DYNAMIC PANEL GMM WITH NEAR UNITY

By

Peter C. B. Phillips

December 2014

COWLES FOUNDATION DISCUSSION PAPER NO. 1962

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.econ.yale.edu/

Dynamic Panel GMM with Roots Near Unity*

Peter C. B. Phillips
Yale University, University of Auckland,
Singapore Management University & University of Southampton

August 5, 2014

Abstract

Limit theory is developed for the dynamic panel GMM estimator in the presence of an autoregressive root near unity. In the unit root case, Anderson-Hsiao lagged variable instruments satisfy orthogonality conditions but are well-known to be irrelevant. For a fixed time series sample size (T) GMM is inconsistent and approaches a shifted Cauchy-distributed random variate as the cross section sample size $n \to \infty$. But when $T \to \infty$, either for fixed n or as $n \to \infty$, GMM is $\sqrt{T}$ consistent and its limit distribution is a ratio of random variables that converges to twice a standard Cauchy as $n \to \infty$. In this case, the usual instruments are uncorrelated with the regressor but irrelevance does not prevent consistent estimation. The same Cauchy limit theory holds sequentially and jointly as $(n, T) \to \infty$ with no restriction on the divergence rates of n and T. When the common autoregressive root $\rho = 1 + c/\sqrt{T}$ the panel comprises a collection of mildly integrated time series. In this case, the GMM estimator is $\sqrt{n}$ consistent for fixed T and $\sqrt{nT}$ consistent with limit distribution $N(0, 4)$ when $n, T \to \infty$ sequentially or jointly. These results are robust for common roots of the form $\rho = 1 + c/T^{\gamma}$ for all $\gamma \in (0, 1)$ and joint convergence holds. Limit normality holds but the variance changes when $\gamma = 1$. When $\gamma > 1$ joint convergence fails and sequential limits differ with different rates of convergence. These findings reveal the fragility of conventional Gaussian GMM asymptotics to persistence in dynamic panel regressions.

Keywords: Cauchy limit theory, Dynamic panel, GMM estimation, Instrumental variable, Irrelevant instruments, Panel unit roots, Persistence.

JEL classification: C230, C360

*This paper originated in a Yale take home examination (Phillips, 2013). Some of the results were presented at a Conference in honor of Richard J. Smith's 65th birthday at Cambridge University in May, 2014. The author acknowledges support from the NSF under Grant No. SES 12-58258.

1 Introduction

The use of instrumental variables (IV) in dynamic panel estimation was suggested by Anderson and Hsiao (1981, 1982) and has led to a substantial theoretical and applied literature on the use of IV and generalized method of moment (GMM) estimation techniques in dynamic panels. The unit root case is well-known to present difficulties for IV/GMM methods because lagged variable instruments satisfy the required orthogonality conditions but fail the relevance condition. The problem was discussed in Blundell and Bond (1998) and Moon and Phillips (2004). It is easy to dismiss the unit root case as unidentified by IV/GMM formulations involving lagged level instruments. As a result there are few analyses of GMM asymptotics in this apparently unidentified case. An important exception is Kruiniger (2009) who considered dynamic panel estimation with persistent data when the cross section sample size $n \to \infty$ and the time series sample size (T) is fixed, showing inconsistency of the GMM estimator of the autoregressive parameter.

The existence of other techniques that do deliver consistent estimation in the unit root case has partly diverted attention from the GMM approach, although these alternative methods also present difficulties such as bias and bias discontinuities in the case of level maximum likelihood (Hahn and Kuersteiner, 2003), likelihood function anomalies in the case of first difference maximum likelihood (Han and Phillips, 2013), and sensitivity to departures from stationary errors under X-differencing (Han, Phillips, and Sul, 2014). In view of these difficulties as well as the convenience of standard software implementation, GMM and its many variants are still heavily used in empirical work with dynamic panels. In such applications, conventional GMM Gaussian asymptotic theory is typically assumed to apply when either or both the cross section sample size (n) and time series sample size (T) tend to infinity. When the autoregressive root lies in the vicinity of unity, these Gaussian asymptotics are inevitably fragile because of failing instrument relevance.

The present paper completes existing theory by providing an asymptotic analysis of GMM in the unit root panel AR(1) model using large n, large T, and joint (n, T) asymptotics. For fixed T, we show that GMM is inconsistent and approaches a shifted and scaled Cauchy distributed random limit variate as $n \to \infty$, which corresponds to the finding in Kruiniger (2009). For fixed n, GMM is $\sqrt{T}$ consistent as $T \to \infty$ and has a limit distribution that involves a ratio of random variables which depends on the distribution of the data, so no invariance principle applies. When $T \to \infty$ as $n \to \infty$, GMM is $\sqrt{T}$ consistent and its limit distribution is two times a standard Cauchy. The same limit theory holds both sequentially, irrespective of the order of divergence of (n, T), and jointly as $(n, T) \to \infty$, irrespective of the relative rates of divergence of n and T. Importantly, the usual instruments are uncorrelated with the regressor in this case, but this irrelevance does not prevent consistent estimation at least as $T \to \infty$. In nonstationary data models even orthogonal instruments can be effective in delivering consistent estimation, as was pointed out in early nonstationary time series work (Phillips and Hansen, 1990). Similar effects arise with panel data in the unit root case even though the model is differenced to remove fixed effects prior to regression. In this event, the differenced regressor is itself stationary and so the relevance effect arises from a sample covariance between a stationary and unit root process giving a random limit with zero mean and positive variance, thereby helping to explain the well-known dispersion of the GMM estimator which applies here even in the limit in the unit root case. The Cauchy form of the asymptotics (and uncertainty reflected in the heavy tailed distribution) is reminiscent of (and related to) the limit theory that applies in unidentified simultaneous equations models when estimated by instrumental variables under conditions of apparent identification (Phillips, 1989).

The paper further investigates near unit root cases where the common autoregressive coefficient lies in the vicinity of unity. We focus primarily on cases where $\rho = 1 + c/\sqrt{T}$, consonant with the $\sqrt{T}$ convergence rate of GMM when $\rho = 1$. Results for large n, large T, sequential, and joint asymptotics are provided. The limit theory leads to a correction of the asymptotic variance reported in Anderson and Hsaio (1981). Extensions of these results are given for common roots of the form $\rho = 1 + c/T^{\gamma}$ for all $\gamma \in (0, 1)$, $\gamma = 1$, and $\gamma > 1$.

The remainder of the paper is organized as follows. Succeeding sections give the limit theory for the panel unit root model under fixed n, fixed T, sequential (n, T), sequential (T, n), and joint $(n, T) \to \infty$ asymptotics. Later sections examine the impact of local to unity parameterizations on the asymptotic theory. Extensions to the multiple instrument and differenced instrument cases are considered in the penultimate section. Section 5 concludes with some further discussion. Proofs and derivations are given in the Appendix. Throughout the paper, we use the notation $(n, T)_{\text{seq}} \to \infty$ to signify $T \to \infty$ followed by $n \to \infty$; correspondingly, $(T, n)_{\text{seq}} \to \infty$ signifies $n \to \infty$ followed by $T \to \infty$; $(n, T) \to \infty$ denotes joint asymptotics where there is no restriction on the passage of n and T to infinity; and $T_j = T - j$ for all integer j.

2 Model Preliminaries

In the dynamic panel regression model

$$y_{it} = \alpha_i (1 - \rho) + \rho y_{it-1} + u_{it}, \quad i = 1, ..., n; t = 1, ...T \tag{2.1}$$

the α_i are fixed effects for which $\sigma_\alpha^2 = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^n \alpha_i^2 < \infty$, the errors u_{it} are $iid\left(0, \sigma^2\right)$ with finite fourth moment across all i and over all t, and the initial conditions $y_{i0} = O_p(1)$ for all i and are independent of the u_{it} for all i and t. Heterogeneity over i may be introduced without disturbing some of the results given below provided large n limit theory applies and uniformity conditions continue to hold for joint (n, T) asymptotics. In order to deliver quick results we will maintain the iid assumption for u_{it} in what follows, while pointing out some of the extensions that apply. We define $u_{is} = 0$ for all $s \leq 0$ and we often assume for simplicity that $y_{i0} = 0$, $a.s.$, although calculations are usually shown for the more general case.

We start by studying the simple linear IV/GMM estimator (Anderson and Hsiao, 1981) which uses instruments y_{it-2} in the differenced regression

$$\Delta y_{it} = \rho \Delta y_{it-1} + \Delta u_{it}, \tag{2.2}$$

leading to the estimator $\rho_{gmm} = \sum_{i=1}^n \sum_{t=2}^T \Delta y_{it} y_{it-2} / \sum_{i=1}^n \sum_{t=2}^T \Delta y_{it-1} y_{it-2}$. When the true autoregressive coefficient in (2.1) is $\rho = 1$ we have

$$\rho_{gmm} - 1 = \frac{\sum_{i=1}^n \sum_{t=2}^T \Delta u_{it} y_{it-2}}{\sum_{i=1}^n \sum_{t=2}^T \Delta y_{it-1} y_{it-2}}. \tag{2.3}$$

With $\rho = 1$ we have $\Delta y_{it} = u_{it}$ whose partial sum solution is $y_{it} = \sum_{s=1}^t u_{is} + y_{i0}$ up to the initial condition y_{i0} and since

$$\mathbb{E}\left(u_{it} y_{it-2}\right) = \mathbb{E}\left(\Delta u_{it} y_{it-2}\right) = 0, \tag{2.4}$$

the instrument y_{it-2} satisfies the orthogonality condition in both (2.1) and (2.2). So instrument orthogonality to the regression error in (2.2) holds. However, orthogonality is generally insufficient for identification and consistent estimation, for which relevance of the instrument (to use the terminology of Phillips, 1989) is typically needed. In the present case, we have

$$\mathbb{E}\left(\Delta y_{it-1}y_{it-2}\right)=\mathbb{E}\left(u_{it-1}y_{it-2}\right)=0,\ \text{ for all } t \text{ and all } i \tag{2.5}$$

so the instrument y_{it-2} is actually orthogonal to the regressor Δy_{it-1} in (2.2) and relevance fails. In this event, the moment conditions (2.4) do not identify the unit root (Kruiniger, 2009)). As is well-known, therefore, the GMM estimator (2.3) is expected to perform poorly in finite samples and to be inconsistent in the limit, as the instrument y_{it-2} is irrelevant for the regressor Δy_{it-1} in (2.2). Similar properties of orthogonality and irrelevance hold for all instrumental variables that take the form of lagged variables $\{y_{is}: s=1,2,...t-2\}$.

3 Asymptotics when $\rho = 1$

3.1 Large n Asymptotics

Start with the case where T is fixed and $n\to\infty$. Consider $\sqrt{n}$ standardized forms of the numerator and denominator of (2.3), viz.,

$$N_{nT}=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{t=2}^{T}\Delta u_{it}y_{it-2},\quad D_{nT}=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{t=2}^{T}\Delta y_{it-1}y_{it-2}.$$

Observe that $\Delta\left(u_{it}y_{it-2}\right)=\Delta u_{it}y_{it-2}+u_{it-1}\Delta y_{it-2}=\Delta u_{it}y_{it-2}+u_{it-1}u_{it-2}$ under $\rho=1$, so by partial summation

$$\sum_{t=3}^{T}\Delta u_{it}y_{it-2}=\left(u_{iT}y_{iT-2}-u_{i2}y_{i0}\right)-\sum_{t=3}^{T}u_{it-1}u_{it-2}. \tag{3.1}$$

Adding $\Delta u_{i2}y_{i0}=u_{i2}y_{i0}-u_{i1}y_{i0}$ to each side gives

$$\sum_{t=2}^{T}\Delta u_{it}y_{it-2}=\left(u_{iT}y_{iT-2}-u_{i1}y_{i0}\right)-\sum_{t=3}^{T}u_{it-1}u_{it-2}. \tag{3.2}$$

Then, $N_{nT}=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(u_{iT}y_{iT-2}-u_{i1}y_{i0}\right)-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{t=3}^{T}u_{it-1}u_{it-2}$, and $D_{nT}=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{t=2}^{T}u_{it-1}y_{it-2}$, for which we have the following limit behavior as $n\to\infty$ when T is fixed.

Theorem 1 *For fixed T as $n\to\infty$*

(i) $(N_{nT}, D_{nT}) \underset{n\to\infty}{\Rightarrow} N(0, V_T), V_T = \sigma^4 T_2 \begin{pmatrix} 2 & -1 \\ -1 & T_1/2 \end{pmatrix}$, *where* $T_j = T - j$;

(ii) $\rho_{gmm} - 1 \underset{n\to\infty}{\Rightarrow} -\frac{2}{T_1} + 2\frac{\left(1-\frac{1}{T_1}\right)^{1/2}}{T_1^{1/2}}\mathbb{C}$, *where* $\mathbb{C}$ *is a standard Cauchy variate.*

Thus, when T is fixed and $n \to \infty$, ρ_{gmm} is inconsistent and converges weakly to a Cauchy distribution centred on $1 - \frac{2}{T_1}$, a result that was earlier obtained in Kruiniger (2009, theorem 1(i)) for the random coefficient case with $T = 3$. The heavy tailed limit distribution arises because the denominator D_{nT} has a random limit and its Gaussian distribution is symmetrically distributed with a positive density at zero, which ensures that no integer moments exist. The random limiting denominator reflects the presence of random information in the GMM signal in the limit.

Next consider sequential asymptotics in which $n \to \infty$ is followed by $T \to \infty$. From Theorem 1(ii) we deduce directly that

$$\sqrt{T_1}\left(\rho_{gmm} - 1 + \frac{2}{T_1}\right) \underset{n\to\infty}{\Rightarrow} 2\left(1 - \frac{1}{T_1}\right)^{1/2}\mathbb{C} \underset{T\to\infty}{\Rightarrow} 2\mathbb{C}, \tag{3.3}$$

giving

$$\sqrt{T}\left(\rho_{gmm} - 1\right) \underset{(T,n)_{\text{seq}}\to\infty}{\Rightarrow} 2\mathbb{C}. \tag{3.4}$$

Evidently, the GMM estimator ρ_{gmm} is consistent as $T \to \infty$, even though the instrumental variable y_{it-2} is irrelevant in the panel regression for all t. The rate of convergence is $\sqrt{T}$, which is slower than the usual rate (T) for (unit root) nonstationary data in time series regression. The explanation for the large T consistency of ρ_{gmm} is that, although the relevance condition fails for all t and $\mathbb{E}(\Delta y_{it-1} y_{it-2}) = 0$, the sample covariance (moment condition) does not have a zero limit as $T \to \infty$. Instead,

$$\frac{1}{T}\sum_{t=2}^{T} \Delta y_{it-1} y_{it-2} = \frac{1}{T}\sum_{t=2}^{T} u_{it-1} y_{it-2} \Rightarrow \int_0^1 B_i dB_i \neq 0 \quad a.s. \tag{3.5}$$

where B_i is Brownian motion with variance σ^2 for all i (Phillips, 1987a). On the other hand, as $n \to \infty$ with T fixed, the sample covariance

$$\frac{1}{n}\sum_{i=1}^{n}\left(\sum_{t=2}^{T} \Delta y_{it-1} y_{it-2}\right) \to_p 0, \tag{3.6}$$

in view of (2.5). The nonzero limit (3.5) ensures some relevance as $T \to \infty$ in the nonstationary instrument y_{it-2} in spite of the fact that $\mathbb{E}(\Delta y_{it-1} y_{it-2}) = 0$. But since the limit (3.5) is random, there is inevitably high variability in the GMM estimate. The variability is sustained in the Cauchy distribution limit for which there are heavy tails and no finite sample integer moments, just as in the fixed T case.

The convergence rate is $\sqrt{T}$ because the IV regression signal is $\sum_{t=2}^{T} \Delta y_{it-1} y_{it-2} = O_p(T)$, the order of a sample covariance between integrated and stationary processes. This $O\left(\sqrt{T}\right)$ rate is slower than the usual $O(T)$ convergence rate for unit root and IV unit root regressions where, with both instrument and regressor integrated, the signal is of order $O\left(T^2\right)$ not $O(T)$.

3.2 Large T Asymptotics

Start with the case where n is fixed. As the following result shows, the limit theory for $T \to \infty$ does not obey an invariance principle and is dependent on the distribution of the data. But when $T \to \infty$ is followed by $n \to \infty$, an invariance principle holds and we again have a Cauchy distribution limit.

Theorem 2

(i) *As $T \to \infty$ with n fixed*

$$\sqrt{T}\left(\rho_{gmm} - 1\right) \underset{T \to \infty}{\Rightarrow} \frac{\sum_{i=1}^{n} \left\{u_{i\infty} B_i(1) - G_i\right\}}{\sum_{i=1}^{n} \frac{1}{2}\left\{B_i(1)^2 - \sigma^2\right\}}. \tag{3.7}$$

where $\{B_i(r)\}_{i=1}^{n}$ are a family of iid Brownian motions with variance σ^2 that are independent of the family of iid Gaussian variates $\{G_i\}_{i=1}^{n}$ each with zero mean and variance σ^4 and all independent of the variate $u_{i\infty}$ which is an identically distributed copy of u_{it}.

(ii) *When $T \to \infty$ followed by $n \to \infty$*

$$\sqrt{T}\left(\rho_{gmm} - 1\right) \underset{(n,T)_{\text{seq}} \to \infty}{\Rightarrow} 2\mathbb{C}. \tag{3.8}$$

Hence, ρ_{gmm} is consistent as $T \to \infty$ when the cross section sample size n is fixed. The explanation is the same as that given above concerning the relevance of the nonstationary instrument y_{it-2}. Observe that (3.7) is a ratio of two random variables each of which is centred on the origin and the denominator has positive probability density at the origin, which ensures that the ratio (3.7) has no finite sample integer moments..

Importantly, as shown in the proof of the theorem, the limit (3.7) involves only a partial application of an invariance principle. The component $u_{i\infty}$ in the numerator of (3.7) is not the outcome of an invariance principle but is instead distribution dependent since $u_{i\infty} =_d u_{it}$ for all t in view of the identical distribution assumption concerning u_{it}.

Evidently, when $T \to \infty$ is followed by $n \to \infty$ the sequential limit distribution is identical to the limit distribution with the reverse order of sequential limits (i.e., $n \to \infty$ followed by $T \to \infty$) as given in (3.4).

3.3 Joint Limit Theory as $(n, T) \to \infty$

The equivalence of the sequential limit results (3.4) and (3.8) suggests that the limit theory is robust to the path of divergence of the respective cross section and time series sample sizes or the relative rates at which $(n, T) \to \infty$. The limit theory under joint sample size expansion $(n, T) \to \infty$ is proved in the following result using the criteria for joint convergence given in Phillips and Moon (1999).

Theorem 3 $\sqrt{T}\left(\rho_{gmm} - 1\right) \underset{(n,T)\to\infty}{\Rightarrow} 2\mathbb{C}$ *and joint convergence applies as* $(n, T) \to \infty$ *irrespective of the order and rates of expansion of the respective sample sizes.*

The heavy tailedness property of the GMM estimator ρ_{gmm} manifested in the joint limit theory to a Cauchy variate applies irrespective of the manner in which the cross section and time series sample sizes diverge to infinity. The rate of convergence is $\sqrt{T}$, as in both forms of sequential asymptotics, and is slower than the usual $O(T)$ rate associated with unit root time series because of the diminished signal from the 'apparently irrelevant' instrument y_{it-2} used in the GMM regression.

4 Local Unit Root Asymptotics when $\rho = 1 + c/T^{\gamma}$, $c < 0$

There are several local unit root (LUR) cases that may be considered. For large n fixed T asymptotics it is possible to consider deviations from unity of the form $\rho = 1 + c/n^{\gamma}$ for $\gamma \in (0, 1)$ as in Kruiniger (2009). This formulation is largely for mathematical convenience in analyzing the effects of local departures from unity in large n asymptotics. Importantly, the autoregressive parameter ρ measures time series dependence in the panel data y_{it}. It is therefore more difficult to justify modeling time series dependence through a parameter whose value $\rho = \rho_n = 1 + c/n^{\gamma}$ depends on the number of cross section observations. In particular, the dependence $\rho_n = 1 + c/n^{\gamma}$ implies that the AR coefficient of an individual time series like y_{1t} in the panel will approach unity simply by increasing the number of panel observations. Given cross section independence in the panel, it seems hard to justify such dependence of ρ_n on n other than for the mathematical convenience of more closely studying limit behavior in the vicinity of unity. One possible justification is that the *commonality* of the AR parameter ρ across individual time series in the panel y_{it} provides a linkage across the panel that rationalizes formulations such as $\rho_n = 1 + c/n^{\gamma}$. Then, raising the number of cross section observations n enables us to model phenomena with common AR time dependence that is increasingly close to unity, even in spite of the cross section independence in the panel. In this case, in view of the commonality of ρ across section, more cross section information may reasonably be expected to enable us to model phenomena with AR time dependence closer to persistence.

By contrast, time series sample size dependences of ρ on T, such as $\rho = \rho_T = 1 + c/T^{\gamma}$, are already commonplace in the time series literature. The classifications used in that literature for measuring departures from unity apply in the same way for panels. Thus, when $\gamma = 1$ the departures are deemed to be local to unity (LUR) concordant with a Pitman drift when the estimation convergence rate is $O(T)$, as is typical in time series regression. When $\gamma \in (0, 1)$, the departures are said to constitute a mild unit root (MUR) and lead to mildly integrated time series in the sense of Phillips and Magdalinos (2007). In both cases, as the time series sample size T increases, the triangular array model formulation allows us to model time series phenomena with AR time dependence that approaches persistence ($\rho = 1$) and differentiates the effects of such parameterizations on the limit theory, thereby bridging part of the large T limit theory gap between fixed stationary and unit root cases.

The justification for using such LUR and MUR formulations of ρ as $T \to \infty$ is now well established in the time series literature. Accordingly, in view of the $\sqrt{T}$ convergence rate of the GMM estimator when $\rho = 1$, this section concentrates largely on MUR asymptotic theory for localizing sequences that are of the form $\rho = 1 + c/\sqrt{T}$. As earlier, we will consider large n and large T asymptotics, sequential limits, and joint convergence. We also look at cases where $\rho = 1 + c/T^{\gamma}$ and develop large n, large T asymptotics that cover the implied wider and narrower vicinities of unity that occur for more general parameterizations with $\gamma > 0$.

4.1 Large n and Sequential $(T,n)_{\text{seq}} \to \infty$ Asymptotics

It is natural to start with the case where $\rho = 1 + \frac{c}{\sqrt{T}}$ with fixed $c < 0$ and fixed T as $n \to \infty$. This localization seems appropriate given that $\sqrt{T}$ asymptotics apply when $\rho = 1$, but more general cases may be considered and these are discussed below. Fixing the parameters (c, T) implies a fixed $|\rho| < 1$ and Gaussian asymptotics apply. Anderson and Hsaio (1981, AH) gave results for the fixed stationary case in a model with random effects as $n, T \to \infty$, but their expression for the asymptotic variance is incorrect.[1] In the fixed effects case, which is closely related, the limit theory is as follows.

Theorem 4 *In model (2.1) with $\rho = 1 + \frac{c}{\sqrt{T}}$ for fixed $c < 0$, we have*

(i) $\sqrt{n}\left(\rho_{gmm} - \rho\right) \underset{n\to\infty}{\Rightarrow} N\left(0, \omega_T^2\right)$,

(ii) $\sqrt{nT}\left(\rho_{gmm} - \rho\right) \underset{(T,n)_{\text{seq}}\to\infty}{\Rightarrow} N\left(0, 4\right)$,

where $\omega_T^2 = \frac{\omega_{NT}}{\omega_{DT}} = \frac{2(1+\rho)}{T_1} + O\left(\frac{1}{T_1^2}\right)$. Explicit expressions for $(\omega_{NT}, \omega_{DT})$ are given in (6.21) and (6.22) in the Appendix. Parts (i) and (ii) continue to hold when $\rho = 1 + \frac{c}{T^\gamma}$ with the same convergence rates $\sqrt{n}$ and $\sqrt{nT}$ *and the same limit variances for all $\gamma \in (0,1)$.*
(iii) *When $\gamma = 1$, the Gaussian limit theory (i) still applies but the limit variance ω_T^2 has the alternate form*

$$\omega_T^2 \sim -\frac{8c}{T_1}\frac{\left(1 - 2c - e^{2c}\right)}{\left(1 + 2c - e^{2c}\right)^2}\left\{1 + o\left(1\right)\right\}, \quad \text{as } T \to \infty \tag{4.1}$$

and

$$\sqrt{nT}\left(\rho_{gmm} - \rho\right) \underset{(T,n)_{\text{seq}}\to\infty}{\Rightarrow} N\left(0, (-8c)\frac{\left(1 - 2c - e^{2c}\right)}{\left(1 + 2c - e^{2c}\right)^2}\right). \tag{4.2}$$

(iv) *When $\gamma > 1$, $\sqrt{nT^{3-2\gamma}}\left(\rho_{gmm} - \rho\right) \underset{(T,n)_{\text{seq}}\to\infty}{\Rightarrow} N\left(0, \frac{8}{c^2}\right)$.*

In the stationary case with fixed $c < 0$, fixed T, and the stationary initial condition $y_{i0} = \alpha_i + \sum_{j=0}^{\infty}\rho^j u_{i,-j}$, we have $y_{it} = \alpha_i + \sum_{j=0}^{\infty}\rho^j u_{t-j}$ and the asymptotic variance when

[1] The Anderson-Hsaio (1981) formula given by equation (8.4) in their paper assumes, incorrectly, that $\Delta u_{it} y_{it-2}$ is a martingale difference, so the formula omits cross product terms in the expression. The correct limiting variance of $\sqrt{nT}\left(\rho_{gmm} - \rho\right)$ is given by $2\left(1+\rho\right)$, as shown in theorem 4 and (4.3). This result applies also in the stationary random effects case as discussed below.

$n \to \infty$ then has the simpler explicit form

$$\omega_T^2 = \frac{2(1+\rho)}{T_1} + 2\frac{(1+\rho)^2 \left\{ \frac{\sigma_\alpha^2/\sigma^2}{(1-\rho)^2} + \frac{\rho}{1-\rho^2} \right\}}{T_1^2} \sim \frac{2(1+\rho)}{T_1} \quad \text{for large } T. \tag{4.3}$$

The same limit theory applies in the stationary, random effects case where $\alpha_i \sim_{iid} \left(0, \sigma_\alpha^2\right)$, which was studied in AH (1981). Expression (4.3) corrects the formula given in AH (equation 8.4) for the limiting variance in the random effects model. See Phillips and Han (2014) for further details.[2]

Different localization rates may be studied in the same way. Importantly, whatever rate $\gamma \in (0,1)$ is used for $\rho = 1 + \frac{c}{T^\gamma}$ to approach unity, the limit variance ω_T^2 continues to apply for all fixed T and fixed c. Correspondingly, since $\omega_T^2 \sim \frac{2(1+\rho)}{T_1}$ for large T, we still get $\sqrt{T}$ convergence and a normal limit theory for these localization coefficients, irrespective of how close to unity $1 + \frac{c}{T^\gamma}$ is. Sequential asymptotics $\sqrt{nT}\left(\rho_{gmm} - \rho\right) \underset{(T,n)_{\text{seq}} \to \infty}{\Rightarrow} N(0,4)$ then hold whenever $n \to \infty$ followed by $T \to \infty$. So, theorem 4 continues to apply for $\rho = 1 + \frac{c}{T^\gamma}$ and all $\gamma \in (0,1)$.

Moreover, as indicated in part (iii) of the theorem, when $\gamma = 1$, sequential Gaussian asymptotics hold but the variance of the limiting distribution changes to $(-8c)\frac{\left(1-2c-e^{2c}\right)}{\left(1+2c-e^{2c}\right)^2}$. Observe that

$$(-8c)\frac{\left(1-2c-e^{2c}\right)}{\left(1+2c-e^{2c}\right)^2} \sim \frac{8}{c^2} \to \infty \quad \text{as } c \to 0 \tag{4.4}$$

so the Gaussian limit theory changes when closer approaches to the unit root occur. Moreover, as shown in the proof of Theorem 4(iv), when $\rho = 1 + \frac{c}{T^\gamma}$ with $c < 0$ fixed and $\gamma > 1$ the limit distribution is given by

$$\sqrt{nT^{3-2\gamma}}\left(\rho_{gmm} - \rho\right) \underset{(T,n)_{\text{seq}} \to \infty}{\Rightarrow} N\left(0, \frac{8}{c^2}\right), \tag{4.5}$$

where the variance corresponds with (4.4) and the rate of convergence depends on γ, changes from $\sqrt{nT}$ to $\sqrt{nT^{3-2\gamma}}$, and reduces as γ increases. Thus, when $\gamma \to \frac{3}{2}$ the rate of convergence approaches $\sqrt{n}$ and when $\gamma > \frac{3}{2}$ there is divergence because the limit variance when $n \to \infty$ is $\omega_T^2 = \frac{8T_2}{c^2 T_1^{2(2-\gamma)}}\{1 + o(1)\}$ and the variance diverges as $T \to \infty$. So sequential $(T,n)_{\text{seq}} \to \infty$ asymptotics fail and the distribution diverges as $T \to \infty$. In

[2] The error in AH (1981, equation 8.4) was noted independently by Yinja (Jeff) Qiu in his take home examination solution (2014) to Phillips (2013).

that case, the convergence rate is effectively slower than $\sqrt{n}$ and the limit theory is not captured by sequential asymptotics where $(T,n)_{\text{seq}} \to \infty$. Instead, as shown below, unit root $\sqrt{T}$ asymptotics apply when $(n,T)_{\text{seq}} \to \infty$ because $\rho = 1 + \frac{c}{T^\gamma}$ is in close proximity to $\rho = 1$ when $\gamma > 1$ and $T \to \infty$ first.

4.2 Large T and Sequential $(n,T)_{\text{seq}} \to \infty$ Asymptotics

We now consider limits in which $T \to \infty$ and $\rho = 1 + \frac{c}{\sqrt{T}}$ differs moderately from unity. In a time series framework, this formulation is a special case of moderate integration in the sense of Phillips and Magdalinos (2007). Again, more general cases where $\rho = 1 + \frac{c}{T^\gamma}$ are considered below. The panel asymptotics are given in the following result.

Theorem 5 *In model (2.1) with $\rho = 1 + \frac{c}{\sqrt{T}}$ for fixed $c < 0$, we have*

(i) $\sqrt{T}\left(\rho_{gmm} - \rho\right) \underset{T\to\infty}{\Rightarrow} \frac{2}{n}\sum_{i=1}^{n} \zeta_i$, where $\zeta_i \sim_{iid} N(0,1)$,

(ii) $\sqrt{nT}\left(\rho_{gmm} - \rho\right) \underset{(n,T)_{\text{seq}}\to\infty}{\Rightarrow} N(0,4)$.

It follows that ρ_{gmm} is $\sqrt{T}$ consistent and $\sqrt{T}\left(\rho_{gmm} - \rho\right)$ is asymptotically Gaussian $N(0,4/n)$ when $T \to \infty$. In sequential limits as $(n,T)_{\text{seq}} \to \infty$, the limit distribution is Gaussian $N(0,4)$ after rescaling, just as in Theorem 4 above when $(T,n)_{\text{seq}} \to \infty$. Joint convergence to $N(0,4)$ then follows in the same manner as Theorem 3 and is given in the following result.

Theorem 6 *When $\rho = 1 + \frac{c}{\sqrt{T}}$ with fixed $c < 0$, we have $\sqrt{T}\left(\rho_{gmm} - \rho\right) \underset{(n,T)\to\infty}{\Rightarrow} N(0,4)$ and joint convergence applies as $(n,T) \to \infty$ irrespective of the order and rates of expansion of the respective sample sizes.*

Next consider the case where $\rho = 1 + \frac{c}{T^\gamma}$ with fixed $c < 0$.

Theorem 7 *Let $\rho = 1 + \frac{c}{T^\gamma}$ with fixed $c < 0$ and $\gamma > 0$.*

(i) *When* $\gamma \in (0,1)$, $\sqrt{T}\left(\rho_{gmm} - \rho\right) \underset{T\to\infty}{\Rightarrow} \frac{2}{n}\sum_{i=1}^{n} \zeta_i$, where $\zeta_i \sim_{iid} N(0,1)$, and

$$\sqrt{nT}\left(\rho_{gmm} - \rho\right) \underset{(n,T)_{\text{seq}}\to\infty}{\Rightarrow} N(0,4). \tag{4.6}$$

(ii) *When* $\gamma = 1$, $\sqrt{T}\left(\rho_{gmm} - \rho\right) \underset{T\to\infty}{\Rightarrow} \frac{\sum_{i=1}^{n}\left\{\left(\sigma^{-1}u_{i\infty}\right)J_{ci}(1)-\zeta_i\right\}}{\sum_{i=1}^{n}\left\{c\int_0^1 J_{ci}(r)^2dr+\int_0^1 J_{ci}(r)dW_i\right\}}$, *where* $J_{ci}(r) = \int_0^r e^{c(r-s)}dW_i(s)$, W_i *are standard Brownian motions, and the* $\zeta_i \sim_{iid} N(0,1)$ *and independent of* W_i *for all* i. *Further when* $\gamma = 1$ *and* $T \to \infty$ *followed by* $n \to \infty$, *we have*

$$\sqrt{nT}\left(\rho_{gmm} - \rho\right) \underset{(n,T)_{\text{seq}}\to\infty}{\Rightarrow} N\left(0, -8c\frac{1-2c-e^{2c}}{\left(e^{2c}-1-2c\right)^2}\right). \tag{4.7}$$

(iii) *When* $\gamma > 1$, $\sqrt{T}\left(\rho_{gmm} - \rho\right) \underset{T\to\infty}{\Rightarrow} \frac{\sum_{i=1}^{n}\zeta_i}{\sum_{i=1}^{n}\int_0^1 W_i dW_i}$ *and then*

$$\sqrt{T}\left(\rho_{gmm} - \rho\right) \underset{(n,T)_{\text{seq}}\to\infty}{\Rightarrow} 2\mathbb{C} \tag{4.8}$$

The $N(0,4)$ sequential limit theory given in (4.6) mirrors Theorem 5(ii), showing that this limit result is robust for all moderately integrated panels with mild integration parameter $\gamma \in (0,1)$. Joint limit theory applies in this case, precisely as in the proof of Theorem 6, so the details are omitted here.

When $\gamma = 1$, the large T limit theory under (ii) involves the standardized diffusion processes J_{ci}, as is usual in local to unity cases. The corresponding sequential $(n,T)_{\text{seq}}$ limit theory in (4.7) retains the $\sqrt{nT}$ convergence rate and has a limit variance that depends on the localizing coefficient c, again as may be expected in the LUR case. Moreover, this limit theory is the same as the $(T,n)_{\text{seq}} \to \infty$ sequential asymptotics given in (4.2) of Theorem 4 and both $\sqrt{nT}$ convergence and limiting normality continue to hold. Again, the limit theory is independent of the direction of the asymptotics and joint convergence holds in the same way as Theorem 6. .

When $\gamma > 1$, the sequential $(n,T)_{\text{seq}}$ limit theory (4.8) corresponds exactly to the panel unit root limit Cauchy distribution since the panel autoregressive root $\rho = 1 + \frac{c}{T^\gamma}$ is closer to unity than the local to unity coefficient $\rho = 1 + \frac{c}{T}$ and $T \to \infty$ first. It is this close proximity of $\rho = 1+\frac{c}{T^\gamma}$ to unity as $T \to \infty$ that ensures that the panel unit root limit theory obtains when $\gamma > 1$. Importantly, joint convergence no longer holds in this case. Instead, directional asymptotics occur and the limit distribution depends on the nature of the sample size expansion. For when $(T,n)_{\text{seq}} \to \infty$ we have $\sqrt{nT^{3-2\gamma}}\left(\rho_{gmm} - \rho\right) \underset{(T,n)_{\text{seq}}\to\infty}{\Rightarrow} N\left(0, \frac{8}{c^2}\right)$ as obtained in (4.5), whereas when $(n,T)_{\text{seq}} \to \infty$ we have $\sqrt{T}\left(\rho_{gmm} - \rho\right) \underset{(n,T)_{\text{seq}}\to\infty}{\Rightarrow} 2\mathbb{C}$ as given in (4.8). In effect, the non-Gaussian Cauchy limit theory cannot be captured in $(T,n)_{\text{seq}}$ directional sequential asymptotics where the limit theory is Gaussian because

$|\rho| < 1$ as $n \to \infty$.

5 Further Discussion

Anderson and Hsiao (1981) also suggested using the lagged differences Δy_{it-2} (rather than the lagged levels y_{it-2}) as an instrumental variable. This estimator has the form

$$\rho_{gmm2} = \frac{\sum_{i=1}^{n}\sum_{t=2}^{T}\Delta y_{it}\Delta y_{it-2}}{\sum_{i=1}^{n}\sum_{t=2}^{T}\Delta y_{it-1}\Delta y_{it-2}} = \frac{\sum_{i=1}^{n}\sum_{t=2}^{T}u_{it}u_{it-2}}{\sum_{i=1}^{n}\sum_{t=2}^{T}u_{it-1}u_{it-2}}, \quad \text{when } \rho = 1.$$

Calculations similar to those given in Section 3 now lead to

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\sum_{t=2}^{T}u_{it}u_{it-2}, \sum_{t=2}^{T}u_{it-1}u_{it-2}\right\} \underset{n\to\infty}{\Rightarrow} N\left(0, \sigma^4 T_1\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}\right),$$

which is invariant to T after rescaling by $\sqrt{T_1}$. It follows that $\rho_{gmm2} \underset{n\to\infty}{\Rightarrow} \mathbb{C}$, showing that ρ_{gmm2} is inconsistent, miscentred around the origin, with a random variable limit that has heavy tails like ρ_{gmm} and is invariant to the time series sample size T. In consequence, sequential asymptotics give the same limit, viz., $\rho_{gmm2} \underset{(T,n)_{\text{seq}}\to\infty}{\Rightarrow} \mathbb{C}$. Similarly,

$$\frac{1}{\sqrt{T}}\sum_{t=2}^{T}\left\{\sum_{i=1}^{n}u_{it}u_{it-2}, \sum_{i=1}^{n}u_{it-1}u_{it-2}\right\} \underset{T\to\infty}{\Rightarrow} N\left(0, \sigma^4 n\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}\right)$$

which is invariant to n after scaling by $\sqrt{n}$. Then $\rho_{gmm2} \underset{T\to\infty}{\Rightarrow} \mathbb{C}$, leading directly to the sequential asymptotics $\rho_{gmm2} \underset{(n,T)_{\text{seq}}\to\infty}{\Rightarrow} \mathbb{C}$. Similar arguments to those given earlier show that this limit theory applies jointly as $(n, T) \to \infty$ irrespective of the rates of divergence of the sample sizes. Use of lagged differences Δy_{it-2} as instruments therefore leads to an inconsistent estimator of ρ in the unit root case for fixed T, fixed n, and joint asymptotics. In this case, both the regressors Δy_{it-1} and the instruments Δy_{it-2} are stationary with covariance $\mathbb{E}(\Delta y_{it-1}\Delta y_{it-2}) = 0$. So the sample signal $n^{-1}\sum_{i=1}^{n}T^{-1}\sum_{t=2}^{T}\Delta y_{it-1}\Delta y_{it-2}$ has zero expectation and zero limit in probability, thereby providing no leverage for consistent estimation. Mildly integrated cases with $\rho = 1 + c/T^{\gamma}$ may also be examined using these methods, as may GMM estimates with more instruments, but they are not considered in the present work and will be reported elsewhere.

6 Appendix

Proof of Theorem 1. Part (i) follows by the Lindeberg Lévy CLT

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\begin{pmatrix}\sum_{t=2}^{T}\Delta u_{it}y_{it-2}\\ \sum_{t=2}^{T}\Delta y_{it-1}y_{it-2}\end{pmatrix}\Rightarrow N\left(0,V_T\right), \tag{6.1}$$

with

$$V_T=\begin{pmatrix}\mathbb{E}\left(\sum_{t=2}^{T}\Delta u_{it}y_{it-2}\right)^2 & \mathbb{E}\left(\sum_{t=2}^{T}\Delta u_{it}y_{it-2}\right)\left(\sum_{t=2}^{T}\Delta y_{it-1}y_{it-2}\right)\\ \mathbb{E}\left(\sum_{t=2}^{T}\Delta u_{it}y_{it-2}\right)\left(\sum_{t=2}^{T}\Delta y_{it-1}y_{it-2}\right) & \mathbb{E}\left(\sum_{t=2}^{T}\Delta y_{it-1}y_{it-2}\right)^2\end{pmatrix}. \tag{6.2}$$

To evaluate, note by partial summation as indicated in (3.1) and (3.2), we have

$$\sum_{t=2}^{T}\Delta u_{it}y_{it-2}=u_{iT}y_{iT-2}-u_{i1}y_{i0}-\sum_{t=3}^{T}u_{it-1}\Delta y_{it-2} \tag{6.3}$$

To compute V_T, note that

$$\begin{aligned}
&\mathbb{E}\left(\sum_{t=2}^{T}\Delta u_{it}y_{it-2}\right)^2=\mathbb{E}\left\{\left(u_{iT}y_{iT-2}-u_{i1}y_{i0}\right)-\sum_{t=3}^{T}u_{it-1}u_{it-2}\right\}^2\\
=&\ \mathbb{E}\left(u_{iT}y_{iT-2}-u_{i1}y_{i0}\right)^2-2\mathbb{E}\left\{\left(u_{iT}y_{iT-2}-u_{i1}y_{i0}\right)\left(\sum_{t=3}^{T}u_{it-1}u_{it-2}\right)\right\}+\mathbb{E}\left(\sum_{t=3}^{T}u_{it-1}u_{it-2}\right)^2\\
=&\ \mathbb{E}u_{iT}^2y_{iT-2}^2+\mathbb{E}u_{i1}^2y_{i0}^2+\sum_{t=3}^{T}\mathbb{E}\left(u_{it-1}^2u_{it-2}^2\right)\\
=&\ \sigma^4T_2+2\sigma^2\mathbb{E}y_{i0}^2+\sigma^4T_2=2\sigma^4T_2,
\end{aligned}$$

the final line following if the initial condition $y_{i0}=0$, which will be assumed in the calculations below. The large n asymptotic results will continue to hold for $y_{i0}=O_p(1)$ even for finite T with some obvious minor adjustments to the variance matrix expressions involving quantities of $O(1)$ in T. Next

$$\begin{aligned}\mathbb{E}\left(\sum_{t=2}^{T}\Delta y_{it-1}y_{it-2}\right)^2 &= \mathbb{E}\left(\sum_{t=2}^{T}u_{it-1}y_{it-2}\right)^2 = \sigma^2\sum_{t=2}^{T}\mathbb{E}y_{it-2}^2 = \sigma^4\sum_{t=2}^{T}(t-2)\\ &= \sigma^4 T_2 T_1/2,\end{aligned}$$

and, with $y_{i0}=0$ (or up to $O(1)$ in T if $y_{i0}\neq 0$)

$$\begin{aligned}&\mathbb{E}\left(\sum_{t=2}^{T}\Delta u_{it}y_{it-2}\right)\left(\sum_{t=2}^{T}u_{it-1}y_{it-2}\right) = \mathbb{E}\left\{\left((u_{iT}y_{iT-2}-u_{i1}y_{i0})-\sum_{t=3}^{T}u_{it-1}u_{it-2}\right)\left(\sum_{t=2}^{T}u_{it-1}y_{it-2}\right)\right\}\\ &= -\mathbb{E}\left(\sum_{t=3}^{T}u_{it-1}u_{it-2}\right)\left(\sum_{t=2}^{T}u_{it-1}y_{it-2}\right)-\mathbb{E}\left(u_{i1}y_{i0}\right)^2\\ &= -\mathbb{E}\left\{\sum_{t=3}^{T}\left(u_{it-1}u_{it-2}\right)\left(u_{it-1}y_{it-2}\right)+\sum_{s,t=3;s\neq t}^{T}\left(u_{it-1}u_{it-2}\right)\left(u_{is-1}y_{is-2}\right)\right\}\\ &= -\sum_{t=3}^{T}\mathbb{E}u_{it-1}^2u_{it-2}^2 = -\sigma^4\sum_{t=3}^{T}1 = -\sigma^4 T_2.\end{aligned}$$

Then

$$V_T = \begin{pmatrix} 2\sigma^4 T_2 & -\sigma^4 T_2\\ -\sigma^4 T_2 & \sigma^4 T_2 T_1/2\end{pmatrix} = \sigma^4 T_2\begin{pmatrix} 2 & -1\\ -1 & T_1/2\end{pmatrix},$$

as stated.

For Part (ii), simply write $\rho_{gmm}-1=\frac{N_{nT}}{D_{nT}}$, and note from (i) that $(N_{nT},D_{nT}) \underset{n\to\infty}{\Rightarrow} \sigma^2 T_2^{1/2}\left(\xi_{N,T},\xi_{D,T}\right)$, where $\left(\xi_{N,T},\xi_{D,T}\right)$ is bivariate $N\left(0,\begin{bmatrix}2 & -1\\ -1 & T_1/2\end{bmatrix}\right)$. Next, decompose $\xi_{N,T}$ as $\xi_{N,T}=\xi_{N.D,T}+\frac{-1}{T_1/2}\xi_{D,T}$ where $\xi_{N.D,T}\equiv N\left(0,2-\frac{(-1)^2}{T_1/2}\right)=N\left(0,2\left(1-\frac{1}{T_1}\right)\right)$ is independent of $\xi_{D,T}$, so that

$$\begin{pmatrix}\xi_{N.D,T}\\ \xi_{D,T}\end{pmatrix}\equiv N\left(0,\begin{bmatrix}2\left(1-\frac{1}{T_1}\right) & 0\\ 0 & T_1/2\end{bmatrix}\right).$$

Combining these results, we have by joint weak convergence and continuous mapping that as $n \to \infty$ with T fixed,

$$\begin{aligned} \rho_{gmm} - 1 &= \frac{N_T}{D_T} \underset{n\to\infty}{\Rightarrow} \frac{\xi_{N,T}}{\xi_{D,T}} = \frac{\xi_{N.D,T} - \frac{2}{T_1}\xi_{D,T}}{\xi_{D,T}} \qquad (6.4) \\ &= -\frac{2}{T_1} + \frac{\xi_{N.D,T}}{\xi_{D,T}} = -\frac{2}{T_1} + \frac{2\left(1 - \frac{1}{T_1}\right)^{1/2}}{T_1^{1/2}} \frac{\zeta_N}{\zeta_D} \\ &\equiv -\frac{2}{T_1} + 2\frac{\left(1 - \frac{1}{T_1}\right)^{1/2}}{T_1^{1/2}} \mathbb{C}, \qquad (6.5) \end{aligned}$$

where $(\zeta_N, \zeta_D) \equiv N(0, I_2)$ and $\mathbb{C}$ is a standard Cauchy variate. Thus

$$\rho_{gmm} - 1 \underset{n\to\infty}{\Rightarrow} -\frac{2}{T_1} + 2\frac{\left(1 - \frac{1}{T_1}\right)^{1/2}}{T_1^{1/2}} \mathbb{C}, \qquad (6.6)$$

yielding the stated result. ■

Proof of Theorem 2. From (2.3) and (3.1) we have

$$\begin{aligned} \rho_{gmm} - 1 &= \frac{\sum_{i=1}^n \sum_{t=2}^T \Delta u_{it} y_{it-2}}{\sum_{i=1}^n \sum_{t=2}^T \Delta y_{it-1} y_{it-2}} \\ &= \frac{\sum_{i=1}^n \left\{ (u_{iT} y_{iT-2} - u_{i1} y_{i0}) - \sum_{t=3}^T u_{it-1} u_{it-2} \right\}}{\sum_{i=1}^n \sum_{t=2}^T u_{it-1} y_{it-2}}, \end{aligned}$$

and rescaling gives

$$\sqrt{T}\left(\rho_{gmm} - 1\right) = \frac{\sum_{i=1}^n \frac{1}{\sqrt{T}} \left\{ (u_{iT} y_{iT-2} - u_{i1} y_{i0}) - \sum_{t=3}^T u_{it-1} u_{it-2} \right\}}{\sum_{i=1}^n \frac{1}{T} \sum_{t=2}^T u_{it-1} y_{it-2}}. \qquad (6.7)$$

By partial summation

$$\sum_{t=1}^T u_{it} y_{it-1} = \sum_{t=1}^T u_{it} \left(\sum_{s=1}^{t-1} u_{is} + y_{i0} \right) = \frac{1}{2} \left\{ \left(\sum_{t=1}^T u_{it} \right)^2 - \sum_{t=1}^T u_{it}^2 \right\} + \sum_{t=1}^T u_{it} y_{i0}.$$

Using the fact that $\mathbb{E}\left(u_{it} u_{is} u_{is-1}\right) = 0$ for all (t, s), we have by standard functional limit theory for $r \in [0,1]$

$$T^{-1/2}\sum_{t=1}^{\lfloor Tr \rfloor}\begin{bmatrix} u_{it} \\ u_{it}u_{it-1}\end{bmatrix} \Rightarrow \begin{bmatrix} B_i(r) \\ G_i(r)\end{bmatrix} \equiv BM\left(\begin{bmatrix}\sigma^2 & 0 \\ 0 & \sigma^4\end{bmatrix}\right),$$

where B_i and G_i are independent Brownian motions for all i. Then, since $y_{i0} = O_p(1)$ and $T^{-1}\sum_{t=1}^{T} u_{it} = o_p(1)$, we deduce the joint weak convergence

$$\begin{bmatrix} T^{-1/2}\sum_{t=1}^{T} u_{it} \\ T^{-1}\sum_{t=1}^{T} u_{it}y_{it-1} \\ T^{-1}\sum_{t=1}^{T} u_{it}u_{it-1}\end{bmatrix} \underset{T\to\infty}{\Rightarrow} \begin{bmatrix} B_i(1) \\ \frac{1}{2}\left\{B_i(1)^2 - \sigma^2\right\} \\ G_i(1)\end{bmatrix} = \begin{bmatrix} B_i(1) \\ \int_0^1 B_i dB_i \\ G_i(1)\end{bmatrix}. \tag{6.8}$$

Since u_{it} is *iid* over t and i, it follows that $u_{iT} \Rightarrow u_{i\infty}$ as $T \to \infty$, where the limit variates $\{u_{i\infty}\}$ are independent over i and have the same distribution as u_{it}. Note that u_{iT} is independent of $\left(T^{-1/2}\sum_{t=1}^{T_1} u_{it}, T^{-1}\sum_{t=1}^{T_1} u_{it}y_{it-1}, T^{-1}\sum_{t=1}^{T_1} u_{it}u_{it-1}\right)$ and, hence, asymptotically independent of $\left(T^{-1/2}\sum_{t=1}^{T} u_{it}, T^{-1}\sum_{t=1}^{T} u_{it}y_{it-1}, T^{-1}\sum_{t=1}^{T} u_{it}u_{it-1}\right)$. It follows that $u_{i\infty}$ is independent of the vector of limit variates (6.8). We therefore have the combined weak convergence

$$\begin{bmatrix} T^{-1/2}\sum_{t=1}^{T} u_{it} \\ T^{-1}\sum_{t=1}^{T} u_{it}y_{it-1} \\ T^{-1/2}\sum_{t=1}^{T} u_{it}u_{it-1} \\ u_{iT}\end{bmatrix} \underset{T\to\infty}{\Rightarrow} \begin{bmatrix} B_i(1) \\ \frac{1}{2}\left\{B_i(1)^2 - \sigma^2\right\} \\ G_i(1) \\ u_{i\infty}\end{bmatrix} = \begin{bmatrix} B_i(1) \\ \int_0^1 B_i dB_i \\ G_i(1) \\ u_{i\infty}\end{bmatrix}. \tag{6.9}$$

Setting $G_i = G_i(1)$, the stated result (3.7)

$$\sqrt{T}\left(\rho_{gmm} - 1\right) \underset{T\to\infty}{\Rightarrow} \frac{\sum_{i=1}^{n}\left\{u_{i\infty}B_i(1) - G_i\right\}}{\sum_{i=1}^{n}\frac{1}{2}\left\{B_i(1)^2 - \sigma^2\right\}} \tag{6.10}$$

follows from (6.7) and (6.9) by continuous mapping.

For part (ii) we consider sequential asymptotics in which $T \to \infty$ is followed by $n \to \infty$. Observe that $u_{i\infty}B_i(1) - G_i$ is *iid* over i with zero mean and variance

$$\mathbb{E}\left\{u_{i\infty}B_i(1) - G_i(1)\right\}^2 = \mathbb{E}\left(u_{i\infty}^2\right)\mathbb{E}\left(B_i(1)^2\right) + \mathbb{E}\left(G_i(1)^2\right) = 2\sigma^4,$$

and is uncorrelated with $B_i(1)^2$. Since $\left\{B_i(1)^2 - \sigma^2\right\}$ is *iid* with zero mean and variance $2\sigma^4$, application of the Lindeberg Lévy CLT as $n \to \infty$ gives

$$\left[\begin{array}{c} n^{-1/2}\sum_{i=1}^{n}\left\{u_{i\infty}B_i(1) - G_i\right\} \\ n^{-1/2}\sum_{i=1}^{n}\frac{1}{2}\left\{B_i(1)^2 - \sigma^2\right\} \end{array}\right] \underset{n\to\infty}{\Rightarrow} \left[\begin{array}{c} \left(2\sigma^4\right)^{1/2}\zeta_N \\ \left(\sigma^4/2\right)^{1/2}\zeta_D \end{array}\right], \tag{6.11}$$

where $(\zeta_N, \zeta_D) \equiv N(0, I_2)$. Hence,

$$\sqrt{T}\left(\rho_{gmm} - 1\right) \underset{(n,T)_{\text{seq}}\to\infty}{\Rightarrow} 2\mathbb{C}, \tag{6.12}$$

giving the required result. ■

Proof of Theorem 3. We proceed by examining a set of sufficient conditions for joint convergence limit theory developed in Phillips and Moon (1999). In particular, we consider conditions that suffice to ensure that sequential convergence as $(n,T)_{\text{seq}} \to \infty$ (i.e., $T \to \infty$ followed by $n \to \infty$) implies joint convergence $(n,T) \to \infty$ where there is no restriction on the diagonal path in which n and T pass to infinity.

We start by defining the vector of standardized components appearing in the numerator and denominator of ρ_{gmm}

$$X_{nT} = \left(n^{-1/2}\sum_{i=1}^{n}\frac{1}{\sqrt{T}}\left\{\left(u_{iT}y_{iT-2} - u_{i1}y_{i0}\right) - \sum_{t=3}^{T}u_{it-1}u_{it-2}\right\}, n^{-1/2}\sum_{i=1}^{n}\left(\frac{1}{T}\sum_{t=2}^{T}u_{it-1}y_{it-2}\right)\right)'. \tag{6.13}$$

From (6.9) and (6.11) we have the sequential convergence

$$\begin{aligned} X_{nT} \underset{T\to\infty}{\Rightarrow} X_n \ &: = \left(n^{-1/2}\sum_{i=1}^{n}\left\{u_{i\infty}B_i(1) - G_i(1)\right\}, n^{-1/2}\sum_{i=1}^{n}\frac{1}{2}\left\{B_i(1)^2 - \sigma^2\right\}\right)' \\ \underset{n\to\infty}{\Rightarrow} X \ &: = \left(\left(2\sigma^4\right)^{1/2}\zeta_N, \left(\frac{\sigma^4}{2}\right)^{1/2}\zeta_D\right), \end{aligned} \tag{6.14}$$

which in turn implies the sequential limit $\sqrt{T}\left(\rho_{gmm} - 1\right) \underset{(n,T)_{\text{seq}}\to\infty}{\Rightarrow} 2\mathbb{C}$ given in (6.12). By Lemma 6(b) of Phillips and Moon (1999), when $X_{nT} \underset{T\to\infty}{\Rightarrow} X_n \underset{n\to\infty}{\Rightarrow} X$ sequentially, joint weak convergence $X_{nT} \Rightarrow X$ as $(n, T) \to \infty$ holds if and only if

$$\limsup_{n,T\to\infty} |\mathbb{E}f(X_{nT}) - \mathbb{E}f(X_n)| = 0 \tag{6.15}$$

for all bounded, continuous real functions f on $\mathbb{R}^2$.

Simple primitive conditions sufficient for (6.15) to hold are available in the case where the components of the random quantity X_{nT} involve averages of *iid* random variables as in the present case where we have $X_{nT} = n^{-1/2}\sum_{i=1}^{n} Y_{iT}$ with the Y_{iT} independent over i. Component-wise we have

$$X_{nT} = (X_{1nT}, X_{2nT})' := \left(n^{-1/2}\sum_{i=1}^{n} Y_{1iT}, n^{-1/2}\sum_{i=1}^{n} Y_{2iT} \right),$$

where $Y_{iT} = (Y_{1iT}, Y_{2iT})'$ with

$$\begin{aligned} Y_{1iT} &= \frac{1}{\sqrt{T}}\left\{ (u_{iT}y_{iT-2} - u_{i1}y_{i0}) - \sum_{t=3}^{T} u_{it-1}u_{it-2} \right\} \underset{T\to\infty}{\Rightarrow} Y_{1i} := u_{i\infty}B_i(1) - G_i(1), \\ Y_{2iT} &= \frac{1}{T}\sum_{t=2}^{T} u_{it-1}y_{it-2} \underset{T\to\infty}{\Rightarrow} Y_{2i} := \frac{1}{2}\left\{ B_i(1)^2 - \sigma^2 \right\}, \end{aligned}$$

for all i. The working probability space can be expanded as needed to ensure that the (limit) random quantities $Y_i := (Y_{1i}, Y_{2i})'$ are defined in the same space for all i so that averages involving $\sum_{i=1}^{n} Y_i$ are meaningful. In this framework we can use a result on joint convergence by Phillips and Moon (1999) – see lemma PM below – to verify condition (6.15). In what follows we use the notation of lemma PM.

We proceed to verify these conditions for Y_{iT} and Y_i. First, Y_{iT} is integrable since

$$\begin{aligned} \mathbb{E}|u_{iT}y_{iT-2}| &\leq \left(\mathbb{E}|u_{iT}|^2 \mathbb{E}|y_{iT-2}|^2 \right)^{1/2} < \infty, \\ \mathbb{E}\left| \sum_{t=2}^{T} u_{it-1}u_{it-2} \right| &\leq T\mathbb{E}|u_{it-1}u_{it-2}| \leq T\mathbb{E}\left(u_{it}^2\right) < \infty, \\ \mathbb{E}\left| \sum_{t=2}^{T} u_{it-1}y_{it-2} \right| &\leq \sum_{t=2}^{T} \mathbb{E}|u_{it-1}y_{it-2}| \leq \sum_{t=2}^{T} \left(\mathbb{E}u_{it-1}^2 \mathbb{E}y_{it-2}^2 \right)^{1/2} < \infty. \end{aligned}$$

To show (i) holds, observe that

$$
\begin{aligned}
\mathbb{E}\left\|Y_{iT}\right\|^2 &= \mathbb{E}Y_{1iT}^2+\mathbb{E}Y_{2iT}^2 \\
&= \frac{1}{T}\mathbb{E}\left\{u_{iT}y_{iT-2}-\sum_{t=3}^{T}u_{it-1}u_{it-2}\right\}^2+\frac{1}{T^2}\mathbb{E}\left(\sum_{t=2}^{T}u_{it-1}y_{it-2}\right)^2 \\
&= 2\sigma^4\frac{T-2}{T}+\sigma^4\frac{1}{T^2}\sum_{t=2}^{T}(t-2)<\infty, \qquad (6.16)
\end{aligned}
$$

when $y_{i0}=0$, with obviously valid extension to the case where $y_{i0}=O_p(1)$ with finite second moments. Then

$$
\limsup_{n,T\to\infty}\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left\|Y_{iT}\right\|=\limsup_{T\to\infty}\mathbb{E}\left\|Y_{iT}\right\|\leq\limsup_{T\to\infty}\left(\mathbb{E}\left\|Y_{iT}\right\|^2\right)^{1/2}<\infty,
$$

as required. To show (ii) holds, simply observe that $\mathbb{E}Y_{iT}=\mathbb{E}Y_i=0$. To show (iii) holds, note that

$$
\limsup_{n,T\to\infty}\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left\|Y_{iT}\right\|\mathbf{1}\left\{\left\|Y_{iT}\right\|>n\epsilon\right\}=\limsup_{T\to\infty}\mathbb{E}\left\|Y_{iT}\right\|\mathbf{1}\left\{\left\|Y_{iT}\right\|>n\epsilon\right\}=0,\ \text{ for all } \epsilon>0,
$$

since $\sup_T\mathbb{E}\left\|Y_{iT}\right\|^2<\infty$ by virtue of (6.16). Finally, note that

$$
\limsup_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left\|Y_i\right\|\mathbf{1}\left\{\left\|Y_i\right\|>n\epsilon\right\}=\limsup_{n\to\infty}\mathbb{E}\left\|Y_i\right\|\mathbf{1}\left\{\left\|Y_i\right\|>n\epsilon\right\}=0,
$$

since $\mathbb{E}\left\|Y_i\right\|^2<\infty$, proving (iv). Hence, condition (6.15) holds and we have joint weak convergence

$$
X_{nT}=n^{-1/2}\sum_{i=1}^{n}Y_{iT}\underset{n,T\to\infty}{\Rightarrow}X:=\left(\left(2\sigma^4\right)^{1/2}\zeta_N,\left(\frac{\sigma^4}{2}\right)^{1/2}\zeta_D\right),
$$

irrespective of the divergence rates of n and T to infinity. By continuous mapping, the required result follows for the GMM estimator so that $\sqrt{T}\left(\rho_{gmm}-1\right)\underset{n,T\to\infty}{\Rightarrow}2\mathbb{C}$ jointly as $(n,T)\to\infty$ irrespective of the order and rates of divergence of the respective sample sizes. ■

Lemma PM (Phillips and Moon, 1999, theorem 1) *Suppose the $m\times 1$ random vec-*

tors Y_{iT} are independent across i for all T and integrable. Assume that $Y_{iT} \Rightarrow Y_i$ as $T \to \infty$ for all i. Then, condition (6.15) holds if the following hold:

(i) $\limsup_{n,T\to\infty} \frac{1}{n}\sum_{i=1}^{n} E\|Y_{iT}\| < \infty$,

(ii) $\limsup_{n,T\to\infty} \frac{1}{n}\sum_{i=1}^{n} \|\mathbb{E}Y_{iT} - \mathbb{E}Y_i\| < \infty$,

(iii) $\limsup_{n,T\to\infty} \frac{1}{n}\sum_{i=1}^{n} E\|Y_{iT}\| 1\{\|Y_{iT}\| > n\epsilon\} = 0$, *for all* $\epsilon > 0$

(iv) $\limsup_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} E\|Y_i\| 1\{\|Y_i\| > n\epsilon\} = 0$, *for all* $\epsilon > 0$

Proof of Theorem 4. In case (i) T is fixed as well as $c < 0$, which implies that $\rho = 1 + \frac{c}{\sqrt{T}}$ is fixed. So large n asymptotics follow as in the (asymptotically) stationary case. From (2.1) we have $y_{it} = \alpha_i(1-\rho) + \rho y_{it-1} + u_{it} = -\frac{\alpha_i c}{\sqrt{T}} + \left(1 + \frac{c}{\sqrt{T}}\right) y_{it-1} + u_{it}$ and $\Delta y_{it} = \rho\Delta y_{it-1} + \Delta u_{it}$ so that $\Delta y_{it} = \alpha_i(1-\rho) + (\rho-1) y_{it-1} + u_{it} = -\frac{\alpha_i c}{\sqrt{T}} + \frac{c}{\sqrt{T}} y_{it-1} + u_{it}$. Then, as usual, $\mathbb{E}(u_{it}y_{it-2}) = \mathbb{E}(\Delta u_{it}y_{it-2}) = 0$ and orthogonality holds. When $y_{i0} = 0$, back substitution gives

$$y_{it} = \alpha_i\left(1-\rho^t\right) + \sum_{j=0}^{t-1} \rho^j u_{it-j},$$

and $\mathbb{E}(y_{it}) = \alpha_i\left(1-\rho^t\right)$, $\mathbb{V}\text{ar}(y_{it}) = \sigma^2\sum_{j=0}^{t-1}\rho^{2j} = \sigma^2\frac{1-\rho^{2t}}{1-\rho^2}$, and $\mathbb{E}\left(y_{it}^2\right) = \sigma^2\frac{1-\rho^{2t}}{1-\rho^2} + \alpha_i^2\left(1-\rho^t\right)^2$. Instrument relevance is determined by the magnitude of the moment

$$\begin{aligned}
\mathbb{E}(\Delta y_{it-1}y_{it-2}) &= \mathbb{E}(\{\alpha_i(1-\rho) + (\rho-1)y_{it-2} + u_{it-1}\}y_{it-2}) \\
&= \alpha_i^2(1-\rho)\left(1-\rho^{t-2}\right) + (\rho-1)\left\{\sigma^2\frac{1-\rho^{2(t-2)}}{1-\rho^2} + \alpha_i^2\left(1-\rho^{t-2}\right)^2\right\} \\
&= -\sigma^2\frac{1-\rho^{2(t-2)}}{1+\rho} - \alpha_i^2(1-\rho)\left(1-\rho^{t-2}\right)\rho^{t-2} \qquad (6.17)
\end{aligned}$$

which is non zero for $c < 0$ and zero when $c = 0$, corresponding to the unit root case ($\rho = 1$) considered earlier. Note that in the fully stationary case where initial conditions are in the infinite past so that $y_{i0} = \alpha_i + \sum_{j=0}^{\infty}\rho^j u_{i,-j}$ and $y_{it} = \alpha_i + \sum_{j=0}^{\infty}\rho^j u_{t-j}$ we have

$$\begin{aligned}
\mathbb{E}(\Delta y_{it-1}y_{it-2}) &= \alpha_i^2(1-\rho) + (\rho-1)\mathbb{E}\left(y_{it}^2\right) = \alpha_i^2(1-\rho) - (1-\rho)\left\{\frac{\sigma^2}{1-\rho^2} + \alpha_i^2\right\} \\
&= -\frac{\sigma^2}{1+\rho},
\end{aligned}$$

which corresponds with the leading term of (6.17) when $t \to \infty$ with $|\rho| < 1$.

Now consider the numerator and denominator of the centred and scaled GMM estimate

$$\sqrt{n}\left(\rho_{gmm} - \rho\right) = \frac{\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{t=2}^{T} \Delta u_{it} y_{it-2}}{\frac{1}{n} \sum_{i=1}^{n} \sum_{t=2}^{T} \Delta y_{it-1} y_{it-2}} =: \frac{N_{nT}}{D_{nT}}. \tag{6.18}$$

First, noting that $\Delta y_{it-1} y_{it-2}$ is quadratic in α_i, and using $T_j = T - j$ and $\sigma_\alpha^2 = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \alpha_i^2$, the denominator of (6.18) takes the following form as $n \to \infty$

$$\begin{aligned}
& \frac{1}{n} \sum_{i=1}^{n} \sum_{t=2}^{T} \Delta y_{it-1} y_{it-2} \to_p \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left(\sum_{t=2}^{T} \Delta y_{it} y_{it-2} \right) \\
= & \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \sum_{t=2}^{T} \left\{ -\sigma^2 \frac{1-\rho^{2(t-2)}}{1+\rho} + \alpha_i^2 (1-\rho)\left(1-\rho^{t-2}\right)\left[1-\left(1-\rho^{t-2}\right)\right] \right\} \\
= & -\frac{\sigma^2}{1+\rho}\left[T_1 - \frac{1-\rho^{2T_1}}{1-\rho^2}\right] + \sigma_\alpha^2 (1-\rho)\left[T_1 - \frac{1-\rho^{T_1}}{1-\rho}\right] - \sigma_\alpha^2 (1-\rho)\left[T_1 - 2\frac{1-\rho^{T_1}}{1-\rho} + \frac{1-\rho^{2T_1}}{1-\rho^2}\right] \\
= & -\frac{\sigma^2}{1+\rho}\left[T_1 - \frac{1-\rho^{2T_1}}{1-\rho^2}\right] + \sigma_\alpha^2 (1-\rho)\left[\frac{1-\rho^{T_1}}{1-\rho} - \frac{1-\rho^{2T_1}}{1-\rho^2}\right] \\
= & -\frac{\sigma^2}{1+\rho}\left[T_1 - \frac{1-\rho^{2T_1}}{1-\rho^2}\right] + \sigma_\alpha^2\left[1-\rho^{T_1} - \frac{1-\rho^{2T_1}}{1+\rho}\right],
\end{aligned} \tag{6.19}$$

which is again zero when $c = 0$ ($\rho = 1$). Turning to the numerator, we have $\mathbb{E}\left(\Delta u_{it} y_{it-2}\right) = 0$ by orthogonality and by a standard CLT argument for fixed T as $n \to \infty$

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\sum_{t=2}^{T} \Delta u_{it} y_{it-2} \right) \Rightarrow N\left(0, v_T\right)$$

with

$$v_T = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left(\sum_{t=2}^{T} \Delta u_{it} y_{it-2} \right)^2$$

We evaluate the above variance as follows. Using (3.2) and $y_{i0} = 0$, we have

$$\sum_{t=2}^{T} \Delta u_{it} y_{it-2} = u_{iT} y_{iT-2} - u_{i1} y_{i0} - \sum_{t=3}^{T} u_{it-1} \Delta y_{it-2} = u_{iT} y_{iT-2} - \sum_{t=3}^{T} u_{it-1} \Delta y_{it-2}, \tag{6.20}$$

with variance

$$\begin{aligned}
\mathbb{E}\left(u_{iT}y_{iT-2}-\sum_{t=3}^{T}u_{it-1}\Delta y_{it-2}\right)^{2} &= \sigma^{2}\mathbb{E}\left(y_{iT-2}\right)^{2}+\sigma^{2}\mathbb{E}\sum_{t=3}^{T}\left(\Delta y_{it-2}\right)^{2} \\
&= \sigma^{4}\frac{1-\rho^{2T_{2}}}{1-\rho^{2}}+\alpha_{i}^{2}\sigma^{2}\left(1-\rho^{T_{2}}\right)^{2}+\sigma^{2}\mathbb{E}\sum_{t=3}^{T}\left(\Delta y_{it-2}\right)^{2}.
\end{aligned}$$

Using $\mathbb{E}\left(y_{it}\right)=\alpha_{i}\left(1-\rho^{t}\right)$, $\mathbb{V}\text{ar}\left(y_{it}\right)=\sigma^{2}\sum_{j=0}^{t-1}\rho^{2j}=\sigma^{2}\frac{1-\rho^{2t}}{1-\rho^{2}}$, $\mathbb{E}\left(y_{it}^{2}\right)=\sigma^{2}\frac{1-\rho^{2t}}{1-\rho^{2}}+\alpha_{i}^{2}\left(1-\rho^{t}\right)^{2}$, and $\Delta y_{it}=\alpha_{i}\left(1-\rho\right)+\left(\rho-1\right)y_{it-1}+u_{it}$, the final term $\sum_{t=3}^{T}\mathbb{E}\left(\Delta y_{it-2}\right)^{2}$ above is

$$\begin{aligned}
& \sum_{t=3}^{T}\left\{\alpha_{i}^{2}\left(1-\rho\right)^{2}+\left(1-\rho\right)^{2}\mathbb{E}\left(y_{it-3}^{2}\right)+\sigma^{2}-2\alpha_{i}^{2}\left(1-\rho\right)^{2}\left(1-\rho^{t-3}\right)\right\} \\
= \;& \sigma^{2}T_{2}-T_{2}\alpha_{i}^{2}\left(1-\rho\right)^{2}+2\alpha_{i}^{2}\left(1-\rho\right)^{2}\left(\frac{1-\rho^{T_{2}}}{1-\rho}\right)+\left(1-\rho\right)^{2}\sum_{t=3}^{T}\mathbb{E}\left(y_{it-3}^{2}\right) \\
= \;& \sigma^{2}T_{2}-T_{2}\alpha_{i}^{2}\left(1-\rho\right)^{2}+2\alpha_{i}^{2}\left(1-\rho\right)\left(1-\rho^{T_{2}}\right)+\left(1-\rho\right)^{2}\frac{\sigma^{2}}{1-\rho^{2}}\left[T_{2}-\frac{1-\rho^{2T_{2}}}{1-\rho^{2}}\right] \\
& +\alpha_{i}^{2}\left(1-\rho\right)^{2}\left[T_{2}-2\frac{1-\rho^{T_{2}}}{1-\rho}+\frac{1-\rho^{2T_{2}}}{1-\rho^{2}}\right] \\
= \;& \sigma^{2}T_{2}\left[1+\frac{\left(1-\rho\right)}{1+\rho}\right]-\frac{\sigma^{2}\left(1-\rho^{2T_{2}}\right)}{\left(1+\rho\right)^{2}}+2\alpha_{i}^{2}\left(1-\rho\right)\left(1-\rho^{T_{2}}\right)+\alpha_{i}^{2}\left(1-\rho\right)^{2}\left[-2\frac{1-\rho^{T_{2}}}{1-\rho}+\frac{1-\rho^{2T_{2}}}{1-\rho^{2}}\right] \\
= \;& \frac{2\sigma^{2}T_{2}}{1+\rho}-\frac{\sigma^{2}\left(1-\rho^{2T_{2}}\right)}{\left(1+\rho\right)^{2}}+\alpha_{i}^{2}\left(1-\rho\right)\frac{1-\rho^{2T_{2}}}{1+\rho}.
\end{aligned}$$

Then

$$\begin{aligned}
& \mathbb{E}\left(u_{iT}y_{iT-2}-\sum_{t=3}^{T}u_{it-1}\Delta y_{it-2}\right)^{2}=\sigma^{4}\frac{1-\rho^{2T_{2}}}{1-\rho^{2}}+\alpha_{i}^{2}\sigma^{2}\left(1-\rho^{T_{2}}\right)^{2}+\sigma^{2}\mathbb{E}\sum_{t=3}^{T}\left(\Delta y_{it-2}\right)^{2} \\
= \;& \sigma^{4}\frac{1-\rho^{2T_{2}}}{1-\rho^{2}}+\alpha_{i}^{2}\sigma^{2}\left(1-\rho^{T_{2}}\right)^{2}+\frac{2\sigma^{4}T_{2}}{1+\rho}-\frac{\sigma^{4}\left(1-\rho^{2T_{2}}\right)}{\left(1+\rho\right)^{2}}+\alpha_{i}^{2}\sigma^{2}\left(1-\rho\right)\frac{1-\rho^{2T_{2}}}{1+\rho} \\
= \;& \frac{2\sigma^{4}T_{2}}{1+\rho}+\sigma^{4}\frac{2\rho\left(1-\rho^{2T_{2}}\right)}{\left(1-\rho^{2}\right)\left(1+\rho\right)}+\alpha_{i}^{2}\sigma^{2}\left(1-\rho^{T_{2}}\right)^{2}+\alpha_{i}^{2}\sigma^{2}\left(1-\rho\right)\frac{\left(1-\rho^{2T_{2}}\right)}{1+\rho} \\
= \;& \frac{2\sigma^{4}T_{2}}{1+\rho}+\sigma^{4}\frac{2\rho\left(1-\rho^{2T_{2}}\right)}{\left(1-\rho^{2}\right)\left(1+\rho\right)}+\alpha_{i}^{2}\sigma^{2}\left(1-\rho^{T_{2}}\right)\left[1-\rho^{T_{2}}+\frac{\left(1-\rho\right)\left(1+\rho^{T_{2}}\right)}{1+\rho}\right],
\end{aligned}$$

and

$$\omega_{NT} = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left(\sum_{t=2}^{T} \Delta u_{it} y_{it-2}\right)^2$$
$$= \frac{2\sigma^4 T_2}{1+\rho} + \sigma^4 \frac{2\rho\left(1-\rho^{2T_2}\right)}{\left(1-\rho^2\right)\left(1+\rho\right)} + \sigma_\alpha^2\sigma^2\left(1-\rho^{T_2}\right)\left[1-\rho^{T_2}+\frac{\left(1-\rho\right)\left(1+\rho^{T_2}\right)}{1+\rho}\right] \tag{6.21}$$

From (6.19) we have

$$\omega_{DT} = \left\{-\frac{\sigma^2}{1+\rho}\left[T_1 - \frac{1-\rho^{2T_1}}{1-\rho^2}\right] + \sigma_\alpha^2\left[1-\rho^{T_1}-\frac{1-\rho^{2T_1}}{1+\rho}\right]\right\}^2 \tag{6.22}$$

which leads to the asymptotic variance

$$\omega_T^2 = \frac{\omega_{NT}}{\omega_{DT}} = \frac{\frac{2\sigma^4 T_2}{1+\rho} + \sigma^4 \frac{2\rho\left(1-\rho^{2T_2}\right)}{\left(1-\rho^2\right)\left(1+\rho\right)} + \sigma_\alpha^2\sigma^2\left(1-\rho^{T_2}\right)\left[1-\rho^{T_2}+\frac{\left(1-\rho\right)\left(1+\rho^{T_2}\right)}{1+\rho}\right]}{\left\{-\frac{\sigma^2}{1+\rho}\left[T_1 - \frac{1-\rho^{2T_1}}{1-\rho^2}\right] + \sigma_\alpha^2\left[1-\rho^{T_1}-\frac{1-\rho^{2T_1}}{1+\rho}\right]\right\}^2}$$
$$= \frac{2\left(1+\rho\right)}{T_1} + O\left(\frac{1}{T_1^{3/2}}\right), \tag{6.23}$$

giving the stated result for (i). The error magnitude as $T\to\infty$ in the asymptotic expansion (6.23) is justified as follows. Since $c<0$ is fixed we have $1-\rho^2 = -2\frac{c}{\sqrt{T}} - \frac{c^2}{T}$ and

$$\frac{1-\rho^{2T}}{1-\rho^2} = \frac{1-\left[1+\frac{c}{\sqrt{T}}\right]^{2T}}{-2\frac{c}{\sqrt{T}}-\frac{c^2}{T}} \sim \frac{1-e^{c\frac{2T}{\sqrt{T}}}}{-2\frac{c}{\sqrt{T}}-\frac{c^2}{T}} \sim \frac{\sqrt{T}}{-2c}. \tag{6.24}$$

Then, by direct calculation as $T\to\infty$

$$\frac{\omega_{NT}}{\omega_{DT}} = \frac{\frac{2\sigma^4 T_2}{1+\rho} + \sigma^4\frac{2\rho}{(1+\rho)}\left(\frac{\sqrt{T_2}}{-2c}\right) + \sigma_\alpha^2\sigma^2\left(1-e^{cT_2/\sqrt{T}}\right)\left[1-e^{cT_2/\sqrt{T}} - c\frac{1+e^{cT_2/\sqrt{T}}}{\sqrt{T}(1+\rho)}\right]}{\left\{-\frac{\sigma^2}{1+\rho}\left[T_1 - \left(\frac{\sqrt{T}}{-2c}\right)\left(1-e^{cT_1/\sqrt{T}}\right)\right] + \sigma_\alpha^2\left[1-e^{cT_1/\sqrt{T}} - \frac{1-e^{2cT_1/\sqrt{T}}}{1+\rho}\right]\right\}^2}$$
$$= \frac{2\left(1+\rho\right)}{T_1} + O\left(\frac{1}{T_1^{3/2}}\right) \tag{6.25}$$

The sequential limit theory (ii) follows directly from (i) and the asymptotic expansion (6.23) of ω_T^2.

If $\rho = 1 + \frac{c}{T^\gamma}$ with $\gamma \in (0,1)$, it is clear that the above fixed (T, c) limit theory as $n \to \infty$ continues to hold. Then, as $T \to \infty$, we have in place of (6.24)

$$\frac{1-\rho^{2T}}{1-\rho^2} = \frac{1-\left[1+\frac{c}{T^\gamma}\right]^{2T}}{-2\frac{c}{T^\gamma} - \frac{c^2}{T^\gamma}} \sim \frac{1-e^{2cT^{1-\gamma}}}{-2\frac{c}{T^\gamma} - \frac{c^2}{T^\gamma}} \sim \frac{T^\gamma}{-2c}$$

leading to

$$\frac{\omega_{NT}}{\omega_{DT}} = \frac{2(1+\rho)}{T_1} + O\left(\frac{1}{T_1^{2-\gamma}}\right) \sim \frac{4}{T_1} + O\left(\frac{1}{T_1^{2-\gamma}}\right)$$

It follows that (ii) continues to hold with the same convergence rate $\sqrt{nT}$ and same limit variance 4 for all $\gamma \in (0,1)$.

When $\gamma = 1$, the sequential normal limit theory in (ii) still holds but the variance of the limiting distribution changes. Observe that in this case

$$\frac{1-\rho^{2T}}{1-\rho^2} = \frac{1-\left[1+\frac{c}{T}\right]^{2T}}{-2\frac{c}{T} - \frac{c^2}{T}} \sim \frac{1-e^{2c}}{-2\frac{c}{T} - \frac{c^2}{T}} \sim \frac{T\left(1-e^{2c}\right)}{-2c}.$$

Using (6.21) we then have the following limit behavior as $T \to \infty$

$$\frac{\omega_{NT}}{\omega_{DT}} \sim \frac{\sigma^4 T_2 + \sigma^4 T_2 \frac{\left(1-e^{2c}\right)}{-2c} + O(1)}{\left\{-\frac{\sigma^2 T_1}{2}\left[1 - \frac{\left(1-e^{2c}\right)}{-2c}\right] + O(1)\right\}^2} = \frac{4}{T_1} \frac{1 + \frac{\left(1-e^{2c}\right)}{-2c}}{\left[1 - \frac{\left(1-e^{2c}\right)}{-2c}\right]^2} \{1 + o(1)\},$$

so that

$$\omega_T^2 = \frac{-8c}{T_1} \frac{\left(1-2c-e^{2c}\right)}{\left(1+2c-e^{2c}\right)^2} \{1 + o(1)\}.$$

Hence

$$\sqrt{nT}\left(\rho_{gmm} - \rho\right) \underset{(T,n)_{\text{seq}} \to \infty}{\Rightarrow} N\left(0, (-8c)\frac{\left(1-2c-e^{2c}\right)}{\left(1+2c-e^{2c}\right)^2}\right), \tag{6.26}$$

so the $\sqrt{nT}$ Gaussian limit theory holds but with a different variance when $\rho = 1 + \frac{c}{T}$. Observe that

$$(-8c)\frac{\left(1-2c-e^{2c}\right)}{\left(1+2c-e^{2c}\right)^2} \sim \frac{8}{c^2} \to \infty \quad \text{as } c \to 0,$$

indicating that the variance in (6.26) diverges and the $\sqrt{nT}$ convergence rate fails as the unit root is approached via $c \to 0$.

Next, examine the case where $\rho = 1 + \frac{c}{T^\gamma}$ with $\gamma > 1$ and $c < 0$, so that ρ is in the immediate vicinity of unity, closer than the LUR case but still satisfying $\rho < 1$ for fixed T. In that case, we still have Gaussian limit theory as $n \to \infty$ because $|\rho| < 1$. To find the limit theory as $(T,n)_{\text{seq}} \to \infty$ we consider the behavior of the numerator and denominator of ω_T. First, note that $\log\left[1+\frac{c}{T^\gamma}\right]^{2T} = \frac{2c}{T^{\gamma-1}} - \frac{c^2}{T^{2\gamma-1}} + O\left(\frac{1}{T^{3\gamma-1}}\right)$ so that $\left[1+\frac{c}{T^\gamma}\right]^{2T} = 1 + \frac{2c}{T^{\gamma-1}} - \frac{c^2}{T^{2\gamma-1}} + \frac{1}{2}\left(\frac{2c}{T^{\gamma-1}}\right)^2 + O\left(\frac{1}{T^{2\gamma-2}}\right)$ giving

$$\begin{aligned}
\frac{1-\rho^{2T}}{1-\rho^2} &= \frac{1-\left[1+\frac{c}{T^\gamma}\right]^{2T}}{-2\frac{c}{T^\gamma}-\frac{c^2}{T^{2\gamma}}} = \frac{1-\left[1+\frac{2c}{T^{\gamma-1}}-\frac{c^2}{T^{2\gamma-1}}+\frac{1}{2}\left(\frac{2c}{T^{\gamma-1}}\right)^2+O\left(\frac{1}{T^{3(\gamma-1)}}\right)\right]}{-2\frac{c}{T^\gamma}-\frac{c^2}{T^{2\gamma}}} \\
&= \frac{-2cT+\frac{c^2}{T^{\gamma-1}}-\frac{2c^2}{T^{\gamma-2}}+o\left(\frac{1}{T^{\gamma-2}}\right)}{-2c\left\{1+\frac{1}{2}\frac{c}{T^\gamma}+O\left(\frac{1}{T^{2\gamma}}\right)\right\}} = \left\{T+cT^{2-\gamma}+O\left(T^{1-\gamma}\right)\right\}\left\{1+\frac{1}{2}\frac{c}{T^\gamma}+O\left(\frac{1}{T^{2\gamma}}\right)\right\}^{-1} \\
&= T+cT^{2-\gamma}+O\left(\frac{1}{T^{\gamma-1}}\right).
\end{aligned}$$

Using this result, and letting $\omega_{DT} = v_{DT}^2$ with $v_{DT} = -\frac{\sigma^2}{1+\rho}\left[T_1 - \frac{1-\rho^{2T_1}}{1-\rho^2}\right] + \sigma_\alpha^2\left[1-\rho^{T_1}-\frac{1-\rho^{2T_1}}{1+\rho}\right]$, we have

$$\begin{aligned}
v_{DT} &= -\frac{\sigma^2}{1+\rho}\left[cT_1^{2-\gamma}\{1+o(1)\}\right] + \sigma_\alpha^2\left\{-\left[\frac{c}{T_1^{\gamma-1}}+\frac{T_1(T_1-1)}{2}\left(\frac{c}{T_1^\gamma}\right)^2+O\left(\frac{1}{T_1^{3(\gamma-1)}}\right)\right]\right. \\
&\qquad \left. -\frac{-\frac{2c}{T_1^{\gamma-1}}-\frac{T_1(T_1-1)}{2}\left(\frac{c}{T_1^\gamma}\right)^2+O\left(\frac{1}{T_1^{3(\gamma-1)}}\right)}{2+\frac{2c}{T_1^{\gamma-1}}\{1+o(1)\}}\right\} \\
&= -\frac{c\sigma^2}{1+\rho}T_1^{2-\gamma}\{1+o(1)\} + \sigma_\alpha^2\left[-\frac{2c}{T_1^{\gamma-1}}-\frac{1}{2}\frac{c^2}{T_1^{2(\gamma-1)}}+O\left(\frac{1}{T_1^{3(\gamma-1)}}\right)\right. \\
&\qquad \left. +\left\{\frac{c}{T_1^{\gamma-1}}+\frac{T_1(T_1-1)}{4}\left(\frac{c}{T_1^\gamma}\right)^2+O\left(\frac{1}{T_1^{3(\gamma-1)}}\right)\right\}\left\{1+\frac{c}{T_1^{\gamma-1}}\{1+o(1)\}\right\}^{-1}\right] \\
&= -\frac{c\sigma^2}{1+\rho}T_1^{2-\gamma}\{1+o(1)\} + \sigma_\alpha^2\left[-\frac{c}{T_1^{\gamma-1}}-\frac{1}{4}\frac{c^2}{T_1^{2(\gamma-1)}}+O\left(\frac{1}{T_1^{3(\gamma-1)}}\right)\right]\{1+o(1)\} \\
&= \left\{-\frac{c\sigma^2}{1+\rho}T_1^{2-\gamma}-c\sigma_\alpha^2\frac{c}{T_1^{\gamma-1}}\right\}\{1+o(1)\} = -\frac{c\sigma^2}{2}T_1^{2-\gamma}\{1+o(1)\},
\end{aligned}$$

so that the denominator is $\omega_{DT}=\frac{c^2\sigma^4}{4}T_1^{4-2\gamma}\left\{1+o\left(1\right)\right\}$. The numerator ω_{NT} is

$$\begin{aligned}
&\frac{2\sigma^4T_2}{1+\rho}+\sigma^4\frac{2\rho\left(1-\rho^{2T_2}\right)}{\left(1-\rho^2\right)\left(1+\rho\right)}+\sigma_\alpha^2\sigma^2\left(1-\rho^{T_2}\right)\left[1-\rho^{T_2}+\frac{\left(1-\rho\right)\left(1+\rho^{T_2}\right)}{1+\rho}\right]\\
=\;&\sigma^4T_2\left\{1-\frac{c}{T_1^{\gamma-1}}\right\}\left\{1+o\left(1\right)\right\}+\sigma^4\frac{\left(2+\frac{2c}{T^\gamma}\right)}{\left(2+\frac{c}{T^\gamma}\right)}\left\{T+cT^{2-\gamma}-\frac{1}{2}\frac{c}{T^{\gamma-1}}+O\left(\frac{T}{T^{2(\gamma-1)}}\right)\right\}\\
&+\sigma_\alpha^2\sigma^2\left\{-2cT_2-\frac{2T_2\left(2T_2-1\right)}{2}\frac{c^2}{T_2^\gamma}+O\left(\frac{T^3}{T^{2\gamma}}\right)\right\}\\
&\times\left\{-\left[\frac{c}{T_2^{\gamma-1}}+\frac{T_2\left(T_2-1\right)}{2}\left(\frac{c}{T_2^\gamma}\right)^2\left\{1+o\left(1\right)\right\}\right]-\frac{\frac{2c}{T^{\gamma-1}}\left[1+\frac{2c}{T_2^{\gamma-1}}+\frac{2T_2(2T_2-1)}{2}\left(\frac{c}{T_2^\gamma}\right)^2\left\{1+o\left(1\right)\right\}\right]}{2+\frac{2c}{T^{\gamma-1}}\left\{1+o\left(1\right)\right\}}\right\}\\
=\;&\left\{2\sigma^4T_2+2\sigma_\alpha^2\sigma^2\left(-c\right)T\left(\frac{-2c}{T_2^{\gamma-1}}\right)\right\}\left[1+o\left(1\right)\right]=2\sigma^4T_2\left[1+o\left(1\right)\right].
\end{aligned}$$

Combining these results we obtain

$$\omega_T^2=\frac{\omega_{NT}}{\omega_{DT}}=\frac{2\sigma^4T_2\left[1+o\left(1\right)\right]}{\left\{\frac{c^2\sigma^4}{2}T_1^{4-2\gamma}\left\{1+o\left(1\right)\right\}\right\}^2\left\{1+o\left(1\right)\right\}}=\frac{8T_2}{c^2T_1^{2(2-\gamma)}}\left\{1+o\left(1\right)\right\}.$$

It now follows that for $\rho=1+\frac{c}{T^\gamma}$ with $c<0$ fixed and $\gamma>1$

$$\sqrt{nT^{3-2\gamma}}\left(\rho_{gmm}-\rho\right)\underset{(T,n)_{\text{seq}}\to\infty}{\Rightarrow}N\left(0,\frac{8}{c^2}\right).$$

Hence when ρ is closer to unity than a local unit root, the $\sqrt{nT}$ rate of convergence is reduced to $\sqrt{n}T^{\frac{3-2\gamma}{2}}$. When $\gamma=\frac{3}{2}$ the rate of convergence is simply $\sqrt{n}$ and for $\gamma>\frac{3}{2}$ the large n Gaussian asymptotic distribution $N\left(0,\omega_T^2\right)$ diverges as $T\to\infty$ because $\omega_T^2=\frac{8T_2}{c^2T_1^{2(2-\gamma)}}\left\{1+o\left(1\right)\right\}$ diverges with T. In this event, sequential $(T,n)_{\text{seq}}\to\infty$ asymptotics fail. In effect, the convergence rate is slower than $\sqrt{n}$ and the non-Gaussian Cauchy limit theory cannot be captured in these $(T,n)_{\text{seq}}$ directional sequential asymptotics even though $\rho=1+\frac{c}{T^\gamma}$ with $\gamma>1$ is closer proximity to a unit root than than the usual local unit root case with $\gamma=1$. ■

Proof of Theorem 5. In the mildly integrated case where $\rho=1+\frac{c}{\sqrt{T}}$ we have $y_{it}=-\frac{\alpha_ic}{\sqrt{T}}+\left(1+\frac{c}{\sqrt{T}}\right)y_{it-1}+u_{it}$ and $\Delta y_{it}=\rho\Delta y_{it-1}+\Delta u_{it}$ so that $\Delta y_{it}=-\frac{\alpha_ic}{\sqrt{T}}+$

$\frac{c}{\sqrt{T}} y_{it-1} + u_{it} = \alpha_i (1-\rho) + (\rho - 1) y_{it-1} + u_{it}$. By partial summation, as shown above in (3.2), we have $\sum_{t=2}^{T} \Delta u_{it} y_{it-2} = u_{iT} y_{iT-2} - u_{i1} y_{i0} - \sum_{t=3}^{T} u_{it-1} \Delta y_{it-2}$, so that

$$\rho_{gmm} - \rho = \frac{\sum_{i=1}^{n} \left\{ (u_{iT} y_{iT-2} - u_{i1} y_{i0}) - \sum_{t=3}^{T} u_{it-1} \left[-\frac{\alpha_i c}{\sqrt{T}} + \frac{c}{\sqrt{T}} y_{it-3} + u_{it-2} \right] \right\}}{\sum_{i=1}^{n} \sum_{t=2}^{T} \left\{ -\frac{\alpha_i c}{\sqrt{T}} + \frac{c}{\sqrt{T}} y_{it-2} + u_{it-1} \right\} y_{it-2}}. \quad (6.27)$$

Rescaling and using $y_{i0} = 0$ gives

$$\sqrt{T} \left(\rho_{gmm} - \rho \right) = \frac{\sum_{i=1}^{n} \frac{1}{\sqrt{T}} \left\{ u_{iT} y_{iT-2} - \sum_{t=3}^{T} u_{it-1} \left[-\frac{\alpha_i c}{\sqrt{T}} + \frac{c}{\sqrt{T}} y_{it-3} + u_{it-2} \right] \right\}}{\sum_{i=1}^{n} \frac{1}{T} \sum_{t=2}^{T} \left\{ -\frac{\alpha_i c}{\sqrt{T}} + \frac{c}{\sqrt{T}} y_{it-2} + u_{it-1} \right\} y_{it-2}}. \quad (6.28)$$

Since $\frac{1}{\sqrt{T}} \sum_{t=2}^{T} u_{it-1} u_{it-2} \Rightarrow G_i \equiv N\left(0, \sigma^4\right)$, $\frac{1}{T} \sum_{t=2}^{T} u_{it-1} = o_p(1)$, and $\frac{1}{T} \sum_{t=2}^{T} u_{it-1} y_{it-3} = o_p(1)$ – see (6.30) below – the numerator is

$$\begin{aligned} & \sum_{i=1}^{n} \frac{1}{\sqrt{T}} \left\{ u_{iT} y_{iT-2} - \sum_{t=3}^{T} u_{it-1} \left[-\frac{\alpha_i c}{\sqrt{T}} + \frac{c}{\sqrt{T}} y_{it-3} + u_{it-2} \right] \right\} \\ = & -\sum_{i=1}^{n} \frac{1}{\sqrt{T}} \sum_{t=3}^{T} u_{it-1} u_{it-2} + o_p(1) \Rightarrow -\sum_{i=1}^{n} G_i(1). \end{aligned} \quad (6.29)$$

Using Phillips and Magdalinos (2007, theorem 3.2) we find that

$$T^{-3/2} \sum_{t=2}^{T} y_{it}^2 \to_p \frac{\sigma^2}{-2c}, T^{-3/4} \sum_{t=2}^{T} y_{it-1} u_{it} \Rightarrow N\left(0, \frac{\sigma^4}{-2c}\right), \text{ and } T^{-3/2} \sum_{t=2}^{T} y_{it} = o_p(1). \quad (6.30)$$

The denominator of (6.28) therefore satisfies

$$\sum_{i=1}^{n} \frac{1}{T} \sum_{t=2}^{T} \left\{ -\frac{\alpha_i c}{\sqrt{T}} + \frac{c}{\sqrt{T}} y_{it-2} + u_{it-1} \right\} y_{it-2} \to_p \sum_{i=1}^{n} \left\{ c \frac{\sigma^2}{-2c} \right\}. \quad (6.31)$$

Hence, using (6.29) and (6.31) we have $\sqrt{T} \left(\rho_{gmm} - \rho \right) \underset{T \to \infty}{\Rightarrow} \frac{2}{n \sigma^2} \sum_{i=1}^{n} G_i = \frac{2}{n} \sum_{i=1}^{n} \zeta_i$ where $\zeta_i \sim_{iid} N(0,1)$. Then

$$\sqrt{nT} \left(\rho_{gmm} - \rho \right) \underset{T \to \infty}{\Rightarrow} \frac{2}{\sqrt{n}} \sum_{i=1}^{n} \zeta_i \underset{n \to \infty}{\Rightarrow} N(0,4), \quad (6.32)$$

which gives (i) and then leads directly to the sequential limit $\sqrt{nT}\left(\rho_{gmm}-\rho\right) \underset{(n,T)_{\text{seq}}\to\infty}{\Rightarrow} N\left(0,4\right)$. ■

Proof of Theorem 6. The proof follows the same lines as the proof of Theorem 3 above. As before, we define the vector of standardized components appearing in the numerator and denominator of $\sqrt{T}\left(\rho_{gmm}-\rho\right)$ in (6.28)

$$X_{nT} = \left(X_{1nT}, X_{2nT}\right)' := \left(n^{-1/2}\sum_{i=1}^{n} Y_{1iT}, n^{-1}\sum_{i=1}^{n} Y_{2iT}\right),$$

where $Y_{iT} = \left(Y_{1iT}, Y_{2iT}\right)'$ with

$$\begin{aligned} Y_{1iT} &= \frac{1}{\sqrt{T}}\left\{u_{iT}y_{iT-2} - \sum_{t=3}^{T} u_{it-1}\left[-\frac{\alpha_i c}{\sqrt{T}} + \frac{c}{\sqrt{T}}y_{it-3} + u_{it-2}\right]\right\} \underset{T\to\infty}{\Rightarrow} Y_{1i} := -G_i\left(1\right), \\ Y_{2iT} &= \frac{1}{T}\sum_{t=2}^{T}\left\{-\frac{\alpha_i c}{\sqrt{T}} + \frac{c}{\sqrt{T}}y_{it-2} + u_{it-1}\right\} y_{it-2} \underset{T\to\infty}{\Rightarrow} Y_{2i} := c\frac{\sigma^2}{-2c}. \end{aligned}$$

From (6.29) and (6.31) we have the sequential convergence

$$\begin{aligned} X_{nT} \underset{T\to\infty}{\Rightarrow} X_n &:= \left(-n^{-1/2}\sum_{i=1}^{n} G_i\left(1\right), n^{-1}\sum_{i=1}^{n}\left\{c\frac{\sigma^2}{-2c}\right\}\right)' \\ \underset{n\to\infty}{\Rightarrow} X &:= \left(\sigma^2\zeta, -\frac{\sigma^2}{2}\right), \text{ where } \zeta = N\left(0,1\right), \end{aligned} \tag{6.33}$$

which in turn implies the sequential limit $\sqrt{nT}\left(\rho_{gmm}-\rho\right) \underset{n\to\infty,T\to\infty}{\Rightarrow} N\left(0,4\right)$ given in (6.12). Since $X_{nT} \underset{T\to\infty}{\Rightarrow} X_n \underset{n\to\infty}{\Rightarrow} X$ sequentially, joint weak convergence $X_{nT} \Rightarrow X$ as $\left(n,T\right)\to\infty$ holds in the same manner as Theorem 3 with only minor definitional changes. First, Y_{iT} is integrable just as before. To show Lemma A(i) holds, observe that

$$\begin{aligned} \mathbb{E}\left\|Y_{iT}\right\|^2 &= \mathbb{E}Y_{1iT}^2 + \mathbb{E}Y_{2iT}^2 \\ &= \frac{1}{T}\mathbb{E}\left\{u_{iT}y_{iT-2} - \sum_{t=3}^{T} u_{it-1}\left(-\frac{\alpha_i c}{\sqrt{T}} + \frac{c}{\sqrt{T}}y_{it-3} + u_{it-2}\right)\right\}^2 \\ &\quad + \frac{1}{T^2}\mathbb{E}\left\{\sum_{t=2}^{T}\left(-\frac{\alpha_i c}{\sqrt{T}} + \frac{c}{\sqrt{T}}y_{it-2} + u_{it-1}\right) y_{it-2}\right\}^2 \end{aligned}$$

$$
\begin{aligned}
&= \frac{\sigma^2}{T}\mathbb{E}y_{iT-2}^2 + \frac{1}{T}\mathbb{E}\left\{\sum_{t=3}^{T} u_{it-1}\left(-\frac{\alpha_i c}{\sqrt{T}} + \frac{c}{\sqrt{T}}y_{it-3} + u_{it-2}\right)\right\}^2 \\
&= \frac{\sigma^2}{T}\mathbb{E}y_{iT-2}^2 + \frac{1}{T}\mathbb{E}\left(\sum_{t=3}^{T} u_{it-1}u_{it-2}\right)^2 + \frac{c^2\alpha_i^2}{T^2}\mathbb{E}\left(\sum_{t=3}^{T} u_{it-1}\right)^2 + \frac{c^2}{T^2}\mathbb{E}\left(\sum_{t=3}^{T} u_{it-1}y_{it-3}\right)^2 \\
&\quad -\frac{2\alpha_i c}{T^{3/2}}\mathbb{E}\left(\sum_{t=3}^{T} u_{it-1}\sum_{s=3}^{T} u_{is-2}\right) - \frac{2\alpha_i c^2}{T^2}\mathbb{E}\left(\sum_{t=3}^{T} u_{it-1}\sum_{s=3}^{T} y_{is-3}\right) + \frac{2c}{T^{3/2}}\mathbb{E}\left(\sum_{t=3}^{T} u_{it-1}u_{it-2}\sum_{s=3}^{T} y_{is-3}\right) \\
&= \frac{\sigma^4}{-2c}\frac{\sqrt{T_2}}{T} + \sigma^4\frac{T_2}{T} + O\left(\frac{1}{T}\right) + \frac{c^2\sigma^2}{T^2}\sum_{t=3}^{T}\mathbb{E}y_{it-3}^2 + O\left(\frac{1}{\sqrt{T}}\right) + O\left(\frac{1}{T}\right) + O\left(\frac{1}{\sqrt{T}}\right) \\
&= \sigma^4\frac{T_2}{T} + o(1),
\end{aligned}
$$

since from (6.24) $\mathbb{E}\left(y_{it}^2\right) = \sigma^2\frac{1-\rho^{2t}}{1-\rho^2} + \alpha_i^2\left(1-\rho^t\right)^2 = \sigma^2\frac{\sqrt{t}}{-2c}\left\{1+o(1)\right\}$. Then $\mathbb{E}\left\|Y_{iT}\right\|^2 < \infty$ and we deduce that

$$
\limsup_{n,T\to\infty}\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left\|Y_{iT}\right\| = \limsup_{T\to\infty}\mathbb{E}\left\|Y_{iT}\right\| \leq \limsup_{T\to\infty}\left(\mathbb{E}\left\|Y_{iT}\right\|^2\right)^{1/2} < \infty,
$$

as required. Condition (ii) holds, as we again have $\mathbb{E}Y_{iT} = \mathbb{E}Y_i = 0$; and condition (iii) and (iv) hold because $\sup_T \mathbb{E}\left\|Y_{iT}\right\|^2 < \infty$ and $\mathbb{E}\left\|Y_i\right\|^2 < \infty$. We then have joint weak convergence

$$
X_{nT} = n^{-1/2}\sum_{i=1}^{n} Y_{iT} \underset{n,T\to\infty}{\Rightarrow} X := \left(\sigma^2\zeta, \frac{\sigma^2}{2}\right),
$$

irrespective of the divergence rates of n and T to infinity. By continuous mapping, the required result follows for the GMM estimator so that $\sqrt{T}\left(\rho_{gmm}-\rho\right) \underset{n,T\to\infty}{\Rightarrow} N(0,4)$ holds jointly as $(n,T)\to\infty$ irrespective of the order and rates of divergence. ■

Proof of Theorem 7. We have $\rho = 1 + \frac{c}{T^\gamma}$ for some fixed $c<0$ and let $T\to\infty$. In this case, $y_{it} = -\frac{\alpha_i c}{T^\gamma} + \left(1+\frac{c}{T^\gamma}\right)y_{it-1} + u_{it}$ and $\Delta y_{it} = \rho\Delta y_{it-1} + \Delta u_{it}$ so that $\Delta y_{it} = -\frac{\alpha_i c}{T^\gamma} + \frac{c}{T^\gamma}y_{it-1} + u_{it}$. As before, we have

$$
\sqrt{T}\left(\rho_{gmm}-\rho\right) = \frac{\frac{1}{\sqrt{T}}\sum_{i=1}^{n}\left\{\left(u_{iT}y_{iT-2} - u_{i1}y_{i0}\right) - \sum_{t=3}^{T}u_{it-1}\left[-\frac{\alpha_i c}{T^\gamma} + \frac{c}{T^\gamma}y_{it-3} + u_{it-2}\right]\right\}}{\frac{1}{T}\sum_{i=1}^{n}\sum_{t=2}^{T}\left\{-\frac{\alpha_i c}{T^\gamma} + \frac{c}{T^\gamma}y_{it-2} + u_{it-1}\right\}y_{it-2}}. \tag{6.34}
$$

We use the following results from Phillips and Magdalinos (2007) and Magdalinos and Phillips (2009), which hold for all $\gamma \in (0,1)$,

$$T^{-1-\gamma}\sum_{t=2}^{T} y_{it}^2 \to_p \frac{\sigma^2}{-2c}, T^{-(1+\gamma)/2}\sum_{t=2}^{T} y_{it-1}u_{it} \Rightarrow N\left(0, \frac{\sigma^4}{-2c}\right), \text{ and } T^{-1/2-\gamma}\sum_{t=2}^{T} y_{it} = O_p(1). \tag{6.35}$$

Then, since $\frac{1}{\sqrt{T}}\sum_{t=2}^{T} u_{it-1}u_{it-2} \Rightarrow G_i \equiv N\left(0,\sigma^4\right)$, $\frac{1}{T^{1/2+\gamma}}\sum_{t=2}^{T} u_{it-1} = o_p(1)$, and $\frac{1}{T^{1/2+\gamma}}\sum_{t=2}^{T} u_{it-1}y_{it-3} = o_p(1)$ when $\gamma \in (0,1)$, the numerator of (6.34) is

$$\begin{aligned} &\sum_{i=1}^{n}\frac{1}{\sqrt{T}}\left\{u_{iT}y_{iT-2} - \sum_{t=3}^{T} u_{it-1}\left[-\frac{\alpha_i c}{T^\gamma} + \frac{c}{T^\gamma}y_{it-3} + u_{it-2}\right]\right\} \\ =\ & -\sum_{i=1}^{n}\frac{1}{\sqrt{T}}\sum_{t=3}^{T} u_{it-1}u_{it-2} + o_p(1) \Rightarrow -\sum_{i=1}^{n} G_i(1). \end{aligned}$$

Using (6.35), we find that the denominator of (6.34) satisfies

$$\sum_{i=1}^{n}\frac{1}{T}\sum_{t=2}^{T}\left\{-\frac{\alpha_i c}{T^\gamma} + \frac{c}{T^\gamma}y_{it-2} + u_{it-1}\right\}y_{it-2} \to_p \sum_{i=1}^{n}\left\{c\frac{\sigma^2}{-2c}\right\} = -\frac{\sigma^2 n}{2}.$$

Hence, as $T \to \infty$

$$\sqrt{T}\left(\rho_{gmm} - \rho\right) \underset{T\to\infty}{\Rightarrow} \frac{-\sum_{i=1}^{n} G_i}{-\frac{\sigma^2}{2}n} = \frac{\sigma^2\sum_{i=1}^{n}\zeta_i}{-\frac{\sigma^2}{2}n}, \quad \text{where } \zeta_i \sim_{iid} N(0,1).$$

Then, as $T \to \infty$ is followed by $n \to$, we have

$$\sqrt{nT}\left(\rho_{gmm} - \rho\right) \underset{T\to\infty}{\Rightarrow} \frac{2}{\sqrt{n}}\sum_{i=1}^{n}\zeta_i \underset{n\to\infty}{\Rightarrow} N(0,4), \text{ for all } \gamma \in (0,1).$$

Next consider the case $\gamma = 1$. The numerator of (6.34) is then

$$\begin{aligned} &\sum_{i=1}^{n}\frac{1}{\sqrt{T}}\left\{u_{iT}y_{iT-2} - \sum_{t=3}^{T} u_{it-1}\left[-\frac{\alpha_i c}{T} + \frac{c}{T}y_{it-3} + u_{it-2}\right]\right\} \\ =\ & \sum_{i=1}^{n}\frac{1}{\sqrt{T}}\left\{u_{iT}y_{iT-2} - \sum_{t=3}^{T} u_{it-1}u_{it-2}\right\} + o_p(1) \underset{T\to\infty}{\Rightarrow} u_{i\infty}K_{ci}(1) - \sum_{i=1}^{n} G_i(1), \end{aligned}$$

since by standard functional limit theory for near integrated processes (Phillips, 1987b) we have

$$\left(\frac{1}{T^{1/2}}y_{iT}, \frac{1}{T}\sum_{t=2}^{T} y_{it-1}u_{it}\right) \underset{T\to\infty}{\Rightarrow} \left(K_{ci}(r), \int_0^1 K_{ci} dB_i\right),$$

where $B_i(r) =: \sigma W_i(r)$ are *iid* Brownian motions with common variance σ^2, and $K_{ci}(r) = \int_0^r e^{c(r-s)} dB_i(s) =: \sigma J_{ci}(r)$ is a linear diffusion. The denominator of (6.34) satisfies

$$\sum_{i=1}^{n}\frac{1}{T}\sum_{t=2}^{T}\left\{-\frac{\alpha_i c}{T} + \frac{c}{T}y_{it-2} + u_{it-1}\right\} y_{it-2} \underset{T\to\infty}{\Rightarrow} \sum_{i=1}^{n}\left\{c\int_0^1 K_{ci}(r)^2 dr + \int_0^1 K_{ci} dB_i\right\}.$$

Hence

$$\sqrt{T}\left(\rho_{gmm} - \rho\right) \underset{T\to\infty}{\Rightarrow} \frac{\sum_{i=1}^{n}\left\{u_{i\infty}K_{ci}(1) - \sum_{i=1}^{n} G_i\right\}}{\sum_{i=1}^{n}\left\{c\int_0^1 K_{ci}(r)^2 dr + \int_0^1 K_{ci}(r) dB_i\right\}} = \frac{\sum_{i=1}^{n}\left\{\left(\sigma^{-1}u_{i\infty}\right) J_{ci}(1) - \zeta_i\right\}}{\sum_{i=1}^{n}\left\{c\int_0^1 J_{ci}(r)^2 dr + \int_0^1 J_{ci} dW_i\right\}}, \tag{6.36}$$

where the $\zeta_i \sim_{iid} N(0,1)$ and are independent of the W_i and $u_{i\infty}$ for all i. This gives the first part of (ii). Scaling the numerator and denominator of (6.36), noting that $\int_0^1 J_{ci}(r) dW_i$ has zero mean and finite variance, and using the independence of ζ_i, $u_{i\infty}$, and W_i, we obtain

$$\begin{aligned}\sqrt{nT}\left(\rho_{gmm} - \rho\right) &\underset{T\to\infty}{\Rightarrow} \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\left(\sigma^{-1}u_{i\infty}\right) J_{ci}(1) - \zeta_i\right\}}{\frac{1}{n}\sum_{i=1}^{n}\left\{c\int_0^1 J_{ci}(r)^2 dr + \int_0^1 J_{ci}(r) dW_i\right\}} \\ \underset{n\to\infty}{\Rightarrow} \frac{N\left(0, \frac{1-2c-e^{2c}}{-2c}\right)}{c\mathbb{E}\left(\int_0^1 J_{ci}(r)^2 dr\right)} &= N\left(0, -8c\frac{1-2c-e^{2c}}{\left(e^{2c}-1-2c\right)^2}\right),\end{aligned}$$

since, using results in Phillips (1987b), we have $\mathbb{E}\left(\int_0^1 J_{ci}(r)^2 dr\right) = \frac{e^{2c}-1-2c}{(2c)^2}$ and

$$\mathbb{E}\left\{\left(\sigma^{-1}u_{i\infty}\right) J_{ci}(1) - \zeta_i\right\}^2 = \mathbb{E}\left(\sigma^{-1}u_{i\infty}\right)^2 \mathbb{E}J_{ci}(1)^2 + \mathbb{E}\zeta_i^2 = 1 + \frac{1-e^{2c}}{-2c} = \frac{1-2c-e^{2c}}{-2c}.$$

Hence, when $\gamma = 1$, we have

$$\sqrt{nT}\left(\rho_{gmm} - \rho\right) \underset{(n,T)\to\infty}{\Rightarrow} N\left(0, (-8c)\frac{1-2c-e^{2c}}{\left(e^{2c}-1-2c\right)^2}\right) \tag{6.37}$$

From Lemma 2 of Phillips (1987b) we have

$$\left((-2c) \int_0^1 J_{ci}(r)^2 dr, (-2c)^{1/2} \int_0^1 J_{ci}(r) dW_i \right) \underset{c \to 0}{\Rightarrow} (1, Z_i), \quad Z_i \sim_{iid} N(0, 1,)$$

and $\frac{1-2c-e^{2c}}{-2c} = 2\{1 + o(1)\}$ as $c \to 0$, so that

$$(-8c) \frac{1 - 2c - e^{2c}}{(e^{2c} - 1 - 2c)^2} \sim (-8c) \frac{(-4c)}{\left\{ \frac{1}{2} (2c)^2 \right\}^2} = \frac{8}{c^2} \text{ for small } c \sim 0 \tag{6.38}$$

which explodes as $c \to 0$, consonant with the unit root case where we only have $\sqrt{T}$ convergence. Observe that both (6.37) and (6.38) correspond to earlier results with the reverse order of sequential convergence $(T, n)_{\text{seq}} \to \infty$.

Next suppose $\gamma > 1$ so that $\rho = 1 + \frac{c}{T^\gamma}$ is closer to unity than the LUR case with $\gamma = 1$. In this case, the numerator and denominator of (6.34) have the same limits as in the unit root case, viz.,

$$\begin{aligned} & \sum_{i=1}^n \frac{1}{\sqrt{T}} \left\{ u_{iT} y_{iT-2} - \sum_{t=3}^T u_{it-1} \left[-\frac{\alpha_i c}{T^\gamma} + \frac{c}{T^\gamma} y_{it-3} + u_{it-2} \right] \right\} \\ = & \sum_{i=1}^n \frac{1}{\sqrt{T}} \left\{ u_{iT} y_{iT-2} - \sum_{t=3}^T u_{it-1} u_{it-2} \right\} + o_p(1) \underset{T \to \infty}{\Rightarrow} \sum_{i=1}^n \{ u_{i\infty} B_i(1) - G_i \}, \end{aligned}$$

and

$$\sum_{i=1}^n \frac{1}{T} \sum_{t=2}^T \left\{ -\frac{\alpha_i c}{T^\gamma} + \frac{c}{T^\gamma} y_{it-2} + u_{it-1} \right\} y_{it-2} \underset{T \to \infty}{\Rightarrow} \sum_{i=1}^n \left\{ \int_0^1 B_i dB_i \right\}.$$

Then

$$\sqrt{T} \left(\rho_{gmm} - \rho \right) \underset{T \to \infty}{\Rightarrow} \frac{\sum_{i=1}^n \{ u_{i\infty} B_i(1) - G_i \}}{\sum_{i=1}^n \int_0^1 B_i dB_i} = \frac{\frac{1}{\sqrt{n}} \sum_{i=1}^n \left\{ \left(\sigma^{-1} u_{i\infty} \right) W_i(1) - \zeta_i \right\}}{\frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^1 W_i dW_i} \underset{n \to \infty}{\Rightarrow} 2\mathbb{C},$$

since $\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \left\{ \left(\sigma^{-1} u_{i\infty} \right) W_i(1) - \zeta_i \right\}, \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^1 W_i dW_i \right) \underset{n \to \infty}{\Rightarrow} N \left(0, \begin{bmatrix} 2 & 0 \\ 0 & 1/2 \end{bmatrix} \right).$

■

7 References

Anderson, T. W. and C. Hsiao (1981), "Estimation of dynamic models with error components", *Journal of the American Statistical Association,* 76, 598-606.

Anderson, T. W. and C. Hsiao (1982), "Formulation and estimation of dynamic models using panel data", *Journal of Econometrics*, 18, 47-82.

Blundell, R. and S. Bond (1998). "Initial conditions and moment restrictions in dynamic panel data models", *Journal of Econometrics*, 87, 115-143.

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, 1639–1657.

Han, C. and P. C. B. Phillips (2013). "First Difference MLE and Dynamic Panel Estimation", *Journal of Econometrics*, 175, 35–45.

Han, C., P. C. B. Phillips and D. Sul, (2014). "X-Differencing and Dynamic Panel Model Estimation," *Econometric Theory*, 30, pp 201-251.

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

Moon, H. R. and P. C. B. Phillips, (2004). "GMM Estimation of Autoregressive Roots Near Unity with Panel Data", *Econometrica*, 72, 467-522.

Phillips, P. C. B. (1987a). "Time Series Regression with a Unit Root," *Econometrica,* 55, 277–302.

Phillips, P. C. B. (1987b). "Towards a Unified Asymptotic Theory for Autoregression," *Biometrika* 74, 535–547.

Phillips, P. C. B. (1989). "Partially identified econometric models," *Econometric Theory* 5, 181–240.

Phillips, P. C. B. (2013). "Econometrics IV: Take Home Examination", Yale University (http://korora.econ.yale.edu/phillips/teach/ex/553a-ex13.pdf).

Phillips, P. C. B. and C. Han (2014). “On the Limit Distributions of FDLS and Anderson Hsiao Estimators”, working paper.

Phillips, P. C. B. and B. E. Hansen (1990). “Statistical inference in instrumental variables regression with I(1) processes,” *Review of Economic Studies* 57, 99–125.

Phillips, P. C. B., T. Magdalinos (2007), "Limit Theory for Moderate Deviations from a Unit Root," *Journal of Econometrics* 136, 115-130.

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

Qiu, Yinjia (Jeff). (2014), “Take Home Examination Solution to Phillips (2013)”.