UvA ECONOMETRICS

Discussion Paper: 2015/02

Iterative Bias Correction Procedures Revisited: A Small Scale Monte Carlo Study

Artūras Juodis

www.ase.uva.nl/uva-econometrics

Amsterdam School of Economics
Roetersstraat 11
1018 WB AMSTERDAM
The Netherlands

UvA UNIVERSITEIT VAN AMSTERDAM

Iterative Bias Correction Procedures Revisited: A Small Scale Monte Carlo Study[☆]

Artūras Juodis[a,b]

[a]*Amsterdam School of Economics, University of Amsterdam, Roeterstraat 11, 1018 WB, Amsterdam, The Netherlands*
[b]*Tinbergen Institute, The Netherlands*

Abstract

This paper considers estimation of general panel data models subject to the incidental parameter problem of Neyman and Scott (1948). Our main focus is on the finite sample behavior of analytical bias corrected Maximum Likelihood estimators as discussed in Hahn and Kuersteiner (2002), Hahn and Newey (2004) and Hahn and Kuersteiner (2011). As it is mentioned in Hahn and Newey (2004) and Arellano and Hahn (2006), in principle it is possible to iterate the bias formula to obtain an estimator that might have better finite sample properties than the one step estimator. In this paper we will investigate merits and limitations of iterative bias correction procedures in finite samples, by considering three examples: Panel AR(1), Panel VAR(1) and Static Panel Probit.

Keywords: Panel Data, Maximum Likelihood, Bias Correction, Large T Consistency, Probit, Monte Carlo Simulation.
JEL: C13, C33.

---

1. Introduction

The inconsistency of the conventional Fixed Effect (FE) estimator in Dynamic Panel Data (DPD) models for fixed T has been one of the leading topics in the DPD literature for the last three decades, see e.g. Nickell (1981) and Kiviet (1995). Because of the inconsistency of the FE estimator, the estimation of linear DPD models has been mainly concentrated within the GMM framework, with the estimators suggested by Arellano and Bond (1991) *inter alia*.

[☆]I would like to thank Maurice Bun and Jan Kiviet for their comments and suggestions. Financial support from the NWO MaGW grant "Likelihood-based inference in dynamic panel data models with endogenous covariates" is gratefully acknowledged.

Email address: a.juodis@uva.nl (Artūras Juodis)

In his seminal paper Kiviet (1995) suggested to bias correct the conventional FE estimator using some initial fixed T consistent estimator of the bias term. Following similar lines Hahn and Kuersteiner (2002) proposed a bias corrected FE estimator using an alternative set of large N and large T asymptotics. In principal, their approach is operational even for small values of T and does not require complicated procedures to obtain standard errors for inference. Later Hahn and Newey (2004) and Hahn and Kuersteiner (2011) described analytical bias correction procedures for non-linear static and dynamic panel data models, with the approach of Hahn and Kuersteiner (2002) as a special case.

In all the aforementioned papers finite sample results are provided for one-step bias corrected estimators. However, as it was mentioned in Hahn and Newey (2004), in principle it is possible to iterate the bias formula to obtain an estimator that might have better finite sample performance without compromising asymptotic properties. In this paper, we will consider three commonly analyzed examples in the panel data literature and investigate whether iterative methods can improve finite sample properties of the bias corrected estimators. As it will turn out, the results are mixed and depend crucially on the particular setup considered.

The rest of this note is structured as follows. In Section 2 we introduce the PVAR(1) model and provide finite sample evidence for both univariate and bivariate systems. We continue in Section 3 with another commonly used example in the panel data literature, namely a static panel probit model. Finally, we conclude in Section 4.

2. Linear models

In this paper we will focus on simple linear dynamic specifications without any additional explanatory variables and/or cross-sectional dependence. Namely, we consider a general (reduced form) Panel Vector Autoregressive models of order one or shortly PVAR(1):

$$\boldsymbol{y}_{i,t} = \boldsymbol{\eta}_i + \boldsymbol{\Phi}\boldsymbol{y}_{i,t-1} + \boldsymbol{\varepsilon}_{i,t}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T, \tag{2.1}$$

where $\boldsymbol{y}_{i,t}$ is an $[m \times 1]$ vector, $\boldsymbol{\Phi}$ is an $[m \times m]$ matrix of parameters to be estimated, $\boldsymbol{\eta}_i$ is an $[m \times 1]$ vector of fixed effects and $\boldsymbol{\varepsilon}_{i,t}$ is an $[m \times 1]$ vector of innovations independent across i, with zero mean and constant covariance matrix $\boldsymbol{\Sigma}$.[1] If we set $m = 1$ the model reduces to the linear DPD model with AR(1) dynamics.

[1] Later in the paper we present the detailed analysis when $\boldsymbol{\Sigma}$ is i specific.

For a prototypical example of (2.1) consider the following bivariate model (see e.g. Bun and Kiviet (2006), Akashi and Kunitomo (2012) and Hsiao and Zhou (2015)):

$$\begin{aligned} y_{i,t} &= \eta_{yi} + \gamma y_{i,t-1} + \beta x_{i,t} + u_{i,t}, \\ x_{i,t} &= \eta_{xi} + \phi y_{i,t-1} + \rho x_{i,t-1} + v_{i,t}, \end{aligned}$$

where $\mathrm{E}[u_{i,t}v_{i,t}] = \sigma_{uv}$. Depending on the parameter values, the process $\{x_{i,t}\}_{t=0}^{T}$ can be either exogenous ($\phi = \sigma_{uv} = 0$), weakly exogenous ($\sigma_{uv} = 0$) or endogenous ($\sigma_{uv} \neq 0$). In the next two sub-sections we will discuss the results for $m = 1$ and $m = 2$, respectively.

2.1. Panel AR(1)

The panel AR(1) model without exogenous regressors is the simplest possible specification among all Dynamic Panel Data models. However, it is also the most commonly studied model in Monte Carlo studies. It is well known that the FE (or equivalently, Maximum Likelihood) estimator is not consistent for a fixed number of time series observations, see e.g. Nickell (1981). Because of the inconsistency of the FE estimator, the estimation of linear DPD models has been mainly concentrated within the GMM framework, with the most popular estimators introduced by Arellano and Bond (1991) and Blundell and Bond (1998).[2]

However, there is a long standing tradition of bias-corrected estimators of this model for fixed T, substantially influenced by the work of Kiviet (1995). The fixed T assumption is crucial for models with few time-series observations. If, on the other hand, the time-series dimension is not negligible then one can consider joint asymptotic approximation for $N, T \to \infty$. Using alternative large N,T asymptotics with $N/T \to \rho$ Hahn and Kuersteiner (2002) proved the following result:

Theorem 2.1.

$$\sqrt{NT}(\hat{\phi}_{FE} - \phi_0) \xrightarrow{d} N\left(-\sqrt{\rho}\frac{\sigma^2}{\sigma_X^2(1-\phi_0)}, \frac{\sigma^2}{\sigma_X^2}\right), \tag{2.2}$$

where

$$\sigma_X^2 = \operatorname*{plim}_{N,T\to\infty} \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} \tilde{y}_{i,t-1}^2 = \frac{\sigma^2}{1-\phi_0^2}. \tag{2.3}$$

[2]Alternatively, there is a growing literature on fixed T consistent likelihood based estimators, e.g. Hsiao et al. (2002), Kruiniger (2013) and Bun et al. (2015).

The non-centrality parameter $-\sqrt{\rho}\frac{\sigma^2}{\sigma_X^2(1-\phi_0)}$ captures the asymptotic bias of order $\mathcal{O}(T^{-1})$ while the remaining higher order terms are asymptotically negligible as $\frac{N}{T^3} \to 0$. Both equations of Theorem 2.1 can be combined to conclude that the non-centrality parameter in the asymptotic distribution is given by:

$$-\sqrt{\rho}(1+\phi_0) \tag{2.4}$$

Remark 2.1. Note that the non-centrality parameter does not change if the white noise assumption is relaxed to the original assumptions of Hahn and Kuersteiner (2002). Furthermore, similarly to Juodis (2014) one can also show that this conclusion is also valid if some degree of time-series heteroscedasticity is allowed.

One can then conclude that bias-corrected FE estimator of the following

$$\hat{\hat{\phi}}_{FE} = \hat{\phi}_{FE} + \frac{1}{T}(1+\check{\phi}),$$

where $\check{\phi}$ is any large N,T consistent estimator of ϕ, is consistent for $N, T \to \infty$ and it is also free from non-centrality term in the limiting distribution. The key difference between approaches based on fixed T and large T approximations, is that in the latter case one can use $\hat{\phi}_{FE}$ as plug-in estimator in the bias formula. Particularly, Hahn and Kuersteiner (2002) suggest a bias corrected fixed effect estimator of the following forms:

$$\hat{\hat{\phi}}_{FE} = \frac{T+1}{T}\hat{\phi}_{FE} + \frac{1}{T}.$$

Finite sample properties of this bias corrected estimator are well documented in the panel data literature, see e.g. Bun and Carree (2005), Hahn and Moon (2006), Moreira (2009) and more recently Everaert (2013).

Several remarks regarding this estimator are worth mentioning. First of all, as it was pointed-out in Hahn and Newey (2004) and Arellano and Hahn (2006), it is possible to iterate the bias formula. The iterative bias corrected estimator $\phi^{(\infty)}$ solves:

$$\hat{\phi}^{(k)} = \left(\hat{\phi}_{FE} + \frac{1}{T}\right) + \frac{1}{T}\hat{\phi}^{(k-1)}, k \to \infty.$$

It is not difficult to see that in this simple case the iterative bias corrected estimator is of the following form:

$$\hat{\phi}^{(\infty)} = \frac{T}{T-1}\hat{\phi}_{FE} + \frac{1}{T-1}.$$

So that

$$\check{\phi} = \frac{T}{T-1}\hat{\phi}_{FE} + \frac{1}{T(T-1)}, \tag{2.5}$$

rather than $\check{\phi} = \hat{\phi}_{FE}$. As $\phi^{(\infty)} - \phi^{(k)} = \mathcal{O}_p(T^{-2})$, for large values of T both estimators are identical, but for small values the discrepancy between the two estimators may be substantial. To discuss the matter in further detail note that under rather general conditions one can show that:[3]

$$\mathrm{E}[\hat{\phi}_{FE}] = \phi_0 - \frac{1}{T}(1+\phi_0) + R,$$

where the remainder term R is of order $\mathcal{O}(T^{-2})$. It then easily follows that:

$$\mathrm{E}\left[\hat{\hat{\phi}}\right] = \phi_0 + \left(R + \frac{1}{T}R - \frac{1+\phi_0}{T^2}\right), \tag{2.6}$$

$$\mathrm{E}\left[\hat{\phi}^{(k)}\right] = \phi_0 + \left(R\sum_{j=0}^{k}\frac{1}{T^j} - \frac{1+\phi_0}{T^{1+k}}\right), \quad \forall k \geq 1 \tag{2.7}$$

$$\mathrm{E}\left[\hat{\phi}^{(\infty)}\right] = \phi_0 + \frac{T}{T-1}R = \phi_0 + \left(R + \frac{1}{T}R\right) + R\sum_{j=2}^{\infty}\frac{1}{T^j}. \tag{2.8}$$

As a result:

$$\mathrm{E}\left[\hat{\phi}^{(\infty)}\right] - \mathrm{E}\left[\hat{\hat{\phi}}\right] = \frac{1+\phi_0}{T^2} + \mathcal{O}(T^{-4}). \tag{2.9}$$

From the above we conclude as follows. From equations (2.6)-(2.8) we can observe the effects of analytical bias correction procedures on the higher order bias terms of the FE estimator. Although Dhaene and Jochmans (2015) made this observation while comparing analytical and Jackknife bias correction procedures, no analytical results were provided in their study.[4]

The way higher order bias terms are altered differs substantially between the one step and fully iterated estimators. The main difference is in the additional $\mathcal{O}(T^{-2})$ order term introduced by the $\hat{\hat{\phi}}$ estimator. As a result the fully iterated estimator does not change the bias term of order $\mathcal{O}(T^{-2})$.[5] Note that for all values of ϕ_0 in the stationary region this term is always negative, see e.g. Kiviet (1995). As equation (2.9) indicates, this term will play the leading role to explain the discrepancy between the two estimator. Lastly, in a situation where $r = 0$ the fully-iterated estimator is exactly unbiased, while the one-step counterpart still contains a bias term of order $\mathcal{O}(T^{-2})$.

[3] Here by the expectation operator we mean the expansion after taking the limit in the cross-sectional dimension N, similar to Dhaene and Jochmans (2015).

[4] In the case of the fully iterated bias corrected estimator it can be easily seen that the bias term of order k is given by $B'_k = \sum_{j=2}^{k} B_j$.

[5] Similar to the one step bias corrected estimator, the Split Panel Jackknife estimator of Dhaene and Jochmans (2015) does affect all higher order terms.

Remark 2.2. Note that for $T = 2$ we have $\phi^{(\infty)} = 2\hat{\phi}_{FE} + 1$. This estimator is nothing but the bias corrected FE estimator under the assumption of covariance stationarity.[6] If this assumption is not violated, this estimator is fixed T consistent. In terms of the results presented in equation (2.6)-(2.8) this case refers to $r = 0$.

Coming back to the results of Theorem 2.1 note that as an estimator of σ^2 we have used[7] the method of moments (or autocovariance) based estimator:

$$\hat{\sigma}^2 = (1 - \check{\phi}^2)\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{y}_{i,t-1}^2.$$

This estimator is consistent for large N and large T, but for small values of T might be sensitive to possible deviations of the initial condition from the equilibrium.[8] Thus alternatively we can use the standard Maximum Likelihood estimator of σ^2:

$$\hat{\sigma}^2(\check{\phi}) = \frac{1}{N(T-1)}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\tilde{y}_{i,t} - \check{\phi}\tilde{y}_{i,t-1}\right)^2.$$

The bias corrected estimator is then defined as:

$$\hat{\phi}_M = \hat{\phi}_{FE} + \frac{1}{T}\frac{\hat{\sigma}^2(\check{\phi})}{(1-\check{\phi})\hat{\sigma}_X^2},$$

with $\check{\phi} = \hat{\phi}_{FE}$ in this paper. The bias corrected estimators of this structure can be especially useful for models with e.g. additional exogenous regressors where not clear correspondence between σ_X^2 and ϕ can be derived. At this stage it might be tempting to consider the iterated version of this estimator, but as it turned out in preliminary Monte Carlo studies this estimator fails to converge on many occasions. Thus by considering only the leading term of the Nickell (1981) bias, convergence issues of the Bun and Carree (2005) iterative bias corrected estimator are not solved.[9]

For a finite sample study we take the original designs from Hahn and Kuersteiner (2002) and for illustration consider bias properties only. Moreover, we replace designs

[6]Similarly, for $T = 2$, one can refer to this estimator as the FDLS estimator of Han and Phillips (2013), that is consistent for any T.

[7]Implicitly in this paper and explicitly in Hahn and Kuersteiner (2002).

[8]Recall that the sufficient condition for this estimator to be $\sqrt{NT}$ consistent is $(1/N)\sum_{i=1}^{N} y_{i,0}^2 = \mathcal{O}_p(1)$.

[9]In a preliminary study other feasible large N, T bias corrected estimators that included additional higher order terms were considered. However as their performance was inferior to estimators presented here, we decided to skip the finite sample results for them.

with $T = 20$ to the ones with $T = 2$. By $\hat{\phi}_{HK}^{(1)}$ and $\hat{\phi}_{HK}^{(\infty)}$ we respectively denote one step and iterated bias corrected estimators of Hahn and Kuersteiner (2002). The one step bias iterated estimator as in Bun and Carree (2005) that is asymptotically equivalent to HK estimator is denoted by $\hat{\phi}_{BC}^{(1)}$.

Table 1: Mean Bias and RMSE AR(1) with stationary initial condition $y_{i,0}$. 100000 Monte Carlo replications.

			Bias					RMSE				
Case	N/T	ϕ_0	$\hat{\phi}_{FE}$	$\hat{\phi}_{HK}^{(1)}$	$\hat{\phi}_{HK}^{(\infty)}$	$\hat{\phi}_M$	$\hat{\phi}_{BC}^{(1)}$	$\hat{\phi}_{FE}$	$\hat{\phi}_{HK}^{(1)}$	$\hat{\phi}_{HK}^{(\infty)}$	$\hat{\phi}_M$	$\hat{\phi}_{BC}^{(1)}$
(1)	100/2	0.00	-0.49	-0.25	0.00	0.27	-0.12	0.51	0.28	0.18	0.30	0.17
(2)	100/2	0.30	-0.65	-0.32	0.00	0.14	-0.21	0.66	0.35	0.19	0.21	0.24
(3)	100/2	0.90	-0.95	-0.47	0.00	0.03	-0.44	0.96	0.50	0.20	0.28	0.47
(4)	100/5	0.00	-0.20	-0.04	0.00	0.00	-0.03	0.20	0.07	0.06	0.05	0.06
(5)	100/5	0.30	-0.28	-0.07	-0.02	-0.02	-0.07	0.28	0.09	0.06	0.06	0.09
(6)	100/5	0.90	-0.47	-0.18	-0.11	0.01	-0.23	0.47	0.18	0.12	0.11	0.24
(7)	100/10	0.00	-0.10	-0.01	0.00	0.00	-0.00	0.10	0.04	0.03	0.04	0.04
(8)	100/10	0.30	-0.14	-0.02	-0.01	0.00	-0.02	0.14	0.04	0.04	0.04	0.04
(9)	100/10	0.90	-0.24	-0.08	-0.06	0.12	-0.11	0.25	0.09	0.07	0.14	0.12

Results in Table 1 indicate that the iterated bias corrected estimator strictly dominates the one-step estimator of Hahn and Kuersteiner (2002) in terms of the finite sample bias. This conclusion is valid irrespective of the time series dimension and degree of persistence in $\{y_{i,t}\}$. It does not come as a surprise that this domination is especially clear for $T = 2$ when the iterated estimator is fixed T consistent. For most scenarios the finite sample bias of $\hat{\phi}_{HK}^{(1)}$ and $\hat{\phi}_{BC}^{(1)}$ estimators is of comparable magnitude and sign. The $\hat{\phi}_M$ estimator, on the other hand, has smaller finite sample bias than $\hat{\phi}_{HK}^{(1)}$ in 7 out of 9 scenarios. However, it is dominated by $\hat{\phi}_{HK}^{(\infty)}$ in the vast majority of cases. Note that results in Table 1 confirm that the difference between $\hat{\phi}_{HK}^{(1)} - \hat{\phi}_{HK}^{(\infty)}$ can be solely explained by the $(1 + \phi_0)/T^2$ term, as shown in (2.9).

2.2. Panel VAR(1)

Now we take one step further in terms of the model complexity and consider the Panel Vector Autoregressive model of order 1 for which the simple panel AR(1) model is just a special case. For this setup the non-centrality parameter in the limiting distribution of the fixed effects estimator is given by:

$$-\sqrt{\rho}\left(\boldsymbol{I}_m \otimes (\boldsymbol{\Sigma}_X^{-1}(\boldsymbol{I}_m - \boldsymbol{\Phi}_0)^{-1})\right)\operatorname{vec}(\boldsymbol{\Sigma}_0). \tag{2.10}$$

Here $\boldsymbol{\Sigma}_X = \sum_{j=0}^{\infty} \boldsymbol{\Phi}_0^j \boldsymbol{\Sigma}_0 (\boldsymbol{\Phi}_0^j)'$ with σ_X^2 as a special case and $\boldsymbol{\Sigma}_0 = \mathrm{E}[\boldsymbol{\varepsilon}_{i,t}\boldsymbol{\varepsilon}_{i,t}']$. Similarly to the univariate case we can use this fact to derive:

$$\mathrm{vec}\,(\boldsymbol{\Sigma}_0) = (\boldsymbol{I}_m \otimes \boldsymbol{I}_m - \boldsymbol{\Phi}_0 \otimes \boldsymbol{\Phi}_0)\,\mathrm{vec}\,\boldsymbol{\Sigma}_X.$$

So that the non-centrality parameter can be rewritten as:

$$-\sqrt{\rho}\,(\boldsymbol{I}_m \otimes ((\boldsymbol{I}_m - \boldsymbol{\Phi}_0)\boldsymbol{\Sigma}_X))^{-1}\,(\boldsymbol{I}_m \otimes \boldsymbol{I}_m - \boldsymbol{\Phi}_0 \otimes \boldsymbol{\Phi}_0)\,\mathrm{vec}\,\boldsymbol{\Sigma}_X$$

However, unlike the univariate case, this expression cannot be further simplified. Hence, the bias corrected estimator for PVAR(1) (with trivial generalization to PVAR(p) with p finite) is given by:

$$\mathrm{vec}\left(\hat{\hat{\boldsymbol{\Phi}}}'_{FE}\right) = \mathrm{vec}\left(\hat{\boldsymbol{\Phi}}'_{FE}\right) + \frac{1}{T}\left(\boldsymbol{I}_m \otimes ((\boldsymbol{I}_m - \check{\boldsymbol{\Phi}})\hat{\boldsymbol{\Sigma}}_X)\right)^{-1}\left(\boldsymbol{I}_m \otimes \boldsymbol{I}_m - \check{\boldsymbol{\Phi}} \otimes \check{\boldsymbol{\Phi}}\right)\mathrm{vec}\,\hat{\boldsymbol{\Sigma}}_X \tag{2.11}$$

for some consistent estimator $\check{\boldsymbol{\Phi}}$. Similarly to the univariate case we can consider an iterative bias corrected estimator based on equation (2.11) and another one based on the Maximum Likelihood estimator $\boldsymbol{\Sigma}$.

One special case is obtained if the population matrices are such that $\boldsymbol{\Phi}_0\boldsymbol{\Sigma}_0 = (\boldsymbol{\Phi}_0\boldsymbol{\Sigma}_0)'$.[10] Then the results that are available for the univariate case are directly applicable for the multivariate case as well. In particular:

$$\boldsymbol{\Phi}_{HK}^{(\infty)} = \frac{T}{T-1}\hat{\boldsymbol{\Phi}}_{FE} + \frac{1}{T-1}\boldsymbol{I}_m.$$

To see this consider equation (2.10), and observe that if $\boldsymbol{\Phi}_0\boldsymbol{\Sigma}_0 = (\boldsymbol{\Phi}_0\boldsymbol{\Sigma}_0)'$, one has

$$\boldsymbol{\Sigma}_X = \sum_{j=0}^{\infty} \boldsymbol{\Phi}_0^{(2j)}\boldsymbol{\Sigma}_0 = (\boldsymbol{I}_m - \boldsymbol{\Phi}_0^2)^{-1}\boldsymbol{\Sigma}_0, \tag{2.12}$$

thus

$$-\sqrt{\rho}\left(\boldsymbol{I}_m \otimes (\boldsymbol{\Sigma}_X^{-1}(\boldsymbol{I}_m - \boldsymbol{\Phi}_0)^{-1})\right)\mathrm{vec}\,(\boldsymbol{\Sigma}_0) = -\sqrt{\rho}\,\mathrm{vec}(\boldsymbol{I}_m + \boldsymbol{\Phi}_0),$$

which is a trivial extension of the univariate case $m = 1$.

To illustrate the applicability of the iterative procedure in bivariate case, we provide two designs that are based on a comprehensive study of Juodis (2014):

$$\boldsymbol{\Phi}_0 = \begin{pmatrix} 0.4 & 0.2 \\ 0.2 & 0.4 \end{pmatrix}, \quad \boldsymbol{\Sigma}_0 = \begin{pmatrix} 0.07 & 0.05 \\ 0.05 & 0.07 \end{pmatrix}.$$

[10] The Monte Carlo designs of this type were considered in the Monte Carlo study of Binder et al. (2005). As discussed by Juodis (2014), in general this condition is highly restrictive because higher order PVAR(p) processes violate it by construction.

and

$$\boldsymbol{\Phi}_0 = \begin{pmatrix} 0.4 & 0.15 \\ -0.1 & 0.6 \end{pmatrix}, \quad \boldsymbol{\Sigma}_0 = \begin{pmatrix} 0.07 & 0.05 \\ 0.05 & 0.07 \end{pmatrix}.$$

In terms of both persistence and population R^2_Δ these two designs are comparable. However, only the first design satisfies the $\boldsymbol{\Phi}_0\boldsymbol{\Sigma}_0 = (\boldsymbol{\Phi}_0\boldsymbol{\Sigma}_0)'$ condition. For ease of exposition we present results for ϕ_{11} and ϕ_{12} elements of $\boldsymbol{\Phi}$. Sample sizes considered are the same as in the univariate case. We define the $\hat{\boldsymbol{\Phi}}_M$ estimator analogously to its univariate counterpart.

At first, let us focus on the results for diagonal element ϕ_{11} in Table 2. Similar to the

Table 2: VAR(1). Bias of diagonal elements

Case	N/T	ϕ_{11}	$\hat{\phi}_{FE}$	$\hat{\phi}^{(1)}_{HK}$	$\hat{\phi}^{(2)}_{HK}$	$\hat{\phi}^{(3)}_{HK}$	$\hat{\phi}_M$
				Design 1			
(1)	100/2	0.4	-0.70	-0.35	-0.18	-0.09	0.01
(2)	100/5	0.4	-0.31	-0.09	-0.04	-0.04	-0.03
(3)	100/10	0.4	-0.15	-0.03	-0.01	-0.01	-0.01
				Design 2			
(4)	100/2	0.4	-0.73	-0.34	-0.14	-0.05	0.07
(5)	100/5	0.4	-0.34	-0.09	-0.04	-0.03	-0.02
(6)	100/10	0.4	-0.17	-0.03	-0.01	-0.01	-0.01

Table 3: VAR(1). Bias of off-diagonal elements

Case	N/T	ϕ_{12}	$\hat{\phi}_{FE}$	$\hat{\phi}^{(1)}_{HK}$	$\hat{\phi}^{(2)}_{HK}$	$\hat{\phi}^{(3)}_{HK}$	$\hat{\phi}_M$
				Design 1			
(1)	100/2	0.2	-0.10	-0.05	-0.02	-0.01	0.00
(2)	100/5	0.2	-0.06	-0.03	-0.02	-0.02	-0.02
(3)	100/10	0.2	-0.03	-0.01	-0.01	-0.01	-0.01
				Design 2			
(4)	100/2	0.15	-0.02	-0.05	-0.06	-0.07	-0.07
(5)	100/5	0.15	0.02	-0.01	-0.02	-0.02	-0.02
(6)	100/10	0.15	0.02	0.00	-0.01	-0.01	-0.01

univariate results in Table 1, we see a substantial reduction in finite sample bias of the HK one-step estimator as compared to the FE estimator. This conclusion is valid irrespective of design and/or sample size considered. More importantly, analogous to the Panel AR(1) model, an iterated version of the HK estimator further reduces the finite sample bias at every iteration step. However, unlike in the univariate case, we do not present results for the fully iterated estimator $\phi_{HK}^{(\infty)}$ because in the second design the iterative procedure does not converge. That is a clear indication that for more-complex problems where no closed form solution for $\hat{\boldsymbol{\Phi}}^{(\infty)}$ is available, the fully iterated bias correction procedures are not feasible. As for the $\hat{\boldsymbol{\Phi}}_M$ estimator, analogously to the AR(1) case it has superior finite sample bias properties as compared to the usual one-step iterated estimator of HK.

The results for the off-diagonal element ϕ_{12} are quite unexpected and qualitatively different to those for ϕ_{11}. At first sight, based on the top part of Table 3 we can conclude that in terms of both patterns and magnitudes there is no clear difference between results for diagonal and off-diagonal elements in $\boldsymbol{\Phi}$. However, this conclusion is true for Design 1 only. First of all, the FE estimator for the off-diagonal elements has a quite small bias in both absolute and relative terms. As a result, in most cases bias correction procedures correct "too much" and in the wrong direction (note the sign of the bias changed for $T = 5, 10$). Hence, for small values of ϕ_{12} this overshooting might result in "sign biased" estimates after the bias correction even in the situations where the FE estimator itself points into the right direction.

3. Static panel probit

In this section we discuss the final example that is commonly used in panel data literature. Originally this example was analyzed by Heckman (1981) and later exploited heavily in numerous Monte Carlo studies to illustrate incidental parameter problem. For simplicity and possibility of comparison we will use the setup of Hahn and Newey (2004). The model of interest can be summarized as:

$$\begin{aligned} y_{i,t}^* &= \alpha_i + \boldsymbol{\beta}'\boldsymbol{x}_{i,t} + \varepsilon_{i,t}, \quad \varepsilon_{i,t} \overset{iid}{\sim} \mathrm{N}(0,1) \\ y_{i,t} &= 1(y_{i,t}^* > 0). \end{aligned}$$

with k strictly exogenous regressors $\boldsymbol{x}_{i,t}$. Denoted by $\boldsymbol{\theta} \equiv (\alpha_1, \ldots, \alpha_N)'$ the (fixed effects) log-likelihood function has the following expression:[11]

$$\ell(\boldsymbol{\beta}, \boldsymbol{\theta}) = \sum_{i=1}^{N} \sum_{t=1}^{T} \log(\Phi \left[(2y_{i,t} - 1)(\alpha_i + \boldsymbol{x}_{i,t}'\boldsymbol{\beta})\right])).$$

[11] We prefer to use this particular expression of the log-likelihood to simplify expressions for derivatives of the log-likelihood function.

The maximum likelihood estimator of $\boldsymbol{\beta}$ suffers from the incidental parameter problem and is not consistent for fixed T (unlike the standard linear model with strictly exogenous regressors). Similarly to linear models presented in previous sections we can reduce asymptotic bias of ML estimator up to order $\mathcal{O}(T^{-2})$ using analytical bias correction formulas of Hahn and Newey (2004). In order to use their approach we need to provide expressions for the first two derivatives of the log-likelihood function $\ell(\boldsymbol{\beta}, \boldsymbol{\theta})$ with respect to α_i and first derivative with respect to $\boldsymbol{\beta}$. At first, for the ease of notation denote $f_{i,t} \equiv (2y_{i,t} - 1)(\alpha_i + \boldsymbol{x}'_{i,t}\boldsymbol{\beta})$ then:

$$
\begin{aligned}
\boldsymbol{u}_{i,t}(\boldsymbol{\beta}) &\equiv \frac{\partial \ell_{i,t}}{\partial \boldsymbol{\beta}}|_{\alpha_i=\alpha_i(\boldsymbol{\beta})} = \frac{\phi(f_{i,t})}{\Phi(f_{i,t})}(2y_{i,t}-1)\boldsymbol{x}_{i,t}, \\
v_{i,t}(\boldsymbol{\beta}) &\equiv \frac{\partial \ell_{i,t}}{\partial \alpha_i}|_{\alpha_i=\alpha_i(\boldsymbol{\beta})} = \frac{\phi(f_{i,t})}{\Phi(f_{i,t})}(2y_{i,t}-1), \\
v_{i,t,\alpha}(\boldsymbol{\beta}) &\equiv \frac{\partial^2 \ell_{i,t}}{\partial \alpha_i^2}|_{\alpha_i=\alpha_i(\boldsymbol{\beta})} = -\left(\frac{2y_{i,t}-1}{\Phi(f_{i,t})}\right)^2 \phi(f_{i,t}) \left[\Phi(f_{i,t}) f_{i,t} + \phi(f_{i,t})\right].
\end{aligned}
$$

The (one-step) bias corrected estimator is then given by:

$$
\begin{aligned}
\boldsymbol{\beta}^{(1)} &= \hat{\boldsymbol{\beta}}_{ML} + \frac{1}{T}\boldsymbol{H}(\check{\boldsymbol{\beta}})^{-1}\boldsymbol{b}(\check{\boldsymbol{\beta}}), \\
\boldsymbol{H}(\boldsymbol{\beta}) &= -\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{U}_{i,t}(\boldsymbol{\beta})\boldsymbol{U}_{i,t}(\boldsymbol{\beta})', \\
\boldsymbol{b}(\boldsymbol{\beta}) &= -\frac{1}{2N}\sum_{i=1}^{N}\left(\sum_{t=1}^{T}\boldsymbol{U}_{i,t}(\boldsymbol{\beta})\boldsymbol{V}_{i,t}(\boldsymbol{\beta}) \Big/ \left(\sum_{s=1}^{T} v_{i,s}(\boldsymbol{\beta})^2\right)\right), \\
\boldsymbol{V}_{i,t}(\boldsymbol{\beta}) &= v_{i,t}(\boldsymbol{\beta})^2 + v_{i,t,\alpha}(\boldsymbol{\beta}), \\
\boldsymbol{U}_{i,t}(\boldsymbol{\beta}) &= \boldsymbol{u}_{i,t}(\boldsymbol{\beta}) - v_{i,t}(\boldsymbol{\beta})\left(\sum_{s=1}^{T}\boldsymbol{u}_{i,s}(\boldsymbol{\beta})v_{i,s}(\boldsymbol{\beta}) \Big/ \left(\sum_{s=1}^{T} v_{i,s}(\boldsymbol{\beta})^2\right)\right).
\end{aligned}
$$

Here, similarly to univariate linear setup, $\check{\boldsymbol{\beta}}$ is any (large N, T) consistent estimator of $\boldsymbol{\beta}$. For any given value of $\check{\boldsymbol{\beta}}$ we obtain $\alpha_i(\check{\boldsymbol{\beta}})$ with simple Newton-Raphson iterations using expressions of derivatives presented above.

For the Monte Carlo analysis we take a setup of Hahn and Newey (2004) without any changes:

$$
\begin{aligned}
&y^*_{i,t} = \alpha_i + \beta_0 \times x_{i,t} + \varepsilon_{i,t}, \quad \varepsilon_{i,t} \overset{iid}{\sim} \mathrm{N}(0,1), \\
&x_{i,t} = 0.1t + 0.5x_{i,t-1} + \varepsilon^{(x)}_{i,t}, \quad \varepsilon^{(x)}_{i,t} \overset{iid}{\sim} \mathrm{U}(-1/2, 1/2), \\
&x_{i,0} \overset{iid}{\sim} \mathrm{U}(-1/2, 1/2), \alpha_i \qquad \overset{iid}{\sim} \mathrm{N}(0,1).
\end{aligned}
$$

We set $N = \{100; 500\}$ and vary $T = \{4; 8\}$ while the number of MC replications is set to $10,000$.[12] Joint log-likelihood function $\ell(\boldsymbol{\beta}, \boldsymbol{\theta})$ is maximized with respect to all parameters using the Newton-Raphson algorithm that exploits sparsity of the Hessian matrix, for further details please refer to e.g. Greene (2004).[13] The iterative bias correction procedure can be summarized by the following algorithm:[14]

Algorithm 1 Iterative Bias Correction Procedure

1. For $k = 1, \ldots, k^{\max}$:
2. Given $\boldsymbol{\beta}^{(k-1)}$ compute $\alpha_i(\boldsymbol{\beta}^{(k-1)})$ $\forall i = 1, \ldots, N$ using Newton-Raphson algorithm,
3. Given $\boldsymbol{\beta}^{(k-1)}$ and $\alpha_i(\boldsymbol{\beta}^{(k-1)})$ compute: $\boldsymbol{\beta}^{(k)} = \hat{\boldsymbol{\beta}}_{ML} + \frac{1}{T}\boldsymbol{H}(\boldsymbol{\beta}^{(k-1)})^{-1}\boldsymbol{b}(\boldsymbol{\beta}^{(k-1)})$;
4. If $\|\boldsymbol{\beta}^{(k)} - \boldsymbol{\beta}^{(k-1)}\| < \epsilon$, stop.

Here we set $\boldsymbol{\beta}^{(0)} = \hat{\boldsymbol{\beta}}_{ML}$.

We present the results with biases of conventional FE and bias corrected estimators in Table 4. In vast majority of the cases $\phi_{HN}^{(\infty)}$ converges after $7-8$ iterations, no non-convergence cases were observed.[15] From the results presented in Table 4 it can be seen

Table 4: Static panel probit $\beta_0 = 1$, with 10000 MC replications

Case	N/T	β_{FE}	$\beta_{HN}^{(1)}$	$\phi_{HN}^{(2)}$	$\phi_{HN}^{(\infty)}$
(1)	100/4	0.42	0.11	0.16	0.15
(2)	100/8	0.19	0.05	0.06	0.06
(3)	500/4	0.41	0.11	0.15	0.15
(4)	500/8	0.18	0.05	0.06	0.06

that the FE-probit estimator is substantially biased in finite sample with bias diminishing in T. Furthermore, as it was already well documented in the Panel Data literature, this estimator tends to have a positive bias. Turning our attention to bias corrected estimators we can see that the estimator of Hahn and Newey (2004) reduced the absolute bias of the FE estimator almost 4 times, while the relative bias is reduced from roughly 40% (20%) to 20% (5%) depending on the sample size $T = 4$ ($T = 8$). The results are identical to those

[12]Note that in the original study of Heckman (1981) only 25 MC replications were performed.

[13]We fix starting values for all replications to $\beta = 0$ and $\alpha_i \overset{iid}{\sim} \mathrm{N}(0, 1)$.

[14]Alternatively, we can fix α_i for all k to $\alpha_i(\boldsymbol{\beta}^{(0)})$.

[15]For larger values of β_0 (> 3) the iterative bias correction procedure failed to converge for several replications, due to the nearly perfect separation of the data.

presented by Hahn and Newey (2004), confirming that no programming error was made. Unlike the one step estimator, two step and fully iterated versions of the bias corrected estimator do not lead to improvements in terms of the finite sample bias. On the contrary, further iterations result in marginally higher bias as compared to the one step iterated estimator.

4. Conclusions

In this paper we have studied finite sample properties of the bias corrected Maximum Likelihood estimators for panel data models. Following suggestions in the Dynamic Panel Literature we have particularly focused on the iterated bias corrected estimators and their finite sample properties. We have shown that in simple panel AR(1) model unlike the one-step iterated estimator of Hahn and Kuersteiner (2002) the fully iterated estimator does not change bias of the order T^{-2} and thus might be preferred for this reason. We illustrated this issue in a short Monte Carlo study. In addition, Monte Carlo results suggest that the use of alternative estimators for variance parameter might lead to finite sample improvements as compared to the one-step iterated estimator as in Hahn and Kuersteiner (2002).

Afterwards we have taken a next step in terms of the model complexity and considered Panel VAR(1) model. We have argued that no explicit analytical formula is available for the fully-iterated bias corrected estimator. Furthermore, no clear results are available regarding the existence of the fully-iterated estimator in this case. In the Monte Carlo part of that section we have provided some evidence on the behavior of the iterative procedures for this type of models. In particular, we have encountered that while iterative procedures tend to reduce bias for diagonal elements of the VAR parameter matrix, no general conclusion can be reached for off-diagonal elements.

As our final example of this paper we have taken static panel probit model, that has long history of finite sample studies in the literature. We have replicated Monte Carlo results of Hahn and Newey (2004) and found that the one step iterated estimator a reduces substantial amount of the finite sample bias. However, our finite sample results suggest that further iterations of the bias formula do not lead to any improvement in terms of bias. On the contrary, we have encountered situations where additional iterations lead to more pronounced bias in comparison to the one step iterated estimator.

References

Akashi, K. and N. Kunitomo (2012): "Some Properties of the LIML Estimator in a Dynamic Panel Structural Equation," *Journal of Econometrics*, 166, 167 – 183.

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 J. Hahn (2006): "Understanding Bias in Nonlinear Panel Models: Some Recent Developments," Mimeo.

Binder, M., C. Hsiao, and M. H. Pesaran (2005): "Estimation and Inference in Short Panel Vector Autoregressions with Unit Root and Cointegration," *Econometric Theory*, 21, 795–837.

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

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

Bun, M. J. G., M. A. Carree, and A. Juodis (2015): "On Maximum Likelihood Estimation of Dynamic Panel Data Models," UvA-Econometrics Working Paper Series.

Bun, M. J. G. and J. F. Kiviet (2006): "The Effects of Dynamic Feedbacks on LS and MM Estimator Accuracy in Panel Data Models," *Journal of Econometrics*, 132, 409–444.

Dhaene, G. and K. Jochmans (2015): "Split-panel Jackknife Estimation of Fixed-Effect Models," *Review of Economic Studies*, (forthcoming).

Everaert, G. (2013): "Using Backward Means to Eliminate Individual Effects from Dynamic Panels," *The Econometrics Journal*, 16, 179–221.

Greene, W. (2004): "The Behaviour of the Maximum Likelihood Estimator of Limited Dependent Variable Models in the Presence of Fixed Effects," *Econometrics Journal*, 7, 98–119.

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

——— (2011): “Bias Reduction for Dynamic Nonlinear Panel Data Models with Fixed Effects,” *Econometric Theory*, 27, 1152–1191.

Hahn, J. and H. R. Moon (2006): “Reducing Bias of MLE in A Dynamic Panel Model,” *Econometric Theory*, 22, 499–512.

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

Han, C. and P. C. B. Phillips (2013): “First Difference Maximum Likelihood and Dynamic Panel Estimation,” *Journal of Econometrics*, 175, 35–45.

Heckman, J. (1981): “Statistical Models for Discrete Panel Data,” in *Structural Analysis of Discrete Data with Econometric Applications*, ed. by C. Manski and D. McFadden, MIT Press.

Hsiao, C., M. H. Pesaran, and A. K. Tahmiscioglu (2002): “Maximum Likelihood Estimation of Fixed Effects Dynamic Panel Data Models Covering Short Time Periods,” *Journal of Econometrics*, 109, 107–150.

Hsiao, C. and Q. Zhou (2015): “Statistical Inference for Panel Dynamic Simultaneous Equations Models,” *Journal of Econometrics*, (forthcoming).

Juodis, A. (2014): “First Difference Transformation in Panel VAR models: Robustness, Estimation and Inference,” UvA-Econometrics working paper 2013/06.

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

Kruiniger, H. (2013): “Quasi ML Estimation of the Panel AR(1) Model with Arbitrary Initial Conditions,” *Journal of Econometrics*, 173, 175–188.

Moreira, M. J. (2009): “A Maximum Likelihood Method for the Incidental Parameter Problem,” *The Annals of Statistics*, 37(6A), 3660–3696.

Neyman, J. and E. L. Scott (1948): “Consistent Estimation from Partially Consistent Observations,” *Econometrica*, 16, 1–32.

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