26.46 [10.72]	25.24 [11.29]

Note: All the specifications include country and year effects. Analytical clustered standard errors at the country level are shown in parentheses. Bootstrap standard errors based on 500 replications are shown in brackets.

Fernández-Val, Iván, and Martin Weidner. 2018. “Fixed Effects Estimation of Large-TPanel Data Models.” *Annual Review of Economics*, 10(1): 109–138.

Hahn, Jinyong, and Guido Kuersteiner. 2011. “Bias reduction for dynamic nonlinear panel models with fixed effects.” *Econometric Theory*, 27(06): 1152–1191.

Hahn, Jinyong, and Guido M. Kuersteiner. 2002. “Asymptotically Unbiased Inference for a Dynamic Panel Model with Fixed Effects When Both N and T Are Large.” *Econometrica*, 70(4): 1639–1657.

Hahn, Jinyong, and Whitney Newey. 2004. “Jackknife and Analytical Bias Reduction for Nonlinear Panel Models.” *Econometrica*, 72(4): 1295–1319.

Hahn, Jinyong, Jerry Hausman, and Guido Kuersteiner. 2005. “Bias Cor-

rected Instrumental Variables Estimation for Dynamic Panel Models with Fixed Effects.” Boston University - Department of Economics Boston University - Department of Economics - Working Papers Series WP2005-024.

Kiviet, Jan F. 1995. “On bias, inconsistency, and efficiency of various estimators in dynamic panel data models.” *J. Econometrics*, 68(1): 53–78.

Newey, Whitney K., and Frank Windmeijer. 2009. “Generalized Method of Moments With Many Weak Moment Conditions.” *Econometrica*, 77(3): 687–719.

Nickell, Stephen. 1981. “Biases in Dynamic Models with Fixed Effects.” *Econometrica*, 49(6): 1417–1426.

Appendix

The online supplemental Appendix contains the data, descriptive statistics, and

code in **R** and **Stata** for the empirical application.

SUPPLEMENTAL APPENDIX

TABLE A1—DESCRIPTIVE STATISTICS

	Mean	SD	Dem = 1	Dem = 0
Democracy	0.62	0.49	1.00	0.00
Log(GDP)	7.58	1.61	8.09	6.75
Number Obs.	3,381	3,381	2,099	1,282