Backward mean transformation in unit root panel data models

Juodis, A.; Poldermans, R.W.

DOI
10.1016/j.econlet.2021.109780

Publication date
2021

Document Version
Final published version

Published in
Economics Letters

Economics Letters 201 (2021) 109780

Contents lists available at ScienceDirect

Economics Letters

journal homepage: www.elsevier.com/locate/ecolet

Backward mean transformation in unit root panel data models☆

Artūras Judois[a,*], Rutger W. Poldermans[b]

[a] Amsterdam School of Economics, University of Amsterdam and Tinbergen Institute, The Netherlands
[b] Department of Economics and Management, University of Luxembourg, Luxembourg

ARTICLE INFO

Article history:
Received 18 January 2021
Accepted 8 February 2021
Available online 11 February 2021

JEL classification:
C12
C13

Keywords:
Dynamic panel data
Unit root
Backward filtering

ABSTRACT

The effectiveness of an orthogonal to backward mean transformation is investigated in the context of a non-stationary panel data model. It is shown that the corresponding estimator is as efficient as Transformed Maximum Likelihood when the autoregressive parameter is equal to unity. Furthermore, a recently introduced bias-corrected version is almost as efficient as the Pooled Least Squares estimator.

1. Introduction

Dynamic panel data models play a prominent role among empirical tools used by applied researchers. As is well-known from Nickell (1981), the conventional Fixed Effects (FE) estimator suffers from a sizeable finite sample bias for small values of T. The bias is in general more noticeable in case of persistent data, which is bad news for most applications involving macroeconomic panels. It motivated some authors to devise alternative estimation techniques capable of overcoming the shortcomings of the FE estimator.

In particular, we investigate the Least Squares (LS) estimators of Choi et al. (2010) and Everaert (2013). Both papers, among other things, establish that the corresponding estimators are nearly (asymptotically) unbiased under stationarity, while they are asymptotically unbiased in case of a unit root. However, neither of the two studies investigate the asymptotic variance in a non-stationary unit-root setup. In this paper we do exactly that.

After introducing the model in Section 2, the main results are presented in Section 3. We show that in a model with the autoregressive parameter equal to unity the estimator of Everaert (2013) is as efficient as the Transformed Maximum Likelihood (TML) estimator studied by Kruiniger (2008). Similar conditions render the estimator of Choi et al. (2010) to have even a smaller asymptotic variance. Consequently, the backward-means based estimators outperform the conventional (bias-corrected) FE estimator when data are persistent.

2. The model

Consider the panel AR(1) model

$$y_{i,t} = \eta_i + \phi y_{i,t-1} + \varepsilon_{i,t}, \quad \text{with } \mathrm{E}[\varepsilon_{i,t}|y_{i,0}, \eta_i] = 0 \tag{1}$$

and data observed over $i = 1, \ldots, N$ cross-sectional units in $t = 1, \ldots, T$ time periods. We assume that idiosyncratic errors $\varepsilon_{i,t}$ are independent over i, while the initial conditions $y_{i,0}$ are assumed to be observed. The FE estimator is not consistent for any T fixed in this model, where the corresponding inconsistency is usually labelled as the "Nickell bias".

An alternative estimator is proposed by Everaert (2013) to mitigate the finite sample bias. In order to introduce the method, it is useful to consider the LS estimator of ϕ from the following augmented regression (e.g. in Mundlak 1978)

$$y_{i,t} = \phi y_{i,t-1} + \delta \bar{y}_{i,-} + \dot{\varepsilon}_{i,t} \tag{2}$$

with the new composite error term given by $\dot{\varepsilon}_{i,t} = \varepsilon_{i,t} + \eta_i - \delta \bar{y}_{i,-}$ and where $\bar{y}_{i,-} = T^{-1}\sum_{t=1}^{T} y_{i,t-1}$. The inclusion of $\bar{y}_{i,-}$ in the regression model (while at the same time ignoring the presence of η_i), ensures that the LS estimator, which is numerical equivalent to the FE estimator, is consistent as $T \to \infty$.

The use of the full sample mean $\bar{y}_{i,-}$ is not without drawbacks, as it is correlated with all $\{\varepsilon_{i,t}\}_{t=1}^{T-1}$. In particular, the sequence of combined error terms $\{\dot{\varepsilon}_{i,t}, \dot{\varepsilon}_{i,t-1}, \ldots\}$ is not a Martingale Difference even for $\eta_i = 0$. Following Everaert (2013), this drawback

☆ Financial support from the Netherlands Organization for Scientific Research (NWO) under research grant number 451-17-002 is gratefully acknowledged by the first author.

* Correspondence to: Roetersstraat 11, Amsterdam School of Economics, University of Amsterdam, 1018 WB Amsterdam, The Netherlands.

E-mail addresses: A.Juodis@uva.nl (A. Juodis), rutger.poldermans@uni.lu (R.W. Poldermans).

https://doi.org/10.1016/j.econlet.2021.109780

is easily fixed if in estimation $\bar{y}_{i,-}$ is replaced by the backward (recursive) mean (of $y_{i,t-1}$)

$$\bar{y}^b_{i,t-1} = \frac{1}{t}\sum_{k=0}^{t-1} y_{i,k}. \tag{3}$$

Unlike the full sample mean, the backward mean by construction is not correlated with the current and the future values of $\varepsilon_{i,t}$.[1] Everaert (2013) used this observation as a motivation to estimate ϕ from the following augmented regression model

$$y_{i,t} = \phi y_{i,t-1} + \delta \bar{y}^b_{i,t-1} + \dot{\varepsilon}_{i,t} \tag{4}$$

with $\dot{\varepsilon}_{i,t}$ appropriately redefined. As with the standard FE estimator, the LS estimator of this type is not consistent for any fixed T but is consistent for T large if the data is stationary (see Everaert 2013 and Juodis 2021). While the proposed estimator is not consistent for T fixed, it has a small bias in case of an AR(1) model as found by Everaert (2013).

From this point onward, we refer to the LS estimator of ϕ from (4) as the Orthogonal to Backward Mean (OBM) estimator $\widehat{\phi}_{OBM}$. As argued by Juodis (2021), the related Recursive Mean Adjustment (RMA) estimator $\widehat{\phi}_{RMA}$ of Choi et al. (2010) in this setup can be seen as the restricted version of the OBM estimator of Everaert (2013) by fixing $\delta = 1 - \phi$.

3. Main results

The use of backward means as the de-trending tool has a long tradition in panel unit root testing, see e.g. Westerlund (2015, 2016). In many cases the use of backward means (or "recursive de-trending") leads to superior power properties. Motivated by this observation, we study the OBM and RMA estimator in a unit root setting in this section, i.e. the model in (1) with $\phi_0 = 1$. Although we formally present results for $\phi_0 = 1$, we suspect that our results can be useful in understanding the general behaviour of the OBM/RMA estimator in models with persistent regressors.

To be specific, attention is restricted to the standard model with a *common dynamic restriction* imposed, i.e. $\eta_i = (1-\phi)\mu_i$. Then

$$y_{i,t} = y_{i,t-1} + \varepsilon_{i,t} \tag{5}$$

for $\phi_0 = 1$. It is assumed that $\varepsilon_{i,t}$ and $y_{i,0}$ satisfy the following set of assumptions.

Assumption 3.1 (*Sampling*)**.**

(i) $\varepsilon_{i,t}$ is independently and identically distributed (iid) across both i and t with $\mathrm{E}[\varepsilon_{i,t}] = 0$, $\mathrm{E}\left[\varepsilon^2_{i,t}\right] = \sigma^2_\varepsilon$ and $\mathrm{E}\left[\varepsilon^4_{i,t}\right] < \infty$.

(ii) $y_{i,0} = \sqrt{C(T)}\ddot{y}_{i,0}$ where $C(T) = o(T)$, and $\mathrm{E}[\ddot{y}_{i,0}] = 0$ and $\mathrm{E}[\ddot{y}^4_{i,0}] < \infty$. $\ddot{y}_{i,0}$ are iid across i, independent of $\varepsilon_{i,t}$.

Note that the second moment of the initial observation is restricted to be of lower order than $\mathcal{O}(T)$, e.g. accommodating the $y_{i,0} = \mathcal{O}_P(1)$ setup of Hahn and Kuersteiner (2002). The next theorem summarizes the asymptotic properties of the OBM estimator in the unit root model.

Theorem 1. *Let Assumption 3.1 be satisfied. Then as $N, T \xrightarrow{j} \infty$:*

$$T\sqrt{N}(\widehat{\phi}_{OBM} - 1) \xrightarrow{d} N(0, 8). \tag{6}$$

Note that we do not impose any restrictions on the relative expansions rates of N, T. More specifically, one can easily show that $\widehat{\phi}_{OBM}$ is actually fixed T consistent. Firstly, this theorem indicates that the OBM estimator is asymptotically unbiased in the non-stationary setup. Our next result shows that this conclusion is also applicable for the RMA estimator.

Corollary 1. *Let Assumption 3.1(i) be satisfied. Then as $N, T \xrightarrow{j} \infty$:*

$$T\sqrt{N}(\widehat{\phi}_{RMA} - 1) \xrightarrow{d} N(0, 6). \tag{7}$$

If also $T/C(T) = o(1)$:

$$T\sqrt{N}(\widehat{\phi}_{OBM} - 1) \xrightarrow{d} N(0, 6). \tag{8}$$

Contrary to the stationary case (see e.g. Juodis 2021), the first result of this corollary shows that the RMA estimator is more efficient as long as the initial condition is bounded. However, the second part suggests that the asymptotic equivalence (in terms of the limiting random variable) can be restored if and only if the initial condition is divergent.

In order to put our results into a perspective of the available asymptotic results in the literature, it is evaluated in terms of the FE estimator. Under the assumption of normally distributed error terms, Hahn and Kuersteiner (2002) showed that asymptotic variance of the FE estimator is equal to 10.2 in the unit root case. Thus the OBM and RMA are not only more efficient than the FE estimator asymptotically, but also, unlike the FE estimator, they are asymptotically unbiased.

To the best of our knowledge there is only one other estimator in the literature with the same asymptotic variance as in Theorem 1. In particular, under similar assumptions Kruiniger (2008) showed that

$$T\sqrt{N}(\widehat{\phi}_{TML} - 1) \xrightarrow{d} N(0, 8), \tag{9}$$

where $\widehat{\phi}_{TML}$ is the TML estimator of Hsiao et al. (2002) and Kruiniger (2008).[2] As a result, the OBM estimator is asymptotically at least as efficient as the TML estimator.

Finally, in Juodis (2021) a bias-corrected OBM estimator is proposed. The motivation for bias-correction originates from the potentially non-negligible bias in case of stationarity. For this estimator, the following result is established.

Proposition 1. *Let Assumption 3.1 be satisfied. Then as $N, T \xrightarrow{j} \infty$:*

$$T\sqrt{N}(\widetilde{\phi}_{OBM} - 1) \xrightarrow{d} N\left(0, \frac{8}{3}\right), \tag{10}$$

where $\widetilde{\phi}_{OBM} = \widehat{\phi}_{OBM} - \widehat{b}_{N,T}$.

Because $2 < 8/3 \ll 8$, the asymptotic variance of the bias-corrected OBM estimator is substantially smaller than that of its original counterpart.[3] This efficiency gain results from non-zero asymptotic covariance between $y_{i,t}$ and $\bar{y}^b_{i,t-1}$. It is yet another manifestation of non-standard asymptotic behaviour of the panel data estimators with a root close to unity.

[1] This observation can be used to motivate the use of $y_{i,t-1} - \bar{y}^b_{i,t-1}$ as an instrument for $y_{i,t-1}$.

[2] Here we consider the restricted version of the TML estimator that in estimation assumes that $\Delta y_{i,t}$ is a covariance stationary process for all t. The unrestricted version of that estimator, as studied e.g. in Kruiniger (2013), Bun et al. (2017), and Juodis (2018a,b) in general does not have a normal asymptotic limit.

[3] Here the lower bound of 2 corresponds to the variance of the Pooled OLS estimator.

4. Concluding remarks

In this note we studied the asymptotic properties of the panel data estimators of Choi et al. (2010) and Everaert (2013). In a model with the autoregressive parameter equal to unity we showed that both estimators have a substantially smaller asymptotic variance than the FE estimator. These results are complementary to those provided in Choi et al. (2010), Everaert (2013), and Juodis (2021) who study asymptotic and finite sample results under stationarity.

Appendix. Proofs

Proof of Theorem 1. For now assume that $y_{i,0} = 0, \forall i$. Later we relax this assumption. Denote by $\boldsymbol{\beta} = (\phi, \rho)'$, with $\rho = \delta - 1$ and $\boldsymbol{\beta}_0 = (1, 0)'$. The scaled and centered version of the estimator takes a usual form

$$T\sqrt{N}\left(\widehat{\boldsymbol{\beta}}_{OBM} - \boldsymbol{\beta}_0\right) = \boldsymbol{A}_{NT}^{-1}\boldsymbol{d}_{NT}, \tag{A.1}$$

where $\boldsymbol{A}_{NT}$ and $\boldsymbol{d}_{NT}$ are of order 2×2 and 2×1 respectively. Each element of the corresponding matrix/vector can be expressed as a weighted cross-sectional average, e.g.

$$\boldsymbol{A}_{NT}^{(1.1)} = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{A}_{iT}^{(1.1)}, \quad \boldsymbol{d}_{NT}^{(1)} = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\boldsymbol{d}_{iT}^{(1)}.$$

To streamline the derivations, observe that

$$\frac{1}{T^{k+1}}\sum_{t=1}^{T}t^k = \int_0^1 x^k dx + \mathcal{O}(T^{-1}) = \frac{1}{k+1} + \mathcal{O}(T^{-1}). \tag{A.2}$$

At first consider

$$\begin{aligned}\mathrm{E}[\boldsymbol{A}_{iT}^{(1.1)}] &= \frac{1}{T^2}\sum_{t=1}^{T}\mathrm{E}[y_{i,t-1}^2] = \sigma_\varepsilon^2\frac{1}{T^2}\sum_{t=1}^{T}(t-1)\\ &= \sigma_\varepsilon^2\int_0^1 xdx + \mathcal{O}(T^{-1}) = \frac{\sigma_\varepsilon^2}{2} + \mathcal{O}(T^{-1}).\end{aligned}$$

Regarding the off-diagonal element

$$\begin{aligned}\mathrm{E}[\boldsymbol{A}_{iT}^{(1.2)}] &= \frac{1}{T^2}\sum_{t=1}^{T}\mathrm{E}[y_{i,t-1}\bar{y}_{i,t-1}^b] = \frac{1}{T^2}\sum_{t=1}^{T}\frac{1}{t}\sum_{s=1}^{t}\mathrm{E}[y_{i,t-1}y_{i,s-1}]\\ &= \sigma_\varepsilon^2\frac{1}{T^2}\sum_{t=1}^{T}t\left(\frac{1}{t^2}\sum_{s=1}^{t}(s-1)\right)\\ &= \sigma_\varepsilon^2\int_0^1 x\left(\int_0^1 ydy\right)dx + \mathcal{O}(T^{-1})\\ &= \frac{\sigma_\varepsilon^2}{2}\int_0^1 xdx + \mathcal{O}(T^{-1}) = \frac{\sigma_\varepsilon^2}{4} + \mathcal{O}(T^{-1}).\end{aligned}$$

The remaining diagonal $\boldsymbol{A}_{iT}^{(2.2)}$ term requires a bit more care, i.e.

$$\begin{aligned}\mathrm{E}[\boldsymbol{A}_{iT}^{(2.2)}] &= \frac{1}{T^2}\sum_{t=1}^{T}\mathrm{E}[(\bar{y}_{i,t-1}^b)^2] = \frac{1}{T^2}\sum_{t=1}^{T}\frac{1}{t^2}\sum_{s=1}^{t}\sum_{k=1}^{t}\mathrm{E}[y_{i,k-1}y_{i,s-1}]\\ &= \sigma_\varepsilon^2\frac{1}{T^2}\sum_{t=1}^{T}\frac{1}{t^2}\sum_{s=1}^{t}\sum_{k=1}^{s}(k-1)\\ &\quad + \sigma_\varepsilon^2\frac{1}{T^2}\sum_{t=1}^{T}\frac{1}{t^2}\sum_{s=1}^{t}\sum_{k=s+1}^{t}(s-1)\\ &= \sigma_\varepsilon^2\int_0^1 x\left[\int_0^1\left(\int_0^y zdz\right)dy + \int_0^1\left(\int_y^1 ydz\right)dy\right]dx\\ &\quad + \mathcal{O}(T^{-1})\\ &= \sigma_\varepsilon^2\int_0^1 x\left[\int_0^1 \frac{1}{2}y^2dy + \int_0^1(y-y^2)dy\right]dx + \mathcal{O}(T^{-1})\\ &= \sigma_\varepsilon^2\int_0^1 x\left[\int_0^1\left(y - \frac{1}{2}y^2\right)dy\right]dx + \mathcal{O}(T^{-1})\\ &= \frac{\sigma_\varepsilon^2}{6}\int_0^1 x[3-1]\,dx + \mathcal{O}(T^{-1}) = \frac{\sigma_\varepsilon^2}{6} + \mathcal{O}(T^{-1}).\end{aligned}$$

Combining all results

$$\operatorname*{plim}_{N,T\to\infty} \boldsymbol{A}_{NT} = \frac{\sigma_\varepsilon^2}{12}\begin{pmatrix} 6 & 3\\ 3 & 2\end{pmatrix},$$

where the conclusion holds using Kolmogorov's Strong Law of Large Numbers for iid data. Similarly, if one considers transformed variables, i.e. $\widetilde{y}_{i,t-1} = y_{i,t-1} - \bar{y}_{i,t-1}^b$ rather than $y_{i,t-1}$, this matrix is of the form

$$\operatorname*{plim}_{N,T\to\infty} \boldsymbol{A}_{NT} = \frac{\sigma_\varepsilon^2}{12}\begin{pmatrix} 2 & 1\\ 1 & 2\end{pmatrix}. \tag{A.3}$$

Next, consider $\boldsymbol{d}_{NT}$, where $\boldsymbol{d}_{iT} = T^{-1}\sum_{t=1}^{T}\boldsymbol{\xi}_{i,t}$. As both $\bar{y}_{i,t-1}^b$ and $y_{i,t-1}$ are independent of $\varepsilon_{i,t}$, vector $\boldsymbol{\xi}_{i,t} = (y_{i,t-1}, \bar{y}_{i,t-1}^b)'\varepsilon_{i,t}$ forms a martingale difference sequence (in the time-series dimension for each i), such that $\mathrm{E}[\boldsymbol{d}_{NT}] = \boldsymbol{0}_2$. The variance–covariance matrix of $\boldsymbol{d}_{iT}$ follows from the expressions for $\mathrm{E}[\boldsymbol{A}_{iT}]$. In particular,

$$\begin{aligned}\mathrm{E}[(\boldsymbol{d}_{iT}^{(1)})^2] &= \frac{1}{T^2}\sum_{t=1}^{T}\sum_{k=1}^{T}\mathrm{E}[\varepsilon_{i,t}\varepsilon_{i,k}y_{i,t-1}y_{i,k-1}]\\ &= \frac{1}{T^2}\sum_{t=1}^{T}\mathrm{E}[\varepsilon_{i,t}\varepsilon_{i,t}y_{i,t-1}y_{i,t-1}]\\ &\quad + \frac{1}{T^2}\sum_{t=1}^{T}\sum_{k>t}^{T}\mathrm{E}[\varepsilon_{i,t}\varepsilon_{i,k}y_{i,t-1}y_{i,k-1}]\\ &= \sigma_\varepsilon^4\left(\int_0^1 xdx\right) + \mathcal{O}(T^{-1}) = \frac{\sigma_\varepsilon^4}{2} + \mathcal{O}(T^{-1}),\end{aligned}$$

and

$$\begin{aligned}\mathrm{E}[(\boldsymbol{d}_{iT}^{(2)})^2] &= \frac{1}{T^2}\sum_{t=1}^{T}\sum_{k=1}^{T}\mathrm{E}[\varepsilon_{i,t}\varepsilon_{i,k}\bar{y}_{i,t-1}^b\bar{y}_{i,k-1}^b]\\ &= \frac{1}{T^2}\sum_{t=1}^{T}\mathrm{E}[\varepsilon_{i,t}^2\bar{y}_{i,t-1}^b\bar{y}_{i,t-1}^b]\\ &\quad + \frac{1}{T^2}\sum_{t=1}^{T}\sum_{k>t}^{T}\mathrm{E}[\varepsilon_{i,t}\varepsilon_{i,k}\bar{y}_{i,t-1}^b\bar{y}_{i,k-1}^b]\\ &= \frac{\sigma_\varepsilon^4}{6} + \mathcal{O}(T^{-1}).\end{aligned}$$

Analogously, $\mathrm{E}[\boldsymbol{d}_{iT}^{(1)}\boldsymbol{d}_{iT}^{(2)}] = \frac{\sigma_\varepsilon^4}{4} + \mathcal{O}(T^{-1})$. This is enough to show sequential convergence where $N \to \infty$ first, followed by $T \to \infty$. The uniform integrability condition sufficient for joint convergence can be verified analogously to Phillips and Moon (1999).

Combination of the above results yields

$$T\sqrt{N}\left(\widehat{\boldsymbol{\beta}}_{OBM} - \boldsymbol{\beta}_0\right) \xrightarrow{d} N\left(\boldsymbol{0}_2, \boldsymbol{\Sigma}_\beta\right), \tag{A.4}$$

as $N, T \xrightarrow{j} \infty$, where

$$\boldsymbol{\Sigma}_\beta = \sigma_\varepsilon^2\left(\operatorname*{plim}_{N,T\to\infty}\boldsymbol{A}_{NT}\right)^{-1} = \begin{pmatrix} 8 & -12\\ -12 & 24\end{pmatrix}. \tag{A.5}$$

Hence the marginal distribution of the autoregressive parameter corresponds with

$$T\sqrt{N}\left(\widehat{\phi}_{OBM}-1\right)\xrightarrow{d}N\left(0,8\right). \tag{A.6}$$

Next, we allow for a non-zero initial condition. Observe how $y_{i,t-1}$ and $\bar{y}^{b}_{i,t-1}$ can be expressed as

$$y_{i,t-1}=y_{i,0}+\sum_{j=0}^{t-2}\varepsilon_{i,t-1-j}$$

$$\bar{y}^{b}_{i,t-1}=y_{i,0}+\frac{1}{t}\sum_{k=2}^{t-1}\sum_{j=0}^{k-2}\varepsilon_{i,k-1-j}.$$

In turn, the elements of $\boldsymbol{A}_{iT}$ are

$$\mathrm{E}[\boldsymbol{A}_{iT}^{(1.1)}]=\frac{\sigma_{\varepsilon}^{2}}{2}+\frac{1}{T}\mathrm{E}[y_{i,0}^{2}]+\mathcal{O}(T^{-1}),$$

$$\mathrm{E}[\boldsymbol{A}_{iT}^{(1.2)}]=\frac{\sigma_{\varepsilon}^{2}}{4}+\frac{1}{T}\mathrm{E}[y_{i,0}^{2}]+\mathcal{O}(T^{-1}),$$

$$\mathrm{E}[\boldsymbol{A}_{iT}^{(2.2)}]=\frac{\sigma_{\varepsilon}^{2}}{6}+\frac{1}{T}\mathrm{E}[y_{i,0}^{2}]+\mathcal{O}(T^{-1}).$$

Under our assumptions $\mathrm{E}[y_{i,0}^{2}]=o(T)$, such that all terms involving $y_{i,0}^{2}$ are asymptotically negligible. □

Proof of Corollary 1. From (A.3) in the proof of Theorem 1 we know that

$$\operatorname*{plim}_{N,T\to\infty}\boldsymbol{A}_{NT}=\frac{\sigma_{\varepsilon}^{2}}{12}\begin{pmatrix}2&1\\1&2\end{pmatrix}.$$

As the RMA estimator has only one regressor, namely $\widetilde{y}_{i,t-1}=y_{i,t-1}-\bar{y}^{b}_{i,t-1}$, it is straightforward to conclude that the variance of this estimator is simply given by the inverse of the first diagonal element of $\operatorname{plim}_{N,T\to\infty}\boldsymbol{A}_{NT}$, i.e. $\sigma_{\varepsilon}^{2}(\operatorname{plim}_{N,T\to\infty}\boldsymbol{A}_{NT,11})^{-1}=6$.

The second part follows from Theorem 1. For non-zero value of $y_{i,0}$ we have ($\boldsymbol{\imath}_2=(1,1)'$)

$$\operatorname*{plim}_{N,T\to\infty}\boldsymbol{A}_{NT}=\frac{\sigma_{\varepsilon}^{2}}{12}\begin{pmatrix}6&3\\3&2\end{pmatrix}+\boldsymbol{\imath}_2\boldsymbol{\imath}_2'\lim_{T\to\infty}\frac{\mathrm{E}[y_{i,0}^{2}]}{T}. \tag{A.7}$$

In this case the estimator is not invariant to the initial condition. However, for $\widetilde{\boldsymbol{\beta}}$ the previous matrix is of the form

$$\operatorname*{plim}_{N,T\to\infty}\boldsymbol{A}_{NT}=\frac{\sigma_{\varepsilon}^{2}}{12}\begin{pmatrix}2&1\\1&2\end{pmatrix}+\begin{pmatrix}0&0\\0&1\end{pmatrix}\lim_{T\to\infty}\frac{\mathrm{E}[y_{i,0}^{2}]}{T}. \tag{A.8}$$

Hence the convergence rate of δ is determined by $\mathrm{E}[y_{i,0}^{2}]$. E.g. for $\mathrm{E}[y_{i,0}^{2}]=\mathcal{O}(C(T))\geq\mathcal{O}(T)$

$$T\sqrt{N}\left(\widehat{\phi}_{OBM}-1\right)\xrightarrow{d}N\left(0,6\left(1-\frac{1}{4}\left(1+\lim_{T\to\infty}\frac{\mathrm{E}[y_{i,0}^{2}]}{T\sigma_{\varepsilon}^{2}}\right)^{-1}\right)^{-1}\right).\quad\square \tag{A.9}$$

Proof of Proposition 1. Here, we use matrix notation for an easier comparison with a more general model. Let

$$\widehat{\boldsymbol{b}}_{NT}=-\left(\frac{1}{NT}\boldsymbol{x}'\boldsymbol{M}_{\boldsymbol{x}_B}\boldsymbol{x}\right)^{-1}\frac{1}{NT}\widetilde{\boldsymbol{x}}'\boldsymbol{P}_{\boldsymbol{x}_B}\left(\boldsymbol{y}-\boldsymbol{x}\widehat{\phi}_{OBM}\right),$$

where vectors $\boldsymbol{y}$ and $\boldsymbol{x}$ are defined to be of order $NT\times 1$. Analogously, one can stack observations in $\bar{\boldsymbol{x}}^{b}_{i,t}$ to obtain $\boldsymbol{x}_B$. Furthermore, let $\boldsymbol{P}_{\boldsymbol{x}_B}=\boldsymbol{x}_B(\boldsymbol{x}_B'\boldsymbol{x}_B)^{-1}\boldsymbol{x}_B'$ and $\boldsymbol{M}_{\boldsymbol{x}_B}=\boldsymbol{I}_{NT}-\boldsymbol{P}_{\boldsymbol{x}_B}$, respectively.

To prove the claim of this proposition, the bias-corrected estimator is rewritten as

$$T\sqrt{N}\left(\widetilde{\phi}_{OBM}-1\right)=\left(1-\left(\frac{1}{NT^2}\widetilde{\boldsymbol{x}}'\boldsymbol{M}_{\boldsymbol{x}_B}\widetilde{\boldsymbol{x}}\right)^{-1}\left(\frac{1}{NT^2}\widetilde{\boldsymbol{x}}'\boldsymbol{P}_{\boldsymbol{x}_B}\boldsymbol{x}\right)\right)\times T\sqrt{N}\left(\widehat{\phi}_{OBM}-1\right)+\left(\frac{1}{NT^2}\widetilde{\boldsymbol{x}}'\boldsymbol{M}_{\boldsymbol{x}_B}\widetilde{\boldsymbol{x}}\right)^{-1}\left(\frac{1}{T\sqrt{N}}\widetilde{\boldsymbol{x}}'\boldsymbol{P}_{\boldsymbol{x}_B}\boldsymbol{\varepsilon}\right) \tag{A.10}$$

Using (A.3) gives

$$\frac{1}{NT^2}\widetilde{\boldsymbol{x}}'\boldsymbol{M}_{\boldsymbol{x}_B}\widetilde{\boldsymbol{x}}=\frac{\sigma_{\varepsilon}^{2}}{8}+o_P(1). \tag{A.11}$$

On the other hand,

$$\frac{1}{NT^2}\widetilde{\boldsymbol{x}}'\boldsymbol{P}_{\boldsymbol{x}_B}\boldsymbol{x}=\frac{1}{NT^2}\widetilde{\boldsymbol{x}}'\boldsymbol{P}_{\boldsymbol{x}_B}\widetilde{\boldsymbol{x}}+\frac{1}{NT^2}\widetilde{\boldsymbol{x}}'\boldsymbol{x}_B.$$

Plugging in the asymptotic approximations for all unknown quantities gives

$$\frac{1}{NT^2}\widetilde{\boldsymbol{x}}'\boldsymbol{P}_{\boldsymbol{x}_B}\boldsymbol{x}=\sigma_{\varepsilon}^{2}\left(\frac{1}{12}\left(\frac{2}{12}\right)^{-1}\frac{1}{12}+\frac{1}{12}\right)+o_P(1)=\sigma_{\varepsilon}^{2}\left(\frac{1}{24}+\frac{1}{12}\right)+o_P(1)=\frac{\sigma_{\varepsilon}^{2}}{8}+o_P(1). \tag{A.12}$$

As a result, the term in front of $T\sqrt{N}\left(\widehat{\phi}_{OBM}-1\right)$ is $o_P(1)$, i.e.

$$\left(1-\left(\frac{1}{NT^2}\widetilde{\boldsymbol{x}}'\boldsymbol{M}_{\boldsymbol{x}_B}\widetilde{\boldsymbol{x}}\right)^{-1}\left(\frac{1}{NT^2}\widetilde{\boldsymbol{x}}'\boldsymbol{P}_{\boldsymbol{x}_B}\boldsymbol{x}\right)\right)=o_P(1). \tag{A.13}$$

Since $T\sqrt{N}\left(\widehat{\phi}_{OBM}-1\right)=\mathcal{O}_P(1)$,

$$\left(1-\left(\frac{1}{NT^2}\widetilde{\boldsymbol{x}}'\boldsymbol{M}_{\boldsymbol{x}_B}\widetilde{\boldsymbol{x}}\right)^{-1}\left(\frac{1}{NT^2}\widetilde{\boldsymbol{x}}'\boldsymbol{P}_{\boldsymbol{x}_B}\boldsymbol{x}\right)\right)T\sqrt{N}\left(\widehat{\phi}_{OBM}-1\right)=o_P(1). \tag{A.14}$$

Plugging this expression back into (A.10) gives

$$T\sqrt{N}\left(\widetilde{\phi}_{OBM}-1\right)=\left(\frac{1}{NT^2}\widetilde{\boldsymbol{x}}'\boldsymbol{M}_{\boldsymbol{x}_B}\widetilde{\boldsymbol{x}}\right)^{-1}\left(\frac{1}{T\sqrt{N}}\widetilde{\boldsymbol{x}}'\boldsymbol{P}_{\boldsymbol{x}_B}\boldsymbol{\varepsilon}\right)+o_P(1).$$

Only the asymptotic variance of $\left(\frac{1}{T\sqrt{N}}\widetilde{\boldsymbol{x}}'\boldsymbol{P}_{\boldsymbol{x}_B}\boldsymbol{\varepsilon}\right)$ remains to be evaluated, which can be easily shown to be equal to

$$\operatorname*{plim}_{N,T\to\infty}\frac{\sigma_{\varepsilon}^{2}}{NT^2}\widetilde{\boldsymbol{x}}'\boldsymbol{P}_{\boldsymbol{x}_B}\widetilde{\boldsymbol{x}}=\frac{\sigma_{\varepsilon}^{4}}{24} \tag{A.15}$$

where the $\sigma_{\varepsilon}^{2}/24$ terms stems from the decomposition in (A.12). Combining all these results we conclude that

$$T\sqrt{N}\left(\widetilde{\phi}_{OBM}-1\right)\xrightarrow{d}N\left(0,\frac{8}{3}\right), \tag{A.16}$$

where $8/3=8\times 24^{-1}\times 8$. □

References

Bun, M.J.G., Carree, M.A., Juodis, A., 2017. On maximum likelihood estimation of dynamic panel data models. Oxf. Bul. Econ. Stat. 79 (4), 463–494.

Choi, C.-Y., Mark, N.C., Sul, D., 2010. Bias reduction in dynamic panel data models by common recursive mean adjustment. Oxf. Bul. Econ. Stat. 72 (5), 567–599.

Everaert, G., 2013. Orthogonal to backward mean transformation for dynamic panel data models. Econom. J. 16 (2), 179–221.

Hahn, J., Kuersteiner, G., 2002. Asymptotically unbiased inference for a dynamic panel model with fixed effects when both N and T are large. Econometrica 70 (4), 1639–1657.

Hsiao, C., Pesaran, M.H., Tahmiscioglu, A.K., 2002. Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods. J. Econometrics 109, 107–150.

Juodis, A., 2018a. First difference transformation in panel VAR models: Robustness, estimation and inference. Econometric Rev. 37 (6), 650–693.

Juodis, A., 2018b. Rank based cointegration testing for dynamic panels with fixed T. Empirical Economics 55 (2), 349–389.

Juodis, A., 2021. Backward mean transformation in panel data with predetermined regressors. In: Essays in Honor of M.H. Pesaran, vol. 44. Emerald Publishing Limited,

Kruiniger, H., 2008. Maximum likelihood estimation and inference methods for the covariance stationary panel AR(1)/unit root model. J. Econometrics 144, 447–464.
Kruiniger, H., 2013. Quasi ML estimation of the panel AR(1) model with arbitrary initial conditions. J. Econometrics 173, 175–188.
Mundlak, Y., 1978. On the pooling of time series and cross section data. Econometrica 46 (1), 69–85.
Nickell, S., 1981. Biases in dynamic models with fixed effects. Econometrica 49, 1417–1426.
Phillips, P.C.B., Moon, H.R., 1999. Linear regression limit theory for nonstationary panel data. Econometrica 67 (5), 1057–1111.
Westerlund, J., 2015. The effect of recursive detrending on panel unit root tests. J. Econometrics 185 (2), 453–467.
Westerlund, J., 2016. An IV test for a unit root in generally trending and correlated panels. Oxf. Bul. Econ. Stat. 78 (5), 752–764.