Mean Average Estimation of Dynamic Panel Models with Nonstationary Initial Condition

John Chao
University of Maryland

Myungsup Kim
University of North Texas

Donggyu Sul
University of Texas at Dallas

This Version, March 2014

Abstract

This paper proposes a new class of estimators for the autoregressive coefficient of a dynamic panel data model with random individual effects and nonstationary initial condition. The new estimators we introduce are weighted averages of the well-known first difference (FD) GMM/IV estimator and the pooled ordinary least squares (POLS) estimator. The proposed procedure seeks to exploit the differing strengths of the FD GMM/IV estimator relative to the pooled OLS estimator. In particular, the latter is inconsistent in the stationary case but is consistent and asymptotically normal with a faster rate of convergence than the former when the underlying panel autoregressive process has a unit root. By averaging the two estimators in an appropriate way, we are able to construct a class of estimators which are consistent and asymptotically standard normal, when suitably standardized, in both the stationary and the unit root case. The results of our simulation study also show that our proposed estimator has favorable finite sample properties when compared to a number of existing estimators.

JEL Classification Code: C33

Keywords: Nonstationary initial condition, Mean average estimator, Smoothing Threshold, POLS, FDIV

1 Introduction

Pathbreaking research on the estimation of dynamic panel data models was done in the early 1980s by Nickell (1981) and by Anderson and Hsiao (1981, 1982). The former analyzed the bias properties of the least-squares dummy variable (LSDV) estimator and the pooled OLS estimator, whereas the latter two papers introduced the first-difference (FD) GMM/IV estimator. The three decades since these pioneering papers have witnessed the rapid development of the literature on dynamic panel regression. Among the important contributions to this literature are those of Arellano and Bond (1991), Ahn and Schmidt (1995), Arellano and Bover (1995), Kiviet (1995), Blundell and Bond (1998), Moon and Phillips (2000), Hahn and Kuersteinder (2002), Hsiao, Pesaran and Tahmiscioglu (2002), Alvarez and Arellano (2003), Bun and Kiviet (2006), Kiviet (2007), Phillips and Sul (2007), and Han, Phillips, and Sul (2011, 2014).

Within this vast literature, the work of Kiviet (2007) and Arellano (2003) have considered dynamic panel data settings with the most general form of nonstationary initial condition. In particular, they study a model of the form

$$y_{it} = a_i (1 - \rho) + \rho y_{it-1} + \varepsilon_{it}, \quad y_{i0} = \alpha_0 + \alpha_1 a_i + \alpha_2 \varepsilon_{i0} \tag{1}$$

where $\varepsilon_{it} \sim i.i.d. \left(0, \sigma^2\right)$ for $t \geq 0$ and $a_i \sim i.i.d. \left(\mu_1, \sigma_a^2\right)$. Kiviet notes that modeling initial condition in this way results in three additional unknown parameters: $\alpha_0, \alpha_1, \alpha_2$. Under this setting, Arellano (2003) shows that the following moment conditions hold regardless of nonstationary initial conditions and the restriction of time series heteroskedasticity.

$$\mathrm{E}\left[y_{it} - \rho y_{it-1} - \mu_1 (1 - \rho)\right] = 0 \tag{2}$$

$$\mathrm{E}\left[y_{it} - \rho y_{it-1} - \mu_1 (1 - \rho)\right]\left[\Delta y_{it-1} - \rho \Delta y_{it-2}\right] = 0 \tag{3}$$

$$\mathrm{E}\left[y_{it-s}\right]\left[\Delta y_{it} - \rho \Delta y_{it-1}\right] = 0 \text{ for } s > 2 \tag{4}$$

Using the moment conditions in (2) through (4) yields consistency of GMM estimator. However, as is well-known, the moment condition becomes weaker as ρ approaches unity. Under stationary initial condition, several alternative estimators have been proposed to overcome the weak instrument problem. For example, Choi, Mark and Sul (2010) use backward recursive mean adjustment to reduce the bias especially when ρ is near to unity. Han, Phillips and Sul (2011, 2014) propose 'X-differencing' method to eliminate the bias completely regardless the value of ρ including unit root case. However under nonstationary initial condition, this paper shows that the X-differencing estimator becomes inconsistent.

The main objective of the current paper is add to this impressive list of contributions by introducing a new class of estimators which can be shown to have good asymptotic properties when used to estimate a dynamic panel data regression with nonstationary initial condition. A main reason for considering nonstationary initial conditions is that they arise naturally in a number of situations where the panel data set is generated from laboratory experiments. Our estimation procedure is motivated by the observation that although it is well-known that the GMM/IV estimator is well-behaved in the case where $|\rho|$ is bounded away from one, this estimator suffers from a slower rate of convergence in the unit root case when $\rho = 1$, due to a weak instrument problem. On the other hand, the pooled OLS estimator, while inconsistent in the case where $|\rho| < 1$, becomes consistent with a faster rate of convergence than the FD GMM/IV estimator in the unit root case when $N, T \to \infty$. The present paper seeks to exploit the differing strengths of GMM/IV vis-à-vis pooled OLS in different regions of the parameter space by introducing a class of mean average estimators which are weighted averages of these two estimators.

Several alternative weighting functions for mean average estimation are considered in our paper. In particular, we generate different weighting functions based on taking alternative transformations of a class of information criteria used to evaluate the validity of the unit root hypothesis. The class of information criteria we consider are variants of the Bayesian Information Criterion (BIC). Alternative transformations which we consider include the logistic-type transformation, which has previously been used in Buckland, Burnham, and Augustin (1997), Burham and Anderson (2002), Hjort and Claeskens (2003), Hansen (2007), and Chao (2013); although none of those papers consider the problem of estimating a dynamic panel data model, as we do here. In addition, we also consider a Gaussian type transformation which was originally introduced in Chao (2013).

The remainder of the paper proceeds as follows. In section 2, we briefly discuss the dynamic panel regression studied in this paper and give the assumptions used in our subsequent large sample analysis. Section 3 of the paper provides some asymptotic results for the FD GMM/IV and the pooled OLS estimators. Section 4 introduces our mean average estimator and provides large sample results for this estimator. The results of a Monte Carlo study comparing different versions of the mean average estimator with several existing estimators are given in section 5. We briefly conclude in section 6 of the paper. Proofs and supporting lemmas are given in the Appendices.

A few words on notations before we proceed. In this paper, we use $\Longrightarrow$ to denote weak convergence and $\xrightarrow{p}$ to denote convergence in probability. In addition, we let χ^2_ν denote a chi-square random variable with ν degrees of freedom, and let $W_i(r)$, $r \in [0, 1]$, denote the Wiener process, or standard Brownian motion.

2 Model and Assumptions

We consider the following dynamic panel data model written in unobserved components form

$$y_{it} = a_i + x_{it}, \quad (5)$$

$$x_{it} = \rho x_{it-1} + \varepsilon_{it}, \quad (6)$$

for $i = 1, ...N$ and $t = 1, ..., T$. Here, $\{y_{it}\}$ denotes the observed data, whereas a_i denotes an (unobserved) random effect and $\{x_{it}\}$ is generated by a latent AR(1) process. The initial condition for this model is given by

$$y_{i0} = a_i + x_{i0}. \quad (7)$$

We make the following assumptions on the model given by equations (5)-(7).

Assumption 1: (Idiosyncratic Errors)

(a) $\{\varepsilon_{it}\} \equiv i.i.d. \left(0, \sigma^2\right)$ across i and t, $\sigma^2 > 0$;

(b) $E\left[\varepsilon_{it}^4\right] < \infty$.

Assumption 2: (Random Effects)

(a) $\{a_i\} \equiv i.i.d. \left(\mu_1, \sigma_a^2\right)$ across i, $\sigma_a^2 > 0$;

(b) $E\left[a_i^4\right] < \infty$.

Assumption 3: (Initial Condition)

(a) $\{x_{i0}\} \equiv i.i.d. \left(m_1, \sigma_x^2\right)$ across i, $\sigma_x^2 > 0$;

(b) $E\left[x_{i0}^4\right] < \infty$.

Assumption 4:

ε_{it}, a_i, and x_{i0} are mutually independent for all i and t.

Remarks:

(i) As Kiviet (2007) suggests, we can model the initial condition as (1). However there are too many parameters to characterize the initial conditions. To simplify the parameterization, we consider the following latent AR(1) model.

$$y_{it} = a_i + x_{it}, \quad x_{it} = \rho x_{it-1} + \varepsilon_{it}, \tag{8}$$

of which regression model at time $t = 1$ can be rewritten as

$$y_{i1} = a_i (1 - \rho) + \rho y_{i0} + \varepsilon_{i1}. \tag{9}$$

Modelling the initial condition based on the latent model in (8) is much easier than doing based on the actual regression model in (9). To see this, let treat y_{i1} as the initial condition by setting $y_{i0} = 0$ in (9), then we have $y_{i1} = a_i (1 - \rho) + \varepsilon_{i1}$ and need to model ε_{i1}. Meanwhile let $x_{i0} = 0$ and use the latent model in (8), then we have $y_{i1} = a_i + x_{i1}$. Hence we need to model x_{i1}. Define the steady state mean of y_{it} as $\lim_{t\to\infty} E(y_{it}) = a_i$. Then the initial condition based on the latent model can be rewritten as $x_{i1} = y_{i1} - \lim_{t\to\infty} E(y_{it})$. Hence it becomes more attractive in terms of economic implication. Now change the time from $t = 1$ to $t = 0$ and treat y_{i0} as the initial condition. Then we have

$$y_{i0} = \lim_{t\to\infty} E(y_{it}) + x_{i0} = a_i + x_{i0} = c_i, \ (say). \tag{10}$$

Note that $x_{i0} = \rho x_{i,-1} + \varepsilon_{i0}$. Hence, without loss of generality, we may let $\varepsilon_{i0} = 0$. In this case, the latent model can be assumed to be

$$y_{it} = a_i + x_{i0}\rho^t + w_{it}, \tag{11}$$

where

$$w_{it} = \sum_{j=0}^{t} \rho^j \varepsilon_{it-j} \text{ where } \varepsilon_{i0} = 0. \tag{12}$$

Hence this approach identifies three unknown parameters in Kiviet (2007)'s model. That is, $\alpha_0 = x_{i0}$, $\alpha_1 = 1$ and $\alpha_2 = 0$. The basic but standard model can be obtained by setting $x_{i0} = 0$.

(ii) We can also allow for heterogeneous variance for ε_{it} across i, but for notational convenience we do not consider this extension here.

(iii) Note that Assumption 1 does not allow for cross-sectional dependence in the idiosyncratic errors $\{\varepsilon_{it}\}$. This assumption is reasonable in experimental games since subjects are randomly selected. It is also important to note that Assumption 1 does not imply cross sectional independence among the experimental outcome y_{it}. Motivated by the possibility of common game setting and learning, the outcome variable is cross-sectionally correlated. In fact, the cross sectional structure can be thought of as an approximate factor model. Observe that equations (5)-(7) imply that

$$y_{it} = a_i + x_{i0}\rho^t + w_{it}$$

where

$$w_{it} = \sum_{j=0}^{t} \rho^j \varepsilon_{it-j}, \text{ with } \varepsilon_{i0} = 0.$$

In this representation, the initial condition parameter x_{i0} can be interpreted as a factor loading coefficient and the learning parameter ρ^t can be taken to be a common factor. However, in the stationary case, the cross-sectional dependence goes away quickly since ρ^t decays to zero geometrically as t increases.

3 Consistency of Existent Estimators

There are many consistent estimators available when the initial condition is stationary. As we discussed in Introduction, only FDIV estimator, which is using the last moment condition in (4), becomes consistent. Other GMM/IV estimator such as forward orthogonal demeaning by Arellano and Bond (1998) does not satisfy the moment conditions. Hence our interest is rather focusing on two existent estimators: Exact mean unbiased (EMU) estimator by Phillips and Sul (2007) and X-differencing estimator proposed by Han, Phillips and Sul (2011,2014). Throughout this subsection, we consider only N asymptotics. That is, we consider a fixed T but let $N \to \infty$.

We assume researchers run the following panel AR(1) regression with fixed effects regardless of the true latent models.

$$y_{it} = \alpha_i + \rho y_{it-1} + \varepsilon_{it}, \quad t = 1, ..., T+1 \tag{13}$$

Even though y_{it} has a time varying mean, the panel AR(1) regression in (13) is not misspecified. To see this, we calculate the quasi-difference in y_{it} as

$$y_{it} - \rho y_{it-1} = \left[a_i + x_{i0}\rho^t + w_{it}\right] - \left[\rho a + x_{i0}\rho^t + \rho w_{it-1}\right] = a_i (1-\rho) + \varepsilon_{it}$$

3.1 Non-Existence of MUE based on Within Group Estimator

First we consider the within group (WG) estimator. It is well known that WG estimator suffers from the asymptotic bias with small T. Here we derive the exact bias form first and then see if the bias function is monotonic over ρ. If so, we may consider the exact mean unbiased estimator considered by Phillips and Sul (2007). First, define the WG estimator

$$\hat{\rho}_{\text{wg}} = \frac{\sum^N \sum^T \tilde{y}_{it-1}\tilde{y}_{it}}{\sum^N \sum^T \tilde{y}_{it-1}^2},$$

and rewrite it as

$$\hat{\rho}_{\text{wg}} - \rho = \frac{\sum^N \sum^T \tilde{y}_{it-1}\tilde{u}_{it}}{\sum^N \sum^T \tilde{y}_{it-1}^2} := \frac{A_{NT}}{B_{NT}}, \tag{14}$$

Let $E_N = \text{plim}_{N\to\infty}$. Then we have the following.

$$E_N A_{NT} = \left(-\frac{1}{1-\rho} + \frac{1-\rho^T}{(1-\rho)^2 T}\right)\sigma^2$$

but

$$E_N B_{NT} = E_N B_{NT} = \sigma_x^2 \left(\frac{\rho^2\left(1-\rho^{2T}\right)}{1-\rho^2} - \frac{\rho^2\left(1-\rho^T\right)^2}{(1-\rho)^2 T}\right) + B_{1T},$$
$$B_{1T} = \sigma^2\left(\frac{T}{1-\rho^2} - \frac{(1+\rho)^2+\rho^2-\rho^{2+2T}}{\left(1-\rho^2\right)^2} + \frac{\rho\left(1-\rho^T\right)\left(2+\rho-\rho^{1+T}\right)}{\left(1-\rho^2\right)(1-\rho)^2 T}\right),$$

where B_{1T} is the expected value of the denominator term with stationary initial condition. Therefore the bias of the WG estimator under nonstationary initial condition is less than that under stationary initial condition since usually $E_N B_{NT}$ is less than B_{1T}. When T is large, the Nickell bias under nonstationary initial condition is given by

$$E_N\left(\hat{\rho}_{\text{wg}} - \rho\right) = -\frac{1+\rho}{T} + O\left(T^{-2}\right),$$

which is the same as the Nickell bias under stationary initial condition. It is due to the local nonstationarity of y_{it}.

Nonetheless, if the first derivative of a function does not change its sign, then the function is monotonic by definition. Alternatively, we can say if either maximum or minimum of the first derivatives of a function is negative or positive, respectively, then the function is monotonic. Let $\mathcal{B}_T = E_N\left(\hat{\rho}_{\text{wg}} - \rho\right)$ and $\mathcal{C}_T = E_N \hat{\rho}_{\text{wg}}$. Then $\mathcal{B}_T$ becomes the asymptotic inconsistency function of the WG estimators meanwhile $\mathcal{C}_T$ becomes the asymptotic mean function. We calculate numerically and find that $\mathcal{B}_T$ and $\mathcal{C}_T$ are not monotonic at all. Due to this non-monotonicity, the exact mean unbiased estimator cannot exist.

3.2 Inconsistency of X-differencing Estimator

The basic reason of why X-differencing method eliminates the bias completely with stationary initial conditions can be seen as follows. First consider the following cross differencing by using the $(s-1)$th and sth observations where $1 \leq s < t-1$.

$$y_{it} - y_{is-1} = \rho\left(y_{it-1} - y_{is}\right) + e_{it}, \text{ for } e_{it} = x_{i0}\left(\rho^{s+1} - \rho^{s-1}\right) + (\rho - 1)\, w_{is-1} + \varepsilon_{is} + \varepsilon_{it}.$$

Under stationary initial condition, $\mathrm{E}e_{it}(y_{it-1}-y_{is})=0$ for all $1\leq s<t-1$. However under nonstationary initial condition, this moment condition becomes invalid. That is,

$$\begin{aligned}&\mathrm{E}e_{it}(y_{it-1}-y_{is})\\=\;&m_2\left(\rho^{s+1}-\rho^{s-1}\right)\left(\rho^{t-1}-\rho^{s}\right)+\sigma^2\left[(\rho-1)\left(\rho^{t+1}-\rho\right)\frac{\rho^2-\rho^{2s-2}}{1-\rho^2}+\rho^{t-s-1}-1\right]\neq 0\end{aligned}$$

for all $1\leq s<t-1$.

To evaluate the inconsistency, we can consider only partial aggregation since full aggregation gives more inconsistency. The probability limit of the numerator term becomes by setting $s=1$.

$$\mathrm{E}_N\sum\nolimits_{t=1}^{T-1}e_{it}(y_{it-1}-y_{i1})=-\left[\sigma^2-m_2\left(1-\rho^2\right)\right]\rho T+O(1)\,. \tag{15}$$

Meanwhile the probability limit of the denominator is given by

$$\mathrm{E}_N\sum\nolimits_{t=1}^{T-1}(y_{it-1}-y_{i1})^2=\frac{m_2\rho^2\left(1-\rho^2\right)+\sigma^2\left(2-\rho^2\right)}{1-\rho^2}T+O\left(T^{-2}\right). \tag{16}$$

Finally the inconsistency of X-differencing estimator, $\hat{\rho}_{\mathrm{x}}$, becomes

$$E_N(\hat{\rho}_{\mathrm{x}}-\rho_T)=\frac{\rho\left(1-\rho^2\right)\left(\sigma^2-m_2\left(1-\rho^2\right)\right)}{m_2\rho^2\left(1-\rho^2\right)+\sigma^2\left(2-\rho^2\right)}+O\left(T^{-1}\right). \tag{17}$$

Therefore the inconsistency of $\hat{\rho}_{\mathrm{x}}$ is dependent on many nuisance values of m_2 and σ^2.

4 Asymptotic Properties of FD GMM/IV and Pooled OLS Estimators

To motivate the construction of our proposed mean average estimator given in the next section, we first provide some preliminary asymptotic results on two well-known estimators, the FD GMM/IV estimator and the pooled OLS estimator. As proposed originally by Anderson and Hsiao (1981, 1982), the FD GMM/IV estimator uses y_{it-2} as an instrument to estimate the first differenced dynamic panel model

$$\Delta y_{it}=\rho\Delta y_{it-1}+\Delta\varepsilon_{it}$$

resulting in an estimator of the form

$$\hat{\rho}_{\mathrm{IV}}=\left(\sum_{i=1}^{N}\sum_{t=2}^{T}y_{it-2}\Delta y_{it-1}\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=2}^{T}y_{it-2}\Delta y_{it}\right). \tag{18}$$

On the other hand, running pooled OLS on the level specification

$$y_{it}=\alpha_i+\rho y_{it-1}+\varepsilon_{it}$$

results in the estimator

$$\widehat{\rho}_{\text{pols}} = \left[\sum_{i=1}^{N}\sum_{t=2}^{T}\left(y_{it-1} - \overline{y}_{-1,N,T}\right)^2\right]^{-1}\left[\sum_{i=1}^{N}\sum_{t=2}^{T}\left(y_{it-1} - \overline{y}_{-1,N,T}\right)\left(y_{it} - \overline{y}_{N,T}\right)\right]$$

where

$$\overline{y}_{-1,N,T} = \frac{1}{N\left(T-1\right)}\sum_{i=1}^{N}\sum_{t=2}^{T} y_{it-1},\ \overline{y}_{N,T} = \frac{1}{N\left(T-1\right)}\sum_{i=1}^{N}\sum_{t=2}^{T} y_{it},$$

The asymptotic results we give below are for the case where N and T are both large, with the order of magnitude of T not exceeding that of N, as this covers many panel data settings of interest to economists. We consider both the case where the panel AR(1) process is stationary with $\rho \in (0,1)$ and also the unit root case where $\rho = 1$.

Theorem 1:

Suppose that Assumptions 1-4 hold. Then, the following statements are true.

(a) *If* $\rho \in (0,1)$,

$$\sqrt{NT}\left(\widehat{\rho}_{IV} - \rho\right) \Longrightarrow N\left(0, 2\left[1+\rho\right]\right)$$

as $N, T \to \infty$ *such that* $T/N \to \kappa \in [0,\infty)$.

(b) *If* $\rho = 1$,

$$\sqrt{T}\left(\widehat{\rho}_{IV} - 1\right) \Longrightarrow 2\frac{\mathcal{Z}_1}{\mathcal{Z}_2}$$

as $N, T \to \infty$ *such that* $T/N \to \kappa \in [0,\infty)$. *Here,* $Z_1 \equiv N\left(0,1\right)$ *and* $Z_2 \equiv N\left(0,1\right)$ *and* Z_1 *and* Z_2 *are independent.*

Theorem 2:

Suppose that Assumptions 1-4 hold. Then, the following statements are true.

(a) *If* $\rho \in (0,1)$,

$$\widehat{\rho}_{\text{pols}} \xrightarrow{p} \rho + \frac{\sigma_a^2\left(1-\rho\right)\left(1-\rho^2\right)}{\sigma_a^2\left(1-\rho^2\right)+\sigma^2}.$$

as $N, T \to \infty$ *such that* $T/N \to \kappa \in [0,\infty)$.

(b) *If* $\rho = 1$,

$$\sqrt{N}T\left(\widehat{\rho}_{\text{pols}} - 1\right) \Longrightarrow N\left(0,2\right).$$

as $N, T \to \infty$ *such that* $T/N \to \kappa \in [0,\infty)$.

Remarks:

(i) Note that Theorems 1 and 2 show that the relative desirability of FD GMM/IV vis-à-vis the pooled OLS estimator depends on whether the panel AR(1) is stationary or has a unit root. In the stationary case, pooled OLS is inconsistent whereas FD GMM/IV is both consistent and asymptotically normal. On the other hand, in the unit root case, both FD GMM/IV and pooled OLS are consistent but the latter has a faster rate of convergence. In addition, in the unit root case, FD GMM/IV, due the weak instrument problem, is no longer asymptotically normal and has a limiting distribution which is a scaled Cauchy distribution, being proportion to the ratio of two independent standard normal variates.

(ii) Note that some of the results reported in the above theorems are not completely new to this paper. In particular, part (b) of Theorem 1 has been shown recently by Phillips (2014). In related results, Kruiniger (2009) has also shown a Cauchy-like limiting distribution for the FD GMM/IV estimator in situations with persistent data where N is large and T is fixed.

5 Mean Average Estimator

More precisely, the construction of our estimator begins with two existing estimators: the first difference IV estimator, $\hat{\rho}_{\mathrm{IV}}$ and the pooled OLS estimator $\hat{\rho}_{\mathrm{pols}}$. Our reason for choosing the pooled OLS estimator for this construction stems in part from its good finite sample properties in cases where ρ is close to unity. Some evidence of $\hat{\rho}_{\mathrm{pols}}$'s good finite sample performance relative to other existing estimators for ρ close to unity is provided in our simulation results reported below.

Given these estimators, our proposed mean average estimator can be written as

$$\hat{\rho}_{\mathrm{avg}} = w_{IC}\hat{\rho}_{\mathrm{IV}} + (1 - w_{IC})\,\hat{\rho}_{\mathrm{pols}}.$$

where w_{IC} is a weighting function based on a BIC-type information criterion, denoted by Δ_{IC} below. When the weighting function is based on the logistic-type transformation, we have the explicit expression

$$w_{IC,1} = \frac{1}{1 + \exp\left\{\frac{1}{2}\Delta_{IC}\right\}}; \tag{19}$$

and when the Gaussian-type tranformation is used, we have

$$w_{IC,2} = \Pr\left(\mathbb{Z} > \Delta_{IC}\right) \tag{20}$$

where $\mathbb{Z}$ denotes the standard normal random variable.

We also consider alternative specifications for Δ_{IC}. More specifically, we consider an information criterion of the general form

$$\Delta_{IC}(\alpha,\beta) \equiv \mathbb{T}_{N,T} + \alpha \ln(N) + \beta \ln(T),$$

where $\alpha \geq 2$ and $\beta \geq 1$ are pre-specified (positive) integers and where $\mathbb{T}_{N,T}$ denotes the usual t-statistic for testing the presence of a unit root in dynamic panel data model, i.e.,

$$\mathbb{T}_{N,T} = \frac{\hat{\rho}_{\text{pols}} - 1}{\sqrt{\widehat{\sigma}^2 / \sum_{i=1}^{N} \sum_{t=1}^{T} \left(y_{it-1} - \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=2}^{T} y_{it-1} \right)^2}}.$$

Next, we analyze the asymptotic properties of this mean average estimator. To do so, we first introduce the following normalization function

$$\nu_{NT} = w_{IC} \left(\frac{2}{NT} \left[1 + \widehat{\rho}_{\text{IV}} \right] \right)^{-1/2} + (1 - w_{IC}) \widehat{\sigma}^{-1} \left[\sum_{i=1}^{N} \sum_{t=2}^{T} (y_{it-1} - \overline{y}_{NT})^2 \right]^{1/2}.$$

Our main asymptotic result is given by the following theorem.

Theorem 3:

Suppose that Assumptions 1-4 hold. Then, the following statements are true.

(a) *If* $\rho \in (0,1)$,

$$\nu_{NT} \left(\widehat{\rho}_{\text{avg}} - \rho \right) \Longrightarrow N(0,1).$$

as $N, T \to \infty$ *such that* $T/N \to \kappa \in [0, \infty)$.

(b) *If* $\rho = 1$,

$$\nu_{NT} \left(\widehat{\rho}_{\text{avg}} - 1 \right) \Longrightarrow N(0,1).$$

as $N, T \to \infty$ *such that* $T/N \to \kappa \in [0, \infty)$.

Note that Theorem 3 shows that the mean average estimator $\widehat{\rho}_{\text{avg}}$ improves upon both the FD GMM/IV and the pooled OLS estimator, in that when appropriately normalized, it has an asymptotic standard normal distribution in both the stationary and the unit root case. It is also easily seen that it has a faster rate of convergence vis-à-vis the FD GMM/IV estimator when the underlying panel AR(1) process has a unit root.

6 Monte Carlo Simulation

For our Monte Carlo study, we consider the following data generating process:

$$x_{i0} \sim i.i.d.\ N(5,1),\ a_i \sim i.i.d.\ N(1,1),\quad \{\varepsilon_{it}\} \sim i.i.d.\ N(0,1).$$

Parameters which have important impact on the finite sample performance of different estimators include ρ, N, T, and $m_1 = E[x_{i0}]$. In our study, we vary $\rho = 0.7$, 0.8, 0.85, 0.9, 0.95; $N = 100$, 200; and $T = 10$, 15, 20, 25, 30, 40, 50, 100. To keep the dimension of the parameter space manageable, we fix the mean of the initial condition, m_1, to be 5 in all our experiments. In unreported simulation results, we have found that our simulation results are not much affected by changes in m_1, except when m_1 is very small.

As mentioned earlier, we consider four versions of the mean average estimator. These alternative versions differ only in terms of the information criterion used in the weighting function. More specifically, given the general form of the information criterion used here, i.e.,

$$\Delta_{IC}(\alpha,\beta) = \mathbb{T}_{N,T} + \alpha \ln(N) + \beta \ln(T),$$

we consider four alternative specifications given by

$$\begin{array}{lll} \text{BIC1:} & \alpha = 1 & \beta = 1 \\ \text{BIC2:} & \alpha = 2 & \beta = 1 \\ \text{BIC3:} & \alpha = 2 & \beta = 2 \\ \text{BIC4:} & \alpha = 2 & \beta = 3 \end{array}.$$

Note that of these four specifications, BIC4 carries the strongest penalty function against overparameterization and, everything else being equal, would tend to favor the more parsimonious unit root hypothesis. On the other hand, BIC1 has the weakest penalty function and, in consequence, most strongly favor the stationary case. It follows that, of the four variants, the mean average estimator based on BIC4 would tend to put the heaviest weight on the POLS estimator in finite sample, whereas the estimator based on BIC1 would put a heavier weight on the FDIV estimator relative to the other three specifications.

Tables 1 and 2 give the sample mean of the simulated estimators over 2,000 replications. Looking at these tables, we see that the WG estimator suffers from serious downward (asymptotic) bias, the POLS estimator shows a mild upward bias, whereas the X-differencing estimator exhibits a more significant upward bias. Also, in terms of bias, all four mean average estimators tend to perform better than WG, X-Dif and POLS; but the estimator with the smallest bias across all values of ρ examined under our experimental design is the FD GMM/IV estimator.

Tables 3 and 4 give the relatively mean squares error (MSE) of the different estimator relative to that of the X-differencing estimator. Note first that variants of the mean average estimator based on BIC3 and BIC4 have uniformly better MSE than the X-differencing estimator across all data generating processes considered in our simulation study. The FD GMM/IV estimator also performs well in MSE and, in fact, does better than BIC3 and BIC4 in cases where $\rho = 0.7$, 0.8. However, the MSE of FD GMM/IV deteriorates drastically as ρ moves closer to unity so that, when $\rho = 0.95$, FD GMM/IV not only performs much worse than BIC3 and BIC4 in terms of MSE but its MSE is also larger than that of the X-differencing estimator in this case. Note also that the mean average estimators also tend to perform better in terms of MSE than POLS, especially in cases where ρ is not so close to unity such as when $\rho = 0.7$ or 0.8.

7 Conclusion

In this paper, we introduce a new class of estimators, which are constructed by taking a weighted average of the FD GMM/IV and the pooled OLS estimators. We show that an advantage of this mean average estimator is that it is consistent and asymptotically normal when appropriately standardized, regardless of whether the underlying AR(1) process is stationary or has a unit root. Our simulation study also shows that this estimator performs well in finite sample relative to a number of other estimators that have been proposed in the panel data literature.

Table 1: Means of Various Estimators ($N = 100$)

ρ	T	WG	Xdif	POLS	FDIV	BIC1	BIC2	BIC3	BIC4
0.95	10	0.695	0.998	0.957	0.953	0.957	0.957	0.957	0.957
0.95	20	0.833	0.987	0.955	0.949	0.955	0.955	0.955	0.955
0.95	30	0.875	0.980	0.954	0.950	0.954	0.954	0.954	0.954
0.95	50	0.907	0.972	0.954	0.951	0.953	0.953	0.954	0.954
0.95	100	0.929	0.962	0.954	0.950	0.951	0.951	0.953	0.954
0.9	10	0.710	1.053	0.917	0.901	0.916	0.916	0.917	0.917
0.9	20	0.814	0.999	0.914	0.899	0.911	0.911	0.913	0.913
0.9	30	0.843	0.974	0.913	0.900	0.909	0.907	0.911	0.913
0.9	50	0.865	0.949	0.913	0.901	0.904	0.903	0.909	0.912
0.9	100	0.882	0.927	0.914	0.900	0.900	0.900	0.902	0.908
0.85	10	0.700	1.082	0.879	0.851	0.876	0.875	0.877	0.878
0.85	20	0.777	0.990	0.875	0.849	0.864	0.862	0.870	0.874
0.85	30	0.800	0.952	0.875	0.850	0.857	0.856	0.865	0.873
0.85	50	0.818	0.917	0.877	0.851	0.852	0.852	0.857	0.868
0.85	100	0.833	0.886	0.879	0.850	0.850	0.850	0.850	0.854
0.8	10	0.671	1.085	0.842	0.800	0.833	0.830	0.837	0.840
0.8	20	0.733	0.968	0.839	0.800	0.813	0.810	0.824	0.834
0.8	30	0.752	0.920	0.841	0.800	0.805	0.804	0.814	0.830
0.8	50	0.769	0.877	0.844	0.801	0.802	0.801	0.804	0.818
0.8	100	0.783	0.841	0.847	0.800	0.800	0.800	0.800	0.801
0.7	10	0.589	1.036	0.773	0.700	0.740	0.733	0.753	0.765
0.7	20	0.637	0.886	0.777	0.700	0.710	0.707	0.723	0.750
0.7	30	0.654	0.829	0.782	0.700	0.703	0.702	0.709	0.733
0.7	50	0.671	0.780	0.787	0.701	0.701	0.701	0.702	0.710
0.7	100	0.684	0.741	0.793	0.700	0.700	0.700	0.700	0.700

Table 2: Means of Various Estimators ($N = 200$)

ρ	T	WG	Xdif	POLS	FDIV	BIC1	BIC2	BIC3	BIC4
0.95	10	0.695	0.999	0.958	0.953	0.958	0.958	0.958	0.958
0.95	20	0.833	0.987	0.955	0.949	0.955	0.955	0.955	0.955
0.95	30	0.876	0.980	0.955	0.950	0.955	0.954	0.955	0.955
0.95	50	0.907	0.972	0.954	0.950	0.953	0.952	0.954	0.954
0.95	100	0.929	0.962	0.954	0.950	0.951	0.950	0.951	0.953
0.9	10	0.711	1.053	0.918	0.901	0.918	0.917	0.918	0.918
0.9	20	0.814	0.999	0.914	0.899	0.912	0.907	0.912	0.914
0.9	30	0.843	0.974	0.914	0.900	0.908	0.903	0.907	0.912
0.9	50	0.865	0.949	0.913	0.900	0.901	0.900	0.901	0.906
0.9	100	0.882	0.927	0.914	0.900	0.900	0.900	0.900	0.900
0.85	10	0.701	1.082	0.880	0.851	0.878	0.872	0.877	0.879
0.85	20	0.777	0.990	0.876	0.849	0.862	0.853	0.860	0.869
0.85	30	0.800	0.953	0.876	0.850	0.853	0.851	0.853	0.861
0.85	50	0.818	0.917	0.877	0.850	0.850	0.850	0.850	0.852
0.85	100	0.833	0.886	0.879	0.850	0.850	0.850	0.850	0.850
0.8	10	0.672	1.086	0.843	0.801	0.835	0.819	0.830	0.838
0.8	20	0.733	0.968	0.840	0.799	0.807	0.801	0.805	0.816
0.8	30	0.753	0.921	0.842	0.800	0.801	0.800	0.801	0.806
0.8	50	0.769	0.877	0.844	0.800	0.800	0.800	0.800	0.801
0.8	100	0.783	0.840	0.848	0.800	0.800	0.800	0.800	0.800
0.7	10	0.590	1.037	0.774	0.700	0.734	0.710	0.723	0.743
0.7	20	0.637	0.886	0.778	0.699	0.702	0.700	0.701	0.708
0.7	30	0.655	0.829	0.783	0.700	0.701	0.700	0.700	0.702
0.7	50	0.670	0.780	0.788	0.700	0.700	0.700	0.700	0.700
0.7	100	0.684	0.741	0.794	0.700	0.700	0.700	0.700	0.700

Table 3: Relative Mean Squares Errors against X-Differencing ($N = 100$)

ρ	T	WG	Xdif	POLS	FDIV	BIC1	BIC2	BIC3	BIC4
0.95	10	18.22	1.000	0.069	2.071	0.069	0.069	0.069	0.069
0.95	20	8.574	1.000	0.055	1.383	0.056	0.057	0.055	0.055
0.95	30	5.734	1.000	0.056	1.329	0.059	0.063	0.056	0.056
0.95	50	3.737	1.000	0.071	1.499	0.107	0.141	0.072	0.070
0.95	100	2.755	1.000	0.149	2.467	1.022	1.288	0.225	0.147
0.9	10	1.508	1.000	0.022	0.120	0.021	0.021	0.021	0.021
0.9	20	0.768	1.000	0.027	0.131	0.026	0.027	0.026	0.027
0.9	30	0.610	1.000	0.041	0.170	0.045	0.054	0.037	0.040
0.9	50	0.513	1.000	0.083	0.250	0.148	0.170	0.079	0.076
0.9	100	0.484	1.000	0.291	0.485	0.464	0.471	0.375	0.204
0.85	10	0.425	1.000	0.020	0.038	0.018	0.017	0.019	0.019
0.85	20	0.276	1.000	0.037	0.057	0.027	0.028	0.029	0.034
0.85	30	0.252	1.000	0.068	0.082	0.052	0.057	0.043	0.057
0.85	50	0.242	1.000	0.168	0.129	0.116	0.120	0.092	0.102
0.85	100	0.250	1.000	0.659	0.260	0.258	0.259	0.252	0.212
0.8	10	0.209	1.000	0.025	0.022	0.019	0.017	0.021	0.023
0.8	20	0.166	1.000	0.059	0.039	0.027	0.028	0.032	0.047
0.8	30	0.165	1.000	0.123	0.059	0.049	0.052	0.045	0.075
0.8	50	0.170	1.000	0.332	0.095	0.092	0.093	0.086	0.100
0.8	100	0.187	1.000	1.360	0.196	0.196	0.196	0.195	0.187
0.7	10	0.114	1.000	0.050	0.015	0.021	0.018	0.030	0.041
0.7	20	0.121	1.000	0.174	0.032	0.029	0.029	0.036	0.083
0.7	30	0.134	1.000	0.410	0.051	0.049	0.050	0.049	0.095
0.7	50	0.149	1.000	1.192	0.085	0.085	0.085	0.084	0.090
0.7	100	0.175	1.000	5.053	0.184	0.184	0.184	0.184	0.183

Table 3: Relative Mean Squares Errors against X-Differencing ($N = 200$)

ρ	T	WG	Xdif	POLS	FDIV	BIC1	BIC2	BIC3	BIC4
0.95	10	21.93	1.000	0.050	1.206	0.050	0.050	0.050	0.050
0.95	20	9.040	1.000	0.040	0.728	0.040	0.042	0.040	0.040
0.95	30	5.694	1.000	0.044	0.679	0.044	0.058	0.044	0.044
0.95	50	3.746	1.000	0.056	0.791	0.075	0.278	0.069	0.056
0.95	100	2.743	1.000	0.133	1.275	0.908	1.196	0.747	0.170
0.9	10	1.503	1.000	0.018	0.059	0.018	0.017	0.018	0.018
0.9	20	0.762	1.000	0.025	0.065	0.021	0.025	0.021	0.023
0.9	30	0.589	1.000	0.039	0.084	0.034	0.061	0.034	0.033
0.9	50	0.503	1.000	0.081	0.131	0.110	0.127	0.106	0.061
0.9	100	0.474	1.000	0.292	0.250	0.249	0.249	0.248	0.236
0.85	10	0.417	1.000	0.018	0.018	0.017	0.013	0.016	0.017
0.85	20	0.272	1.000	0.036	0.028	0.018	0.023	0.018	0.024
0.85	30	0.242	1.000	0.067	0.040	0.033	0.039	0.034	0.029
0.85	50	0.235	1.000	0.167	0.068	0.067	0.068	0.067	0.061
0.85	100	0.243	1.000	0.670	0.134	0.134	0.134	0.134	0.133
0.8	10	0.203	1.000	0.024	0.011	0.017	0.010	0.014	0.019
0.8	20	0.162	1.000	0.059	0.019	0.016	0.018	0.016	0.019
0.8	30	0.158	1.000	0.123	0.029	0.027	0.028	0.027	0.025
0.8	50	0.165	1.000	0.334	0.050	0.050	0.050	0.050	0.049
0.8	100	0.180	1.000	1.393	0.101	0.101	0.101	0.101	0.101
0.7	10	0.110	1.000	0.050	0.007	0.014	0.007	0.010	0.020
0.7	20	0.118	1.000	0.177	0.016	0.015	0.016	0.015	0.016
0.7	30	0.127	1.000	0.411	0.025	0.025	0.025	0.025	0.025
0.7	50	0.144	1.000	1.211	0.046	0.046	0.046	0.046	0.046
0.7	100	0.166	1.000	5.218	0.096	0.096	0.096	0.096	0.096

References

[1] Ahn, S.C. and P. Schmidt (1995). Efficient Estimation of Models for Dynamic Panel Data. *Journal of Econometrics*, 68, 29–51.

[2] Alvarez, J. and M. Arellano (2003). The Time Series and Cross-Section Asymptotics of Dynamic Panel Data Estimators. *Econometrica*, 71, 1121–1159.

[3] Anderson, T. W. and C. Hsiao (1981). Estimation of Dynamic Models with Error Components. *Journal of American Statistical Association*, 76, 598–606.

[4] Anderson, T. W. and C. Hsiao (1982). Formulation and Estimation of Dynamic Models Using Panel Data. *Journal of Econometrics*, 18, 47–82 .

[5] Arellano, M. (2003). *Panel Data Econometrics*. Oxford University Press, New York.

[6] Arellano, M. and S. Bond (1991). Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *Review of Economic Studies*, 58, 277–297.

[7] Arellano, M. and O. Bover (1995). Another Look at the Instrumental Variable Estimation of Error-components Models. *Journal of Econometrics*, 68, 29–51.

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

[9] Buckland, S.T., K. P. Burnham, and N. H. Augustin (1997). Model Selection: An integral Part of Inference. *Biometrics*, 53, 603-618.

[10] Bun, M. J. G. and J. F. Kiviet (2006). The Effects of Dynamic Feedbacks on LS and MM Estimator Accuracy in Panel Data Models. *Journal of Econometrics*, 132(2), 409-444.

[11] Burnham, K. P. and D. R. Anderson (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer-Verlag, Berlin.

[12] Chao, J. C. (2013). Panel Structural Modeling with Weak Instrumentation and Covariance Restrictions. *Econometric Theory*, forthcoming.

[13] Choi, C.Y., N.C. Mark and D. Sul (2010). Bias Reduction in Dynamic Panel Data Models by Recursive Mean Adjustment under Cross Section Dependence, *Oxford Bulletin of Economics and Statistics,* 72(7), 567–599.

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

[15] Hansen, B. E. (2007). Least Squares Model Averaging. *Econometrica*, 75, 1175-1189.

[16] Han, C., P.C.B. Phillips and D. Sul (2011). Uniform Asymptotic Normality in Stationary and Unit Root Autoregression, *Econometric Theory*, 27(6), 1117–1151.

[17] Han, C., P.C.B. Phillips and D. Sul (2014). X-Differencing and Dynamic Panel Model Estimation, *Econometric Theory*, 30 (1), 201–251.

[18] Hayakawa, K. (2008). On the Effect of Nonstationary Initial Conditions in Dynamic Panel Data Models. Hi-Stat DP series, No. 245, Hitotsubashi University.

[19] Hayakawa, K. (2009). On the Effect of Mean-Nonstationarity in Dynamic Panel Data Models. *Journal of Econometrics*, 153, 133–135.

[20] Hayakawa, K. (2010). The Effects of Dynamic Feedbacks on LS and MM Estimator Accuracy in Panel Data Models: Some Additional Results. *Journal of Econometrics*, 159(1), 202-208.

[21] Hjort, N. L. and G. Claeskens (2003). Frequentist Model Average Estimators. *Journal of the American Statistical Association*, 98, 879-899.

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

[23] Kiviet, J. F. (1995). On Bias, Inconsistency, and Effciency of Various Etimators in Dynamic-Panel Data Models. *Journal of Econometrics*, 68, 53–78.

[24] Kiviet, J. F. (2007). Judging Contending Estimators by Simulation: Tournaments in Dynamic Panel Data Models. In: Phillips, G.D.A., and Tzavalis, E. (Eds.), *The Refinement of Econometric Estimation and Test Procedures*. Cambridge University Press, Cambridge, 282–318.

[25] Kruiniger, H. (2009). GMM Estimation of Dynamic Panel Data Models with Persistent Data. *Econometric Theory*, 25, 1348-1391.

[26] Moon, H. R. and P. C. B. Phillips (2000). Estimation of Autoregressive Roots Near Unity Using Panel Data. *Econometric Theory*, 16, 927-997,

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

[28] Phillips, P. C. B. and H. R. Moon (1999). Linear Regression Limit Theory for Nonstationary Panel Data. *Econometrica*, 67(5), 1057-1111.

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

[30] Phillips, P. C. B. (2014). Dynamic Panel GMM with Unit Root and Irrelevant Instruments: Exam Solution for Question A. Unpublished Manuscript, Yale University.

8 Appendix

This appendix is divided into two parts. In the first part, we prove the main results of the paper. In the second part, we state and prove a number of supporting lemmas which are used in proving the main results. Throughout this appendix, we let C denote a generic positive constant that may be different in different uses.

8.1 Proofs of Theorems

Proof of Theorem 1:

To show part (a), first write

$$\sqrt{NT}\left(\widehat{\rho}_{IV}-\rho\right)=\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}y_{it-2}\Delta y_{it-1}\right)^{-1}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=2}^{T}y_{it-2}\Delta\varepsilon_{it}.$$

Note that by tedious but straightforward calculations, we can show that

$$\begin{aligned}&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}y_{it-2}\Delta y_{it-1}\\ =\;&-(1-\rho)\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}\sum_{j=0}^{t-3}\sum_{k=0}^{t-3}\rho^{j+k}\varepsilon_{it-2-j}\varepsilon_{it-2-k}+O_p\left(T^{-1}\right)+O_p\left(N^{-1}\right)\\ =\;&-(1-\rho)\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}\sum_{j=0}^{t-3}\sum_{k=0}^{t-3}\rho^{j+k}\varepsilon_{it-2-j}\varepsilon_{it-2-k}+O_p\left(T^{-1}\right)\end{aligned}$$

and

$$-(1-\rho)\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}\sum_{j=0}^{t-3}\sum_{k=0}^{t-3}\rho^{j+k}\varepsilon_{it-2-j}\varepsilon_{it-2-k}\xrightarrow{p}-\left[\frac{\sigma^2}{1+\rho}\right]$$

as $N,T\to\infty$ such that $T/N\to\kappa\in[0,\infty)$. It follows that

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}y_{it-2}\Delta y_{it-1}\xrightarrow{p}-\left[\frac{\sigma^2}{1+\rho}\right]$$

as $N,T\to\infty$ such that $T/N\to\kappa\in[0,\infty)$.

Next, note that

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=2}^{T}y_{it-2}\Delta\varepsilon_{it}=\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=3}^{T}w_{it-2}\Delta\varepsilon_{it}+O_p\left(T^{-1/2}\right)$$

where

$$w_{it-2}=\sum_{k=0}^{t-3}\rho^{k}\varepsilon_{it-2-k}.$$

From the proof of Lemma A1 below, it is evident that

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=3}^{T}w_{it-2}\Delta\varepsilon_{it}\Longrightarrow N\left(0,\frac{2\sigma^4}{1+\rho}\right).$$

as $N,T\to\infty$ such that $T/N\to\kappa\in[0,\infty)$.

It follows by the Cramér Convergence Theorem that

$$\sqrt{NT}\left(\widehat{\rho}_{IV}-\rho\right)\Longrightarrow-\left(\frac{1+\rho}{\sigma^2}\right)N\left(0,\frac{2\sigma^4}{1+\rho}\right)\equiv N\left(0,2\left[1+\rho\right]\right),$$

as required.

Part (b) can be shown by modifying the argument given in Phillips (2014) for the case with Gaussian errors. For brevity, the proof is omitted here. □

Proof of Theorem 2:

To show part (a), write

$$\begin{aligned}
&\widehat{\rho}_{\text{pols}}\\
=&\left[\sum_{i=1}^{N}\sum_{t=2}^{T}\left(y_{it-1}-\overline{y}_{-1,N,T}\right)^2\right]^{-1}\\
&\times\left[\sum_{i=1}^{N}\sum_{t=2}^{T}\left(y_{it-1}-\overline{y}_{-1,N,T}\right)\left(y_{it}-\overline{y}_{N,T}\right)\right]\\
=&\left[\sum_{i=1}^{N}\sum_{t=2}^{T}\left(y_{it-1}-\overline{y}_{-1,N,T}\right)^2\right]^{-1}\\
&\times\sum_{i=1}^{N}\sum_{t=2}^{T}\left(y_{it-1}-\overline{y}_{-1,N,T}\right)\left(a_i\left(1-\rho\right)+\rho y_{it-1}+\varepsilon_{it}-\frac{1}{N\left(T-1\right)}\sum_{i=1}^{N}\sum_{t=2}^{T}\left[a_i\left(1-\rho\right)+\rho y_{it-1}+\varepsilon_{it}\right]\right)
\end{aligned}$$

where

$$\overline{y}_{-1,N,T}=\frac{1}{N\left(T-1\right)}\sum_{i=1}^{N}\sum_{t=2}^{T}y_{it-1},\ \overline{y}_{N,T}=\frac{1}{N\left(T-1\right)}\sum_{i=1}^{N}\sum_{t=2}^{T}y_{it},$$

It follows that

$$\begin{aligned}
&\widehat{\rho}_{\text{pols}}-\rho\\
=&\left[\sum_{i=1}^{N}\sum_{t=2}^{T}\left(y_{it-1}-\overline{y}_{N,T}\right)^2\right]^{-1}\sum_{i=1}^{N}\sum_{t=2}^{T}\left(y_{it-1}-\overline{y}_{N,T}\right)\left[a_i-\overline{a}_N\right]\left(1-\rho\right)\\
&+\left[\sum_{i=1}^{N}\sum_{t=2}^{T}\left(y_{it-1}-\overline{y}_{-1,N,T}\right)^2\right]^{-1}\sum_{i=1}^{N}\sum_{t=2}^{T}\left(y_{it-1}-\overline{y}_{-1,N,T}\right)\left(\varepsilon_{it}-\overline{\varepsilon}_{N,T}\right)\\
=&\left[\sum_{i=1}^{N}\sum_{t=2}^{T}\left(y_{it-1}-\overline{y}_{-1,N,T}\right)^2\right]^{-1}\sum_{i=1}^{N}\sum_{t=2}^{T}\left(y_{it-1}-\overline{y}_{-1,N,T}\right)\left[a_i\left(1-\rho\right)+\varepsilon_{it}\right]
\end{aligned}$$

where

$$\overline{a}_N = \frac{1}{N}\sum_{i=1}^{N} a_i,\ \overline{\varepsilon}_{N,T} = \frac{1}{N(T-1)}\sum_{i=1}^{N}\sum_{t=2}^{T}\varepsilon_{it}.$$

We can show by tedious but straightforward calculations that, as $N, T \to \infty$ such that $T/N \to \kappa \in [0, \infty)$,

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}\left(y_{it-1} - \overline{y}_{-1,N,T}\right)^2 = \sigma_a^2 + \frac{\sigma^2}{1-\rho^2} + O_p\left(T^{-1}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right)$$

and

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}\left(y_{it-1} - \overline{y}_{-1,N,T}\right)\left[a_i(1-\rho) + \varepsilon_{it}\right] = (1-\rho)\sigma_a^2 + O_p\left(\frac{1}{T}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right).$$

It follows by Slutsky's theorem that

$$\begin{aligned} & \widehat{\rho}_{\text{pols}} - \rho \\ = & \left[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}\left(y_{it-1} - \overline{y}_{-1,N,T}\right)^2\right]^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}\left(y_{it-1} - \overline{y}_{-1,N,T}\right)\left[a_i(1-\rho) + \varepsilon_{it}\right] \\ = & \left[\sigma_a^2 + \frac{\sigma^2}{1-\rho^2}\right]^{-1}(1-\rho)\sigma_a^2\left[1 + O_p\left(\frac{1}{T}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right)\right] \end{aligned}$$

so that

$$\widehat{\rho}_{\text{pols}} \xrightarrow{p} \left[\sigma_a^2 + \frac{\sigma^2}{1-\rho^2}\right]^{-1}(1-\rho)\sigma_a^2 = \frac{\sigma_a^2(1-\rho)\left(1-\rho^2\right)}{\sigma_a^2(1-\rho^2) + \sigma^2}$$

as required.

To show part (b), write

$$\begin{aligned} & \sqrt{N}T\left(\widehat{\rho}_{\text{pols}} - 1\right) \\ = & \left[\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T}\left(y_{it-1} - \overline{y}_{N,T}\right)^2\right]^{-1}\frac{1}{\sqrt{N}T}\sum_{i=1}^{N}\sum_{t=2}^{T}\left(y_{it-1} - \overline{y}_{N,T}\right)\varepsilon_{it} \end{aligned}$$

Define

$$w_{it-2} = \sum_{j=1}^{t-2}\varepsilon_{ij}$$

and note that by direct calculation and using Lemma A6 below, we obtain, as $N, T \to \infty$ such that $T/N \to \kappa \in [0, \infty)$,

$$
\begin{aligned}
& \left[\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T}\left(y_{it-1}-\overline{y}_{N,T}\right)^2\right]^{-1}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=2}^{T}\left(y_{it-1}-\overline{y}_{N,T}\right)\varepsilon_{it} \\
= & \left[\frac{\sigma^2}{2}+O_p\left(\frac{1}{T}\right)\right]^{-1}\left[\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=2}^{T}w_{it-1}\varepsilon_{it}+O_p\left(\frac{1}{\sqrt{T}}\right)\right] \\
= & \frac{2}{\sigma^2}\left[\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=2}^{T}w_{it-1}\varepsilon_{it}+O_p\left(\frac{1}{\sqrt{T}}\right)\right]\left[1+O_p\left(\frac{1}{T}\right)\right] \\
= & \frac{2}{\sigma^2}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=2}^{T}w_{it-1}\varepsilon_{it}\left[1+O_p\left(\frac{1}{\sqrt{T}}\right)\right] \\
\Longrightarrow & \frac{2}{\sigma^2}N\left(0,\frac{\sigma^4}{2}\right) \\
\equiv & N\left(0,2\right).
\end{aligned}
$$

as required. □

Proof of Theorem 3:

We will prove this result only for the case where the weighting function is based on a logistic type transformation, i.e.,

$$
w_{IC,1}=\frac{1}{1+\exp\left\{\frac{1}{2}\Delta_{IC}\right\}},
$$

Asymptotic results when the weight function is of the form given by (20) can be proved similarly; hence, for purposes of brevity, we omit its proof.

To proceed, consider first part (a), where we assume that $\rho\in(0,1)$. In this case, note that by part (a) of Theorems 1 and 2 and Lemma A7, we have that

$$
\sqrt{NT}\left(\widehat{\rho}_{IV}-\rho\right)\Longrightarrow N\left(0,2\left[1+\rho\right]\right),
$$

$$
\widehat{\rho}_{\text{pols}}=\rho+\frac{\sigma_a^2\left(1-\rho\right)\left(1-\rho^2\right)}{\sigma_a^2\left(1-\rho^2\right)+\sigma^2}+O_p\left(T^{-1}\right),
$$

and

$$
\frac{\mathbb{T}_{NT}}{\left(NT\right)^{\frac{1}{2}-\epsilon}}\xrightarrow{p}-\infty
$$

for any $\epsilon>0$, as $N,T\to\infty$ such that $T/N\to\kappa\in[0,\infty)$.

It follows that in this case,

$$
\Delta_{IC}=\mathbb{T}_{NT}+\alpha\ln\left(N\right)+\beta\ln\left(T\right)\sim-\sqrt{NT}
$$

which implies that

$$
w_{IC}=\frac{1}{1+\exp\left\{\frac{1}{2}\Delta_{IC}\right\}}=1+O_p\left(\exp\left\{-\frac{1}{2}\sqrt{NT}\right\}\right).
$$

Hence,

$$
\begin{aligned}
& \nu_{NT}\left(\widehat{\rho}_{\text{avg}}-\rho\right) \\
= \quad & \left\{ w_{IC}\left(\frac{2}{NT}\left[1+\widehat{\rho}_{\text{IV}}\right]\right)^{-1/2} + \left(1-w_{IC}\right)\widehat{\sigma}^{-1}\left[\sum_{i=1}^{N}\sum_{t=2}^{T}\left(y_{it-1}-\overline{y}_{NT}\right)^{2}\right]^{1/2}\right\} \\
& \times\left(w_{IC}\widehat{\rho}_{\text{IV}}+\left(1-w_{IC}\right)\widehat{\rho}_{\text{pols}}-\rho\right) \\
= \quad & \left\{\sqrt{NT}\left[1+O_{p}\left(\exp\left\{-\frac{1}{2}\sqrt{NT}\right\}\right)\right]\left(2\left[1+\widehat{\rho}_{\text{IV}}\right]\right)^{-1/2}\right. \\
& \left.+\left[1-1+O_{p}\left(\exp\left\{-\frac{1}{2}\sqrt{NT}\right\}\right)\right]O_{p}\left(\sqrt{NT}\right)\right\} \\
& \times\left\{\left[1+O_{p}\left(\exp\left\{-\frac{1}{2}\sqrt{NT}\right\}\right)\right]\widehat{\rho}_{\text{IV}}-\rho+\left[1-1+O_{p}\left(\exp\left\{-\frac{1}{2}\sqrt{NT}\right\}\right)\right]O_{p}\left(1\right)\right\} \\
= \quad & \left[\sqrt{NT}\left(2\left[1+\widehat{\rho}_{\text{IV}}\right]\right)^{-1/2}+O_{p}\left(\sqrt{NT}\exp\left\{-\frac{1}{2}\sqrt{NT}\right\}\right)\right] \\
& \times\left[\widehat{\rho}_{\text{IV}}-\rho+O_{p}\left(\exp\left\{-\frac{1}{2}\sqrt{NT}\right\}\right)\right] \\
= \quad & \left(2\left[1+\widehat{\rho}_{\text{IV}}\right]\right)^{-1/2}\sqrt{NT}\left(\widehat{\rho}_{\text{IV}}-\rho_{T}\right)+O_{p}\left(\sqrt{NT}\exp\left\{-\frac{1}{2}\sqrt{NT}\right\}\right) \\
\Longrightarrow \quad & N\left(0,1\right)
\end{aligned}
$$

Next, consider part (b), where we assume that $\rho = 1$. Now, by Theorem 1 part (b), we have, as $N, T \to \infty$ such that $T/N \to \kappa \in [0, \infty)$,

$$\sqrt{T}\left(\widehat{\rho}_{IV}-\rho_{N,T}\right)\Longrightarrow\frac{2\mathcal{Z}_{1}}{\mathcal{Z}_{2}}$$

where $\mathcal{Z}_1 \equiv N(0,1)$, $\mathcal{Z}_2 \equiv N(0,1)$, and $\mathcal{Z}_1$ and $\mathcal{Z}_2$ are independent. Moreover, by Theorem 2 part (b) and Lemma A7,

$$
\begin{aligned}
\sqrt{N}T\left(\widehat{\rho}_{\text{pols}}-\rho_{T}\right) &\Longrightarrow N\left(0,2\right), \\
\mathbb{T}_{NT} &= O_{p}\left(1\right)
\end{aligned}
$$

as $N, T \to \infty$ such that $T/N \to \kappa \in [0, \infty)$.

It follows that in this case

$$\Delta_{IC}=\mathbb{T}_{NT}+\alpha\ln\left(N\right)+\beta\ln\left(T\right)=O_{p}\left(\max\left\{\alpha\ln\left(N\right),\beta\ln\left(T\right)\right\}\right)$$

which implies that

$$
\begin{aligned}
w_{IC} &= \frac{1}{1+\exp\left\{\frac{1}{2}\Delta_{IC}\right\}} \\
&= \frac{1}{1+O_p\left(\exp\left\{\max\left\{\alpha\ln\left(N\right),\beta\ln\left(T\right)\right\}\right\}\right)} \\
&= \frac{1}{1+O_p\left(\max\left\{N^{\alpha},T^{\beta}\right\}\right)} \\
&= O_p\left(\frac{1}{\max\left\{N^{\alpha},T^{\beta}\right\}}\right)=o_p\left(1\right).
\end{aligned}
$$

It follows that

$$
\begin{aligned}
& \nu_{NT}\left(\widehat{\rho}_{\text{avg}}-1\right) \\
= & \left\{w_{IC}\left(\frac{2}{NT}\left[1+\widehat{\rho}_{\text{IV}}\right]\right)^{-1/2}+\left(1-w_{IC}\right)\widehat{\sigma}^{-1}\left[\sum_{i=1}^{N}\sum_{t=2}^{T}\left(y_{it-1}-\overline{y}_{NT}\right)^2\right]^{1/2}\right\} \\
& \times\left(w_{IC}\widehat{\rho}_{\text{IV}}+\left(1-w_{IC}\right)\widehat{\rho}_{\text{pols}}-1\right) \\
= & \left\{O_p\left(\frac{1}{\max\left\{N^{\alpha},T^{\beta}\right\}}\sqrt{NT}\right)\right. \\
& \left.+\left(1-O_p\left(\frac{1}{\max\left\{N^{\alpha},T^{\beta}\right\}}\right)\right)\widehat{\sigma}^{-1}\left[\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T}\left(y_{it-1}-\overline{y}_{NT}\right)^2\right]^{1/2}\sqrt{N}T\right\} \\
& \times\left[\widehat{\rho}_{\text{pols}}-1+O_p\left(\frac{1}{\max\left\{N^{\alpha},T^{\beta}\right\}}\right)\right] \\
= & \widehat{\sigma}^{-1}\left[\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T}\left(y_{it-1}-\overline{y}_{NT}\right)^2\right]^{1/2}\sqrt{N}T\left(\widehat{\rho}_{\text{pols}}-\rho_T\right)+O_p\left(\frac{1}{N^{\alpha-3/2}}\right) \\
= & \widehat{\sigma}^{-1}\left[\frac{\sigma^2}{2}+O_p\left(\frac{1}{T}\right)\right]^{1/2}\frac{2}{\sigma^2}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=2}^{T}w_{it-1}\varepsilon_{it} \\
& +\widehat{\sigma}^{-1}\left[\frac{\sigma^2}{2}+O_p\left(\frac{1}{T}\right)\right]^{1/2}O_p\left(\frac{1}{\sqrt{T}}\right)+O_p\left(\frac{1}{N^{\alpha-3/2}}\right)
\end{aligned}
$$

so that

$$
\begin{aligned}
\nu_{NT}\left(\widehat{\rho}_{\text{avg}}-1\right) &= \frac{1}{\sqrt{2}}\frac{2}{\sigma^2}\frac{\sigma^2}{\sqrt{2}}\mathcal{Z}+O_p\left(\frac{1}{\sqrt{T}}\right) \\
&= \mathcal{Z}+O_p\left(\frac{1}{\sqrt{T}}\right) \\
&\Longrightarrow \mathcal{Z},
\end{aligned}
$$

where $\mathcal{Z}\equiv N\left(0,1\right)$, as required. $\square$

8.2 Supporting Lemmas

Lemma A1:

Suppose that Assumptions 1-4 hold, and let $\rho \in (0,1)$. Then, as $N, T \to \infty$ such that $T/N \to \kappa \in [0, \infty)$,

$$\frac{1}{\omega_{N,T}} \sum_{i=1}^{N} \sum_{t=3}^{T} w_{it-2} \Delta\varepsilon_{it} \Longrightarrow N(0,1)$$

where

$$\begin{aligned} w_{it-2} &= \sum_{k=0}^{t-3} \rho^k \varepsilon_{it-2-k}, \\ \omega_{N,T}^2 &= \sum_{i=1}^{N} \omega_{i,T}^2 = \sum_{i=1}^{N} E\left[(X_{i,T} + Y_{i,T})^2 \right] \end{aligned}$$

and

$$\begin{aligned} X_{i,T} &= \frac{1}{\sqrt{T}} \sum_{t=3}^{T} \varepsilon_{it-2} \varepsilon_{it-1}, \\ Y_{i,T} &= (\rho - 1) \frac{1}{\sqrt{T}} \sum_{t=3}^{T} w_{it-3} \varepsilon_{it-1} \end{aligned}$$

Proof:

To proceed, note first that, by Abel's partial summation formula,

$$\begin{aligned} & \sum_{i=1}^{N} \sum_{t=3}^{T} w_{it-2} \Delta\varepsilon_{it} \\ = & \sum_{i=1}^{N} \left\{ \sum_{t=3}^{T} (w_{it-3} - w_{it-2})\, \varepsilon_{it-1} \right\} + \sum_{i=1}^{N} w_{iT-2} \varepsilon_{iT} - \sum_{i=1}^{N} w_{i1} \varepsilon_{i2} \\ = & \sum_{i=1}^{N} \sum_{t=4}^{T} \left(\sum_{k=0}^{t-4} \rho^k \varepsilon_{it-3-k} - \sum_{k=0}^{t-3} \rho^k \varepsilon_{it-2-k} \right) \varepsilon_{it-1} + \sum_{i=1}^{N} \sum_{k=0}^{T-3} \rho^k \varepsilon_{iT-2-k} \varepsilon_{iT} - \sum_{i=1}^{N} \varepsilon_{i1} \varepsilon_{i2} \end{aligned}$$

Note that

$$\begin{aligned} & E\left(\sum_{i=1}^{N} \sum_{k=0}^{T-3} \rho^k \varepsilon_{iT-2-k} \varepsilon_{iT} \right)^2 \\ = & \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=0}^{T-3} \sum_{\ell=0}^{T-3} \rho^k \rho^\ell E\left[\varepsilon_{iT-2-k} \varepsilon_{jT-2-\ell} \varepsilon_{iT} \varepsilon_{jT} \right] \\ = & \sigma^4 \sum_{i=1}^{N} \sum_{k=0}^{T-3} \rho^{2k} \\ = & \sigma^4 N \left[\frac{1 - \rho^{2(T-2)}}{1 - \rho^2} \right] \end{aligned}$$

so that

$$\sum_{i=1}^{N}\sum_{k=0}^{T-3}\rho^k\varepsilon_{iT-2-k}\varepsilon_{iT}=O_p\left(\sqrt{N}\right)$$

Moreover,

$$E\left(\sum_{i=1}^{N}\varepsilon_{i1}\varepsilon_{i2}\right)^2=\sum_{i=1}^{N}\sum_{j=1}^{N}E\left[\varepsilon_{i1}\varepsilon_{j1}\varepsilon_{i2}\varepsilon_{j2}\right]=\sigma^4N$$

so that

$$\sum_{i=1}^{N}\varepsilon_{i1}\varepsilon_{i2}=O_p\left(\sqrt{N}\right).$$

It follows that

$$\begin{aligned}
&\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=3}^{T}w_{it-2}\Delta\varepsilon_{it}\\
=&\ \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=3}^{T}\left(w_{it-3}-w_{it-2}\right)\varepsilon_{it-1}+O_p\left(\frac{1}{\sqrt{T}}\right)\\
=&\ -\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=3}^{T}\left(\varepsilon_{it-2}-w_{it-3}+\rho w_{it-3}\right)\varepsilon_{it-1}+O_p\left(\frac{1}{\sqrt{T}}\right)\\
=&\ -\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=3}^{T}\left[\varepsilon_{it-2}+\left(\rho-1\right)w_{it-3}\right]\varepsilon_{it-1}+O_p\left(\frac{1}{\sqrt{T}}\right).
\end{aligned}$$

Now, let

$$\underline{U}_{N,T}=-\sum_{i=1}^{N}\left(X_{i,T}+Y_{i,T}\right)$$

where $X_{i,T}$ and $Y_{i,T}$ are as defined in the statement of the lemma. Next, note that

$$\begin{aligned}
E\left[X_{i,T}\right] &= \frac{1}{\sqrt{T}}\sum_{t=4}^{T}E\left[\varepsilon_{it-2}\varepsilon_{it-1}\right]=0\\
E\left[Y_{i,T}\right] &= \left(\rho-1\right)\frac{1}{\sqrt{T}}\sum_{t=3}^{T}\sum_{k=0}^{t-3}\rho^kE\left[\varepsilon_{it-3-k}\varepsilon_{it-1}\right]\\
&= \left(\rho-1\right)\frac{1}{\sqrt{T}}\sum_{t=3}^{T}\sum_{k=0}^{t-3}\rho^kE\left[\varepsilon_{it-3-k}\right]E\left[\varepsilon_{it-1}\right]\\
&= 0
\end{aligned}$$

and, thus,

$$E\left[\underline{U}_{N,T}\right]=-\sum_{i=1}^{N}\left(E\left[X_{i,T}\right]+E\left[Y_{i,T}\right]\right)=0$$

In addition, note that

$$\begin{aligned}
E\left[X_{i,T}^2\right] &= \frac{1}{T}\sum_{t=4}^{T}\sum_{s=4}^{T} E\left[\varepsilon_{it-2}\varepsilon_{it-1}\varepsilon_{is-2}\varepsilon_{is-1}\right] = \frac{1}{T}\sum_{t=4}^{T} E\left[\varepsilon_{it-2}^2\right] E\left[\varepsilon_{it-1}^2\right] \\
&= \sigma^4\left(\frac{T-3}{T}\right) \\
&= \sigma^4 + O\left(\frac{1}{T}\right),
\end{aligned}$$

$$\begin{aligned}
E\left[Y_{i,T}^2\right] &= (\rho-1)^2\frac{1}{T}\sum_{t=3}^{T}\sum_{s=3}^{T}\sum_{k=0}^{t-3}\sum_{\ell=0}^{s-3}\rho^k\rho^\ell E\left[\varepsilon_{it-3-k}\varepsilon_{is-3-\ell}\varepsilon_{it-1}\varepsilon_{is-1}\right] \\
&= \sigma^4(1-\rho)^2\frac{1}{T}\sum_{t=3}^{T}\sum_{k=0}^{t-3}\rho^{2k} \\
&= \sigma^4(1-\rho)^2\frac{1}{T}\sum_{t=3}^{T}\frac{1-\rho^{2(t-2)}}{1-\rho^2} \\
&= \frac{\sigma^4}{T}\frac{(1-\rho)^2}{1-\rho^2}\left\{(T-2)-\rho^2\frac{1-\rho^{2(T-2)}}{1-\rho^2}\right\} \\
&= \frac{\sigma^4}{T}\frac{1-\rho}{1+\rho}\left\{(T-2)-\rho^2\frac{1-\rho^{2(T-2)}}{1-\rho^2}\right\} \\
&= \sigma^4\left(\frac{1-\rho}{1+\rho}\right)+O\left(\frac{1}{T}\right)
\end{aligned}$$

and

$$E\left[X_{i,T}Y_{i,T}\right] = (\rho-1)\frac{1}{T}\sum_{t=3}^{T}\sum_{s=3}^{T}\sum_{k=0}^{t-3}\rho^k E\left[\varepsilon_{it-3-k}\varepsilon_{it-1}\varepsilon_{is-2}\varepsilon_{is-1}\right] = 0$$

It follows that

$$\begin{aligned}
\omega_{i,T}^2 &= E\left[\left(X_{i,T}+Y_{i,T}\right)^2\right] \\
&= \sigma^4+\sigma^4\left(\frac{1-\rho}{1+\rho}\right)+O\left(\frac{1}{T}\right) \\
&= \sigma^4\frac{1+\rho+1-\rho}{1+\rho}+O\left(\frac{1}{T}\right) \\
&= \frac{2\sigma^4}{1+\rho}+O\left(\frac{1}{T}\right) \\
\omega_{N,T}^2 &= \sum_{i=1}^{N}\omega_{i,T}^2 = \frac{2\sigma^4}{1+\rho}N\left[1+O\left(\frac{1}{T}\right)\right]
\end{aligned}$$

Hence, to prove the required result, it suffices to show that

$$\frac{U_{N,T}}{\omega_{N,T}} = \frac{1}{\omega_{N,T}/\sqrt{N}}\frac{-1}{\sqrt{N}}\sum_{i=1}^{N}\left(X_{i,T}+Y_{i,T}\right)\xrightarrow{d} N(0,1)$$

as $N, T \to \infty$ such that $T/N \to \kappa \in [0, \infty)$. To show this, write

$$U_{N,T} = \frac{1}{\omega_{N,T}/\sqrt{N}} \frac{-1}{\sqrt{N}} \sum_{i=1}^{N} (X_{i,T} + Y_{i,T})$$

Next, note that by direct calculations, we have

$$\frac{\omega_{N,T}^2}{N} = \frac{1}{N}\sum_{i=1}^{N}\omega_{i,T}^2 = \frac{2\sigma^4}{1+\rho}\frac{1}{N}N\left[1 + O\left(\frac{1}{T}\right)\right] = \frac{2\sigma^4}{1+\rho} + O\left(\frac{1}{T}\right).$$

Since $0 < \sigma^4 < \infty$ by assumption, it follows that there exists a positive constant C

$$0 < \frac{1}{C} \leq \frac{\omega_{N,T}}{\sqrt{N}} \leq C < \infty \text{ eventually as } N, T \to \infty.$$

Hence, to show the asymptotic normality of $U_{N,T}$, it suffices to show that

$$\lim_{N,T\to\infty} \sum_{i=1}^{N} E\left[\frac{1}{\sqrt{N}} U_{N,T}\right]^4 = \lim_{N,T\to\infty} \frac{1}{N^2}\sum_{i=1}^{N} E\left[(X_{i,T} + Y_{i,T})^4\right] = 0$$

To show this, note first that by Loève's c_r inequality, we have that

$$\begin{aligned}
&\frac{1}{N^2}\sum_{i=1}^{N} E\left[(X_{i,T} + Y_{i,T})^4\right] \\
\leq\ & \frac{1}{N^2}\sum_{i=1}^{N} 8\left\{E\left[\left(\frac{1}{\sqrt{T}}\sum_{t=4}^{T}\varepsilon_{it-2}\varepsilon_{it-1}\right)^4\right] + E\left[\left((\rho-1)\frac{1}{\sqrt{T}}\sum_{t=3}^{T}\sum_{k=0}^{t-3}\rho^k\varepsilon_{it-2-k}\varepsilon_{it-1}\right)^4\right]\right\} \\
=\ & 8\frac{1}{N^2T^2}\sum_{i=1}^{N} E\left[\left(\sum_{t=4}^{T}\varepsilon_{it-2}\varepsilon_{it-1}\right)^4\right] + 8\frac{(\rho-1)^4}{N^2T^2}\sum_{i=1}^{N} E\left[\left(\sum_{t=3}^{T}\sum_{k=0}^{t-3}\rho^k\varepsilon_{it-2-k}\varepsilon_{it-1}\right)^4\right]
\end{aligned}$$

Now,

$$
\begin{aligned}
& E\left[\left(\sum_{t=4}^{T}\varepsilon_{it-2}\varepsilon_{it-1}\right)^4\right] \\
= \; & \sum_{r=4}^{T}\sum_{s=4}^{T}\sum_{t=4}^{T}\sum_{u=4}^{T}E\left[\varepsilon_{itr-2}\varepsilon_{is-2}\varepsilon_{it-2}\varepsilon_{iu-2}\varepsilon_{ir-1}\varepsilon_{is-1}\varepsilon_{it-1}\varepsilon_{iu-1}\right] \\
= \; & \sum_{t=4}^{T}E\left[\varepsilon_{it-2}^4\right]E\left[\varepsilon_{it-1}^4\right]+6\sum_{s=6}^{T}\sum_{t=4}^{s-2}E\left[\varepsilon_{is-2}^2\right]E\left[\varepsilon_{is-1}^2\right]E\left[\varepsilon_{it-2}^2\right]E\left[\varepsilon_{it-1}^2\right] \\
& +6\sum_{t=4}^{T-1}E\left[\varepsilon_{it}^2\right]E\left[\varepsilon_{it-1}^4\right]E\left[\varepsilon_{it-2}^2\right] \\
= \; & \left(E\left[\varepsilon_{it-1}^4\right]\right)^2(T-3)+6\sigma^8\sum_{s=6}^{T}(s-5)+6E\left[\varepsilon_{it-1}^4\right]\sigma^4(T-4) \\
= \; & 6\sigma^8\frac{(T-5)(T-4)}{2}+\left(E\left[\varepsilon_{it-1}^4\right]\right)^2(T-3)+6E\left[\varepsilon_{it-1}^4\right]\sigma^4(T-4) \\
= \; & 3\sigma^8T^2\left[1+O\left(\frac{1}{T}\right)\right]
\end{aligned}
$$

Moreover,

$$
\begin{aligned}
& E\left[\left(\sum_{t=3}^{T}\sum_{k=0}^{t-3}\rho^k\varepsilon_{it-2-k}\varepsilon_{it-1}\right)^4\right] \\
= \; & \sum_{r=3}^{T}\sum_{s=3}^{T}\sum_{t=3}^{T}\sum_{u=3}^{T}\sum_{h=0}^{r-3}\sum_{j=0}^{s-3}\sum_{k=0}^{t-3}\sum_{\ell=0}^{u-3}\rho^h\rho^j\rho^k\rho^\ell E\left[\varepsilon_{ir-2-h}\varepsilon_{is-2-j}\varepsilon_{it-2-k}\varepsilon_{iu-2-\ell}\varepsilon_{ir-1}\varepsilon_{is-1}\varepsilon_{it-1}\varepsilon_{iu-1}\right] \\
\leq \; & \overline{C}\left\{\sum_{t=3}^{T}\sum_{k=0}^{t-3}\rho^{4k}E\left[\varepsilon_{it-2-k}^4\right]E\left[\varepsilon_{it-1}^4\right]+\sum_{t=4}^{T}\sum_{s=3}^{t-1}\sum_{\ell=0}^{s-3}\rho^{2\ell}\rho^{2(t-s-1)}E\left[\varepsilon_{is-2-\ell}^2\right]E\left[\varepsilon_{is-1}^4\right]E\left[\varepsilon_{it-1}^2\right]\right. \\
& +\sum_{t=4}^{T}\sum_{s=3}^{t-1}\sum_{\ell=0}^{s-3}\rho^{2\ell}\rho^{2(t-s+\ell)}E\left[\varepsilon_{is-2-\ell}^4\right]E\left[\varepsilon_{is-1}^2\right]E\left[\varepsilon_{it-1}^2\right] \\
& +\sigma^8\sum_{t=5}^{T}\sum_{s=4}^{t-1}\sum_{r=3}^{s-1}\sum_{h=0}^{r-3}\rho^h\rho^{t-s-1}\rho^{s-r-1}\rho^{t-r+h} \\
& \left|+\sum_{t=5}^{T}\sum_{r=3}^{t-1}\sum_{h=0}^{r-3}\rho^h\rho^{2(t-r-1)}\rho^{t-r+h}\left|E\left[\varepsilon_{it-1}^3\right]\right|\left|E\left[\varepsilon_{ir-1}^3\right]\right|\sigma^2\right\}
\end{aligned}
$$

for some positive constant $\overline{C}$. Thus, given Assumption 1, there exists a positive constant C

$$
\begin{aligned}
& E\left[\left(\sum_{t=3}^{T}\sum_{k=0}^{t-3}\rho^{k}\varepsilon_{it-2-k}\varepsilon_{it-1}\right)^{4}\right] \\
\leq\ & C\left[\sum_{t=3}^{T}\sum_{k=0}^{t-3}\rho^{4k}+\sum_{t=4}^{T}\sum_{s=3}^{t-1}\rho^{2(t-s-1)}\sum_{\ell=0}^{s-3}\rho^{2\ell}+\sum_{t=4}^{T}\sum_{s=3}^{t-1}\rho^{2(t-s)}\sum_{\ell=0}^{s-3}\rho^{4\ell}\right. \\
& \left.+\sum_{t=5}^{T}\sum_{s=4}^{t-1}\sum_{r=3}^{s-1}\sum_{h=0}^{r-3}\rho^{2(t-r-1+h)}+\sum_{t=5}^{T}\sum_{r=3}^{t-1}\sum_{h=0}^{r-3}\rho^{h}\rho^{2(t-r-1)}\rho^{t-r+h}\right] \\
=\ & C\left[\sum_{t=3}^{T}\frac{1-\rho^{4(t-2)}}{1-\rho^{4}}+\sum_{t=4}^{T}\sum_{s=3}^{t-1}\rho^{2(t-1-s)}\frac{1-\rho^{2(s-2)}}{1-\rho^{2}}+\sum_{t=4}^{T}\sum_{s=3}^{t-1}\rho^{2(t-s)}\frac{1-\rho^{4(s-2)}}{1-\rho^{4}}\right. \\
& \left.+\sum_{t=5}^{T}\sum_{s=4}^{t-1}\rho^{2(t-s)}\sum_{r=3}^{s-1}\rho^{2(s-1-r)}\frac{1-\rho^{2(r-2)}}{1-\rho^{2}}+\sum_{t=5}^{T}\rho^{3t}\sum_{r=3}^{t-1}\rho^{-(3r+2)}\frac{1-\rho^{2(r-2)}}{1-\rho^{2}}\right]
\end{aligned}
$$

$$
\begin{aligned}
=\ & C\left[\sum_{t=3}^{T}\frac{1-\rho^{4(t-2)}}{1-\rho^{4}}+\frac{1}{(1-\rho^{2})^{2}}\sum_{t=4}^{T}\left(1-\rho^{2(t-3)}\right)\right. \\
& -\frac{1}{(1-\rho^{2})}\sum_{t=4}^{T}(t-3)\,\rho^{2(t-3)}+\frac{1}{(1-\rho^{4})(1-\rho^{2})}\sum_{t=4}^{T}\rho^{2}\left(1-\rho^{2(t-3)}\right) \\
& -\frac{1}{(1-\rho^{4})(1-\rho^{2})}\sum_{t=4}^{T}\rho^{2(t-1)}\left(1-\rho^{2(t-3)}\right)+\frac{1}{(1-\rho^{2})^{2}}\sum_{t=5}^{T}\sum_{s=4}^{t-1}\rho^{2(t-s)}\left(1-\rho^{2(s-3)}\right) \\
& -\frac{1}{(1-\rho^{2})}\sum_{t=5}^{T}\rho^{2(t-3)}\frac{(t-4)(t-3)}{2}+\frac{1}{(1-\rho^{3})(1-\rho^{2})}\sum_{t=5}^{T}\rho\left(1-\rho^{3(t-3)}\right) \\
& \left.-\frac{1}{(1-\rho^{2})(1-\rho)}\sum_{t=5}^{T}\rho^{2(t-1)-3}\left(1-\rho^{(t-3)}\right)\right] \\
=\ & C\left[I_1+I_2+I_3+I_4+I_5+I_6+I_7+I_8+I_9\right]
\end{aligned}
$$

Next, note that using parts (a), (b), and (c) of Lemma A2, we obtain

$$
\begin{aligned}
& I_3 \\
=\ & -\frac{1}{(1-\rho^{2})}\sum_{t=4}^{T}(t-3)\,\rho^{2(t-3)} \\
=\ & -\frac{\rho}{2(1-\rho^{2})}\left\{\frac{2\rho-[2(T-3)+1]\,\rho^{2(T-3)}}{1-\rho}+\frac{\rho^{2}\left(1-\rho^{2(T-3)-1}\right)}{(1-\rho)^{2}}\right\} \\
=\ & O(1)
\end{aligned}
$$

$$
\begin{aligned}
& I_6 \\
= \; & \frac{1}{\left(1-\rho^2\right)^2} \sum_{t=5}^{T} \sum_{s=4}^{t-1} \rho^{2(t-s)} \left(1-\rho^{2(s-3)}\right) \\
= \; & \frac{\rho^2}{\left(1-\rho^2\right)^3} \sum_{t=5}^{T} \left(1-\rho^{2(t-4)}\right) - \frac{1}{\left(1-\rho^2\right)^2} \sum_{t=5}^{T} \rho^{2(t-3)} (t-4) \\
= \; & \frac{\rho^2}{\left(1-\rho^2\right)^3} \left\{ (T-4) - \rho^2 \frac{1-\rho^{2(T-4)}}{1-\rho^2} \right\} - \frac{1}{\left(1-\rho^2\right)^2} \sum_{t=5}^{T} \rho^{2(t-3)} (t-4) \\
= \; & \frac{\rho^2}{\left(1-\rho^2\right)^3} \left\{ (T-4) - \rho^2 \frac{1-\rho^{2(T-4)}}{1-\rho^2} \right\} \\
& - \frac{\rho}{2\left(1-\rho^2\right)^2} \left\{ \frac{4\rho^3 - \left[2(T-3)+1\right]\rho^{2(T-3)}}{1-\rho} + \frac{\rho^4\left(1-\rho^{2T-9}\right)}{(1-\rho)^2} - \frac{\rho^3\left(1-\rho^{2(T-4)}\right)}{\left(1-\rho^2\right)} \right\} \\
= \; & O(T)
\end{aligned}
$$

$$
\begin{aligned}
& I_7 \\
= \; & -\frac{1}{\left(1-\rho^2\right)} \sum_{t=5}^{T} \rho^{2(t-3)} \frac{(t-4)(t-3)}{2} \\
= \; & -\frac{1}{2\left(1-\rho^2\right)} \frac{\rho^2}{4} \left\{ \frac{12\rho^2 - \left[2(T-3)+1\right] 2(T-3) \rho^{2(T-3)-1}}{1-\rho} + \frac{8\rho^3 - 2\left[2(T-3)+1\right]\rho^{2(T-3)}}{(1-\rho)^2} \right. \\
& \left. + \frac{2\left[\rho^4 - \rho^{2(T-3)+1}\right]}{(1-\rho)^3} \right\} + \frac{1}{2\left(1-\rho^2\right)} \frac{\rho}{4} \left\{ \frac{4\rho^3 - \left[2(T-3)+1\right]\rho^{2(T-3)}}{1-\rho} + \frac{\rho^4 - \rho^{2(T-3)+1}}{(1-\rho)^2} \right\} \\
= \; & -\frac{\rho^2}{8\left(1-\rho^2\right)} \left\{ \frac{12\rho^2 - \left[2(T-3)+1\right] 2(T-3) \rho^{2(T-3)-1}}{1-\rho} + \frac{8\rho^3 - 2\left[2(T-3)+1\right]\rho^{2(T-3)}}{(1-\rho)^2} \right. \\
& \left. + \frac{2\left[\rho^4 - \rho^{2(T-3)+1}\right]}{(1-\rho)^3} \right\} + \frac{\rho}{8\left(1-\rho^2\right)} \left\{ \frac{4\rho^3 - \left[2(T-3)+1\right]\rho^{2(T-3)}}{1-\rho} + \frac{\rho^4 - \rho^{2(T-3)+1}}{(1-\rho)^2} \right\} \\
= \; & O(1).
\end{aligned}
$$

In addition, by direct calculations, we have

$$
I_1 = \sum_{t=3}^{T} \frac{1-\rho^{4(t-2)}}{1-\rho^4} = \frac{1}{1-\rho^4} \left\{ (T-2) - \rho^4 \frac{1-\rho^{4(T-2)}}{1-\rho^4} \right\} = O(T)
$$

$$
\begin{aligned}
I_2 \; = \; & \frac{1}{\left(1-\rho^2\right)^2} \sum_{t=4}^{T} \left(1-\rho^{2(t-3)}\right) \\
= \; & \frac{1}{\left(1-\rho^2\right)^2} \left\{ (T-3) - \rho^2 \frac{1-\rho^{2(T-3)}}{1-\rho^2} \right\} \\
= \; & O(T)
\end{aligned}
$$

$$
\begin{aligned}
I_4 &= \frac{1}{(1-\rho^4)(1-\rho^2)} \sum_{t=4}^{T} \rho^2 \left(1-\rho^{2(t-3)}\right) \\
&= \frac{\rho^2}{(1-\rho^4)(1-\rho^2)} \left\{ (T-3) - \rho^2 \frac{1-\rho^{2(T-3)}}{1-\rho^2} \right\} \\
&= O(T)
\end{aligned}
$$

$$
\begin{aligned}
I_5 &= -\frac{1}{(1-\rho^4)(1-\rho^2)} \sum_{t=4}^{T} \rho^{2(t-1)} \left(1-\rho^{2(t-3)}\right) \\
&= -\frac{\rho^6}{(1-\rho^4)(1-\rho^2)^2} \left(1-\rho^{2(T-3)}\right) + \frac{\rho^8}{(1-\rho^4)^2(1-\rho^2)} \left(1-\rho^{4(T-3)}\right) \\
&= O(1)
\end{aligned}
$$

$$
\begin{aligned}
I_8 &= \frac{1}{(1-\rho^3)(1-\rho^2)} \sum_{t=5}^{T} \rho \left(1-\rho^{3(t-3)}\right) \\
&= \frac{\rho}{(1-\rho^3)(1-\rho^2)} \left\{ (T-4) - \rho^6 \frac{1-\rho^{3(T-4)}}{1-\rho^3} \right\} \\
&= O(T)
\end{aligned}
$$

$$
\begin{aligned}
I_9 &= -\frac{1}{(1-\rho^2)(1-\rho)} \sum_{t=5}^{T} \rho^{2(t-1)-3} \left(1-\rho^{(t-3)}\right) \\
&= -\frac{\rho^5\left(1-\rho^{2(T-4)}\right)}{(1-\rho^2)^2(1-\rho)} + \frac{\rho^7\left(1-\rho^{3(T-4)}\right)}{(1-\rho^3)(1-\rho^2)(1-\rho)} \\
&= O(1)
\end{aligned}
$$

Putting these results together, we have

$$
E\left[\left(\sum_{t=3}^{T}\sum_{k=0}^{t-3} \rho^k \varepsilon_{it-2-k}\varepsilon_{it-1}\right)^4\right] = O(T)
$$

Hence,

$$
\begin{aligned}
&\frac{1}{N^2}\sum_{i=1}^{N} E\left[(X_{i,T}+Y_{i,T})^4\right] \\
\leq\ & 8\frac{1}{N^2T^2}\sum_{i=1}^{N} E\left[\left(\sum_{t=4}^{T}\varepsilon_{it-2}\varepsilon_{it-1}\right)^4\right] + 8\frac{(\rho-1)^4}{N^2T^2}\sum_{i=1}^{N} E\left[\left(\sum_{t=3}^{T}\sum_{k=0}^{t-3}\rho^k\varepsilon_{it-2-k}\varepsilon_{it-1}\right)^4\right] \\
=\ & O\left(\frac{1}{N^2T^2}NT^2\right) + O\left(\frac{1}{N^2T^2}NT\right) \\
=\ & O\left(N^{-1}\right),
\end{aligned}
$$

as required. □

Lemma A2:

(a)

$$\sum_{t=4}^{T}(t-1)\rho^{2(t-3)}$$
$$= \frac{\rho}{2}\left\{\frac{2\rho-[2(T-3)+1]\rho^{2(T-3)}}{1-\rho}+\frac{\rho^2\left(1-\rho^{2(T-3)-1}\right)}{(1-\rho)^2}\right\}+\frac{2\rho^2\left(1-\rho^{2(T-3)}\right)}{1-\rho^2}$$

(b)

$$\sum_{t=5}^{T}\rho^{2(t-3)}(t-4)$$
$$= \frac{\rho}{2}\left\{\frac{4\rho^3-[2(T-3)+1]\rho^{2(T-3)}}{1-\rho}+\frac{\rho^4\left(1-\rho^{2T-9}\right)}{(1-\rho)^2}\right\}-\frac{\rho^4\left(1-\rho^{2(T-4)}\right)}{2(1-\rho^2)}$$
$$= \frac{\rho}{2}\left\{\frac{4\rho^3-[2(T-3)+1]\rho^{2(T-3)}}{1-\rho}+\frac{\rho^4\left(1-\rho^{2T-9}\right)}{(1-\rho)^2}-\frac{\rho^3\left(1-\rho^{2(T-4)}\right)}{1-\rho^2}\right\}$$

(c)

$$\begin{aligned}\sum_{t=5}^{T}(t-4)(t-3)\rho^{2(t-3)} = \; & \frac{\rho^2}{4}\left\{\frac{12\rho^2-[2(T-3)+1]\,2(T-3)\,\rho^{2(T-3)-1}}{1-\rho}\right.\\ & \left.+\frac{8\rho^3-2[2(T-3)+1]\rho^{2(T-3)}}{(1-\rho)^2}+\frac{2\left[\rho^4-\rho^{2(T-3)+1}\right]}{(1-\rho)^3}\right\}\\ & -\frac{\rho}{4}\left\{\frac{4\rho^3-[2(T-3)+1]\rho^{2(T-3)}}{1-\rho}+\frac{\rho^4-\rho^{2(T-3)+1}}{(1-\rho)^2}\right\}\end{aligned}$$

Proof:

To show part (a), note that

$$\begin{aligned}
&\sum_{t=4}^{T}(t-3)\,\rho^{2(t-3)} \\
=\;& \frac{\rho}{2}\sum_{t=4}^{T}2\,(t-3)\,\rho^{2(t-3)-1} \\
=\;& \frac{\rho}{2}\sum_{s=2}^{2(T-3)} s\rho^{s-1} \\
=\;& \frac{\rho}{2}\frac{\partial}{\partial\rho}\left\{\rho^{2}\sum_{s=2}^{2(T-3)}\rho^{s-2}\right\} \\
=\;& \frac{\rho}{2}\frac{\partial}{\partial\rho}\left\{\rho^{2}\frac{1-\rho^{2(T-3)-1}}{1-\rho}\right\} \\
=\;& \frac{\rho}{2}\frac{\partial}{\partial\rho}\left\{\frac{\rho^{2}-\rho^{2(T-3)+1}}{1-\rho}\right\} \\
=\;& \frac{\rho}{2}\left\{\frac{2\rho-[2\,(T-3)+1]\,\rho^{2(T-3)}}{1-\rho}+\frac{\rho^{2}\left(1-\rho^{2(T-3)-1}\right)}{(1-\rho)^{2}}\right\}
\end{aligned}$$

To show part (b), note that

$$\begin{aligned}
&\sum_{t=5}^{T}(t-4)\,\rho^{2(t-3)} \\
=\;& \frac{\rho}{2}\sum_{t=5}^{T}2\,(t-3-1)\,\rho^{2(t-3)-1} \\
=\;& \frac{\rho}{2}\sum_{t=5}^{T}2\,(t-3)\,\rho^{2(t-3)-1}-\frac{\rho}{2}\sum_{t=5}^{T}\rho^{2(t-3)-1} \\
=\;& \frac{\rho}{2}\sum_{s=4}^{2(T-3)} s\rho^{s-1}-\frac{\rho^{4}}{2}\sum_{t=5}^{T}\rho^{2(t-5)} \\
=\;& \frac{\rho}{2}\frac{\partial}{\partial\rho}\left\{\rho^{4}\sum_{s=4}^{2(T-3)}\rho^{s-4}\right\}-\frac{\rho^{4}\left(1-\rho^{2(T-4)}\right)}{2\,(1-\rho^{2})} \\
=\;& \frac{\rho}{2}\frac{\partial}{\partial\rho}\left\{\rho^{4}\frac{1-\rho^{2(T-3)-3}}{1-\rho}\right\}-\frac{\rho^{4}\left(1-\rho^{2(T-4)}\right)}{2\,(1-\rho^{2})} \\
=\;& \frac{\rho}{2}\left\{\frac{4\rho^{3}-[2\,(T-3)+1]\,\rho^{2(T-3)}}{1-\rho}+\frac{\rho^{4}\left(1-\rho^{2T-9}\right)}{(1-\rho)^{2}}\right\}-\frac{\rho^{4}\left(1-\rho^{2(T-4)}\right)}{2\,(1-\rho^{2})} \\
=\;& \frac{\rho}{2}\left\{\frac{4\rho^{3}-[2\,(T-3)+1]\,\rho^{2(T-3)}}{1-\rho}+\frac{\rho^{4}\left(1-\rho^{2T-9}\right)}{(1-\rho)^{2}}-\frac{\rho^{3}\left(1-\rho^{2(T-4)}\right)}{1-\rho^{2}}\right\}
\end{aligned}$$

Finally, to show part (c), note that

$$
\begin{aligned}
& \sum_{t=5}^{T} (t-4)(t-3)\rho^{2(t-3)} \\
= & \frac{\rho^2}{4}\sum_{t=5}^{T}[2(t-3)-1]\,2(t-3)\rho^{2(t-3)-2} - \frac{\rho^2}{4}\sum_{t=5}^{T} 2(t-3)\rho^{2(t-3)-2} \\
= & \frac{\rho^2}{4}\sum_{t=5}^{T}[2(t-3)-1]\,2(t-3)\rho^{2(t-3)-2} - \frac{\rho}{4}\sum_{t=5}^{T} 2(t-3)\rho^{2(t-3)-1} \\
= & \frac{\rho^2}{4}\sum_{s=4}^{2(T-3)}[s-1]\,s\rho^{s-2} - \frac{\rho}{4}\sum_{s=4}^{2(T-3)} s\rho^{s-1} \\
= & \frac{\rho^2}{4}\frac{\partial^2}{\partial\rho^2}\left\{\rho^4\sum_{t=4}^{2(T-3)}\rho^{s-4}\right\} - \frac{\rho}{4}\frac{\partial}{\partial\rho}\left\{\rho^4\sum_{s=4}^{2(T-3)}\rho^{s-4}\right\} \\
= & \frac{\rho^2}{4}\frac{\partial^2}{\partial\rho^2}\left\{\frac{\rho^4\left(1-\rho^{2(T-3)-3}\right)}{1-\rho}\right\} - \frac{\rho}{4}\frac{\partial}{\partial\rho}\left\{\frac{\rho^4\left(1-\rho^{2(T-3)-3}\right)}{1-\rho}\right\} \\
= & \frac{\rho^2}{4}\left\{\frac{12\rho^2-[2(T-3)+1]\,2(T-3)\rho^{2(T-3)-1}}{1-\rho} + \frac{4\rho^3-[2(T-3)+1]\rho^{2(T-3)}}{(1-\rho)^2}\right. \\
& \left. + \frac{4\rho^3-[2(T-3)+1]\rho^{2(T-3)}}{(1-\rho)^2} + \frac{2\left[\rho^4-\rho^{2(T-3)+1}\right]}{(1-\rho)^3}\right\} \\
& - \frac{\rho}{4}\left\{\frac{4\rho^3-[2(T-3)+1]\rho^{2(T-3)}}{1-\rho} + \frac{\rho^4-\rho^{2(T-3)+1}}{(1-\rho)^2}\right\} \\
= & \frac{\rho^2}{4}\left\{\frac{12\rho^2-[2(T-3)+1]\,2(T-3)\rho^{2(T-3)-1}}{1-\rho} + \frac{8\rho^3-2[2(T-3)+1]\rho^{2(T-3)}}{(1-\rho)^2}\right. \\
& \left. + \frac{2\left[\rho^4-\rho^{2(T-3)+1}\right]}{(1-\rho)^3}\right\} - \frac{\rho}{4}\left\{\frac{4\rho^3-[2(T-3)+1]\rho^{2(T-3)}}{1-\rho} + \frac{\rho^4-\rho^{2(T-3)+1}}{(1-\rho)^2}\right\}
\end{aligned}
$$

Lemma A3 (Phillips and Moon, 1999, Theorem 3): Suppose that $Y_{i,T} = C_i Q_{i,T}$, where the $(m\times 1)$ random vectors $Q_{i,T}$ are $i.i.d.\,(0,\Sigma_T)$ across i for all T and the C_i are $(m\times m)$ nonzero and nonrandom matrices. Assume the following conditions hold.

(i) Let $\sigma_T^2 = \lambda_{\min}(\Sigma_T)$ and $\liminf_T \sigma_T^2 > 0$;

(ii) $\max_{1\le i\le n}\|C_i\|^2/\lambda_{\min}\left(\sum_{i=1}^{N} C_i C_i'\right) = O(1/N)$ as $N\to\infty$;

(iii) $\|Q_{i,T}\|^2$ are uniformly integrable in T;

(iv) $\lim_{N,T}(1/N)\sum_{i=1}^{N} C_i\Sigma_T C_i' = \Omega > 0$.

Then,

$$X_{N,T} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} Y_{i,T} \Longrightarrow N\left(0, \Omega\right), \text{ as } N, T \to \infty.$$

Lemma A4 (Phillips and Moon, 1999, Corollary 1): Suppose that $Y_{i,T} = C_i Q_{i,T}$, where $Q_{i,T}$ are *i.i.d.* across i for all T, and the C_i are $(m \times m)$ nonrandom matrices for all i. Assume that $Q_{i,T}$ are integrable for all T and $Q_{i,T} \Longrightarrow Q_i$ as $T \to \infty$. Assume that $C = \lim_N \left(1/N\right) \sum_{i=1}^{N} C_i$ exists. If $\|Q_{i,T}\|$ is uniformly integrable in T for all i, and if $\sup_i \|C_i\| < \infty$, then

$$\frac{1}{N} \sum_{i=1}^{N} Y_{i,T} \xrightarrow{p} CE\left[Q_i\right]$$

as $N, T \to \infty$.

Lemma A5: Suppose that Assumption 1 holds and define

$$w_{it-1} = \sum_{k=0}^{t-1} \varepsilon_{it-1-k}$$

with $\varepsilon_{i0} = 0$. Then,

(a)

$$\frac{1}{T} \sum_{t=2}^{T} w_{it-1} \varepsilon_{it} \Longrightarrow \frac{\sigma^2}{2} \left[\chi_1^2 - 1\right] = \sigma^2 \int_0^1 W_i\left(r\right) dW_i\left(r\right),$$

as $T \to \infty$.

(b)

$$\frac{1}{T^2} \sum_{t=2}^{T} w_{it-1}^2 \Longrightarrow \sigma^2 \int_0^1 \left[W_i\left(r\right)\right]^2 dr$$

as $T \to \infty$.

Joint weak convergence of (a) and (b) also apply.

Proof:

These results are well-known from the unit root literature, so we omit the proof. □

Lemma A6: Suppose that Assumption 1 holds. Then,

(a)

$$\frac{1}{T} \sum_{t=2}^{T} w_{it-1} \varepsilon_{it} \Longrightarrow N\left(0, \frac{\sigma^4}{2}\right),$$

as $N, T \to \infty$ such that $T/N \to \kappa \in [0, \infty)$.

(b)

$$\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T} w_{it-1}^2 \xrightarrow{p} \frac{\sigma^2}{2}$$

as $N, T \to \infty$ such that $T/N \to \kappa \in [0, \infty)$

Proof:

To show part (a), we verify the conditions of Theorem 3 of Phillips and Moon (1999), given above as Lemma A3. To proceed, first define

$$Q_{i,T} = \frac{1}{T}\sum_{t=2}^{T} w_{it-1}\varepsilon_{it}.$$

and note that by direct calculation

$$\begin{aligned}
\liminf_{T\to\infty} \sigma_T^2 &= \liminf_{T\to\infty} E\left[Q_{i,T}^2\right] \\
&= \lim_{T\to\infty} E\left[Q_{i,T}^2\right] \\
&= \lim_{T\to\infty} \frac{1}{T^2}\sum_{t=2}^{T}\sum_{s=2}^{T}\sum_{k=1}^{t-1}\sum_{\ell=1}^{s-1} E\left[\varepsilon_{ik}\varepsilon_{i\ell}\varepsilon_{it}\varepsilon_{is}\right] \\
&= \lim_{T\to\infty} \frac{1}{T^2}\sum_{t=2}^{T}\sum_{k=1}^{t-1} E\left[\varepsilon_{ik}^2\right] E\left[\varepsilon_{it}^2\right] \\
&= \sigma^4 \lim_{T\to\infty} \frac{T(T-1)}{2T^2} \\
&= \frac{\sigma^4}{2} > 0
\end{aligned}$$

so that condition (i) of Lemma A3 is satisfied. Moreover, in this case, we have

$$C_i = 1 \text{ for all } i$$

so that

$$\max_{1\leq i\leq n} \|C_i\|^2 / \lambda_{\min}\left(\sum_{i=1}^{N} C_i C_i'\right) = 1/N = O\left(1/N\right)$$

as required by condition (ii). Next, observe that by part (a) of Lemma A5, we have that as $T \to \infty$,

$$Q_{i,T} = \frac{1}{T}\sum_{t=2}^{T} w_{it-1}\varepsilon_{it} \Longrightarrow \frac{\sigma^2}{2}\left[\chi_1^2 - 1\right] \equiv Q$$

so that by the continuous mapping theorem

$$Q_{i,T}^2 \Longrightarrow Q^2 \equiv \frac{\sigma^4}{4}\left[\chi_1^2 - 1\right]^2, \text{ as } T \to \infty$$

In addition,

$$\begin{aligned} E\left[Q^2\right] &= \frac{\sigma^4}{4}\left\{E\left[\chi_1^2\right]^2 - 2E\left[\chi_1^2\right] + 1\right\} \\ &= \frac{\sigma^4}{4}\left\{Var\left[\chi_1^2\right] + \left(E\left[\chi_1^2\right]\right)^2 - 2E\left[\chi_1^2\right] + 1\right\} \\ &= \frac{\sigma^4}{4}\left\{2+1-2+1\right\} \\ &= \frac{\sigma^4}{2}, \end{aligned}$$

and note that, as $T \to \infty$,

$$\lim_{T\to\infty} E\left[Q_{i,T}^2\right] = \sigma^4 \lim_{T\to\infty} \frac{T(T-1)}{2T^2} = \frac{\sigma^4}{2} = E\left[Q^2\right].$$

It follows from Theorem 5.4 of Billingsley (1968) that $\left\{Q_{i,T}^2\right\}$ is uniformly integrable, so that condition (iii) of Lemma A3 is satisfied. Finally, note that in this case

$$\lim_{N,T}(1/N)\sum\nolimits_{i=1}^{N} C_i \Sigma_T C_i' = \lim_{N,T}(1/N)\sum\nolimits_{i=1}^{N} E\left[Q_{i,T}^2\right] = \frac{\sigma^4}{2} > 0$$

so that condition (iv) is satisfied as well. It follows then from Lemma A3 that

$$Q_{i,T} = \frac{1}{T}\sum_{t=2}^{T} w_{it-1}\varepsilon_{it} \Longrightarrow N\left(0, \frac{\sigma^4}{2}\right)$$

as $N, T \to \infty$.

To show part (b), we verify the conditions of Corollary 1 of Phillips and Moon (1999), given here as Lemma A4. To proceed, define now

$$Q_{i,T} = \frac{1}{T^2}\sum_{t=2}^{T} w_{it-1}^2.$$

Note that $Q_{i,T}$ is integrable in light of Assumption 1 and by part (b) of Lemma A5, we have that, as $T \to \infty$

$$Q_{i,T} = \frac{1}{T^2}\sum_{t=2}^{T} w_{it-1}^2 \Longrightarrow \sigma^2 \int_0^1 \left[W_i(r)\right]^2 dr = Q_i.$$

Again, we have in this case

$$C_i = 1 \text{ for all } i$$

so that trivially,

$$C = \lim_{N}(1/N)\sum\nolimits_{i=1}^{N} C_i = 1 < \infty.$$

and

$$\sup_i \|C_i\| = 1 < \infty.$$

Moreover, in this case,

$$\begin{aligned} E\left[|Q_{i,T}|\right] &= E\left[Q_{i,T}\right] \\ &= \frac{1}{T^2}\sum_{t=2}^{T} E\left[w_{it-1}^2\right] \\ &= \frac{1}{T^2}\sum_{t=2}^{T}\sum_{k=1}^{t-1}\sum_{\ell=1}^{t-1} E\left[\varepsilon_{ik}\varepsilon_{i\ell}\right] \\ &= \frac{1}{T^2}\sum_{t=2}^{T}\sum_{k=1}^{t-1} E\left[\varepsilon_{ik}^2\right] \\ &= \sigma^2\frac{T(T-1)}{2T^2} \end{aligned}$$

so that, as $T \to \infty$,

$$\lim_{T\to\infty} E\left[|Q_{i,T}|\right] = \sigma^2 \lim_{T\to\infty}\frac{T(T-1)}{2T^2} = \frac{\sigma^2}{2} = \sigma^2\int_0^1 E\left[W_i(r)\right]^2 dr = E\left[Q_i\right] \text{ for all } i.$$

It follows from Theorem 5.4 of Billingsley (1968) that $\{|Q_{i,T}|\}$ is uniformly integrable in T for all i. Hence, all the conditions of Lemma A4 are satisfied, and we deduce from this lemma that

$$\frac{1}{N}\sum_{i=1}^{N} Q_{i,T} = \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=2}^{T} w_{it-1}^2 \xrightarrow{p} CE\left[Q_i\right] = \frac{\sigma^2}{2}$$

as $N, T \to \infty$. □

Lemma A7: Under Assumptions 1-4, the following statements are true.

(a) If $\rho \in (0,1)$,

$$\frac{\mathbb{T}_{NT}}{(NT)^{\frac{1}{2}-\epsilon}} \xrightarrow{p} -\infty$$

for any $\epsilon > 0$, as $N, T \to \infty$ such that $T/N \to \kappa \in [0, \infty)$.

(b) If $\rho = 1$,

$$\mathbb{T}_{NT} \Longrightarrow N(0,1),$$

as $N, T \to \infty$ such that $T/N \to \kappa \in [0, \infty)$.

Proof:

To show (a), note from part (a) of Theorem 2 that in this case

$$\widehat{\rho}_{\text{pols}} = \rho + \frac{\sigma_a^2(1-\rho)\left(1-\rho^2\right)}{\sigma_a^2(1-\rho^2)+\sigma^2} + O_p\left(T^{-1}\right),$$

so that

$$\widehat{\rho}_{\text{pols}} - 1 \xrightarrow{p} -(1-\rho) + \frac{\sigma_a^2 (1-\rho)\left(1-\rho^2\right)}{\sigma_a^2 (1-\rho^2) + \sigma^2} = -\frac{\sigma^2 (1-\rho)}{\sigma_a^2 (1-\rho^2) + \sigma^2} < 0$$

as $N, T \to \infty$ such that $T/N \to \kappa \in [0, \infty)$. Note also that

$$\sum_{i=1}^{N} \sum_{t=2}^{T} \left(y_{it-1} - \overline{y}_{-1,NT}\right)^2 = O_p(NT)$$

where

$$\overline{y}_{-1,NT} = \frac{1}{N(T-1)} \sum_{i=1}^{N} \sum_{t=2}^{T} y_{it-1}$$

It follows that for any $\epsilon > 0$

$$\begin{aligned}
& \frac{\mathbb{T}_{NT}}{(NT)^{\frac{1}{2}-\epsilon}} \\
= \; & \frac{\widehat{\rho}_{\text{pols}} - 1}{(NT)^{\frac{1}{2}-\epsilon} \sqrt{\widehat{\sigma}^2 \left[\sum_{i=1}^{N} \sum_{t=2}^{T} \left(y_{it-1} - \overline{y}_{NT}\right)^2\right]^{-1}}} \\
= \; & \widehat{\sigma}^{-1} \left[\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=2}^{T} \left(y_{it-1} - \overline{y}_{NT}\right)^2\right]^{1/2} \frac{\sqrt{NT}\left(\widehat{\rho}_{\text{pols}} - 1\right)}{(NT)^{\frac{1}{2}-\epsilon}} \\
= \; & \widehat{\sigma}^{-1} \left[\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=2}^{T} \left(y_{it-1} - \overline{y}_{NT}\right)^2\right]^{1/2} (NT)^{\epsilon} \left(\widehat{\rho}_{\text{pols}} - 1\right) \xrightarrow{p} -\infty
\end{aligned}$$

Next, consider part (b), where we assume that $\rho = 1$. In this case, by part (b) Theorem 2, we have

$$\sqrt{N}T\left(\widehat{\rho}_{\text{pols}} - \rho_T\right) \Longrightarrow N(0, 2)$$

as $N, T \to \infty$ such that $T/N \to \kappa \in [0, \infty)$. Moreover, in this case, we have

$$\frac{1}{NT^2} \sum_{i=1}^{N} \sum_{t=2}^{T} \left(y_{it-1} - \overline{y}_{NT}\right)^2 = \frac{1}{NT^2} \sum_{i=1}^{N} \sum_{t=2}^{T} w_{it-1}^2 + O_p\left(\frac{1}{T}\right) \xrightarrow{p} \frac{\sigma^2}{2}$$

It follows that

$$\begin{aligned}
& \mathbb{T}_{NT} \\
= \; & \frac{\widehat{\rho}_{\text{pols}} - 1}{\sqrt{\widehat{\sigma}^2 \left[\sum_{i=1}^{N} \sum_{t=2}^{T} \left(y_{it-1} - \overline{y}_{NT}\right)^2\right]^{-1}}} \\
= \; & \left[\frac{1}{NT^2} \sum_{i=1}^{N} \sum_{t=2}^{T} \left(y_{it-1} - \overline{y}_{NT}\right)^2\right]^{1/2} \frac{\sqrt{N}T\left(\widehat{\rho}_{\text{pols}} - 1\right)}{\widehat{\sigma}} \Longrightarrow \frac{\sigma}{\sqrt{2}} \frac{N(0,2)}{\sigma} \equiv N(0,1)
\end{aligned}$$

as $N, T \to \infty$ such that $T/N \to \kappa \in [0, \infty)$, thus, establishing the required result. □