Bias-corrected estimation of linear dynamic panel data models

Sebastian Kripfganz[1] Jörg Breitung[2]

[1]University of Exeter Business School, Department of Economics, Exeter, UK

[2]University of Cologne, Faculty of Management, Economics and Social Sciences, Institute of Econometrics and Statistics, Cologne, Germany

28th UK Stata Conference

London, September 8, 2022

```
ssc install xtdpdbc
net install xtdpdbc, from(http://www.kripfganz.de/stata/)
```

Estimation of linear dynamic panel data models

- In the presence of unobserved group-specific heterogeneity, the traditional "fixed effects" (FE) / "random effects" (RE) estimators for linear panel data models with a lagged dependent variable are biased when the time horizon is short (Nickell, 1981).
- Existing solutions include generalized method of moments (GMM) estimators (e.g. `xtdpdgmm`; Kripfganz, 2019) and quasi-maximum likelihood (QML) estimators (e.g. `xtdpdqml`; Kripfganz, 2016).
  - GMM estimators can flexibly accommodate predetermined or endogenous regressors by utilizing appropriate instrumental variables.
  - QML estimators can be much more efficient when all regressors are strictly exogenous, but they require additional assumptions on the initial observations.

Bias-corrected estimation of dynamic panel models

- We present a new easy-to-use bias-corrected (BC) estimator with attractive finite-sample properties (Breitung, Kripfganz, and Hayakawa, 2021): `xtdpdbc`
  - Since the analytical form of the bias is known, the BC estimator can correct it directly at the source by adjusting the respective moment conditions.
  - The small variance of FE/RE estimators is retained.
  - Higher-order autoregressive models can be accommodated.
  - Both FE and RE versions are available.
  - Our BC estimator is a method of moments estimator with known asymptotic distribution. Consequently, standard errors can be readily computed.
  - Standard errors can be adjusted to be robust to cross-sectional dependence.

Bias-corrected estimation of dynamic panel models

- The FE version of our BC estimator is equivalent to the Dhaene and Jochmans (2016) adjusted profile likelihood estimator. When there is only a single lag of the dependent variable, it is also equivalent to the iterative BC estimator of Bun and Carree (2005).
- In addition to the mentioned advantages, our BC estimator does not require a preliminary consistent estimator, as is needed for the bias approximation of Kiviet (1995) (`xtlsdvc`; Bruno, 2005).

Linear dynamic panel data model

$$y_{it} = \sum_{j=1}^{p} \lambda_j y_{i,t-j} + \mathbf{x}'_{it}\beta + \underbrace{\alpha_i + u_{it}}_{=e_{it}}, \quad i = 1, 2, \ldots, N, \quad t = 1, 2, \ldots, T$$

- Basic model assumptions:
  - (Higher-order) autoregressive model with p lags of the dependent variable and only minimal regularity conditions on the initial observations
  - Strictly exogenous regressors $\mathbf{x}_{it}$ with respect to the idiosyncratic error term: $E[\mathbf{x}_{it} u_{is}] = \mathbf{0}$ for all t and s
  - Unobserved group-specific "fixed effects", $E[\mathbf{x}_{it}\alpha_i] \neq \mathbf{0}$, or "random effects", $E[\mathbf{x}_{it}\alpha_i] = \mathbf{0}$
  - Serially uncorrelated idiosyncratic errors, $E[u_{it} u_{is}] = 0$ for all t and s, but possibly heteroskedastic, $E[u_{it}^2] = \sigma_i^2$

Bias-corrected method of moments estimator

- For simplicity, let $p = 1$ and define $\boldsymbol{\theta} = (\lambda_1, \boldsymbol{\beta}')'$.
- The just-identified FE-BC estimator solves

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \left(\sum_{i=1}^{N} \mathbf{m}_i(\boldsymbol{\theta}) \right)' \left(\sum_{i=1}^{N} \mathbf{m}_i(\boldsymbol{\theta}) \right)$$

with moment functions

$$\mathbf{m}_i(\boldsymbol{\theta}) = \frac{1}{T} \sum_{t=1}^{T} \left[\begin{pmatrix} y_{i,t-1} - \bar{y}_{-1,i} \\ \mathbf{x}_{it} - \bar{x}_i \end{pmatrix} - \begin{pmatrix} \frac{T}{T-1} b(\lambda_1)(e_{it} - \bar{e}_i) \\ \mathbf{0} \end{pmatrix} \right] e_{it}$$

where $\bar{x}_i = \frac{1}{T} \sum_{t=1}^{T} \mathbf{x}_{it}$ and $b(\lambda_1) = -\frac{1}{T^2} \sum_{t=0}^{T-2} \sum_{s=0}^{t} \lambda_1^s$,
such that $E[\mathbf{m}_i(\boldsymbol{\theta})] = \mathbf{0}$.

- See Breitung, Kripfganz, and Hayakawa (2021) for details.

Bias-corrected method of moments estimator

- The overidentified RE-BC estimator solves

$$\hat{\boldsymbol{\theta}}^{(j)} = \arg\min_{\boldsymbol{\theta}} \left(\sum_{i=1}^{N} \mathbf{m}_i(\boldsymbol{\theta}) \right)' \mathbf{W} \left(\sum_{i=1}^{N} \mathbf{m}_i(\boldsymbol{\theta}) \right)$$

with moment functions

$$\mathbf{m}_i(\boldsymbol{\theta}) = \frac{1}{T} \sum_{t=1}^{T} \left[\underbrace{\begin{pmatrix} y_{i,t-1} - \bar{y}_{-1,i} \\ \mathbf{x}_{it} - \bar{x}_i \\ \mathbf{x}_{it} \end{pmatrix}}_{=\mathbf{z}_{it}} - \begin{pmatrix} \frac{T}{T-1} b(\lambda_1)(e_{it} - \bar{e}_i) \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix} \right] e_{it}$$

and weighting matrix $\mathbf{W} = \left(\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbf{z}_{it}\mathbf{z}_{it}' \right)^{-1}$ for the 1-step estimator $\hat{\boldsymbol{\theta}}^{(1)}$, or $\mathbf{W} = \left(\frac{1}{N} \sum_{i=1}^{N} \mathbf{m}_i(\hat{\boldsymbol{\theta}}^{(1)})\mathbf{m}_i(\hat{\boldsymbol{\theta}}^{(1)})' \right)^{-1}$ for the 2-step estimator $\hat{\boldsymbol{\theta}}^{(2)}$.

Bias-corrected fixed-effects estimator

```
. webuse abdata
. xtdpdbc n w k, fe vce(robust)

Bias-corrected estimation
Iteration 0:   f(b) =  .00268095
Iteration 1:   f(b) =  9.158e-06
Iteration 2:   f(b) =  1.024e-08
Iteration 3:   f(b) =  1.829e-14

Group variable: id                            Number of obs        =       891
Time variable: year                           Number of groups     =       140

Fixed-effects model                           Obs per group:    min =         6
                                                                avg =  6.364286
                                                                max =         8

                                   (Std. err. adjusted for clustering on id)
------------------------------------------------------------------------------
             |               Robust
           n | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
           n |
         L1. |   .7795513   .1171015     6.66   0.000     .5500366    1.009066
             |
           w |  -.4609536   .1117199    -4.13   0.000    -.6799206   -.2419865
           k |   .2429143   .0580169     4.19   0.000     .1292033    .3566253
       _cons |   1.750505   .4455191     3.93   0.000     .8773034    2.623706
------------------------------------------------------------------------------

. estimates store fe
```

Bias-corrected random-effects estimator

```
. xtdpdbc n w k, re vce(robust)

Bias-corrected estimation

Step 1:
(iteration log partly omitted)
Iteration 3:   f(b) =  .00483265

Step 2:
(iteration log partly omitted)
Iteration 4:   f(b) =  .07006667

Group variable: id                          Number of obs         =       891
Time variable: year                         Number of groups      =       140

Random-effects model                        Obs per group:    min =         6
                                                              avg =  6.364286
                                                              max =         8

                                     (Std. err. adjusted for clustering on id)
------------------------------------------------------------------------------
             |               Robust
           n | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
           n |
         L1. |   .6424014   .0559358    11.48   0.000     .5327693    .7520336
             |
           w |  -.3010359   .0775378    -3.88   0.000    -.4530073   -.1490646
           k |   .3075345    .047089     6.53   0.000     .2152418    .3998271
       _cons |   1.427429   .2735669     5.22   0.000     .8912479    1.963611
------------------------------------------------------------------------------
```

Postestimation specification tests

- Hansen (1982) overidentification test (RE versus FE):

```
. estat overid

Hansen test of the overidentifying restrictions        chi2(2)      =    9.8093
H0: overidentifying restrictions are valid             Prob > chi2 =    0.0074
```

- Generalized Hausman (1978) test (RE versus FE):

```
. estat hausman fe

Generalized Hausman test                               chi2(2)      =    9.0373
H0: coefficients do not systematically differ          Prob > chi2 =    0.0109
```

- Arellano and Bond (1991) serial-correlation test:

```
. estat serial

Arellano-Bond test for autocorrelation of the first-differenced residuals
H0: no autocorrelation of order 1:     z =    -2.6032   Prob > |z|  =    0.0092
H0: no autocorrelation of order 2:     z =    -0.5955   Prob > |z|  =    0.5515
```

Higher-order autoregressive model and time effects

```
. xtdpdbc n w k, fe vce(robust) lags(2) teffects nolog

Bias-corrected estimation

Group variable: id                          Number of obs        =       751
Time variable: year                         Number of groups     =       140

Fixed-effects model                         Obs per group:   min =         5
                                                             avg =  5.364286
                                                             max =         7

                                      (Std. err. adjusted for clustering on id)
------------------------------------------------------------------------------
             |               Robust
           n | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
           n |
         L1. |   .8497413   .1276216     6.66   0.000     .5996076    1.099875
         L2. |  -.1058313   .1069847    -0.99   0.323    -.3155175    .1038548
             |
           w |  -.4105421   .1694169    -2.42   0.015     -.742593   -.0784911
           k |   .2569002   .0590054     4.35   0.000     .1412516    .3725487
             |
        year |
       1979  |   .0001341   .0090099     0.01   0.988    -.0175249    .0177932
       1980  |  -.0310339    .010839    -2.86   0.004    -.0522781   -.0097898
       1981  |    -.07454     .01572    -4.74   0.000    -.1053507   -.0437294
       1982  |  -.0341935   .0160473    -2.13   0.033    -.0656455   -.0027414
       1983  |    .009513   .0192666     0.49   0.621    -.0282489    .0472749
       1984  |   .0338537   .0309918     1.09   0.275    -.0268891    .0945966
             |
       _cons |    1.65538   .5712723     2.90   0.004     .5357066    2.775053
```

Time effects and postestimation specification tests

```
. estimates store fe

. xtdpdbc n w k, re vce(robust) lags(2) teffects nolog eigtolerance(0.1)
(output omitted)

. estat overid
note: degrees of freedom adjusted for time effects in unbalanced panels

Hansen test of the overidentifying restrictions        chi2(2)      =   13.2684
H0: overidentifying restrictions are valid             Prob > chi2 =    0.0013

. estat hausman fe
note: degrees of freedom might be incorrect -- use option df()

Generalized Hausman test                               chi2(6)      =   19.2573
H0: coefficients do not systematically differ          Prob > chi2 =    0.0038

. estat hausman fe, df(2)

Generalized Hausman test                               chi2(2)      =   19.2573
H0: coefficients do not systematically differ          Prob > chi2 =    0.0001
```

- In unbalanced panels, the degrees of freedom need to be adjusted because the additional moment conditions for the time effects used by the RE-BC estimator are irrelevant for the tested hypothesis. This is only detected automatically if time effects are specified with the `teffects` option.

Technical comment on multiple solutions

- Due to the nonlinearity of the bias-corrected moment function, there are multiple solutions and the numerical algorithm occasionally converges to an incorrect solution. The command automatically re-initializes the optimization with alternative initial parameter values until a correct solution is found (Breitung, Kripfganz, and Hayakawa, 2021).

```
. set seed 20220908
. xtdpdbc n w k, fe vce(robust) from(0.99 0 0 0, copy)

Bias-corrected estimation
Iteration 0:   f(b) =  .02226985
Iteration 1:   f(b) =  .01587185
Iteration 2:   f(b) =   .0006813
Iteration 3:   f(b) =  7.378e-06
Iteration 4:   f(b) =  9.699e-09
Iteration 5:   f(b) =  3.249e-14
incorrect solution -- reinitialize
Iteration 0:   f(b) =  7.0030413
Iteration 1:   f(b) =  .00300671
Iteration 2:   f(b) =  .00001076
Iteration 3:   f(b) =  1.375e-08
Iteration 4:   f(b) =  3.285e-14

(output omitted)
```

Summary

- xtdpdbc provides an easy-to-use implementation of a bias-corrected method of moments estimator for linear dynamic panel data models with fixed or random effects:
  - Attractive finite-sample properties
  - Higher-order autoregressive models supported
  - Analytical standard errors available
  - Time-invariant regressors can be included under the random-effects assumption
  - Specification tests for random effects versus fixed effects readily available

```
ssc install xtdpdbc
net install xtdpdbc, from(http://www.kripfganz.de/stata/)

help xtdpdbc
help xtdpdbc postestimation
```

References

- Arellano, M., and S. R. Bond (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies 58* (2): 277–297.
- Breitung, J., S. Kripfganz, and K. Hayakawa (2021). Bias-corrected method of moments estimators for dynamic panel data models. *Econometrics and Statistics*: forthcoming.
- Bruno, G. S. F. (2005). Estimation and inference in dynamic unbalanced panel-data models with a small number of individuals. *Stata Journal 5* (4): 473–500.
- Bun, M. J. G., and M. A. Carree (2005). Bias-corrected estimation in dynamic panel data models. *Journal of Business & Economic Statistics 23* (2): 200–210.
- Dhaene, G., and K. Jochmans (2016). Likelihood inference in an autoregression with fixed effects. *Econometric Theory 32* (5): 1178–1215.
- Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica 50* (4): 1029–1054.
- Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica 46* (6): 1251–1271.
- Kiviet, J. F. (1995). On bias, inconsistency, and efficiency of various estimators in dynamic panel data models. *Journal of Econometrics 68* (1): 53–78.
- Kripfganz, S. (2016). Quasi-maximum likelihood estimation of linear dynamic short-T panel-data models. *Stata Journal 16* (4): 1013–1038.
- Kripfganz, S. (2019). Generalized method of moments estimation of linear dynamic panel data models. *Proceedings of the 2019 London Stata Conference*.
- Nickel, S. (1981). Biases in dynamic models with fixed effects. *Econometrica 49* (6): 1417–1426.