GMM ESTIMATION FOR DYNAMIC PANELS WITH FIXED EFFECTS AND STRONG INSTRUMENTS AT UNITY

BY

Chirok Han and Peter C.B. Phillips

COWLES FOUNDATION PAPER NO. 1290

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

2010

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

Econometric Theory, **26**, 2010, 119–151.
doi:10.1017/S026646660909063X

GMM ESTIMATION FOR DYNAMIC PANELS WITH FIXED EFFECTS AND STRONG INSTRUMENTS AT UNITY

CHIROK HAN
Korea University

AND

PETER C.B. PHILLIPS
Yale University, University of Auckland, University of York, and Singapore Management University

This paper develops new estimation and inference procedures for dynamic panel data models with fixed effects and incidental trends. A simple consistent GMM estimation method is proposed that avoids the weak moment condition problem that is known to affect conventional GMM estimation when the autoregressive coefficient (ρ) is near unity. In both panel and time series cases, the estimator has standard Gaussian asymptotics for all values of $\rho \in (-1, 1]$ irrespective of how the composite cross-section and time series sample sizes pass to infinity. Simulations reveal that the estimator has little bias even in very small samples. The approach is applied to panel unit root testing.

1. INTRODUCTION

In simple dynamic panel models, it is well known that the usual fixed effects estimator is inconsistent when the time span is small (Nickell, 1981), as is the ordinary least squares (OLS) estimator based on first differences. In such cases, the instrumental variable (IV) estimator (Anderson and Hsiao, 1981) and generalized method of moments (GMM) estimator (Arellano and Bond, 1991) are both widely used. However, as noted by Blundell and Bond (1998), these estimators both suffer from a weak instrument problem when the dynamic panel autoregressive coefficient (ρ) approaches unity. When $\rho = 1$, the moment conditions are completely irrelevant for the true parameter ρ, and the nature of the behavior of the estimator depends on T. When T is small, the estimators are asymptotically random, and when T is large the unweighted GMM estimator may be inconsistent and the

Our thanks go to two referees for helpful comments on the original version. Phillips acknowledges partial support from a Kelly Fellowship and the NSF under Grant Nos. SES 04-142254 and SES 06-47086. Address correspondence to Peter C.B. Phillips, Cowles Foundation, Yale University, P.O. Box 208281, New Haven, CT 06520, USA; e-mail: peter.phillips@yale.edu; and Chirok Han, Department of Economics, Korea University, Anam-dong, Seongbuk-gu, Korea; e-mail: chirokhan@korea.ac.kr.

 0266-4666/10 $15.00

efficient two-step estimator (including the two-stage least squares estimator) may behave in a nonstandard manner. Some special cases of such situations are studied in Staiger and Stock (1997) and Stock and Wright (2000), among others, and Han and Phillips (2006), the latter in a general context that includes some panel cases.

Methods to avoid these problems were developed in Arellano and Bover (1995), Blundell and Bond (1998), and more recently in Hsiao, Pesaran, and Tahmiscioglu (2002). Arellano and Bover (1995) and Blundel and Bond (1998) propose a system GMM procedure that uses moment conditions based on the level equations together with the usual Arellano and Bond type orthogonality conditions. Hsiao et al. (2002), on the other hand, consider direct maximum likelihood estimation based on the differenced data under assumed normality for the idiosyncratic errors. Both approaches yield consistent estimators for all ρ values, but there are remaining issues that have yet to be determined in regard to the limit distribution when ρ is unity and T is large.

In a recent paper dealing with the time series case, Phillips and Han (2008) introduced a differencing-based estimator in an AR(1) model for which asymptotic Gaussian-based inference is valid for all values of $\rho \in (-1, 1]$. The present paper applies those ideas to dynamic panel data models, where we show that significant advantages occur. In panels, the estimator again has a standard Gaussian limit for all ρ values including unity, it has virtually no bias, and it completely avoids the usual weak instrument problem for ρ in the vicinity of unity.

As discussed later, this panel estimator makes use of moment conditions that are strong for all values of $\rho \in (-1, 1]$ under the assumption that the errors are white noise over time. (The white noise condition is stronger than that on which the usual IV/GMM approaches by Anderson and Hsiao, 1981, or Arellano and Bond, 1991, are based.) Under this condition, the proposed estimator is consistent, supports asymptotically valid Gaussian inference even with highly persistent panel data, and is free of initial conditions on levels. These advantages stem from the following properties: (i) The limit distribution is continuous as the autoregressive coefficient passes through unity; (ii) the rate of convergence is the same for stationary and nonstationary panels; and (iii) differencing transformations essentially eliminate dependence on level initial conditions.

Furthermore, there are no restrictions on the number of the cross-sectional units (n) and the time span (T) other than the simple requirement that $nT \to \infty$. Thus, neither large T nor large n is required for the limit theory to hold. Gaussian asymptotics apply irrespective of how the composite sample size $nT \to \infty$, including both fixed T and fixed n cases, as well as any diagonal path and relative rate of divergence for these sample dimensions. This robust feature of the asymptotics is unique to our approach and differs substantially from the existing literature, including recent contributions by Hahn and Kuersteiner (2002), Alvarez and Arellano (2003), and Moon, Perron, and Phillips (2005), who analyze various cases with large n and large T. Apart from the fact that the asymptotic variance of our proposed estimator can be better estimated by different methods when n is large and T is small (because the variance evolves with T), no other modification or consideration is required in the implementation of our approach, so it is well suited to practical implementation. This wide applicability does come at a cost in efficiency for the fixed effects model and a loss of power for the incidental trends model compared with existing methods. For example, the convergence rate $\sqrt{nT}$ is slower than the $\sqrt{n}T$ rate obtained in Levin, Lin, and Chu (2002) for a bias-corrected OLS estimator.

In what follows, Section 2 considers the model and estimator for a simple dynamic model with fixed effects, where the basic idea of our transformation is explained. Section 3 deals with a dynamic panel model where exogenous variables are present, and Section 4 studies the case with incidental trends. Section 5 applies the new approach to panel unit root testing. Section 6 contains some concluding remarks. Proofs are in Appendix A and additional formulas are given in Appendix B. Throughout the paper we define $0^0 = 1$ and use T_j to denote $\max(T - j, 0)$. We assume that data are observed for $t = 0, 1, \ldots, T$.

2. SIMPLE DYNAMIC PANELS

2.1. A New Estimator and Limit Theory

We consider the simple dynamic panel model

$$y_{it} = \alpha_i + u_{it}, \quad u_{it} = \rho u_{it-1} + \varepsilon_{it}, \quad \rho \in (-1, 1],$$

implying

$$y_{it} = (1-\rho)\alpha_i + \rho y_{it-1} + \varepsilon_{it}, \tag{1}$$

where α_i are unobservable individual effects and $\varepsilon_{it} \sim$ i.i.d. $(0, \sigma^2)$ with finite fourth moments.

This model differs slightly in its components form from the usual dynamic panel model $y_{it} = \alpha_i + \rho y_{it-1} + \varepsilon_{it}$ in that the individual effects disappear when $\rho = 1$. This formulation is made only to guarantee continuity in the model formulation and the asymptotics at $\rho = 1$. When $|\rho| < 1$ the two models are not distinguishable. Without this alternative parametrization, the data generating process has a discontinuity at $\rho = 1$, at which point the individual effects become individual trends.

As is well known, the OLS estimator based on the 'within' transformation yields an inconsistent estimator because the transformed regressor and the corresponding error are correlated—see Nickell (1981), among others. This bias is also not corrected by first differencing

$$\Delta y_{it} = \rho \Delta y_{it-1} + \Delta \varepsilon_{it}, \tag{2}$$

because the transformation induces a correlation between Δy_{it-1} and $\Delta \varepsilon_{it}$. Instead, following Phillips and Han (2008), we transform (2) further into the form

$$2\Delta y_{it} + \Delta y_{it-1} = \rho \Delta y_{it-1} + \eta_{it}, \quad \eta_{it} = 2\Delta y_{it} + (1-\rho)\Delta y_{it-1}. \tag{3}$$

Then, this formulation produces the following key moment conditions.[1]

LEMMA 1. *If* $E\varepsilon_{it}^2 = \sigma^2$ *for all* t *and* $E\varepsilon_{is}\varepsilon_{it} = 0$ *for all* $s \neq t$*, then*

$$Eg_{it}(\rho) = E\Delta y_{it-1}[2\Delta y_{it} + (1-\rho)\Delta y_{it-1}] = 0, \qquad t = 2, \ldots, T \tag{4}$$

for every $\rho \in (-1, 1]$.

It is worth noting that the moment conditions in (4) are based on the pattern of covariogram of $u_{it} = y_{it} - \alpha_i$ implied by the AR(1) assumption. In particular, when u_{it} is stationary, the key conditions are $Eu_{it}^2 = \sigma^2/(1-\rho^2)$ and $Eu_{it}u_{it+1} = \rho Eu_{it}^2$. More precisely, we have

$$\begin{aligned} g_{it}(\rho) &= \Delta y_{it-1}[2\Delta\varepsilon_{it} + (1+\rho)\Delta y_{it-1}] \\ &= -2y_{it-2}\Delta\varepsilon_{it} + 2y_{it-1}\varepsilon_{it} - 2y_{it-1}\varepsilon_{it-1} + (1+\rho)(\Delta y_{it-1})^2. \end{aligned}$$

When $|\rho| < 1$ and $Eu_{it}\alpha_i = 0$, the first two terms have zero mean under Ahn and Schmidt's (1995) "standard assumptions." But in order for the sum of the remaining two terms to have zero mean, it is required that $E(\Delta u_{it})^2 = 2\sigma^2/(1+\rho)$, which is implied by $Eu_{it}^2 = \sigma^2/(1-\rho^2)$ and $Eu_{it}u_{it-1} = \rho Eu_{it}^2$. These neither imply nor are implied by the assumptions of Ahn and Schmidt (1995).

The T_1 (i.e., $T-1$) moment conditions in (4) are strong for all $\rho \in (-1, 1]$ in the sense that the expected derivatives of the moment functions $g_{it}(\rho)$ differ from zero for all ρ as long as Δy_{it-1} has enough variation across i. This is easily verified by the calculation $E\partial g_{it}(\rho)/\partial\rho = -E(\Delta y_{it-1})^2$.

There are many ways to make use of these T_1 moment conditions. The simplest is to use pooled least squares estimation of (3), which leads to

$$\hat{\rho}_{ols} = \frac{\sum_{i=1}^n \sum_{t=2}^T \Delta y_{it-1}(2\Delta y_{it} + \Delta y_{it-1})}{\sum_{i=1}^n \sum_{t=2}^T (\Delta y_{it-1})^2},$$

which we call the *first difference least squares* (FDLS) estimator. This estimator has the following limit distribution.

THEOREM 1. *For each* T, $\sqrt{nT_1}(\hat{\rho}_{ols} - \rho) \Rightarrow N(0, V_{ols,T})$ *as* $n \to \infty$ *for all* $\rho \in (-1, 1]$*, where*

$$V_{ols,T} = \frac{ET_1^{-1}\left(\sum_{t=2}^T \Delta y_{it-1}\eta_{it}\right)^2}{\left[ET_1^{-1}\sum_{t=2}^T (\Delta y_{it-1})^2\right]^2}.$$

As $T \to \infty$, $V_{ols,T} \to 2(1+\rho)$*, and furthermore,* $\sqrt{nT_1}(\hat{\rho}_{ols} - \rho) \Rightarrow N(0, 2(1+\rho))$.

The Gaussian limit theory is valid for any n/T ratio as long as $nT_1 \to \infty$, including finite T values for which the variance $V_{ols,T}$ evolves with T. Most remarkably, the joint limit as both n and T pass to infinity is identical to the limit where $T \to \infty$ individually, or the sequential limit as $T \to \infty$ and then $n \to \infty$

or the sequential limit as $n \to \infty$ and then $T \to \infty$. As a result, the limit theory is remarkably robust to different sample size constellations of (n, T), and simulations support the resulting intuition that testing based on Theorem 2 should show little size distortion.

We remark that the fourth-moment condition $E\varepsilon_{it}^4 < \infty$ is required for the limit theory to hold. For small T, the variance $V_{ols,T}$ can be expressed directly in terms of the parameters, using the general formula given in (B.7) in Appendix B. For example, if $\varepsilon_{it} \sim N(0, \sigma^2)$ or more weakly $E\varepsilon_{it}^4 = 3\sigma^4$ and if $T = 2$, then we have $V_{ols,2} = (1+\rho)(3-\rho)$. For $T > 2$ (and fixed) the variances $V_{ols,T}$ are plotted in Figure 1. The expression is quite complicated for general ρ, and is unlikely to be practically useful because $V_{ols,T}$ depends on the nuisance fourth moment of ε_{it}, except for $\rho = 1$ (see Corollary 1 below), and because $V_{ols,T}$ can be readily estimated by just replacing the expectation operators with averaging over i and the error process η_{it} with the residuals $\hat{\eta}_{it}$ from the regression of (3). More specifically, when n is large, $V_{ols,T}$ is estimated by

$$\hat{V}_{ols,T} = \frac{(nT_1)^{-1}\sum_{i=1}^n \left(\sum_{t=2}^T \Delta y_{it-1}\hat{\eta}_{it}\right)^2}{\left[(nT_1)^{-1}\sum_{i=1}^n \sum_{t=2}^T (\Delta y_{it-1})^2\right]^2}, \tag{5}$$

where $\hat{\eta}_{it} = 2\Delta y_{it} + \Delta y_{it-1} - \hat{\rho}_{ols}\Delta y_{it-1}$. The corresponding standard error for $\hat{\rho}_{ols}$ is

$$se(\hat{\rho}_{ols}) = \left[\sum_{i=1}^n \left(\sum_{t=2}^T \Delta y_{it-1}\hat{\eta}_{it}\right)^2\right]^{1/2} \Bigg/ \sum_{i=1}^n \sum_{t=2}^T (\Delta y_{it-1})^2. \tag{6}$$

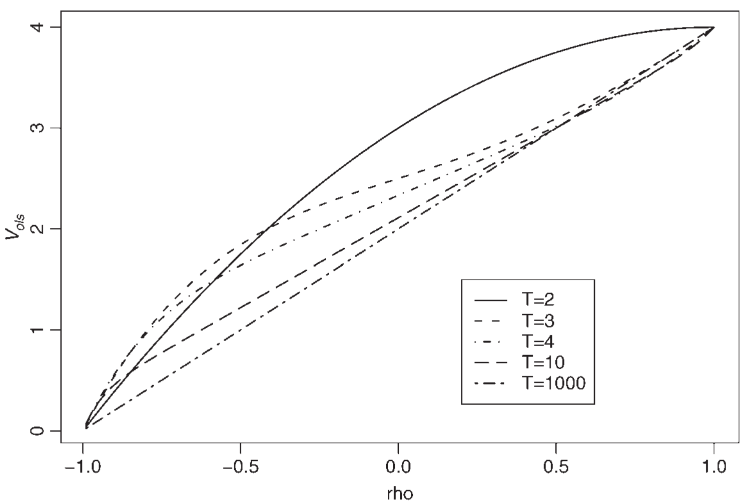

FIGURE 1. $V_{ols,T}$ for various T with normal errors. As $T \to \infty$, $V_{ols,T}$ approaches $2(1+\rho)$. The convergence is fast when $\rho > 0$.

As $T \to \infty$, both $V_{ols,T}$ and $\hat{V}_{ols,T}$ converge to $2(1+\rho)$, and the $N(0, 2(1+\rho))$ limit holds if $T \to \infty$ whether or not $n \to \infty$. Technically speaking, the joint limit as both $n \to \infty$ and $T \to \infty$ coincides with the sequential limit as $n \to \infty$ followed by $T \to \infty$ or the sequential limit as $T \to \infty$ followed by $n \to \infty$. In this case, both (5) and $2(1+\hat{\rho}_{ols})$ are consistent for $V_{ols,T}$, so either formula can be used. This undiscriminating feature is a characteristic of the new approach.

If n is small and T is large, then performance of (5) may be poor, as it relies on the law of large numbers across cross-sectional units. But in this case the $2(1+\hat{\rho}_{ols})$ formula approximates the actual variances quite well. This good performance of the asymptotic theory has been confirmed in earlier simulations reported in Phillips and Han (2008) for the time series case where $n = 1$.

When $\rho = 1$, the differenced data (Δy_{it}) are i.i.d. over both cross-sectional and time-series dimensions, resulting in the same Gaussian limit holding as $nT \to \infty$ (more precisely, $nT_1 \to \infty$) irrespective of the n/T ratio, as given in the following result:

COROLLARY 1. *If* $\rho = 1$, *then* $(nT_1)^{1/2}(\hat{\rho}_{ols} - 1) \Rightarrow N(0, 4)$ *as* $nT_1 \to \infty$.

This specific asymptotic distribution can also be derived from Bond et al. (2005).

Simulation results are given in Table 1 for $T = 2$ and $T = 24$. The exact form of $V_{ols,T}$ is provided in (B.7) in Appendix B. The simulated test size based on t-ratios with the standard errors obtained by (6) are given in the "size" columns. The results generally reflect the asymptotic theory well, including the consistency of both the estimator and the standard error. When n is relatively small, the standard errors according to (6) slightly underestimate the true variance. The asymptotic variance formula $2(1+\rho)$ for $\sqrt{nT_1}(\hat{\rho}_{ols} - \rho)$ from Theorem 1, which is appropriate when $T \to \infty$, obviously does not perform as well when T is as small as it is in this experiment, although the formula is surprisingly good when ρ is close to unity. Also, when $T \geq 3$, the formula $2(1+\rho)$ is rather close to the true variance for reasonably large ρ values (e.g., $\rho \geq 0.5$), at least if the errors are normally distributed. For error distributions with thicker tails, one might expect that larger values of T might be needed to get a good correspondence with the asymptotic formula based on $T \to \infty$.

Table 2 presents a brief comparison of FDLS, first-differenced GMM (DiffGMM), system GMM (SysGMM), and first-differenced MLE (DiffMLE) for $y_{it} = \alpha_i + u_{it}$ and $u_{it} = \rho u_{it-1} + \varepsilon_{it}$. We observe that the small sample properties of system GMM depend on the variance of α_i as noted by Hayakawa (2007). Both first-differenced MLE and FDLS are consistent, and obviously the MLE is more efficient than FDLS. (This efficiency gap becomes larger as T increases. For example, when $N = 50$ and $T = 10$ with $\rho = 0.5$, the variance of the FDLS estimator is approximately 2.5 times as large as that of the MLE, according to unreported simulations.) It is worth noting that the comparison of GMM estimates and FDLS should be interpreted with caution because they are based on different sets of moment conditions.

TABLE 1. FDLS for $\varepsilon_{it} \sim N(0,1)$. Simulations conducted using Gauss with 10,000 iterations. The limit variance $V_{ols,T}$ (denoted by V in this table) is calculated by (B.7) in Appendix B. The sizes of test based on the t-ratios using (6) are listed in the "size" columns.

	$T=2$								
	$\rho=0$ ($V=3$)			$\rho=-0.5$ ($V=1.75$)			$\rho=-0.9$ ($V=0.39$)		
n	mean	nT_1var	size	mean	nT_1var	size	mean	nT_1var	size
50	0.000	3.189	0.077	−0.499	1.857	0.076	−0.900	0.401	0.071
100	0.002	3.081	0.063	−0.500	1.822	0.063	−0.900	0.387	0.058
200	0.000	3.031	0.054	−0.500	1.734	0.054	−0.900	0.392	0.051
400	0.000	3.032	0.053	−0.499	1.779	0.054	−0.900	0.387	0.055
	$\rho=0.5$ ($V=3.75$)			$\rho=0.9$ ($V=3.99$)			$\rho=1$ ($V=4$)		
n	mean	nT_1var	size	mean	nT_1var	size	mean	nT_1var	size
50	0.498	3.968	0.075	0.901	4.049	0.067	1.001	4.066	0.068
100	0.500	3.808	0.064	0.900	4.041	0.061	1.001	3.980	0.058
200	0.501	3.823	0.057	0.899	3.962	0.055	1.000	4.047	0.057
400	0.500	3.846	0.054	0.900	3.930	0.052	1.001	4.098	0.058
	$T=24$								
	$\rho=0$ ($V=2.043$)			$\rho=-0.5$ ($V=1.074$)			$\rho=-0.9$ ($V=0.313$)		
n	mean	nT_1var	size	mean	nT_1var	size	mean	nT_1var	size
50	0.001	2.057	0.061	−0.498	1.091	0.062	−0.899	0.329	0.069
100	0.000	2.077	0.059	−0.499	1.095	0.056	−0.899	0.321	0.060
200	0.000	2.039	0.052	−0.500	1.053	0.049	−0.900	0.325	0.057
400	0.000	2.035	0.050	−0.500	1.079	0.051	−0.900	0.324	0.050
	$\rho=0.5$ ($V=2.987$)			$\rho=0.9$ ($V=3.758$)			$\rho=1$ ($V=4$)		
n	mean	nT_1var	size	mean	nT_1var	size	mean	nT_1var	size
50	0.500	2.986	0.059	0.900	3.735	0.059	1.000	3.988	0.059
100	0.500	3.036	0.054	0.900	3.723	0.053	1.000	4.039	0.057
200	0.500	3.036	0.056	0.900	3.685	0.050	1.000	3.841	0.045
400	0.500	3.073	0.054	0.900	3.817	0.053	1.000	4.071	0.055

As the moment functions are correlated over t, optimal GMM, which we call *first difference GMM* (FDGMM), is possibly a more efficient alternative to pooled OLS. The efficiency gain, however, looks marginal in our case. When $\rho = 1$, it can be shown that OLS equals the optimal GMM. For other ρ values, the variance ratio V_{gmm}/V_{ols} (where V_{gmm} and V_{ols} are the variances of FDGMM and FDOLS, respectively) is evaluated in Figure 2 in the case of standard normal errors, with $\rho = -0.5, 0, 0.5, 0.9, 1$ and $T = 2, 3, \ldots, 100$. The lowest variance ratio is

TABLE 2. Comparison of various estimates (10,000 iterations). DGP: $y_{it} = \alpha_i + u_{it}$, $u_{it} = \rho u_{it-1} + \varepsilon_{it}$, $\varepsilon_{it} \sim N(0,1)$, $\alpha_i \sim N(1, \sigma_\alpha^2)$; $\mathrm{sd}(u_{i,-1}) = (1-\rho^2)^{-1/2}\{|\rho| < 1\} + 25\{\rho = 1\}$, $N = 200$, $T = 3$.

	$\sigma_\alpha = 1, \sigma_\varepsilon = 1$			
ρ	DiffGMM	SysGMM	DiffMLE	FDLS
0.5	0.487 (0.155)	0.497 (0.090)	0.499 (0.079)	0.501 (0.089)
0.7	0.679 (0.188)	0.687 (0.093)	0.698 (0.082)	0.700 (0.093)
0.9	0.848 (0.294)	0.875 (0.107)	0.896 (0.083)	0.899 (0.097)
1.0	0.000 (0.940)	0.945 (0.138)	0.996 (0.082)	0.998 (0.100)
	$\sigma_\alpha = 5, \sigma_\varepsilon = 1$			
ρ	DiffGMM	SysGMM	DiffMLE	FDLS
0.5	0.448 (0.298)	0.507 (0.151)	0.499 (0.079)	0.501 (0.089)
0.7	0.572 (0.453)	0.668 (0.167)	0.698 (0.082)	0.700 (0.093)
0.9	0.579 (0.707)	0.816 (0.200)	0.896 (0.083)	0.899 (0.097)
1.0	−0.004 (0.931)	0.884 (0.237)	0.996 (0.082)	0.998 (0.100)

approximately 0.99, which is obtained at $\rho = 0.5$ and $T = 5$, indicating that the efficiency gain of optimal GMM over OLS is marginal. But note that this simulation result applies only to normally distributed errors. From additional experiments (not reported here) it was found that the efficiency of GMM over OLS is responsive to kurtosis, but for reasonable degrees of kurtosis, the efficiency gain of FDGMM remains marginal. For example, when $\mathrm{var}((\varepsilon_i/\sigma)^2) = 7$, the minimal V_{gmm}/V_{ols} ratio is approximately 0.98.

Because the performance of the feasible two-step GMM estimator may deteriorate due to inaccurate estimation of the covariance matrix, the two-step efficient GMM may yield a poorer estimator than OLS when the efficiency gain of the infeasible optimal GMM is marginal. When ε_{it} is normally distributed, this is likely to be the case. According to simulations not reported here, the two-step efficient GMM (using OLS as the first step estimator) looks less efficient than OLS for a wide range of ρ and T values up to quite a large n. So we generally recommend FDLS over FDGMM for practical use.

It is interesting to view FDLS in the context of method of moments and compare it with other consistent estimators. For the model $y_{it} = \alpha_i + u_{it}$, $u_{it} = \rho u_{it-1} + \varepsilon_{it}$, under the further assumption that α_i is uncorrelated with ε_{it} (and $u_{i,-1}$), we get the moments

$$Ey_{it}y_{is} = E\alpha_i^2 + \sigma^2\rho^{|t-s|}/(1-\rho^2), \quad |\rho| < 1,$$

$$Ey_{it}y_{is} = E\alpha_i^2 + Eu_{i,-1}^2 + \sigma^2(s \wedge t + 1), \quad \rho = 1,$$

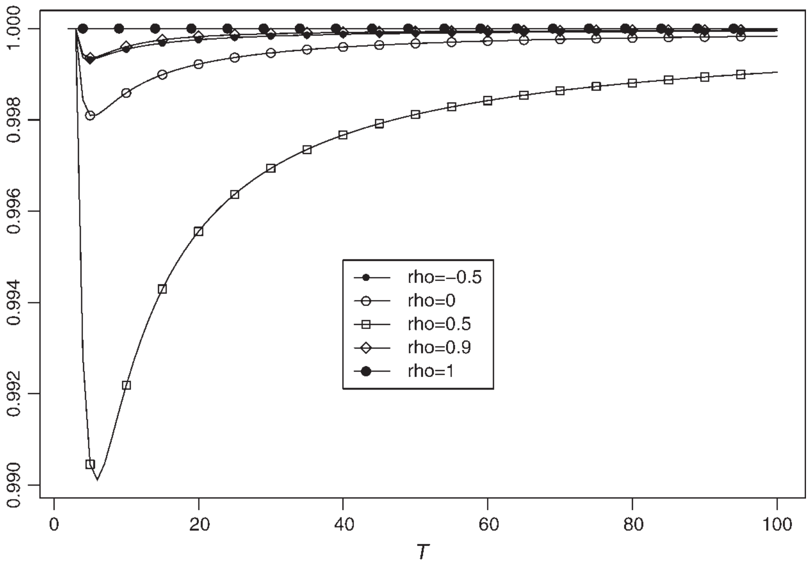

FIGURE 2. Variance ratio V_{gmm}/V_{ols} for normal errors for $T = 2, 3, \ldots, 100$. The minimum efficiency of FDLS relative to FDGMM is approximately 0.99, with the low point being attained at $\rho = 0.5$ and $T = 4$. The efficiency gain of FDGMM over FDLS is marginal.

for all $t, s = 0, 1, \ldots, T$, which in turn provide $(T+1)(T+2)/2$ distinct moments. Next, rewrite the moments in terms of $(y_{i0}, \Delta y_i')' = (y_{i0}, \Delta y_{i1}, \ldots, \Delta y_{iT})'$ as

$$Ey_{i0}^2 = E\alpha_i^2 + \rho^2 Eu_{i,-1}^2 + \sigma^2,$$

$$Ey_{i0}\Delta y_{it} = -\sigma^2\rho^{t-1}/(1+\rho)\cdot 1_{\{\rho<1\}}, \quad t \geq 1,$$

$$E(\Delta y_{it})^2 = 2\sigma^2/(1+\rho), \quad \text{and}$$

$$E\Delta y_{it}\Delta y_{is} = -\sigma^2\rho^{|t-s|-1}(1-\rho)/(1+\rho), \quad t \neq s,$$

or in matrix form as

$$E\begin{bmatrix} y_{i0} \\ \Delta y_i \end{bmatrix}\begin{bmatrix} y_{i0} \\ \Delta y_i \end{bmatrix}' = \frac{\sigma^2}{1+\rho}\begin{bmatrix} \xi_\rho & -1_{\{\rho<1\}} & -\rho_{\{\rho<1\}} & \cdots & -\rho_{\{\rho<1\}}^{T_1} \\ -1_{\{\rho<1\}} & 2 & -(1-\rho) & \cdots & -\rho^{T_2}(1-\rho) \\ -\rho_{\{\rho<1\}} & -(1-\rho) & 2 & \cdots & -\rho^{T_3}(1-\rho) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\rho_{\{\rho<1\}}^{T_1} & -\rho^{T_2}(1-\rho) & -\rho^{T_3}(1-\rho) & \cdots & 2 \end{bmatrix}, \quad (7)$$

where $\xi_\rho = (E\alpha_i^2 + \rho^2 E u_{i,-1}^2 + \sigma^2)(1+\rho)/\sigma^2$. (Note that the moments Ey_{i0}^2 and $E\Delta y_i y_{i0}$ depend on the condition that the α_i are uncorrelated with $u_{i,-1}$ and ε_{it} but the moments $E\Delta y_i \Delta y_i$ do not.) Rewriting the moments $Ey_i y_i'$ as (7) does not waste any information because we can recover the original moments $Ey_i y_i'$ by a linear transformation. Now, among these $(T+1)(T+2)/2$ distinct moments, Ey_{i0}^2 contains the nuisance parameters $E\alpha_i^2$ and $Eu_{i,-1}^2$ and, in fact, only Ey_{i0}^2 does so. Thus this element does not contribute to the estimation of ρ and can be safely ignored. Now, in what follows, we will show that each conventional (consistent) estimator can be derived as a (generalized) method of moments estimator using a subset of the above moment conditions.

Let us first consider the conventional IV/GMM estimators such as those of Anderson and Hsiao (1981) and Arellano and Bond (1991). The moment conditions

$$Ef_{i1t}(\rho) = Ey_{i0}(\Delta y_{it} - \rho \Delta y_{it-1}) = 0, \quad t \geq 2, \tag{8}$$

are obtained by combining any two consecutive elements of the first column (except for Ey_{i0}^2). Furthermore, any lower off-diagonal element of $E\Delta y_i \Delta y_i'$ and the element below it provides the moment condition

$$Eh_{ist}(\rho) = E\Delta y_{is}(\Delta y_{it} - \rho \Delta y_{it-1}) = 0, \quad s < t-1, \tag{9}$$

for some s and t. These moment conditions are strong when $\rho < 1$, so the IV/GMM estimators are consistent. But if $\rho \simeq 1$, then the moment conditions are weakly identifying, and in case $\rho = 1$, each moment condition in (8) and (9) fails to identify the true parameter because then $Ef_{i1t}(\rho) \equiv 0$ and $Eh_{ist}(\rho) \equiv 0$. When $\rho = 1$, if T is small and fixed, then the limit distribution of the GMM estimator is nondegenerate due to the lack of identification (see, e.g., Phillips, 1989; Staiger and Stock, 1997). But if T is large, then the unweighted GMM using these moment conditions converges to zero (under some regularity conditions for $u_{i,-1} = y_{i,-1} - \alpha_i$), because the unweighted criterion function satisfies the convergence

$$q_n^{-1} \left\{ \sum_{t=2}^{T} \left[n^{-1/2} \sum_{i=1}^{n} f_{i1t}(\rho) \right]^2 + \sum_{s<t-1} \left[n^{-1/2} \sum_{i=1}^{n} h_{ist}(\rho) \right]^2 \right\} \to_p (1+\rho^2)\sigma^4,$$

(if the initial condition that $(nT)^{-1/2} \sum_{i=1}^{n} u_{i,-1} \to_p 0$ holds under other regularity conditions) due to the accumulating signal variability and despite the fact that each moment condition fails to identify any ρ value, and this limit is minimized at zero (see Han and Phillips, 2006, for a detailed study of such situations), where $q_n = (T-1) + (T-1)(T-2)/2$ is the total number of moment conditions. So IV/GMM estimation based on (8) and (9) can hardly be used successfully when ρ may take values near unity. The behavior of the two-step efficient GMM estimator in this case has not been determined.

Unlike this IV/GMM estimator, which uses the off-diagonal elements of (7), our approach works from the moment conditions on the diagonal elements of $E\Delta y_i \Delta y_i'$. Each diagonal element of $E\Delta y_i \Delta y_i'$ in (7) and the element right below it construct a moment condition in (4), i.e., $(1-\rho)E(\Delta y_{it-1})^2 + 2E\Delta y_{it-1}\Delta y_{it} = 0$, or equivalently, $E\Delta y_{it-1}[2\Delta y_{it} + (1-\rho)\Delta y_{it-1}] = 0$. It is also interesting that more moment conditions can be obtained by combining the diagonal elements and their left elements, viz.,

$$Eg_{it}^*(\rho) = E\Delta y_{it}[2\Delta y_{it-1} + (1-\rho)\Delta y_{it}] = 0, \quad t = 2, \ldots, T. \tag{10}$$

These moment conditions are symmetric to (4) and are obtained by swapping the roles of Δy_{it} and Δy_{it-1}. We may consider GMM estimation or OLS estimation using these additional moment conditions together with those in (4). Simulations illustrate that the efficiency gain over $\hat{\rho}_{ols}$ by considering (10) is considerable when $T = 2$, especially for negative ρ, but the contribution of these additional moment conditions diminishes with T. Furthermore, when $\rho = 1$, $g_{it}(\rho)$ of (4) and $g_{it}^*(\rho)$ of (10) are asymptotically identical when evaluated at the true parameter, so the moment conditions are singular and the traditional feasible optimal GMM procedure does not provide standard asymptotics should we use both (4) and (10). In that case, procedures based on analytically calculated optimal weights (so the weights are a function of ρ) could avoid the singularity, but this sort of general analytic weighting scheme cannot be implemented because the optimal weighting matrix depends on the nuisance parameter $E\varepsilon_{it}^4$. When $\varepsilon_{it} \sim N(0,\sigma^2)$, however, the variances of $g_{it}(\rho)$ and $g_{it}^*(\rho)$ are almost equal except for $\rho \simeq -1$ according to simulations, implying that the unweighted sum $g_{it}(\rho) + g_{it}^*(\rho)$ is an (almost) optimal transformation of $g_{it}(\rho)$ and $g_{it}^*(\rho)$. Especially, when $\sum_{t=2}^T g_{it}(\rho)$ is used (as in the derivation of $\hat{\rho}_{ols}$) instead of each $g_{it}(\rho)$ separately, the unweighted sum $\sum_{t=2}^T [g_{it}(\rho) + g_{it}^*(\rho)]$ is an almost optimal exactly identifying moment condition with normal errors. This leads to a natural OLS regression of the pooled dependent variable $(2\Delta y_{it} + \Delta y_{it-1}, 2\Delta y_{it-1} + \Delta y_{it})'$ on the correspondingly pooled independent variable $(\Delta y_{it-1}, \Delta y_{it})'$. Let us denote this estimator by $\hat{\rho}_{ols}^{**}$. According to simulations, when $\rho = 0$ and $T = 2$, the variance ratio $\text{var}(\hat{\rho}_{ols}^{**})/\text{var}(\hat{\rho}_{ols})$ is about 0.75, meaning that by additionally using the moment condition that $Eg_{i2}^*(\rho) = 0$, we can reduce 25% of the variance. For other values of ρ and T, Table 3 reports the ratio of the variance of the OLS estimator based on $g_{it}(\rho)$ to the variance of the OLS estimator based on both $g_{it}(\rho)$ and $g_{it}^*(\rho)$. Note again that this result applies only to normal errors, and when the tails of the error distribution are thicker (e.g., the t_5 distribution), $g_{it}^*(\rho)$ will have larger variance than $g_{it}(\rho)$, and as a result, the estimator $\hat{\rho}_{ols}^{**}$ may be even less efficient than $\hat{\rho}_{ols}$ because of the failure of optimal weighting.

In summary, the loss of efficiency from using OLS rather than GMM is marginal, and the gain from adding $g_{it}^*(\rho)$ is not big enough compared with the possible risk of singularity (in case of GMM) and efficiency loss (in case of pooled OLS). In view of these many considerations, the original FDLS method (which yields $\hat{\rho}_{ols}$) is again recommended for practical use.

TABLE 3. Efficiency gain by adding $g_{it}^*(\rho)$. Variance ratios $\mathrm{var}(\hat{\rho}_{ols}^{**})/\mathrm{var}(\hat{\rho}_{ols})$ are reported.

$\rho \setminus T$	2	3	4	5	7	10	20
−0.5	0.44	0.47	0.56	0.62	0.73	0.79	0.90
0.0	0.75	0.78	0.86	0.88	0.92	0.95	0.98
0.5	0.93	0.95	0.97	0.98	0.99	0.99	1.00
0.9	0.99	1.00	1.00	1.00	1.00	1.00	1.00
1.0	1.00	1.00	1.00	1.00	1.00	1.00	1.00

2.2. The Explosive Case

When $\rho > 1$, the differences Δy_{it} continue to manifest explosive behavior and all the good properties of FDLS do not hold. The estimator is inconsistent and the limit depends on the n/T ratio and the initial status $u_{i,-1}$. The following lemma is indicative of what happens in this case.

LEMMA 2. *If* $\rho > 1$, *then*

$$E\frac{1}{T_1}\sum_{t=2}^{T}(\Delta y_{it-1})^2 = (\rho+1)^{-1}[2+\nu(\rho,T)]\sigma^2, \quad \text{and}$$

$$E\frac{1}{T_1}\sum_{t=2}^{T}\Delta y_{it-1}\eta_{it} = \nu(\rho,T)\sigma^2,$$

where

$$\nu(\rho,T) = \left[\frac{\rho^2(\rho^{2T_1}-1)\delta_\rho}{T_1(\rho+1)}\right], \qquad \delta_\rho = (\rho^2-1)Eu_{i,-1}^2/\sigma^2+1.$$

When T is fixed, this lemma implies, by the law of large numbers, that

$$\operatorname*{plim}_{n\to\infty}\hat{\rho}_{ols} = \rho + \frac{(\rho+1)\nu(\rho,T)}{2+\nu(\rho,T)}.$$

Thus, in the case where ρ is close to unity and $Eu_{i,-1}^2$ is small so that $\nu(\rho,T)$ is negligible, the inconsistency of $\hat{\rho}_{ols}$ is also small. For a given ρ, the bias of $\hat{\rho}_{ols}$ is bigger when T is large than when T is small. If ρ is even closer to unity and such that $\sqrt{n}\nu(\rho,T) \to 0$, then the bias of $\sqrt{nT_1}(\hat{\rho}_{ols}-\rho)$ is negligible, leading to a Gaussian limit whose variance changes continuously as ρ deviates slightly from unity into the explosive area. This observation implies that the limit distribution of Theorem 1 is continuous as ρ passes through unity to very mildly explosive cases.

The case with large T and fixed n can be analyzed by Theorems 2 and 3 of Phillips and Han (2008), who consider the time series case. Again, the asymptotics are continuous as ρ passes through unity under the initial condition that $(\rho_T - 1)u_{i,-1}^2 \to_p 0$ where $\rho = \rho_T \downarrow 1$.

If both n and T are large, the $N(0, 2(1+\rho))$ asymptotics are still continuous as ρ passes through the boundary of unity into the explosive area, though the boundary of ρ for the continuous asymptotics is then correspondingly narrower on the explosive side of unity.

2.3. Heterogeneity and Cross-Section Dependence

The remainder of this section considers issues of cross-sectional heteroskedasticity, heterogeneity, and cross-section dependence.

If the error variance $E\varepsilon_{it}^2$ is different across i, then Theorem 1 still applies as long as the Lindeberg condition holds, with the only modification to $V_{ols,T}$ being the more general expression

$$\lim_{n\to\infty} \frac{(nT_1)^{-1}\sum_{i=1}^n E(\sum_{t=2}^T \Delta y_{it-1}\eta_{it})^2}{\left[(nT_1)^{-1}\sum_{i=1}^n E\sum_{t=2}^T (\Delta y_{it-1})^2\right]^2}.$$

Computation of standard errors by (6) remains valid.

If the AR coefficient changes across i so that the model involves $u_{it} = \rho_i u_{it-1} + \varepsilon_{it}$, then we have the limit

$$\hat{\rho}_{ols} \to_p \operatorname*{plim}_{n\to\infty} \sum_{i=1}^n w_i \rho_i$$

if the limit on the right-hand side exists, where

$$w_i = \frac{\sum_{t=2}^T (\Delta y_{it-1})^2}{\sum_{i=1}^n \sum_{t=2}^T (\Delta y_{it-1})^2}.$$

Noting that $E(\Delta y_{it-1})^2 = 2\sigma_i^2/(1+\rho_i)$, we see that an individual unit with smaller ρ_i is given a bigger weight if there is no heterogeneity.

To allow for cross-section dependence, let $\varepsilon_{it} = \sum_{k=1}^K \lambda_{ik} f_{kt} + v_{it}$, where the f_{kt} are common shocks i.i.d. over t and independent of other shocks, λ_{ik} are the factor loadings, and the v_{it} are i.i.d. Since the compounded error ε_{it} is still white noise, the moment conditions (4) still hold. However, consistency does not hold unless $T \to \infty$ or $K \to \infty$ (where K is the number of common factors). The case $K = 1$ is exemplary. Here $(\Delta y_{it-1})^2$ involves $(\lambda_{i1} f_{1t})^2$, and the randomness in the double sum

$$(nT_1)^{-1} \sum_{i=1}^n \sum_{t=2}^T (\lambda_{i1} f_{1t})^2 = n^{-1} \sum_{i=1}^n \lambda_{i1}^2 \cdot T_1^{-1} \sum_{t=2}^T f_{1t}^2$$

persists unless T is large (cf. Phillips and Sul, 2007). When T is small, as is the case in typical microeconometric work, K may be considered large and each factor loading is assumed to be sufficiently small to ensure convergence. (Then the variation of the cross-section dependence term $\sum_{k=1}^{K} \lambda_{ik} f_{kt}$ is correspondingly controlled.) As long as the common factors f_{kt} are white noise over time, the moment conditions (4) hold, and under further regularity to ensure convergence, the FDLS estimator would be consistent. See Phillips and Sul (2007) for details on panel models with these characteristics.

3. DYNAMIC PANELS WITH EXOGENOUS VARIABLES

This section considers the model with exogenous variables $y_{it} = \alpha_i + \beta' x_{it} + u_{it}$, where $u_{it} = \rho u_{it-1} + \varepsilon_{it}$ for $\rho \in (-1, 1]$, which may be transformed to

$$y_{it} = (1-\rho)\alpha_i + \beta'(x_{it} - \rho x_{it-1}) + \rho y_{it-1} + \varepsilon_{it}. \tag{11}$$

Let $z_{it} = y_{it} - \beta' x_{it}$. Then the model (11) is written as $z_{it} = (1-\rho)\alpha_i + \rho z_{it-1} + \varepsilon_{it}$, which is the same as (1) with y_{it} replaced with z_{it}. By applying the same transformation, we get

$$2\Delta z_{it} + \Delta z_{it-1} = \rho \Delta z_{it-1} + \eta_{it}, \quad \eta_{it} = 2\Delta \varepsilon_{it} + (1+\rho)\Delta z_{it-1}.$$

For all ρ, Δz_{it-1} and η_{it} are uncorrelated and these moment conditions are strong for all ρ, as before. Next, if we allow α_i to be arbitrarily correlated with x_{it}, then we can apply the within transformation to (11), giving

$$\ddot{y}_{it} - \rho \dot{y}_{it-1} = (\ddot{x}_{it} - \rho \dot{x}_{it-1})'\beta + \ddot{\varepsilon}_{it},$$

where $\ddot{y}_{it} = y_{it} - T^{-1} \sum_{s=1}^{T} y_{is}$, $\dot{y}_{it-1} = y_{it-1} - T^{-1} \sum_{s=1}^{T} y_{is-1}$, and so on. From the fact that the within-group estimator is efficient when α_i is allowed to be correlated with x_{it} in the usual linear panel data model (see Im, Ahn, Schmidt, and Wooldridge, 1999), we propose the following exactly identifying moment conditions:

$$E \sum_{t=2}^{T} \Delta z_{it-1}[(2\Delta z_{it} + \Delta z_{it-1}) - \rho \Delta z_{it-1}] = 0, \tag{12}$$

$$E \sum_{t=1}^{T} (\ddot{x}_{it} - \rho \dot{x}_{it-1}) \left[(\ddot{y}_{it} - \rho \dot{y}_{it-1}) - (\ddot{x}_{it} - \rho \dot{x}_{it-1})'\beta\right] = 0, \tag{13}$$

where $\Delta z_{it} = \Delta y_{it} - \Delta x_{it}'\beta$. Note that when there is no exogenous variable, the first moment condition (12) leads to the FDLS estimator of Section 2. We can estimate ρ and β by the method of moments. In practice, after (12) and (13) are rewritten in terms of the sample moment conditions, the parameters can be estimated by estimating ρ and β iteratively between (12) and (13) if the procedure converges.

For the asymptotic variance, we can use the usual "$(D'\Omega^{-1}D)^{-1}$" formula, where D contains the expected scores of (12) and (13) and Ω is the variance matrix of those moment conditions, both evaluated at the true parameter. Conveniently, D is block diagonal, and if ε_{it} has zero third moment, then so is the Ω matrix, separating the estimation of ρ from the estimation of β. (See Appendix B for further details.) As a result, we can treat the final estimator $\hat{\beta}$ as the true β parameter in computing the Δz_{it}'s and then estimate ρ and compute its standard error; similarly, we can treat the final $\hat{\rho}$ as the true ρ parameter and compute the standard errors of $\hat{\beta}$ using usual within-group estimation technique after transforming x_{it} and y_{it} to $x_{it} - \hat{\rho}x_{it-1}$ and $y_{it} - \hat{\rho}y_{it-1}$, respectively.

4. INCIDENTAL TRENDS

When the model includes incidental trends so that $y_{it} = \alpha_i + \gamma_i t + u_{it}$, $u_{it} = \rho u_{it-1} + \varepsilon_{it}$, it may be written in the form

$$y_{it} = (1-\rho)\alpha_i + \rho\gamma_i + (1-\rho)\gamma_i t + \rho y_{it-1} + \varepsilon_{it}. \tag{14}$$

Double differencing eliminates the combined fixed effects, giving

$$\Delta^2 y_{it} = \rho\Delta^2 y_{it-1} + \Delta^2\varepsilon_{it}, \tag{15}$$

which implies that

$$\Delta^2 y_{it} = \sum_{j=0}^{\infty}\rho^j\Delta^2\varepsilon_{it-j} = \varepsilon_{it} - (2-\rho)\varepsilon_{it-1} + (1-\rho)^2\sum_{j=0}^{\infty}\rho^j\varepsilon_{it-j-2}.$$

When $\rho = 1$, we have $\Delta^2 y_{it} = \Delta\varepsilon_{it}$, so $E(\Delta^2 y_{it-1})^2 = 2\sigma^2$ and $E\Delta^2 y_{it-1}\Delta^2\varepsilon_{it} = -3\sigma^2$. When $|\rho| < 1$, we have

$$E(\Delta^2 y_{it-1})^2 = \left[1 + (2-\rho)^2 + \frac{(1-\rho)^4}{1-\rho^2}\right]\sigma^2 = \frac{2(3-\rho)\sigma^2}{1+\rho}$$

and

$$E\Delta^2 y_{it-1}\Delta^2\varepsilon_{it} = -(4-\rho)\sigma^2,$$

so these formulas cover the case of $\rho = 1$. Thus,

$$E\Delta^2 y_{it-1}\tilde{\eta}_{it} = 0, \quad \tilde{\eta}_{it} = 2\Delta^2\varepsilon_{it} + \frac{(1+\rho)(4-\rho)}{3-\rho}\Delta^2 y_{it-1}.$$

This corresponds to transforming (15) to the model in differences

$$2\Delta^2 y_{it} + \Delta^2 y_{it-1} = \theta\Delta^2 y_{it-1} + \tilde{\eta}_{it}, \quad \theta = -\frac{(1-\rho)^2}{3-\rho}. \tag{16}$$

Correspondingly, the *double difference least squares* (DDLS) estimator $\hat{\theta}$ is

$$\hat{\theta} = \frac{\sum_{i=1}^{n}\sum_{t=3}^{T}\Delta^2 y_{it-1}(2\Delta^2 y_{it}+\Delta^2 y_{it-1})}{\sum_{i=1}^{n}\sum_{t=3}^{T}(\Delta^2 y_{it-1})^2}.$$

Note that $\theta \in (-1,0]$ for all $\rho \in (-1,1]$ and $\theta = 0$ if $\rho = 1$. When point estimation of ρ is of interest, we can simply run pooled OLS on (16) to get $\hat{\theta}$ and censor at 0 and -1, and then recover $\hat{\rho}$ from $\hat{\theta}$ by $\hat{\rho} = \frac{1}{2}[2+\hat{\theta}-(\hat{\theta}^2-8\hat{\theta})^{1/2}]$. Alternatively, we may also consider method of moments estimation based on (16) using the moment condition

$$E\sum_{t=3}^{T}\Delta^2 y_{it-1}\left(2\Delta^2 y_{it}+\left[1+\frac{(1-\rho)^2}{3-\rho}\right]\Delta y_{it-1}\right)=0.$$

Some caution is needed here because the parameter $\theta = -(1-\rho)^2/(3-\rho)$ can never exceed zero for all $\rho \in (-1,1]$, and therefore the sample moment function may not attain zero at any parameter value.

The asymptotic distribution of the DDLS estimator for θ is given as follows.

THEOREM 2. *For each T, $\sqrt{nT_2}(\hat{\theta}-\theta) \Rightarrow N(0, W_{ols,T})$ for some positive $W_{ols,T}$. Furthermore, $W_{ols,T} \to W_{ols}$ for some $W_{ols} > 0$ as $T \to \infty$. Whether or not $n \to \infty$, $\sqrt{nT_2}(\hat{\theta}-\theta) \Rightarrow N(0, W_{ols})$.*

The variance $W_{ols,T}$ is presented in Appendix B for the special cases of $\rho = 0$ and $\rho = 1$ and is given in general form because of the complexity. An expression for W_{ols} is given in Theorem 4 of Phillips and Han (2008).

For testing, we can rely on this asymptotic distribution. Just as for the case without incidental trends, the limit theory is continuous and joint as $n \to \infty$ or $T \to \infty$ or both pass to infinity with $W_{ols,T}$ evolving with T. However, as is apparent from the relation $\theta = -(1-\rho)^2/(3-\rho)$, an $O\left(n^{-1/4}T^{-1/4}\right)$ neighborhood of $\rho = 1$ corresponds to an $O\left(n^{-1/2}T^{-1/2}\right)$ neighborhood of $\theta = 0$, and for $\rho = 1$, it is easily seen that the rate of convergence of $\hat{\rho}$ is at the slower $n^{1/4}T_2^{1/4}$ rate, while that of $\hat{\theta}$ is $n^{1/2}T_2^{1/2}$. For $\rho < 1$, the rates of convergence of $\hat{\rho}$ and $\hat{\theta}$ are both $n^{1/2}T_2^{1/2}$. Hence, there is a deficiency in the convergence rate for $\hat{\rho}$ around unity.

When T is small, $W_{ols,T}$ depends on $E\varepsilon_{it}^4$, and the algebraic form of $W_{ols,T}$ for small T is too complicated to be of interest. But when n is large, as in typical microeconometric projects, the standard error of $\hat{\theta}$ is easily calculated by

$$se(\hat{\theta}) = \frac{\sqrt{\sum_{i=1}^{n}(\sum_{t=3}^{T}\Delta^2 y_{it-1}\hat{\eta}_{it})^2}}{\sum_{i=1}^{n}\sum_{t=3}^{T}(\Delta^2 y_{it-1})^2}, \tag{17}$$

with $\hat{\eta}_{it}$ denoting the residuals from the regression of (16). On the other hand, in macroeconometrics, where T is often moderately large, then $\sqrt{nT_2}(\hat{\theta}-\theta) \Rightarrow$

$N(0, \lim_{T\to\infty} W_{ols,T})$, where an expression for $\lim_{T\to\infty} W_{ols,T}$ is given in Theorem 4 of Phillips and Han (2008). As presented in Theorems A.1 and A.2 in Appendix A, the convergence is uniform in the sense that the limit of the variance is the variance of the limit distribution as $T \to \infty$. Note that when $\rho < 1$, the asymptotic variance of $\hat{\rho}$ (indirectly obtained from $\hat{\theta}$) may be obtained using the delta method.

As a special case, if $\rho = 1$, then

$$\sqrt{nT_2}\hat{\theta} \Rightarrow N(0, 2+\kappa_4/2T_2) \tag{18}$$

for all T as $n \to \infty$, where $\kappa_4 = \text{var}(\varepsilon_{it}^2)/\sigma^4$ (this variance is calculated in Appendix B), and the limit distribution as $T \to \infty$ is simply $N(0,2)$ whether n is large or small.

5. PANEL UNIT ROOT TESTING

5.1. Fixed Effects Model

The inferential apparatus and its limit theory may be applied directly to panel unit root testing. Consider the fixed effects panel $y_{it} = (1-\rho)\alpha_i + \rho_i y_{it-1} + \varepsilon_{it}$ where the ε_{it} are i.i.d. The unit root null hypothesis is that $\rho_i = 1$ for all i. The OLS estimator $\hat{\rho}_{ols}$ and its limit theory in Theorem 1 and Corollary 1 can form the basis of a statistical test. More precisely, the test statistic is $\hat{\tau}_0 := (nT_1)^{1/2}(\hat{\rho}_{ols} - 1)/2$. This test statistic is derived under the assumption of cross-sectional homoskedasticity. If the variances $\sigma_i^2 := E\varepsilon_{it}^2$ differ across i and $\rho_i \equiv 1$, then the standard deviation of the limit distribution of $\hat{\rho}_{ols}$ is approximated by

$$\frac{2\left(\sum_{i=1}^n \sigma_i^4\right)^{1/2}}{T_1^{1/2}\sum_{i=1}^n \sigma_i^2} = \frac{2}{(nT_1)^{1/2}} \times \frac{\left(n^{-1}\sum_{i=1}^n \sigma_i^4\right)^{1/2}}{n^{-1}\sum_{i=1}^n \sigma_i^2}, \tag{19}$$

which is larger than the simple standard error form $2/(nT_1)^{1/2}$ obtained for homoskedastic data. Under heteroskedasticity, the $\hat{\tau}_0$ statistic can be computed by using formula (6) for the standard error, or by replacing σ_i^2 in (19) with the estimate $\hat{\sigma}_i^2 = T^{-1}\sum_{t=1}^T (\Delta y_{it})^2$. Under the null hypothesis, $\hat{\tau}_0 \Rightarrow N(0,1)$, and under the alternative hypothesis that $\rho_i < 1$ for some i (where the number of such i is a nonnegligible fraction of n), $\text{plim}\,\hat{\rho}_{ols} < 1$, and as a result, $\hat{\tau}_0 \to_p -\infty$ as $nT \to \infty$. So, the test is consistent for any passage of $nT \to \infty$.

Unlike Levin and Lin (1992) or Im, Pesaran, and Shin (1997), this test does not require any restriction on the path to infinity such as $T \to \infty$ and $n/T \to 0$. Only the composite divergence $nT \to \infty$ is required, and there is virtually no size distortion when T is small. On the other hand, while the point optimal test for a unit root has local power in a neighborhood of unity that shrinks at the rate $n^{-1/2}T^{-1}$ (see Moon and Perron, 2004; Moon, Perron, and Phillips, 2006b), the $\hat{\tau}_0$ test has only trivial asymptotic power in $n^{-1/2}T^{-1}$ neighborhoods and nontrivial local asymptotic power in $n^{-1/2}T^{-1/2}$ neighborhoods of unity. So in this model at least, there is an infinite power deficiency in neighborhoods of order $n^{-1/2}T^{-1}$.

This deficiency is partly due to the fact that $\hat{\rho}_{ols}$ depends only on differenced data, which thereby reduces the signal relative to that of point optimal and other tests that make direct use of the nonstationary regressor. But that is not the only reason, and, interestingly, a common point optimal test based on the differenced data may attain the optimal rate of $n^{1/2}T$. To illustrate this possibility, consider the common point likelihood ratio test for the case where the ε_{it} are i.i.d. $N(0,\sigma^2)$ and σ^2 is known. The test for $H_0 : n^{1/2}T(1-\rho) = 0$ against the alternative $H_1 : n^{1/2}T(1-\rho) = c$ using the differenced data Δy_{it} is based on the likelihood ratio

$$U_{nT}(c) = 2\left[\log L\left(1 - n^{-1/2}T^{-1}c\right) - \log L(1)\right], \tag{20}$$

where $\log L(\cdot)$ is calculated based on the joint distribution of $\Delta y_i = (\Delta y_{i1},\ldots, \Delta y_{nT})'$, i.e.,

$$\log L(\rho) = -\frac{nT}{2}\log(2\pi) - \frac{nT}{2}\log\sigma^2 - \frac{n}{2}\log|\Omega_T(\rho)| - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\Delta y_i'\Omega_T(\rho)^{-1}\Delta y_i,$$

with

$$\Omega_T(\rho) = (1+\rho)^{-1}\text{Toeplitz}\left\{2, -(1-\rho), -\rho(1-\rho), \ldots, -\rho^{T_2}(1-\rho)\right\}. \tag{21}$$

In the above, the notation Toeplitz$\{a_0,\ldots,a_{m-1}\}$ denotes the $m \times m$ symmetric Toeplitz matrix whose (i, j) element is $a_{|i-j|}$. As shown in the Appendix, when $n^{1/2}T(1-\rho) = c$ (i.e., under the alternative hypothesis), we have

$$U_{nT}(c) \to_d N(c^2/8, c^2/2). \tag{22}$$

So if the null hypothesis of $n^{1/2}T(1-\rho) = 0$ is tested against the alternative that $n^{1/2}T(1-\rho) = \tilde{c}$ such that the null hypothesis is rejected for $U_{nT}(\tilde{c}) \geq \bar{z}_\alpha c/\sqrt{2}$ where $\Phi(-\bar{z}_\alpha) = \alpha$, and if the $\tilde{c}$ happens to equal the true c, then the local power of this test is

$$\Phi\left(\frac{c}{4\sqrt{2}} - \bar{z}_\alpha\right).$$

For all $c > 0$, this local power function resides below the power envelope $\Phi(c/\sqrt{2} - \bar{z}_\alpha)$ based on the level data for large T obtained by Moon Perron, and Phillips (2006b). Nonetheless, it is remarkable that the optimal rate can be attained by using differenced data.

As discussed so far, the deficiency of testing using $\hat{\rho}_{ols}$ comes partly from differencing and partly from inefficient use of the moment conditions. This fact may at first seem to contradict our earlier observation that $\hat{\rho}_{ols}$ is almost as good as the (infeasible) optimal GMM estimator based on the differenced data (e.g.,

Figure 2). Nonetheless, it seems that maximum likelihood estimation on the differenced model combines the moment conditions so cleverly that at $\rho = 1$ (at which point the levels-based MLE is superconsistent), the otherwise useless moment conditions contribute to estimating $\rho = 1$. (See the last part of Section 2.3.) A full analysis and comparison of panel MLE in levels and differences that explores this issue will be useful and interesting, and deserves a separate research paper.

To return to testing based on the FDLS estimator, the deficiency in power based on this procedure may be interpreted as a cost arising from the simplicity of the $\hat{\tau}_0$ test, the uniform convergence rate of the estimator, and its robustness to the asymptotic expansion path for (n, T). Table 4 reports the simulated size and power of the $\hat{\tau}_0$ test, in comparison to Im et al.'s (1997) test, for the data generating process $y_{it} = (1-\rho_i)\alpha_i + \rho_i y_{it-1} + \sigma_i \varepsilon_{it}$ with alternative parameter settings, where α_i and ε_i are standard normal and $\sigma_i \sim U(0.5, 1.5)$. To simulate power, the cases $\rho_i \equiv 0.9$ and $\rho_i \sim U(0.9, 1)$ are considered. Panel length is chosen to be $T = 6$ and $T = 25$, choices that roughly illustrate size and power for small and moderate T. ($T = 6$ is the smallest value covered by Table 1 of Im et al. (2003).) Note that the $\hat{\tau}_0$ test does not require bias adjustment. It is also remarkable that the $\hat{\tau}_0$ test seems to have better power than the IPS test when T is small. But with larger T ($T = 25$ in the simulation), the IPS test has better power, which is related to the $O(\sqrt{nT})$ convergence rate of the FDLS estimator.

For panels with small T, Bond et al. (2005) report size and power for various tests such as OLS in levels, Breitung and Meyer's (1994) test, Harris and Tzavalis's (1999) test based on LSDV, a test based on Blundell and Bond's (1998) system GMM, a test using Hsiao et al.'s (2002) MLE in differences, the test based on FDLS (labeled "FD" in their paper), and others. See their paper for more on the comparison. Note that FDLS, system GMM, and MLE in differences are consistent both under the null and the alternative hypotheses, while others are based on estimators consistent only under the null hypothesis.

5.2. Incidental Trends Model

Next consider the case where incidental trends are present, as laid out in Section 4. Let $\hat{\theta}$ be the pooled OLS estimator from the regression of (16). Noting that $\rho = 1$ corresponds to $\theta = 0$, we can base the panel unit root test on the statistic $\hat{\tau}_1 := \hat{\theta}/se(\hat{\theta}) \Rightarrow N(0,1)$, where $se(\hat{\theta})$ is given in (17) when n is large or $se(\hat{\theta}) = \sqrt{2/nT_2}$ when T is large. (Again note that (17) is robust to the presence of cross-sectional heteroskedasticity.) The null hypothesis $H_0 : \rho_i = 1$ for all i is rejected if $\hat{\tau}_1$ is less than the left-tailed critical value from the standard normal distribution. When some ρ_i are smaller than unity (again with the number of such i comparable to n), or equivalently when some θ_i are smaller than 0, $\hat{\theta}$ converges in probability to the limit of $\sum_{i=1}^n \tilde{\omega}_i \theta_i$, where $\tilde{\omega}_i = \sum_{t=3}^T (\Delta^2 y_{it-1})^2 / \sum_{i=1}^n \sum_{t=3}^T (\Delta^2 y_{it-1})^2$. As long as a nonnegligible portion of individual units i have $\rho_i < 1$, this limit is strictly negative, and hence the test

TABLE 4. Simulated size and power of unit root tests with incidental intercepts case. DGP: $y_{it} = \alpha_i + u_{it}, \quad u_{it} = \rho_i u_{it-1} + \sigma_i \varepsilon_{it}$, $\alpha_i, \varepsilon_{it} \sim$ i.i.d. $N(0,1)$, $\sigma_i \sim$ i.i.d. $U(0.5, 1.5)$. Significance level: 5%

	$T=6$					
	$\rho_i \equiv 1$		$\rho_i \equiv 0.9$		$\rho_i \sim U(0.9,1)$	
n	HP	IPS	HP	IPS	HP	IPS
50	6.09	5.83	20.13	10.76	12.13	7.95
100	6.00	5.47	28.37	14.77	13.37	9.20
200	5.30	5.56	42.88	22.05	18.49	11.52
400	5.54	5.97	66.30	35.02	26.77	15.63
	$T=25$					
	$\rho_i \equiv 1$		$\rho_i \equiv 0.9$		$\rho_i \sim U(0.9,1)$	
n	HP	IPS	HP	IPS	HP	IPS
50	6.17	5.14	49.56	84.26	22.66	33.61
100	5.89	5.34	72.64	98.63	30.01	54.62
200	5.05	5.31	93.00	99.99	47.09	81.03
400	4.91	5.70	99.73	100.0	71.05	97.83

statistic $\hat{\tau}_1$ diverges to $-\infty$ in probability. Thus, the test $\hat{\tau}_1$ is consistent regardless of the existence of incidental trends.

Local power of the test is relatively weak, which can be explained by the definition of θ in (16). Since $\hat{\theta}$ has an $(nT_2)^{1/2}$ rate of convergence, the test should have some local power when θ is in an $O(n^{-1/2}T_2^{-1/2})$ neighborhood of zero. But that is the case when ρ is in an $O(n^{-1/4}T_2^{-1/4})$ neighborhood of unity, which shrinks at a far slower rate than the optimal $n^{-1/4}T^{-1}$ rate attained by a point optimal test when the incidental trends are extracted from the panel data as described in Moon, Perron, and Phillips (2006b).

Interestingly there may exist a unit root test based on the double-differenced data that has local power in a neighborhood of unity shrinking at the $n^{-1/2}T_1^{-1/2}$ rate when $T_1/\sqrt{n} \to 0$ and at a correspondingly faster rate when T grows faster. (An earlier version of this paper, Han and Phillips, 2007, contains an analysis regarding this point.) Compared to large-T point optimal tests (such as Ploberger and Phillips, 2002) which have nontrivial power under the $O(n^{-1/4}T^{-1})$ local alternatives, this $\sqrt{nT}$ rate would imply the possible existence of more efficient estimators using double-differenced data for panel data with large N and small T. This interesting issue presents a major challenge for future research.

6. CONCLUSION

This paper develops a simple GMM estimator for dynamic panel data models, which is largely free from bias as the AR coefficient approaches unity, and which yields standard Gaussian asymptotics for all values of ρ and without any discontinuity at unity. The limit theory is also robust in the sense that it performs well under all possible passages to infinity, including $n \to \infty$, $T \to \infty$, and all diagonal paths. The method also extends in a straightforward manner to cases with exogenous variables, cross-section dependence, and incidental trends.

The approach leads to standard Gaussian panel unit root tests. These tests do not suffer from size distortion regardless of the n/T ratio. Illustration of power properties of some infeasible likelihood ratio tests indicates that the optimal convergence rate can perhaps be achieved using (double) differenced data while at the same time preserving standard Gaussian limits. Such tests can be expected to be particularly useful when n is large and T is small or moderate, and to outperform existing point optimal tests and exceed the usual power envelope for such sample size configurations. Extension of the present line of research in this direction is a major challenge and is left for future work.

NOTE

1. These moment conditions are also found in Kruiniger (2007). Our approach is different from his in that we use only globally strong (strong at all ρ values) moment conditions and combine them to form a least squares estimator, which yields simple asymptotics. Bond, Nauges, and Windmeijer (2005) consider this least squares estimator for micro-panel unit root testing.

REFERENCES

Ahn, S.C. & P. Schmidt (1995) Efficient estimation of models for dynamic panel data. *Journal of Econometrics* 68, 5–27.

Alvarez, J. & M. Arellano (2003) The time series and cross-section asymptotics of dynamic panel data estimators. *Econometrica* 71(4), 1121–1159.

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

Arellano, M. & S. Bond (1991) Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *Review of Economic Studies* 58, 277–297.

Arellano, M. & O. Bover (1995) Another look at the instrumental variable estimation of error-component models. *Journal of Econometrics* 68, 29–51.

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

Bond, S., C. Nauges, & F. Windmeijer (2005) Unit Roots: Identification and Testing in Micro Panels. IFS Working paper CWP07/05.

Hahn, J. & 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.

Han, C. (2007) Determinants of covariance matrices of differenced AR(1) processes. *Econometric Theory* 23, 1248–1253.

Han, C. & P.C.B. Phillips (2006) GMM with many moment conditions. *Econometrica* 74, 147–192.

Han, C. & P.C.B. Phillips (2007) GMM Estimation for Dynamic Panels with Fixed Effects and Strong Instruments at Unity. Mimeo, University of Auckland.

Hayakawa, K. (2007) Small sample bias properties of the system GMM estimator in dynamic panel data models. *Economics Letters* 95, 32–78.
Hsiao, C., M.H. Pesaran, & A.K. Tahmiscioglu (2002) Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods. *Journal of Econometrics* 109, 107–150.
Im, K.S., S.C. Ahn, P. Schmidt, & J. Wooldridge (1999) Efficient estimation of panel data models with strictly exogenous explanatory variables—A Monte Carlo study. *Journal of Econometrics* 93(1), 177–201.
Im, K.S., M.H. Pesaran, & Y. Shin (1997) Testing for Unit Roots in Heterogeneous Panels. Working paper, University of Cambridge.
Kallenberg, O. (2002) *Foundations of Modern Probability*, 2nd ed. Springer.
Kruiniger, H. (2007) An efficient linear GMM estimator for the covariance stationary AR(1)/unit root model for panel data. *Econometric Theory* 23, 519–535.
Levin, A. & C.F. Lin (1992) Unit Root Test in Panel Data: Asymptotic and Finite Sample Properties. Discussion paper #92-93, University of California at San Diego.
Levin, A., C.F. Lin, & C.S.J. Chu (2002). Unit root tests in panel data: Asymptotic and finite-sample properties. *Journal of Econometrics* 108, 1–24.
Moon, H.R. & B. Perron (2004) Asymptotic Local Power of Pooled t-Ratio Tests for Unit Roots in Panels with Fixed Effects. Mimeo, University of Southern California.
Moon, H.R., B. Perron, & P.C.B. Phillips (2006a) On the Breitung test for panel unit roots and local asymptotic power. *Econometric Theory* 22, 1179–1190.
Moon, H.R., B. Perron, & P.C.B. Phillips (2006b) Incidental Trends and the Power of Panel Unit Root Tests. Working paper.
Nickell, S. (1981) Biases in dynamic models with fixed effects. *Econometrica* 49, 1417–1426.
Phillips, P.C.B. (1989) Partially identified econometric models. *Econometric Theory* 5, 181–240.
Phillips, P.C.B. & C. Han (2008). Gaussian inference in AR(1) times series with or without unit root. *Econometric Theory* 24, 631–650.
Phillips, P.C.B. & H.R. Moon (1999) Linear regression limit theory for nonstationary panel data. *Econometrica* 67(5), 1057–1111.
Phillips, P.C.B. & V. Solo (1992) Asymptotics for linear processes. *Annals of Statistics* 20(2), 971–1001.
Phillips, P.C.B. & D. Sul (2007) Bias in dynamic panel estimation with fixed effects, incidental trends and cross section dependence. *Journal of Econometrics* 137, 162–188.
Ploberger, W. & P.C.B. Phillips (2002) Optimal Testing for Unit Roots in Panel Data. Mimeo, Yale University.
Staiger, D. & J.H. Stock (1997) Instrumental variables regression with weak instruments. *Econometrica* 65(3), 557–586.
Stock, J.H. & J.H. Wright (2000) GMM with weak identification. *Econometrica* 68(5), 1055–1096.

APPENDIX A: Proofs

The assumed model in the fixed effects case is $y_{it} = \alpha_i + u_{it}$, $u_{it} = \rho u_{it-1} + \varepsilon_{it}$, where $\varepsilon_{it} \sim$ i.i.d. $(0, \sigma^2)$. Obviously we can express, for all $\rho \in (-1, 1]$,

$$\Delta y_{it-1} = \sum_{j=0}^{\infty} \rho^j \Delta\varepsilon_{it-1-j} = \varepsilon_{it-1} - (1-\rho) \sum_{j=1}^{\infty} \rho^{j-1} \varepsilon_{it-j-1}, \tag{A.1}$$

and

$$\eta_{it} = 2\Delta\varepsilon_{it} + (1+\rho)\Delta y_{it-1} = 2\varepsilon_{it} - (1-\rho)\varepsilon_{it-1} - (1-\rho^2) \sum_{j=1}^{\infty} \rho^{j-1} \varepsilon_{it-j-1}, \tag{A.2}$$

where $0 \cdot \pm\infty = 0$.

Proof of Lemma 1. If $\rho = 1$, then $\Delta y_{it} = \varepsilon_{it}$, and $\eta_{it} = 2\varepsilon_{it}$. So we have $E\Delta y_{it-1}\eta_{it} = 2E\varepsilon_{it-1}\varepsilon_{it} = 0$. If $|\rho| < 1$, then $E\Delta y_{it-1}\eta_{it} = -(1-\rho)\sigma^2 + (1-\rho)(1-\rho^2)(1-\rho^2)^{-1}\sigma^2 = 0$. ■

Next, we present some lemmas that are useful in proving the theorems (Theorem 1 in particular). Let $T_m = \max(T-m, 0)$. Let $X_{it} = \sum_0^\infty c_j \varepsilon_{it-j}$ and $Y_{it} = \sum_0^\infty d_j \varepsilon_{it-j}$ where $\varepsilon_{it} \sim$ i.i.d. $(0, \sigma^2)$. We will frequently assume that

$$\sum_0^\infty (|c_s| + |d_s|) < \infty \quad \text{and} \tag{A.3}$$

$$\sum_0^\infty s(c_s^2 + d_s^2) < \infty. \tag{A.4}$$

THEOREM A.1 *If $E\varepsilon_{it}^2 < \infty$, then under (A.4),*

$$\frac{1}{nT_m}\sum_{i=1}^n \sum_{t=m+1}^T X_{it}^2 \to_p \sigma^2 \sum_0^\infty c_j^2$$

for any small m (e.g., 2 or 3), as $nT_m \to \infty$.

Of course, the part of condition (A.4) related to d_s is not relevant in this case.

Proof. If T is fixed and $n \to \infty$, then by Khinchine's law of large numbers,

$$\frac{1}{T_m}\sum_{t=m+1}^T \left[\frac{1}{n}\sum_{i=1}^n X_{it}^2\right] \to_p \frac{1}{T_m}\sum_{t=m+1}^T EX_{it}^2 = EX_{it}^2 = \sigma^2 \sum_0^\infty c_j^2,$$

where the first equality on the right hand side holds because of stationarity. If n is fixed and $T \to \infty$, then the result follows from the convergence $T_m^{-1}\sum_{t=m+1}^T X_{it}^2 \to_p \sigma^2 \sum_0^\infty c_j^2$ for each i under the stated conditions (Thm. 3.7 of Phillips and Solo, 1992) and the cross-sectional i.i.d. assumption. If both n and T increase to infinity, by Theorem 1 of Phillips and Moon (1999), it suffices to show that (a) $\limsup_{n,T} n^{-1}\sum_{i=1}^n EZ_{iT}\{Z_{iT} > n\delta\} = 0$ for all $\delta > 0$ where $Z_{iT} = T_m^{-1}\sum_{t=m+1}^T X_{it}^2$ (a notation that is used only in this proof), because all other conditions in Phillips and Moon's (1999) theorem are obviously satisfied. Because the Z_{iT} are i.i.d. across i, condition (a) is equivalent to $EZ_{1T}\{Z_{1T} > n\delta\} = 0$ for all $\delta > 0$, which is implied by the uniform integrability of Z_{1T} (over T). Since $Z_{1T} \to_p \sigma^2 \sum_0^\infty c_j^2$, the uniform integrability of Z_{1T} (over T) is equivalent to the convergence $EZ_{1T} \to \sigma^2 \sum_0^\infty c_j^2$, which is obviously true because $EZ_{1T} = \sigma^2 \sum_0^\infty c_j^2$ for all T. ■

Next, we establish a panel CLT for the sample covariance of X_{it} and Y_{it}. We assume that X_{it} and Y_{it} are uncorrelated by imposing the condition

$$\sum_0^\infty c_j d_j = 0. \tag{A.5}$$

THEOREM A.2 *Let $\Pi_{j,r} = c_j d_{j+r} + c_{j+r} d_j$. If $E\varepsilon_{it}^4 < \infty$, then under (A.3), (A.4), and (A.5), for any fixed T and small m (e.g., 2 or 3),*

$$U_{nT} := (nT_m)^{-1/2}\sum_{i=1}^n \sum_{t=m+1}^T X_{it}Y_{it} \Rightarrow N(0, V_T) \tag{A.6}$$

as $n \to \infty$, where

$$V_T = A_T \operatorname{var}(\varepsilon_{it}^2) + B_T \sigma^4,$$

with

$$A_T = \sum_0^{\infty} c_j^2 d_j^2 + \frac{2}{T_m} \sum_{t=m+2}^{T} \sum_{r=1}^{t-m-1} \sum_{j=0}^{\infty} c_j d_j c_{j+r} d_{j+r}, \tag{A.7}$$

$$B_T = \sum_{j=0}^{\infty} \sum_{r=1}^{\infty} \Pi_{j,r}^2 + \frac{2}{T_m} \sum_{t=m+2}^{T} \sum_{k=1}^{t-m-1} \sum_{j=0}^{\infty} \sum_{r=1}^{\infty} \Pi_{j,r} \Pi_{j+k,r}. \tag{A.8}$$

Furthermore,

$$V_T \to \sigma^4 B = \sigma^4 \sum_{r=1}^{\infty} \left(\sum_{j=0}^{\infty} \Pi_{j,r} \right)^2 \tag{A.9}$$

as $T \to \infty$. Whether or not $n \to \infty$, $U_{nT} \Rightarrow N(0, \sigma^4 B)$ as $T \to \infty$.

Proof. When T is small, (A.6) follows from the central limit law for i.i.d. variates because fourth moments are finite. The variance V_T is computed as follows. Since

$$X_{it} Y_{it} = \sum_{j=0}^{\infty} c_j d_j \varepsilon_{it-j}^2 + \sum_{j=0}^{\infty} \sum_{r=1}^{\infty} \Pi_{j,r} \varepsilon_{it-j} \varepsilon_{it-j-r}, \tag{A.10}$$

we have

$$C_0 = \operatorname{var}(X_{it} Y_{it}) = \left(\sum_{j=0}^{\infty} c_j^2 d_j^2 \right) \operatorname{var}(\varepsilon_{it}^2) + \left(\sum_{j=0}^{\infty} \sum_{r=1}^{\infty} \Pi_{j,r}^2 \right) \sigma^4. \tag{A.11}$$

Also

$$X_{it-k} Y_{it-k} = \sum_{j=0}^{\infty} c_j d_j \varepsilon_{it-k-j}^2 + \sum_{j=0}^{\infty} \sum_{r=1}^{\infty} \Pi_{j,r} \varepsilon_{it-k-j} \varepsilon_{it-k-j-r}, \tag{A.12}$$

and from (A.10) and (A.12), $C_k := \operatorname{cov}(X_{it} Y_{it}, X_{it-k} Y_{it-k})$ is

$$C_k = \left(\sum_{j=0}^{\infty} c_j d_j c_{j+k} d_{j+k} \right) \operatorname{var}(\varepsilon_{it}^2) + \left(\sum_{j=0}^{\infty} \sum_{r=1}^{\infty} \Pi_{j,r} \Pi_{j+k,r} \right) \sigma^4.$$

Now

$$T_m^{-1} \operatorname{var}\left(\sum_{t=m+1}^{T} X_{it} Y_{it} \right) = C_0 + \frac{2}{T_m} \sum_{t=m+2}^{T} \sum_{k=1}^{t-m-1} C_k = A_T \operatorname{var}(\varepsilon_{it}^2) + B_T \sigma^4, \tag{A.13}$$

as stated. Lemmas A.3 and A.4 below, respectively, show that $A_T \to 0$ and $B_T \to B$ under (A.3) and (A.4), and thus the convergence (A.9) is obvious from (A.13).

Now we prove the central limit theory as $T \to \infty$. If n is fixed, then the result follows from Theorem 6 of Phillips and Han (2008), implied by (A.4) and the finiteness of the second moments. For the case where both n and T increase, we will apply the Lindeberg central limit theorem to the row-wise independent array $\{n^{-1/2}W_{Ti}\}$, where $W_{Ti} = T_m^{-1/2}\sum_{t=m+1}^{T} X_{it}Y_{it}$ (notation that is used only in this proof). The Lindeberg condition is

$$\frac{1}{n}\sum_{i=1}^{n} EW_{Ti}^2\{W_{Ti}^2 > n\epsilon\} \to 0, \quad \text{for all } \epsilon > 0 \tag{A.14}$$

(e.g., Kallenberg, 2002, Thm. 5.12). Because the random variables are i.i.d. across i, (A.14) reduces to $EW_{T1}^2\{W_{T1}^2 > n\epsilon\} \to 0$ for all $\epsilon > 0$ (note that T depends on n), which is again implied by

$$\sup_T EW_{T1}^2\{W_{T1}^2 > n\epsilon\} \to 0 \quad \text{for all} \quad \epsilon > 0.$$

This last condition holds if W_{T1}^2 (as a sequence indexed by T) is uniformly integrable over T. When a positive random variable converges in distribution, uniform convergence is equivalent to convergence of the means (Kallenberg, 2002, Lem. 4.11). In the case of W_{T1}^2, we have $W_{T1} \to_d W_1 \sim N(0, \sigma^4 B)$ by Theorem 6 of Phillips and Han (2008), and by the continuous mapping theorem, $W_{T1}^2 \to_d W_1^2$. But by the first part of the theorem,

$$EW_{T1}^2 = A_T \text{var}(\varepsilon_{it}^2) + \sigma^4 B_T \to \sigma^4 B_T = EW_1^2,$$

and the joint limit theory follows straightforwardly. ■

The next two lemmas, respectively, show that $A_T \to 0$ and $B_T \to B$ as $T \to \infty$, as indicated above.

LEMMA A.3. *Under (A.4) and (A.5),* $\lim_{T\to\infty} A_T = 0$.

Proof. Let $f_j = c_j d_j$ (notation used only in this proof). For any sequence $\{a_r\}$,

$$\sum_{t=m+2}^{T}\sum_{r=1}^{t-m-1} a_r = \sum_{s=1}^{T_m-1}(T_m - s)a_s, \tag{A.15}$$

and thus we have

$$\begin{aligned}
A_T &= \sum_0^{\infty} f_j^2 + \frac{2}{T_m}\sum_{j=0}^{\infty}\sum_{s=1}^{T_m-1}(T_m - s) f_j f_{j+s} \\
&= \left[\sum_0^{\infty} f_j^2 + 2\sum_{j=0}^{\infty}\sum_{s=1}^{T_m-1} f_j f_{j+s}\right] - 2\left[\frac{1}{T_m}\sum_{j=0}^{\infty}\sum_{s=1}^{T_m-1} s f_j f_{j+s}\right] \\
&= \left[\sum_0^{\infty} f_j^2 + 2\sum_{j=0}^{\infty}\sum_{s=1}^{\infty} f_j f_{j+s}\right] - 2\sum_{j=0}^{\infty}\sum_{s=T_m}^{\infty} f_j f_{j+s} - 2\left[\frac{1}{T_m}\sum_{j=0}^{\infty}\sum_{s=1}^{T_m-1} s f_j f_{j+s}\right] \\
&= A_{1T} - 2A_{2T} - 2A_{3T}, \text{ say.}
\end{aligned}$$

Here $A_{1T} = (\sum_0^{\infty} f_j)^2 = (\sum_0^{\infty} c_j d_j)^2 = 0$ because of (A.5), and therefore it suffices to show that $A_{2T} \to 0$ and $A_{3T} \to 0$ as $T \to \infty$. First,

$$|A_{2T}| \le \sum_{j=0}^{\infty}\sum_{s=T_m}^{\infty} |f_j f_{j+s}| \le \sum_{j=0}^{\infty}|f_j| \sum_{s=T_m}^{\infty}|f_s| \to 0,$$

because

$$\sum_0^\infty |f_j| = \sum_0^\infty |c_j d_j| \le \left(\sum_0^\infty c_j^2\right)^{1/2} \left(\sum_0^\infty d_j^2\right)^{1/2} < \infty, \tag{A.16}$$

by (A.4). Next,

$$|A_{3T}| \le T_m^{-1} \sum_{j=0}^\infty \sum_{s=1}^{T_m} s|f_j f_{j+s}| = T_m^{-1} \sum_{j=0}^\infty |f_j| \sum_{s=1}^{T_m} s|f_{j+s}| \le T_m^{-1} \left(\sum_0^\infty |f_j|\right) \sum_0^\infty s|f_s|.$$

But $\sum_0^\infty |f_j| < \infty$ by (A.16), and $\sum_0^\infty s|f_s| \le (\sum_0^\infty sc_s^2)^{1/2}(\sum_0^\infty sd_s^2)^{1/2} < \infty$ by (A.4). So $A_{3T} = O(T_m^{-1}) \to 0$. ■

LEMMA A.4. *Under (A.3) and (A.4),* $\lim_{T\to\infty} B_T = B = \sum_{r=1}^\infty \left(\sum_{j=0}^\infty \Pi_{j,r}\right)^2 < \infty$.

Proof. The finiteness of B is proved in Theorem 6 of Phillips and Han (2008). To establish convergence, note that

$$B = \sum_{r=1}^\infty \left(\sum_{j=0}^\infty \Pi_{j,r}\right)^2 = \sum_{r=1}^\infty \sum_{j=0}^\infty \Pi_{j,r}^2 + 2\sum_{r=1}^\infty \sum_{j=0}^\infty \sum_{k=1}^\infty \Pi_{j,r}\Pi_{j+k,r},$$

so we have

$$B_T - B = -2\left(\sum_{k=1}^\infty \sum_{j=0}^\infty \sum_{r=1}^\infty \Pi_{j,r}\Pi_{j+k,r} - T_m^{-1} \sum_{t=m+2}^{T} \sum_{k=1}^{t-m-1} \sum_{j=0}^\infty \sum_{r=1}^\infty \Pi_{j,r}\Pi_{j+k,r}\right) = -2D_T,$$

say. Using (A.15) again, we have

$$\begin{aligned} D_T &= \sum_{k=1}^\infty \sum_{j=0}^\infty \sum_{r=1}^\infty \Pi_{j,r}\Pi_{j+k,r} - \sum_{k=1}^{T_m-1} \sum_{j=0}^\infty \sum_{r=1}^\infty \Pi_{j,r}\Pi_{j+k,r} + \frac{1}{T_m} \sum_{k=1}^{T_m-1} k \sum_{j=0}^\infty \sum_{r=1}^\infty \Pi_{j,r}\Pi_{j+k,r} \\ &= \sum_{k=T_m}^\infty \sum_{j=0}^\infty \sum_{r=1}^\infty \Pi_{j,r}\Pi_{j+k,r} + \frac{1}{T_m} \sum_{k=1}^{T_m-1} \sum_{j=0}^\infty \sum_{r=1}^\infty k\Pi_{j,r}\Pi_{j+k,r} = D_{1T} + D_{2T}, \text{ say.} \end{aligned}$$

As shown below, we have

$$\sum_{k=1}^\infty \sum_{j=0}^\infty \sum_{r=1}^\infty \left|\Pi_{j,r}\Pi_{j+k,r}\right| < \infty \quad \text{and} \quad \sum_{k=1}^\infty \sum_{j=0}^\infty \sum_{r=1}^\infty k\left|\Pi_{j,r}\Pi_{j+k,r}\right| < \infty, \tag{A.17}$$

which imply that $D_{1T} \to 0$ and $D_{2T} = O(T_m^{-1}) \to 0$.

Now we prove (A.17). For the first part of (A.17), note that $\sum_{k=1}^\infty |\Pi_{j+k,r}| \le \sum_{k=1}^\infty |\Pi_{k,r}| \le \sum_{j=0}^\infty |\Pi_{j,r}|$, so

$$\sum_{k=1}^\infty \sum_{j=0}^\infty \sum_{r=1}^\infty \left|\Pi_{j,r}\Pi_{j+k,r}\right| \le \sum_{r=1}^\infty \left(\sum_{j=0}^\infty |\Pi_{j,r}|\right)^2. \tag{A.18}$$

Now let $a_j = |c_j| + |d_j|$, implying that $|\Pi_{j,r}| \le |c_j d_{j+r}| + |c_{j+r} d_j| \le a_j a_{j+r}$. Then (A.3) and (A.4) imply that

$$\sum_0^\infty a_s < \infty, \quad \sum_0^\infty a_s^2 < \infty, \quad \sum_0^\infty s a_s^2 < \infty, \tag{A.19}$$

where the second and third results hold because $a_j^2 \le 2(c_j^2 + d_j^2)$. Now, the right-hand side of (A.18) is bounded by

$$\sum_{r=1}^\infty \left(\sum_{j=0}^\infty a_j a_{j+r} \right)^2 \le \sum_{r=1}^\infty \left(\sum_{j=0}^\infty a_j^2 \right) \left(\sum_{j=0}^\infty a_{j+r}^2 \right) = \left(\sum_0^\infty a_j^2 \right) \left(\sum_1^\infty s a_s^2 \right) < \infty,$$

by (A.19). So the first part of (A.17) is proved.

For the second part of (A.17),

$$\sum_{k=1}^\infty k |\Pi_{j+k,r}| \le \sum_{k=1}^\infty k a_{j+k} a_{j+k+r} \le \left(\sum_{k=1}^\infty k a_{j+k}^2 \right)^{1/2} \left(\sum_{k=1}^\infty k a_{j+k+r}^2 \right)^{1/2}$$

$$\le \sum_{k=1}^\infty k a_{j+k}^2 \le \sum_{k=1}^\infty (k+j) a_{k+j}^2 \le \sum_1^\infty k a_k^2,$$

and therefore the second part of (A.17) is

$$\sum_{k=1}^\infty \sum_{j=0}^\infty \sum_{r=1}^\infty k |\Pi_{j,r} \Pi_{j+k,r}| \le \sum_{j=0}^\infty \sum_{r=1}^\infty |\Pi_{j,r}| \left(\sum_1^\infty k a_k^2 \right) \le \left(\sum_0^\infty a_j \right)^2 \left(\sum_1^\infty k a_k^2 \right) < \infty$$

by (A.19). ■

We apply Theorems A.1 and A.2 to the components of the FDLS estimator $\hat{\rho}_{ols}$.

LEMMA A.5. $(nT_1)^{-1} \sum_{i=1}^n \sum_{t=2}^T (\Delta y_{it-1})^2 \to_p 2\sigma^2/(1+\rho)$ *as* $nT_1 \to \infty$.

Proof. Because of (A.1), we invoke Theorem A.1 with $c_0 = 0$, $c_1 = 1$, and $c_j = -(1-\rho)\rho^{j-2}$ for $j \ge 2$. The calculation of $\sum_0^\infty c_j^2$ for the limit is then obvious. ■

For the central limit theorem, we have the following lemma.

LEMMA A.6. *If* $E\varepsilon_{it}^4 < \infty$, *then as* $n \to \infty$, $(nT_1)^{-1/2} \sum_{i=1}^n \sum_{t=2}^T \Delta y_{it-1} \eta_{it}$ *converges to a normal distribution with variance* $8\sigma^4/(1+\rho)$, *and, furthermore,* $(nT_1)^{-1/2} \sum_{i=1}^n \sum_{t=2}^T \Delta y_{it-1} \eta_{it} \Rightarrow N(0, 8\sigma^4/(1+\rho))$ *whether* $n \to \infty$ *or not.*

Proof. Because the coefficients of the lag polynomials for Δy_{it-1} and η_{it} decay exponentially, conditions (A.3) and (A.4) are satisfied. Condition (A.5) holds by Lemma 1. The results follow from Theorem A.2. The variances for finite T and infinite T are calculated in Appendix B. ■

Proof of Theorem 1. This follows from Lemmas A.5 and A.6. ■

Proof of Lemma 2. Recall that the model is $y_{it} = \alpha_i + u_{it}$ with $u_{it} = \rho u_{it-1} + \varepsilon_{it}$. Let the sequence be initialized at $u_{i,-1}$. Then

$$u_{it} = \rho^{t+1} u_{i,-1} + \sum_{j=0}^t \rho^j \varepsilon_{it-j},$$

and therefore

$$\Delta y_{it-1} = \Delta u_{it-1} = \rho^{t-1}(\rho-1)u_{i,-1} + \varepsilon_{it-1} + (\rho-1)\sum_{j=0}^{t-2}\rho^j \varepsilon_{it-j-2},$$

and

$$E(\Delta y_{it-1})^2 = (\rho-1)^2\rho^{2(t-1)}Eu_{i,-1}^2 + \left[1+(\rho-1)^2\left(1+\rho^2+\cdots\rho^{2(t-2)}\right)\right]\sigma^2$$
$$= (\rho-1)^2\rho^{2(t-1)}Eu_{i,-1}^2 + (\rho+1)^{-1}\left[2+(\rho-1)\rho^{2(t-1)}\right]\sigma^2$$
$$= 2\sigma^2/(\rho+1) + \left(\frac{\rho-1}{\rho+1}\right)\rho^{2(t-1)}\sigma^2\delta_\rho, \tag{A.20}$$

where $\delta_\rho = (\rho^2-1)Eu_{i,-1}^2/\sigma^2 + 1$. Next, because $\eta_{it} = 2\Delta\varepsilon_{it} + (1+\rho)\Delta y_{it-1}$, we have

$$E\Delta y_{it-1}\eta_{it} = (\rho-1)\rho^{2(t-1)}\sigma^2\delta_\rho. \tag{A.21}$$

The results follow from (A.20) and (A.21). ■

Proof of Theorem 2. Because the moving average coefficients of the double differenced process decay exponentially as lag order increases, the conditions for Theorems A.1 and A.2 are all satisfied. The result follows from Theorems A.1 and A.2. ■

Next we show that the moment conditions (12) and (13) make the expected first derivative matrix (D) block diagonal, and if $E\varepsilon_{it}^3 = 0$ then the variance-covariance matrix (Ω) is block diagonal too when evaluated at the true parameter.

Proof that D and Ω Are Block Diagonal. For convenience, we write (12) and (13) as

$$EA_{it}(\rho,\beta) = E\sum_{t=2}^{T}\Delta z_{it-1}[2\Delta z_{it} + (1-\rho)\Delta z_{it}] = 0,$$

$$EB_{it}(\rho,\beta) = E\sum_{t=1}^{T}(\ddot{x}_{it} - \rho\dot{x}_{it-1})\left[(\ddot{y}_{it} - \rho\dot{y}_{it-1}) - (\ddot{x}_{it} - \rho\dot{x}_{it-1})'\beta\right] = 0,$$

where $\Delta z_{it} = \Delta y_{it} - \Delta x_{it}'\beta$.

For the D matrix, we need to show that $E\partial A_{it}/\partial\beta = 0$ and $E\partial B_{it}/\partial\rho = 0$ when evaluated at the true parameter. Because $\Delta z_{it} = \Delta u_{it}$ and $\ddot{y}_{it} = (\ddot{x}_{it} - \rho\dot{x}_{it-1})'\beta + \rho\dot{y}_{it-1} + \ddot{\varepsilon}_{it}$ when evaluated at the true parameter, we have

$$E\frac{\partial A_{it}}{\partial\beta} = -E\sum_{t=2}^{T}\left[\Delta x_{it-1}\eta_{it} + \Delta u_{it}\{2\Delta x_{it} + (1-\rho)\Delta x_{it-1}\}\right] = 0,$$

$$E\frac{\partial B_{it}}{\partial\rho} = -E\sum_{t=1}^{T}\left[\dot{x}_{it-1}\ddot{\varepsilon}_{it} + (\ddot{x}_{it} - \rho\dot{x}_{it-1})\dot{u}_{it-1}\right] = 0,$$

where $\eta_{it} = 2\Delta u_{it} + (1-\rho)\Delta u_{it-1}$ and $\dot{u}_{it-1} = u_{it-1} - T^{-1}\sum_{s=1}^{T}u_{is-1} = \dot{y}_{it-1} - \dot{x}_{it-1}'\beta$. So the D matrix is block diagonal. Next, for the Ω matrix, at the true parameter,

$$EB_{it}A_{it} = E\left[\sum_{t=1}^{T}(\ddot{x}_{it} - \rho\dot{x}_{it-1})\ddot{\varepsilon}_{it}\right]\left[\sum_{t=2}^{T}\Delta u_{it-1}\eta_{it}\right] = 0,$$

when $E\varepsilon_{it} = 0$, $E\varepsilon_{it}^3 = 0$, and ε_{it} are i.i.d. So the Ω matrix is also block diagonal under the zero third-moment assumption.

Now let us obtain the limit behavior of $U_{nT}(c)$ of (20) under the local to unity setting $\rho = \rho_{nT} = 1 - n^{-1/2}T^{-1}c$ as $n \to \infty$ (whether T is fixed or $T \to \infty$).

Proof of (22). We omit the nT subscript from ρ_{nT} for notational brevity. Clearly,

$$U_{nT}(c) = -n \log \left(|\Omega_T(\rho)| / |\Omega_T(1)| \right) - \frac{1}{\sigma^2} \sum_{i=1}^{n} (\Delta y_i)' \left[\Omega_T(\rho)^{-1} - \Omega_T(1)^{-1} \right] (\Delta y_i),$$

where $\Omega_T(\cdot)$ is defined in (21). Under the alternative hypothesis that $n^{1/2}T(1-\rho) = c$,

$$\begin{aligned} EU_{nT}(c) &= n \left(\operatorname{tr} \Omega_T(\rho) - T - \log |\Omega_T(\rho)| \right) \\ &= n \left(\frac{(1-\rho)T}{1+\rho} - \log \left[1 + \frac{(1-\rho)T}{1+\rho} \right] \right), \end{aligned}$$

where $|\Omega_T(\rho)| = 1 + T(1-\rho)/(1+\rho)$ (see Han, 2007, Thm. 1). Because $x - \log(1+x) = x^2/2 - o(x^2)$ when x is close to zero, we have

$$EU_{nT}(c) = \frac{nT^2(1-\rho)^2}{2(1+\rho)^2} + o\left(\frac{nT^2(1-\rho)^2}{(1+\rho)^2} \right) = \frac{c^2}{2(1+\rho)^2} + o(1) \to \frac{c^2}{8}.$$

The variance of $U_{nT}(c)$ is calculated from the fact that $\operatorname{var}(z'Az) = 2\operatorname{tr}(A'A)$ if $z \sim N(0, I)$ and A is symmetric. In our case,

$$\sigma^{-2} \Delta y_i' \left[\Omega_T(\rho)^{-1} - \Omega_T(1)^{-1} \right] \Delta y_i = z_i' [I_T - \Omega_T(\rho)] z_i = z_i' Q_T z_i, \quad \text{say,}$$

where $z_i = \sigma^{-1} \Omega_T(\rho)^{-1/2} \Delta y_i \sim N(0, I)$ and

$$Q_T = \text{Toeplitz}\{-1, 1, \rho, \ldots, \rho^{T_2}\}(1-\rho)/(1+\rho),$$

so that

$$\begin{aligned} \operatorname{var}\left(U_{nT}(c) \right) &= 2n \operatorname{tr}(Q_T' Q_T) = \frac{2n(1-\rho)^2}{(1+\rho)^2} \left[T + 2 \sum_{k=1}^{T_1} T_k \rho^{2(k-1)} \right] \\ &= \frac{2nT^2(1-\rho)^2}{(1+\rho)^2} \left[\frac{1}{T} + \frac{2}{T^2} \sum_{k=1}^{T_1} T_k - \frac{2}{T^2} \sum_{k=1}^{T_1} T_k (1 - \rho^{2(k-1)}) \right] \\ &= \frac{2c^2}{(1+\rho)^2} \left[1 - \frac{2}{T^2} \sum_{k=1}^{T_1} T_k (1 - \rho^{2(k-1)}) \right] = \frac{2c^2}{(1+\rho)^2} [1 - o(1)] \to \frac{c^2}{2} \end{aligned}$$

as $n \to \infty$. Above, the last $o(1)$ order follows from

$$\begin{aligned} \frac{1}{T^2} \sum_{k=1}^{T_1} T_k [1 - \rho^{2(k-1)}] &= \frac{1}{T^2} \sum_{k=1}^{T_2} T_{k+1} (1 - \rho^{2k}) = \frac{1}{T^2} \sum_{k=1}^{T_2} T_{k+1} (1-\rho) \left(\sum_{j=0}^{k} \rho^j \right) (1 + \rho^k) \\ &\le \frac{2}{T^2} \sum_{k=1}^{T_2} T_{k+1} k (1-\rho) = \frac{2c}{n^{1/2}T^2} \sum_{k=1}^{T_3} k \to 0 \end{aligned}$$

as $n \to \infty$, whether T is fixed or $T \to \infty$.

So far, we have obtained that $EU_{nT}(c) \to c^2/8$ and $\mathrm{var}(U_{nT}(c)) \to c^2/2$. In the derivation of the variance, we can also verify that it is uniformly bounded, which is sufficient to invoke the Lindeberg CLT for $U_{nT}(c)$. Thus,

$$U_{nT}(c) \to_d N(c^2/8, c^2/2)$$

as claimed. ■

APPENDIX B: Asymptotic Variance Formulas

The Variance of FDLS. We calculate the variance of the FDLS estimator $\hat{\rho}_{ols}$ in terms of the parameters using the expressions in Theorem A.2. Let $\rho_1 = 1 - \rho$ and $\rho_2 = 1 - \rho^2$. The lag polynomial coefficients of Δy_{t-1} and η_t are tabulated in Table B1, and the corresponding $\Pi_{j,r}$ terms are tabulated in Table B2. So

$$\sum_0^\infty c_j^2 d_j^2 = \rho_1^2 + \rho_1^2 \rho_2^2 (1 + \rho^4 + \cdots) = 2(1-\rho)^2/(1+\rho^2), \tag{B.1}$$

$$\sum_0^\infty c_j d_j c_{j+k} d_{j+k} = -\rho^{2(k-1)} \left[\rho_1^2 \rho^2 + \rho^2 \rho_1^2 \rho_2^2 (1 + \rho^4 + \cdots) \right]$$
$$= -\rho^{2(k-1)} (1-\rho)^2 (1-\rho^2)/(1+\rho^2), \quad k \geq 1, \tag{B.2}$$

TABLE B1. Coefficients of lag polynomials for Δy_{it-1} and η_{it}

	0	1	2	3	4	j
Δy_{t-1}	0	1	$-\rho_1$	$-\rho\rho_1$	$-\rho^2\rho_1$	$-\rho^{j-2}\rho_1$
η_t	2	$-\rho_1$	$-\rho_2$	$-\rho\rho_2$	$-\rho^2\rho_2$	$-\rho^{j-2}\rho_2$

$\rho_1 = 1 - \rho$, $\rho_2 = 1 - \rho^2$

TABLE B2. The $\Pi_{j,r}$ terms

$r \setminus j$	0	1	2	3	4	5
1	2	$-2\rho\rho_1$	$2\rho\rho_1\rho_2$	$2\rho^3\rho_1\rho_2$	$2\rho^5\rho_1\rho_2$	$\cdots$
2	$-2\rho_1$	$-2\rho^2\rho_1$	$2\rho^2\rho_1\rho_2$	$2\rho^4\rho_1\rho_2$	$2\rho^6\rho_1\rho_2$	$\cdots$
3	$-2\rho\rho_1$	$-2\rho^3\rho_1$	$2\rho^3\rho_1\rho_2$	$2\rho^5\rho_1\rho_2$	$2\rho^7\rho_1\rho_2$	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	

where $0^0 = 1$. Also,

$$\sum_{j=0}^{\infty}\sum_{r=1}^{\infty}\Pi_{j,r}^2 = [4+4\rho_1^2(1+\rho^2+\cdots)]+4\rho^2\rho_1^2(1+\rho^2+\cdots)$$

$$+4\rho^2\rho_1^2\rho_2^2(1+\rho^2+\cdots)(1+\rho^4+\cdots)$$

$$=4\left[1+\frac{(1+\rho^2)(1-\rho)}{1+\rho}+\frac{\rho^2(1-\rho)^2}{1+\rho^2}\right], \quad \textbf{(B.3)}$$

and (after some algebra)

$$\sum_{j=0}^{\infty}\sum_{r=1}^{\infty}\Pi_{j,r}\Pi_{j+k,r} = -4\rho(1-\rho)/(1+\rho)\{k=1\}+4\rho^{2k-3}(1-\rho)^2\{k>1\}$$

$$-4\rho^{2k}(1-\rho)^2/(1+\rho^2), \quad k\geq 1. \quad \textbf{(B.4)}$$

The calculation of A_T and B_T of Theorem A.2 (with $m=1$) is tedious. We will use

$$\sum_{t=3}^{T}\sum_{k=1}^{t-2}\rho^{2(k-1)} = \frac{T_2}{1-\rho^2}-\left(\frac{\rho^2}{1-\rho^2}\right)\frac{1-\rho^{2T_2}}{1-\rho^2}$$

and

$$\sum_{t=3}^{T}\sum_{k=2}^{t-2}\rho^{2(k-2)} = \sum_{t=4}^{T}\sum_{k=2}^{t-2}\rho^{2(k-2)} = \frac{T_3}{1-\rho^2}-\left(\frac{\rho^2}{1-\rho^2}\right)\frac{1-\rho^{2T_3}}{1-\rho^2}.$$

From the definition of A_T, (B.1) and (B.2), we have

$$A_T = \frac{2(1-\rho)^2}{1+\rho^2}-\frac{2}{T_1}\left[\frac{T_2}{1-\rho^2}-\left(\frac{\rho^2}{1-\rho^2}\right)\frac{1-\rho^{2T_2}}{1-\rho^2}\right]\frac{(1-\rho)^2(1-\rho^2)}{1+\rho^2}$$

$$=\frac{2}{T_1}\left[\frac{(1-\rho)^2}{1+\rho^2}+\frac{\rho^2(1-\rho)(1-\rho^{2T_2})}{(1+\rho)(1+\rho^2)}\right], \quad \textbf{(B.5)}$$

and

$$B_T = 4\left[1+\frac{(1+\rho^2)(1-\rho)}{1+\rho}+\frac{\rho^2(1-\rho)^2}{1+\rho^2}\right]$$

$$+\frac{8}{T_1}\left\{-\frac{T_2\rho(1-\rho)}{1+\rho}\{T>3\}+\left[\frac{T_3(1-\rho)}{1+\rho}-\frac{\rho^2(1-\rho^{2T_3})}{(1+\rho)^2}\right]\rho\right.$$

$$\left.-\left[\frac{T_2(1-\rho)}{1+\rho}-\frac{\rho^2(1-\rho^{2T_2})}{(1+\rho)^2}\right]\frac{\rho^2}{1+\rho^2}\right\},$$

which equals

$$B_T = 4\left[1 + \frac{(1+\rho^2)(1-\rho)}{1+\rho} + \frac{\rho^2(1-\rho)^2}{1+\rho^2} - \left(\frac{2}{T_1}\right)\frac{\rho(1-\rho)}{1+\rho}\{T \geq 3\} - 2\left(\frac{T_2}{T_1}\right)\frac{\rho^2(1-\rho)}{(1+\rho)(1+\rho^2)} + \left(\frac{2}{T_1}\right)\frac{\rho^4(1-\rho^{2T_2})}{(1+\rho)^2(1+\rho^2)} - \left(\frac{2}{T_1}\right)\frac{\rho^3(1-\rho^{2T_3})}{(1+\rho)^2}\right], \quad \textbf{(B.6)}$$

where A_T and B_T are defined in Theorem A.2 and $m = 1$ is used.

The limit variance (as $n \to \infty$) of $(nT_1)^{1/2}(\hat{\rho}_{ols} - \rho)$ is now obtained by multiplying $A_T \operatorname{var}(\varepsilon_{it}^2) + B_T\sigma^4$ by $(1+\rho)^2/4\sigma^4$, viz.,

$$\begin{aligned} V_{ols,T} = \frac{1}{T_1}\Bigg[&\frac{(1-\rho^2)^2}{1+\rho^2} - \frac{\rho^2(1-\rho^2)(1-2\rho^{2T_2})}{1+\rho^2}\Bigg](1/2)\operatorname{var}(\varepsilon_{it}^2/\sigma^2) \\ &+ (1+\rho)^2 + (1+\rho^2)(1-\rho^2) + \frac{\rho^2(1-\rho^2)^2}{1+\rho^2} - \left(\frac{2}{T_1}\right)\rho(1-\rho^2)\{T \geq 3\} \\ &- 2\left(\frac{T_2}{T_1}\right)\frac{\rho^2(1-\rho^2)}{1+\rho^2} + \left(\frac{2}{T_1}\right)\frac{\rho^4(1-\rho^{2T_2})}{1+\rho^2} \\ &- \left(\frac{2}{T_1}\right)\rho^3(1-\rho^{2T_3}). \end{aligned} \quad \textbf{(B.7)}$$

As a special case, if $T = 2$ and $\varepsilon_{it} \sim N(0,\sigma^2)$, then $\operatorname{var}(\varepsilon_{it}^2/\sigma^2) = 2$ and $V_{ols,T}$ is simplified to

$$V_{ols,T} = (3-\rho)(1+\rho), \quad T = 2,\ \varepsilon_{it} \sim N(0,\sigma^2). \quad \textbf{(B.8)}$$

It is also easily verified that $V_{ols,T} \to 2(1+\rho)$ as $T \to \infty$.

The Variance of DDLS when $\rho = 1$. Next, we calculate the variance of the DDLS (double-differencing least squares) estimator $\hat{\theta}$ for $\rho = 1$. According to Phillips and Han (2008), $\Delta^2 y_{it-1} = \sum_0^\infty c_j \varepsilon_{it-j}$ and $\tilde{\eta}_{it} = \sum_0^\infty d_j \varepsilon_{it-j}$, where

$$c_0 = 0,\ c_1 = 1,\ c_2 = -(2-\rho),\ c_k = \rho^{k-3}(1-\rho)^2,\ k \geq 3, \quad \textbf{(B.9)}$$

$$d_0 = 2,\ d_1 = -4 + \phi c_1,\ d_2 = 2 + \phi c_2,\ d_k = \phi c_k,\ k \geq 3, \quad \textbf{(B.10)}$$

with $\phi = (4-\rho)(1+\rho)/(3-\rho)$ and $0^0 = 1$. When $\rho = 1$, we have

$$c_0 = 0,\ c_1 = 1,\ c_2 = -1,\ c_k = 0, \quad k \geq 3,$$

$$d_0 = 2,\ d_1 = -1,\ d_2 = -1,\ d_k = 0, \quad k \geq 3.$$

So

$$\sum_0^\infty c_j^2 d_j^2 = 2 \quad \text{and} \quad \sum_{t=m+2}^{T}\sum_{r=1}^{t-m-1}\sum_{j=0}^{\infty} c_j d_j c_{j+r} d_{j+r} = -T_{m+1},$$

and because

$$\Pi_{0,1} = 2, \quad \Pi_{0,2} = -2, \quad \text{and} \quad \Pi_{j,r} = 0 \quad \text{for all other } j \text{ and } r,$$

we have

$$\sum_{j=0}^{\infty} \sum_{r=1}^{\infty} \Pi_{j,r}^2 = 8, \qquad \sum_{t=m+2}^{T} \sum_{k=1}^{t-m-1} \sum_{j=0}^{\infty} \sum_{r=1}^{\infty} \Pi_{j,r} \Pi_{j+k,r} = 0.$$

For the DDLS estimator $\hat{\theta}$, we use $m = 2$, and thus, by Theorem A.2, when $\rho = 1$ and $T > 2$,

$$(nT_2)^{-1/2} \sum_{i=1}^{n} \sum_{t=3}^{T} \Delta^2 y_{it-1} \tilde{\eta}_{it} \to_d N\Big(0, 2\,\mathrm{var}(\varepsilon_{it}^2)/T_2 + 8\sigma^4\Big),$$

and because $\mathrm{plim}_{nT\to\infty}(nT_2)^{-1} \sum_{i=1}^{n} \sum_{t=3}^{T} (\Delta^2 y_{it-1})^2 = 2\sigma^4$ by Theorem A.1, we have $\theta = 1$ and as $nT \to \infty$,

$$(nT_2)^{1/2} \hat{\theta} \to_d N\Big(0, \mathrm{var}(\varepsilon_{it}^2/\sigma^2)/2T_2 + 2\Big).$$