Pooled Bewley Estimator of Long-Run Relationships in Dynamic Heterogenous Panels

Alexander Chudik, M. Hashem Pesaran and Ron P. Smith

Pooled Bewley Estimator of Long-Run Relationships in Dynamic Heterogenous Panels[*]

Alexander Chudik[†], M. Hashem Pesaran[‡] and Ron P. Smith[§]

May 27, 2021

Abstract

This paper, using the Bewley (1979) transformation of the autoregressive distributed lag model, proposes a pooled Bewley (PB) estimator of long-run coefficients for dynamic panels with heterogeneous short-run dynamics, in the same setting as the widely used Pooled Mean Group (PMG) estimator. The Bewley transform enables us to obtain an analytical closed form expression for the PB, which is not available when using the maximum likelihood approach. This lets us establish asymptotic normality of PB as $n, T \to \infty$ jointly, allowing for applications with n and T large and of the same order of magnitude, but excluding panels where T is short relative to n. In contrast, asymptotic distribution of PMG estimator was obtained for n fixed and $T \to \infty$. Allowing for both n and T large seems to be the more relevant empirical setting, as revealed by numerous applications of the PMG estimator in the literature. Dynamic panel estimators are biased when T is not sufficiently large. Three bias corrections (simulation based, split-panel jackknife, and a combined procedure) are investigated using Monte Carlo experiments, of which the combined procedure works best in reducing bias. In contrast to PMG, PB does not weight by estimated variances, which can make it more robust in small samples, though less efficient asymptotically. The PB estimator is illustrated with an application to the aggregate consumption function estimated in the original PMG paper.

Keywords: Heterogeneous dynamic panels, I(1) regressors, pooled mean group estimator (PMG), Autoregressive-Distributed Lag model (ARDL), Bewley transform, bias correction, split-panel jackknife.

JEL Classifications: C12, C13, C23, C33.

1 Introduction

Estimation of cointegrating relationships in panels with heterogeneous short-run dynamics is important for a number of applications in open economy macroeconomics as well as in other fields in economics. Existing estimators in the literature are Fully Modified OLS (FMOLS) by Pedroni (2001), panel Dynamic OLS (PDOLS) by Mark and Sul (2003), likelihood based Pooled Mean Group (PMG) estimator by Pesaran, Shin, and Smith (1999), and the parametric approach by Breitung (2005).[1] In this paper, we propose a pooled Bewley (PB) estimator of long-run relationships under a similar setting as PMG, relying on the Bewley transform of ARDL model (Bewley, 1979).[2] PB estimator is computed analytically using a simple formula, and it does not rely on numerical maximization of the complex likelihood function of the PMG estimator. In contrast to PMG, we also adopt robust equal weighting in pooling of the long-run coefficients, and therefore our estimator will not be as efficient as PMG in general, but our simulations suggest very small efficiency losses when time dimension (T) is large ($T = 200$), and gains for smaller values of T (notably $T = 30$), in the relevant case with cross sectional error heteroskedasticity and heterogeneous speed of convergence towards the long-run relationships, which is when PMG estimator has asymptotic advantage.

We derive the asymptotic distribution of the PB estimator when the cross-section dimension (n) and the time dimension diverge to infinity jointly such that $\sup_{n,T} \sqrt{n}/T^{1-\epsilon} < K$, for some small $\epsilon > 0$ and a fixed positive constant, K, which allows for the relevant empirical setting where both n and T are large and of similar order of magnitude, whilst it excludes panels where T is short relative to n. In contrast, asymptotic results for the PMG, PDOLS, FMOLS and Breitung's estimators have been developed in the case with n fixed and $T \to \infty$ and/or a sequential asymptotics $T \to \infty$ followed by $n \to \infty$. Allowing for n, T to increase concurrently is more relevant for applications where n and T are both large.

Like PMG, FMOLS, PDOLS as well as Breitung's estimator, the proposed PB estimator will suffer from a small-T bias in panels where T is not sufficiently large in relation to n. Our simulations suggest this bias can be important for finite samples of interest, and therefore we also propose three

---

[1] There are numerous applications in the literature adopting these estimators. We do not provide a review here. The referenced four papers have a total 8,364 citations in Google Scholar as of 21 May 2021.

[2] See Wickens and Breusch (1988) for a discussion of the Bewley transform.

bias-corrected PB estimators, relying either on stochastic simulations and/or split panel jackknife approaches. While all methods perform well in reducing the overall bias, a combined procedure where the split panel estimates are combined together in a data-dependent way tends to outperform the others in terms of bias, for our design and sample sizes. The usefulness of the proposed estimator is also illustrated in the context of a consumption function application for OECD economies taken from Pesaran, Shin, and Smith (1999).

The remainder of this paper is organized as follows. Section 2 presents the model and assumptions, introduces the PB estimator, provides asymptotic results, and proposes bias-corrected estimators. Section 3 presents Monte Carlo evidence. Section 4 revisits the aggregate consumption function empirical application in Pesaran, Shin, and Smith (1999). Section 5 concludes. Some of the mathematical derivations and proofs are presented in an Appendix. The online Supplement provides a description of bias-corrected PMG estimators.

2 Pooled Bewley estimator of long-run relationships

Our setup is similar to Pesaran, Shin, and Smith (1999). Let $\mathbf{z}_{it} = (y_{it}, x_{it})'$ and consider the following illustrative model

$$\Delta y_{it} = c_i - \alpha_i \left(y_{i,t-1} - \beta x_{i,t-1} \right) + u_{y,it} \tag{1}$$

$$\Delta x_{it} = u_{x,it}, \tag{2}$$

for $i = 1, 2, ..., n$, and $t = 1, 2, ..., T$. Extension to include additional lags and regressors is relatively straightforward. We keep the model and notations simple for expositional purposes. The following assumptions are postulated.

Assumption 1 *(Coefficients)* *There exists $\epsilon > 0$ such that $\epsilon < \alpha_i < 1$ for all i.*

Assumption 2 *(Innovations)* $u_{x,it} \sim IID \left(0, \sigma_{xi}^2 \right)$, *and $u_{y,it}$ is given by*

$$u_{y,it} = \delta_i u_{x,it} + v_{it}, \tag{3}$$

for all i and t, where $v_{it} \sim IID \left(0, \sigma_{vi}^2 \right)$, and $u_{x,it}$ is independently distributed of $v_{i't'}$ for all i, i', t,

and t'. In addition, $\sup_{i,t} E\,|v_{it}|^{16} < K$ and $\sup_{i,t} E\,|u_{x,it}|^{8} < K$, and limits $\lim_{n\to\infty} n^{-1}\sum_{i=1}^{n}\sigma_{xi}^{2} = \sigma_{x}^{2}$ and $\lim_{n\to\infty} n^{-1}\sum_{i=1}^{n}\sigma_{xi}^{2}\sigma_{vi}^{2}/\left(6\alpha_{i}^{2}\right) = \omega_{v}^{2}$ exist.

Assumption 3 *(Initial values and deterministic terms)* $\mathbf{z}_{i,0} = (y_{i,0}, x_{i,0})'$ *is given by*

$$\mathbf{z}_{i0} = \boldsymbol{\mu}_i + \mathbf{C}_i^{*}\left(L\right)\mathbf{u}_0, \tag{4}$$

for all i and t, and $c_i = \alpha_i\mu_{i,1} - \alpha_i\beta\mu_{i,2}$ for all i, where $\mathbf{u}_0 = (u_{y,i,0}, u_{x,i,0})'$, $\boldsymbol{\mu}_i = \left(\mu_{i,1}, \mu_{i,2}\right)'$, $\|\boldsymbol{\mu}_i\| < K$, and $\mathbf{C}_i^{*}\left(L\right)$ is defined in Section A.1 in Appendix.

Remark 1 *Assumption 1 rules out zero as a limit point of $\{\alpha_i, i \in \mathbb{N}\}$, where we use $\mathbb{N}$ to denote the set of natural numbers. Assumption 2 allows for $u_{x,it}$ to be correlated with $u_{y,it}$. Cross-section dependence of $\mathbf{u}_{it}$ is ruled out. Assumption 3 (together with the remaining assumptions) ensure $\Delta\mathbf{z}_{it}$ and $(y_{it} - \beta x_{it})$ are covariance stationary.*

Substituting first (3) for $u_{y,it}$ in (1), and then substituting $u_{x,it} = \Delta x_{it}$, we obtain the following ARDL representation for y_{it}

$$\Delta y_{it} = c_i - \alpha_i\left(y_{i,t-1} - \beta x_{i,t-1}\right) + \delta_i\Delta x_{it} + v_{it}. \tag{5}$$

The pooled Bewley estimator takes advantage of the Bewley transform (Bewley, 1979). Subtracting $(1 - \alpha_i)\,y_{it}$ from both sides of (5) and re-arranging, we have

$$\alpha_i y_{it} = c_i - (1 - \alpha_i)\,\Delta y_{it} + \alpha_i\beta x_{it} + \delta_i\Delta x_{it} + v_{it}, \tag{6}$$

or (noting that $\alpha_i > 0$ for all i and multiplying the equation above by α_i^{-1})

$$y_{it} = \alpha_i^{-1}c_i + \beta x_{it} + \boldsymbol{\psi}_i'\Delta\mathbf{z}_{it} + \alpha_i^{-1}v_{it}, \tag{7}$$

where $\Delta\mathbf{z}_{it} = (\Delta y_{it}, \Delta x_{it})'$, and $\boldsymbol{\psi}_i = \left(-\frac{1-\alpha_i}{\alpha_i}, \frac{\delta_i}{\alpha_i}\right)'$. Further, stacking (7) for $t = 1, 2, ..., T$, we have

$$\mathbf{y}_i = \alpha_i^{-1}c_i\boldsymbol{\tau}_T + \mathbf{x}_i\beta + \Delta\mathbf{Z}_i\boldsymbol{\psi}_i + \alpha_i^{-1}\mathbf{v}_i, \tag{8}$$

3

where $\mathbf{y}_i = (y_{i1}, y_{i2}, ..., y_{iT})'$, $\mathbf{x}_i = (x_{i1}, x_{i2}, ..., x_{iT})'$, $\Delta\mathbf{Z}_i = \left(\Delta\mathbf{z}'_{i,1}, \Delta\mathbf{z}'_{i,2}, ..., \Delta\mathbf{z}'_{i,T}\right)'$, $\mathbf{v}_i = (v_{i,1}, v_{i,2}, ..., v_{i,T})'$, and $\boldsymbol{\tau}_T$ is $T \times 1$ vector of ones. Define projection matrix $\mathbf{M}_\tau = \mathbf{I}_T - \boldsymbol{\tau}_T\boldsymbol{\tau}'_T$. This projection matrix subtracts the period average. Let $\tilde{\mathbf{y}}_i = (\tilde{y}_{i1}, \tilde{y}_{i2}, ..., \tilde{y}_{iT})' = \mathbf{M}_\tau\mathbf{y}_i$, and similarly $\tilde{\mathbf{x}}_i = (\tilde{x}_{i1}, \tilde{x}_{i2}, ..., \tilde{x}_{iT})' = \mathbf{M}_\tau\mathbf{x}_i$, $\Delta\tilde{\mathbf{Z}}_i = \mathbf{M}_\tau\Delta\mathbf{Z}_i$, and $\tilde{\mathbf{v}}_i = \mathbf{M}_\tau\mathbf{v}_i$. Multiplying (8) by $\mathbf{M}_\tau$, we have

$$\tilde{\mathbf{y}}_i = \tilde{\mathbf{x}}_i\beta + \Delta\tilde{\mathbf{Z}}_i\boldsymbol{\psi}_i + \alpha_i^{-1}\tilde{\mathbf{v}}_i,$$

Consider the matrix of instruments

$$\tilde{\mathbf{H}}_i = (\tilde{\mathbf{y}}_{i,-1}, \tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_{i,-1}) = \mathbf{M}_\tau\mathbf{H}_i, \ \mathbf{H}_i = (\mathbf{y}_{i,-1}, \mathbf{x}_i, \mathbf{x}_{i,-1}), \tag{9}$$

where $\mathbf{y}_{i,-1} = (y_{i,1}, y_{i,1}, ..., y_{i,T-1})'$ is the data vector on the first lag of y_{it}, similarly $\mathbf{x}_{i,-1} = (x_{i,1}, x_{i,1}, ..., x_{i,T-1})'$. The PB estimator of β is given by

$$\hat{\beta} = \left(\sum_{i=1}^n \tilde{\mathbf{x}}'_i\mathbf{M}_i\tilde{\mathbf{x}}_i\right)^{-1}\left(\sum_{i=1}^n \tilde{\mathbf{x}}'_i\mathbf{M}_i\tilde{\mathbf{y}}_i\right), \tag{10}$$

where

$$\mathbf{M}_i = \mathbf{P}_i - \mathbf{P}_i\Delta\tilde{\mathbf{Z}}_i\left(\Delta\tilde{\mathbf{Z}}'_i\mathbf{P}_i\Delta\tilde{\mathbf{Z}}_i\right)^{-1}\Delta\tilde{\mathbf{Z}}'_i\mathbf{P}_i, \tag{11}$$

and

$$\mathbf{P}_i = \tilde{\mathbf{H}}_i\left(\tilde{\mathbf{H}}'_i\tilde{\mathbf{H}}_i\right)^{-1}\tilde{\mathbf{H}}_i \tag{12}$$

is the projection matrix associated with $\tilde{\mathbf{H}}_i$.

In addition to Assumptions 1-3, we also require the following high-level conditions to hold in the derivations of the asymptotic distribution of the PB estimator under the joint asymptotics $n, T \to \infty$.

Assumption 4 *There exists $T_0 \in \mathbb{N}$ such that the following conditions are satisfied:*

(i) $\sup_{i\in\mathbb{N}, \ T>T_0} E\left[\lambda_{\min}^{-2}(\mathbf{B}_{iT})\right] < K$, where $\mathbf{B}_{iT} = \Delta\tilde{\mathbf{Z}}'_i\mathbf{P}_i\Delta\tilde{\mathbf{Z}}_i/T$, $\mathbf{P}_i$ is given by (12), and $\Delta\tilde{\mathbf{Z}}_i$ is defined below (8).

(ii) $\sup_{i \in \mathbb{N},\ T > T_0} E\left[\lambda_{\min}^{-2}\left(\mathbf{A}_T \tilde{\mathbf{H}}_i^{*\prime} \tilde{\mathbf{H}}_i^* \mathbf{A}_T\right)\right] < K$, where

$$\mathbf{A}_T = \begin{pmatrix} T^{-1} & 0 \\ 0 & T^{-1/2} \end{pmatrix},\ \mathbf{H}_i^* = \left(\tilde{\mathbf{x}}_i, \Delta \tilde{\mathbf{x}}_i, \tilde{\boldsymbol{\xi}}_{i,-1}\right), \tag{13}$$

$\tilde{\boldsymbol{\xi}}_{i,-1} = \left(\tilde{\xi}_{i,0}, \tilde{\xi}_{i,1}, ..., \tilde{\xi}_{i,T-1}\right)'$, $\tilde{\xi}_{i,t-1} = \tilde{y}_{i,t-1} - \beta \tilde{x}_{i,t-1}$, $\tilde{y}_{it}$ and $\tilde{x}_{it}$ are defined below (8), $\tilde{\mathbf{x}}_i = (\tilde{x}_{i1}, \tilde{x}_{i2}, ..., \tilde{x}_{iT})'$, and $\Delta \tilde{\mathbf{x}}_i = (\Delta \tilde{x}_{i1}, \Delta \tilde{x}_{i2}, ..., \Delta \tilde{x}_{iT})'$.

Remark 2 *Under Assumptions 1-3 (and without Assumption 4), we have* $\text{plim}_{T \to \infty} \mathbf{B}_{i,T} = \mathbf{B}_i$, *where* $\mathbf{B}_i$ *is nonsingular (see Lemma A.7 in Appendix). Similarly, it can be shown that Assumptions 1-3 are sufficient for* $\text{plim}_{T \to \infty} \mathbf{A}_T \tilde{\mathbf{H}}_i^{*\prime} \tilde{\mathbf{H}}_i^* \mathbf{A}_T$ *to exist and to be nonsingular. However, these results are not sufficient for the moments of* $\left\|\mathbf{B}_i^{-1}\right\|$ *and* $\left\|(\mathbf{A}_T \mathbf{H}_i^{*\prime} \mathbf{H}_i^* \mathbf{A}_T)^{-1}\right\|$ *to exist, which we require for the derivations of the asymptotic distribution of the PB estimator. This is ensured by Assumption 4.*

2.1 Asymptotic results

The following theorem establishes the asymptotic distribution of $\hat{\beta}$.

Theorem 1 *Let* (y_{it}, x_{it}) *be generated by model (1)-(2) and suppose Assumptions 1-4 hold. Consider the PB estimator* $\hat{\beta}$ *given by (10). Then,*

$$T\sqrt{n}\left(\hat{\beta} - \beta\right) \to_d N\left(0, \Omega\right),\ \Omega = \omega_x^{-4}\omega_v^2, \tag{14}$$

as $n, T \to \infty$ *such that* $\sup_{n,T} \sqrt{n}/T^{1-\epsilon} < K$, *for some* $\epsilon > 0$, *where* $\omega_x^2 = \sigma_x^2/6$, $\sigma_x^2 = \lim_{n \to \infty} n^{-1} \sum_{i=1}^n \sigma_{xi}^2$ *and* $\omega_v^2 = \lim_{n \to \infty} n^{-1} \sum_{i=1}^n \sigma_{xi}^2 \sigma_{vi}^2/\left(6\alpha_i^2\right)$.

All proofs are provided in Appendix.

Remark 3 *Like the PMG estimator in Pesaran, Shin, and Smith (1999), the PB estimator will also work when variables are integrated of order 0 (the I(0) case), which is not pursued in this paper. In the I(0) case, the PB estimator converges at rate* $\sqrt{nT}$.

To conduct inference, let

$$\hat{\omega}_x^2 = n^{-1} \sum_{i=1}^{n} \frac{\mathbf{x}_i' \mathbf{M}_i \mathbf{x}_i}{T^2},$$ (15)

and

$$\hat{\omega}_v^2 = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{\mathbf{x}_i' \mathbf{M}_i \hat{\mathbf{v}}_i^*}{T} \right)^2,$$ (16)

where $\hat{\mathbf{v}}_i^*$ is the vector of residuals from (8), namely

$$\hat{\mathbf{v}}_i^* = \mathbf{M}_i \left(\mathbf{y}_i - \hat{\beta} \mathbf{x}_i \right).$$ (17)

We propose the following estimator of Ω:

$$\hat{\Omega} = \hat{\omega}_x^{-4} \hat{\omega}_v^2.$$ (18)

2.2 Bias mitigation

When n is not sufficiently small relative to T, specifically when $\sqrt{n}/T \to K > 0$, then $\sqrt{n}T \left(\hat{\beta} - \beta \right)$ is no longer asymptotically distributed with zero mean. The asymptotic bias is due nonzero mean of $\tilde{\mathbf{x}}_i' \mathbf{M}_i \tilde{\mathbf{v}}_i$, and it can have important consequences for finite sample performance, as the Monte Carlo evidence in Section 3 illustrates. We consider a simulation based and split-panel jackknife methods to mitigate this bias.[3]

2.2.1 Simulation-based bias reduction

Once an estimate of the bias of $\hat{\beta}$ is available, denoted as $\hat{b}$, then the bias-corrected PB estimator is given by

$$\tilde{\beta} = \hat{\beta} - \hat{b}.$$ (19)

One possibility of estimating the bias in the literature is by stochastic simulation. We consider the following algorithm.

1. Compute $\hat{\beta}$. Given pooled estimate $\hat{\beta}$, estimate the remaining unknown coefficients of elements of (1)-(2) by least squares, and compute residuals $\hat{u}_{y,it}, \hat{u}_{x,it}$.

---

[3]There are numerous approaches that could be considered for bias reduction, besides the three methods considered in this paper. Comprehensive comparison of different bias-reduction methods is outside the scope of this paper.

2. For each $r = 1, 2, ..., R$, generate new draws for $\hat{u}_{y,it}^{(r)} = a_{y,it}^{(r)} \hat{u}_{y,it}$, and $\hat{u}_{x,it}^{(r)} = a_{x,it}^{(r)} \hat{u}_{x,it}$, where $a_{y,it}^{(r)}, a_{x,it}^{(r)}$ are randomly drawn from Rademacher distribution (Davidson and Flachaire, 2008),

$$a_{h,it}^{(r)} = \begin{cases} -1, & \text{with probability } 1/2 \\ 1, & \text{with probability } 1/2 \end{cases},$$

for $h = y, x$. Given the estimated parameters of (1)-(2) from Step 1, and initial values y_{i1}, x_{i1} generate simulated data $y_{it}^{(r)}, x_{it}^{(r)}$ for $t = 2, 3, ..., T$ and $i = 1, 2, ..., n$. Using the generated data compute $\hat{\beta}^{(r)}$.

3. Compute $\hat{b} = \left[R^{-1} \sum_{r=1}^{R} \hat{\beta}^{(r)} - \hat{\beta} \right]$.

The above procedure can be iterated by using the bias-corrected estimator, $\tilde{\beta}$, in Step 1. This is not considered in this paper.

We conduct inference using bootstrapped critical values instead of asymptotic critical values to make more accurate small sample inference. In particular, the α percent critical values are computed using the $1 - \alpha$ percent quantile of $\left\{ \left| t^{(r)} \right| \right\}_{r=1}^{R}$, where $t^{(r)} = \tilde{\beta}^{(r)} / se\left(\tilde{\beta}^{(r)} \right)$, $\tilde{\beta}^{(r)} = \hat{\beta}^{(r)} - \hat{b}$ is the bias-corrected estimate of β using the r-th draw of the simulated data, $se\left(\tilde{\beta}^{(r)} \right) = T^{-1} n^{-1/2} \hat{\Omega}^{(r)}$ is the corresponding standard error estimate, and $\hat{\Omega}^{(r)}$ is computed in the same way as $\hat{\Omega}$ in (18) but using the simulated data.

2.2.2 Jackknife bias reduction

We consider half-panel jackknife bias correction methods,[4] which can be written as

$$\tilde{\beta}_{jk} = \tilde{\beta}_{jk}(\kappa) = \hat{\beta} - \kappa \left(\frac{\hat{\beta}_a + \hat{\beta}_b}{2} - \hat{\beta} \right), \tag{20}$$

where $\hat{\beta}$ is the full sample PB estimator, $\hat{\beta}_a$ and $\hat{\beta}_b$ are the first and the second half sub-sample PB estimators, and κ is a suitably chosen weighting parameter. In stationary setting, where the bias is of order $O\left(T^{-1} \right)$, κ is chosen to be one, so that $\frac{K}{T} - \kappa \cdot \left(\frac{K}{T/2} - \frac{K}{T} \right) = 0$ for any arbitrary K.

---

[4] For other panel applications of split-panel jackknife methods, see for example Dhaene and Jochmans (2015) and Chudik, Pesaran, and Yang (2018).

In general, when the bias is of order $O\left(T^{-\epsilon}\right)$ for some $\epsilon > 0$, then κ can be chosen to solve $\frac{K}{T^\epsilon} - \kappa \cdot \left(\frac{K}{(T/2)^\epsilon} - \frac{K}{T^\epsilon}\right) = 0$, which yields $\kappa = 1/\left(2^\epsilon - 1\right)$. Under our setup with I(1) variables, we need to correct $\hat{\beta}$ for its $O\left(T^{-2}\right)$ bias, namely $\epsilon = 2$, which yields $\kappa = 1/3$.

Asymptotic arguments need not perform well for some T, therefore we also consider a simulation-based adaptive jackknife correction where $\kappa = \hat{\kappa}$ is data-dependent and computed by stochastic simulation,

$$\hat{\kappa} = \frac{\hat{b}}{\hat{b}_{a,b} - \hat{b}}, \tag{21}$$

where $\hat{b} = R^{-1} \sum_{r=1}^{R} \hat{\beta}^{(r)} - \hat{\beta}$, and $\hat{b}_{a,b} = \left(\hat{b}_a + \hat{b}_b\right)/2$, $\hat{b}_a = R^{-1} \sum_{r=1}^{R} \hat{\beta}_a^{(r)} - \hat{\beta}_a$, $\hat{b}_b = R^{-1} \sum_{r=1}^{R} \hat{\beta}_b^{(r)} - \hat{\beta}_b$.

Inference using $\tilde{\beta}^{jk}$ can be conducted based on (18) but with $\hat{\omega}_v^2$ replaced by

$$\tilde{\omega}_v^2 = \hat{\omega}_v^2 = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{\left[(1 + \kappa)\,\mathbf{x}_i' \mathbf{M}_i - 2\kappa \mathbf{x}_{ab,i}' \mathbf{M}_{ab,i}\right] \tilde{\mathbf{v}}_i^*}{T}\right)^2, \tag{22}$$

where $\tilde{\mathbf{v}}_i^* = \mathbf{M}_i \left(\mathbf{y}_i - \tilde{\beta}^{jk} \mathbf{x}_i\right)$,

$$\mathbf{x}_{ab,i}' = \begin{pmatrix} \mathbf{x}_{a,i}' \\ \mathbf{x}_{b,i}' \end{pmatrix}, \ \mathbf{M}_{ab,i} = \begin{pmatrix} \mathbf{M}_{a,i} \\ \mathbf{M}_{b,i} \end{pmatrix},$$

$\mathbf{x}_{a,i}' \left(\mathbf{x}_{b,i}'\right)$ and $\mathbf{M}_{a,i} \left(\mathbf{M}_{b,i}\right)$ are defined in the same way as $\mathbf{x}_i$, and $\mathbf{M}_i$ but using only the first (second) half of the sample.

We use bootstrapped critical values to conduct more accurate small sample inference, for both choices of κ (1/3 and $\hat{\kappa}$). Specifically, the α percent critical value is computed as the $1 - \alpha$ percent quantile of $\left\{\left|t_{jk}^{(r)}\right|\right\}_{r=1}^{R}$, where $t_{jk}^{(r)} = \tilde{\beta}_{jk}^{(r)} / se\left(\tilde{\beta}_{jk}^{(r)}\right)$, $\tilde{\beta}_{jk}^{(r)}$ is the jackknife estimate of β using the r-th draw of the simulated data generated using the algorithm described in Subsection 2.2.1, $se\left(\tilde{\beta}_{jk}^{(r)}\right)$ is the corresponding standard error estimate, namely $se\left(\tilde{\beta}_{jk}^{(r)}\right) = T^{-1} n^{-1/2} \hat{\Omega}_{jk}^{(r)}$, $\hat{\Omega}_{jk}^{(r)} = \hat{\omega}_{x,(r)}^{-4} \tilde{\omega}_{v,(r)}^2$, in which $\tilde{\omega}_{v,(r)}$ and $\hat{\omega}_{x,(r)}^2$ are computed using the simulated data, based on expressions (22) and (15), respectively.

3 Monte Carlo Evidence

3.1 Design

The Data Generating Process (DGP) is given by (1)-(2), for $i = 1, 2, ..., n$, $T = 1, 2, ..., T$, with starting values satisfying Assumption 3 with $\boldsymbol{\mu}_i \sim IIDN(\boldsymbol{\tau}_2, \mathbf{I}_2)$, and $c_i = \alpha_i \mu_{i,1} - \alpha_i \beta \mu_{i,2}$. We generate $u_{y,it} = \sigma_{y,i} e_{y,it}$, $u_{x,it} = \sigma_{x,i} e_{x,it}$, $\sigma_{y,i}^2, \sigma_{x,i}^2 \sim IIDU[0.8, 1.2]$,

$$\begin{pmatrix} e_{y,it} \\ e_{x,it} \end{pmatrix} \sim IIDN(\mathbf{0}_2, \boldsymbol{\Sigma}_e), \; \boldsymbol{\Sigma}_e \sim \begin{pmatrix} 1 & \rho_i \\ \rho_i & 1 \end{pmatrix}, \text{ and } \rho_i \sim IIDU[0.3, 0.7].$$

This setup features heteroskedastic (over i) and correlated (over y & x equations) errors, namely $E\left(u_{y,it}^2\right) = \sigma_{y,i}^2$, $E\left(u_{x,it}^2\right) = \sigma_{x,i}^2$, and $cov(u_{y,it}, u_{x,it}) = \rho_i$. We generate $\alpha_i \sim IIDU[0.2, 0.3]$. We consider $n, T = 30, 50, 100, 200$ and compute $R_{MC} = 2000$ Monte Carlo replications.

3.2 Objectives

We report bias, root mean square error (RMSE), size ($H_0 : \beta = 1$, 5% nominal level) and power ($H_1 : \beta = 0.98$, 5% nominal level) findings for the PB estimator $\hat{\beta}$ given by (10), with variance estimated using (18). Moreover, we also report findings for the three bias corrected versions of PB estimator as described in Subsection 2.2. We compare the performance of PB estimators with the PMG estimator proposed by Pesaran, Shin, and Smith (1999) and its bias-corrected versions proposed in the online supplement.[5]

3.3 Findings

Table 1 reports the findings for all estimators. The top panel reports results for PB and PMG estimators uncorrected for their small-T bias. Uncorrected PB estimator features notable negative bias, which declines with T, and does not change much with n. This bias contributes to oversized inference when T is small relative to n. These findings illustrate an important scope for bias-correction methods. The bias of PMG estimator is about 50% smaller as compared to that of the PB estimator. Despite the differences in the bias, the reported RMSE values of the two estimators

---

[5]We use $R = 5000$ replications for PB bias correction methods described in Subsection 2.2 and for PMG bias correction methods described in the online supplement.

are quite similar for sample sizes where $T > n$. When $n > T$, PMG tends to dominate in terms of RMSE due to its lower bias. Interestingly, for the smallest sample size considered, $n = T = 30$, the RMSE of PB estimator is smaller compared with PMG (0.0719 vs. 0.0749), despite almost twice larger bias (-0.0515 vs. -0.0312). Both, PMG and PB estimators are grossly oversized when T is not sufficiently large relative to n, in part also due to underestimation of standard errors in small samples (in addition to the consequences of the bias for inference).

Bias-corrected methods are quite successful in reducing the bias, and in the majority of cases also RMSE of the PB and PMG estimators. The best performing bias reduction method in reducing the bias is split-panel jackknife with $\kappa = \hat{\kappa}_{NT}$ chosen by simulations as opposed to asymptotic considerations ($\kappa = 1/3$). In terms of RMSE, the best performing method is bias reduction by stochastic simulations. For $T = 30$, application of any bias-reduction method also resulted in improved RMSE compared with the uncorrected estimators. Bias-corrected PB (using any of the three bias-reduction methods considered) achieved lower RMSE values compared with PMG (bias-corrected or uncorrected) for $T = 30$, and all choices of n.

Another important observation is the dramatic improvement in size performance. Notably, the size is very good for jackknife-corrected PB estimators (both choices of κ) for all sample sizes considered, whereas only a relatively moderate size distortions for smaller choices of T (30 and 50) are observed for bias-corrected PMG estimators.

We conclude that bias-corrected PB estimators can perform better (notably in terms of RMSE) compared with corrected or uncorrected PMG estimators for smaller values of T (especially for $T = 30$). For large values of T (≥ 200) there does not seem to be any particular advantage of PB over PMG, as both estimators seemingly perform very close with PMG performing slightly better due to its asymptotically efficient weighting of cross-section units, as to be expected. For large values of T (≥ 200) and n/T sufficiently small, there also does not seem to be any notable benefit of bias reduction methods, since both PB and PMG perform without any noticeable drawbacks. Bias-corrected PB estimators are consequently useful addition to the literature as a complement of PMG estimator, considering that the sample size in terms of time periods is often quite limited in many applications in economics.

4 Empirical Application

This section revisits consumption function empirical application undertaken by Pesaran, Shin, and Smith (1999), hereafter PSS. The long-run consumption function is assumed to be given by

$$c_{it} = d_i + \beta_1 y_{it}^d + \beta_2 \pi_{it} + \vartheta_{it},$$

for country $i = 1, 2, ..., n$, where c_{it} is the logarithm of real consumption per capita, y_{it}^d is the logarithm of real per capita disposable income, π_{it} is the rate of inflation, and ϑ_{it} is an $I(0)$ process. We take the dataset from PSS, which consists of $N = 24$ countries and a slightly unbalanced time period covering 1960-1993.[6] PSS assume all variables are $I(1)$ and cointegrated; and they estimate the coefficients β_1 and β_2 using an ARDL(1,1,1) specification, which can be written as the following error-correcting equation

$$\Delta c_{it} = -\alpha_i \left(c_{i,t-1} - d_i - \beta_1 y_{i,t-1}^d - \beta_2 \pi_{i,t-1} \right) + \delta_{i1} \Delta y_{it}^d + \delta_{i2} \Delta \pi_{it} + v_{it}, \tag{23}$$

where all coefficients, except the long-run coefficients β_1 and β_2 are country-specific.

Table 2 presents alternative estimates of the long-run coefficients. The first column reports the PMG estimates, the second column reports the PB estimates, and the subsequent columns report bias-corrected versions of these two estimators. Uncorrected PMG estimates are $\hat{\beta}_{1,PMG} = 0.904$ and $\hat{\beta}_{2,PMG} = -0.466$, for y_{it}^d, and π_{it}, respectively. Bias-corrected PMG estimates are not too far from the uncorrected PMG estimates, suggesting that the bias is small. PB estimates for the income elasticity (β_1), are slightly larger but generally very close compared with the PMG estimates. The uncorrected PB estimate for the income elasticity is $\hat{\beta}_1 = 0.912$, and the bias-corrected PB estimates are slightly larger in the range 0.918 to 0.926. In contrast, PB estimates for the inflation effect coefficient (β_2) are all substantially smaller compared with the PMG estimates. PB estimates of β_2 lie in the relatively narrow range -0.153 to -0.120, compared with the range of PMG estimates -0.474 to -0.403. While the income elasticity PB estimates are very close to the PMG estimates, PB estimators suggest much smaller long-run inflation effect.

---

[6]We have downloaded data at http://www.econ.cam.ac.uk/people-files/emeritus/mhp1/pmge_prog.zip. Codes for Monte Carlo and empirical applications in this paper is available from authors' websites.

5 Conclusion

This paper proposed a simple alternative to the Pesaran, Shin, and Smith (1999) PMG estimator of long-run relationships in heterogeneous dynamic panels. Taking advantage of Bewley transform, the proposed PB estimator has an analytical closed-form expression, and since it does not weight by estimated variances, it is more robust in small samples, though less efficient asymptotically. Since dynamic panel estimators are biased when T is small relative to n, this paper also considered bias-correction methods for the PB and PMG estimators. Monte Carlo experiments show good small sample performance of bias-corrected estimators with (corrected) PB estimators achieving better RMSE compared with (corrected) PMG for small T (in particular $T = 30$), whereas PMG slightly outperforms PB estimator for large values of T (≥ 200). The usefulness of the PB estimator was also illustrated by revisiting the aggregate consumption function estimated in the original PMG paper, where we found similar income elasticity, but substantially smaller inflation effect.

Table 1: MC findings for the estimation of long-run coefficient β

$n\backslash T$	Bias ($\times$ 100)				RMSE ($\times$ 100)				Size (5% level)				Power (5% level)			
	30	50	100	200	30	50	100	200	30	50	100	200	30	50	100	200
PB																
30	-5.15	-2.18	-0.58	-0.18	7.19	3.91	1.74	0.81	24.70	15.75	10.45	7.65	15.15	8.35	19.65	66.30
50	-5.34	-2.26	-0.61	-0.17	6.63	3.42	1.43	0.66	33.90	18.60	10.00	7.40	18.10	8.00	23.10	84.15
100	-5.08	-2.17	-0.58	-0.17	5.77	2.77	1.06	0.46	53.15	27.80	12.10	7.45	25.00	6.70	39.80	98.70
200	-5.04	-2.10	-0.57	-0.14	5.38	2.41	0.83	0.34	78.65	45.75	16.70	8.35	41.65	6.45	67.45	100.00
PMG																
30	-3.12	-1.14	-0.29	-0.10	7.47	3.89	1.75	0.80	39.40	23.85	14.40	8.50	35.35	23.25	31.45	74.25
50	-3.04	-1.16	-0.29	-0.08	6.09	3.18	1.42	0.65	41.20	25.35	14.65	8.20	35.00	25.15	41.80	90.30
100	-2.70	-1.09	-0.26	-0.08	4.56	2.28	0.97	0.44	45.85	29.50	15.10	8.80	36.25	25.95	63.45	99.70
200	-2.68	-1.05	-0.26	-0.06	3.68	1.74	0.70	0.32	57.45	34.05	16.25	9.95	34.95	31.95	88.30	100.00
Bias-corrected PB estimators																
Jackknife-corrected PB using $\kappa = 1/3$																
30	-2.31	-0.67	-0.08	-0.04	6.16	3.66	1.76	0.84	7.30	5.65	5.75	4.90	4.95	7.00	19.10	58.85
50	-2.37	-0.66	-0.10	-0.02	5.03	2.92	1.39	0.68	6.85	5.85	5.25	5.70	3.80	7.10	28.10	81.30
100	-2.14	-0.58	-0.08	-0.02	3.75	2.00	0.95	0.47	6.15	5.00	5.65	5.00	2.45	9.95	52.35	98.25
200	-2.14	-0.55	-0.06	0.00	3.03	1.42	0.64	0.33	7.60	5.15	4.40	5.50	1.50	15.70	84.10	100.00
Jackknife-corrected PB, using $\kappa = \hat{\kappa}_{NT}$																
30	-0.11	-0.04	0.00	-0.03	6.52	3.82	1.79	0.85	6.10	6.05	5.60	4.95	5.85	8.65	19.85	58.25
50	0.01	0.02	-0.02	-0.01	5.08	3.03	1.41	0.69	5.00	5.50	5.20	5.85	5.70	10.85	29.35	81.20
100	0.16	0.07	0.00	-0.02	3.53	2.04	0.96	0.47	4.50	5.10	5.50	4.90	8.45	16.95	54.20	98.10
200	0.09	0.06	0.00	0.00	2.45	1.38	0.65	0.33	4.75	4.80	4.85	5.55	12.75	31.10	85.45	100.00
Bias-corrected PB using stochastic simulations																
30	-1.71	-0.47	-0.06	-0.04	5.65	3.40	1.66	0.79	7.60	6.35	5.65	5.15	6.55	8.45	22.35	65.35
50	-1.73	-0.46	-0.06	-0.02	4.58	2.70	1.31	0.64	8.50	6.70	5.05	5.45	6.80	10.60	32.40	86.70
100	-1.53	-0.41	-0.05	-0.02	3.31	1.84	0.90	0.44	9.95	6.65	5.10	4.75	6.50	16.05	60.05	99.20
200	-1.53	-0.38	-0.04	0.00	2.54	1.30	0.61	0.31	13.50	6.90	5.20	5.80	6.65	27.85	88.75	100.00
Bias-corrected PMG estimators																
Jackknife-corrected PMG using $\kappa = 1/3$																
30	-1.35	-0.28	-0.03	-0.03	8.16	4.24	1.88	0.85	14.30	9.75	6.80	4.85	13.10	12.15	21.40	64.05
50	-1.11	-0.22	-0.04	-0.02	6.40	3.40	1.50	0.68	13.95	8.95	6.90	5.80	13.40	13.15	29.80	82.65
100	-0.85	-0.17	-0.02	-0.01	4.42	2.28	1.01	0.47	12.90	8.70	5.75	5.25	12.55	16.75	54.00	99.00
200	-0.86	-0.17	-0.02	0.00	3.08	1.56	0.71	0.33	11.55	6.75	6.95	6.60	13.00	24.55	82.55	100.00
Jackknife-corrected PMG using $\kappa = \hat{\kappa}_{NT}$																
30	-1.19	-0.19	-0.02	-0.03	8.30	4.31	1.89	0.85	14.50	9.75	6.85	4.65	13.55	12.45	21.50	63.75
50	-0.84	-0.14	-0.03	-0.02	6.56	3.46	1.51	0.68	14.05	9.15	6.65	5.75	14.15	14.05	29.45	82.90
100	-0.58	-0.10	-0.02	-0.01	4.52	2.31	1.01	0.47	12.95	8.70	5.80	5.15	13.80	17.55	53.40	98.95
200	-0.59	-0.11	-0.01	0.00	3.11	1.58	0.71	0.33	11.80	6.55	7.05	6.55	14.80	26.20	82.80	100.00
Bias-corrected PMG using stochastic simulations																
30	-2.08	-0.56	-0.09	-0.04	7.27	3.81	1.74	0.80	14.10	8.95	6.45	4.55	11.60	9.65	22.00	68.40
50	-1.97	-0.56	-0.08	-0.02	5.76	3.06	1.40	0.64	14.80	8.70	7.00	5.35	11.45	11.15	32.00	86.75
100	-1.66	-0.51	-0.07	-0.02	4.12	2.09	0.94	0.43	16.05	8.35	5.50	4.20	11.70	14.55	56.60	99.65
200	-1.66	-0.49	-0.07	-0.01	3.08	1.49	0.66	0.31	18.75	9.85	6.45	5.30	11.95	23.70	86.35	100.00

Notes: DGP is given by $\Delta y_{it} = c_i - \alpha_i (y_{i,t-1} - \beta x_{i,t-1}) + u_{y,it}$ and $\Delta x_{it} = u_{x,it}$, for $i = 1, 2, ..., n$, $T = 1, 2, ..., T$, with $\beta = 1$ and $\alpha_i \sim IIDU\,[0.2, 0.3]$. See Section 3.1 for complete description of the DGP. The pooled Bewley estimator is given by (10), with variance estimated using (18). PMG is the Pooled Mean Group estimator proposed by Pesaran, Shin, and Smith (1999). Bias-corrected versions of the PB estimator are described in Subsection 2.2. Bias-corrected versions of the PMG estimator are described in the online supplement. The size and power findings are computed using 5% nominal level and the reported power is the rejection frequency for testing the hypothesis $\beta = 0.98$.

Table 2: Estimated consumption function coefficients for OECD countries

	(1)	(2)	Bias-corrected estimators					
			Jackknife, $\kappa = 1/3$		Jackknife, $\hat{\kappa}$		Stochastic simul.	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	PMG*	PB**	PMG	PB	PMG	PB	PMG	PB
β_1: Income	.904	.912	.915	.926	.901	.926	.904	0.918
95% Conf. Int.	[.889,.919]	[.845,.980]	[.885,.945]	[.846,1.006]	[.877,.926]	[.848,1.005]	[.879,.929]	[.852,.984]
β_2: Inflation	-.466	-.134	-.403	-0.120	-.432	-.153	-.474	-0.126
95% Conf. Int.	[-.566,-.365]	[-.260,-.008]	[-.583,-.222]	[-.211,-.029]	[-.603,-.262]	[-.326,.020]	[-.637,-.310]	[-.320,-.067]

*PMG stand for Pooled Mean Group estimator, **PB stands for pooled Bewley estimator

Notes: This table revisits empirical application in Table 1 of Pesaran, Shin, and Smith (1999). Column (1) of this table reports the PMG estimates of long-run income elasticity (β_1) and inflation effect (β_2) coefficients and their 95% confidence intervals in the ARDL(1,1,1) consumption functions (23) for OECD countries using the dataset from Pesaran, Shin, and Smith (1999). Column (2) reports the PB estimator. Columns (3)-(8) report bias-corrected versions of the PMG and PB estimator. Jackknife bias correction using $\kappa = 1/3$ is reported in columns (3)-(4), jackknife bias correction using simulated value $\hat{\kappa}$ is reported in columns (5)-(6), and simulation-based bias correction is reported in columns (7)-(8). Description of bias correction methods is provided in Subsection 2.2 for PB estimator and in the online supplement for PMG estimator.

References

Bewley, R. A. (1979). The direct estimation of the equilibrium response in a linear dynamic model. *Economics Letters 3*, 357–361.

Breitung, J. (2005). A parametric approach to the estimation of cointegration vectors in panel data. *Econometric Reviews 24*, 151–173.

Chudik, A. and M. H. Pesaran (2013). Econometric analysis of high dimensional VARs featuring a dominant unit. *Econometric Reviews 32*, 592–649.

Chudik, A., M. H. Pesaran, and J.-C. Yang (2018). Half-panel jackknife fixed-effects estimation of linear panels with weakly exogenous regressors. *Journal of Applied Econometrics 33*(6), 816–836.

Davidson, R. and E. Flachaire (2008). The wild bootstrap, tamed at last. *Journal of Econometrics 146*, 162–169.

Dhaene, G. and K. Jochmans (2015). Split-panel jackknife estimation of fixed-effect models. *Review of Economic Studies 82*(3), 991–1030.

Mark, N. C. and D. Sul (2003). Cointegration vector estimation by panel DOLS and long-run money demand. *Oxford Bulletin of Economics and Statistics*.

Pedroni, P. (2001). Fully modified OLS for heterogeneous cointegrated panels. In B. Baltagi, T. Fomby, and R. C. Hill (Eds.), *Nonstationary Panels, Panel Cointegration, and Dynamic Panels, (Advances in Econometrics, Vol. 15)*, pp. 93–130. Emerald Group Publishing Limited, Bingley.

Pesaran, M. H., Y. Shin, and R. P. Smith (1999). Pooled mean group estimation of dynamic heterogeneous panels. *Journal of the American Statistical Association 94*, 621–634.

Phillips, P. C. B. and H. R. Moon (1999). Linear regression limit theory for nonstationary panel data. *Econometrica 67*, 1057–1011.

Wickens, M. R. and T. S. Breusch (1988). Dynamic specification, the long-run and the estimation of transformed regression models. *The Economic Journal 98*, 189–205.

A Appendix

This Appendix is organized in three sections. Section A.1 introduces some notations and definitions. Section A.2 presents lemmas and proofs needed for the proof of Theorem 1. Section A.3 presents proof of Theorem 1.

A.1 Notations and definitions

Define $\mathbf{C}_i\left(L\right) = \sum_{\ell=0}^{\infty} \mathbf{C}_{i\ell} L^\ell$ and $\mathbf{C}_i^*\left(L\right) = \sum_{\ell=0}^{\infty} \mathbf{C}_{i\ell}^* L^\ell$, where

$$
\begin{aligned}
\mathbf{C}_{i0} &= \mathbf{I}_2, \\
\mathbf{C}_{i\ell} &= \left(\mathbf{\Phi}_i - \mathbf{I}_2\right) \mathbf{\Phi}_i^{\ell-1}, \ \ell = 1, 2, ...,
\end{aligned}
$$

$$
\mathbf{\Phi}_i = \left(
\begin{array}{cc}
1 - \alpha_i & \alpha_i \beta \\
0 & 1
\end{array}
\right),
\tag{A.1}
$$

$$
\mathbf{C}_i(1) = \mathbf{C}_{i0} + \mathbf{C}_{i1} + ... = \lim_{\ell \to \infty} \mathbf{\Phi}_i^\ell = \left(
\begin{array}{cc}
0 & \beta \\
0 & 1
\end{array}
\right)
$$

and

$$
\begin{aligned}
\mathbf{C}_{i0}^* &= \mathbf{C}_{i0} - \mathbf{C}_i(1) = \left(
\begin{array}{cc}
1 & -\beta \\
0 & 0
\end{array}
\right) \\
\mathbf{C}_{i\ell}^* &= \mathbf{C}_{i,\ell-1}^* + \mathbf{C}_{i\ell} = \left(
\begin{array}{cc}
(1-\alpha_i)^\ell & -(1-\alpha_i)^\ell \beta \\
0 & 0
\end{array}
\right), \text{ for } \ell = 1, 2, ...
\end{aligned}
$$

Model (1)-(2) can be equivalently written as

$$
\mathbf{\Phi}_i\left(L\right) \mathbf{z}_{it} = \mathbf{c}_i + \mathbf{u}_{it},
$$

for $i = 1, 2, ..., n$ and $t = 1, 2, ..., T$, where $\mathbf{c}_i = (c_i, 0)'$,

$$
\mathbf{\Phi}_i\left(L\right) = \mathbf{I}_2 - \mathbf{\Phi}_i L,
\tag{A.2}
$$

and $\mathbf{I}_2$ is a 2×2 identity matrix. The lag polynomial $\mathbf{\Phi}_i\left(L\right)$ can be re-written in the following (error correcting) form

$$
\mathbf{\Phi}_i\left(L\right) = -\Pi_i L + (1 - L)\mathbf{I}_2,
\tag{A.3}
$$

where

$$
\Pi_i = -\left(\mathbf{I}_2 - \mathbf{\Phi}_i\right) = \left(
\begin{array}{cc}
-\alpha_i & \alpha_i \beta \\
0 & 0
\end{array}
\right).
\tag{A.4}
$$

A1

The VAR model (A.5) can be also rewritten in the following form

$$\mathbf{\Phi}_i\left(L\right)\left(\mathbf{z}_{it} - \boldsymbol{\mu}_i\right) = \mathbf{u}_{it}, \tag{A.5}$$

where $\mathbf{c}_i = -\Pi_i\boldsymbol{\mu}_i = (c_i, 0)'$, namely $c_i = \alpha_i\mu_{i,1} - \alpha_i\beta\mu_{i,2}$.

Using Granger representation theorem, the process $\mathbf{z}_{it}$ under the assumptions 1-3 has representation

$$
\begin{aligned}
y_{it} &= \mu_{yi} + \beta s_{it} + \sum_{\ell=0}^{\infty}\left(1-\alpha_i\right)^{\ell}\left(u_{y,i,t-\ell} - \beta u_{x,i,t-\ell}\right), &\tag{A.6}\\
x_{it} &= \mu_{xi} + s_{it}, &\tag{A.7}
\end{aligned}
$$

where

$$s_{it} = \sum_{\ell=1}^{t} u_{x,it}, \tag{A.8}$$

is the stochastic trend.

A.2 Lemmas: Statements and proofs

Lemma A.1 *Suppose Assumptions 2 and 3 hold, and consider* $\tilde{\mathbf{x}}_i = \left(\tilde{x}_{i,1}, \tilde{x}_{i,2}, ..., \tilde{x}_{i,T}\right)'$, *where* $\tilde{x}_{it} = x_{it} - \bar{x}_i$, $x_{it} = \sum_{s=1}^{t} u_{x,it}$, *and* $\bar{x}_i = T^{-1}\sum_{t=1}^{T} x_{it}$. *Then*

$$n^{-1}\sum_{i=1}^{n}\frac{\tilde{\mathbf{x}}_i'\tilde{\mathbf{x}}_i}{T^2} \rightarrow_p \omega_x^2 = \frac{\sigma_x^2}{6}, \text{ as } n, T \rightarrow \infty, \tag{A.9}$$

where $\sigma_x^2 = \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n}\sigma_{xi}^2$.

Proof. Recall $\mathbf{M}_\tau = \mathbf{I}_T - \boldsymbol{\tau}_T\boldsymbol{\tau}_T'$, where $\mathbf{I}_T$ is $T \times T$ identity matrix and $\boldsymbol{\tau}_T$ is $T \times 1$ vector of ones. Since $\tilde{\mathbf{x}}_i = \mathbf{M}_\tau\mathbf{x}_i$, and $\mathbf{M}_\tau$ is symmetric and idempotent ($\mathbf{M}_\tau'\mathbf{M}_\tau = \mathbf{M}_\tau = \mathbf{M}_\tau'$) we can write $\tilde{\mathbf{x}}_i'\tilde{\mathbf{x}}_i$ as $\tilde{\mathbf{x}}_i'\tilde{\mathbf{x}}_i = \mathbf{x}_i'\mathbf{M}_\tau'\mathbf{M}_\tau\mathbf{x}_i = \mathbf{x}_i'\mathbf{M}_\tau'\mathbf{x}_i = \tilde{\mathbf{x}}_i'\mathbf{x}_i$. Denote $S_{i,T} = \tilde{\mathbf{x}}_i'\mathbf{x}_i/T^2$. We have

$$n^{-1}\sum_{i=1}^{n}\frac{\tilde{\mathbf{x}}_i'\tilde{\mathbf{x}}_i}{T^2} = n^{-1}\sum_{i=1}^{n}S_{i,T} = n^{-1}\sum_{i=1}^{n}E\left(S_{i,T}\right) + n^{-1}\sum_{i=1}^{n}\left[S_{i,T} - E\left(S_{i,T}\right)\right]. \tag{A.10}$$

Consider $E\left(S_{i,T}\right)$ first. Noting that $\tilde{x}_{it} = \sum_{s=1}^{t}u_{x,it} - \bar{x}_i$, $\bar{x}_i = T^{-1}\sum_{s=1}^{T}\left(T - s + 1\right)u_{st}$, and

A2

$x_{it} = \sum_{s=1}^{t} u_{x,it}$, $S_{i,T}$ can be written as

$$
\begin{aligned}
S_{i,T} &= \frac{1}{T^2} \sum_{t=1}^{T} \tilde{x}_{it} x_{it} \\
&= \frac{1}{T^2} \sum_{t=1}^{T} \left[\left(\sum_{s=1}^{t} u_{x,is} \right)^2 - \bar{x}_i \sum_{s=1}^{t} u_{x,is} \right] \\
&= \frac{1}{T^2} \sum_{t=1}^{T} \left[\left(\sum_{s=1}^{t} u_{x,is} \right)^2 - \sum_{s=1}^{t} \frac{T-s+1}{T} u_{x,is} \cdot \sum_{s=1}^{t} u_{x,is} \right]
\end{aligned}
$$

Taking expectations, we obtain

$$
E(S_{i,T}) = \frac{\sigma_{xi}^2}{T^2} \sum_{t=1}^{T} \left[t - \sum_{s=1}^{t} \frac{T-s+1}{T} \right].
$$

Using $\sum_{s=1}^{t} \frac{T-s+1}{T} = \sum_{s=1}^{t} (1 - s/T + 1/T) = t - (t+1)t/(2T) + t/T$, we have

$$
E(S_{i,T}) = \frac{\sigma_{xi}^2}{T^2} \sum_{t=1}^{T} \left[t - t + \frac{(t+1)t}{2T} - \frac{t}{T} \right] = \frac{\sigma_{xi}^2}{T^2} \sum_{t=1}^{T} \frac{(t+1)t}{2T} - \frac{t}{T}.
$$

Finally, noting that $\sum_{t=1}^{T} (t+1)t = (T+2)(T+1)T/3$, and $\sum_{t=1}^{T} t = (T+1)T/2$, we obtain

$$
E(S_{i,T}) = \sigma_{xi}^2 \varkappa_T < K < \infty, \tag{A.11}
$$

for all $T > 0$, where

$$
\varkappa_T = \left[\frac{(T+2)(T+1)T}{6T^3} - \frac{(T+1)T}{2T^3} \right]. \tag{A.12}
$$

In addition, $\varkappa_T \to 1/6$, as $T \to \infty$, and

$$
\frac{1}{n} \sum_{i=1}^{n} E(S_{i,T}) = \varkappa_T \frac{1}{n} \sum_{i=1}^{n} \sigma_{xi}^2 \to \frac{\sigma_x^2}{6},
$$

as $n, T \to \infty$. This establishes the limit of the first term on the right side of (A.10). Consider the second term next. Since $E[S_{i,T} - E(S_{i,T})] = 0$, and $S_{i,T}$ is independent over i, we have

$$
E \left\{ n^{-1} \sum_{i=1}^{n} [S_{i,T} - E(S_{i,T})] \right\}^2 = \frac{1}{n^2} \sum_{i=1}^{n} E(S_{i,T}^2) - \frac{1}{n^2} \sum_{i=1}^{n} [E(S_{i,T})]^2.
$$

But it follows from (A.11) that there exist finite positive constant $K_1 < \infty$ (which does not depend on n, T) such that $[E(S_{i,T})]^2 < K_1$. In addition, due to existence of uniformly bounded fourth moments of $u_{x,it}$, it also can be shown that $E\left(S_{i,T}^2 \right) < K_2 < \infty$. Hence, $E\left\{ n^{-1} \sum_{i=1}^{n} [S_{i,T} - E(S_{i,T})] \right\}^2 =$

$O\left(n^{-1}\right)$, which implies $n^{-1}\sum_{i=1}^{n}\left[S_{i,T}-E\left(S_{i,T}\right)\right]\to_{p}0$, as $n,T\to\infty$. This completes the proof. ∎

Lemma A.2 *Suppose Assumptions 1-2 hold. Then there exists finite positive constant K that does not depend on i and/or T such that*

$$E\left(\frac{1}{T}\sum_{t=1}^{T}u_{x,it}\tilde{x}_{it}\right)^{\varrho}<K, \tag{A.13}$$

and

$$E\left(\frac{1}{T}\sum_{t=1}^{T}\Delta y_{it}\tilde{x}_{it}\right)^{\varrho}<K, \tag{A.14}$$

for $\varrho=4$, where $\tilde{x}_{it}=x_{it}-\bar{x}_{i}$, $x_{it}=\sum_{s=1}^{t}u_{x,it}$, $\bar{x}_{i}=T^{-1}\sum_{t=1}^{T}x_{it}$, and $\Delta y_{it}=\delta_{i}u_{x,it}+v_{it}-\alpha_{i}\sum_{\ell=1}^{\infty}(1-\alpha_{i})^{\ell-1}[v_{i,t-\ell}+(\delta_{i}-\beta)u_{x,i,t-\ell}]$.

Proof. Consider $\sum_{t=1}^{T}u_{it}\tilde{x}_{it}/T$ and $\varrho=2$ first, and note that $\tilde{x}_{it}=\sum_{s=1}^{t}u_{x,is}-\bar{x}_{i}$, where $\bar{x}_{i}=T^{-1}\sum_{s=1}^{T}(T-s+1)u_{x,is}$. We have

$$
\begin{aligned}
\left(\frac{1}{T}\sum_{t=1}^{T}u_{x,it}\tilde{x}_{it}\right)^{2} &= \frac{1}{T^{2}}\sum_{t=1}^{T}\sum_{t'=1}^{T}u_{x,it}u_{x,it'}\tilde{x}_{it}\tilde{x}_{it'} \\
&= \frac{1}{T^{2}}\sum_{t=1}^{T}\sum_{t'=1}^{T}u_{it}u_{it'}\left(\sum_{s=1}^{t}u_{x,is}-\bar{x}_{i}\right)\left(\sum_{s=1}^{t'}u_{x,is}-\bar{x}_{i}\right) \\
&= A_{i,T,1}+A_{i,T,2}-A_{i,T,3}-A_{i,T,4},
\end{aligned}
$$

where

$$
\begin{aligned}
A_{i,T,1} &= \frac{1}{T^{2}}\sum_{t=1}^{T}\sum_{t'=1}^{T}u_{x,it}u_{x,it'}\left(\sum_{s=1}^{t}u_{x,is}\right)\left(\sum_{s=1}^{t'}u_{x,is}\right) \\
A_{i,T,2} &= \frac{1}{T^{2}}\sum_{t=1}^{T}\sum_{t'=1}^{T}u_{x,it}u_{x,it'}\bar{x}_{i}^{2}, \\
A_{i,T,3} &= \frac{1}{T^{2}}\sum_{t=1}^{T}\sum_{t'=1}^{T}u_{x,it}u_{x,it'}\bar{x}_{i}\sum_{s=1}^{t}u_{x,is}, \\
A_{i,T,4} &= \frac{1}{T^{2}}\sum_{t=1}^{T}\sum_{t'=1}^{T}u_{x,it}u_{x,it'}\bar{x}_{i}\sum_{s=1}^{t'}u_{x,is}.
\end{aligned}
$$

Taking expectations and noting that $u_{x,it}$ is independent of $u_{x,it'}$ for any $t\neq t'$, we have

$$E\left(A_{i,T,1}\right)=\frac{1}{T^{2}}\left(\sum_{t=1}^{T}\sum_{t'=1}^{t-1}\sigma_{ix}^{2}+\sum_{t=1}^{T}E\left(u_{x,it}^{4}\right)+\sum_{t=1}^{T}\sum_{t'=1}^{t-1}\sigma_{ix}^{2}\right).$$

A4

Under Assumption 2, there exists a finite constant K that does not depend on i and/or t, such that $\sigma^2_{ix} < K$ and $E\left(u^4_{x,it}\right) < K$. Hence $|E\left(A_{i,T,1}\right)| < K$. Similarly, we can bound the remaining elements, $|E\left(A_{i,T,j}\right)| < K$, for $j = 2, 3, 4$. It now follows that $E\left(\frac{1}{T}\sum_{t=1}^{T} u_{x,it}\tilde{x}_{it}\right)^2 < K$, where the upper bound K does not depend on i or T. This establishes (A.13) hold for $\varrho = 2$. Sufficient condition for (A.13) to hold for $\varrho = 4$ are:[7] $E\left(A^2_{i,T,j}\right) < K$ for $j = 1, 2, 3, 4$. These conditions follow from uniformly bounded eights moments of $u_{x,it}$. This completes the proof of (A.13). Result (A.14) can be established in the same way by using the first difference of representation (A.6). ∎

Lemma A.3 *Suppose Assumptions 1-4 hold, and consider s_{iT} given by*

$$s_{iT} = \tilde{\mathbf{x}}'_i \Delta\tilde{\mathbf{Z}}_i \left(\Delta\tilde{\mathbf{Z}}'_i \mathbf{P}_i \Delta\tilde{\mathbf{Z}}_i\right)^{-1} \Delta\tilde{\mathbf{Z}}'_i \tilde{\mathbf{x}}_i, \tag{A.15}$$

where $\mathbf{P}_i$ is given by (12), and $\tilde{\mathbf{x}}_i$ and $\Delta\tilde{\mathbf{Z}}_i$ are defined below (8). Then,

$$n^{-1}\sum_{i=1}^{n} \frac{s_{iT}}{T^2} \to_p 0, \text{ as } n, T \to \infty. \tag{A.16}$$

Proof. Consider $s_{i,T}/T$, which can be written as

$$\frac{s_{iT}}{T} \leq \mathbf{a}'_{iT}\mathbf{B}^{-1}_{iT}\mathbf{a}_{iT}, \tag{A.17}$$

where

$$\mathbf{a}_{iT} = \frac{\Delta\tilde{\mathbf{Z}}'_i \tilde{\mathbf{x}}_i}{T} = \frac{\Delta\mathbf{Z}'_i \tilde{\mathbf{x}}_i}{T}, \tag{A.18}$$

and

$$\mathbf{B}_{iT} = \frac{\Delta\tilde{\mathbf{Z}}'_i \mathbf{P}_i \Delta\tilde{\mathbf{Z}}_i}{T}. \tag{A.19}$$

Using these notations, we have

$$E\left|\frac{1}{n}\sum_{i=1}^{n} \frac{s_{iT}}{T^2}\right| \leq \frac{1}{nT}\sum_{i=1}^{n} E\left|\mathbf{a}'_{iT}\mathbf{B}^{-1}_{iT}\mathbf{a}_{iT}\right|.$$

Using $\left|\mathbf{a}'_{iT}\mathbf{B}^{-1}_{iT}\mathbf{a}_{iT}\right| \leq \lambda^{-1}_{\min}(\mathbf{B}_{iT})\mathbf{a}'_{iT}\mathbf{a}_{iT}$, and Cauchy-Schwarz inequality, we obtain

$$E\left|\frac{1}{n}\sum_{i=1}^{n} \frac{s_{iT}}{T^2}\right| \leq \frac{1}{nT}\sum_{i=1}^{n} \sqrt{E\left[\left(\mathbf{a}'_{iT}\mathbf{a}_{iT}\right)^2\right]}\sqrt{E\left[\lambda^{-2}_{\min}(\mathbf{B}_{iT})\right]}.$$

Lemma A.2 implies the fourth moments of the individual elements of $\mathbf{a}_{i,T}$ are uniformly bounded in i and T, which is sufficient for $E\left[\left(\mathbf{a}'_{iT}\mathbf{a}_{iT}\right)^2\right] < K$. In addition, $E\left[\lambda^{-2}_{\min}(\mathbf{B}_{iT})\right] < K$ by Assumption

---

[7]For the cross-product terms, note that $|E\left(A_{i,T,j}A_{i,T,s}\right)| \leq \sqrt{E\left(A^2_{i,T,j}\right)}\sqrt{E\left(A^2_{i,T,s}\right)}$. Hence, $E\left(A^2_{i,T,j}\right) < K$, for $j = 1, 2, 3, 4$, is sufficient.

4. Hence, there exists $K < \infty$, which does not depend on (n, T) such that

$$E \left| n^{-1} \sum_{i=1}^{n} \frac{s_{iT}}{T^2} \right| < \frac{K}{T}, \tag{A.20}$$

and result (A.16) follows. ■

Lemma A.4 *Suppose Assumptions 1-4 hold. Then*

$$n^{-1} \sum_{i=1}^{n} \frac{\tilde{\mathbf{x}}_i' \mathbf{M}_i \tilde{\mathbf{x}}_i}{T^2} \to_p \omega_x^2 = \frac{\sigma_x^2}{6}, \text{ as } n, T \to \infty, \tag{A.21}$$

where $\sigma_x^2 = \lim_{n \to \infty} n^{-1} \sum_{i=1}^{n} \sigma_{xi}^2$, $\mathbf{M}_i$ is defined in (11) and $\tilde{\mathbf{x}}_i$ is defined below (8).

Proof. Noting that $\tilde{\mathbf{x}}_i$ is one of the column vectors of $\mathbf{H}_i$, we have $\mathbf{P}_i \tilde{\mathbf{x}}_i = \tilde{\mathbf{x}}_i$, and $\tilde{\mathbf{x}}_i' \mathbf{M}_i \tilde{\mathbf{x}}_i$ can be written as

$$\tilde{\mathbf{x}}_i' \mathbf{M}_i \tilde{\mathbf{x}}_i = \tilde{\mathbf{x}}_i' \tilde{\mathbf{x}}_i - s_{i,T}, \tag{A.22}$$

where $s_{i,T}$ is given by (A.15). Sufficient conditions for result (A.21) are:

$$n^{-1} \sum_{i=1}^{n} \frac{\tilde{\mathbf{x}}_i' \tilde{\mathbf{x}}_i}{T^2} \to_p \omega_x^2 = \frac{\sigma_x^2}{6}, \text{ as } n, T \to \infty, \tag{A.23}$$

and

$$n^{-1} \sum_{i=1}^{n} \frac{s_{i,T}}{T^2} \to_p 0, \text{ as } n, T \to \infty. \tag{A.24}$$

Condition (A.23) is established by Lemma A.1, and condition (A.24) is established by Lemma A.3. ■

Lemma A.5 *Let Assumptions 1-3 hold. Then*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\tilde{\mathbf{x}}_i' \tilde{\mathbf{v}}_i}{T \alpha_i} \to_d N \left(0, \omega_v^2 \right), \text{ as } n, T \to \infty, \tag{A.25}$$

where $\omega_v^2 = \lim_{n \to \infty} n^{-1} \sum_{i=1}^{n} \sigma_{xi}^2 \sigma_{vi}^2 / \left(6 \alpha_i^2 \right)$, and $\tilde{\mathbf{x}}_i$ and $\tilde{\mathbf{v}}_i$ are defined below (8).

Proof. Recall $\mathbf{M}_\tau = \mathbf{I}_T - \boldsymbol{\tau}_T \boldsymbol{\tau}_T'$, where $\mathbf{I}_T$ is $T \times T$ identity matrix and $\boldsymbol{\tau}_T$ is $T \times 1$ vector of ones. Since $\mathbf{M}_\tau' \mathbf{M}_\tau = \mathbf{M}_\tau'$, we have

$$\tilde{\mathbf{x}}_i' \tilde{\mathbf{v}}_i = \mathbf{x}_i' \mathbf{M}_\tau' \mathbf{M}_\tau \mathbf{v}_i = \mathbf{x}_i' \mathbf{M}_\tau' \mathbf{v}_i = \tilde{\mathbf{x}}_i' \mathbf{v}_i$$

Let $C_i = \frac{\sigma_{xi}\sigma_{vi}}{\alpha_i}$ and $Q_{i,T} = C_i^{-1} \frac{\tilde{\mathbf{x}}_i' \mathbf{v}_i}{T\alpha_i}$. We have $E(Q_{i,T}) = 0$, and (under independence of v_{it} over t and independence of v_{it} and $u_{x,it'}$ for any t, t')

$$E\left[\left(\frac{\tilde{\mathbf{x}}_i' \mathbf{v}_i}{T\alpha_i}\right)^2\right] = \frac{1}{T^2 \alpha_i^2} \sum_{t=1}^{T} E\left(\tilde{x}_{it}^2\right) E\left(v_{it}^2\right),$$

where $E\left(v_{it}^2\right) = \sigma_{vi}^2$. In addition, (A.11) established that $\frac{1}{T^2} \sum_{t=1}^{T} E\left(\tilde{x}_{it}^2\right) = \sigma_{xi}^2 \varkappa_T$, where $\varkappa_T$ is given by (A.12). Hence,

$$E\left[\left(\frac{\tilde{\mathbf{x}}_i' \mathbf{v}_i}{T\alpha_i}\right)^2\right] = \frac{\sigma_{vi}^2 \sigma_{xi}^2}{\alpha_i^2} \varkappa_T = C_i^2 \varkappa_T.$$

It follows that

$$E\left(Q_{i,T}^2\right) = \varkappa_T$$

where $\varkappa_T \to 1/6 < \infty$. Finite fourth moments of $u_{x,it}$ and v_{it} imply $Q_{i,T}^4$ is uniformly bounded in T, and therefore $Q_{i,T}^2$ is uniformly integrable in T. We can apply Theorem 3 of Phillips and Moon (1999) to obtain

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} C_i Q_{i,T} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\tilde{\mathbf{x}}_i' \mathbf{v}_i}{T\alpha_i} \to_d N\left(0, \omega_v^2\right), \text{ as } n, T \to \infty$$

where $\omega_v^2 = \lim_{n\to\infty} C_i^2 \varkappa_T = \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} \sigma_{xi}^2 \sigma_{vi}^2 / \left(6\alpha_i^2\right)$. ∎

Lemma A.6 *Suppose Assumptions 1-4 hold, and consider* $\mathbf{q}_{iT} = \alpha_i^{-1} \Delta\tilde{\mathbf{Z}}_i' \mathbf{P}_i \tilde{\mathbf{v}}_i / \sqrt{T}$. *Then,*

$$E\|\mathbf{q}_{iT}\|_2^4 < K, \tag{A.26}$$

and

$$|E(\mathbf{q}_{iT})| < \frac{K}{\sqrt{T}}. \tag{A.27}$$

Proof. Denote the individual elements of 2×1 vector $\mathbf{q}_{iT}$ as $q_{iT,j}$, $j = 1, 2$. Sufficient conditions for (A.26) to hold are

$$E(q_{iT,j})^4 < K, \text{ for } j = 1, 2. \tag{A.28}$$

We establish (A.26) for $j = 1$ first. We have

$$q_{iT,1} = \frac{\Delta\tilde{\mathbf{y}}_i' \mathbf{P}_i \tilde{\mathbf{v}}_i}{\alpha_i \sqrt{T}},$$

where $\Delta\mathbf{y}_i$ can be written as

$$\Delta\tilde{\mathbf{y}}_i = -\alpha_i \tilde{\boldsymbol{\xi}}_{i,-1} + \delta_i \Delta\tilde{\mathbf{x}}_i + \mathbf{v}_i, \tag{A.29}$$

where $\tilde{\boldsymbol{\xi}}_{i,-1} = \tilde{\mathbf{y}}_{i,-1} - \tilde{\mathbf{x}}_{i,-1}$. Note that $\mathbf{P}_i = \tilde{\mathbf{H}}_i \left(\tilde{\mathbf{H}}_i' \tilde{\mathbf{H}}_i\right)^{-1} \tilde{\mathbf{H}}_i'$ and $\tilde{\mathbf{H}}_i = (\tilde{\mathbf{y}}_{i,-1}, \tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_{i,-1})$. Hence

$\Delta\tilde{\mathbf{x}}_i'\mathbf{P}_i = \Delta\tilde{\mathbf{x}}_i'$ and $\tilde{\boldsymbol{\xi}}_{i,-1}'\mathbf{P}_i = \tilde{\boldsymbol{\xi}}_{i,-1}'$, since $\Delta\tilde{\mathbf{x}}_i$ and $\tilde{\boldsymbol{\xi}}_{i,-1}$ can be both obtained as a linear combinations of the column vectors of $\tilde{\mathbf{H}}_i$. Hence

$$q_{iT,1} = -\frac{\tilde{\boldsymbol{\xi}}_{i,-1}'\tilde{\mathbf{v}}_i}{\sqrt{T}} + \frac{\Delta\tilde{\mathbf{x}}_i'\tilde{\mathbf{v}}_i}{\alpha_i\sqrt{T}} + \frac{\tilde{\mathbf{v}}_i'\mathbf{P}_i\tilde{\mathbf{v}}_i}{\alpha_i\sqrt{T}} \equiv \varsigma_{a,iT} + \varsigma_{b,iT} + \varsigma_{c,iT}, \tag{A.30}$$

where we simplified notations by introducing $\varsigma_{a,iT} = -\tilde{\boldsymbol{\xi}}_{i,-1}'\tilde{\mathbf{v}}_i/\sqrt{T}$, $\varsigma_{b,iT} = \alpha_i^{-1}\Delta\tilde{\mathbf{x}}_i'\tilde{\mathbf{v}}_i/\sqrt{T}$ and $\varsigma_{c,iT} = \alpha_i^{-1}\Delta\tilde{\mathbf{v}}_i'\mathbf{P}_i\tilde{\mathbf{v}}_i/\sqrt{T}$ to denote the individual terms in the expression (A.30) for $q_{iT,1}$. Sufficient conditions for $E\left(q_{iT,1}^4\right) < K$ are $E\left(\varsigma_{s,iT}^4\right) < K$ for $s \in \{a, b, c\}$.

For $s = a$, we have

$$\varsigma_{a,iT} = -\frac{\tilde{\boldsymbol{\xi}}_{i,-1}'\tilde{\mathbf{v}}_i}{\sqrt{T}} = -\frac{1}{\sqrt{T}}\sum_{i=1}^{T}\left(\xi_{i,t-1} - \bar{\xi}_{i,-1}\right)\left(v_{it} - \bar{v}_i\right) = -\frac{1}{\sqrt{T}}\sum_{i=1}^{T}\xi_{i,t-1}v_{it} + \sqrt{T}\bar{\xi}_{i,-1}\bar{v}_i,$$

where $\bar{\xi}_{i,-1} = T^{-1}\sum_{t=1}^{T}\xi_{i,t-1}$, and

$$\begin{aligned}
\xi_{it} &= \sum_{\ell=0}^{\infty}(1-\alpha_i)^{\ell}\left(u_{y,i,t-\ell} - \beta u_{x,i,t-\ell}\right) \\
&= \sum_{\ell=0}^{\infty}(1-\alpha_i)^{\ell}\left(\delta_i - \beta\right)u_{x,i,t-\ell} + \sum_{\ell=0}^{\infty}(1-\alpha_i)^{\ell}v_{it}.
\end{aligned}$$

Noting that $\sup_i(1-\alpha_i) < 1$ under Assumption 1, and fourth moments of $u_{x,i,t}$ and eights moments of v_{it} are bounded, we obtain

$$E\left[\left(\frac{1}{\sqrt{T}}\sum_{i=1}^{T}\xi_{i,t-1}v_{it}\right)^4\right] \le K,$$

and

$$T^2 \cdot E\left(\bar{\xi}_{i,-1}^4\bar{v}_i^4\right) \le K,$$

which are sufficient conditions for $E\left(\varsigma_{a,iT}^4\right) \le K$.

For $s = b$, we have

$$\varsigma_{b,iT} = \frac{\Delta\tilde{\mathbf{x}}_i'\tilde{\mathbf{v}}_i}{\alpha_i\sqrt{T}} = \frac{1}{\alpha_i\sqrt{T}}\sum_{t=1}^{T}\left(u_{x,it} - \bar{u}_{x,i}\right)\left(v_{it} - \bar{v}_i\right) = \frac{1}{\alpha_i\sqrt{T}}\sum_{t=1}^{T}u_{x,it}v_{it} - \frac{\sqrt{T}}{\alpha_i}\bar{u}_{x,i}\bar{v}_i.$$

Using Assumption 2, we obtain the following upper bound

$$E\left(\varsigma_{b,iT}^4\right) \le |\alpha_i^{-4}|\frac{1}{T}\sum_{t=1}^{T}E\left(u_{x,it}^4\right)E\left(v_{it}^4\right) + |\alpha_i^{-4}|TE\left(\bar{u}_{x,i}^4\right)E\left(\bar{v}_i^4\right) \le K, \tag{A.31}$$

A8

where $\left|\alpha_i^{-4}\right| < K$, $E\left(u_{x,it}^4\right) < K$, $E\left(v_{it}^4\right) < K$, $E\left(\bar{u}_{x,i}^4\right) < K/T^2$, and $E\left(\bar{v}_i^4\right) < K/T^2$.

For $s = c$, we have

$$\varsigma_{c,iT} = \frac{\tilde{\mathbf{v}}_i' \mathbf{P}_i \tilde{\mathbf{v}}_i}{\alpha_i \sqrt{T}} = \frac{\Delta \tilde{\mathbf{v}}_i' \tilde{\mathbf{H}}_i \left(\tilde{\mathbf{H}}_i' \tilde{\mathbf{H}}_i\right)^{-1} \tilde{\mathbf{H}}_i' \tilde{\mathbf{v}}_i}{\alpha_i \sqrt{T}}$$

Consider $\tilde{\mathbf{H}}_i^* = \left(\tilde{\mathbf{x}}_i, \Delta \tilde{\mathbf{x}}_i, \tilde{\boldsymbol{\xi}}_{i,-1}\right)$ and note that

$$\tilde{\mathbf{H}}_i = (\tilde{\mathbf{y}}_{i,-1}, \tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_{i,-1}) = \mathbf{B}^* \tilde{\mathbf{H}}_i^*,$$

where

$$\mathbf{B}^* = \begin{pmatrix} -\beta & \beta & 1 \\ 1 & 0 & 0 \\ -1 & 1 & 0 \end{pmatrix}$$

is nonsingular (for any β). Hence $\mathbf{P}_i = \tilde{\mathbf{H}}_i \left(\tilde{\mathbf{H}}_i' \tilde{\mathbf{H}}_i\right)^{-1} \tilde{\mathbf{H}}_i' = \tilde{\mathbf{H}}_i^* \left(\tilde{\mathbf{H}}_i^{*\prime} \tilde{\mathbf{H}}_i^*\right)^{-1} \tilde{\mathbf{H}}_i^{*\prime}$, and we can write $\varsigma_{c,iT}$ as

$$\varsigma_{c,iT} = \frac{\tilde{\mathbf{v}}_i' \tilde{\mathbf{H}}_i^* \left(\tilde{\mathbf{H}}_i^{*\prime} \tilde{\mathbf{H}}_i^*\right)^{-1} \tilde{\mathbf{H}}_i^{*\prime} \tilde{\mathbf{v}}_i}{\alpha_i \sqrt{T}}.$$

Consider the scaling matrix

$$\mathbf{A}_T = \begin{pmatrix} T^{-1} & 0 & 0 \\ 0 & T^{-1/2} & 0 \\ 0 & 0 & T^{-1/2} \end{pmatrix}. \tag{A.32}$$

We have

$$\varsigma_{c,iT} = \frac{1}{\alpha_i \sqrt{T}} \tilde{\mathbf{v}}_i' \tilde{\mathbf{H}}_i^* \mathbf{A}_T \left(\mathbf{A}_T \tilde{\mathbf{H}}_i^{*\prime} \tilde{\mathbf{H}}_i^* \mathbf{A}_T\right)^{-1} \mathbf{A}_T \tilde{\mathbf{H}}_i^{*\prime} \tilde{\mathbf{v}}_i \geq 0.$$

Using the inequality $\mathbf{x}' \mathbf{A}^{-1} \mathbf{x} \leq \lambda_{\min}(\mathbf{A}) \|\mathbf{x}\|^2$, we have

$$0 \leq \varsigma_{c,iT} \leq \frac{1}{\alpha_i \sqrt{T}} \lambda_{\min}^{-1}\left(\mathbf{A}_T \tilde{\mathbf{H}}_i^{*\prime} \tilde{\mathbf{H}}_i^* \mathbf{A}_T\right) \left\|\mathbf{A}_T \tilde{\mathbf{H}}_i^{*\prime} \tilde{\mathbf{v}}_i\right\|_2^2.$$

Using Cauchy-Schwarz inequality, we obtain

$$E\left(\varsigma_{c,iT}^4\right) \leq \frac{1}{\alpha_i^4 T^2} \sqrt{E\left[\lambda_{\min}^{-4}\left(\mathbf{A}_T \tilde{\mathbf{H}}_i^{*\prime} \tilde{\mathbf{H}}_i^* \mathbf{A}_T\right)\right]} \sqrt{E\left\|\mathbf{A}_T \tilde{\mathbf{H}}_i^{*\prime} \tilde{\mathbf{v}}_i\right\|_2^8}.$$

But $\alpha_i^{-4} < K$ under Assumption 1, and $E\left[\lambda_{\min}^{-4}\left(\mathbf{A}_T \tilde{\mathbf{H}}_i^{*\prime} \tilde{\mathbf{H}}_i^* \mathbf{A}_T\right)\right] < K$ under Assumption 4. It follows

$$E\left(\varsigma_{c,iT}^4\right) \leq \frac{K}{T^2} \sqrt{E\left\|\mathbf{A}_T \tilde{\mathbf{H}}_i^{*\prime} \tilde{\mathbf{v}}_i\right\|_2^8}.$$

A9

Let $\mathbf{A}_T\tilde{\mathbf{H}}_i^{*\prime}\tilde{\mathbf{v}}_i \equiv \mathbf{h}_{viT}$ and consider the individual elements of $\mathbf{h}_{viT}$, denoted as $h_{viT,j}$ for $j = 1, 2, 3$,

$$\mathbf{h}_{viT} = \mathbf{A}_T\tilde{\mathbf{H}}_i^{*\prime}\tilde{\mathbf{v}}_i = \begin{pmatrix} h_{viT,1} \\ h_{viT,2} \\ h_{viT,3} \end{pmatrix} = \begin{pmatrix} \frac{1}{T}\sum_{t=1}^{T} \tilde{x}_{it}\tilde{v}_{it} \\ \frac{1}{\sqrt{T}}\sum_{t=1}^{T} \tilde{u}_{it}\tilde{v}_{it} \\ \frac{1}{\sqrt{T}}\sum_{t=1}^{T} \bar{\xi}_{i,t-1}\tilde{v}_{it} \end{pmatrix}.$$

Under Assumption 2, it can be shown that

$$E\left(h_{viT,j}^{8}\right) < K, \text{ for } j = 1, 2, 3,$$

which is sufficient for $E\left\|\mathbf{A}_T\tilde{\mathbf{H}}_i^{*\prime}\tilde{\mathbf{v}}_i\right\|_2^8 < K$. It follows that

$$E\left(\varsigma_{c,iT}^4\right) < \frac{K}{T^2}. \tag{A.33}$$

This completes the proof of (A.26) for $j = 1$.

Consider next (A.26) for $j = 2$, and note $q_{iT,2}$ is the same as $\varsigma_{b,iT}$, namely

$$q_{iT,2} = \frac{\Delta\tilde{\mathbf{x}}_i'\mathbf{P}_i\tilde{\mathbf{v}}_i}{\alpha_i\sqrt{T}} = \frac{\Delta\tilde{\mathbf{x}}_i'\tilde{\mathbf{v}}_i}{\alpha_i\sqrt{T}} = \varsigma_{b,iT}.$$

But $E\left(\varsigma_{b,iT}^4\right) < K$, see (A.31). This completes the proof of (A.26).

We establish (A.27) next. As before we consider the individual elements of 2×1 vector $\mathbf{q}_{iT}$, denoted as $q_{iT,s}$ for $s = 1, 2$, separately. For $s = 1$ we have (using the individual terms in expression (A.30))

$$|E\left(q_{iT,1}\right)| = \left|E\left(-\frac{\bar{\xi}_{i,-1}'\tilde{\mathbf{v}}_i}{\sqrt{T}} + \frac{\Delta\tilde{\mathbf{x}}_i'\tilde{\mathbf{v}}_i}{\alpha_i\sqrt{T}} + \frac{\Delta\tilde{\mathbf{v}}_i'\mathbf{P}_i\tilde{\mathbf{v}}_i}{\alpha_i\sqrt{T}}\right)\right| \leq |E\left(\varsigma_{a,iT}\right)| + |E\left(\varsigma_{b,iT}\right)| + |E\left(\varsigma_{c,iT}\right)|. \tag{A.34}$$

For the first term in (A.34), we obtain

$$|E\left(\varsigma_{a,iT}\right)| = \left|E\left[-\frac{\bar{\xi}_{i,-1}'\tilde{\mathbf{v}}_i}{\sqrt{T}} = -\frac{1}{\sqrt{T}}\sum_{i=1}^{T}\left(\xi_{i,t-1} - \bar{\xi}_{i,-1}\right)\left(v_{it} - \bar{v}_i\right)\right]\right|$$

$$\leq \frac{1}{\sqrt{T}}\sum_{i=1}^{T}\left|E\left(\xi_{i,t-1}v_{it}\right)\right| + \sqrt{T}E\left|\bar{\xi}_{i,-1}\bar{v}_i\right|.$$

But $E\left(\xi_{i,t-1}v_{it}\right) = 0$ and $E\left|\bar{\xi}_{i,-1}\bar{v}_i\right| < K/T$ under Assumptions 1-2. Hence,

$$|E\left(\varsigma_{a,iT}\right)| \leq \frac{K}{\sqrt{T}}.$$

For the second term in (A.34), we obtain

$$
\begin{aligned}
|E\left(\varsigma_{b,iT}\right)| &= \left| E\left(\frac{1}{\alpha_i \sqrt{T}} \sum_{t=1}^{T} u_{x,it} v_{it} - \frac{\sqrt{T}}{\alpha_i} \bar{u}_{x,i} \bar{v}_i \right) \right| \\
&\leq K \frac{1}{\sqrt{T}} \sum_{t=1}^{T} |E\left(u_{x,it} v_{it}\right)| + K\sqrt{T}\, |E\left(\bar{u}_{x,i} \bar{v}_i\right)|\,.
\end{aligned}
$$

But $E\left(u_{x,it} v_{it}\right) = 0$ and $E\left(\bar{u}_{x,i} \bar{v}_i\right) = 0$ under Assumption 2. Hence

$$
|E\left(\varsigma_{b,iT}\right)| = 0.
$$

Finally, for the last term we note that

$$
|E\left(\varsigma_{c,iT}\right)| \leq E\,|\varsigma_{c,iT}| \leq \sqrt{E\left(\varsigma_{c,iT}^2\right)},
$$

and using result (A.33), we obtain

$$
|E\left(\varsigma_{c,iT}\right)| < \frac{K}{\sqrt{T}}.
$$

It now follows that $|E\left(q_{iT,1}\right)| < K/\sqrt{T}$, as desired.

Consider $|E\left(q_{iT,s}\right)|$ for $s = 2$ next. We have

$$
|E\left(q_{iT,2}\right)| = |E\left(\varsigma_{b,iT}\right)| = 0.
$$

This completes the proof of result (A.27).   ∎

Lemma A.7 *Let Assumptions 1-4 hold, and consider $\mathbf{B}_{iT}$ defined by (A.19). Then we have*

$$
T^{\alpha/2}\,\|\mathbf{B}_{iT} - \mathbf{B}_i\| \to_p 0 \text{ as } T \to \infty, \text{ for any } \alpha < 1/2, \tag{A.35}
$$

where

$$
\mathbf{B}_i = \operatorname*{plim}_{T\to\infty} \mathbf{B}_{iT} = \begin{pmatrix} \alpha_i^2 E\left(\xi_{it}^2\right) + \delta_i^2 \sigma_{xi}^2 & \delta_i \sigma_{xi}^2 \\ \delta_i \sigma_{xi}^2 & \sigma_{xi}^2 \end{pmatrix}, \tag{A.36}
$$

and $\xi_{it} = \sum_{\ell=0}^{\infty} (1 - \alpha_i)^\ell \left(u_{y,i,t-\ell} - \beta u_{x,i,t-\ell}\right)$.

Proof. We have

$$
\mathbf{B}_{iT} = \frac{\Delta\tilde{\mathbf{Z}}_i' \mathbf{P}_i \Delta\tilde{\mathbf{Z}}_i}{T} = \frac{1}{T} \begin{pmatrix} \Delta\tilde{\mathbf{y}}_i' \mathbf{P}_i \Delta\tilde{\mathbf{y}}_i & \Delta\tilde{\mathbf{y}}_i' \mathbf{P}_i \Delta\tilde{\mathbf{x}}_i \\ \Delta\tilde{\mathbf{x}}_i' \mathbf{P}_i \Delta\tilde{\mathbf{y}}_i & \Delta\tilde{\mathbf{x}}_i' \mathbf{P}_i \Delta\tilde{\mathbf{x}}_i \end{pmatrix} = \begin{pmatrix} b_{iT,11} & b_{iT,12} \\ b_{iT,21} & b_{iT,22} \end{pmatrix}.
$$

A11

Consider the element $b_{iT,22}$ first. Since $\Delta\tilde{\mathbf{x}}_i'\mathbf{P}_i = \Delta\tilde{\mathbf{x}}_i'$, and $\Delta\tilde{x}_{it} = u_{x,it} - \bar{u}_{x,i}$, we have

$$b_{iT,12} = \left(\frac{1}{T}\sum_{t=1}^{T} u_{x,it}^2\right) - \bar{u}_{x,i}^2.$$

Under Assumption 2, $u_{x,it} \sim IID\left(0, \sigma_{xi}^2\right)$ with finite fourth order moments, and therefore

$$T^\alpha \left(\frac{1}{T}\sum_{t=1}^{T} u_{x,it}^2 - \sigma_{xi}^2\right) \xrightarrow{p} 0, \text{ for any } \alpha < 1/2.$$

In addition, $E\left(\bar{u}_{x,i}^2\right) < K/T$, which implies $T^\alpha \bar{u}_{x,i}^2 \xrightarrow{p} 0$, for any $\alpha < 1/2$. It follows

$$T^\alpha \left(b_{iT,22} - \sigma_{xi}^2\right) \xrightarrow{p} 0, \text{ for any } \alpha < 1/2. \tag{A.37}$$

Consider the element $b_{iT,11}$ next. We will use similar arguments as in the proof of Lemma A.6. In particular, $\Delta\tilde{\mathbf{y}}_i$ can be written as in (A.29), and, since $\mathbf{P}_i\tilde{\boldsymbol{\xi}}_{i,-1} = \tilde{\boldsymbol{\xi}}_{i,-1}$ and $\mathbf{P}_i\Delta\tilde{\mathbf{x}}_i = \Delta\tilde{\mathbf{x}}_i$, we have

$$b_{i,T,11} = \frac{\Delta\tilde{\mathbf{y}}_i'\mathbf{P}_i\Delta\tilde{\mathbf{y}}_i}{T} = \zeta_{aa,iT} + \zeta_{bb,iT} + \zeta_{cc,iT} + 2\zeta_{ab,iT} + 2\zeta_{ac,iT} + 2\zeta_{bc,iT}, \tag{A.38}$$

where

$$\zeta_{aa,iT} = \alpha_i^2 \frac{\tilde{\boldsymbol{\xi}}_{i,-1}'\tilde{\boldsymbol{\xi}}_{i,-1}}{T},$$
$$\zeta_{bb,iT} = \delta_i^2 \frac{\Delta\tilde{\mathbf{x}}_i'\Delta\tilde{\mathbf{x}}_i}{T},$$
$$\zeta_{cc,iT} = \frac{\tilde{\mathbf{v}}_i'\mathbf{P}_i\tilde{\mathbf{v}}_i}{T}$$

and the cross-product terms are

$$\zeta_{ab,iT} = \alpha_i\delta_i \frac{\tilde{\boldsymbol{\xi}}_{i,-1}'\Delta\tilde{\mathbf{x}}_i}{T}$$
$$\zeta_{ac,iT} = \alpha_i \frac{\tilde{\boldsymbol{\xi}}_{i,-1}'\tilde{\mathbf{v}}_i}{T}, \text{ and}$$
$$\zeta_{bc,iT} = \delta_i \frac{\Delta\tilde{\mathbf{x}}_i'\tilde{\mathbf{v}}_i}{T}.$$

We consider these individual terms ζ next. Note that

$$\xi_{it} = \sum_{\ell=0}^{\infty}(1-\alpha_i)^\ell \left(u_{y,i,t-\ell} - \beta u_{x,i,t-\ell}\right)$$
$$= \sum_{\ell=0}^{\infty}(1-\alpha_i)^\ell v_{i,t-\ell} + (\delta_i - \beta)\sum_{\ell=0}^{\infty}(1-\alpha_i)^\ell u_{x,i,t-\ell},$$

A12

where $\sup_i (1 - \alpha_i) < 1$ under Assumption 1, and innovations v_{it} and u_{xit} have finite fourth order moments under Assumption 2. Hence, $T^\alpha \left[T^{-1} \sum_{t=1}^T \xi_{i,t-1}^2 - E\left(\xi_{i,t-1}^2\right) \right] \to_p 0$, $E\left(\bar\xi_{i,-1}^2\right) < K/T$, and we obtain

$$T^\alpha \left[\zeta_{aa,iT} - \alpha_i^2 E\left(\bar\xi_{i,-1}^2\right) \right] \to_p 0, \text{ for any } \alpha < 1/2. \tag{A.39}$$

Noting that $\zeta_{bb,iT} = \delta_i^2 b_{i,T,12}$, and using result (A.37), we have

$$T^\alpha \left[\zeta_{bb,iT} - \delta_i^2 \sigma_{xi}^2 \right] \to_p 0, \text{ for any } \alpha < 1/2. \tag{A.40}$$

Consider $\zeta_{cc,iT}$ and note that $\zeta_{cc,iT} = \frac{\alpha_i}{\sqrt{T}} \varsigma_{c,iT}$, where $\varsigma_{c,iT} = \alpha_i^{-1} \Delta \tilde{\mathbf{v}}_i' \mathbf{P}_i \tilde{\mathbf{v}}_i / \sqrt{T}$ was introduced in (A.30) in proof of Lemma A.6. But $E\left(\varsigma_{c,iT}^2\right) < \frac{K}{T}$ by (A.33), and it follows

$$T^\alpha \zeta_{cc,iT} \to_p 0, \text{ for any } \alpha < 1/2. \tag{A.41}$$

Using similar arguments, we obtain for the cross-product terms,

$$T^\alpha \zeta_{ab,iT} \to_p 0, \ T^\alpha \zeta_{ac,iT} \to_p 0, \text{ and } T^\alpha \zeta_{bc,iT} \to_p 0, \text{ for any } \alpha < 1/2, \text{ as } T \to \infty. \tag{A.42}$$

Using (A.39)-(A.42) in (A.38), we obtain

$$T^\alpha \left(b_{i,T,11} - \alpha_i^2 E\left(\xi_{it}^2\right) - \delta_i^2 \sigma_{xi}^2 \right) \xrightarrow{p} 0, \text{ for any } \alpha < 1/2. \tag{A.43}$$

Using the same arguments for the last term $b_{i,T,12} = b_{i,T,21}$, we obtain

$$T^\alpha \left(b_{i,T,12} - \delta_i \sigma_{xi}^2 \right) \xrightarrow{p} 0, \text{ for any } \alpha < 1/2.$$

This completes the proof of (A.35). ∎

Lemma A.8 *Let Assumptions 1-4 hold, and consider $\mathbf{B}_{iT}$ defined by (A.19) and $\mathbf{B}_i = \text{plim}_{T \to \infty} \mathbf{B}_{iT}$ defined by (A.36). Then we have*

$$T^{\alpha/2} \left\| \mathbf{B}_{iT}^{-1} - \mathbf{B}_i^{-1} \right\| \to_p 0 \text{ as } T \to \infty, \text{ for any } \alpha < 1/2. \tag{A.44}$$

Proof. This proof closely follows proof of Lemma A.8 in Chudik and Pesaran (2013). Let $p = \left\| \mathbf{B}_i^{-1} \right\|$, $q = \left\| \mathbf{B}_{iT}^{-1} - \mathbf{B}_i^{-1} \right\|$, and $r = \left\| \mathbf{B}_{iT} - \mathbf{B}_i \right\|$. We suppressed subscripts i, T to simplify the notations, but it is understood that the terms p, q, r depend on (i, T). Using the triangle inequality

and the submultiplicative property of matrix norm $\|.\|$, we have

$$
\begin{aligned}
q &= \left\| \mathbf{B}_{iT}^{-1} \left(\mathbf{B}_i - \mathbf{B}_{iT} \right) \mathbf{B}_i^{-1} \right\|, \\
&\leq \left\| \mathbf{B}_{iT}^{-1} \right\| rp, \\
&\leq \left\| \left(\mathbf{B}_{iT}^{-1} - \mathbf{B}_i^{-1} \right) + \mathbf{B}_i^{-1} \right\| rp, \\
&\leq \left(p + q \right) rp.
\end{aligned}
$$

Subtracting rpq from both sides and multiplying by $T^{\alpha/2}$, we have, for any $\alpha < 1/2$,

$$
(1 - rp) \left(T^{\alpha/2} q \right) \leq p^2 \left(T^{\alpha/2} r \right). \tag{A.45}
$$

Note that $T^{\alpha/2} r \xrightarrow{p} 0$ by Lemma A.7, and $|p| < K$ since $\mathbf{B}_i$ is invertible and $\lambda_{\min} (\mathbf{B}_i)$ is bounded away from zero[8]. It follows

$$
(1 - rp) \xrightarrow{p} 1, \tag{A.46}
$$

and

$$
p^2 \left(T^{\alpha/2} r \right) \xrightarrow{p} 0. \tag{A.47}
$$

(A.45)-(A.47) imply $T^{\alpha/2} q \xrightarrow{p} 0$. This establishes result (A.44). ∎

Lemma A.9 *Let Assumptions 1-4 hold, and consider ξ_{iT} defined by*

$$
\xi_{iT} = \frac{\tilde{\mathbf{x}}_i' \Delta \tilde{\mathbf{Z}}_i}{T} \left(\frac{\Delta \tilde{\mathbf{Z}}_i' \mathbf{P}_i \Delta \tilde{\mathbf{Z}}_i}{T} \right)^{-1} \frac{\Delta \tilde{\mathbf{Z}}_i' \mathbf{P}_i \tilde{\mathbf{v}}_i}{\alpha_i \sqrt{T}}.
$$

where $\mathbf{P}_i$ is given by (12), and $\tilde{\mathbf{x}}_i$ and $\Delta \tilde{\mathbf{Z}}_i$ are defined below (8). Then

$$
\frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \xi_{iT} \to_p 0, \tag{A.48}
$$

as $n, T \to \infty$ such that $\sup_{n,T} \sqrt{n} / T^{1-\epsilon} < K$ for some $\epsilon > 0$.

Proof. Term ξ_{iT} can be written as

$$
\xi_{iT} = \mathbf{a}_{iT}' \mathbf{B}_{iT}^{-1} \mathbf{q}_{iT}, \tag{A.49}
$$

where $\mathbf{a}_{iT}$ is given by (A.18), $\mathbf{B}_{iT}$ is given by (A.19), and

$$
\mathbf{q}_{iT} = \frac{\Delta \tilde{\mathbf{Z}}_i' \mathbf{P}_i \tilde{\mathbf{v}}_i}{\alpha_i \sqrt{T}}. \tag{A.50}
$$

---

[8] This follows from observing that both σ_{xi}^2 and $E \left(\xi_{it}^2 \right)$ as well as α_i^2 in (A.36) are bounded away from zero.

We have

$$\frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \xi_{i,T} = \frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \mathbf{a}_{iT}' \left(\mathbf{B}_{iT}^{-1} - \mathbf{B}_i^{-1} \right) \mathbf{q}_{iT} + \frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \mathbf{a}_{iT}' \mathbf{B}_i^{-1} \mathbf{q}_{iT}. \tag{A.51}$$

Consider the two terms on the right side of (A.51) in turn. Lemma A.2 established fourth moments of $\mathbf{a}_{iT}$ are bounded, which is sufficient for $\|\mathbf{a}_{iT}\| = O_p(1)$. Result (A.26) of Lemma A.6 established second moments of individual elements of $\mathbf{q}_{iT}$ are bounded, which is sufficient for $\|\mathbf{q}_{iT}\| = O_p(1)$. In addition, Lemma A.8 established

$$T^{\alpha/2} \left\| \mathbf{B}_{iT}^{-1} - \mathbf{B}_i^{-1} \right\| \to_p 0 \text{ as } T \to \infty, \text{ for any } \alpha < 1/2.$$

Let $\epsilon = (1 - \alpha)/2$ so that $1 - \epsilon = 1/2 + \alpha/2$. Then we obtain

$$\frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \mathbf{a}_{iT}' \left(\mathbf{B}_{iT}^{-1} - \mathbf{B}_i^{-1} \right) \mathbf{q}_{iT} = \frac{\sqrt{n}}{\sqrt{T}} \frac{1}{T^{\alpha/2}} \frac{1}{n} \sum_{i=1}^{n} \mathbf{a}_{iT}' \left[T^{\alpha/2} \left(\mathbf{B}_{iT}^{-1} - \mathbf{B}_i^{-1} \right) \right] \mathbf{q}_{iT}$$

$$\leq \frac{\sqrt{n}}{T^{1-\epsilon}} \left(\frac{1}{n} \sum_{i=1}^{n} \|\mathbf{a}_{iT}\| \left(T^{\alpha/2} \left\| \mathbf{B}_{iT}^{-1} - \mathbf{B}_i^{-1} \right\| \right) \|\mathbf{q}_{iT}\| \right) \to_p 0 \tag{A.52}$$

as $n, T \to \infty$ such that $\sup_{n,T} \frac{\sqrt{n}}{T^{1-\epsilon}} < K$ for some $\epsilon > 0$.

Consider next the second term on the right side of (A.51). Let $\mu_{iT}^* = E\left(\mathbf{a}_{iT}' \mathbf{B}_i^{-1} \mathbf{q}_{iT} \right)$, and consider the variance of $(nT)^{-1/2} \sum_{i=1}^{n} \mathbf{a}_{iT}' \mathbf{B}_i^{-1} \mathbf{q}_{iT}$. By independence of $\mathbf{a}_{iT}' \mathbf{B}_i^{-1} \mathbf{q}_{iT}$ across i,

$$Var\left(\frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \mathbf{a}_{iT}' \mathbf{B}_i^{-1} \mathbf{q}_{iT} \right) = \frac{1}{nT} \sum_{i=1}^{n} Var\left(\mathbf{a}_{iT}' \mathbf{B}_i^{-1} \mathbf{q}_{iT} \right)$$

$$\leq \frac{1}{nT} \sum_{i=1}^{n} E\left(\mathbf{a}_{iT}' \mathbf{B}_i^{-1} \mathbf{q}_{iT} \right)^2. \tag{A.53}$$

Denoting individual elements of $\mathbf{B}_i^{-1}$ as $b_{i,sj}^-$, individual elements of $\mathbf{a}_{iT}$ as $a_{iT,j}$, and individual elements of $\mathbf{q}_{iT}$ as $q_{iT,s}$, for $s, j = 1, 2$, we have

$$\mathbf{a}_{iT}' \mathbf{B}_i^{-1} \mathbf{q}_{iT} = \sum_{s=1}^{2} \sum_{j=1}^{2} b_{i,sj}^- a_{iT,s} q_{iT,j}$$

$$= b_{i,11}^- a_{iT,1} q_{iT,1} + b_{i,21}^- a_{iT,2} q_{iT,1} + b_{i,12}^- a_{iT,1} q_{iT,2} + b_{i,22}^- a_{iT,2} q_{iT,2}, \tag{A.54}$$

where

$$a_{iT,1} = \frac{1}{T} \sum_{t=1}^{T} \tilde{x}_{it} \Delta \tilde{y}_{it} = \frac{1}{T} \sum_{t=1}^{T} (x_{it} - \bar{x}_i) \Delta y_{it}, \tag{A.55}$$

$$a_{iT,2} = \frac{1}{T} \sum_{t=1}^{T} \tilde{x}_{it} \Delta \tilde{x}_{it} = \frac{1}{T} \sum_{t=1}^{T} (x_{it} - \bar{x}_i) u_{x,it}, \tag{A.56}$$

A15

$$q_{iT,1} = \frac{\Delta \tilde{\mathbf{y}}_i' \mathbf{P}_i \tilde{\mathbf{v}}_i}{\alpha_i \sqrt{T}}, \tag{A.57}$$

and

$$q_{iT,2} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{\tilde{u}_{x,it} \tilde{v}_{it}}{\alpha_i}. \tag{A.58}$$

Note that $\sup_i \left\| \mathbf{B}_i^{-1} \right\| < K$,[9] and therefore $\left| \left(b_{i,sj}^- \right)^2 \right| < K$. Using this result and Cauchy-Schwarz inequality for the individual summands on the right side of (A.54), we obtain

$$E \left(\mathbf{a}_{iT}' \mathbf{B}_i^{-1} \mathbf{q}_{iT} \right)^2 \leq K \sum_{s=1}^{2} \sum_{j=1}^{2} \sqrt{E \left(a_{iT,s}^4 \right)} \sqrt{E \left(q_{iT,j}^4 \right)} < K, \tag{A.59}$$

where $E \left(a_{iT,s}^4 \right) < K$ by Lemma A.2, and $E \left(q_{iT,j}^4 \right) < K$ by result (A.26) of Lemma A.6. Using (A.59) in (A.53), it follows that

$$Var \left(\frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \mathbf{a}_{iT}' \mathbf{B}_i^{-1} \mathbf{q}_{iT} \right) < \frac{K}{T},$$

and therefore

$$\frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \left(\mathbf{a}_{iT}' \mathbf{B}_i^{-1} \mathbf{q}_{iT} - \mu_{iT}^* \right) \to_{q.m.} 0 \text{ as } n, T \to \infty. \tag{A.60}$$

We establish an upper bound for $|\mu_{iT}^*|$ next. We have (using (A.54) and noting that $\left| b_{i,sj}^- \right| < K$)

$$|\mu_{iT}^*| < K \cdot \sum_{s=1}^{2} \sum_{j=1}^{2} |E \left(a_{iT,s} q_{iT,j} \right)|.$$

It follows that if we can show that

$$|E \left(a_{iT,s} q_{iT,j} \right)| < \frac{K}{\sqrt{T}}, \tag{A.61}$$

holds for all $s, j = 1, 2$, then

$$|\mu_{iT}^*| < \frac{K}{\sqrt{T}}, \tag{A.62}$$

hold. We establish (A.61) for $s = j = 2$, first, which is the most convenient case to consider. We have

$$E \left(a_{iT,2} q_{iT,2} \right) = E \left(\frac{1}{T} \sum_{t=1}^{T} \left(x_{it} - \bar{x}_i \right) u_{x,it} \cdot \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{u_{x,it} v_{it}}{\alpha_i} \right) = 0, \tag{A.63}$$

---

[9] $\mathbf{B}_i$ is invertible and $\inf_i \lambda_{\min} \left(\mathbf{B}_i \right)$ is bounded away from zero. This follows from observing that both σ_{xi}^2 and $E \left(\xi_{it}^2 \right)$ as well as α_i^2 in (A.36) are bounded away from zero.

since v_{it} is independently distributed of $u_{x,it'}$ for any t, t'. Consider next $s = 1$, $j = 2$. We have

$$E\left(a_{iT,1}q_{iT,2}\right) = E\left(\frac{1}{T}\sum_{t=1}^{T}\left(x_{it} - \bar{x}_i\right)\Delta y_{it} \cdot \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{u_{x,it}v_{it}}{\alpha_i}\right), \tag{A.64}$$

where (first-differencing (A.6) and substituting (3))

$$\begin{aligned}
\Delta y_{it} &= \delta_i u_{x,it} + v_{it} - \alpha_i \sum_{\ell=1}^{\infty}\left(1 - \alpha_i\right)^{\ell-1}\left[v_{i,t-\ell} + \left(\delta_i - \beta\right)u_{x,i,t-\ell}\right], \\
&= \eta_{u,it} + \eta_{v,it}, \tag{A.65}
\end{aligned}$$

in which

$$\eta_{u,it} = \delta_i u_{x,it} - \alpha_i \sum_{\ell=1}^{\infty}\left(1 - \alpha_i\right)^{\ell-1}\left(\delta_i - \beta\right)u_{x,i,t-\ell}, \tag{A.66}$$

and

$$\eta_{v,it} = v_{it} - \alpha_i \sum_{\ell=1}^{\infty}\left(1 - \alpha_i\right)^{\ell-1}v_{i,t-\ell}. \tag{A.67}$$

Hence, $E\left(a_{iT,1}q_{iT,2}\right)$ can be written as

$$\begin{aligned}
E\left(a_{iT,1}q_{iT,2}\right) &= E\left(\frac{1}{T}\sum_{t=1}^{T}\left(x_{it} - \bar{x}_i\right)\eta_{u,it} \cdot \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{u_{x,it}v_{it}}{\alpha_i}\right) \\
&\quad + E\left(\frac{1}{T}\sum_{t=1}^{T}\left(x_{it} - \bar{x}_i\right)\eta_{v,it} \cdot \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{u_{x,it}v_{it}}{\alpha_i}\right).
\end{aligned}$$

The first term is equal to 0, since v_{it} is independently distributed of $u_{x,it'}$ for any t, t'. Consider the second term. Noting that $E\left[\left(x_{it} - \bar{x}_i\right)u_{x,is}\right] < K$ and $\left|\alpha_i^{-1}\right| < K$ for any i, t, s, we obtain

$$\begin{aligned}
E\left(\frac{1}{T}\sum_{t=1}^{T}\left(x_{it} - \bar{x}_i\right)\eta_{v,it} \cdot \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{u_{x,it}v_{it}}{\alpha_i}\right) &= \frac{1}{T^{3/2}}\sum_{t=1}^{T}\sum_{s=1}^{T}\alpha_i^{-1}E\left[\left(x_{it} - \bar{x}_i\right)u_{x,is}\right]E\left(\eta_{v,is}v_{it}\right), \\
&\leq \frac{K}{T^{3/2}}\sum_{t=1}^{T}\sum_{s=1}^{T}E\left(\eta_{v,is}v_{it}\right).
\end{aligned}$$

But

$$E\left(\eta_{v,is}v_{it}\right) = \begin{cases} 0, \text{ for } s < t, \\ \sigma_{vi}^2 < K, \text{ for } s = t, \\ \leq K\rho^{s-t}, \text{ for } s > t, \end{cases}$$

where $\rho \equiv \sup_i\left(1 - \alpha_i\right) < 1$ by Assumption 1. Hence $\left|\sum_{s=1}^{T}E\left(\eta_{v,is}v_{it}\right)\right| < K$ for any $t = 1, 2, ...T$, and

$$\left|E\left(\frac{1}{T}\sum_{t=1}^{T}\left(x_{it} - \bar{x}_i\right)\eta_{v,it} \cdot \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{u_{x,it}v_{it}}{\alpha_i}\right)\right| \leq \frac{K}{\sqrt{T}}, \tag{A.68}$$

A17

as desired. This establish (A.61) hold for $s = 1$, $j = 2$.

Consider next (A.61) for $s \in \{1, 2\}$ and $j = 1$. Using expression (A.30), we can write $a_{iT,s}q_{iT,1}$, for $s = 1, 2$, as

$$a_{iT,s}q_{iT,1} = a_{iT,s}\varsigma_{a,iT} + a_{iT,s}\varsigma_{b,iT} + a_{iT,s}\varsigma_{c,iT}, \tag{A.69}$$

where as in the proof of Lemma A.6 $\varsigma_{a,iT} = -\tilde{\boldsymbol{\xi}}'_{i,-1}\tilde{\mathbf{v}}_i/\sqrt{T}$, $\varsigma_{b,iT} = \alpha_i^{-1}\Delta\tilde{\mathbf{x}}'_i\tilde{\mathbf{v}}_i/\sqrt{T}$ and $\varsigma_{c,iT} = \alpha_i^{-1}\Delta\tilde{\mathbf{v}}'_i\mathbf{P}_i\tilde{\mathbf{v}}_i/\sqrt{T}$. Using similar arguments as in establishing (A.68), we obtain

$$|E\left(a_{iT,s}\varsigma_{a,iT}\right)| < \frac{K}{\sqrt{T}}, \text{ for } s = 1, 2.$$

Noting next that $\varsigma_{b,iT} = \alpha_i^{-1}q_{i,T,2}$, it directly follows from results (A.63) and (A.68) that

$$|E\left(a_{iT,s}\varsigma_{b,iT}\right)| < \frac{K}{\sqrt{T}}, \text{ for } s = 1, 2.$$

Consider the last term, $a_{i,T,s}\varsigma_{c,iT}$, for $s = 1, 2$. Using Cauchy-Schwarz inequality we have

$$|E\left(a_{iT,s}\varsigma_{c,iT}\right)| \leq \sqrt{E\left(a_{iT,s}^2\right)}\sqrt{E\left(\varsigma_{c,iT}^2\right)}, \text{ for } s = 1, 2.$$

But $E\left(a_{iT,s}^2\right) < K$, for $s = 1, 2$ by Lemma A.2, and $E\left(\varsigma_{c,iT}^2\right) < K/T$ is implied by (A.33). Hence

$$|E\left(a_{iT,s}\varsigma_{c,iT}\right)| \leq \frac{K}{\sqrt{T}}, \text{ for } s = 1, 2.$$

This completes the proof of (A.61) for all $s, j = 1, 2$, and therefore (A.62) holds. Using (A.62), we

$$\left|\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\mu_{iT}^*\right| \leq \frac{1}{\sqrt{nT}}\sum_{i=1}^{n}|\mu_{iT}^*| < \frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\frac{K}{\sqrt{T}} = K\frac{\sqrt{n}}{T} \to 0, \tag{A.70}$$

as $n, T \to \infty$ such that $\sqrt{n}/T \to 0$. Results (A.60) and (A.70) imply

$$\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\mathbf{a}'_{iT}\mathbf{B}_i^{-1}\mathbf{q}_{iT} \to_p 0, \tag{A.71}$$

as $n, T \to \infty$ such that $\sqrt{n}/T \to 0$. Finally, using (A.52) and (A.71) in (A.51), we obtain (A.60), as desired. ∎

A.3 Proof of Theorem

Proof of Theorem 1. Substituting $\tilde{\mathbf{y}}_i = \tilde{\mathbf{x}}_i \beta + \Delta \tilde{\mathbf{Z}}_i \boldsymbol{\psi}_i + \alpha_i^{-1} \tilde{\mathbf{v}}_i$ in (10), and using $\mathbf{M}_i \tilde{\mathbf{x}}_i = \tilde{\mathbf{x}}_i$, and $\mathbf{M}_i \Delta \tilde{\mathbf{Z}}_i = \mathbf{0}$, we have

$$T\sqrt{n}\left(\hat{\beta} - \beta\right) = \left(\frac{1}{n}\sum_{i=1}^{n} \frac{\tilde{\mathbf{x}}_i' \mathbf{M}_i \tilde{\mathbf{x}}_i}{T^2}\right)^{-1} \left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \frac{\tilde{\mathbf{x}}_i' \mathbf{M}_i \tilde{\mathbf{v}}_i}{T\alpha_i}\right). \tag{A.72}$$

Consider the first term on the right side of (A.72) first. Lemma A.4 establishes

$$\frac{1}{n}\sum_{i=1}^{n} \frac{\tilde{\mathbf{x}}_i' \mathbf{M}_i \tilde{\mathbf{x}}_i}{T^2} \to_p \omega_x^2 > 0, \text{ as } n, T \to \infty, \tag{A.73}$$

where $\omega_x^2 = \sigma_x^2/6$, $\sigma_x^2 = \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n}\sigma_{xi}^2$. Consider the second term on the right side of (A.72). Noting that $\tilde{\mathbf{x}}_i' \mathbf{P}_i = \tilde{\mathbf{x}}_i'$, we have

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \frac{\tilde{\mathbf{x}}_i' \mathbf{M}_i \tilde{\mathbf{v}}_i}{\alpha_i T} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \frac{\tilde{\mathbf{x}}_i' \tilde{\mathbf{v}}_i}{\alpha_i T} + \frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\xi_{i,T}, \tag{A.74}$$

where

$$\xi_{iT} = \frac{\tilde{\mathbf{x}}_i' \Delta \tilde{\mathbf{Z}}_i}{T} \left(\frac{\Delta \tilde{\mathbf{Z}}_i' \mathbf{P}_i \Delta \tilde{\mathbf{Z}}_i}{T}\right)^{-1} \frac{\Delta \tilde{\mathbf{Z}}_i' \mathbf{P}_i \tilde{\mathbf{v}}_i}{\alpha_i \sqrt{T}}. \tag{A.75}$$

Using Lemma A.5 (for the first term on the right side of (A.74)), and Lemma A.9 (for the second term on the right side of (A.74)), we obtain

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \frac{\tilde{\mathbf{x}}_i' \mathbf{M}_i \tilde{\mathbf{v}}_i}{\alpha_i T} \to_d N\left(0, \omega_v^2\right), \tag{A.76}$$

as $n, T \to \infty$ and $\sup_{n,T} \sqrt{n}/T^{1-\epsilon} < K$, for some $\epsilon > 0$, where $\omega_v^2 = \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n}\sigma_{xi}^2\sigma_{vi}^2/\left(6\alpha_i^2\right)$. Using (A.73) and (A.76) in (A.72) establishes (14). ∎