UNIFORM INFERENCE IN PANEL AUTOREGRESSION

By

John C. Chao and Peter C. B. Phillips

January 2017

COWLES FOUNDATION DISCUSSION PAPER NO. 2071

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

http://cowles.yale.edu/

Uniform Inference in Panel Autoregression*

John C. Chao[†] and Peter C. B. Phillips[‡]

January 6, 2017

Abstract

This paper considers estimation and inference concerning the autoregressive coefficient (ρ) in a panel autoregression for which the degree of persistence in the time dimension is unknown. The main objective is to construct confidence intervals for ρ that are asymptotically valid, having asymptotic coverage probability at least that of the nominal level uniformly over the parameter space. It is shown that a properly normalized statistic based on the Anderson-Hsiao IV procedure, which we call the $\mathbb{M}$ statistic, is uniformly convergent and can be inverted to obtain asymptotically valid interval estimates. In the unit root case confidence intervals based on this procedure are unsatisfactorily wide and uninformative. To sharpen the intervals a new procedure is developed using information from unit root pretests to select alternative confidence intervals. Two sequential tests are used to assess how close ρ is to unity and to correspondingly tailor intervals near the unit root region. When ρ is close to unity, the width of these intervals shrinks to zero at a faster rate than that of the confidence interval based on the $\mathbb{M}$ statistic. Only when both tests reject the unit root hypothesis does the construction revert to the $\mathbb{M}$ statistic intervals, whose width has the optimal $N^{-1/2}T^{-1/2}$ rate of shrinkage when the underlying process is stable. The asymptotic properties of this pretest-based procedure show that it produces confidence intervals with at least the prescribed coverage probability in large samples. Simulations confirm that the proposed interval estimation methods perform well in finite samples and are easy to implement in practice. A supplement to the paper provides an extensive set of new results on the asymptotic behavior of panel IV estimators in weak instrument settings.

*The original draft of this paper was prepared in December 2015. We thank participants at the New York Camp Econometrics XI, the 2016 Australasia Meeting of the Econometric Society, and the 2016 NBER/NSF Time Series Conference, as well as workshop participants at Penn State and Boston College for comments. Phillips acknowledges support from the NSF under Grant SES 1258258.

[†]University of Maryland. Email: Chao@econ.umd.edu

[‡]Yale University, University of Auckland, University of Southampton and Singapore Management University. Email: peter.phillips@yale.edu

Keywords: Confidence interval, Dynamic panel data models, panel IV, pooled OLS, Pretesting, Uniform inference.

JEL classification: C230, C360

1 Introduction

Due to the many challenges that arise in estimating and conducting statistical inference for dynamic panel data models, a vast literature has emerged studying these models over the past three decades. Much has been learnt about the large sample properties and finite sample performance of various estimation procedures in stable dynamic panel models, not only in univariate but also in multivariate contexts. Important contributions to this literature began with Nickell (1981) and Anderson and Hsiao (1981,1982), followed by Arellano and Bond (1991), Ahn and Schmidt (1995), Arellano and Bover (1995), Kiviet (1995), Blundell and Bond (1998), Hahn and Kuersteiner (2002), Alvarez and Arellano (2003), amongst many others. Progress has also been made recently in studying such models when more persistent behavior, such as unit root or near unit root behavior, is present. Phillips and Moon (1999) provided methods that opened up the rigorous development of asymptotics in such models for both stationary and nonstationary cases and with multidimensional joint and sequential limits. Many subsequent contributions to this nonstationary panel literature have considered more complex regressions, analyzing the effects of incidental trends, serial dependence and cross section dependence (e.g., Phillips and Sul, 2003; Moon and Phillips, 2004; Moon et al, 2014a,b; Pesaran, 2006; Pesaran and Tosetti, 2011).

While this literature has greatly enhanced our understanding of the panel data sampling behavior of point estimators and of associated test statistics, such as the Studentized t statistic or the Wald statistic, what has not been studied are confidence interval procedures which are asymptotically valid in the sense that asymptotic coverage probabilities are at least that of the nominal level uniformly over the parameter space. The development of theoretically justified confidence intervals is especially important in cases where the empirical researcher may not have good prior information about the degree of persistence in the data, since in such situations interval estimates can serve as indispensible supplements to point estimates by providing additional information about sampling uncertainty and about the range of possible values of the autoregressive parameter ρ that are consistent with the observed data. Moreover, we know from the unit root time series literature that constructing an asymptotically valid confidence interval for the autoregressive parameter of an $AR(1)$ process is a challenging task when the parameter space is taken to be large enough to include both the stable and the unit root cases. This is because the Studentized statistic based on OLS estimation is not uniformly convergent in this case, so that an asymptotically correct confidence interval cannot be constructed by inverting the Studentized statistic in the usual way. To address this problem in the time series literature, Stock (1991) proposed a confidence procedure based on local-to-unity asymptotics, while simulation and bootstrap type methods have been introduced by Andrews (1993) and Hansen (1999). Recent results by Mikusheva (2007, 2012) and by Phillips (2014) have shown that the methods of Andrews (1993) and Hansen (1999) as well as a recentered version of Stock's method all give the correct asymptotic coverage probability uniformly over the parameter space. Extending these procedures to the panel data setting does not seem to be straightforward, and panel data versions of these methods are currently unavailable.

To address this need, the present paper proposes simple, asymptotically correct confidence procedures for the autoregressive coefficient of a panel autoregression.[1] We start by showing that a properly normalized statistic based on the estimating equation of the Anderson-Hsiao (1981, 1982) IV procedure is uniformly convergent over the parameter space $\Theta_\rho = \{\rho : \rho \in (-1, 1]\}$. This statistic, which we refer to as the $\mathbb{M}$ statistic since it is based on the (empirical) IV moment function, can be easily and analytically inverted to obtain an asymptotically correct confidence interval. However, a drawback of this procedure is that in unit root and very near unit root cases these confidence intervals are often not very informative in the sense that they may be wide in finite samples and, asymptotically, their width shrinks toward zero at the slow rate of $T^{-1/2}$ even when both the cross section (N) and time series (T) sample sizes approach infinity. A similar drawback applies to the GMM procedure of of Han and Phillips (2010), which achieves uniform inference with shrinkage rate $(NT)^{-1/2}$ over the full domain Θ_ρ.

To obtain more informative interval estimates, we introduce a new confidence procedure which uses information from two different unit root tests, with different

[1]We do not consider in this paper issues related to incidental trends, cross section dependence, and slope parameter heterogeneity discussed earlier. While these complications are important and empirically relevant, they are beyond the scope of the current paper and considering them here would divert from the main point of this paper which concerns the development of uniform inference procedures.

power properties, to assess the proximity of the true autoregressive parameter from the exact unit root null hypothesis $H_0 : \rho = 1$. More precisely, we infer that the true parameter value is unity or very close to unity if the more powerful of the two unit root tests fails to reject H_0, and we use, in this case, an interval that is localized at $\rho = 1$, with width that shrinks at a faster $N^{-1/2}T^{-1}$ rate.[2] Second, if the more powerful test rejects H_0 but the less powerful test fails to reject, we use another interval that is still localized at $\rho = 1$ but with greater width which shrinks at the rate $N^{-1/2}T^{-1/2}$, a rate that is still faster than that of the width of the confidence interval based on the $\mathbb{M}$ statistic in the vicinity of $\rho = 1$. Finally, if both tests reject H_0, then we conclude that the true parameter value is far enough away from unity that we can use the confidence interval based on the $\mathbb{M}$ statistic, whose width shrinks at the optimal $N^{-1/2}T^{-1/2}$ rate in the stable region of the parameter space. We show that the asymptotic size of this pretest based procedure can be uniformly controlled, so that this procedure is asymptotically valid, albeit slightly conservative when the underlying process is stable. The degree of conservatism under our procedure is also controllable and can be kept small by carefully controlling the probability of a Type II error under a local-to-unity parameter sequence. Moreover, in addition to providing informative and asymptotically correct confidence intervals, our procedure has the further advantage that it is given in analytical form and, hence, is computationally simple and extremely easy to implement.

The remainder of the paper proceeds as follows. Section 2 briefly describes the model, assumptions, and notation. Section 3 introduces two new ways of constructing uniform confidence intervals for the parameter ρ. The first is based on inverting the $\mathbb{M}$ statistic, and the second is the pre-test based confidence interval. Results given in this section show that both confidence procedures are asymptotically valid. Section 4 reports the results of a Monte Carlo study comparing our proposed confidence procedures with some alternative procedures. We provide a brief conclusion in section 5. Proofs of the main theorems are given in the Appendix to this paper. Proofs of additional supporting lemmas as well as additional Monte Carlo results are reported

[2] Other approaches for achieving uniform inference in estimation have been proposed recently in the time series literature by Han et al. (2011) using partial aggregation methods and by Gorodnichenko et al. (2012) using quasi-differencing. In the unit root and very near unit root cases, extending these approaches to the panel data setting leads to confidence intervals whose width shrinks at a slower rate than the optimal $N^{-1/2}T^{-1}$ rate obtained here. Han et al (2014) developed a panel estimator using X-differencing which has good bias properties and limit theory but has different limit theory in unit root and stationary cases, complicating uniform inference.

in a supplement to this paper (Chao and Phillips, 2016). The supplement provides an extensive set of results for panel estimation limit theory in unit root and near unit root cases that are helpful in obtaining the main results in the paper but are of wider interest regarding asymptotic behavior of panel IV estimators, particularly in weak instrument settings.

A word on notation. We use $\Rightarrow$ for convergence in distribution or weak convergence, $\xrightarrow{p}$ for convergence in probability, χ^2_ν denotes a chi-square random variable with ν degrees of freedom, and $W_i(r)$ is standard Brownian motion on the unit interval $[0,1]$ for each i. In addition, for two sequences $\{X_T\}$ and $\{Y_T\}$, we take $X_T \ll Y_T$ to mean $X_T/Y_T = o(1)$ and $X_T \sim Y_T$ to mean that $X_T/Y_T = O(1)$ and $Y_T/X_T = O(1)$, as $T \to \infty$. Finally, we use wid($\mathbb{C}$) to denote the width of the confidence interval $\mathbb{C}$.

2 Model, Assumptions, and Point Estimation

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

$$y_{it} = a_i + w_{it}, \tag{1}$$
$$w_{it} = \rho_T w_{it-1} + \varepsilon_{it}, \tag{2}$$

for $i = 1, ...N$ and $t = 1, ..., T$. Here, $\{y_{it}\}$ is the observed data, $\{w_{it}\}$ is generated by a latent AR(1) process, and a_i denotes an (unobserved) random effect. Since the AR process (2) depends on an indexed parameter ρ_T, as opposed to a fixed autoregressive parameter ρ, the observed data and the latent process are strictly triangular indexed arrays $\{y_{it,T}, w_{it,T}\}$ but the additional dependence is suppressed for notational convenience. However, the analysis that follows studies limit behavior under a variety of parameter sequences, including both exact unit root processes and a general class of local-to-unity sequences given by the parameterization $\rho_T = \exp\{-1/q(T)\}$, where $q(T)$ is a non-negative function of T such that $q(T) \to \infty$ as $T \to \infty$. Parameter sequences for stable AR processes can also be written in this general form by considering the collection of sequences $\{\rho_T\}$ belonging to

$$\mathcal{G}_{\mathrm{St}} = \left\{ |\rho_T| = \exp\left\{-\frac{1}{q(T)}\right\} : q(T) \geq 0 \text{ and } q(T) = O(1) \text{ as } T \to \infty \right\}.$$

It is sometimes convenient to rewrite the model (1)-(2) in the alternate familiar form as a first-order autoregressive process in y_{it}, viz.,

$$y_{it} = a_i (1 - \rho_T) + \rho_T y_{it-1} + \varepsilon_{it} = \eta_{i,T} + \rho_T y_{it-1} + \varepsilon_{it}, \tag{3}$$

where $\eta_{i,T} = a_i (1 - \rho_T)$. The following assumptions are made on the model.

Assumption 1 (Errors): (a) $\{\varepsilon_{it}\} \equiv i.i.d.\,(0, \sigma^2)$ *across* i *and* t, $\sigma^2 > 0$; (b) $E[\varepsilon_{it}^4] < \infty$.

Assumption 2 (Random Effects): (a) $\{a_i\} \equiv i.i.d.\,(\mu_a, \sigma_a^2)$ *across* i, $\sigma_a^2 > 0$; (b) $E[a_i^4] < \infty$.

Assumption 3: ε_{it} *and* a_j, *are mutually independent for all* $i, j = 1, 2, ..., N$ *and for all* $t = 1, 2, ..., T$.

Assumption 4: (Initialization): *Let* $y_{i0} = a_i + w_{i0}$. *Suppose that there exists a positive constant* C *such that* $\sup_i E[w_{i0}^2] \leq C < \infty$, *and suppose that* w_{i0} *and* ε_{jt} *are independent for all* $i, j = 1, 2, ..., N$ *and for all* $t = 1, 2, ..., T$.

Assumption 4 on the initial condition does not impose mean stationarity, i.e., the condition that $E[y_{i0}|a_i] = \eta_i / (1 - \rho) = a_i \;\; a.s.$, which in our setup is equivalent to the restriction that $E[w_{i0}|a_i] = 0 \; a.s.$ In addition, observe that Assumption 4 allows for the case where the initial condition is fixed, i.e., $w_{i0} = c_i$ for some sequence of constants $\{c_i\}$ such that $\sup_i c_i < \infty$. It is also general enough to cover the case where we may specify w_{i0} to be fixed in the unit root case but allow w_{i0} to be a draw from its unconditional distribution with variance $\sigma^2 / (1 - \rho^2)$ when the underlying process is stationary.

Before proceeding to a discussion of confidence procedures, we introduce a new point estimator for the autoregressive parameter ρ. Although the focus of this paper is not on point estimation, we need as part of our confidence procedures a point estimator for ρ with a sufficiently fast rate of convergence in the exact unit root and near unit root cases. For this purpose, the following estimator based on averaging the Anderson-Hsiao IV estimator and the pooled ordinary least squares (POLS) estimator has the desired properties. More precisely, we construct the estimator

$$\hat{\rho}_{\text{AIP}} = w_{IC} \hat{\rho}_{\text{IVD}} + (1 - w_{IC}) \hat{\rho}_{\text{pols}}, \tag{4}$$

where $w_{IC} = \left[1 + \exp\left\{\frac{1}{2}\Delta_{IC}\right\}\right]^{-1}$ and $\Delta_{IC} = \mathbb{T}_{NT} + \sqrt{N}L(T)$ and where

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

are, respectively, the Anderson-Hsiao IV estimator and the POLS estimator[3]. Here, $L(T)$ denotes a slowly varying function such that $L(T) \to \infty$ as $T \to \infty$. In addition, let

$$\mathbb{T}_{N,T} = \frac{\hat{\rho}_{\text{pols}} - 1}{\sqrt{\widehat{\sigma}^2 / \sum_{i=1}^{N}\sum_{t=2}^{T}\left(y_{it-1} - \overline{y}_{-1,NT}\right)^2}}, \tag{5}$$

be the Studentized statistic for testing the unit root null hypothesis, where $\widehat{\sigma}^2 = N^{-1}T_1^{-1}\sum_{i=1}^{N}\sum_{t=2}^{T}\left(y_{it} - \overline{y}_i - \widehat{\rho}_{\text{pre}}\left[y_{it-1} - \overline{y}_{i,-1}\right]\right)^2$, $\overline{y}_i = T_1^{-1}\sum_{t=2}^{T} y_{it}$, and $\overline{y}_{i,-1} = T_1^{-1}\sum_{t=2}^{T} y_{it-1}$, and $T_1 = T - 1$. Our estimator of the scale parameter σ^2 requires a preliminary estimator $\widehat{\rho}_{\text{pre}}$, which we also take to be a weighted average of $\hat{\rho}_{\text{IVD}}$ and $\hat{\rho}_{\text{pols}}$, just like $\hat{\rho}_{\text{AIP}}$, except that we ignore the estimation of σ^2, essentially setting $\widehat{\sigma}^2 = 1$ in the formula for the studentized statistic given in the expression (5).[4] Ignoring the scale parameter σ^2 does not have an adverse effect on the consistency and rate of convergence of this preliminary estimator.[5] However, we do expect $\hat{\rho}_{\text{AIP}}$ to have better finite sample properties relative to $\widehat{\rho}_{\text{pre}}$.

The following theorem summarizes the consistency and rate of convergence of $\hat{\rho}_{\text{AIP}}$ under different parameter sequences $\{\rho_T\}$.

---

[3] We use the notation $\widehat{\rho}_{\text{IVD}}$ to denote the Anderson-Hsiao IV estimator because it is a procedure where IV estimation is performed on a first-differenced equation. Later, we use $\widehat{\rho}_{\text{IVL}}$ to denote the IV estimator introduced by Arellano and Bover (1995) since, in that procedure, IV is performed on the panel autoregression in levels.

[4] More specifically, we define the preliminary estimator to be

$$\widehat{\rho}_{\text{pre}} = \overline{w}_{IC}\hat{\rho}_{\text{IV}} + \left(1 - \overline{w}_{IC}\right)\hat{\rho}_{\text{pols}}.$$

where $\overline{w}_{IC} = 1/1 + \exp\left\{\frac{1}{2}\overline{\Delta}_{IC}\right\}$, $\overline{\Delta}_{IC} = \overline{\mathbb{T}}_{NT} + \sqrt{N}L(T)$, and $\overline{\mathbb{T}}_{N,T} = \left[\sum_{i=1}^{N}\sum_{t=2}^{T}\left(y_{it-1} - \overline{y}_{-1,NT}\right)^2\right]^{1/2}\left(\hat{\rho}_{\text{pols}} - 1\right)$

[5] See Lemma SD-11 and its proof in Appendix SD of the supplement for an asymptotic analysis of $\widehat{\rho}_{\text{pre}}$.

Theorem 2.1:

Suppose Assumptions 1-4 hold. The following statements are true as $N, T \to \infty$ such that $N^\kappa/T = \tau$ for $\kappa \in \left(\frac{1}{2}, \infty\right)$ and $\tau \in (0, \infty)$.

(a) *If $\rho_T = 1$ for all T sufficiently large, then $\widehat{\rho}_{\text{AIP}} - \rho_T = O_p\left(\frac{1}{T\sqrt{N}}\right)$.*

(b) *If $\rho_T = \exp\{-1/q(T)\}$ such that $T/q(T) = O(1)$, then $\widehat{\rho}_{\text{AIP}} - \rho_T = O_p\left(\frac{1}{T\sqrt{N}}\right)$.*

(c) *If $\rho_T = \exp\{-1/q(T)\}$ such that $T/(L(T))^2 \ll q(T) \ll T$, then $\widehat{\rho}_{\text{AIP}} - \rho_T = O_p\left(\frac{1}{\sqrt{NTq(T)}}\right)$.*

(d) *If $\rho_T = \exp\{-1/q(T)\}$ such that $q(T) \to \infty$ but $q(T)(L(T))^2/T = O(1)$, then $\widehat{\rho}_{\text{AIP}} - \rho_T = O_p\left(\frac{1}{\sqrt{NT}}\right)$.*

(e) *If $\rho_T \in \mathcal{G}_{\text{St}} = \left\{|\rho_T| = \exp\left\{-\frac{1}{q(T)}\right\} : q(T) \geq 0 \text{ and } q(T) = O(1) \text{ as } T \to \infty\right\}$, then $\widehat{\rho}_{\text{AIP}} - \rho_T = O_p\left(\frac{1}{\sqrt{NT}}\right)$.*

Remarks:

(i) From Theorem SA-1 given in the supplement, we see that the IV estimator tends to perform well when $q(T) = O(T)$, i.e., when the sequence of autoregressive parameters is relatively further away from unity. On the other hand, Theorem SA-2 in the supplement shows that the POLS estimator does better when the parameter sequence is close to unity than when it is not. More precisely, POLS is not only consistent and asymptotically normal but is also free of second-order bias when $T^{\frac{1+\kappa}{3\kappa}} \ll q(T)$, but it is inconsistent when the underlying process is stable. $\widehat{\rho}_{\text{AIP}}$ is designed so that it takes advantage of the differential strength of the IV estimator vis-à-vis the POLS estimator in different parts of the parameter space. In particular, this estimator behaves like IV in the "more stable" region of the parameter space characterized by parameter sequences such that $q(T) = O(T)$, but behaves like POLS in the "more persistent" region characterized by parameter sequences such that $T^{\frac{1+\kappa}{3\kappa}} \ll q(T)$ for $\kappa \in \left(\frac{1}{2}, \infty\right)$. Note that the weight function w_{IC} used to construct $\widehat{\rho}_{\text{AIP}}$ depends on an information criterion type statistic $\Delta_{IC} = \mathbb{T}_{NT} + \sqrt{N}L(T)$, where the "penalty" component $\sqrt{N}L(T)$ is constructed so that the transition from IV to POLS or vice versa takes place in the region $T^{\frac{1+\kappa}{3\kappa}} \ll q(T) \ll T$. An important consequence of formulating the weight function in this way is that, in addition to being consistent, $\widehat{\rho}_{\text{AIP}}$ is also free of second-order bias. See Appendix SA in the supplement to this paper for more dicussion on $\widehat{\rho}_{\text{AIP}}$.

(ii) Theorem 2.1 above shows that $\widehat{\rho}_{\text{AIP}}$ has certain robustness properties in that it is consistent not only for parameter sequences that are local-to-unity but also for those which characterize $I(0)$ processes. Moreover, its rate of convergence matches that of the POLS estimator for parameter sequences that are close to unity (more precisely, cases where $T/(L(T))^2 \ll q(T)$) whereas, for parameter sequences further away from unity (i.e., cases where $q(T)(L(T))^2/T = O(1)$), its convergence rate matches that of the IV estimator.[6]

3 Uniform Asymptotic Confidence Intervals

3.1 Confidence intervals based on the Anderson-Hsiao IV procedure

A primary objective of this paper is to develop confidence procedures with asymptotic coverage probability that is at least that of the nominal level uniformly over the parameter space $\rho \in (-1, 1]$. As a first step, we consider a statistic based on the empirical moment function of the Anderson-Hsiao IV procedure, but properly standardized by an appropriate estimator of the scale parameter. In particular, let

$$\mathbb{M}(\rho) = \frac{1}{\widehat{\omega}\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=3}^{T} y_{it-2} \left(\Delta y_{it} - \rho \Delta y_{it-1}\right),$$

where $\widehat{\omega}^2 = \widetilde{\sigma}^2 \left[N^{-1}T^{-1} \sum_{i=1}^{N} \sum_{t=4}^{T} (y_{it-3} - y_{it-2})^2 + N^{-1}T^{-1} \sum_{i=1}^{N} y_{iT-2}^2\right]$ and $\widetilde{\sigma}^2 = N^{-1}T_1^{-1} \sum_{i=1}^{N} \sum_{t=2}^{T} \left(y_{it} - \overline{y}_i - \widehat{\rho}_{\text{AIP}} \left[y_{it-1} - \overline{y}_{i,-1}\right]\right)$ and where $\overline{y}_i$, $\overline{y}_{i,-1}$, and $\widehat{\rho}_{\text{AIP}}$ are as defined in the previous section. The asymptotic properties of $\mathbb{M}(\rho)$ under different parameter sequences $\{\rho_T\}$ are given by the following result.

[6]In Appendix SF of the supplement to this paper, we provide additional Monte Carlo results comparing the finite sample performance of the AIP estimator with the bias-corrected within-group (BCWG) estimator of Hahn and Kuersteiner (2002), the POLS estimator, the Anderson-Hsiao IV estimator, the X-differencing estimator of Han, Phillips, and Sul (2014), and the Arellano-Bover IV estimator on the basis of median bias and 0.05-0.95 quantile range.

Theorem 3.1:

Let Assumptions 1-4 hold. The following statements hold as $N, T \to \infty$ such that $N^{\kappa}/T = \tau$, for $\kappa \in \left(\frac{1}{2}, \infty\right)$ and $\tau \in (0, \infty)$.

(a) *If* $\rho_T \in G_1^{\mathbb{M}} = \{\rho_T : \rho_T = 1 \text{ for all } T \text{ sufficiently large}\}$, *then*

$$\begin{aligned} & \mathbb{M}(\rho_T) \\ = & -\frac{1}{\sigma^2\sqrt{2}\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=4}^{T}\varepsilon_{it-2}\varepsilon_{it-1} + \frac{1}{\sigma^2\sqrt{2}\sqrt{NT}}\sum_{i=1}^{N} w_{iT-2}\varepsilon_{iT} \\ & + O_p\left(\max\left\{\frac{1}{\sqrt{N}}, \frac{1}{\sqrt{T}}\right\}\right) \\ \Rightarrow & N(0,1). \end{aligned}$$

(b) *If* $\rho_T \in \mathcal{G}_2^{\mathbb{M}} = \{\rho_T = \exp\{-1/q(T)\} : T/q(T) \to 0\}$, *then*

$$\begin{aligned} & \mathbb{M}(\rho_T) \\ = & -\frac{1}{\sigma^2\sqrt{2}\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=4}^{T}\varepsilon_{it-2}\varepsilon_{it-1} + \frac{1}{\sigma^2\sqrt{2}\sqrt{NT}}\sum_{i=1}^{N} w_{iT-2}\varepsilon_{iT} \\ & + O_p\left(\max\left\{\frac{1}{\sqrt{N}}, \frac{1}{\sqrt{T}}, \frac{T}{q(T)}\right\}\right) \\ \Rightarrow & N(0,1) \end{aligned}$$

(c) *Suppose* $\rho_T \in \mathcal{G}_3^{\mathbb{M}} = \{\rho_T = \exp\{-1/q(T)\} : q(T) \sim T\}$. *Then,*

$$\begin{aligned} & \mathbb{M}(\rho_T) \\ = & -\frac{1}{\omega_T\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=4}^{T}\varepsilon_{it-2}\varepsilon_{it-1} + \frac{1}{\omega_T\sqrt{NT}}\sum_{i=1}^{N} w_{iT-2}\varepsilon_{iT} \\ & + O_p\left(\max\left\{\frac{1}{\sqrt{N}}, \frac{1}{\sqrt{T}}\right\}\right) \\ \Rightarrow & N(0,1) \end{aligned}$$

where $\omega_T = \sigma^2\sqrt{1 + \frac{q(T)}{2T}\left[1 - \exp\left\{-\frac{2T}{q(T)}\right\}\right]}$.

(d) *If* $\rho_T \in \mathcal{G}_4^{\mathbb{M}} = \{\rho_T = \exp\{-1/q(T)\} : q(T) \to \infty \text{ such that } q(T)/T \to 0\}$, *then*

$$
\begin{aligned}
& \mathbb{M}\left(\rho_T\right) \\
= & -\frac{1}{\sigma^2 \sqrt{NT}} \sum_{i=1}^{N} \sum_{t=4}^{T} \varepsilon_{it-2} \varepsilon_{it-1} + O_p\left(\max\left\{\frac{1}{N}, \sqrt{\frac{q(T)}{T}}, \frac{1}{\sqrt{q(T)}}\right\}\right) \\
\Rightarrow & \; N(0,1).
\end{aligned}
$$

(e) *If*

$$
\rho_T \in \mathcal{G}_5^{\mathbb{M}} = \left\{ \left|\rho_T\right| = \exp\left\{-\frac{1}{q(T)}\right\} : q(T) \geq 0 \text{ and } q(T) = O(1) \text{ as } T \to \infty \right\},
$$

then

$$
\begin{aligned}
& \mathbb{M}\left(\rho_T\right) \\
= & -\sqrt{\frac{1+\rho_T}{2\sigma^4}} \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=4}^{T} \varepsilon_{it-2} \varepsilon_{it-1} \\
& + \sqrt{\frac{1+\rho_T}{2\sigma^4}} \frac{\left(1-\rho_T\right)}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=4}^{T} w_{it-3} \varepsilon_{it-1} + o_p(1) \\
\Rightarrow & \; N(0,1).
\end{aligned}
$$

Let $\mathbb{M}^*(\rho) = \widehat{\omega}\mathbb{M}(\rho)$ be the unstandardized version of $\mathbb{M}(\rho)$. It is evident from Theorem 3.1 and its proof (in the Appendix) that $\mathbb{M}^*(\rho)$ can be decomposed into several terms whose orders of magnitude change depending on how close the parameter sequence $\{\rho_T\}$ is to unity. In consequence, the lead term of $\mathbb{M}^*(\rho)$ is not the same in the stable (panel) autoregression case as it would be in the case where ρ_T is very close to unity. On the other hand, when appropriately normalized, this statistic will converge to a standard normal distribution in each case. But this requires an estimator that will adapt to variation in the normalization factor under alternative parameter sequences. The scale estimator $\widehat{\omega}$ turns out to have these adaptive properties, as shown in Lemma SD-13 given in the supplement to this paper.

The following theorem shows the uniform convergence of the statistic $\mathbb{M}(\rho)$ over the parameter space $\rho \in (-1, 1]$.

Theorem 3.2:

Let $\Phi(x)$ denote the cdf of a standard normal random variable. Suppose that Assumptions 1-4 hold. Then, for each $x \in \mathbb{R}$,

$$\sup_{\rho \in (-1,1]} \left| P_\rho \left(\mathbb{M}(\rho) \leq x \right) - \Phi(x) \right| \to 0,$$

as $N, T \to \infty$ such that $N^\kappa / T = \tau$ for constants $\kappa \in \left(\frac{1}{2}, \infty\right)$ and $\tau \in (0, \infty)$.

Remarks 3.1:

(i) Let z_α denote the $1 - \alpha$ quantile of the standard normal distribution. A level $1 - \alpha$ confidence interval based on the statistic $\mathbb{M}(\rho)$ can be taken to be

$$\mathbb{C}_\alpha^{\mathbb{M}} = \left\{ \rho \in (-1, 1] : -z_{\alpha/2} \leq \mathbb{M}(\rho) \leq z_{\alpha/2} \right\} \tag{6}$$

It is immediate from Theorem 3.2 that the confidence procedure defined by (6) is asymptotically valid in the sense that its coverage probability is equal to the nominal level $1 - \alpha$ in large samples, uniformly over the parameter space $\rho \in (-1, 1]$.

(ii) The uniform limit result given in Theorem 3.2 above is established under a pathwise asymptotic scheme where we take $N, T \to \infty$ such that $N^\kappa / T = \tau$ for constants $\kappa \in \left(\frac{1}{2}, \infty\right)$ and $\tau \in (0, \infty)$. Note that the asymptotic framework employed here does not restrict N and T to follow a specific diagonal expansion path, but rather allows a range of possible paths indexed by $\kappa \in \left(\frac{1}{2}, \infty\right)$. Given that many large sample results currently available in the panel data literature require some restriction on the relative orders of magnitude of N and T, such as $N/T \to \tau > 0$ or $N/T \to 0$, the asymptotic scheme we used here covers a wide range of practically important cases, allowing N to grow faster than T (as long as $\sqrt{N}/T \to 0$) as well as cases where T grows faster or at the same rate as N. In fact, under this framework, the order of magnitude of N can be an arbitrary small (positive) power of T, so that the range of situations for which our asymptotic approximation is useful comes close to the multivariate time series setting where N is fixed and only T goes to infinity.

(iii) As is evident from the proof of Theorem 3.2 given in the Appendix, uniform convergence is established using an approach discussed in Lehmann (1999) which demonstrates uniform convergence for a statistic by showing convergence for that statistic under every parameter sequence in the parameter space (see Lehmann, 1999, Lemma 2.6.2). Important recent extension and applications of this approach to a variety of econometric models and inferential procedures have been made in Andrews and Guggenberger (2009) and Andrews, Cheng, and Guggenberger (2011).

(iv) A primary reason why the $\mathbb{M}$ statistic is well-behaved is that the (empirical) IV moment function is well-centered as an unbiased estimating equation. In this sense, our approach relates to early work by Durbin (1960) on unbiased estimating equations which was applied to time series $AR(1)$ regression in his original study. Importantly, in dynamic panel data models with individual effects, estimating equations associated with least squares procedures tend not to be as well-centered as the IV estimating equations explaining the need for IV in this context (c.f., Han and Phillips, 2010).

(v) A drawback of $\mathbb{C}_\alpha^{\mathbb{M}}$ is that the rate at which the width of this confidence interval shrinks toward zero as sample sizes grow is relatively slow for parameter sequences that are very close to unity. This is due to the well-known 'weak instrument' problem which induces a slow rate of convergence for the Anderson-Hsiao IV procedure in this case. More precisely, using the results given in Lemmas SA-1, SD-1, and SD-13 in the supplement to this paper, we can easily show that $\text{wid}\left(\mathbb{C}_\alpha^{\mathbb{M}}\right) = O_p\left(T^{-1/2}\right)$ when $\rho_T = \exp\left\{-1/q\left(T\right)\right\}$ such that $\sqrt{N}T/q\left\{T\right\} = O\left(1\right)$, so that the rate of shrinkage here does not even depend on N, even as both N and T go to infinity (see also Phillips, 2015). This slower rate of convergence is also reflected in the Monte Carlo results reported in section 4 below, as the results there show that the average interval width of $\mathbb{C}_\alpha^{\mathbb{M}}$ is fairly wide when $\rho_0 = 1$. To improve on the performance of $\mathbb{C}_\alpha^{\mathbb{M}}$, the next subsection introduces a pretest-based confidence procedure which is similarly asymptotically valid but which in addition provides more informative intervals when the underlying process has a unit root or a near unit root.

3.2 A Pretest-Based Confidence Procedure

To enhance informativeness in the procedure when there is a unit root, we use a pretest approach. The idea is to apply two different unit root tests sequentially to assess the proximity of ρ to unity and then implement different confidence intervals depending on the information about the location of ρ that emerges from these tests. More precisely, we propose the following level $1-\alpha$ confidence interval of the form

$$\begin{aligned}\mathbb{C}_{\gamma,\alpha,N,T} &= \mathbb{I}\left\{\mathbb{T}_1 \leq -z_{\gamma_1}\right\}\mathbb{I}\left\{\mathbb{T}_2 \leq -z_{\gamma_2}\right\}\mathbb{C}^{\mathbb{M}}_{\alpha_1} + \mathbb{I}\left\{\mathbb{T}_1 > -z_{\gamma_1}\right\}\mathbb{C}^{\mathrm{UR1}}_{\gamma_1,\alpha_2} \\ &\quad + \mathbb{I}\left\{\mathbb{T}_1 \leq -z_{\gamma_1}\right\}\mathbb{I}\left\{\mathbb{T}_2 > -z_{\gamma_2}\right\}\mathbb{C}^{\mathrm{UR2}}_{\gamma_2,\alpha_2}, \end{aligned} \tag{7}$$

where $\mathbb{C}^{\mathbb{M}}_{\alpha_1}$ is as defined in (6) above,

$$\mathbb{C}^{\mathrm{UR1}}_{\gamma_1,\alpha_2} = \left\{\rho \in (-1,1] : 1 - \frac{\sqrt{2}\left(z_{\gamma_1} + z_{\alpha_2}\right)}{T\sqrt{N}} \leq \rho \leq 1\right\}, \text{ and} \tag{8}$$

$$\mathbb{C}^{\mathrm{UR2}}_{\gamma_2,\alpha_2} = \left\{\rho \in (-1,1] : 1 - \frac{2\left(z_{\gamma_2} + z_{\alpha_2}\right)}{\sqrt{NT}} \leq \rho \leq 1\right\} \tag{9}$$

and where $\gamma = (\gamma_1, \gamma_2)$, $\alpha = \alpha_1 + \alpha_2$, $\mathbb{I}$ is an indicator function, and we again take z_{γ_1} to be the $1-\gamma_1$ quantile of a standard normal distribution for some $\gamma_1 \in (0, 0.5]$, with z_{γ_2}, $z_{\alpha_1/2}$, and z_{α_2} similarly defined. In addition, we take

$$\mathbb{T}_1 = \frac{M_{yy}^{1/2}\left(\widehat{\rho}_{\mathrm{pols}} - 1\right)}{\widetilde{\sigma}},$$

with $M_{yy} = \sum_{i=1}^{N}\sum_{t=2}^{T}\left(y_{it-1} - \overline{y}_{-1,NT}\right)^2$, to be the unit root test statistic based on the POLS estimator; and

$$\mathbb{T}_2 = \widehat{\omega}_{\mathrm{IVL}}\left(\widehat{\rho}_{\mathrm{IVL}} - 1\right),$$

with $\widehat{\omega}_{\mathrm{IVL}} = \widetilde{\sigma}^{-2}N^{-1/2}T^{-1/2}\sum_{i=1}^{N}\sum_{t=3}^{T}\Delta y_{it-1}y_{it-1}$, is a unit root test statistic based on the IV estimator

$$\widehat{\rho}_{\mathrm{IVL}} = \left[\sum_{i=1}^{N}\sum_{t=3}^{T}\Delta y_{it-1}y_{it-1}\right]^{-1}\sum_{i=1}^{N}\sum_{t=3}^{T}\Delta y_{it-1}y_{it}$$

which was introduced by Arellano and Bover (1995) and further analyzed in Blundell and Bond (1998). From expression (7), it is apparent that the confidence procedure follows a sequential tree structure. We first pretest for the presence of a unit root using $\mathbb{T}_1$. If the result of this first test fails to reject the unit root null hypothesis, then we employ the tighter unit root interval $\mathbb{C}^{\text{UR1}}_{\gamma_1,\alpha_2}$. Otherwise, we conduct a second test of the unit root null hypothesis using a less powerful test $\mathbb{T}_2$. If this second test fails to reject the null hypothesis, we use the wider unit root interval $\mathbb{C}^{\text{UR2}}_{\gamma_2,\alpha_2}$. On the other hand, if both tests reject the unit root null hypothesis, we then use the interval $\mathbb{C}^{\mathbb{M}}_{\alpha_1}$, which is asymptotically valid but less informative unless the true value of ρ is sufficiently far away from unity.

The next theorem shows that this confidence procedure is asymptotically valid in the sense that its non-converage probability is at most the nominal significance level α uniformly over the parameter space under pathwise asymptotics.

Theorem 3.3:

Let $\alpha \in (0, 0.5]$ be the specified significance level and let $N, T \to \infty$ such that $N^{\kappa}/T = \tau$ for constants $\kappa \in \left(\frac{1}{2}, \infty\right)$ and $\tau \in (0, \infty)$. Set $N = N(T) = (\tau T)^{1/\kappa}$ and $C_{\gamma,\alpha,N,T} = C_{\gamma,\alpha,N(T),T} = C_{\alpha,T}$. Then, under Assumptions 1-4,

$$\limsup_{T\to\infty} \sup_{\rho^0 \in (-1,1]} \Pr\left(\rho \notin \mathbb{C}_{\gamma,\alpha,T} | \rho = \rho^0\right) \leq \alpha.$$

Remarks 3.2:

(i) The pre-test based confidence procedure proposed here is inspired by work of Lepski (1999) who used information from a test procedure to increase the accuracy of confidence sets. The original Lepski paper and subsequent extensions of that paper focused on problems of nonparametric function estimation and canonical versions of such problems, as represented by the many normal means model. Because we deal with a model different from the one studied in Lepski (1999), the construction and analysis of our procedure also differ, even though we use the same idea to improve set estimation accuracy.

(ii) Since

$$\begin{aligned}
& \limsup_{T\to\infty} \sup_{\rho^0\in(-1,1]} \Pr\left(\rho \notin \mathbb{C}_{\gamma,\alpha,T} | \rho = \rho^0\right) \\
= & \limsup_{T\to\infty} \left[1 - \inf_{\rho^0\in(-1,1]} \Pr\left(\rho \in \mathbb{C}_{\gamma,\alpha,T} | \rho = \rho^0\right)\right] \\
= & 1 - \liminf_{T\to\infty} \inf_{\rho^0\in(-1,1]} \Pr\left(\rho \in \mathbb{C}_{\gamma,\alpha,T} | \rho = \rho^0\right),
\end{aligned}$$

it follows that the result obtained in Theorem 3.3, i.e., $\limsup_{T\to\infty} \sup_{\rho^0\in(-1,1]} \Pr\left(\rho \notin \mathbb{C}_{\gamma,\alpha,T} | \rho = \rho^0\right) \leq \alpha$, is equivalent to $\liminf_{T\to\infty} \inf_{\rho^0\in(-1,1]} \Pr\left(\rho \in \mathbb{C}_{\gamma,\alpha,T} | \rho = \rho^0\right) \geq 1-\alpha$, so that the proposed confidence interval has asymptotic coverage probability that is at least the nominal level $1-\alpha$ uniformly over the parameter space $\rho^0 \in (-1,1]$.

(iii) In the procedure given by (7), α_1 is the significance level for the confidence interval $\mathbb{C}^{\mathbb{M}}_{\alpha_1}$. It is, of course, also the asymptotic non-coverage probability of $\mathbb{C}^{\mathbb{M}}_{\alpha_1}$, since $\mathbb{C}^{\mathbb{M}}_{\alpha_1}$ is asymptotically valid.

(iv) As noted in Remark 3.1 (v) above, a drawback of $\mathbb{C}^{\mathbb{M}}_{\alpha_1}$ is that its width shrinks slowly for parameter sequences that are very close to unity. The pre-test confidence procedure seeks to improve on this rate by applying two different unit root tests sequentially and by using the information from these tests to determine whether to use local-to-unity intervals whose width shrinks at a faster rate than $\mathbb{C}^{\mathbb{M}}_{\alpha_1}$ when the autoregressive parameter value is in close proximity to unity. To see how this improvement is achieved, note that when the true parameter value is within an $N^{-1/2}T^{-1}$ neighborhood of unity then, aside from the relatively small probability event of a Type I error, the first unit root test $\mathbb{T}_1$ will fail to reject $H_0 : \rho_0 = 1$, resulting in the use of the interval $\mathbb{C}^{\mathrm{UR1}}_{\gamma_1,\alpha_2}$. When the parameter is this close to unity, $\mathrm{wid}\left(\mathbb{C}^{\mathrm{UR1}}_{\gamma_1,\alpha_2}\right) = O_p\left(N^{-1/2}T^{-1}\right)$ whereas $\mathrm{wid}\left(\mathbb{C}^{\mathbb{M}}_{\alpha_1}\right) = O_p\left(T^{-1/2}\right)$, so that the use of $\mathbb{C}^{\mathrm{UR1}}_{\gamma_1,\alpha_2}$ leads to significant improvement over $\mathbb{C}^{\mathbb{M}}_{\alpha_1}$. The reason for a second unit root test using the statistic $\mathbb{T}_2$ is that for parameter sequences $\rho_T = \exp\{-1/q\{T\}\}$ such that $\max\left\{T, \sqrt{NT}\right\} \ll q(T) \ll \sqrt{N}T$, the first unit root test $\mathbb{T}_1$ will reject H_0 with probability approaching one as sample sizes grow, but the less powerful unit root test based on $\mathbb{T}_2$ will not, subject again to the relatively small probability event of a Type I error. For parameter sequences in this region, $\mathrm{wid}\left(\mathbb{C}_{\alpha_1}^{\mathbb{M}}\right) = O_p\left(N^{-1/2}T^{-3/2}q\left(T\right)\right)$. The result is that we can make further improvement by using the interval $\mathbb{C}_{\gamma_2,\alpha_2}^{\mathrm{UR2}}$ which has width $\mathrm{wid}\left(\mathbb{C}_{\gamma_2,\alpha_2}^{\mathrm{UR2}}\right) = O_p\left(N^{-1/2}T^{-1/2}\right) = o_p\left(N^{-1/2}T^{-3/2}q\left(T\right)\right)$. Finally, if both these unit root tests reject H_0, then our procedure will infer that the parameter is far enough away from unity to use $\mathbb{C}_{\alpha_1}^{\mathbb{M}}$. Of course, the two unit root tests are subject to Type II errors; but, as explained in Remark 3.2(vi) below, the probability of Type II errors could also be properly controlled under our procedure.[7]

(v) γ_1 and γ_2, on the other hand, are the significance levels for the unit root tests based on $\mathbb{T}_1$ and $\mathbb{T}_2$. Note that, especially in large samples, the specification of γ_1 and γ_2 really has more of an impact on the width of the resulting interval than it does on the coverage probability, so that γ_1 and γ_2 are not significance levels in the traditional sense. For example, consider the choice of γ_1. Observe that a smaller value of γ_1 leads to a wider $\mathbb{C}_{\gamma_1,\alpha_2}^{\mathrm{UR1}}$. However, the effect of γ_1 on the width of the interval adopted by the overall procedure could be ambiguous, since, if the null hypothesis of an exact unit root is true, an increase in γ_1 would reduce the width of $\mathbb{C}_{\gamma_1,\alpha_2}^{\mathrm{UR1}}$ but could also lead to a greater chance that $\mathbb{T}_1$ will falsely reject the null hypothesis and switch to either $\mathbb{C}_{\gamma_2,\alpha_2}^{\mathrm{UR2}}$ or $\mathbb{C}_{\alpha_1}^{\mathbb{M}}$, both of which are wider than $\mathbb{C}_{\gamma_1,\alpha_2}^{\mathrm{UR1}}$ in large samples. A similar argument shows that it is also difficult to predict *a priori* the effect of varying γ_2 on the width of the resulting interval. On the other hand, note that, except for pathological specifications where $\gamma_1 = 0$ and/or $\gamma_2 = 0$ (ruled out by our assumption), varying either γ_1 or γ_2 or both does not lead to a material distortion in the (asymptotic) coverage probability of the proposed procedure. To see why this is so, consider the case where the unit root specification is true. Then, even

[7] An interesting recent paper by Bun and Kleibergen (2014) also considers, amongst other things, combining elements of the approach of Anderson and Hsiao (1981, 1982) and Arellano and Bond (1991), which uses lagged levels of y_{it} as instruments for equations in first differences, with the approach by Arellano and Bover (1995) and Blundell and Bond (1998) which uses lagged differences of y_{it} as instruments for equations in levels. However, the focus of the Bun and Kleibergen (2014) paper differs substantially from that of our paper. In particular, they consider test procedures which attain the maximal attainable power curve under worst case setting of the variance of the initial conditions, whereas our procedure uses pretest based information to aggressively increase the power of our procedure in certain regions of the parameter space. Moreover, unlike our paper, they do not provide results on confidence procedures whose asymptotic coverage probability is explicitly shown to be at least that of the nominal level uniformly over the parameter space; and their analysis is conducted within a fixed T framework.

when both γ_1 and γ_2 are set to be large so that the null hypothesis is falsely rejected with high probability leading to the use of $\mathbb{C}_{\alpha_1}^{\mathbb{M}}$, we will still end up with asymptotic coverage probability greater than the nominal level $1-\alpha$ since $\mathbb{C}_{\alpha_1}^{\mathbb{M}}$ is asymptotically valid and, by design, $\alpha_1 < \alpha = \alpha_1 + \alpha_2$. On the other hand, if the underlying process is stable then, both of the unit root tests will reject the null hypothesis with probability approaching one asymptotically, as long as neither γ_1 nor γ_2 is set equal to zero, and our procedure will switch to $\mathbb{C}_{\alpha_1}^{\mathbb{M}}$ which controls the asymptotic coverage probability properly.

(vi) Pre-testing leads to the possibility of errors whose probability needs to be controlled. In particular, there may be parameter sequences which lie just outside of $\mathbb{C}_{\gamma_1,\alpha_2}^{\mathrm{UR1}}$, for which $\mathbb{T}_1$ may fail to reject $H_0 : \rho_0 = 1$ even in large samples. In addition, there may be parameter sequences which lie just outside of $\mathbb{C}_{\gamma_2,\alpha_2}^{\mathrm{UR2}}$, for which H_0 is rejected by $\mathbb{T}_1$ but for which $\mathbb{T}_2$ may not reject H_0 even in large samples. In both of these scenarios, there is the possibility that none of our intervals will cover the true parameter sequence. However, in the proof of Lemma A1 given in the Appendix SB of the technical supplement, we show that, under our procedure, the probability of committing such Type II errors can be no greater than α_2 asymptotically[8]. Hence, by constructing $\mathbb{C}_{\gamma_1,\alpha_2}^{\mathrm{UR1}}$ and $\mathbb{C}_{\gamma_2,\alpha_2}^{\mathrm{UR2}}$ in the manner suggested above, we can properly control the probability of not switching to $\mathbb{C}_{\alpha_1}^{\mathbb{M}}$ when it is preferable to make that switch. In consequence, the asymptotic non-coverage probability under our procedure is always less than or equal to $\alpha = \alpha_1 + \alpha_2$. Given a particular significance level α, different combinations of α_1 and α_2 involve trade-offs where a smaller α_2 leads to a smaller probability of commtting a Type II error but also leads to a larger α_1 and, thus, to $\mathbb{C}_{\alpha_1}^{\mathbb{M}}$ having a smaller asymptotic coverage probability.

(vii) An advantage of our pretest based confidence procedure is its computational simplicity, as it is given in analytical form and, thus, does not require the use of bootstrap or other types of simulation-based methods for its computation. Moreover, the fact that $\mathbb{C}_{\alpha_1}^{\mathbb{M}}$, the interval used under our procedure in the stable case, is based on the Anderson-Hsiao procedure has the further benefit that its validity does not depend on imposing the assumption of mean stationarity of

[8]It should be noted that Lemma A1 itself is given in the Appendix of the main paper, but its proof is rather lengthy; and, hence, we have placed it in the technical supplement.

the initial condition. Hence, the design of our procedure has involved certain trade-offs on the competing goals of interval accuracy, computational simplicity, and relaxation of the assumption of initial condition stationarity.

4 Monte Carlo Study

This section reports the results of a Monte Carlo study comparing the finite sample performance of alternative confidence procedures. For the simulation study, we consider data generating processes of the form

$$\begin{aligned} y_{it} &= a_i + w_{it}, \\ w_{it} &= \rho_0 w_{it-1} + \varepsilon_{it}, \text{ for } i = 1, ..., N \text{ and } t = 1, ..., T; \end{aligned}$$

where $\{\varepsilon_{it}\} \equiv i.i.d.N(0,1)$ and $\{a_i\} \equiv i.i.d.N(2,1)$. We vary $\rho_0 = 1.00$, 0.99, 0.95, 0.90, 0.80, and 0.60 and $w_{i0} = 0, 2$. In addition, we let $N = 100, 200$. When $N = 100$, we take $T = 50$, 100; and when $N = 200$, we consider $T = 100$, 200. We take $\alpha = 0.05$ throughout, so that the (nominal) confidence level is always kept at 95%. Four versions of the pre-test based confidence interval (PCI) given by expression (7) above are considered, with different specifications of γ_1, γ_2, α_1, and α_2, as summarized in the following table:

	γ_1	γ_2	α_1	α_2
$\mathbb{C}_{\text{PCI1}}$	0.01	0.01	0.025	0.025
$\mathbb{C}_{\text{PCI2}}$	0.01	0.01	0.04	0.01
$\mathbb{C}_{\text{PCI3}}$	0.05	0.05	0.025	0.025
$\mathbb{C}_{\text{PCI4}}$	0.05	0.05	0.04	0.01

Tables 1-8 below provide simulation results comparing the four PCI procedures described above with the $\mathbb{C}_{0.05}^{\mathbb{M}}$ procedure given in (6) and with confidence intervals obtained by inverting Studentized statistics associated with the POLS and IVD estimators. More specifically, Tables 1-4 give the empirical coverage probabilities while Tables 5-8 report the average width of the confidence intervals under each of forty-eight experimental setttings, obtained by varying ρ_0, N, T, and w_{i0}. Glancing at Tables 1-4, we see that, consistent with our theory, the empirical coverage probabilities of the the $\mathbb{C}^{\mathbb{M}}_{0.05}$ procedure show the greatest degree of uniformity across different experiments. A clear deficiency of the $\mathbb{C}^{\mathbb{M}}_{0.05}$ procedure as shown in Tables 5-8 is that the average widths of these intervals are much wider than that of the other procedures when $\rho_0 = 1$. Outside of the unit root case, however, the results of Tables 5-8 do show $\mathbb{C}^{\mathbb{M}}_{0.05}$ to be informative.

Turning our attention to the pre-test based procedures, we see that all four PCIs have empirical coverage probabilities that are uniformly better than the $\mathbb{C}^{\mathbb{M}}_{0.05}$ procedure across all forty-eight experiments. An intuitive explanation for this result can be given as follows. When the unit root null hypothesis is true, application of the pre-test procedure will lead to the use of either $\mathbb{C}^{UR1}_{\gamma_1,\alpha_2}$ or $\mathbb{C}^{UR2}_{\gamma_2,\alpha_2}$, except in the small probability event where a Type I error is committed by both of the unit root tests $\mathbb{T}_1$ and $\mathbb{T}_2$. Since both of these intervals cover the point $\rho_0 = 1$ by construction, the overall procedure in this case should cover this point with very high probability. On the other hand, when the unit root hypothesis is false, the pre-test procedure switches to the interval $\mathbb{C}^{\mathbb{M}}_{\alpha_1}$ but with α_1 set at a level strictly less than 0.05, resulting again in coverage probabilities which are greater than that of the $\mathbb{C}^{\mathbb{M}}_{0.05}$ procedure.

From the reported simulation results, it does not seem that there are great differences in the performance of the four alternative specifications of PCI, although some minor trade-offs in coverage probability vis-à-vis average interval width can be discerned. For example, looking at PCI1, we see that this procedure provides very tight intervals in the case where $\rho_0 = 1$. In fact, the average interval width for this procedure in the unit root case is ≤ 0.0070, except in the smaller sample size case with $N = 100$ and $T = 50$, where it is still around 0.0133. Moreover, amongst the seven procedures examined in our study, the empirical coverage probability of PCI1 is the highest, or is at least tied for the highest, almost across the board, for the 48 experiments whose results are reported in Tables 1-4. Although the higher coverage probability of PCI1 in the stable region is due at least in part to the fact that it is designed to be conservative with $\alpha_1 = 0.025$ when the true process is stable, it should be noted that the informativeness of PCI1, as measured by its average width, does not seem to have suffered significantly as a result. Note, in particular, that, over the 48 experiments, the widest average interval width recorded for PCI1 was only 0.1474, or approximately 7% of the width of the entire parameter space $(-1, 1]$; and this occurred with the smaller sample sizes of $N = 100$ and $T = 50$. In addition, PCI1 has average width strictly less than 0.1 in 38 of the experiments. On the other hand, PCI2 sets $\alpha_1 = 0.04$ and is, thus, less conservative relative to PCI1, particularly in the stable region. In consequence, PCI2 tends to have not only smaller interval widths but also lower coverage probabilities relative to PCI1 when the underlying process is stable. The results for PCI1 and PCI2 are illustrative of how the pre-test procedures can greatly improve upon $\mathbb{C}^{\mathbb{M}}_{0.05}$ in terms of accuracy in the unit root and near unit root cases while maintaining coverage probability at a high level throughout the parameter space, with the only trade-off being that they yield slightly wider intervals when the true process is stable.

Tables 1-4 also show that confidence intervals constructed by inverting Studentized statistics associated with $\widehat{\rho}_{\text{POLS}}$ and $\widehat{\rho}_{\text{IVD}}$ are decidedly inferior to the pre-test based confidence procedures. Consistent with our theory, Tables 1-4 show that these confidence intervals have highly non-uniform coverage probabilities across different (true) parameter values ρ_0. More specifically, the coverage probabilities of the IV-based confidence intervals are especially poor when ρ_0 is unity or near-unity, whereas the coverage probabilities for the POLS-based confidence intervals begin to deviate dramatically from the nominal level when $\rho_0 = 0.95$ or less.

Table 1: Coverage Probabilities (nominal level=0.95)
$N = 100$, $w_{i0} = 0$

ρ_0	T	CI_{POLS}	CI_{IVD}	$CI_{\mathbb{M}}$	CI_{PCI1}	CI_{PCI2}	CI_{PCI3}	CI_{PCI4}
1.00	50	0.9490	0.1229	0.9430	1.0000	1.0000	0.9999	0.9998
1.00	100	0.9518	0.1251	0.9411	1.0000	1.0000	0.9999	0.9999
0.99	50	0.9476	0.3874	0.9385	0.9957	0.9943	0.9891	0.9857
0.99	100	0.9443	0.6239	0.9448	0.9918	0.9874	0.9872	0.9802
0.95	50	0.7995	0.8046	0.9369	0.9839	0.9733	0.9839	0.9733
0.95	100	0.6816	0.8911	0.9445	0.9874	0.9791	0.9874	0.9791
0.90	50	0.2384	0.8738	0.9376	0.9833	0.9713	0.9758	0.9585
0.90	100	0.0507	0.9223	0.9465	0.9715	0.9560	0.9715	0.9560
0.80	50	0.0002	0.9055	0.9378	0.9677	0.9488	0.9677	0.9488
0.80	100	0.0000	0.9254	0.9421	0.9705	0.9522	0.9705	0.9522
0.60	50	0.0000	0.9162	0.9351	0.9650	0.9485	0.9650	0.9485
0.60	100	0.0000	0.9335	0.9425	0.9700	0.9549	0.9700	0.9549

Results based on 10,000 simulations

Table 2: Coverage Probabilities (nominal level=0.95)
$N = 100$, $w_{i0} = 2$

ρ_0	T	CI_{POLS}	CI_{IVD}	$CI_{\mathbb{M}}$	CI_{PCI1}	CI_{PCI2}	CI_{PCI3}	CI_{PCI4}
1.00	50	0.9490	0.1345	0.9311	1.0000	1.0000	0.9998	0.9996
1.00	100	0.9518	0.1285	0.9339	1.0000	1.0000	0.9999	0.9999
0.99	50	0.9493	0.4537	0.9226	0.9947	0.9938	0.9878	0.9830
0.99	100	0.9449	0.6648	0.9370	0.9910	0.9875	0.9856	0.9783
0.95	50	0.8056	0.8720	0.9164	0.9752	0.9651	0.9748	0.9640
0.95	100	0.6864	0.9172	0.9326	0.9842	0.9757	0.9839	0.9740
0.90	50	0.2498	0.9227	0.9150	0.9715	0.9560	0.9624	0.9428
0.90	100	0.0546	0.9376	0.9359	0.9632	0.9461	0.9634	0.9461
0.80	50	0.0002	0.9353	0.9198	0.9567	0.9333	0.9567	0.9333
0.80	100	0.0000	0.9393	0.9313	0.9644	0.9454	0.9644	0.9454
0.60	50	0.0000	0.9357	0.9225	0.9563	0.9387	0.9563	0.9387
0.60	100	0.0000	0.9414	0.9358	0.9657	0.9499	0.9657	0.9499

Results based on 10,000 simulations

Table 3: Coverage Probabilities (nominal level=0.95)
$N = 200$, $w_{i0} = 0$

ρ_0	T	CI_{POLS}	CI_{IVD}	$CI_{\mathbb{M}}$	CI_{PCI1}	CI_{PCI2}	CI_{PCI3}	CI_{PCI4}
1.00	100	0.9494	0.0921	0.9455	1.0000	1.0000	0.9998	0.9997
1.00	200	0.9458	0.0879	0.9499	0.9999	0.9999	0.9998	0.9998
0.99	100	0.9468	0.6346	0.9482	0.9875	0.9815	0.9856	0.9783
0.99	200	0.9409	0.8101	0.9483	0.9867	0.9797	0.9863	0.9790
0.95	100	0.4377	0.8949	0.9436	0.9844	0.9758	0.9782	0.9647
0.95	200	0.1796	0.9243	0.9444	0.9736	0.9567	0.9736	0.9567
0.90	100	0.0010	0.9186	0.9451	0.9705	0.9539	0.9705	0.9539
0.90	200	0.0000	0.9318	0.9434	0.9698	0.9541	0.9698	0.9541
0.80	100	0.0000	0.9320	0.9447	0.9740	0.9569	0.9740	0.9569
0.80	200	0.0000	0.9353	0.9422	0.9715	0.9532	0.9715	0.9532
0.60	100	0.0000	0.9368	0.9452	0.9707	0.9559	0.9707	0.9559
0.60	200	0.0000	0.9439	0.9482	0.9732	0.9570	0.9732	0.9570

Results based on 10,000 simulations

Table 4: Coverage Probabilities (nominal level=0.95) $N = 200$, $w_{i0} = 2$								
ρ_0	T	CI_{POLS}	CI_{IVD}	$CI_{\mathbb{M}}$	CI_{PCI1}	CI_{PCI2}	CI_{PCI3}	CI_{PCI4}
1.00	100	0.9494	0.0958	0.9370	1.0000	1.0000	1.0000	0.9998
1.00	200	0.9458	0.0911	0.9468	0.9999	0.9999	0.9998	0.9998
0.99	100	0.9471	0.6731	0.9398	0.9850	0.9800	0.9824	0.9758
0.99	200	0.9421	0.8297	0.9441	0.9857	0.9791	0.9847	0.9774
0.95	100	0.4523	0.9170	0.9334	0.9806	0.9688	0.9730	0.9566
0.95	200	0.1850	0.9325	0.9401	0.9702	0.9515	0.9702	0.9515
0.90	100	0.0009	0.9351	0.9341	0.9622	0.9458	0.9622	0.9458
0.90	200	0.0000	0.9420	0.9411	0.9703	0.9527	0.9703	0.9527
0.80	100	0.0000	0.9436	0.9364	0.9683	0.9483	0.9683	0.9483
0.80	200	0.0000	0.9433	0.9398	0.9678	0.9498	0.9678	0.9498
0.60	100	0.0000	0.9419	0.9374	0.9669	0.9496	0.9669	0.9496
0.60	200	0.0000	0.9469	0.9447	0.9710	0.9542	0.9710	0.9542

Results based on 10,000 simulations

Table 5: Average Width of Confidence Intervals $N = 100$, $w_{i0} = 0$								
ρ_0	T	CI_{POLS}	CI_{IVD}	$CI_{\mathbb{M}}$	CI_{PCI1}	CI_{PCI2}	CI_{PCI3}	CI_{PCI4}
1.00	50	0.0111	0.1172	19.5356	0.0133	0.0144	0.0167	0.0180
1.00	100	0.0056	0.0817	5.4271	0.0070	0.0076	0.0093	0.0100
0.99	50	0.0127	0.1109	0.5139	0.1020	0.1089	0.1363	0.1402
0.99	100	0.0073	0.0782	0.1713	0.0910	0.0972	0.0932	0.0965
0.95	50	0.0184	0.1095	0.1558	0.1448	0.1446	0.1529	0.1463
0.95	100	0.0123	0.0774	0.0919	0.0985	0.0951	0.1004	0.0939
0.90	50	0.0234	0.1081	0.1305	0.1474	0.1364	0.1486	0.1364
0.90	100	0.0161	0.0764	0.0838	0.0958	0.0878	0.0958	0.0878
0.80	50	0.0297	0.1052	0.1171	0.1339	0.1227	0.1339	0.1227
0.80	100	0.0207	0.0744	0.0784	0.0897	0.0821	0.0897	0.0821
0.60	50	0.0369	0.0992	0.1057	0.1209	0.1107	0.1209	0.1107
0.60	100	0.0259	0.0701	0.0724	0.0828	0.0758	0.0828	0.0758

Results based on 10,000 simulations

Table 6: Average Width of Confidence Intervals $N = 100$, $w_{i0} = 2$								
ρ_0	T	CI_{POLS}	CI_{IVD}	$CI_{\mathbb{M}}$	CI_{PCI1}	CI_{PCI2}	CI_{PCI3}	CI_{PCI4}
1.00	50	0.0111	0.1176	8.5408	0.0132	0.0143	0.0157	0.0170
1.00	100	0.0056	0.0819	6.2519	0.0070	0.0076	0.0091	0.0098
0.99	50	0.0127	0.1107	0.3590	0.1023	0.1093	0.1278	0.1325
0.99	100	0.0073	0.0782	0.1504	0.0898	0.0961	0.0887	0.0924
0.95	50	0.0182	0.1095	0.1236	0.1286	0.1308	0.1266	0.1231
0.95	100	0.0122	0.0774	0.0818	0.0906	0.0884	0.0901	0.0847
0.90	50	0.0230	0.1081	0.1063	0.1215	0.1133	0.1211	0.1114
0.90	100	0.0160	0.0764	0.0757	0.0866	0.0794	0.0866	0.0793
0.80	50	0.0292	0.1052	0.0995	0.1137	0.1042	0.1137	0.1042
0.80	100	0.0206	0.0744	0.0722	0.0826	0.0757	0.0826	0.0757
0.60	50	0.0366	0.0992	0.0954	0.1091	0.0999	0.1091	0.0999
0.60	100	0.0258	0.0701	0.0688	0.0786	0.0720	0.0786	0.0720

Results based on 10,000 simulations

Table 7: Average Width of Confidence Intervals $N = 200$, $w_{i0} = 0$								
ρ_0	T	CI_{POLS}	CI_{IVD}	$CI_{\mathbb{M}}$	CI_{PCI1}	CI_{PCI2}	CI_{PCI3}	CI_{PCI4}
1.00	100	0.0039	0.0582	7.7623	0.0048	0.0052	0.0070	0.0075
1.00	200	0.0020	0.0408	5.0962	0.0027	0.0029	0.0039	0.0042
0.99	100	0.0051	0.0553	0.1201	0.0678	0.0714	0.0730	0.0739
0.99	200	0.0032	0.0391	0.0585	0.0461	0.0485	0.0464	0.0470
0.95	100	0.0087	0.0547	0.0649	0.0734	0.0679	0.0739	0.0679
0.95	200	0.0060	0.0387	0.0421	0.0481	0.0441	0.0481	0.0441
0.90	100	0.0114	0.0540	0.0592	0.0677	0.0621	0.0677	0.0621
0.90	200	0.0080	0.0382	0.0400	0.0457	0.0419	0.0457	0.0419
0.80	100	0.0147	0.0526	0.0554	0.0634	0.0581	0.0634	0.0581
0.80	200	0.0103	0.0372	0.0382	0.0437	0.0400	0.0437	0.0400
0.60	100	0.0183	0.0496	0.0512	0.0585	0.0536	0.0585	0.0536
0.60	200	0.0129	0.0351	0.0356	0.0407	0.0373	0.0407	0.0373

Results based on 10,000 simulations

Table 8: Average Width of Confidence Intervals $N = 200$, $w_{i0} = 2$								
ρ_0	T	CI_{POLS}	CI_{IVD}	$CI_{\mathbb{M}}$	CI_{PCI1}	CI_{PCI2}	CI_{PCI3}	CI_{PCI4}
1.00	100	0.0039	0.0575	5.9026	0.0048	0.0052	0.0069	0.0073
1.00	200	0.0020	0.0403	3.5947	0.0027	0.0029	0.0039	0.0042
0.99	100	0.0051	0.0553	0.1056	0.0661	0.0698	0.0683	0.0696
0.99	200	0.0031	0.0391	0.0547	0.0454	0.0479	0.0448	0.0455
0.95	100	0.0086	0.0547	0.0578	0.0657	0.0610	0.0658	0.0605
0.95	200	0.0060	0.0387	0.0398	0.0455	0.0417	0.0455	0.0417
0.90	100	0.0113	0.0540	0.0535	0.0612	0.0561	0.0612	0.0561
0.90	200	0.0079	0.0382	0.0380	0.0435	0.0398	0.0435	0.0398
0.80	100	0.0145	0.0526	0.0511	0.0584	0.0535	0.0584	0.0535
0.80	200	0.0103	0.0372	0.0366	0.0419	0.0384	0.0419	0.0384
0.60	100	0.0182	0.0496	0.0486	0.0556	0.0509	0.0556	0.0509
0.60	200	0.0129	0.0351	0.0347	0.0397	0.0364	0.0397	0.0364

Results based on 10,000 simulations

5 Conclusion

The uniform inference procedure proposed here utilizes information from pretesting the unit root hypothesis to aid the construction of confidence intervals in panel autoregression by means of data-based selection among intervals that are well suited to particular regions of the parameter space. The construction is asymptotically valid in the sense that the large sample coverage probability is at least that of the nominal level uniformly over the parameter space. The method is particularly simple to implement in practical work and simulations provide encouraging evidence that the method produces confidence intervals with good finite sample accuracy, as measured by the combination of empirical coverage probability and average interval width.

References

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

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

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

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

[5] Andrews, D.W.K. (1993). "Exactly Median-Unbiased Estimation of First-Order Autoregressive/Unit Root Models", *Econometrica*, 61, 139-165.

[6] Andrews, D.W.K. and P. Guggenberger (2009). "Validity of Subsampling and 'Plug-in Asymptotic' Inference for Parameters Defined by Moment Inequalities", *Econometric Theory*, 25, 669-709.

[7] Andrews, D.W.K., X. Cheng, and P. Guggenberger (2011). "Generic Results for Establishing the Asymptotic Size of Confidence Sets and Tests", Cowles Foundation Discussion Paper No. 1813.

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

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

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

[11] Bun, M. J. G. and F. Kleibergen (2014). "Identication and Inference in Moments Based Analysis of Linear Dynamic Panel Data Models", UvA Econometrics Discussion Paper.

[12] Chao, J. and P. C. B. Phillips (2016). "Technical Supplement to 'Uniform Inference in Panel Autoregression' ", Working Paper, University of Maryland.

[13] Durbin, J. (1960). "Estimating of Parameters in Time Series Regression Models", *Journal of the Royal Statistical Society*, *Series B*, 22, 139-153.

[14] Gorodnichenko, Y., A. Mikusheva, and S. Ng (2012). "Estimators for Persistent and Possibly Nonstationary Data with Classical Properties", *Econometric Theory*,28, 1003-1036.

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

[16] Han, C. and P. C. B. Phillips (2010). "GMM Estimation for Dynamic Panels with Fixed Effects and Strong Instruments at Unity", *Econometric Theory*, 26, 119-151.

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

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

[19] Hansen, B.E. (1999). "The Grid Bootstrap and the Autoregressive Model", *Review of Economics and Statistics*, 81, 594-607.

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

[21] Lehmann, E. L. (1999). *Elements of Large Sample Theory* (New York: Springer).

[22] Lepski, O.V. (1999). "How to Improve the Accuracy of Estimation", *Mathematical Methods in Statistics*, 8, 441-486.

[23] Mikusheva, A. (2007). "Uniform Inference in Autoregressive Models", *Econometrica*, 75, 1411-1452.

[24] Mikusheva, A. (2012) "One-dimensional Inference in Autoregressive Models with the Potential Presence of a Unit Root", *Econometrica*, 80, 173-212.

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

[26] Moon, H. R., B. Perron and P. C. B. Phillips (2014a). "Incidental Parameters and Dynamic Panel Modeling" in B. Baltagi (Ed.) Handbook of Panel Data, Oxford University Press (forthcoming).

[27] Moon, H. R., B. Perron and P. C. B. Phillips (2014b). "Point Optimal Panel Unit Root Tests with Serially Correlated Data", *Econometrics Journal* , 17, 338-372.

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

[29] Pesaran, M. H. (2006). "Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure", *Econometrica*, 74, 967-1012.

[30] Pesaran, M. H. and E. Tosetti (2011). "Large Panels with Common Factors and Spatial Correlation", *Journal of Econometrics*, 161, 182-202.

[31] Phillips, P. C. B. (1987). "Time Series Regression with a Unit Root", *Econometrica*, 55, 277-301.

[32] Phillips, P. C. B. (2014). “On Confidence Intervals for Autoregressive Roots and Predictive Regression”, *Econometrica*, 82, 1177-1195.

[33] Phillips, P. C. B. (2015). “Dynamic Panel Anderson-Hsiao Estimation With Roots Near Unity,”. *Econometric Theory*, first published online, September 2015, pp. 1-24. DOI: https://doi.org/10.1017/S0266466615000298.

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

[35] Phillips, P. C. B. and D. Sul (2003). “Dynamic Panel Estimation and Homogeneity Testing under Cross Sectional Dependence”, *Econometrics Journal,* 6, 217-259.

[36] Stock, J. (1999). “Confidence Intervals for the Largest Autoregressive Root in US Macroeconomic Time Series”, *Journal of Monetary Economics*, 28, 435-459.

Appendix: Proofs of the Main Results

The proofs given here rely on a large number of technical results that are established in the Online Supplement (Chao and Phillips, 2016). These results are designated in the derivations that follow by use of the prefix S.

Proof of Theorem 2.1:

The proof is almost identical to the proof for Lemma SD-11, which shows the consistency and establishes the rate of convergence for the preliminary estimator $\widehat{\rho}_{\text{pre}}$ under alternative parameter sequences. Indeed, the only difference is that the corresponding proof for $\widehat{\rho}_{\text{AIP}}$ requires in addition the consistency of the variance estimator $\widehat{\sigma}^2$, which we have shown in Lemma SD-12. Hence, to avoid redundancy, we do not replicate the argument here and instead refer interested reader to the proofs of Lemmas SD-11 and SD-12 in Appendix SD of the supplement. $\square$

Proof of Theorem 3.1:

Let $\Delta\varepsilon_{it}(\rho_T) = \Delta y_{it} - \rho_T \Delta y_{it-1}$, and note that

$$\begin{aligned}
\mathbb{M}(\rho_T) &= \frac{1}{\widehat{\omega}\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=3}^{T} y_{it-2} \left(\Delta y_{it} - \rho_T \Delta y_{it-1}\right) \\
&= \frac{1}{\widehat{\omega}\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=3}^{T} a_i \Delta\varepsilon_{it}(\rho_T) + \frac{1}{\widehat{\omega}\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=3}^{T} w_{it-2} \Delta\varepsilon_{it}(\rho_T).
\end{aligned}$$

Applying partial summation, we have

$$\begin{aligned}
&\frac{1}{\widehat{\omega}\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=3}^{T}w_{it-2}\Delta\varepsilon_{it}\left(\rho_T\right) \\
=\ & \frac{1}{\widehat{\omega}\sqrt{NT}}\sum_{i=1}^{N}\left[\sum_{t=4}^{T}\left(w_{it-3}-w_{it-2}\right)\varepsilon_{it-1}+w_{iT-2}\varepsilon_{iT}-w_{i1}\varepsilon_{i2}\right] \\
=\ & \left(1-\exp\left\{-\frac{1}{q\left(T\right)}\right\}\right)\frac{1}{\widehat{\omega}\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=4}^{T}w_{it-3}\varepsilon_{it-1}-\frac{1}{\widehat{\omega}\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=4}^{T}\varepsilon_{it-2}\varepsilon_{it-1} \\
&+\frac{1}{\widehat{\omega}\sqrt{NT}}\sum_{i=1}^{N}w_{iT-2}\varepsilon_{iT}-\frac{\rho_T}{\widehat{\omega}\sqrt{NT}}\sum_{i=1}^{N}w_{i0}\varepsilon_{i2}-\frac{1}{\widehat{\omega}\sqrt{NT}}\sum_{i=1}^{N}\varepsilon_{i1}\varepsilon_{i2},
\end{aligned}$$

so that

$$\begin{aligned}
&\mathbb{M}\left(\rho_T\right) \\
=\ & -\frac{1}{\widehat{\omega}\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=4}^{T}\varepsilon_{it-2}\varepsilon_{it-1}+\frac{1}{\widehat{\omega}\sqrt{NT}}\sum_{i=1}^{N}w_{iT-2}\varepsilon_{iT} \\
&+\left(1-\rho_T\right)\frac{1}{\widehat{\omega}\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=4}^{T}w_{it-3}\varepsilon_{it-1}+\frac{1}{\widehat{\omega}\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=2}^{T}a_i\Delta\varepsilon_{it} \\
&-\frac{\rho_T}{\widehat{\omega}\sqrt{NT}}\sum_{i=1}^{N}w_{i0}\varepsilon_{i2}-\frac{1}{\widehat{\omega}\sqrt{NT}}\sum_{i=1}^{N}\varepsilon_{i1}\varepsilon_{i2}.
\end{aligned}$$

We turn first to part (a). In this case, by assumption, $\rho_T = 1$ for all T sufficiently large. Applying parts (g) and (i) of Lemma SE-11, part (a) of Lemma SE-25, Lemma

SE-35, and part (a) of Lemma SD-13; we have

$$\begin{aligned}
&\mathbb{M}\left(\rho_T\right) \\
= &-\frac{1}{\widehat{\omega}\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=4}^{T}\varepsilon_{it-2}\varepsilon_{it-1}+\frac{1}{\widehat{\omega}\sqrt{NT}}\sum_{i=1}^{N}w_{iT-2}\varepsilon_{iT}+\frac{1}{\widehat{\omega}\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=2}^{T}a_i\Delta\varepsilon_{it} \\
&-\frac{\rho_T}{\widehat{\omega}\sqrt{NT}}\sum_{i=1}^{N}w_{i0}\varepsilon_{i2}-\frac{1}{\widehat{\omega}\sqrt{NT}}\sum_{i=1}^{N}\varepsilon_{i1}\varepsilon_{i2}+\left(1-\rho_T\right)\frac{1}{\widehat{\omega}\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=4}^{T}w_{it-3}\varepsilon_{it-1} \\
= &-\frac{1}{\widehat{\omega}\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=4}^{T}\varepsilon_{it-2}\varepsilon_{it-1}+\frac{1}{\widehat{\omega}\sqrt{NT}}\sum_{i=1}^{N}w_{iT-2}\varepsilon_{iT}+O_p\left(\frac{1}{\sqrt{T}}\right) \\
&+O_p\left(\left(1-\rho_T\right)\sqrt{T}\right) \\
= &-\frac{1}{\sigma^2\sqrt{2}}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=4}^{T}\varepsilon_{it-2}\varepsilon_{it-1}+\frac{1}{\sigma^2\sqrt{2}}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}w_{iT-2}\varepsilon_{iT} \\
&+O_p\left(\max\left\{\frac{1}{\sqrt{N}},\frac{1}{\sqrt{T}}\right\}\right).
\end{aligned}$$

It follows from applying Lemma SE-24 that $\mathbb{M}\left(\rho_T\right)\Rightarrow N\left(0,1\right)$, as required.

Next consider part (b), where we take $\rho_T=\exp\left\{-1/q\left(T\right)\right\}$ such that $T/q\left(T\right)\rightarrow 0$. In this case, using the results in parts (g) and (i) of Lemma SE-11, part (b) of Lemma SE-25, Lemma SE-35, and part (b) of Lemma SD-13, we deduce that

$$\begin{aligned}
&\mathbb{M}\left(\rho_T\right) \\
= &-\frac{1}{\widehat{\omega}\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=4}^{T}\varepsilon_{it-2}\varepsilon_{it-1}+\frac{1}{\widehat{\omega}\sqrt{NT}}\sum_{i=1}^{N}w_{iT-2}\varepsilon_{iT}+O_p\left(\frac{1}{\sqrt{T}}\right) \\
&+O_p\left(\frac{\sqrt{T}}{q\left(T\right)}\right) \\
= &-\frac{1}{\sigma^2\sqrt{2}}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=4}^{T}\varepsilon_{it-2}\varepsilon_{it-1}+\frac{1}{\sigma^2\sqrt{2}}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}w_{iT-2}\varepsilon_{iT} \\
&+O_p\left(\max\left\{\frac{1}{\sqrt{T}},\frac{1}{\sqrt{N}},\frac{T}{q\left(T\right)}\right\}\right)
\end{aligned}$$

It follows from part (a) of Lemma SE-22 that $\mathbb{M}\left(\rho_T\right)\Rightarrow N\left(0,1\right)$.

Consider part (c), where we take $\rho_T=\exp\left\{-1/q\left(T\right)\right\}$ such that $q\left(T\right)\sim T$. Here, we apply parts (g) and (i) of Lemma SE-11, part (c) of Lemma SE-25, Lemma SE-35, and part (c) of Lemma SD-13 to deduce that

$$
\begin{aligned}
& \mathbb{M}\left(\rho_T\right) \\
= & -\frac{1}{\widehat{\omega}\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=4}^{T}\varepsilon_{it-2}\varepsilon_{it-1}+\frac{1}{\widehat{\omega}\sqrt{NT}}\sum_{i=1}^{N}w_{iT-2}\varepsilon_{iT}+O_p\left(\frac{1}{\sqrt{T}}\right) \\
= & -\frac{1}{\overline{\omega}_T\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=4}^{T}\varepsilon_{it-2}\varepsilon_{it-1}+\frac{1}{\overline{\omega}_T\sqrt{NT}}\sum_{i=1}^{N}w_{iT-2}\varepsilon_{iT} \\
& +O_p\left(\max\left\{\frac{1}{\sqrt{N}},\frac{1}{\sqrt{T}}\right\}\right),
\end{aligned}
$$

where $\overline{\omega}_T=\sigma^2\left\{1+\left[q\left(T\right)/2T\right]\left[1-\exp\left\{-2T/q\left(T\right)\right\}\right]\right\}^{1/2}$. It follows from part (b) of Lemma SE-22 that $\mathbb{M}\left(\rho_T\right)\Rightarrow N\left(0,1\right)$.

For part (d), where we assume that $\rho_T=\exp\left\{-1/q\left(T\right)\right\}$ such that $q\left(T\right)\to\infty$ but $q\left(T\right)/T\to 0$, we first apply part (d) of Lemma SD-13 and part (d) of Lemma SE-21 to obtain

$$
\begin{aligned}
& \frac{1}{\widehat{\omega}\sqrt{NT}}\sum_{i=1}^{N}w_{iT-2}\varepsilon_{iT} \\
= & \frac{1}{\sigma^2\sqrt{NT}}\sum_{i=1}^{N}w_{iT-2}\varepsilon_{iT}\left[1+O_p\left(\max\left\{\frac{1}{N},\frac{q\left(T\right)}{T},\frac{1}{q\left(T\right)}\right\}\right)\right]=O_p\left(\sqrt{\frac{q\left(T\right)}{T}}\right).
\end{aligned}
$$

Hence, applying parts (g) and (i) of Lemma SE-11, part (d) of Lemma SE-25, and Lemma SE-35, we further obtain

$$
\begin{aligned}
& \mathbb{M}\left(\rho_T\right) \\
= & -\frac{1}{\widehat{\omega}\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=4}^{T}\varepsilon_{it-2}\varepsilon_{it-1}+O_p\left(\sqrt{\frac{q\left(T\right)}{T}}\right)+O_p\left(\frac{1}{\sqrt{T}}\right)+O_p\left(\frac{1}{\sqrt{q\left(T\right)}}\right) \\
= & -\frac{1}{\sigma^2\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=4}^{T}\varepsilon_{it-2}\varepsilon_{it-1}+O_p\left(\max\left\{\frac{1}{N},\sqrt{\frac{q\left(T\right)}{T}},\frac{1}{\sqrt{q\left(T\right)}}\right\}\right).
\end{aligned}
$$

By part (c) of Lemma SE-22, we then deduce that $\mathbb{M}\left(\rho_T\right)\Rightarrow N(0,1)$, as required for (d).

Finally, to show part (e), note first that by applying parts (g) and (i) of Lemma SE-11, part (e) of Lemma SE-21, Lemma SE-35, and part (e) of Lemma SD-13, we obtain

$$
\begin{aligned}
& \frac{1}{\widehat{\omega}\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=3}^{T} w_{it-2}\Delta\varepsilon_{it}\left(\rho_T\right) \\
= & -\frac{1}{\widehat{\omega}\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=4}^{T}\varepsilon_{it-2}\varepsilon_{it-1} + \frac{\left(1-\rho_T\right)}{\widehat{\omega}\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=4}^{T} w_{it-3}\varepsilon_{it-1} + O_p\left(\frac{1}{\sqrt{T}}\right) \\
= & \sqrt{\frac{1+\rho_T}{2\sigma^4}}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(X_{i,T}+Y_{i,T}\right) + o_p\left(1\right),
\end{aligned}
$$

where $X_{i,T} = -T^{-1/2}\sum_{t=4}^{T}\varepsilon_{it-2}\varepsilon_{it-1}$ and $Y_{i,T} = \left(1-\rho_T\right)T^{-1/2}\sum_{t=4}^{T} w_{it-3}\varepsilon_{it-1}$. It follows by Lemma SE-23 that $\mathbb{M}\left(\rho_T\right) = \sqrt{\left(1+\rho_T\right)/\left(2\sigma^4\right)}N^{-1/2}\sum_{i=1}^{N}\left(X_{i,T}+Y_{i,T}\right) + o_p\left(1\right) \Rightarrow N\left(0,1\right)$. $\square$

Proof of Theorem 3.2:

To proceed, note that, in the pathwise asymptotics considered here, N grows as a monotonically increasing function of T, so that the asymptotics can be taken to be single-indexed with $T \to \infty$. Now, let $\left\{\mathcal{G}_j^{\mathbb{M}} : j = 1,2,3,4,5\right\}$ be the collections of parameter sequences defined in the statement of Theorem 3.1. Moreover, let $\left\{\rho_{k,T}\right\} \in \mathcal{G}_{s_k}^{\mathbb{M}}$ (for $k = 1,2,...,5$), i.e., $\left\{\rho_{k,T}\right\}$ is a sequence belonging to the collection $\mathcal{G}_{s_k}^{\mathbb{M}}$.[9] Define $T_k = f_k\left(T\right)$ $(k = 1,...,d)$, with $d \leq 5$, where $f_k\left(\cdot\right) : \mathbb{N} \to \mathbb{N}$ is an increasing function in its argument, and let $\left\{\rho_{k,T_k}\right\}$ denote a subsequence of $\left\{\rho_{k,T}\right\}$. Note that every parameter sequence $\rho_T \in (-1,1]$ can be represented as $\left\{\rho_T\right\} = \bigcup_{j=1}^{d}\left\{\rho_{j,T_j}\right\}$, where $\left\{\rho_{1,T_1}\right\} \in \mathcal{G}_{s_1}^{\mathbb{M}}, ..., \left\{\rho_{d,T_d}\right\} \in \mathcal{G}_{s_d}^{\mathbb{M}}$, with $\mathcal{G}_{s_k}^{\mathbb{M}} \neq \mathcal{G}_{s_\ell}^{\mathbb{M}}$ for $k \neq \ell$ and where $\mathbb{N} = \bigcup_{k=1}^{d}\left\{T_k = f_k\left(T\right) : T \in \mathbb{N}\right\}$, with $\mathbb{N}$ denoting the set of natural numbers.

Next, note that $\Pr\left(\rho \notin \mathbb{C}_{\alpha_1,T}^{\mathbb{M}} | \rho = \rho_{k,T}\right) = \Pr\left(\left|\mathbb{M}_T\right| > z_{\alpha_1/2} | \rho = \rho_{k,T}\right)$. Theorem 3.1 implies that, for any $\varepsilon > 0$ and for each $k \in \{1,..,d\}$, there exists positive integer M_k such that for every positive integer $T \geq M_k$,

$$
\left|\Pr\left(\left|\mathbb{M}_T\right| > z_{\alpha_1/2} | \rho = \rho_{k,T}\right) - \Pr\left(\left|\mathcal{Z}\right| > z_{\alpha_1/2}\right)\right| < \varepsilon.
$$

Moreover, for any positive integer $T \geq M_k$, we have $T_k = f_k\left(T\right) \geq T \geq M_k$ by Lemma

[9] The reason for using the notation $\mathcal{G}_{s_k}^{\mathbb{M}}$, as opposed to $\mathcal{G}_k^{\mathbb{M}}$, is so that we can refer to a particular collection of sequences amongst $\left\{\mathcal{G}_j^{\mathbb{M}} : j = 1,2,...,5\right\}$ without $\mathcal{G}_{s_1}^{\mathbb{M}}$ necessarily being $\mathcal{G}_1^{\mathbb{M}}$, for example.

SE-33 (given in Appendix SE in the technical supplement to this paper), from which we further deduce that

$$\left|\Pr\left(|\mathbb{M}_{T_k}| > z_{\alpha_1/2}|\rho = \rho_{k,T_k}\right) - \Pr\left(|\mathcal{Z}| > z_{\alpha_1/2}\right)\right| < \varepsilon.$$

Next, let $\overline{M} = \max\{f_1(M_1), ..., f_k(M_d)\}$. Consider any positive integer $T \geq \overline{M}$; we must have $T = f_k(T^*)$ for some $k = 1, ..., d$ and for some $T^* \in \mathbb{N}$. Since $T = f_k(T^*) \geq \overline{M} \geq f_k(M_k)$ and since $f_k(\cdot)$ is an increasing function of its argument by Lemma SE-33, we deduce that $T \geq T^* \geq M_k$, from which it follows that

$$\begin{aligned}
& \left|\Pr\left(|\mathbb{M}_T| > z_{\alpha_1/2}|\rho = \rho_T\right) - \Pr\left(|\mathcal{Z}| > z_{\alpha_1/2}\right)\right| \\
= \; & \left|\Pr\left(\left|\mathbb{M}_{f_k(T^*)}\right| > z_{\alpha_1/2}|\rho = \rho_{k,f_k(T^*)}\right) - \Pr\left(|\mathcal{Z}| > z_{\alpha_1/2}\right)\right| < \varepsilon.
\end{aligned}$$

The desired result then follows from Lehmann (1999) Lemma 2.6.2. □

Lemma A1:

Suppose that Assumptions 1-4 hold. Then, the following statements are true as $N, T \to \infty$ such that $N^\kappa/T = \tau$, for constants $\kappa \in (1/2, \infty)$ and $\tau \in (0, \infty)$.

(a) Suppose that $\rho_T \in \mathcal{G}_1^{\mathrm{P}}$, where

$$\mathcal{G}_1^{\mathrm{P}} = \{\rho_T : \rho_T = 1 \text{ for all } T \text{ sufficiently large}\}.$$

Set $N = N(T) = (\tau T)^{1/\kappa}$ and $\mathbb{C}_{\gamma,\alpha,N,T} = \mathbb{C}_{\gamma,\alpha,N(T),T} = \mathbb{C}_{\gamma,\alpha,T}$. Then,

$$\limsup_{T\to\infty} \Pr\left(\rho \notin \mathbb{C}_{\gamma,\alpha,T}|\rho = \rho_T \in \mathcal{G}_1^{\mathrm{P}}\right) \leq \alpha_1 < \alpha.$$

(b) Suppose that $\rho_T \in \mathcal{G}_2^{\mathrm{P}}$, where

$$\mathcal{G}_2^{\mathrm{P}} = \left\{\rho_T = \exp\left\{-\frac{1}{q(T)}\right\} : \sqrt{N}T \sim T^{\frac{1}{2\kappa}+1} \ll q(T)\right\}.$$

Set $N = N(T) = (\tau T)^{1/\kappa}$ and $\mathbb{C}_{\gamma,\alpha,N,T} = \mathbb{C}_{\gamma,\alpha,N(T),T} = \mathbb{C}_{\gamma,\alpha,T}$. Then,

$$\limsup_{T\to\infty} \Pr\left(\rho \notin \mathbb{C}_{\gamma,\alpha,T}|\rho = \rho_T \in \mathcal{G}_2^{\mathrm{P}}\right) \leq \alpha_1 < \alpha.$$

(c) Suppose that $\rho_T \in \mathcal{G}_3^{\mathrm{P}}$, where

$$\mathcal{G}_3^{\mathrm{P}} = \left\{ \rho_T = \exp\left\{ -\frac{1}{q(T)} \right\} : \sqrt{N}T \sim q(T) \cap \rho_T \geq 1 - \frac{\left(z_{\gamma_1} + z_{\alpha_2}\right)\sqrt{2}}{\sqrt{N}T} \text{ eventually} \right\}.$$

Set $N = N(T) = (\tau T)^{1/\kappa}$ and $\mathbb{C}_{\gamma,\alpha,N,T} = \mathbb{C}_{\gamma,\alpha,N(T),T} = \mathbb{C}_{\gamma,\alpha,T}$. Then,

$$\limsup_{T\to\infty} \Pr\left(\rho \notin \mathbb{C}_{\gamma,\alpha,T} | \rho = \rho_T \in \mathcal{G}_3^{\mathrm{P}}\right) \leq \alpha_1 < \alpha.$$

(d) Suppose that $\rho_T \in \mathcal{G}_4^{\mathrm{P}}$, where

$$\mathcal{G}_4^{\mathrm{P}} = \left\{ \rho_T = \exp\left\{ -\frac{1}{q(T)} \right\} : \sqrt{N}T \sim q(T) \cap \rho_T < 1 - \frac{\left(z_{\gamma_1} + z_{\alpha_2}\right)\sqrt{2}}{\sqrt{N}T} \text{ eventually} \right\}.$$

Set $N = N(T) = (\tau T)^{1/\kappa}$ and $\mathbb{C}_{\gamma,\alpha,N,T} = \mathbb{C}_{\gamma,\alpha,N(T),T} = \mathbb{C}_{\gamma,\alpha,T}$. Then,

$$\limsup_{T\to\infty} \Pr\left(\rho \notin \mathbb{C}_{\gamma,\alpha,T} | \rho = \rho_T \in \mathcal{G}_4^{\mathrm{P}}\right) \leq \alpha_1 + \alpha_2 = \alpha.$$

(e) Suppose that $\rho_T \in \mathcal{G}_5^{\mathrm{P}}$, where

$$\mathcal{G}_5^{\mathrm{P}} = \left\{ \rho_T = \exp\left\{ -\frac{1}{q(T)} \right\} : \left(\sqrt{NT} \ll q(T)\right) \cap \left(T \ll q(T) \ll \sqrt{N}T\right) \right\}.$$

. Set $N = N(T) = (\tau T)^{1/\kappa}$ and $\mathbb{C}_{\gamma,\alpha,N,T} = \mathbb{C}_{\gamma,\alpha,N(T),T} = \mathbb{C}_{\gamma,\alpha,T}$. Then,

$$\limsup_{T\to\infty} \Pr\left(\rho \notin \mathbb{C}_{\gamma,\alpha,T} | \rho = \rho_T \in \mathcal{G}_5^{\mathrm{P}}\right) \leq \alpha_1 < \alpha.$$

(f) Suppose that $\rho_T \in \mathcal{G}_6^{\mathrm{P}}$, where

$$\mathcal{G}_6^{\mathrm{P}} = \left\{ \rho_T = \exp\left\{ -\frac{1}{q(T)} \right\} : T \ll q(T) \sim \sqrt{NT} \cap \rho_T \geq 1 - \frac{2\left(z_{\gamma_2} + z_{\alpha_2}\right)}{\sqrt{NT}} \text{ eventually} \right\}.$$

Set $N = N(T) = (\tau T)^{1/\kappa}$ and $\mathbb{C}_{\gamma,\alpha,N,T} = \mathbb{C}_{\gamma,\alpha,N(T),T} = \mathbb{C}_{\gamma,\alpha,T}$. Then,

$$\limsup_{T\to\infty} \Pr\left(\rho \notin \mathbb{C}_{\gamma,\alpha,T} | \rho = \rho_T \in \mathcal{G}_6^{\mathrm{P}}\right) \leq \alpha_1 < \alpha.$$

(g) Suppose that $\rho_T \in \mathcal{G}_7^{\mathrm{P}}$, where

$$\mathcal{G}_7^{\mathrm{P}} = \left\{ \rho_T = \exp\left\{ -\frac{1}{q(T)} \right\} : T \ll q(T) \sim \sqrt{NT} \cap \rho_T < 1 - \frac{2\left(z_{\gamma_2} + z_{\alpha_2}\right)}{\sqrt{NT}} \text{ eventually} \right\}.$$

Set $N = N(T) = (\tau T)^{1/\kappa}$ and $\mathbb{C}_{\gamma,\alpha,N,T} = \mathbb{C}_{\gamma,\alpha,N(T),T} = \mathbb{C}_{\gamma,\alpha,T}$. Then,

$$\limsup_{T\to\infty} \Pr\left(\rho \notin \mathbb{C}_{\gamma,\alpha,T} | \rho = \rho_T \in \mathcal{G}_7^{\mathrm{P}}\right) \leq \alpha_1 + \alpha_2 = \alpha.$$

(h) Suppose that $\rho_T \in \mathcal{G}_8^{\mathrm{P}}$, where

$$\mathcal{G}_8^{\mathrm{P}} = \left\{ \rho_T = \exp\left\{ -\frac{1}{q(T)} \right\} : \sqrt{NT} \ll q(T) \sim T \right\}.$$

Set $N = N(T) = (\tau T)^{1/\kappa}$ and $\mathbb{C}_{\gamma,\alpha,N,T} = \mathbb{C}_{\gamma,\alpha,N(T),T} = \mathbb{C}_{\gamma,\alpha,T}$. Then,

$$\limsup_{T\to\infty} \Pr\left(\rho \notin \mathbb{C}_{\gamma,\alpha,T} | \rho = \rho_T \in \mathcal{G}_8^{\mathrm{P}}\right) \leq \alpha_1 < \alpha$$

(i) Suppose that $\rho_T \in \mathcal{G}_9^{\mathrm{P}}$, where

$$\mathcal{G}_9^{\mathrm{P}} = \left\{ \rho_T = \exp\left\{ -\frac{1}{q(T)} \right\} : \sqrt{NT} \sim T^{\frac{1}{2}\left(1+\frac{1}{\kappa}\right)} \ll q(T) \ll T \right\}.$$

Set $N = N(T) = (\tau T)^{1/\kappa}$ and $\mathbb{C}_{\gamma,\alpha,N,T} = \mathbb{C}_{\gamma,\alpha,N(T),T} = \mathbb{C}_{\gamma,\alpha,T}$. Then,

$$\limsup_{T\to\infty} \Pr\left(\rho \notin \mathbb{C}_{\gamma,\alpha,T} | \rho = \rho_T \in \mathcal{G}_9^{\mathrm{P}}\right) \leq \alpha_1 < \alpha.$$

(j) Suppose that $\rho_T \in \mathcal{G}_{10}^{\mathrm{P}}$, where

$$\mathcal{G}_{10}^{\mathrm{P}} = \left\{ \rho_T = \exp\left\{ -\frac{1}{q(T)} \right\} : q(T) \sim \sqrt{NT} \sim T \cap \rho_T \geq 1 - \frac{2\left(z_{\gamma_2} + z_{\alpha_2}\right)}{\sqrt{NT}} \text{ eventually} \right\}.$$

Set $N = N(T) = (\tau T)^{1/\kappa}$ and $\mathbb{C}_{\gamma,\alpha,N,T} = \mathbb{C}_{\gamma,\alpha,N(T),T} = \mathbb{C}_{\gamma,\alpha,T}$. Then,

$$\limsup_{T\to\infty} \Pr\left(\rho \notin \mathbb{C}_{\gamma,\alpha,T} | \rho = \rho_T \in \mathcal{G}_{10}^{\mathrm{P}}\right) \leq \alpha_1 < \alpha.$$

(k) Suppose that $\rho_T \in \mathcal{G}^{\mathrm{P}}_{11}$, where

$$\mathcal{G}^{\mathrm{P}}_{11} = \left\{\rho_T = \exp\left\{-\frac{1}{q(T)}\right\} : q(T) \sim \sqrt{NT} \sim T \cap \rho_T < 1 - \frac{2\left(z_{\gamma_2} + z_{\alpha_2}\right)}{\sqrt{NT}} \text{ eventually}\right\}.$$

Set $N = N(T) = (\tau T)^{1/\kappa}$ and $\mathbb{C}_{\gamma,\alpha,N,T} = \mathbb{C}_{\gamma,\alpha,N(T),T} = \mathbb{C}_{\gamma,\alpha,T}$. Then,

$$\limsup_{T\to\infty} \Pr\left(\rho \notin \mathbb{C}_{\gamma,\alpha,T} | \rho = \rho_T \in \mathcal{G}^{\mathrm{P}}_{11}\right) \leq \alpha_1 + \alpha_2 = \alpha.$$

(l) Suppose that $\rho_T \in \mathcal{G}^{\mathrm{P}}_{12}$, where

$$\mathcal{G}^{\mathrm{P}}_{12} = \left\{\rho_T = \exp\left\{-\frac{1}{q(T)}\right\} : q(T) \sim \sqrt{NT} \ll T \cap \rho_T \geq 1 - \frac{2\left(z_{\gamma_2} + z_{\alpha_3}\right)}{\sqrt{NT}} \text{ eventually}\right\}.$$

Set $N = N(T) = (\tau T)^{1/\kappa}$ and $\mathbb{C}_{\gamma,\alpha,N,T} = \mathbb{C}_{\gamma,\alpha,N(T),T} = \mathbb{C}_{\gamma,\alpha,T}$. Then,

$$\limsup_{T\to\infty} \Pr\left(\rho \notin \mathbb{C}_{\gamma,\alpha,T} | \rho = \rho_T \in \mathcal{G}^{\mathrm{P}}_{12}\right) \leq \alpha_1 < \alpha.$$

(m) Suppose that $\rho_T \in \mathcal{G}^{\mathrm{P}}_{13}$, where

$$\mathcal{G}^{\mathrm{P}}_{13} = \left\{\rho_T = \exp\left\{-\frac{1}{q(T)}\right\} : q(T) \sim \sqrt{NT} \ll T \cap \rho_T < 1 - \frac{2\left(z_{\gamma_2} + z_{\alpha_2}\right)}{\sqrt{NT}} \text{ eventually}\right\}.$$

Set $N = N(T) = (\tau T)^{1/\kappa}$ and $\mathbb{C}_{\gamma,\alpha,N,T} = \mathbb{C}_{\gamma,\alpha,N(T),T} = \mathbb{C}_{\gamma,\alpha,T}$. Then,

$$\limsup_{T\to\infty} \Pr\left(\rho \notin \mathbb{C}_{\gamma,\alpha,T} | \rho = \rho_T \in \mathcal{G}^{\mathrm{P}}_{13}\right) \leq \alpha_1 + \alpha_2 = \alpha.$$

(n) Suppose that $\rho_T \in \mathcal{G}^{\mathrm{P}}_{14}$, where

$$\mathcal{G}^{\mathrm{P}}_{14} = \left\{\rho_T = \exp\left\{-\frac{1}{q(T)}\right\} : T \ll q(T) \ll \sqrt{NT}\right\}.$$

Set $N = N(T) = (\tau T)^{1/\kappa}$ and $\mathbb{C}_{\gamma,\alpha,N,T} = \mathbb{C}_{\gamma,\alpha,N(T),T} = \mathbb{C}_{\gamma,\alpha,T}$. Then,

$$\limsup_{T\to\infty} \Pr\left(\rho \notin \mathbb{C}_{\gamma,\alpha,T} | \rho = \rho_T \in \mathcal{G}^{\mathrm{P}}_{14}\right) \leq \alpha_1 < \alpha.$$

(o) Suppose that $\rho_T \in \mathcal{G}_{15}^{\mathrm{P}}$, where

$$\mathcal{G}_{15}^{\mathrm{P}} = \left\{ \rho_T = \exp\left\{ -\frac{1}{q(T)} \right\} : N^{1/3}T^{1/3} \ll q(T) \sim T \ll \sqrt{NT} \right\}.$$

Set $N = N(T) = (\tau T)^{1/\kappa}$ and $\mathbb{C}_{\gamma,\alpha,N,T} = \mathbb{C}_{\gamma,\alpha,N(T),T} = \mathbb{C}_{\gamma,\alpha,T}$. Then,

$$\limsup_{T\to\infty} \Pr\left(\rho \notin \mathbb{C}_{\gamma,\alpha,T} | \rho = \rho_T \in \mathcal{G}_{15}^{\mathrm{P}} \right) \le \alpha_1 < \alpha.$$

(p) Suppose that $\rho_T \in \mathcal{G}_{16}^{\mathrm{P}}$, where

$$\mathcal{G}_{16}^{\mathrm{P}} = \left\{ \rho_T = \exp\left\{ -\frac{1}{q(T)} \right\} : \left(N^{1/3}T^{1/3} \ll q(T) \ll \sqrt{NT} \right) \cap \left(q(T) \ll T \right) \right\}.$$

Set $N = N(T) = (\tau T)^{1/\kappa}$ and $\mathbb{C}_{\gamma,\alpha,N,T} = \mathbb{C}_{\gamma,\alpha,N(T),T} = \mathbb{C}_{\gamma,\alpha,T}$. Then,

$$\limsup_{T\to\infty} \Pr\left(\rho \notin \mathbb{C}_{\gamma,\alpha,T} | \rho = \rho_T \in \mathcal{G}_{16}^{\mathrm{P}} \right) \le \alpha_1 < \alpha.$$

(q) Suppose that $\rho_T \in \mathcal{G}_{17}^{\mathrm{P}}$, where

$$\mathcal{G}_{17}^{\mathrm{P}} = \left\{ \rho_T = \exp\left\{ -\frac{1}{q(T)} \right\} : q(T) \sim T^{\frac{1+\kappa}{3\kappa}} \sim N^{1/3}T^{1/3} \right\}.$$

Set $N = N(T) = (\tau T)^{1/\kappa}$ and $\mathbb{C}_{\gamma,\alpha,N,T} = \mathbb{C}_{\gamma,\alpha,N(T),T} = \mathbb{C}_{\gamma,\alpha,T}$. Then,

$$\limsup_{T\to\infty} \Pr\left(\rho \notin \mathbb{C}_{\gamma,\alpha,T} | \rho = \rho_T \in \mathcal{G}_{17}^{\mathrm{P}} \right) \le \alpha_1 < \alpha.$$

(r) Suppose that $\rho_T \in \mathcal{G}_{18}^{\mathrm{P}}$, where

$$\mathcal{G}_{18}^{\mathrm{P}} = \left\{ \rho_T = \exp\left\{ -\frac{1}{q(T)} \right\} : q(T) \to \infty \text{ such that } q(T) / T^{\frac{1+\kappa}{3\kappa}} \to 0 \right\}.$$

Set $N = N(T) = (\tau T)^{1/\kappa}$ and $\mathbb{C}_{\gamma,\alpha,N,T} = \mathbb{C}_{\gamma,\alpha,N(T),T} = \mathbb{C}_{\gamma,\alpha,T}$. Then,

$$\limsup_{T\to\infty} \Pr\left(\rho \notin \mathbb{C}_{\gamma,\alpha,T} | \rho = \rho_T \in \mathcal{G}_{18}^{\mathrm{P}} \right) \le \alpha_1 < \alpha.$$

(s) Suppose that $\rho_T \in \mathcal{G}_{19}^{\mathrm{P}}$, where

$$\mathcal{G}_{19}^{\mathrm{P}} = \left\{ |\rho_T| = \exp\left\{ -\frac{1}{q(T)} \right\} : q(T) \geq 0 \text{ and } q(T) = O(1) \text{ as } T \to \infty \right\}.$$

Set $N = N(T) = (\tau T)^{1/\kappa}$ and $\mathbb{C}_{\gamma,\alpha,N,T} = \mathbb{C}_{\gamma,\alpha,N(T),T} = \mathbb{C}_{\gamma,\alpha,T}$. Then,

$$\limsup_{T\to\infty} \Pr\left(\rho \notin \mathbb{C}_{\gamma,\alpha,T} | \rho = \rho_T \in \mathcal{G}_{19}^{\mathrm{P}}\right) \leq \alpha_1 < \alpha.$$

The proof of Lemma A1 is given in Appendix SB of the technical supplement.

Proof of Theorem 3.3:

In the pathwise asymptotics considered here, N grows as a monotonically increasing function of T, so that the asymptotics can be taken to be single-indexed with $T \to \infty$. Hence, we can set $N = (\tau T)^{1/\kappa}$ and simplify notation by writing $\mathbb{C}_{\gamma,\alpha,N,T} = \mathbb{C}_{\gamma,\alpha,N(T),T} = \mathbb{C}_{\gamma,\alpha,T}$.

To proceed, note that, by property of a supremum, there exists a sequence $\{\rho_T \in (-1,1] : T \geq 1\}$ such that

$$\limsup_{T\to\infty} \Pr\left(\rho \notin \mathbb{C}_{\gamma,\alpha,T} | \rho = \rho_T\right) = \limsup_{T\to\infty} \sup_{\rho^0 \in (-1,1]} \Pr\left(\rho \notin \mathbb{C}_{\gamma,\alpha,T} | \rho = \rho^0\right)$$

Thus, for some fixed significance level $\alpha \in (0, 0.5]$, to show that

$$\limsup_{T\to\infty} \sup_{\rho^0 \in (-1,1]} \Pr\left(\rho \notin \mathbb{C}_{\gamma,\alpha,T} | \rho = \rho^0\right) \leq \alpha,$$

it suffices to show that

$$\limsup_{T\to\infty} \Pr\left(\rho \notin \mathbb{C}_{\gamma,\alpha,T} | \rho = \rho_T\right) \leq \alpha$$

for every sequence $\{\rho_T \in (-1,1] : T \geq 1\}$. To proceed, let $\left\{\mathcal{G}_j^{\mathrm{P}} : j = 1, 2, ..., 19\right\}$ be the collections of parameter sequences defined in the statement of Lemma A1 given above. Moreover, let $\left\{\rho_{k,T}\right\} \in \mathcal{G}_{s_k}^{\mathrm{P}}$ (for $k = 1, ..., 19$), i.e., $\left\{\rho_{k,T}\right\}$ is a sequence belonging to the collection $\mathcal{G}_{s_k}^{\mathrm{P}}$. Define $T_k = f_k(T)$ $(k = 1, ..., d)$, with $d \leq 19$, where $f_k(\cdot) : \mathbb{N} \to \mathbb{N}$ is an increasing function in its argument, and let $\left\{\rho_{k,T_k}\right\}$ denote a subsequence of $\left\{\rho_{k,T}\right\}$. Note that every parameter sequence $\rho_T \in (-1,1]$ can be represented as $\{\rho_T\} = \bigcup_{j=1}^{d} \left\{\rho_{j,T_j}\right\}$, where $\left\{\rho_{1,T_1}\right\} \in \mathcal{G}_{s_1}^{\mathrm{P}}, ..., \left\{\rho_{d,T_d}\right\} \in \mathcal{G}_{s_d}^{\mathrm{P}}$, with $\mathcal{G}_{s_k}^{\mathrm{P}} \neq \mathcal{G}_{s_\ell}^{\mathrm{P}}$ for $k \neq \ell$ and where

$$\mathbb{N} = \bigcup_{k=1}^{d} \{T_k = f_k(T) : T \in \mathbb{N}\} \tag{10}$$

with $\mathbb{N}$ denoting the set of natural numbers $\{1, 2, ...\}$. Moreover, define

$$\upsilon_{k,T} = \sup_{m \geq T} \Pr\left(\rho \notin \mathbb{C}_{\gamma,\alpha,m} | \rho = \rho_{k,m} \in \mathcal{G}_{s_k}^{\mathrm{P}}\right)$$

and

$$\overline{p}_k = \lim \sup_{T \to \infty} \Pr\left(\rho \notin \mathbb{C}_{\gamma,\alpha,T} | \rho = \rho_{k,T} \in \mathcal{G}_{s_k}^{\mathrm{P}}\right).$$

It is clear from the definition of $\upsilon_{k,T}$ and $\overline{p}_k$ that $\lim_{T \to \infty} \upsilon_{k,T} = \overline{p}_k$ for each $k \in \{1, 2, ..., d\}$; or, more formally, for any $\varepsilon > 0$, there exists positive integer L_k such that, for all $T \geq L_k$, $|\upsilon_{k,T} - \overline{p}_k| < \varepsilon$, from which it follows, using the results of Lemma A1, that, for any $\varepsilon > 0$ and for each $k \in \{1, 2, ..., d\}$, there exists a positive integer L_k such that, for all $T \geq L_k$, $\upsilon_{k,T} < \overline{p}_k + \varepsilon \leq \alpha + \varepsilon$. Now, for any $k \in \{1, ..., d\}$ and for any positive integer $T \geq L_k$, we have, by Lemma SE-33 given in Appendix SE of the technical supplement to this paper, that $T_k = f_k(T) \geq L_k$, so that $|\upsilon_{k,T_k} - \overline{p}_k| < \varepsilon$, for any subsequence $\{\upsilon_{k,T_k}\}$ of $\{\upsilon_{k,T}\}$, from which we further deduce that

$$\upsilon_{k,T_k} = \sup_{m \geq T_k} \Pr\left(\rho \notin \mathbb{C}_{\gamma,\alpha,m} | \rho = \rho_{k,m} \in \mathcal{G}_{s_k}^{\mathrm{P}}\right) < \overline{p}_k + \varepsilon \leq \alpha + \varepsilon.$$

Next, let $L^{\max} = \max\{f_1(L_1), ..., f_d(L_d)\}$. Consider any positive integer $T \geq L^{\max}$; then, (10) implies that $T = f_k(T^*)$ for some $k = 1, ..., d$ and for some $T^* \in \mathbb{N}$. By the fact that $f_k(\cdot)$ is an increasing function of its argument, we have that $T = f_k(T^*) \geq L^{\max} \geq f_k(L_k) \geq L_k$, from which it follows that for every positive integer $T \geq L^{\max}$

$$\begin{aligned} \sup_{m \geq T} \Pr(\rho \notin \mathbb{C}_{\gamma,\alpha,m} | \rho = \rho_m) &= \sup_{m \geq f_k(T^*)} \Pr(\rho \notin \mathbb{C}_{\gamma,\alpha,m} | \rho = \rho_m) \\ &\leq \sup_{m \geq L^{\max}} \Pr(\rho \notin \mathbb{C}_{\gamma,\alpha,m} | \rho = \rho_m) < \alpha + \varepsilon \end{aligned}$$

Hence, for any sequence $\rho_T \in (-1, 1]$,

$$\limsup_{T\to\infty} \Pr\left(\rho \notin \mathbb{C}_{\gamma,\alpha,T} | \rho = \rho_T\right) = \inf_{T\geq 1} \sup_{m\geq T} \Pr\left(\rho \notin \mathbb{C}_{\gamma,\alpha,m} | \rho = \rho_m\right) < \alpha + \varepsilon.$$

Since ε is arbitrary, we deduce that

$$\limsup_{T\to\infty} \Pr\left(\rho \notin \mathbb{C}_{\gamma,\alpha,T} | \rho = \rho_T\right) \leq \alpha$$

for any sequence $\rho_T \in (-1, 1]$, which gives the desired conclusion. $\square$