Bootstrap-based bias correction and inference for dynamic panels with fixed effects

Ignace De Vos[1]
Gerdie Everaert[2]
Ilse Ruyssen[2,3]

[1]Dep. of Economics, Lund University
[2]Dep. of Economics, Ghent University
[3]UNU-CRIS

Stata Conference 2021, June 7

Inconsistency of FE estimator in dynamic panels

A major advantage of panel data is that repeated observations on the same units allows to analyze individual dynamics, typically modeled by adding lagged dependent variables to the individual effects panel model specification

Inconsistency of FE estimator in dynamic panels

A major advantage of panel data is that repeated observations on the same units allows to analyze individual dynamics, typically modeled by adding lagged dependent variables to the individual effects panel model specification

BUT the standard Fixed Effects (FE) estimator is inconsistent when $N \to \infty$ while T is fixed (Nickell, 1981)

Inconsistency of FE estimator in dynamic panels

A major advantage of panel data is that repeated observations on the same units allows to analyze individual dynamics, typically modeled by adding lagged dependent variables to the individual effects panel model specification

BUT the standard Fixed Effects (FE) estimator is inconsistent when $N \to \infty$ while T is fixed (Nickell, 1981)

$\Rightarrow$ Various alternative estimators have been proposed

Generalized method of moments estimators

Difference GMM by Arellano & Bond (1991); System GMM by Arellano & Bover (1995) and Blundell & Bond (1998)

Advantage: Under appropriate assumptions: asymptotically unbiased when N tends to infinity and T is finite

Generalized method of moments estimators

Difference GMM by Arellano & Bond (1991); System GMM by Arellano & Bover (1995) and Blundell & Bond (1998)

Advantage: Under appropriate assumptions: asymptotically unbiased when N tends to infinity and T is finite

Disadvantage: Poor small-sample properties given instrumental-variables technique:

- relatively large standard deviation compared with the FE estimator (Arellano & Bond, 1991; Kiviet, 1995)
- finite-sample bias due to weak-instrument problems (Ziliak, 1997; Bun & Windmeijer, 2010)
- highly unstable GMM estimates over alternative instrument sets (Roodman, 2009)

Kiviet (1995) bias-corrected FE estimator

Based on an analytical approximation of the standard FE estimator's small sample bias in a first-order dynamic panel data model (see `xtlsdvc` by Bruno, 2015)

Advantage: Superior small sample properties compared to GMM estimators (removes most of the bias of the FE estimator while maintaining its relatively small coefficient uncertainty)

Kiviet (1995) bias-corrected FE estimator

Based on an analytical approximation of the standard FE estimator's small sample bias in a first-order dynamic panel data model (see `xtlsdvc` by Bruno, 2015)

Advantage: Superior small sample properties compared to GMM estimators (removes most of the bias of the FE estimator while maintaining its relatively small coefficient uncertainty)

Disadvantage: Bias expression of the FE estimator is derived under strict set of assumptions (homoscedasticity, etc.)
$\rightarrow$ Correction procedure needs to be re-derived to be applicable in less restrictive settings (see e.g. Bun, 2003, or Bun & Carree, 2006)

Everaert & Pozzi (2007) bias-corrected FE estimator

Develop a bootstrap-based bias correction procedure with similar small sample properties as those of Kiviet's (1995) analytical bias-corrected FE estimator

Advantages:

- Does not require an analytical expression for the bias of the FE estimator as this is numerically evaluated using bootstrap resampling
- Applicable in non-standard cases through an adequate modification of the bootstrap resampling scheme

Contribution of the paper

Stata routine, `xtbcfe`, that executes a bootstrap-based bias-corrected FE (BCFE) estimator building on Everaert & Pozzi (2007), yet

- simplifying the core of their bootstrap algorithm
- extending the algorithm to allow for higher order and unbalanced panels
- inference can be carried out using either a parametric or non-parametric bootstrapped variance-covariance matrix or percentile intervals
- allowing for a variety of initialization and resampling schemes to accommodate general heteroscedasticity patterns and error cross-sectional dependence (CSD)

Error (re)sampling schemes

To accommodate various distributional assumptions about the error term ε_{it}, our bootstrap algorithm includes several parametric error sampling and non-parametric error resampling options

All of these rely in some way on the rescaled error terms $\widehat{\varepsilon}^{r}_{it}$

$$\widehat{\varepsilon}^{r}_{it} = \widehat{\varepsilon}_{it}\sqrt{\frac{NT}{NT-k-N}} \tag{1}$$

Parametric sampling schemes

Draw ε_{it}^{b} from the i.i.d. $\mathcal{N}\left(0, \widehat{\sigma}_{it}^{2}\right)$ distribution, and allowing for

- cross-sectional heteroscedasticity: $\sigma_{it}^{2} = \widehat{\sigma}_{i}^{2} = \frac{1}{T}\sum_{t=1}^{T}\left(\widehat{\varepsilon}_{it}^{r}\right)^{2}$
- temporal heteroscedasticity: $\sigma_{it}^{2} = \widehat{\sigma}_{t}^{2} = \frac{1}{N}\sum_{i=1}^{N}\left(\widehat{\varepsilon}_{it}^{r}\right)^{2}$
- assuming homoskedasticity $\widehat{\sigma}_{it}^{2} = \widehat{\sigma}^{2}$

No account of general heteroscedastiticy $\left(\sigma_{it}^{2}\right)$ or error CSD $\left(\sigma_{ijt} \neq 0\right)$ (would require specific assumptions about the functional form of these error structures)

Non-parametric resampling schemes

Obtain ε_{it}^{b} by resampling the rescaled error terms $\widehat{\varepsilon}_{it}^{r}$

- no distributional assumptions required about ε_{it} while its covariance structure can be preserved by an appropriate design of the resampling scheme

j_{it} and s_{it} denote cross-section and time series bootstrap indices drawn specifically for cross-section i at time t

The way these indices are drawn (with replacement) from the cross-section index $(1, \ldots, N)$ and the time index $(1, \ldots, T)$ is aligned with the alleged covariance structure in ε_{it}

Non-parametric sampling schemes II

1. **Homoscedasticity** $(\sigma_{it}^2 = \sigma^2)$: resample $\widehat{\varepsilon}_{it}^r$ both over cross-sections and time, i.e. draw j_{it} from $(1, \ldots, N)$ and s_{it} from $(1, \ldots, T)$
2. **Pure cross-sectional heteroscedasticity** $(\sigma_{it}^2 = \sigma_i^2)$: resample $\widehat{\varepsilon}_{it}^r$ over time within cross-sections, i.e. draw s_{it} from $(1, \ldots, T)$ while:
   - 2.1 if σ_i^2 random over cross-sections $\rightarrow$ draw entire cross-sections and resample over time within cross-sections: $j_{it} = j_i$
   - 2.2 if σ_i^2 cross-section specific $\rightarrow$ resample over time within cross-sections: $j_{it} = i$
3. **Pure temporal heteroscedasticity** $(\sigma_{it}^2 = \sigma_t^2)$: resample $\widehat{\varepsilon}_{it}^r$ over cross-sections within time periods, i.e. draw j_{it} from $(1, \ldots, N)$ while
   - 3.1 **Unconditional** $\rightarrow$ draw entire time periods and resample over cross-sections within time periods: $s_{it} = s_t$
   - 3.2 **Conditional** $\rightarrow$ resample over the cross-sectional dimension: $s_{it} = t$

Non-parametric sampling schemes III

4 **General heteroscedasticity** (σ_{it}^2): use wild bootstrap suggested by Liu (1988) and Mammen (1993) to preserve both the cross-sectional and the temporal structure of the error terms

- 4.1 If unconditional variance σ_i^2 is random over cross-sections, first resample entire cross-sections and next apply the wild bootstrap, i.e. $\varepsilon_{it}^b = \iota_{it}\widehat{\varepsilon}_{j_i,t}^r$
- 4.2 If unconditional variance σ_i^2 is cross-section specific, apply a pure wild bootstrap, i.e. $\varepsilon_{it}^b = \iota_{it}\widehat{\varepsilon}_{it}^r$

5 **Error CSD** $(\sigma_{ijt} \neq 0)$: the covariance between ε_{it} and ε_{jt} is non-zero and may be different at each point in time:

- 5.1 Under global CSD: resample over cross-sections within time periods
- 5.2 Under local CSD: resample over time in the same way for each cross-section, i.e. restrict $j_{it} = i$ as under 2.2 and $s_{it} = s_t$ as under 3.2

Initialization

The calculation of the bootstrap data y_{it}^b of the algorithm requires initial values for the lags of the dependent variable $(y_{i,-(p-1)}^b, \ldots, y_{i0}^b)$

How these initial values are chosen to be generated depends implicitly on the decision about the initial conditions of the data

The initialization choice will influence

- the statistical properties of the estimator
- numerical properties of the algorithm in small datasets

Deterministic initialization

Fastest and most straightforward: set $(y^b_{i,-(p-1)}, \ldots, y^b_{i0})$ equal to the observed (centered) initial values $(\widetilde{y}_{i,-(p-1)}, \ldots, \widetilde{y}_{i0})$ in each bootstrap sample

Advantages

- No assumptions needed about how the initial conditions are generated (Everaert & Pozzi, 2007)
- Avoid generating initial conditions when the data is not rich enough

But risk of inducing a spurious dependency over bootstrap samples, especially if T is small!

Analytic inititialization

Draw initial observations from the multivariate normal distribution

$$\left(y_{i0}^{b},\ldots,y_{i,-(p-1)}^{b}\right)\sim\mathcal{N}\left(\widehat{\mu}_{i}^{0},\widehat{\Sigma}_{i}^{0}\right)$$

In the case of a single lagged dependent variable $(p=1)$, for instance:

$$\widehat{\Sigma}_{i}^{0}=\frac{1}{T}\sum_{t=1}^{T}\left(\widetilde{y}_{it}-\frac{\widetilde{X}_{it}\widehat{\beta}}{1-\widehat{\gamma}_{1}}\right)^{2},$$

which is the variance of y_{it} around its unconditional mean $\widetilde{X}_{it}\widehat{\beta}\,/(1-\widehat{\gamma}_{1})$ observed over the sample

Burn-in initialization

Alternative: start in the distant past from initial values set to zero, e.g. $(y^b_{i,-50-p+1} = 0, \ldots, y^b_{i,-50} = 0)$, and then generate the series y^b_{il}, with $l = -49, \ldots, 0$, setting $\widetilde{X}_{il} = \widetilde{X}_{i0}$

Then simply use $(y^b_{i,-(p-1)}, \ldots, y^b_{i0})$ as initial values and discard the earlier generated values

Advantage: does not require a distributional assumption for the initial conditions plus the error resampling scheme used to generate the actual sample can also be used to generate the initial values

Inference

The small sample distribution of the BCFE estimator can be simulated by resampling the original data and applying the bootstrap bias-correction to the FE estimates obtained in each of the constructed samples

From this simulated distribution we then calculate standard errors and confidence intervals

The resampling of the original data can be done using a parametric or a non-parametric approach

Parametric approach

In the last iteration over the bias-correction procedure, we already obtained J bootstrap samples from a population where our bias-corrected FE estimate $\widehat{\delta}^{bc}$ is used as a proxy for the population parameter vector δ

$\Rightarrow$ the distribution of the BCFE estimator can be obtained by applying the bias-correction procedure to the J FE estimates $\widehat{\delta}_j^b$ obtained in the iterative bootstrap procedure

Advantage: the resampling of the data used to obtain the small sample distribution of the BCFE estimator is exactly the same as the resampling of the data used to bias-correct the FE estimator

Non-parametric approach

As suggested by Kapetanios (2008), resample the original data for cross-sectional units as a whole with replacement

Advantages

- Preserves the dynamic panel structure without the need to make parametric assumptions
- Valid under general heteroscedasticity patterns and a global CSD structure in the data (e.g. a common factor structure)

Yet not valid under local CSD (e.g. a spatial panel structure)

Xtbcfe syntax

The bootstrap procedures presented and tested in this paper are all contained in the `xtbcfe` routine. The basic syntax is as follows:

`xtbcfe` *depvar* [*indepvars*] [*if*] [, `lags(#)`
`resampling(`*string*`)` `initialization(`*string*`)` `bciters(#)`
`criterion(#)` `inference(`*string*`)` `infiters(#)`
`distribution(`*string*`)` `level(#)` `param` `te`]

The program adds the lagged dependent variable(s) as the first explanatory variable(s) and can fit the simple autoregressive model without covariates

Xtbcfe options I

`resampling(`*scheme*`)` specifies the residual resampling scheme to be used in the bootstrap procedure. The default is `resampling(mcho)`.

scheme	Description
`mcho`	drawing from the normal distribution with estimated homogeneous variance; the default
`mche`	drawing from the normal distribution with estimated heterogeneous (cross-section-specific) variance
`mcthe`	drawing from the normal distribution with period (t)-specific estimated variance
`iid`	for resampling independently over both cross-sections and time
`cshet`	for resampling within cross-sections (cross-sectional heteroskedasticity)
`cshet_r`	for resampling within cross-sections with randomized indices (random cross-sectional heteroskedasticity)
`thet`	for resampling within time periods (temporal heteroskedasticity)
`thet_r`	for resampling within time periods with permuted t (random temporal heteroskedasticity)
`wboot`	for wild bootstrap, that is, error terms multiplied by 1 or -1 (general heteroskedasticity)
`wboot_r`	for randomized wild bootstrap, that is, permuted cross-section indices and error terms multiplied by 1 or -1 (random general heteroskedasticity, balanced panels only)
`csd`	for resampling identically over cross-sections (cross-sectional dependence, balanced panels only)

Xtbcfe options II

`lags(#)` specifies the number of lags of the dependent variable to be included among the predictors. The default is `lags(1)`.

`initialization(`*initial*`)` determines the initialization scheme for the bootstrapped lagged dependent variables $(y^b_{i,-(p-1)}, \ldots, y^b_{i0})$. The default is `initialization(det)`.

initial	Description
`det`	deterministic initialization, that is, $(y^b_{i,-(p-1)}, \ldots, y^b_{i0}) = (\widetilde{y}_{i,-(p-1)}, \ldots, \widetilde{y}_{i0})$; the default
`bi`	burn-in initialization using the resampling scheme defined by `resampling()` over the burn-in sample
`aho`	analytical homogeneous initiation $(y^b_{i,-(p-1)}, \ldots, y^b_{i0}) \sim \mathcal{N}\left(\widehat{\mu}^0_i, \widehat{\boldsymbol{\Sigma}}^0\right)$
`ahe`	analytical heterogeneous initiation $(y^b_{i,-(p-1)}, \ldots, y^b_{i0}) \sim \mathcal{N}\left(\widehat{\mu}^0_i, \widehat{\boldsymbol{\Sigma}}^0_i\right)$

Xtbcfe options III

`bciters(#)` sets the number of bootstrap iterations used for the construction of the bias-corrected FE estimator (at least 50). The default is `bciters(250)`.

`criterion(#)` alters the convergence criterion used in the estimation algorithm. The default is `criterion(0.005)`. The specified number will be multiplied by the number of lags (p) of the dependent variable.

`inference(`*option*`)` specifies the type of SEs and CIs. Under the `inference(inf_se)` option, SEs are bootstrapped and are then used to calculate CIs using the Student t distribution. Alternatively, because this distributional assumption may be violated, especially in small datasets with high temporal dependence, the `inference(inf_ci)` option calculates CIs directly from the bootstrap distribution. This approach does not make a distributional assumption but is much more computationally intensive because, compared with calculating SEs, adequate calculation of the desired percentiles requires more bootstrap samples. Finally, the `inference(inf_appr)` option is a fast alternative that approximates SEs by calculating the dispersion of the FE estimator over the bootstrap iterations. While this is much faster than the other options, the resulting SEs are expected to be downward biased, so they should only be used as a rough approximation. We report some Monte Carlo results in section 4 to indicate the relative accuracy of the different inference methods.

Xtbcfe options IV

`infiters(#)` specifies the number of bootstrap iterations to be used for inference. The default is `infiters(250)` for all choices of `inference()`. It is recommended to have at least 50 iterations for bootstrapping SEs and 1,000 iterations for bootstrapping percentile intervals. The number of iterations cannot be smaller than 100 when the `inference(inf_ci)` option is used.

`distribution(`*histogram*`)` requests that the bootstrap distribution of `xtbcfe` obtained by the inference procedures be saved in `e(dist_bcfe)`. This option allows users to inspect the bootstrap distribution and calculate additional statistics from it. If this option is omitted, the distribution will be deleted after estimation. Use `distribution(none)` to save the bootstrap coefficient matrix in `e(dist_bcfe)`. Specifying `distribution(sum)` will additionally display a histogram of the bootstrap distribution for the sum of AR coefficients. The `distribution(all)` option adds histograms for all AR coefficients separately.

`level(#)` specifies the confidence level used to construct CIs. The default is `level(95)`.

`param` requests that inference procedures be initiated using the parametric bootstrap instead of the nonparametric default (see section 2.5).

`te` requests the addition of time effects to the specification. Time dummies are generated and named according to the time indicator used in the `xtset` command. User-specified variables bearing the same name will be overwritten. Time dummies included in *indepvars* will be removed.

Xtbcfe stored results

xtbcfe stores the following in **e()**:

Scalars

e(N)	number of observations
e(N_g)	number of groups
e(k)	number of exogenous regressors
e(df_r)	residual degrees of freedom
e(t_min)	minimum number of time periods
e(t_max)	maximum number of time periods
e(t_avg)	average number of time periods
e(irr)	number of cross-sections removed because of irregular spacing or lack of observations
e(conv)	convergence of the bootstrap algorithm

Macros

e(cmd)	**xtbcfe**
e(depvar)	name of dependent variable
e(predict)	program used to implement **predict**
e(ivar)	panel variable
e(tvar)	time variable

Matrices

e(b)	**xtbcfe** estimates
e(V)	variance–covariance matrix of the estimators
e(dist_bcfe)	**xtbcfe** bootstrap distribution if **distribution()** is specified
e(res_bcfe)	**xtbcfe** error terms

Functions

e(sample)	marks estimation sample

Xtbcfe post-estimation

The `xtbcfe` command supports the postestimation command `predict` (see [R] **predict**) to compute fitted values and residuals. The syntax for `predict` following `xtbcfe` is

`predict` [*type*] [*newvar*] [*if*] [`,` *statistic*]

statistic	Description
`xb`	$\sum_{s=1}^{p} \widehat{\gamma}_p y_{i,t-s} + \mathbf{x}_{it}\widehat{\boldsymbol{\beta}}$, the fitted values; the default
`ue`	$\widehat{\alpha}_i + \widehat{\varepsilon}_{it}$, the combined residuals
`xbu`	$\sum_{s=1}^{p} \widehat{\gamma}_s y_{i,t-s} + \mathbf{x}_{it}\widehat{\boldsymbol{\beta}} + \widehat{\alpha}_i$, the prediction including fixed effect
`u`	$\widehat{\alpha}_i$, the fixed effect
`e`	$\widehat{\varepsilon}_{it}$, the observation-specific error component

The `xb` and `ue` statistics are available both in and out of sample; type `predict ... if e(sample) ...` to restrict statistics to the estimation sample. The `xbu`, `u`, and `e` statistics are calculated only for the estimation sample, even when `if e(sample)` is not specified.

Monte Carlo experiments BCFE

Using Monte Carlo simulations, Everaert & Pozzi (2007) show that the BCFE estimator

- outperforms the difference and system GMM estimators, both in terms of bias and inference, in samples with small to moderate T
- is insensitive to non-normality of the errors, conditional heteroscedasticity or non-stationary initial conditions
- has a bias comparable to the analytical bias corrections of Kiviet (1995) and Bun and Carree (2005)

Further Monte Carlo simulation results illustrate the finite sample properties of our simplified BCFE bootstrap algorithm and its extension to higher-order dynamic models and error CSD

MC data generation

Data are generated from (**??**) with x_{it} restricted to be a single exogenous explanatory variable, generated as

$$x_{it} = \rho x_{i,t-1} + \xi_{it}, \qquad \xi_{it} \sim \text{i.i.d. } \mathcal{N}(0, \sigma_\xi^2) \tag{2}$$

- Normalize the long-run impact of x_{it} to one by setting $\beta = 1 - \sum_{s=1}^{p} \gamma_s$
- Each experiment is based on 1000 iterations, where in each sample we generate $50 + T$ periods and discard the first 50 observations
- The BCFE estimator is implemented setting the number of bootstrap iterations (`bciters`) to 250
- Analyze the performance of alternative initialization schemes and adjust the bootstrap resampling scheme according to the properties of the DGP of y_{it}

MC reporting

We report

(i) mean bias ($bias$)

(ii) standard error (se)

(iii) mean estimated standard error ($\widehat{se}$)

(iv) real size ($size$)

We also include results for Pooled OLS (POLS), FE and for the analytical correction (BCFE_{an}) implemented in the `xtlsdvc` routine developed by Bruno(2005) initiated with the Anderson-Hsiao estimator and standard errors obtained through 200 bootstrap iterations

Simplification bootstrap algorithm

Monte Carlo results for an $AR(1)$ model with $\gamma_1 = 0.8$

	bias	*se*	$\widehat{se}$			*size*			*bias*	*se*	$\widehat{se}$			*size*		
			appr	1000	50	*appr*	*se*	*ci*			*appr*	1000	50	*appr*	*se*	*ci*
	$T = 4, N = 20$								$T = 9, N = 20$							
POLS	0.04	0.06	0.06	-	-	0.12	-	-	0.04	0.04	0.04	-	-	0.22	-	-
FE	−0.51	0.13	0.12	-	-	0.97	-	-	−0.24	0.07	0.07	-	-	0.96	-	-
$BCFE_{an}$	−0.18	0.17	0.16	-	-	0.21	-	-	−0.05	0.08	0.08	-	-	0.09	-	-
$BCFE_{or}$	−0.14	0.15	-	-	-	-	-	0.25	−0.04	0.09	-	-	-	-	-	0.11
$BCFE_{de}$	0.07	0.17	0.13	0.17	0.17	0.15	0.16	0.09	0.03	0.10	0.07	0.09	0.08	0.21	0.14	0.08
$BCFE_{an}$	0.00	0.16	0.13	0.16	0.15	0.08	0.09	0.05	0.00	0.09	0.07	0.08	0.08	0.11	0.09	0.08
$BCFE_{bi}$	−0.04	0.17	0.13	0.16	0.16	0.13	0.10	0.09	−0.01	0.09	0.07	0.08	0.08	0.11	0.09	0.10
	$T = 4, N = 100$								$T = 9, N = 100$							
POLS	0.05	0.03	0.03	-	-	0.47	-	-	0.05	0.02	0.02	-	-	0.76	-	-
FE	−0.51	0.06	0.06	-	-	1.00	-	-	−0.23	0.03	0.03	-	-	1.00	-	-
$BCFE_{an}$	−0.13	0.08	0.09	-	-	0.30	-	-	−0.03	0.04	0.04	-	-	0.14	-	-
$BCFE_{or}$	−0.13	0.07	-	-	-	-	-	0.80	−0.04	0.04	-	-	-	-	-	0.35
$BCFE_{de}$	0.09	0.07	0.06	0.07	0.07	0.40	0.31	0.20	0.03	0.05	0.03	0.05	0.04	0.32	0.13	0.07
$BCFE_{an}$	0.04	0.08	0.06	0.07	0.07	0.20	0.15	0.07	0.00	0.04	0.03	0.04	0.04	0.14	0.08	0.07
$BCFE_{bi}$	−0.02	0.09	0.06	0.09	0.08	0.21	0.07	0.05	−0.01	0.04	0.03	0.04	0.04	0.13	0.06	0.07

Error CSD

Non-standard scenario with cross-sectionally dependent errors: focus on a pure $(\beta = 0)$ first-order autoregressive model with $\gamma_1 = 0.8$ and assume that the error term ε_{it} has the following common factor structure

$$\varepsilon_{it} = \lambda_i F_t + \epsilon_{it},$$

with $F_t \sim$ i.i.d. $\mathcal{N}(0,1)$ and $\epsilon_{it} \sim$ i.i.d. $\mathcal{N}(0,1)$

We follow Sarafidis & Robertson (2009) and generate the factor loadings as $\lambda_i \sim$i.i.d.$\mathcal{U}(1,4)$ and set the individual effect variance to $\sigma_\alpha^2 = (1-\gamma_1)(1+\gamma_1)^{-1}(\mu_\lambda^2 + \sigma_\lambda^2 + 1)$, with μ_λ and σ_λ^2 being the mean and variance of the factor loading distribution

Error CSD

Monte Carlo results for an $AR(1)$ model with $\gamma_1 = 0.8$

	$bias$	se	$\widehat{se}$	$rmse$	$size_t$	$size_{ci}$	$bias$	se	$\widehat{se}$	$rmse$	$size_t$	$size_{ci}$
	$T = 5, N = 20$						$T = 10, N = 20$					
POLS	0.088	0.049	0.046	0.101	0.53	-	0.090	0.037	0.032	0.098	0.77	-
FE	−0.443	0.120	0.104	0.459	0.98	-	−0.226	0.075	0.061	0.238	0.94	-
$BCFE_{an}$	−0.169	0.148	0.134	0.225	0.25	-	−0.053	0.087	0.077	0.102	0.13	-
$BCFE_{csd}$	0.041	0.159	0.143	0.165	0.17	0.09	0.019	0.100	0.084	0.102	0.15	0.10
$BCFE_{thet}$	−0.020	0.167	0.146	0.168	0.15	0.10	−0.002	0.097	0.084	0.097	0.13	0.11
	$T = 5, N = 100$						$T = 10, N = 100$					
POLS	0.098	0.020	0.020	0.101	0.99	-	0.098	0.014	0.014	0.099	1.00	-
FE	−0.430	0.056	0.046	0.434	1.00	-	−0.220	0.034	0.027	0.222	1.00	-
$BCFE_{an}$	−0.105	0.076	0.074	0.130	0.30	-	−0.029	0.044	0.038	0.053	0.16	-
$BCFE_{csd}$	0.070	0.086	0.078	0.111	0.24	0.13	0.022	0.047	0.044	0.052	0.09	0.07
$BCFE_{thet}$	−0.003	0.085	0.080	0.085	0.09	0.07	−0.001	0.044	0.042	0.044	0.07	0.08

Second-order dynamic model

Subsequently, assume $\alpha_i \sim$ i.i.d. $\mathcal{N}(0,1)$ and $\varepsilon_{it} \sim$ i.i.d. $\mathcal{N}(0,1)$, with x_{it} generated from (2) setting $\rho = 0.5$ and assuming $\xi_{it} \sim$ i.i.d. $\mathcal{N}(0,1)$

Results for a series with strong temporal dependence, setting either

- $\gamma_1 = 0.6$ and $\gamma_2 = 0.2 \rightarrow$ an unbiased estimator is expected to lie between POLS and FE, but probably closer to the former than to the latter
- $\gamma_1 = 1.1$ but maintain the stationarity assumption by setting γ_2 to -0.2 $\rightarrow$ an unbiased estimator is expected to lie closer to the POLS estimator for γ_1 but closer to the FE estimator for γ_2

Conclusion: xtbcfe performs well in the 2nd order dynamic model with near perfect test size as N becomes large

Second-order dynamic model

Monte Carlo resuts for an $AR(2)$ model

Case 1: $\gamma_1 = 0.6$ and $\gamma_2 = 0.2$

	γ_1				γ_2				γ_1				γ_2			
	bias	*se*	$\widehat{se}$	$size_t$	*bias*	*se*	$\widehat{se}$	$size_t$	*bias*	*se*	$\widehat{se}$	$size_t$	*bias*	*se*	$\widehat{se}$	$size_t$
	$T=5, N=20$								$T=10, N=20$							
POLS	0.08	0.09	0.10	0.10	0.10	0.09	0.10	0.17	0.09	0.07	0.07	0.25	0.09	0.07	0.07	0.26
FE	−0.39	0.12	0.11	0.92	−0.20	0.11	0.11	0.40	−0.18	0.08	0.07	0.63	−0.11	0.07	0.07	0.33
BCFE	−0.03	0.14	0.13	0.09	−0.02	0.13	0.13	0.07	−0.01	0.09	0.08	0.09	−0.01	0.08	0.08	0.08
	$T=5, N=100$								$T=10, N=100$							
POLS	0.09	0.04	0.04	0.59	0.09	0.04	0.04	0.58	0.09	0.03	0.03	0.87	0.09	0.03	0.03	0.87
FE	−0.38	0.05	0.05	1.00	−0.19	0.05	0.05	0.97	−0.17	0.04	0.03	1.00	−0.11	0.03	0.03	0.90
BCFE	−0.01	0.07	0.07	0.07	−0.01	0.06	0.06	0.06	−0.01	0.04	0.04	0.08	−0.01	0.04	0.04	0.07

Case 2: $\gamma_1 = 1.1$ and $\gamma_2 = -0.2$

	γ_1				γ_2				γ_1				γ_2			
	bias	*se*	$\widehat{se}$	$size_t$	*bias*	*se*	$\widehat{se}$	$size_t$	*bias*	*se*	$\widehat{se}$	$size_t$	*bias*	*se*	$\widehat{se}$	$size_t$
	$T=5, N=20$								$T=10, N=20$							
POLS	0.02	0.10	0.10	0.05	0.07	0.10	0.10	0.09	0.04	0.07	0.07	0.07	0.05	0.07	0.07	0.10
FE	−0.42	0.12	0.11	0.95	−0.02	0.11	0.11	0.06	−0.18	0.08	0.07	0.66	−0.04	0.08	0.07	0.09
BCFE	−0.05	0.13	0.13	0.08	0.00	0.13	0.13	0.09	−0.00	0.09	0.08	0.10	−0.01	0.08	0.08	0.09
	$T=5, N=100$								$T=10, N=100$							
POLS	0.04	0.04	0.04	0.16	0.05	0.04	0.04	0.19	0.05	0.03	0.03	0.28	0.05	0.03	0.03	0.33
FE	−0.40	0.06	0.05	1.00	−0.02	0.05	0.05	0.08	−0.18	0.04	0.03	1.00	−0.04	0.03	0.03	0.21
BCFE	−0.01	0.06	0.06	0.06	−0.00	0.06	0.06	0.05	−0.00	0.04	0.04	0.06	−0.00	0.03	0.04	0.05

"Xtbcfe" Take-Home

Iterative bootstrap-based bias-corrected FE estimator for dynamic panels building on Everaert & Pozzi (2007)

Various bootstrap error resampling schemes to account for general heteroscedasticity and contemporaneous CSD $\rightarrow$ choose the alternative that incorporates the highest degree of randomness in the resampling process

Inference using parametric or non-parametric bootstrapped variance-covariance matrices or percentile intervals

MC: the simplification of the original algorithm results in a BCFE estimator that is virtually unbiased for very small T; support the BCFE in higher order dynamic panels and panels with contemporaneous error CSD

Thank you!

Questions? Comments? Suggestions?

Ilse Ruyssen
Department of Economics, Ghent University & UNU-CRIS
Sint-Pietersplein 6, B-9000 Ghent, Belgium
Ilse.Ruyssen@Ugent.be

The model

Homogeneous dynamic panel data model of order p

$$y_{it} = \alpha_i + \sum_{s=1}^{p} \gamma_s y_{i,t-s} + x_{it}\beta + \varepsilon_{it}, \tag{3}$$

with

- $i = 1, ..., N$ and $t = 1, ..., T$ being the cross-section and time-series dimension, respectively
- y_{it} is the dependent variable
- x_{it} is a $(1 \times (k - p))$ vector of strictly exogenous explanatory variables, where k is the total number of time-varying regressors
- α_i is an unobserved individual effect that may be correlated with x_{it}

Assumptions

Assumptions regarding the error term ε_{it}:

$$\begin{aligned} &\text{(i)} \; E\left[\varepsilon_{it}\varepsilon_{js}\right] = 0, && \forall i, j \text{ and } t \neq s, \\ &\text{(ii)} \; E\left[\varepsilon_{it}^2\right] = \sigma_{it}^2, && \forall i, t, \\ &\text{(iii)} \; E\left[\varepsilon_{it}\varepsilon_{jt}\right] = \sigma_{ijt}, && \forall i, j, t \text{ and } i \neq j, \end{aligned}$$

Initial values $(y_{i,-(p-1)}, \ldots, y_{i0})$ are observed such that T is the actual time series dimension available for estimation

While the bias-correction algorithm allows for an unbalanced dataset, we present the methodology with a balanced data set for simplicity

Stacked version

Stacking observations over time and cross-sections we obtain

$$y = W\delta + D\alpha + \varepsilon, \tag{4}$$

where

- y is the $(NT \times 1)$ vector stacking the observations y_{it}
- $W = (y_{-1}, \ldots, y_{-p}, X)$ is the $(NT \times k)$ matrix stacking observations on the lags of the dependent variable $(y_{i,t-1}, \ldots, y_{i,t-p})$ and the exogenous explanatory variables x_{it}
- $\delta = (\gamma', \beta')'$ is the $k \times 1$ parameter vector of interest
- D is a $NT \times N$ dummy variable matrix calculated as $D = I_N \otimes \iota_T$ with ι_T a $T \times 1$ vector of ones
- the variance-covariance matrix of ε is denoted Σ

The FE estimator

Let $M_D = I_N \otimes \left(I_T - D(D'D)^{-1}D'\right)$ denote the symmetric and idempotent matrix that transforms the data into deviations from individual specific sample means

Since $M_D D = 0$, the individual effects α can be eliminated from the model by multiplying equation (**??**) by M_D

$$\begin{aligned} M_D y =& M_D W\delta + M_D D\alpha + M_D\varepsilon, \\ \widetilde{y} =& \widetilde{W}\delta + \widetilde{\varepsilon}, \end{aligned} \tag{5}$$

where $\widetilde{y} = M_D y$ denotes the centered dependent variable and similarly for the other variables. The least squares estimator for δ in model (5) defines the FE estimator:

$$\widehat{\delta} = \left(\widetilde{W}'\widetilde{W}\right)^{-1}\widetilde{W}'\widetilde{y} = \left(W'M_D W\right)^{-1}W'M_D y. \tag{6}$$

An unbiased estimator

The FE estimator $\widehat{\delta}$ is biased but still an unknown function of the true parameter vector, i.e.

$$\mathrm{E}(\widehat{\delta}\,|\delta,\Sigma,T\,)=\int_{-\infty}^{+\infty}\widehat{\delta}f\left(\widehat{\delta}\,|\delta,\Sigma,T\right)d\widehat{\delta}\neq\delta, \tag{7}$$

If we are able to generate a sequence $\left(\widehat{\delta}_1,\ldots,\widehat{\delta}_J\,|\delta,\Sigma,T\right)$ of J biased FE estimates $\widehat{\delta}$ for δ, the integral in equation (7) can be written as

$$\mathrm{E}(\widehat{\delta}\,|\delta,\Sigma,T\,)=\lim_{J\to\infty}\frac{1}{J}\sum_{j=1}^{J}\widehat{\delta}_j\,|\delta,\Sigma,T\,. \tag{8}$$

$\Rightarrow \widehat{\delta}^{bc}$ is an unbiased estimator for δ if it satisfies

$$\widehat{\delta}=\lim_{J\to\infty}\frac{1}{J}\sum_{j=1}^{J}\widehat{\delta}_j\left|\widehat{\delta}^{bc},\Sigma,T\,.\right. \tag{9}$$

Proposition in Everaert & Pozzi (2007)

For any specific value of δ^*, the condition in equation (9) can be evaluated by generating J bootstrap samples from the data generating process in equation (**??**) and applying FE to each of the samples to obtain the sequence $\left(\widehat{\delta}_1, \ldots, \widehat{\delta}_J \,|\delta^*, \Sigma, T\right)$

The bias-corrected $\widehat{\delta}^{bc}$ can then be obtained by searching over different parameter values δ^* until equation (9) is satisfied

The search for $\widehat{\delta}^{bc}$ can be performed efficiently by iteratively updating the parameter vector δ^* used for the creation of bootstrap samples, taking the original biased FE estimate as the best initial guess $(\delta^*_{(0)} = \widehat{\delta})$.

Outline bootstrap algorithm

The iterative bootstrap bias-correction procedure is given by the following steps:

1. Using equation (5) and the original centered data, calculate the residuals as $\widehat{\varepsilon} = \widetilde{y} - \widetilde{W}\delta^*_{(\kappa)}$
2. Obtain J bootstrap samples, where in each sample $j = 1, \ldots, J$:
   - 2.1 Draw a bootstrap sample ε^b from $\widehat{\varepsilon}$ according to a specified (re)sampling scheme.
   - 2.2 Calculate the bootstrap sample $y^b = W^b\delta^*_{(\kappa)} + \varepsilon^b$ where $W^b = (y^b_{-1}, \ldots, y^b_{-p}, \widetilde{X})$
   - 2.3 Use FE to estimate $\widehat{\delta}^b_j = (W^{b\prime} M_D W^b)^{-1} W^{b\prime} M_D y^b$
3. Calculate $\omega_{(k)} = \widehat{\delta} - \frac{1}{J}\sum_{j=1}^{J} \widehat{\delta}^b_j$
4. Update the parameter vector $\delta^*_{(\kappa+1)} = \delta^*_{(\kappa)} + \omega_{(\kappa)}$

Notes

The bias of the FE estimator is invariant to the variance of the individual effects α as these are effectively wiped out by centering the data$\rightarrow$ we simplify the bootstrap algorithm as there is no need to estimate the individual effects α and use them to generate the data

In step 2(a) the bootstrap errors ε^b should be drawn consistently with the variance-covariance structure in the population error terms ε, as represented by $\Sigma \rightarrow$ Various (re)sampling schemes are discussed in the paper

Furthermore, the calculation of the bootstrap data y_{it}^b in step 2(b) requires initial values for $(y_{i,-(p-1)}^b, \ldots, y_{i0}^b) \rightarrow$ The choice of how these initial values should be generated implicitly entails a decision about the initial conditions of the data