FACULTEIT ECONOMIE
EN BEDRIJFSKUNDE

TWEEKERKENSTRAAT 2
B-9000 GENT
Tel. : 32 - (0)9 – 264.34.61
Fax. : 32 - (0)9 – 264.35.92

WORKING PAPER

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

Forthcoming in the *Stata Journal*

Ignace De Vos, Gerdie Everaert and Ilse Ruyssen

SHERPPA, Ghent University

April 2015
2015/906

D/2015/7012/08

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

Ignace De Vos
SHERPPA, UGent
Ghent, Belgium
ignace.devos@UGent.be

Gerdie Everaert
SHERPPA, UGent
Ghent, Belgium
gerdie.everaert@UGent.be

Ilse Ruyssen
SHERPPA,UGent
Ghent, Belgium
ilse.ruyssen@UGent.be

Abstract. This article describes a new Stata routine, `xtbcfe`, that performs the iterative bootstrap-based bias correction for the fixed effects (FE) estimator in dynamic panels proposed by Everaert and Pozzi (Journal of Economic Dynamics and Control, 2007). We first simplify the core of their algorithm using the invariance principle and subsequently extend it to allow for unbalanced and higher order dynamic panels. We implement various bootstrap error resampling schemes to account for general heteroscedasticity and contemporaneous cross-sectional dependence. Inference can be performed using a bootstrapped variance-covariance matrix or percentile intervals. Monte Carlo simulations show that the simplification of the original algorithm results in a further bias reduction for very small T. The Monte Carlo results also support the bootstrap-based bias correction in higher order dynamic panels and panels with cross-sectional dependence. We illustrate the routine with an empirical example estimating a dynamic labour demand function.

Keywords: st0001, xtbcfe, bootstrap-based bias correction, dynamic panel data, unbalanced, higher order, heteroscedasticity, cross-sectional dependence, Monte Carlo, labour demand

1 Introduction

Many empirical relationships are dynamic in nature as decision makers are not always able or willing to respond immediately to changes in their environment due to e.g. contract lock-up periods, capacity or technological constraints, slowly changing habits, etc. A major advantage of panel data is that repeated observations on the same units allows to analyze individual dynamics. These dynamic relations are typically modeled by adding lagged dependent variables to the individual effects panel model specification. Although the dynamic panel specification may seem straightforward, the combination of individual effects and lagged dependent variables poses major econometric challenges.

Nickell (1981) has shown that the standard Fixed Effects (FE) estimator is inconsistent when the number of cross section units N goes to infinity while the number of time periods T is fixed. Only when T goes to infinity, FE is consistent. Given that the (asymptotic) bias may be quite sizable in many cases relevant to applied research, various alternative estimators have been proposed. Particularly popular are a variety of generalized method of moments (GMM) estimators, most notably the difference GMM (Arellano and Bond 1991) and system GMM (Arellano and Bover 1995; Blundell and

Bond 1998) estimators. These GMM estimators are, under appropriate assumptions, asymptotically unbiased (when N tends to infinity and T is finite), but the fact that they make use of an instrumental variables technique to avoid the dynamic panel data bias often leads to poor small sample properties. First, Monte Carlo simulations show that the GMM estimators have a relatively large standard deviation compared to the FE estimator (see e.g. Arellano and Bond 1991; Kiviet 1995). Second, they may suffer from a substantial finite sample bias due to weak instrument problems (see e.g. Ziliak 1997; Bun and Kiviet 2006; Bun and Windmeijer 2010). Third, GMM estimators require decisions on which and how many instruments to use. When T is relatively large compared to N, many valid instruments are available but this instrument proliferation may render the GMM estimator invalid even though instruments are individually valid (Roodman 2009). In practice, this typically leads to highly unstable GMM estimates over alternative instrument sets.

Motivated by these disadvantages, Kiviet (1995) derived a bias-corrected FE estimator using an analytical approximation of its small sample bias in a first-order dynamic panel data model. Using Monte Carlo simulations, this bias-corrected FE estimator is shown to have superior small sample properties compared to GMM estimators, i.e. it is able to remove most of the bias of the FE estimator while maintaining its relatively small coefficient uncertainty. An extended version of this bias-corrected FE estimator is implemented in the `xtlsdvc` Stata routine written by Bruno (2005). A practical downside of Kiviet's correction however is the strict set of assumptions (homoscedasticity, etc.) under which the bias expression of the FE estimator is derived. These are often violated in practice such that the correction procedure needs to be re-derived to be applicable in less restrictive settings (see e.g. Bun 2003 for higher-order dynamic panels, cross-sectional correlation and cross-sectional heteroscedasticity, and Bun and Carree 2006 for cross-sectional and unconditional temporal heteroscedasticity). Everaert and Pozzi (2007) address this issue by using a bootstrap-based bias correction procedure. The main advantage of their approach is that it does not require an analytical expression for the bias of the FE estimator as this is numerically evaluated using bootstrap resampling. Monte Carlo studies show that the small sample properties of their bootstrap-based bias-corrected FE estimator are similar to those of the Kiviet correction. However, it has the potential to be applicable in non-standard cases through an adequate modification of the bootstrap resampling scheme.

This paper describes a new Stata routine, `xtbcfe`, that executes a bootstrap-based bias-corrected FE (BCFE) estimator building on Everaert and Pozzi (2007). We first simplify the core of their bootstrap algorithm using the fact that the bias of the FE estimator is invariant to the variance of the individual effects such that these can be ignored when generating bootstrap samples. Monte Carlo simulations show that this simplification results in a further bias reduction for very small T, implying the BCFE to be virtually unbiased for all sample sizes in a standard setting. Next, we extend 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. We allow for a variety of initialization and resampling schemes to accommodate general heteroscedasticity patterns and error cross-sectional dependence (CSD). Especially the latter is important as Phillips and Sul (2007) and Everaert and De Groote (2014) have shown that error CSD implies a substantial increase in the small T bias of the FE estimator in a dynamic model. When the CSD in the error terms is restricted to be only contemporaneous, the FE is still consistent as $T \to \infty$, though. However, for an intertemporal CSD pattern, Everaert and De Groote (2014) show that the FE estimator is inconsistent even when $T \to \infty$. The bootstrap algorithm implemented in `xtbcfe` to obtain the BCFE estimator can only account for contemporaneous CSD. We leave the extension to an intertemporal CSD pattern for future research. Using Monte Carlo simulations we show that our extended BCFE estimator also has adequate small sample properties in higher order dynamic models and panels with contemporaneous error CSD.

The remainder of this paper is structured as follows. Section 2 outlines the model and the bootstrap algorithm together with the various initialization and resampling schemes. In section 3 we provide the basic syntax for the `xtbcfe` routine. Some basic Monte Carlo results are discussed in section 4 and an application of the new routine to estimate a labour demand function is presented in section 5. Section 6 concludes.

2 Bootstrap-based bias correction for FE

2.1 Model, assumptions and FE estimator

Consider a homogeneous dynamic panel data model of order p

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

with $i = 1, ..., N$ and $t = 1, ..., T$ being the cross-section and time-series dimension respectively and where 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, and α_i is an unobserved individual effect that may be correlated with x_{it}. Regarding the error term ε_{it} we make the following assumptions

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

The first assumption states that the error terms are serially uncorrelated, both within and over cross-sections. This should not be highly restrictive as it can be accommodated by including a sufficient number of lagged values of y_{it} amongst the regressors. The second assumption allows for a general heteroscedasticity pattern, including cross-sectional heteroscedasticity ($\sigma_{it}^2 = \sigma_i^2$), temporal heteroscedasticity ($\sigma_{it}^2 = \sigma_t^2$) and general heteroscedasticity (σ_{it}^2). Note that the latter two cases not only allow for unconditional but also for conditional temporal heteroscedasticity, like e.g. generalized autoregressive conditional heteroscedasticity (GARCH). Assumption (iii) allows for a general pattern of contemporaneous CSD. This includes global CSD induced by a common factor structure (as in e.g. Stock and Watson 2002; Coakley et al. 2002; Pesaran 2006) and local CSD induced by spatial dependence (as in e.g. Anselin 1988; Kapoor 2007). Note that intertemporal cross-sectional dependence is ruled out by assumption (i).

For notational convenience we assume that the initial values $(y_{i,-(p-1)}, \ldots, y_{i0})$ are observed such that T is the actual time series dimension available for estimation. Furthermore, the bias-correction algorithm presented below allows for an unbalanced data set where the time series dimension is possibly different over cross-sections, i.e. $t = \tau_i, ..., T_i$ with τ_i and T_i respectively the first and last observed time period for individual i. We present the methodology with a balanced data set for simplicity (in notation) sake. The developed Stata routine will however automatically recognize and deal with unbalanced panels.

Stacking observations over time and cross-sections we obtain

$$y \quad = \quad W\delta + D\alpha + \varepsilon, \tag{2}$$

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 and 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, with elements defined by the assumptions (i)-(iii) above.

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 (2) by M_D

$$\begin{aligned} M_D y =& M_D W\delta + M_D D\alpha + M_D\varepsilon, \\ \widetilde{y} =& \widetilde{W}\delta + \widetilde{\varepsilon}, \end{aligned} \tag{3}$$

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 (3) defines the FE estimator:

$$\widehat{\delta} = \left(\widetilde{W}'\widetilde{W}\right)^{-1}\widetilde{W}'\widetilde{y} = \left(W'M_D W\right)^{-1}W'M_D y. \tag{4}$$

2.2 Outline bootstrap algorithm

The bootstrap algorithm implemented in the `xtbcfe` routine to correct the bias of the FE estimator is an extended version of the approach presented in Everaert and Pozzi (2007). The underlying idea is that the FE estimator $\widehat{\delta}$ is biased but still an unknown function of the true parameter vector, i.e.

$$\mathrm{E}(\widehat{\delta}\,|\delta, \Sigma, T\,) = \int_{-\infty}^{+\infty} \widehat{\delta} f\left(\widehat{\delta}\,|\delta, \Sigma, T\right) d\widehat{\delta} \neq \delta, \tag{5}$$

with E being the expected value and f the probability distribution of $\widehat{\delta}$ for given population parameter vector δ, covariance matrix of the error terms Σ and sample size T. 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 (5) can be written as

$$\mathrm{E}(\widehat{\delta}\,|\delta, \Sigma, T) = \lim_{J\to\infty} \frac{1}{J} \sum_{j=1}^{J} \widehat{\delta}_j \,|\delta, \Sigma, T \ . \tag{6}$$

Equation (6) shows that an unbiased estimator for δ can be obtained as the value $\widehat{\delta}^{bc}$ that yields the FE to have a mean of $\widehat{\delta}$ over the J repeated samples. Formally, $\widehat{\delta}^{bc}$ is an unbiased estimator for δ if it satisfies

$$\widehat{\delta} = \lim_{J\to\infty} \frac{1}{J} \sum_{j=1}^{J} \widehat{\delta}_j \,\Big|\widehat{\delta}^{bc}, \Sigma, T \ . \tag{7}$$

The proposition in Everaert and Pozzi (2007) is that for any specific value of δ^*, the condition in equation (7) can be evaluated by generating J bootstrap samples from the data generating process in equation (2) 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 (7) is satisfied. Everaert and Pozzi further suggest that 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 ($\delta^*_{(0)} = \widehat{\delta}$). The iterative bootstrap bias-correction procedure is given by the following steps:

1. Using equation (3) 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 $j = 1, \ldots, J$:
   a. Draw a bootstrap sample ε^b from $\widehat{\varepsilon}$ according to a specified (re)sampling scheme.
   b. Calculate the bootstrap sample $y^b = W^b\delta^*_{(\kappa)} + \varepsilon^b$ where $W^b = (y^b_{-1}, \ldots, y^b_{-p}, \widetilde{X})$.
   c. Use FE to estimate $\widehat{\delta}^b_j = (W^{b'} M_D W^b)^{-1} W^{b'} M_D y^b$.
3. Calculate $\omega_{(k)} = \widehat{\delta} - \frac{1}{J}\sum_{j=1}^{J} \widehat{\delta}^b_j$.
4. Update the parameter vector $\delta^*_{(\kappa+1)} = \delta^*_{(\kappa)} + \omega_{(\kappa)}$.

In other words, step 3 evaluates to what degree the condition in equation (7) is satisfied when $\delta^*_{(\kappa)}$ is used to generate bootstrap samples, with $\omega_{(\kappa)}$ being the approximation error. When $\omega_{(\kappa)}$ is positive (negative), this means that $\delta^*_{(\kappa)}$ and the resulting average of the FE estimates $\frac{1}{J}\sum_{j=1}^{J} \widehat{\delta}^b_j$ over the bootstrap samples is too low (high) for equation

(7) to be satisfied. In step 4 we therefore update $\delta^*_{(\kappa)}$ by adding $\omega_{(\kappa)}$ to obtain $\delta^*_{(\kappa+1)}$ as a new guess for $\widehat{\delta}^{bc}$. The algorithm (steps 1-4) is then iterated until equation (7) is satisfied up to a tolerable degree, i.e. $\omega_{(\kappa)}$ is sufficiently close to zero.

In step 2 of the algorithm, it is crucial to obtain a sequence $\left(\widehat{\delta}^b_1, \ldots, \widehat{\delta}^b_J\right)$ that provides an adequate proxy for the bias of the FE estimator, i.e. the average of the FE estimates $\widehat{\delta}^b_j$ as calculated in equation (6) should be a good approximation to the integral in equation (5). To this end, the sampling of the bootstrap data should be consistent with the properties of the underlying data generating process. A few comments are in order.

First, 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. In fact, it is the centering itself that causes the bias as it induces correlation between $\widetilde{W}$ and $\widetilde{\varepsilon}$ in equation (3). As such, in contrast to Everaert and Pozzi (2007) who calculate bootstrap data from equation (2), we generate bootstrap samples $\left(y^b, W^b\right)$ using equation (3) in step 2(b) of the algorithm. In step 2(c), the use of the FE estimator then still implies centering of the data which causes its bias. The main practical advantage of this is that it simplifies the bootstrap algorithm as there is no need to estimate the individual effects α and use them to generate the data. The simplification is also favorable in terms of properties of the BCFE estimator (see Monte Carlo simulation results in section 5) as it avoids the nuisance induced by estimating α in combination with deterministic initialization.

Second, 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 Σ. Various (re)sampling schemes are discussed in section 2.3. Furthermore, the calculation of the bootstrap data y^b_{it} in step 2(b) requires initial values for $(y^b_{i,-(p-1)}, \ldots, y^b_{i0})$. The choice of how these initial values should be generated implicitly entails a decision about the initial conditions of the data. The possible initialization options are outlined in section 2.4.

2.3 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 in step 2(a). All of these rely in some way on the rescaled error terms $\widehat{\varepsilon}^r_{it}$

$$\widehat{\varepsilon}^r_{it} = \widehat{\varepsilon}_{it}\sqrt{\frac{NT}{NT-k-N}}, \tag{8}$$

where rescaling is necessary to correct for the fact that the estimated error terms $\widehat{\varepsilon}_{it}$, obtained in step 1 of the bootstrap algorithm outlined above, have a lower variance than the population error terms ε_{it}.

Parametric sampling schemes

In the parametric schemes, we draw ε_{it}^{b} from the i.i.d. $\mathcal{N}\left(0, \widehat{\sigma}_{it}^{2}\right)$ distribution, where we set $\sigma_{it}^{2} = \widehat{\sigma}_{i}^{2} = \frac{1}{T}\sum_{t=1}^{T}\left(\widehat{\varepsilon}_{it}^{r}\right)^{2}$ to allow for cross-sectional heteroscedasticity or $\sigma_{it}^{2} = \widehat{\sigma}_{t}^{2} = \frac{1}{N}\sum_{i=1}^{N}\left(\widehat{\varepsilon}_{it}^{r}\right)^{2}$ to allow for temporal heteroscedasticity. Under the assumption of homoscedasticity, we set $\widehat{\sigma}_{it}^{2} = \widehat{\sigma}^{2}$, which can then be calculated as $\widehat{\sigma}^{2} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\widehat{\varepsilon}_{it}^{r}\right)^{2}$. Note that the parametric schemes do not take into account general heteroscedastiticy (σ_{it}^{2}) or error CSD ($\sigma_{ijt} \neq 0$) as this would require specific assumptions about the functional form of these error structures.

Non-parametric resampling schemes

In the non-parametric schemes, ε_{it}^{b} is obtained by resampling the rescaled error terms $\widehat{\varepsilon}_{it}^{r}$. This has the advantage that it does not require distributional assumptions about ε_{it} while its covariance structure can be preserved by an appropriate design of the resampling scheme. In general notation, this implies setting $\varepsilon_{it}^{b} = \widehat{\varepsilon}_{j_{it},s_{it}}^{r}$ with j_{it} and s_{it} denoting 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 $(1, \dots, N)$ and the time index $(1, \dots, T)$ is aligned with the alleged covariance structure in ε_{it}. We allow for the following cases[1]:

1. Under homoscedasticity ($\sigma_{it}^{2} = \sigma^{2}$), $\widehat{\varepsilon}_{it}^{r}$ can be resampled both over cross-sections and time, i.e. j_{it} is drawn from $(1, \dots, N)$ and s_{it} from $(1, \dots, T)$.
2. Under pure cross-sectional heteroscedasticity ($\sigma_{it}^{2} = \sigma_{i}^{2}$), $\widehat{\varepsilon}_{it}^{r}$ can be resampled over time within cross-sections, i.e. s_{it} is drawn from $(1, \dots, T)$ while for j_{it} we consider two cases:
   a. When σ_{i}^{2} is random over cross-sections, we can draw entire cross-sections and resample over time within cross-sections, i.e. restrict $j_{it} = j_{i}$ which implies drawing a cross-section indicator j_{i} for each i from $(1, \dots, N)$ and using this in every time period t.
   b. When σ_{i}^{2} is cross-section specific, we can only resample over time within cross-sections, i.e. restrict $j_{it} = i$.
3. Under pure temporal heteroscedasticity ($\sigma_{it}^{2} = \sigma_{t}^{2}$), $\widehat{\varepsilon}_{it}^{r}$ can be resampled over cross-sections within time periods, i.e. j_{it} is drawn from $(1, \dots, N)$ while for s_{it} we consider two cases:
   a. When the temporal heteroscedasticity pattern is unconditional, we can draw entire time periods and resample over cross-sections within time periods, i.e. restrict $s_{it} = s_{t}$ which implies drawing a time indicator s_{t} for each t from $(1, \dots, T)$ and using this for every cross-section i.

1. Note that the downward bias of the FE estimator induces a serial correlation pattern in the estimated error terms $\widehat{\varepsilon}_{it}$ that is not present in the population error terms ε_{it}. As such, any resampling scheme should remove this spurious serial correlation pattern in the rescaled estimated error terms $\widehat{\varepsilon}_{it}^{r}$. This implies that we cannot resample blocks or entire cross-sections of these errors.

   b. When the temporal heteroscedasticity pattern is conditional, we can only resample over the cross-sectional dimension, i.e. restrict $s_{it} = t$.

4. Under general heteroscedasticity (σ_{it}^2), both the cross-sectional and the temporal structure of the error terms need to be preserved. To meet this challenge, we use the wild bootstrap suggested by Liu (1988) and Mammen (1993). The idea is to resample the residuals by multiplying them by a binomial random variable ι_{it} that is equal to -1 with probability 0.5 and equal to 1 with probability 0.5. We consider two cases:

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

   b. When the unconditional variance σ_i^2 is cross-section specific, we cannot resample over cross-sections and therefore apply a pure wild bootstrap, i.e. $\varepsilon_{it}^b = \iota_{it}\widehat{\varepsilon}_{it}^r$.

5. Under 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. We consider two cases:

   a. Under global CSD we can still resample over cross-sections within time periods. As such, both resampling schemes 3a and 3b can be used.

   b. Under local CSD we can only resample over time in the same way for each cross-section, i.e. we restrict $j_{it} = i$ as under 2b and $s_{it} = s_t$ as under 3a.

Each of the above resampling schemes has been generalized to unbalanced datasets, except for the randomized wild bootstrap (4a) and local CSD resampling (5b) which are only possible in balanced panels.

2.4 Initialization

As mentioned above, the calculation of the bootstrap data y_{it}^b in step 2(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 (see section 4) and tends to play an important role for the numerical properties of the algorithm in small datasets. Below we outline several possibilities that differ in the degree of randomness in generating the initial values. In section 9.1 of the appendix we provide some additional details about convergence and its relation to the initialization schemes.

Deterministic initialization

The fastest and most straightforward way of initializing the series y_{it}^b is by setting $(y_{i,-(p-1)}^b, \ldots, y_{i0}^b)$ equal to the observed (centered) initial values $(\widetilde{y}_{i,-(p-1)}, \ldots, \widetilde{y}_{i0})$ in each bootstrap sample. The advantage of this initialization is that we do not have to make assumptions about how the initial conditions are generated. In fact, this is the initialization used by Everaert and Pozzi (2007). Their Monte Carlo simulations show that it works well for both stationary and non-stationary initial conditions. Moreover, it has the practical advantage that one can avoid generating initial conditions when the data is not rich enough (see section 9.1). However, if initial conditions are random, a deterministic initialization has the obvious risk that it induces a spurious dependency over bootstrap samples, especially when the time series dimension is short. Therefore, we further extend the original bootstap procedure of Everaert and Pozzi (2007) by allowing for random initialization schemes. These assume that initial conditions are in the infinite past. They are outlined below.

Analytic inititialization

In the analytic initialization scheme, the initial observations are drawn from the multivariate normal distribution

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

where $\widehat{\mu}_i^0 = \widetilde{X}_{i0}\widehat{\beta}/(1 - \sum_{s=1}^p \widehat{\gamma}_s)$ is the unconditional expected value of y_{i0} for fixed values of the exogenous variables $\widetilde{X}_{i0}$ and the unconditional variance-covariance matrix $\widehat{\Sigma}_i^0$ is estimated as $\widehat{\Sigma}_i^0 = T^{-1}\sum_{t=1}^T z_{it}' z_{it}$ with $z_{it} = \left(y_{it}^*, \ldots, y_{i,t-p+1}^*\right)$ and where $y_{it}^* = \widetilde{y}_{it} - \widetilde{X}_{it}\widehat{\beta}/(1 - \sum_{s=1}^p \widehat{\gamma}_s)$ is the deviation of y_{it} from its unconditional mean. 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. As $\widehat{\Sigma}_i^0$ is estimated for each cross-section individually, this initialization takes into account cross-sectional heteroscedasticity. Under the assumption of homoscedasticity, $\widehat{\Sigma}_i^0$ can be replaced by $\widehat{\Sigma}^0 = N^{-1}\sum_{i=1}^N \widehat{\Sigma}_i^0$ which is the cross-sectional average of $\widehat{\Sigma}_i^0$.

Burn-in initialization

As an alternative to treating the initial observations as fixed or drawing them from the normal distribution, one may start in the distant past from initial values set to zero, e.g. $(y_{i,-50-p+1}^b = 0, \ldots, y_{i,-50}^b = 0)$, and then generate the series y_{il}^b, with $l = -49, \ldots, 0$, as in step 2(b) of the bootstrap algorithm, setting $\widetilde{X}_{il} = \widetilde{X}_{i0}$ and with bootstrap error terms obtained as in step 2(a). We can then simply use $(y_{i,-(p-1)}^b, \ldots, y_{i0}^b)$ as initial values and discard the earlier generated values. The advantage of this approach is that it does not require a distributional assumption for the initial conditions and that the error resampling scheme used to generate the actual sample can also be used to generate the initial values.

2.5 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.

The parametric approach makes use of the fact that 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. As such, the distribution of the BCFE estimator can be obtained by applying the bias-correction procedure to the J FE estimates $\widehat{\delta}^{b}_{j}$ obtained in step 2(c) of the iterative bootstrap procedure using $\delta^{*}_{(\kappa)} = \widehat{\delta}^{bc}$. The advantage of the parametric approach is that 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. As such, each of the above mentioned resampling and initialization schemes can be used.

In the non-parametric approach, as suggested by Kapetanios (2008), we resample the original data for cross-sectional units as a whole with replacement. The advantage of this resampling scheme is that it preserves the dynamic panel structure without the need to make parametric assumptions. Moreover, it is valid under general heteroscedasticity patterns and a global cross-sectional dependence structure in the data (induced for instance by a common factor structure). It is however not valid under local cross-sectional dependence (induced for instance by a spatial panel structure).

3 The xtbcfe routine

3.1 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`]

`xtbcfe` requires that the data are `xtset` before estimation. The program adds the lagged dependent variable(s) as the first explanatory variable(s) and can fit the simple autoregressive model without covariates. Cross-sections that are irregularly spaced along the time dimension are automatically reduced in size so that the largest block of uninterrupted observations are maintained (see Millimet and McDonough 2013). Cross-sections with too few (≤ 1) usable observations (after lagging) are removed. The `xtbcfe` routine requires that the `moremata`, `estout` (Jann 2005a,b) and `distinct` (Cox and

Longton 2008) packages are installed before use[2].

3.2 Options

`lags(#)` sets the number of lags, p, of the dependent variable to be included. The default is `lags(1)`.

`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
`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 both over cross-sections and time
`cshet`	for resampling within cross-sections (cross-sectional heteroscedasticity)
`cshet_r`	for resampling within cross-sections with randomized indices (random cross-sectional heteroscedasticity)
`thet`	for resampling within time periods (temporal heteroscedasticity)
`thet_r`	for resampling within time periods with permuted t (random temporal heteroscedasticity)
`wboot`	for wild bootstrap, i.e. error terms multiplied by '1' or '-1' (general heteroscedasticity)
`wboot_r`	for randomized wild bootstrap, i.e. permuted cross-section indices and error terms multiplied by '1' or '-1' (random general heteroscedasticity, balanced panels only)
`csd`	for resampling identically over cross-sections (cross-sectional dependence, balanced panels only)

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

2. These packages are easily installed by typing `ssc install moremata`, `ssc install estout` and `ssc install distinct`.

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

When the burn-in initialization is combined with the wild bootstrap (`wboot`), temporal heteroscedasticity (`thet`) or Monte Carlo temporal heteroscedasticity (`mcthe`) (re)sampling schemes, these become blocked variants. This implies that the resampling pattern used for the creation of bootstrapped data is copied (several times) to generate the initial values over the burn-in period.

`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(`*string*`)` specifies the type of standard errors and confidence intervals. Under the `inference(inf_se)` option, standard errors are bootstrapped and are then used to calculate confidence intervals using the Student t-distribution. Alternatively, as this distributional assumption may be violated, especially in small datasets with high temporal dependence, the `inference(inf_ci)` option calculates confidence intervals directly from the bootstrap distribution. This approach does not make a distributional assumption but is much more computationally intensive as, compared to calculating standard errors, adequate calculation of the desired percentiles requires more bootstrap samples. Finally, the `inference(inf_appr)` option is a fast alternative that approximates standard errors by calculating the dispersion of the FE estimator over the bootstrap iterations. While this is much faster than other options, the resulting standard errors are expected to be downward biased so that 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.

`infiters(`*#*`)` specifies the number of bootstrap iterations to be used for inference. The default is `infiters(250)` for all choices of `inference(`*string*`)`. It is recommended to have at least 50 iterations for bootstrapping standard errors and 1000 iterations for bootstrapping percentile intervals. The number of iterations cannot be smaller than 100 when the `inference(inf_ci)` option is selected.

`distribution(`*string*`)` requests that the bootstrap distribution of `xtbcfe` obtained by the inference procedures is 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. Select `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 autoregressive coefficients. The `distribution(all)` option adds histograms for all autoregressive coefficients separately.

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

`param` requests that inference procedures are initiated using the parametric bootstrap instead of the non-parametric 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.

Once all options are specified, the `xtbcfe` routine will remove any time invariant or collinear variables and move on to the main estimation `bcfe_ub` (mata) subroutine.

3.3 Saved results

`xtbcfe` saves the following results in `e()`:

Scalars			
`e(t_min)`	min number of time periods	`e(N)`	number of observations
`e(t_avg)`	average number of time periods	`e(N_g)`	number of groups
`e(t_max)`	max number of time periods	`e(k)`	number of exogenous regressors
`e(irr)`	number of cross-sections removed due to irregular spacing and/or lack of observations	`e(conv)`	convergence of the bootstrap algorithm
		`e(df_r)`	degrees of freedom
Macros			
`e(cmd)`	`xtbcfe`	`e(depvar)`	name of the dependent variable
`e(ivar)`	panel variable	`e(tvar)`	time variable
`e(predict)`	`xtbcfe_p`		
Matrices			
`e(b)`	`xtbcfe` estimates	`e(V)`	variance–covariance matrix of the `xtbcfe` estimator
`e(res_bcfe)`	`xtbcfe` error terms		
`e(dist_bcfe)`	`xtbcfe` bootstrap distribution if `dist()` option is selected		
Functions			
`e(sample)`	marks estimation sample		

3.4 Postestimation

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

<u>`predict`</u> [*type*] [*newvarname*] [*if*] [`, statistic`]

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

Unstarred statistics are available both in and out of sample, use `predict ... if e(sample) ...` for restricting statistics to the estimation sample. Starred statistics are calculated only for the estimation sample, even when `if e(sample)` is not specified.

4 Monte Carlo Experiments

Using Monte Carlo simulations, Everaert and 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. Furthermore, the BCFE is found to be insensitive to non-normality of the errors, conditional heteroscedasticity or non-stationary initial conditions and has a bias comparable to the analytical bias corrections of Kiviet (1995) and Bun and Carree (2005).

In this section we present some further Monte Carlo simulation results to illustrate the finite sample properties of our simplified BCFE bootstrap algorithm and its extension to higher-order dynamic models and error CSD. Data are generated from (1) with x_{it} restricted to be a single exogenous explanatory variable, generated as

$$x_{it} = \rho x_{i,t-1} + \xi_{it}, \qquad \xi_{it} \sim \text{i.i.d. } \mathcal{N}(0, \sigma_{\xi}^2). \tag{9}$$

We 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. We analyze the performance of alternative initialization schemes and adjust the bootstrap resampling scheme according to the properties of the data generating process of y_{it}. We report (i) mean bias (*bias*), which is the average of the deviation of the estimates $\widehat{\gamma}$ from the true population parameter γ, (ii) standard error (*se*), which is the standard error of the estimates $\widehat{\gamma}$, (iii) mean estimated standard error ($\widehat{se}$), which is the average of the estimated standard errors, and (iv) real size (*size*), which is the probability of incorrectly rejecting the correct null hypothesis using a two-sided t-test at the 5% nominal level of significance. 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. All simulations were performed using the Ghent University High Performance Computing infrastructure.

4.1 Simplification using the invariance principle

In Table 1 we compare the performance of the original algorithm (BCFE_{or}) of Everaert and Pozzi (2007) to our simplified algorithm presented in section 2. We use the high temporal dependence setting reported in their Table 2 as this is the case where the original BCFE estimator still exhibits some small sample bias.[3] This setting corresponds to generating y_{it} from a first-order ($p = 1$) version of equation (1), setting $\gamma_1 = 0.8$ and assuming $\alpha_i \sim$ i.i.d. $\mathcal{N}\left(0, (1-\gamma_1)^2\right)$ and $\varepsilon_{it} \sim$ i.i.d. $\mathcal{N}(0,1)$, with x_{it} generated from (9) setting $\rho = 0.5$ and assuming $\xi_{it} \sim$ i.i.d. $\mathcal{N}(0, 0.65)$. The BCFE estimator is implemented setting the number of bootstrap iterations (`bciters`) to 200. As there is no structure in the error terms, each of the BCFE estimators uses the `iid` resampling scheme. We further use this simulation design to shed some light on the relative performance of the various initialization schemes and alternative approaches to inference. As such, we report results for three alternative initialization schemes: `det` (BCFE_{de}), `aho` (BCFE_{an}) and `bi` (BCFE_{bi}). Next, approximate standard errors ($\widehat{se}_{appr}$) obtained from the bootstrapped distribution of the FE estimator using the `inf_appr` option are compared to standard errors from the bootstrapped distribution of the BCFE estimator using the `inf_se` option. For the first option, which is relatively fast, we set the number of bootstrap iterations (`infiters`) to 1000. For the computationally more intensive second option, we analyze the importance of the number of iterations by reporting results for 1000 ($\widehat{se}_{1000}$) and 50 ($\widehat{se}_{50}$) iterations. Finally, we calculate real test sizes using the above obtained approximate FE ($size_{appr}$) and BCFE ($size_{se}$) standard errors and the bootstrapped percentile confidence interval using the `inf_ci` option ($size_{ci}$). These are all based on 1000 bootstrap iterations. As estimates for β are more or less unbiased for all estimators, we only report results for estimating γ.

The simulation results show that our simplified algorithm yields a considerable improvement over the original estimator. Under the analytical and burn-in initialization schemes, the BCFE estimator is nearly unbiased and bootstrapped standard errors ($\widehat{se}_{1000}$) are close to the true standard error. As a result, these versions of the BCFE estimator have a more or less correct real size, even for very small T. Under the deterministic initialization, a small bias remains for $T = 4$. Overall we consider the BCFE initiated with the burn-in initiation the superior alternative because of its low bias and adequate results in terms of inference. Each of the BCFE variants also displays adequate convergence rates with 100% convergence for the deterministic initiation and 97.9% and 98.7% for the analytical and burn-in initiations respectively in the $N = 20$, $T = 4$ case. Any other sample size resulted in 100% convergence for all initiations.

Given that the standard errors based on 1000 bootstrap iterations in the $\widehat{se}_{1000}$ column are computationally very intensive, an assessment of the performance of the less time consuming alternatives is of particular interest for practitioners. As expected, the approximated standard errors $\widehat{se}_{appr}$ have a downward bias that has a detrimental effect on the real test size. The $\widehat{se}_{50}$ column however reveals that on average, the difference between using 1000 and 50 bootstrap iterations for computing standard errors is only marginal. This suggests that 50 bootstrap iterations is a reasonable lower bound for

3. Similar results are obtained for other parameter values.

Table 1: Monte Carlo results for an $AR(1)$ model with $\gamma_1 = 0.8$: simplification bootstrap algorithm

	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	-	-
$\mathrm{BCFE_{an}}$	−0.18	0.17	0.16	-	-	0.21	-	-	−0.05	0.08	0.08	-	-	0.09	-	-
$\mathrm{BCFE_{or}}$	−0.14	0.15	-	-	-	-	-	0.25	−0.04	0.09	-	-	-	-	-	0.11
$\mathrm{BCFE_{de}}$	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
$\mathrm{BCFE_{an}}$	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
$\mathrm{BCFE_{bi}}$	−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	-	-
$\mathrm{BCFE_{an}}$	−0.13	0.08	0.09	-	-	0.30	-	-	−0.03	0.04	0.04	-	-	0.14	-	-
$\mathrm{BCFE_{or}}$	−0.13	0.07	-	-	-	-	-	0.80	−0.04	0.04	-	-	-	-	-	0.35
$\mathrm{BCFE_{de}}$	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
$\mathrm{BCFE_{an}}$	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
$\mathrm{BCFE_{bi}}$	−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

Notes:

(i) Data for y_{it} are generated from a first-order ($p = 1$) version of equation (1), setting $\gamma_1 = 0.8$, $\beta = 0.2$ and assuming $\alpha_i \sim$ i.i.d. $\mathcal{N}\left(0, (1-\gamma_1)^2\right)$ and $\varepsilon_{it} \sim$ i.i.d. $\mathcal{N}(0,1)$, with x_{it} generated from (9) setting $\rho = 0.5$ and assuming $\xi_{it} \sim$ i.i.d. $\mathcal{N}(0, 0.65)$. Note that as we assume y_{i0} to be observed, the sample sizes $T = 5$ and $T = 10$ in Everaert and Pozzi (2007) correspond to $T = 4$ and $T = 9$ in our notation.

(ii) Reported results are for estimating γ_1. POLS and FE refer to the Pooled OLS and Fixed Effects estimator respectively. $\mathrm{BCFE_{an}}$ is the analytical bias-corrected FE estimator implemented in the `xtlsdvc` routine developed by Bruno (2005) initiated with the Anderson-Hsiao estimator and standard errors obtained through 200 bootstrap iterations. $\mathrm{BCFE_{or}}$ is the original bootstrap-based bias-corrected FE estimator of Everaert and Pozzi (2007). Results are taken from their Table 2. $\mathrm{BCFE_{de}}$, $\mathrm{BCFE_{an}}$ and $\mathrm{BCFE_{bi}}$ refer to the simplified BCFE estimator presented in section 2, with 200 bootstrap samples (`bciters`) and the deterministic (`det`), homogeneous analytical (`aho`) and burn-in (`bi`) initialization, respectively. Each of the BCFE estimators uses the `iid` resampling scheme.

(iii) The *bias* is the deviation of the estimates $\widehat{\gamma}_1$ from the population parameter γ_1 while *se* is the standard error of the distribution of $\widehat{\gamma}_1$ over the Monte Carlo draws. The estimated standard errors $\widehat{se}$ are obtained in 3 different ways: $\widehat{se}_{appr}$ are approximate standard errors based on the bootstrapped FE distribution (`inf_appr`), while $\widehat{se}_{1000}$ and $\widehat{se}_{50}$ are bootstrapped standard errors (`inf_se`) using respectively 1000 or 50 bootstrap iterations (`infiters`). The real size (*size*) is the probability of incorrectly rejecting the correct null hypothesis using a two-sided t-test at the 5% nominal level of significance. The sizes reported as $size_{appr}$ and $size_{se}$ are calculated using the standard errors $\widehat{se}_{appr}$ and $\widehat{se}_{1000}$ respectively, while for $size_{ci}$ the bootstrap percentile interval `inf_ci` option is used. Standard errors for the $\mathrm{BCFE_{an}}$ estimator are obtained using a bootstrap with 200 iterations (`infiters`). The standard analytical formulas are used for calculating the standard errors of the POLS and FE estimators.

standard error estimation.

4.2 Error CSD

In Table 2 we analyze the small sample performance of the BCFE estimator in a non-standard scenario with cross-sectionally dependent errors. To this end, we focus on a pure ($\beta = 0$) first-order autoregressive model with $\gamma_1 = 0.8$ and assume that the error term ε_{it} in equation (1) has the following common factor structure

$$\varepsilon_{it} = \lambda_i F_t + \epsilon_{it}, \tag{10}$$

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 and 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. We use the burn-in (`bi`) initiation for the BCFE estimator together with the cross sectional dependence `csd` ($\mathrm{BCFE_{csd}}$) and randomized temporal heteroscedasticity `thet_r` ($\mathrm{BCFE_{thet}}$) resampling schemes. Although `csd` resampling allows for a more general CSD pattern than `thet_r` resampling, both are valid given the common factor structure in equation (10). Next to bias and standard errors, we report real test sizes from t-tests ($size_t$) and confidence intervals ($size_{ci}$) based on 200 iterations (`infiters`). The root mean squared error ($rmse$) is provided as a performance measure that takes both bias and variance into account.

Table 2: Monte Carlo results for an $AR(1)$ model with $\gamma_1 = 0.8$: error CSD

	$bias$	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	-
$\mathrm{BCFE_{an}}$	−0.169	0.148	0.134	0.225	0.25	-	−0.053	0.087	0.077	0.102	0.13	-
$\mathrm{BCFE_{csd}}$	0.041	0.159	0.143	0.165	0.17	0.09	0.019	0.100	0.084	0.102	0.15	0.10
$\mathrm{BCFE_{thet}}$	−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	-
$\mathrm{BCFE_{an}}$	−0.105	0.076	0.074	0.130	0.30	-	−0.029	0.044	0.038	0.053	0.16	-
$\mathrm{BCFE_{csd}}$	0.070	0.086	0.078	0.111	0.24	0.13	0.022	0.047	0.044	0.052	0.09	0.07
$\mathrm{BCFE_{thet}}$	−0.003	0.085	0.080	0.085	0.09	0.07	−0.001	0.044	0.042	0.044	0.07	0.08

Notes:

(i) Data for y_{it} are generated from a first-order ($p = 1$) version of model (1) with $\gamma_1 = 0.8$, $\beta = 0$ and errors generated from the common factor structure in equation (10). We generate loadings as $\lambda_i \sim$i.i.d.$\mathcal{U}(1, 4)$ and set $\sigma_\alpha^2 = (1-\gamma_1)(1+\gamma_1)^{-1}(\mu_\lambda^2 + \sigma_\lambda^2 + 1)$.

(ii) Reported results are for estimating γ_1. POLS and FE refer to the Pooled OLS and Fixed Effects estimator respectively. $\mathrm{BCFE_{an}}$ is the analytical bias-corrected FE estimator implemented in the `xtlsdvc` routine developed by Bruno (2005) initiated with the Anderson-Hsiao estimator and standard errors obtained through 200 bootstrap iterations. $\mathrm{BCFE_{csd}}$ refers to the bootstrap-based bias-corrected FE estimator presented in section 2, with 250 bootstrap samples (`bciters`), burn-in (`bi`) initialization and the `csd` resampling scheme. $\mathrm{BCFE_{thet}}$ is the alternative that uses the random temporal heteroscedasticity (`thet_r`) scheme.

(iii) The *bias* is the deviation of the estimates $\widehat{\gamma}_1$ from the population parameter γ_1, *se* the standard error of the distribution of $\widehat{\gamma}_1$ over the Monte Carlo draws and $rmse = \sqrt{(E(\widehat{\gamma}_1) - \gamma_1)^2 + \sigma^2_{\widehat{\gamma}_1}}$ is the root mean squared error. The estimated standard errors $\widehat{se}$ are obtained using the non-parametric bootstrap resampling scheme with 200 iterations (`infiters`). The real size (*size*) is the probability of incorrectly rejecting the correct null hypothesis using a two-sided test at the 5% nominal level of significance, with $size_t$ based on t-statistics with estimated standard errors using the `inf_se` option and $size_{ci}$ based on the confidence interval (`inf_ci`) option.

The results reveal a general deterioration of the estimated standard errors compared to the real standard errors for all estimators. BCFE standard errors suffer as well but compared to the other estimators this does not result in large size distortions. Note however that the size of the confidence intervals approach is better than that of the t-test approach, especially for the `csd` resampling option in smaller sample sizes. This is due to the skewness of the BCFE distribution caused by the relatively large value of γ_1. The resulting asymmetry renders normal approximations very inaccurate and leads to size distortions. The confidence interval approach, in contrast, is not based on any distributional assumption and therefore has a more appropriate size.

There is also a clear difference in performance between our two alternative resampling schemes. Bias and real size for `thet_r` resampling are generally superior to `csd` resampling. This can be explained by the fact that under the `thet_r` option, errors are resampled both over time and cross-sections (within time periods). Especially when the cross-sectional dimension N is large, this results in more randomness in the bootstrap samples compared to the `csd` scheme which only resamples over time. In small T datasets, the latter offers only a very limited number of reshuffling options and therefore induces a dependency over bootstrap samples that leads to bias and an increased real test size. These results suggest that researchers should in practice opt for the most random resampling scheme among the appropriate alternatives.

4.3 Second-order dynamic model

In Tables 3 and 4 we analyze the small sample performance of the BCFE estimator in a second-order ($p = 2$) version of model (1). We 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 (9) setting $\rho = 0.5$ and assuming $\xi_{it} \sim$ i.i.d. $\mathcal{N}(0,1)$. We report results for estimating γ_1 and γ_2. The BCFE estimator is implemented with `iid` resampling, burn-in (`bi`) initiation and inference using bootstrapped standard errors (`inf_se`) and t-tests based on non-parametric bootstrapping. Note that the BCFE_{an} is not included because the `xtlsdvc` routine does not support higher-order models.

Table 3 reports results for a series with strong temporal dependence, setting $\gamma_1 = 0.6$ and $\gamma_2 = 0.2$. The BCFE estimator again appears as a very effective correction for FE. Its bias is virtually zero at the cost of only a small increase in variance. Standard errors are estimated well and the resulting real test size is near the nominal 5% level. In line with results from a first-order model, the standard pooled OLS estimator has a small but positive bias for both γ_1 and γ_2 for every combination of N and T while the FE estimator is strongly downward biased for both γ_1 and γ_2 for small T. This suggests that in this setting, an unbiased estimator is expected to lie between POLS and FE, but probably closer to the former than to the latter.

In Table 4 we set γ_1 to 1.1 but maintain the stationarity assumption by setting γ_2 to -0.2. The hump-shaped pattern implied by this parameter combination is often encountered in practice (see e.g. the application in section 5) but seldom included in simulation studies. The BCFE estimator is again almost unbiased in all settings with real test sizes close to the desired nominal level. In line with the results in Table 3, the POLS estimator has a small upward bias for both γ_1 and γ_2. For the FE we note an important difference, though. While the FE estimator for γ_1 is still strongly downward biased, it is much less biased for γ_2. In this setting, an unbiased estimator is expected to lie closer to the POLS estimator for γ_1 but closer to the FE estimator for γ_2.

5 Application

In this section, we illustrate the use of the `xtbcfe` routine by reporting some estimation results for labour demand by U.K. firms using the Arellano and Bond (1991) dataset (`abdata.dta`). This has become a prominent example in dynamic panel data modeling as labour demand is known to react very slowly to movements in its explanatory variables due to for instance considerable adjustment costs. Typically, lags of the depen-

Table 3: Monte Carlo results for an $AR(2)$ model with $\gamma_1 = 0.6$ 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

Notes:

(i) Data for y_{it} are generated from a second-order ($p = 2$) version of equation (1), setting $\gamma_1 = 0.6$, $\gamma_2 = 0.2$, $\beta = 0.2$ and assuming $\alpha_i \sim$ i.i.d. $\mathcal{N}(0,1)$ and $\varepsilon_{it} \sim$ i.i.d. $\mathcal{N}(0,1)$, with x_{it} generated from (9) setting $\rho = 0.5$ and assuming $\xi_{it} \sim$ i.i.d. $\mathcal{N}(0,1)$.

(ii) Reported results are for estimating γ_1 and γ_2. POLS and FE refer to the Pooled OLS and Fixed Effects estimator respectively. BCFE refers to the bootstrap-based bias-corrected FE estimator presented in section 2, with 250 bootstrap samples (`bciters`), burn-in (`bi`) initialization and `iid` resampling scheme.

(iii) The *bias* is the deviation of the estimates $\widehat{\gamma}$ from the population parameter γ while *se* is the standard error of the distribution of $\widehat{\gamma}$ over the Monte Carlo draws. The estimated standard errors $\widehat{se}$ are obtained using the non-parametric resampling scheme with 200 iterations (`infiters`). The real size ($size_t$) is the probability of incorrectly rejecting the correct null hypothesis using a two-sided t-test with estimated standard errors (`inf_se`) at the 5% nominal level of significance.

Table 4: Monte Carlo results for an $AR(2)$ model with $\gamma_1 = 1.1$ 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

Notes:

(i) Data for y_{it} are generated from a second-order ($p = 2$) version of equation (1), setting $\gamma_1 = 1.1$, $\gamma_2 = -0.2$, $\beta = 0.1$ and assuming $\alpha_i \sim$ i.i.d. $\mathcal{N}(0,1)$ and $\varepsilon_{it} \sim$ i.i.d. $\mathcal{N}(0,1)$, with x_{it} generated from (9) setting $\rho = 0.5$ and assuming $\xi_{it} \sim$ i.i.d. $\mathcal{N}(0,1)$.

(ii) Reported results are for estimating γ_1 and γ_2. POLS and FE refer to the Pooled OLS and Fixed Effects estimator respectively. BCFE refers to the bootstrap-based bias-corrected FE estimator presented in section 2, with 250 bootstrap samples (`bciters`), burn-in (`bi`) initialization and `iid` resampling scheme.

(iii) The *bias* is the deviation of the estimates $\widehat{\gamma}$ from the population parameter γ while *se* is the standard error of the distribution of $\widehat{\gamma}$ over the Monte Carlo draws. The estimated standard errors $\widehat{se}$ are obtained using the non-parametric resampling scheme with 200 iterations (`infiters`). The real size ($size_t$) is the probability of incorrectly rejecting the correct null hypothesis using a two-sided t-test with estimated standard errors (`inf_se`) at the 5% nominal level of significance.

dent variable are added to the explanatory variables to capture this adjustment process. However, as the dataset has a moderately large cross-section (140 U.K. companies) but only a relatively short time series dimension (max 9 observations between 1976 and 1984), the standard FE estimator is expected to be strongly downward biased. As such, Arellano and Bond (1991) use this example to advocate the use of their GMM estimator as an alternative to FE. They suggest two lags of log employment (n_{it}) and further model the dynamics by adding a single lag for log wages (w_{it}) and two lags for the logs of industry output (ys_{it}) and capital (k_{it}). This yields the following specification

$$n_{it} = \sum_{s=1}^{2} \gamma_s n_{i,t-s} + \sum_{q=0}^{1} \beta_{w,q} w_{i,t-q} + \sum_{r=0}^{2} (\beta_{k,r} k_{i,t-r} + \beta_{ys,r} ys_{i,t-r}) + \alpha_i + \lambda_t + \varepsilon_{it}, \quad (11)$$

where α_i is included to capture individual effects and λ_t is a time dummy that serves to capture aggregate demand shocks. The data is mildly unbalanced with a minimum of 7 observations (prior to lagging) and no gaps.

Table 5 reports estimation results for the POLS, FE, difference GMM (dGMM), system GMM (sGMM) and BCFE estimators. Looking first at the POLS and FE estimates, the coefficient on the first lag is much bigger for the POLS than for the FE estimator, although still relatively high for the latter. The coefficient on the second lag is small and negative for both estimators. For the POLS estimator it is clearly not significantly different from zero, while for FE it is somewhat more negative and significant at the 7% level of significance. This pattern is in line with expectations as in general, the POLS estimator is expected to be upward biased, as not accounting for individual effects implies positive correlation between the error terms and the lagged dependent variable, while the FE estimator is expected to be biased downwards, as the centering used to wipe out the individual effects results in negative correlation between the centered lagged dependent variables and the error terms. The simulation results in section 4.3 show, more specifically, that in a second-order dynamic model like equation (11), the POLS estimator has a more or less equal small upward bias for the coefficients on the first and second lag while the FE has a strong downward bias for the coefficient on the first lag but is much less biased for the coefficient on the second lag. Hence an unbiased estimate for the coefficient on $n_{i,t-1}$ is expected to lie somewhere in between the FE estimate 0.734 and the POLS estimate 1.045, but closer to the POLS than to FE estimate. An unbiased estimate for the coefficient on $n_{i,t-2}$ is expected to lie close to the FE estimate of -0.141 and below the POLS estimate of -0.077. Note that this only holds in expectations; sampling error can still imply that an unbiased estimator results in estimates that are outside the POLS-FE bounds in a specific sample. This risk is more pronounced in a higher order dynamic model as the different lags of the dependent variables are typically highly correlated. This multicollinearity problem tends to increase the variance of the estimates. Therefore, we also report the sum of the autoregressive coefficients ($\gamma_1 + \gamma_2$) as a rough measure of overall temporal dependence. For the POLS estimator the sum is 0.97, which is close to non-stationarity, while for the FE estimator this is much lower at 0.59.

Looking at the GMM results, the dGMM estimator behaves rather poorly as the coefficient estimate of 0.686 on $n_{i,t-1}$ is even lower than the downward biased FE estimate while the coefficient estimate of -0.085 on $n_{i,t-2}$ is close to the upward biased POLS estimate. The sum of the dGMM AR term estimates equals 0.601, which is highly similar to the sum of 0.593 implied by the downward biased FE estimates. This suggests that

Table 5: Estimated employment equations: full sample

Dependent variable: n_{it}										Sample period: 1976-1984, 140 U.K. firms
	POLS		FE		dGMM		sGMM		BCFE	
$n_{i,t-1}$	1.045	(0.051)	0.734	(0.058)	0.686	(0.145)	0.914	(0.127)	1.008	(0.057)
$n_{i,t-2}$	−0.077	(0.048)	−0.141	(0.077)	−0.085	(0.056)	−0.068	(0.055)	−0.161	(0.069)
w_{it}	−0.524	(0.172)	−0.557	(0.155)	−0.608	(0.178)	−0.652	(0.182)	−0.560	(0.163)
$w_{i,t-1}$	0.477	(0.169)	0.326	(0.143)	0.393	(0.168)	0.524	(0.168)	0.495	(0.192)
k_{it}	0.343	(0.048)	0.385	(0.056)	0.357	(0.059)	0.341	(0.062)	0.385	(0.051)
$k_{i,t-1}$	−0.202	(0.064)	−0.084	(0.053)	−0.058	(0.073)	−0.148	(0.075)	−0.202	(0.060)
$k_{i,t-2}$	−0.116	(0.035)	−0.025	(0.042)	−0.020	(0.033)	−0.059	(0.039)	−0.053	(0.037)
ys_{it}	0.433	(0.176)	0.521	(0.193)	0.608	(0.172)	0.660	(0.178)	0.455	(0.178)
$ys_{i,t-1}$	−0.768	(0.248)	−0.659	(0.208)	−0.711	(0.232)	−0.836	(0.234)	−0.746	(0.271)
$ys_{i,t-2}$	0.312	(0.130)	0.001	(0.139)	0.106	(0.141)	0.111	(0.158)	0.133	(0.171)
No. of obs	751		751		611		751		751	
Sum AR	0.968	(0.007)	0.593	(0.067)	0.601	(0.125)	0.846	(0.100)	0.847	(0.050)

Notes:

(i) Pooled OLS (POLS), Fixed Effects (FE) and difference GMM (dGMM) estimates are taken from Arellano and Bond (1991). System GMM (sGMM) estimates are obtained using the `xtdpdsys` Stata routine with the `vce(robust)` option to calculate standard errors.

(ii) The instrument sets used by the GMM estimators are constructed under the assumption that all regressors, except the lagged dependent variables, are strictly exogenous. The reported GMM estimates are one-step results.

(iii) The BCFE estimator uses 250 bootstrap samples (`bciters`) with a burn-in (`bi`) initialization and the wild bootstrap (`wboot`) to allow for general heteroskedasticity.

(iv) Estimated standard errors are reported in brackets. They are robust to general cross-section and time-series heteroskedasticity. For the BCFE they are calculated using 50 bootstrap iterations (`infiters`).

(v) Sum AR is the sum of the estimated AR coefficients $\widehat{\gamma}_1$ and $\widehat{\gamma}_2$.

(vi) Time dummies are included in every specification but not reported.

the temporal dependence implied by the dGMM estimates is also downward biased. Moreover, the standard deviation of the dGMM estimator is much bigger than that of the FE estimator, especially for the coefficient on $n_{i,t-1}$. The sGMM estimator improves on these results as the coefficient of 0.914 on $n_{i,t-1}$ is now between the POLS and FE estimates, as is expected for an unbiased estimator. Moreover, the overall temporal dependence of 0.846 is higher than that implied by the downward biased FE estimates. However, the coefficient of -0.068 on $n_{i,t-2}$ is now even higher than the upward biased POLS estimate and statistically not significantly different from zero. Furthermore, the standard deviation has decreased a little but is still much higher than that of the FE estimator.

The last column in Table 5 reports BCFE estimates. Since firms operating in different industries may have considerably different error variances, which are also likely to change over time, we use the wild bootstrap (`wboot`) resampling scheme as this is robust to heteroskedasticity. We further use the burn-in (`bi`) initialization, which is the most flexible approach. The Stata commands to obtain the BCFE results and full estimation output are reported in section 9.2 of the Appendix. This appendix also contains estimation results for alternative resampling schemes. Turning to the estimation results, in line with what we expect, the coefficient of 1.008 on $n_{i,t-1}$ is closer to the POLS estimate than to the FE estimate while the coefficient of -0.161 on $n_{i,t-2}$ is close to the FE estimate. Moreover, the standard errors of the BCFE are close to that of the FE estimator and much lower than that of the GMM estimators, especially for the coefficient on $n_{i,t-1}$. Also note that although its standard error is higher than that of the sGMM estimator, $n_{i,t-2}$ now even shows up as significantly negative at the 5% level of significance.

As the error terms are potentially correlated over cross-sections, Table 6 reports BCFE estimates using the bootstrap resampling scheme `csd`, which reshuffles error terms using the same time index for each cross-section in order to preserve a general type of contemporaneous error CSD. It also takes into account cross-sectional and unconditional temporal heteroskedasticity. As this resampling scheme requires a balanced panel, we take a subset of the original data which includes 80 firms over the period 1978-1982. As a benchmark, Table 6 also contains the POLS and FE estimates for this reduced dataset. The BCFE estimate of 1.179 on $n_{i,t-1}$ is now somewhat above the POLS estimate of 1.104 while the BCFE estimate of -0.319 on $n_{i,t-2}$ is now slightly below the FE estimate of -0.229. However, the overall temporal dependence, as measured by the sum of the AR coefficients, of 0.86 for the BCFE estimator is still in the range [0.535,0.974] implied by the POLS and FE estimates.

Table 6: Estimated employment equations: balanced panel

Dependent variable: n_{it} Sample period: 1978-1982, 80 U.K. firms

	POLS		FE		BCFE	
$n_{i,t-1}$	1.104	(0.048)	0.764	(0.048)	1.179	(0.058)
$n_{i,t-2}$	−0.130	(0.047)	−0.229	(0.064)	−0.319	(0.063)
w_{it}	−0.087	(0.084)	−0.108	(0.116)	−0.107	(0.125)
$w_{i,t-1}$	0.049	(0.088)	−0.021	(0.120)	0.049	(0.169)
k_{it}	0.326	(0.044)	0.376	(0.054)	0.383	(0.058)
$k_{i,t-1}$	−0.221	(0.059)	−0.090	(0.054)	−0.269	(0.075)
$k_{i,t-2}$	−0.083	(0.036)	0.001	(0.043)	−0.015	(0.036)
ys_{it}	0.095	(0.187)	0.034	(0.204)	0.034	(0.228)
$ys_{i,t-1}$	−0.385	(0.208)	−0.326	(0.194)	−0.375	(0.284)
$ys_{i,t-2}$	0.257	(0.123)	0.305	(0.176)	0.417	(0.220)
Sum AR	0.974	(0.009)	0.535	(0.070)	0.860	(0.046)

Notes:

(i) Pooled OLS (POLS) and Fixed Effects (FE) estimates are obtained using the Stata routines `regress` and `xtreg` respectively with the `vce(robust)` option to calculate standard errors.

(ii) The BCFE estimator uses 250 bootstrap samples (`bciters`) with a burn-in (`bi`) initialization and the `csd` resampling option to allow for general error CSD and cross-sectional as well as unconditional temporal heteroskedasticity.

(iii) Estimated standard errors are reported in brackets. They are robust to general cross-section and time-series heteroskedasticity. For the BCFE they are calculated using 50 bootstrap iterations (`infiters`).

(iv) Sum AR is the sum of the estimated AR coefficients $\widehat{\gamma}_1$ and $\widehat{\gamma}_2$.

(v) Time dummies are included in every specification but not reported.

6 Conclusion

This paper has described a new Stata routine, `xtbcfe`, that executes an iterative bootstrap-based bias-corrected fixed effects estimator for dynamic panels building on Everaert and Pozzi (2007). We first simplify the core of their algorithm using the invariance principle and next extend it to allow for unbalanced and higher-order dynamic panels. We implement various bootstrap error resampling schemes to account for general heteroscedasticity and contemporaneous cross-sectional dependence and include several options for the initial conditions. The choice of an appropriate resampling scheme is important to preserve the structure of the error terms in the resampling process. Several resampling options will often be applicable in practice but tend to imply a different dependency over bootstrap iterations in small datasets. As the `xtbcfe` algorithm performs better when the generated samples are independent, researchers are advised to choose the alternative that incorporates the highest degree of randomness in the resampling process.

Inference can be carried out using either parametric or non-parametric bootstrapped variance-covariance matrices or percentile intervals. The latter have the advantage of not making any distributional assumptions and may be more suited in smaller datasets. Monte Carlo simulations show that the simplification of the original algorithm results in a BCFE estimator that is virtually unbiased for very small T. The Monte Carlo results also support the BCFE in higher order dynamic panels and panels with contemporaneous error CSD.

Future extensions of the code will include allowing for predetermined and endogenous covariates and for intertemporal cross-sectional dependence.

7 Acknowledgements

The authors thank the editor, an anonymous referee and Yannick Thuy for useful comments and suggestions. The computational resources (Stevin Supercomputer Infrastructure) and services used in this work were provided by the VSC (Flemish Supercomputer Center), funded by Ghent University, the Hercules Foundation and the EWI department of the Flemish Government. Ignace De Vos and Ilse Ruyssen gratefully acknowledge financial support from the Ghent University BOF research fund. Ignace De Vos and Gerdie Everaert further acknowledge financial support from the National Bank of Belgium mecenas project.

8 References

Anselin, L. 1988. *Spatial Econometrics*. Dordrecht: Kluwer.

Arellano, M., and S. 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.

Arellano, M., and O. Bover. 1995. Another look at the Instrumental Variable estimation of Error-Components Models. *Journal of Econometrics* 68: 29–51.

Blundell, R., and S. Bond. 1998. Initial Conditions and Moment Restrictions in Dynamic Panel Data Models. *Journal of Econometrics* 87: 115–143.

Bruno, G. S. F. 2005. Estimation and inference in dynamic unbalanced panel data models with a small number of individuals. *The Stata Journal* 5(4): 473–500.

Bun, M. 2003. Bias correction in the dynamic panel data model with a nonscalar disturbance covariance matrix. *Econometric Reviews* 22: 29–58.

Bun, M., and M. Carree. 2005. Bias-Corrected Estimation in Dynamic Panel Data Models. *Journal of Business and Economic Statistics* 23(2): 200–210.

———. 2006. Bias-Corrected Estimation in Dynamic Panel Data Models with Heteroscedasticity. *Economics Letters, 2006* 92(2): 220–227.

Bun, M., and J. Kiviet. 2006. The effects of dynamic feedbacks on LS and MM estimator accuracy in panel data models. *Journal of Econometrics* 132(2): 409–444.

Bun, M. J. G., and F. Windmeijer. 2010. The Weak Instrument Problem of the System GMM Estimator in Dynamic Panel Data Models. *Econometrics Journal* 13(1): 95–126.

Coakley, J., A. Fuertes, and R. Smith. 2002. A Principal Components Approach to Cross-Section Dependence in Panels. In *10th International Conference on Panel Data, Berlin, July 5-6, number B5-3.*

Cox, N. J., and G. M. Longton. 2008. Speaking Stata: Distinct observations. *Stata Journal* 8: 557–568.

Everaert, G., and T. De Groote. 2014. Common Correlated Effects Estimation of Dynamic Panels with Cross-Sectional Dependence. *Econometric Reviews* .

Everaert, G., and L. Pozzi. 2007. Bootstrap-based bias correction for dynamic panels. *Journal of Economic Dynamics and Control* 31(4): 1160–1184.

Jann, B. 2005a. moremata: Stata module (Mata) to provide various functions. Available from http://ideas.repec.org/c/boc/bocode/s455001.html.

———. 2005b. Making regression tables from stored estimates. *Stata Journal* 5: 288–308.

Kapetanios, G. 2008. A bootstrap procedure for panel data sets with many cross-sectional units. *Econometrics Journal* 11(2): 377–395.

Kapoor, H. P. I., M; Kelejian. 2007. Panel data models with spatially correlated error components. *Journal of econometrics* 140: 97–130.

Kiviet, J. 1995. On Bias, Inconsistency, and Efficiency of Various Estimators in Dynamic Panel Data Models. *Journal of Econometrics* 68: 53–78.

Liu, R. 1988. Bootstrap procedures under some non-i.i.d. Models. *Annals of Statistics* 16: 1696–1708.

Mammen, E. 1993. Bootstrap and Wild Bootstrap for High Dimensional Models. *Annals of Statistics* 21: 255–285.

Millimet, D. L., and I. K. McDonough. 2013. Dynamic Panel Data Models with Irregular Spacing: With Applications to Early Childhood Development. Iza discussion papers 7359, Institute for the Study of Labor (IZA).

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

Pesaran, M. 2006. Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure. *Econometrica* 74(4): 967–1012.

Phillips, P. C. B., and D. Sul. 2007. Bias in Dynamic Panel Estimation with Fixed Effects, Incidental Trends and Cross Section Dependence. *Journal of Econometrics* 137(1): 162–188.

Roodman, D. 2009. A Note on the Theme of Too Many Instruments. *Oxford Bulletin of Economics and Statistics* 71(1): 135–158.

Sarafidis, V., and D. Robertson. 2009. On the Impact of Error Cross-Sectional Dependence in Short Dynamic Panel Estimation. *Econometrics Journal* 12(1): 62–81.

Stock, J. H., and M. W. Watson. 2002. Macroeconomic forecasting using diffusion indexes. *Journal of Business and Economic Statistics* 20: 147–162.

Ziliak, J. 1997. Efficient Estimation with Panel Data when Instruments are Predetermined: an Empirical Comparison of Moment-Condition Estimators. *Journal of Business and Economic Statistics* 14(4): 419–431.

9 Appendix

9.1 Convergence, initiation and non-stationarity

In this subsection we provide some additional technical details regarding the `xtbcfe` routine. As this is an iterative bias-correction procedure, an important issue is that of convergence. When evaluating equation (7), the convergence criterion is by default set to 0.005. Point estimates emerging from a divergent estimator will, in general, not have appropriate statistical properties and hence are not reliable for inference (the routine will therefore not initiate the inference sequence in this case). However, the relatively strict convergence criterion may also cause the algorithm to alternate indefinitely within a very small band without ever converging. We accommodate this issue by altering the criterion after a few iterations to measure the difference in the average over the last 4 iterations and the average over the previous 4 iterations. This results in a significant increase in the speed of convergence without a material impact on the statistical properties. Additionally, the criterion is more difficult to satisfy in bigger models. It is therefore automatically adjusted to be more lenient as more lags enter the model, i.e. the criterion 0.005 is multiplied by the number of lags p. Finally, users are also able to specify their own criterion with the `criterion()` option.

Non-convergence in itself also entails important information about the model considered by the researcher. Our (Monte Carlo) experiments have shown that the algorithm has good convergence properties, even in small datasets, when the model is correctly specified. However, if the model is misspecified these properties tend to deteriorate, especially when the considered lag length for the dependent variable is set much too high. As such, failure to converge can be seen as a rough indication for the model being too large and can be used as a tool for model building.

If the researcher is confident about the specified model, a divergent estimator may be remedied by an alternative initialization scheme. Generally speaking, the stability of the algorithm tends to increase when more restrictions are put on the initial conditions. A purely data-driven initialization like the burn-in tends to be less stable whereas the `aho`/`ahe` options and especially the `det` option impose more structure and therefore more likely lead to convergence. This is of particular importance for small datasets where the data may be (nearly) non-stationary or very noisy. Parameter estimates may imply non-stationarity in which case a burn-in initialization can result in generated initial conditions that are close to infinity and not of practical use[4]. Similarly, the original data may not be rich enough to allow meaningful estimation of the initial condition covariance matrix Σ used in the analytical initializations `aho`/`ahe`.[5] In case the generated initial values are unreasonably large, the `xtbcfe` estimator will issue a warning alerting the user of numerical problems that may follow. A less data-driven initiation like the deterministic (`det`) option should then be considered as an alternative.

9.2 Commands and estimation output

We obtained the results for the `xtbcfe` routine from section 5 using the commands and output outlined below. First we load the dataset

```
. webuse abdata
```

4. The burn-in generates initial conditions from the model with estimated parameters. Therefore, if parameter roots imply non-stationarity, the unrestricted burn-in would generate observations from a non-stationary autoregressive process and quickly obtain very large numbers that cause numerical problems. We have therefore adjusted the burn-in to attenuate this issue by imposing stationarity over the burn-in period.
5. To ensure positive definiteness, the estimation of Σ occurs in a restricted manner. We start from a diagonal matrix (estimating variances) and fill in the kth diagonal (estimating covariances) only if the resulting matrix remains positive definite. If this is not the case, all the remaining diagonals (k up to p) are kept at zero.

As this dataset is already `xtset` we do not need to do so again. We specify to generate 250 bootstrap samples with wild bootstrap resampling in combination with the burn-in initiation, 2 lags, 50 iterations for bootstrapped standard errors and the inclusion of time dummies

```
. xtbcfe n w wL1 k kL1 kL2 ys ysL1 ysL2, bciters(250) res(wboot) ini(bi) infer(
> inf_se) infit(50) lags(2) te
25% of inference iterations performed...
50% of inference iterations performed...
75% of inference iterations performed...
95% of inference iterations performed...

Bootstrap corrected dynamic FE regression      Number of obs      =       751
Group variable : id                            Number of groups   =       140

Resample       : Wild bootstrap                Obs per group: min =         5
Initialization : Burn-in                                      avg =       5.4
Convergence    : Yes                                          max =         7

Dependent variable : n
------------------------------------------------------------------------------
                Results
                 Coefs.  Std. Err.          t      P>|t|  [95% Conf.  Interval]
------------------------------------------------------------------------------
L.n           1.0080990  0.0574874      17.54      0.000  0.8951962  1.1210019
L2.n         -0.1610846  0.0694129      -2.32      0.021 -0.2974086 -0.0247606
w            -0.5601488  0.1625968      -3.45      0.001 -0.8794822 -0.2408154
wL1           0.4952296  0.1922564       2.58      0.010  0.1176460  0.8728132
k             0.3849128  0.0507612       7.58      0.000  0.2852199  0.4846056
kL1          -0.2016635  0.0595062      -3.39      0.001 -0.3185311 -0.0847958
kL2          -0.0530621  0.0378414      -1.40      0.161 -0.1273810  0.0212568
ys            0.4548348  0.1783124       2.55      0.011  0.1046367  0.8050330
ysL1         -0.7455434  0.2705431      -2.76      0.006 -1.2768789 -0.2142079
ysL2          0.1329351  0.1708564       0.78      0.437 -0.2026200  0.4684901
year4         0.0146121  0.0128807       1.13      0.257 -0.0106851  0.0399094
year5         0.0265182  0.0199924       1.33      0.185 -0.0127460  0.0657824
year6        -0.0088682  0.0264345      -0.34      0.737 -0.0607846  0.0430481
year7        -0.0117055  0.0208444      -0.56      0.575 -0.0526430  0.0292320
year8         0.0010984  0.0224391       0.05      0.961 -0.0429711  0.0451679
year9         0.0187045  0.0247321       0.76      0.450 -0.0298683  0.0672773
------------------------------------------------------------------------------
Notes:
- Bootstrapped standard errors
- Confidence bounds for the t- distribution calculated with bootstrapped
  standard errors
- Inference performed with non-parametric bootstrap
```

where after estimation we obtain the covariance matrix (only partly displayed here)

```
. matrix list e(V)

symmetric e(V)[16,16]
                L.          L2.
                 n            n            w          wL1            k
  L.n    .0033048
 L2.n  -.00278868    .00481814
    w   .00384405  -.00582258    .02643771
  wL1  -.00368057    .00500156  -.02946648    .03696251
    k  -.00127285    .00008722  -.00001881  -.00065258     .0025767
```

(*output omitted*)

We next select the randomized temporal heteroscedasticity resampling scheme (`thet_r`). We use this to account for cross-sectional dependence without having to balance the data (the `csd` resampling scheme requires a balanced dataset). Moreover, it has the advantage that even though the time series dimension is short, we make use of the large cross-section size to limit the dependency over bootstrap samples and maintain the favorable properties of our estimator (see section 4.2).

```
. xtbcfe n w wL1 k kL1 kL2 ys ysL1 ysL2, bciters(250) res(thet_r) ini(bi) infer
> (inf_se) infit(50) lags(2) te
25% of inference iterations performed...
50% of inference iterations performed...
75% of inference iterations performed...
95% of inference iterations performed...

Bootstrap corrected dynamic FE regression       Number of obs      =       751
Group variable : id                             Number of groups   =       140

Resample       : random T-Heteroscedasticity    Obs per group: min =         5
Initialization : Burn-in                                       avg =       5.4
Convergence    : Yes                                           max =         7

Dependent variable : n
------------------------------------------------------------------------------
                Results
                 Coefs.  Std. Err.          t      P>|t|  [95% Conf.  Interval]
------------------------------------------------------------------------------

L.n           1.0497798  0.0771825      13.60      0.000  0.8981965  1.2013630
L2.n         -0.1679384  0.0674475      -2.49      0.013 -0.3004025 -0.0354743
w            -0.5560476  0.1495334      -3.72      0.000 -0.8497251 -0.2623700
wL1           0.5086207  0.1781491       2.86      0.004  0.1587433  0.8584982
k             0.3810623  0.0660147       5.77      0.000  0.2514121  0.5107125
kL1          -0.2214609  0.0607838      -3.64      0.000 -0.3408378 -0.1020840
kL2          -0.0446566  0.0323679      -1.38      0.168 -0.1082259  0.0189126
ys            0.4662594  0.1775042       2.63      0.009  0.1176485  0.8148704
ysL1         -0.7721300  0.2518845      -3.07      0.002 -1.2668208 -0.2774392
ysL2          0.1531533  0.1304085       1.17      0.241 -0.1029637  0.4092703
year4         0.0203758  0.0109208       1.87      0.063 -0.0010721  0.0418238
year5         0.0344659  0.0178722       1.93      0.054 -0.0006345  0.0695662
year6        -0.0014030  0.0268697      -0.05      0.958 -0.0541740  0.0513681
year7         0.0000237  0.0192105       0.00      0.999 -0.0377049  0.0377524
year8         0.0137072  0.0189288       0.72      0.469 -0.0234681  0.0508825
year9         0.0316466  0.0244391       1.29      0.196 -0.0163509  0.0796440
------------------------------------------------------------------------------

Notes:
- Bootstrapped standard errors
- Confidence bounds for the t- distribution calculated with bootstrapped
  standard errors
- Inference performed with non-parametric bootstrap
```

Subsequently, we estimate the model with the `csd` resampling option. As this requires a balanced panel, we balance the data using the `xtbalance` package (`xtbalance, range(1976 1982)`). We keep the burn-in initiation to also incorporate cross-sectional dependence in the generation of the initial conditions.

```
. xtbcfe n w wL1 k kL1 kL2 ys ysL1 ysL2, bciters(250) res(csd) ini(bi) infer(in
> f_se) infit(50) lags(2) te
25% of inference iterations performed...
50% of inference iterations performed...
75% of inference iterations performed...
95% of inference iterations performed...

Bootstrap corrected dynamic FE regression       Number of obs      =       400
Group variable : id                             Number of groups   =        80

Resample       : Cross-section dependence       Obs per group: min =         5
Initialization : Burn-in                                       avg =       5.0
Convergence    : Yes                                           max =         5

Dependent variable : n
------------------------------------------------------------------------------
                Results
                 Coefs.  Std. Err.          t      P>|t|  [95% Conf.  Interval]
------------------------------------------------------------------------------
L.n          1.1791647  0.0577193      20.43      0.000  1.0655878  1.2927416
L2.n        -0.3189932  0.0629156      -5.07      0.000 -0.4427952 -0.1951913
w           -0.1072298  0.1253633      -0.86      0.393 -0.3539129  0.1394533
wL1          0.0496606  0.1687568       0.29      0.769 -0.2824100  0.3817313
k            0.3833026  0.0581229       6.59      0.000  0.2689314  0.4976739
kL1         -0.2695174  0.0752560      -3.58      0.000 -0.4176022 -0.1214326
kL2         -0.0146591  0.0368312      -0.40      0.691 -0.0871336  0.0578154
ys           0.0337995  0.2279966       0.15      0.882 -0.4148401  0.4824391
ysL1        -0.3751031  0.2839324      -1.32      0.187 -0.9338102  0.1836039
ysL2         0.4174060  0.2198262       1.90      0.059 -0.0151563  0.8499684
year4        0.0136099  0.0125602       1.08      0.279 -0.0111054  0.0383251
year5       -0.0313750  0.0310202      -1.01      0.313 -0.0924148  0.0296649
year6       -0.0992146  0.0384908      -2.58      0.010 -0.1749547 -0.0234745
year7       -0.0195502  0.0195095      -1.00      0.317 -0.0579399  0.0188395
------------------------------------------------------------------------------
Notes:
- Bootstrapped standard errors
- Confidence bounds for the t- distribution calculated with bootstrapped
  standard errors
- Inference performed with non-parametric bootstrap

. matrix list e(V)

symmetric e(V)[14,14]
                L.          L2.
                 n           n           w         wL1           k
 L.n     .00333152
L2.n     -.0025691   .00395837
   w     .00030945  -.00043054   .01571595
 wL1    -.00114287   .00126991  -.01784427   .02847886
   k    -.00162874   .00034465   .00114095  -.00028964   .00337828

  (output omitted)
```

As the time series dimension is now shortened, the distribution of the `xtbcfe` routine may be poorly approximated by the normal distribution. A percentile interval may be the better choice for inference here because it does not make any symmetry or normality assumptions. We select it by specifying the `inference(inf_ci)` option and increase the number of inference iterations to 200.

```
. xtbcfe n w wL1 k kL1 kL2 ys ysL1 ysL2, bciters(250) res(csd) ini(bi) infer(in
```

```
> f_ci) infit(200) lags(2) te
25% of inference iterations performed...
50% of inference iterations performed...
75% of inference iterations performed...
95% of inference iterations performed...

Bootstrap corrected dynamic FE regression       Number of obs      =       400
Group variable : id                             Number of groups   =        80

Resample       : Cross-section dependence       Obs per group: min =         5
Initialization : Burn-in                                       avg =       5.0
Convergence    : Yes                                           max =         5

Dependent variable : n
------------------------------------------------------------------------------
                Results
                  Coefs.  Std. Err.          t      P>|t|  [95% Conf.  Interval]
------------------------------------------------------------------------------
L.n           1.1791647  0.0604333      19.51      0.000  1.0527685  1.2954533
L2.n         -0.3189932  0.0672792      -4.74      0.000 -0.4449091 -0.1578264
w            -0.1072298  0.1208303      -0.89      0.376 -0.3784196  0.0993305
wL1           0.0496606  0.1584258       0.31      0.754 -0.1936102  0.3992755
k             0.3833026  0.0532620       7.20      0.000  0.2632709  0.4742108
kL1          -0.2695174  0.0746871      -3.61      0.000 -0.3728123 -0.1027051
kL2          -0.0146591  0.0415927      -0.35      0.725 -0.1086166  0.0629152
ys            0.0337995  0.2057357       0.16      0.870 -0.2999481  0.4710875
ysL1         -0.3751031  0.2672561      -1.40      0.161 -0.9478765  0.0263027
ysL2          0.4174060  0.2055512       2.03      0.043  0.0411414  0.8130541
year4         0.0136099  0.0135002       1.01      0.314 -0.0144991  0.0360641
year5        -0.0313750  0.0292467      -1.07      0.284 -0.0912637  0.0245360
year6        -0.0992146  0.0347852      -2.85      0.005 -0.1616807 -0.0307033
year7        -0.0195502  0.0200393      -0.98      0.330 -0.0542770  0.0118444
------------------------------------------------------------------------------
Notes:
- Bootstrapped standard errors
- Bootstrap 95% (percentile-based) confidence intervals
- Inference performed with non-parametric bootstrap
```

Given the limited time series size, this resampling scheme may also suffer from correlated bootstrap samples. To alleviate this issue, we use randomized temporal heteroscedasticity resampling.

```
. xtbcfe n w wL1 k kL1 kL2 ys ysL1 ysL2, bciters(250) res(thet_r) ini(bi) infer
> (inf_ci) infit(200) lags(2) te
25% of inference iterations performed...
50% of inference iterations performed...
75% of inference iterations performed...
95% of inference iterations performed...

Bootstrap corrected dynamic FE regression       Number of obs      =       400
Group variable : id                             Number of groups   =        80

Resample       : random T-Heteroscedasticity    Obs per group: min =         5
Initialization : Burn-in                                       avg =       5.0
Convergence    : Yes                                           max =         5

Dependent variable : n
------------------------------------------------------------------------------
                Results
```

	Coefs.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
L.n	1.1283840	0.0602031	18.74	0.000	0.9849712	1.2351810
L2.n	-0.2800163	0.0630316	-4.44	0.000	-0.3680919	-0.1256024
w	-0.1139698	0.1065288	-1.07	0.286	-0.3373439	0.0698350
wL1	0.0493259	0.1495614	0.33	0.742	-0.1886080	0.4095369
k	0.3815278	0.0543022	7.03	0.000	0.2784973	0.4824990
kL1	-0.2431629	0.0695407	-3.50	0.001	-0.3617617	-0.0987648
kL2	-0.0229568	0.0396791	-0.58	0.563	-0.1457693	0.0426496
ys	0.0409369	0.1990656	0.21	0.837	-0.3281545	0.4940364
ysL1	-0.3802078	0.2439174	-1.56	0.120	-0.9901919	0.0201739
ysL2	0.4098277	0.1773858	2.31	0.022	0.0536915	0.7672914
year4	0.0123800	0.0111948	1.11	0.270	-0.0086366	0.0371148
year5	-0.0311606	0.0256305	-1.22	0.225	-0.0820865	0.0169481
year6	-0.1005774	0.0299376	-3.36	0.001	-0.1591851	-0.0424375
year7	-0.0241097	0.0161215	-1.50	0.136	-0.0619703	0.0002694

```
Notes:
- Bootstrapped standard errors
- Bootstrap 95% (percentile-based) confidence intervals
- Inference performed with non-parametric bootstrap
```

About the authors

Ignace De Vos, Gerdie Everaert and Ilse Ruyssen are researchers at the Study Hive for Economic Research and Public Policy Analysis (SHERPPA), Ghent University.