Uniform Inference in Panel Autoregression

John C. Chao*and Peter C. B. Phillips†

First Draft: December 11, 2015 *This Version*: November 16, 2019

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. Our 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. The starting point for our confidence procedure is the estimating equation of the Anderson-Hsiao (AH) IV procedure. It is well known that the AH IV estimation suffers from weak instrumentation when ρ is near unity. But it is not so well known that AH IV estimation is still consistent when $\rho = 1$. In fact, the AH estimating equation is very well-centered and is an unbiased estimating equation in the sense of Durbin (1960), a feature that is especially useful in confidence interval construction. We show that a properly normalized statistic based on the AH estimating equation, which we call the $\mathbb{M}$ statistic, is uniformly convergent and can be inverted to obtain asymptotically valid interval estimates. To further improve the informativeness of our confidence procedure in the unit root and near unit root regions and to alleviate the problem that the AH procedure has greater variation in these regions, we use information from unit root pretesting to select among alternative confidence intervals. Two sequential tests are used to assess how close ρ is to unity, and different intervals are applied depending on whether the test results indicate ρ to be near or far away from unity. When ρ is relatively close to unity, our procedure activates intervals whose width shrinks to zero at a faster rate than that of the confidence interval based on the $\mathbb{M}$ statistic. Only when both of our unit root tests reject the null hypothesis does our procedure turn to the $\mathbb{M}$ statistic interval, whose width has the optimal $N^{-1/2}T^{-1/2}$ rate of shrinkage when the underlying process is stable. Our asymptotic analysis shows this pretest-based confidence procedure to have coverage probability that is at least the nominal level in large samples uniformly over the parameter space. 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.

*Department of Economics, University of Maryland, College Park, MD 20742. Email: jcchao@umd.edu.

†Yale University, University of Auckland, University of Southampton and Singapore Management University. Address: Department of Economics, Yale University, New Haven, CT 06520, USA. Email: peter.phillips@yale.edu.

Keywords: Confidence interval, Dynamic panel data models, panel IV, pooled OLS, Uniform inference.
JEL classification: C23, C26

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), Chang (2002, 2004), Moon and Phillips (2004), Moon et al (2014a,b), Pesaran (2006), and 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 take as our starting point the estimating equation of the Anderson-Hsiao (1981, 1982) IV procedure. Although much has been written about the Anderson-Hsiao IV estimator having a weak instrument problem when ρ is unity or very nearly unity, it should be noted that this estimator is still consistent even when $\rho = 1$ and that the weak instrument problem primarily manifests itself in the form of the asymptotic distribution having greater dispersion[2]. In fact, the Anderson-Hsiao estimating equation is very well-centered and is an unbiased estimating equation in the sense of Durbin (1960), a property that is of particular importance in constructing asymptotically valid confidence intervals. Exploiting this unbiasedness property, we then show that a properly normalized statistic based on this estimating equation is uniformly convergent over the parameter space $\Theta_\rho = (-1, 1]$. 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, because of the weak instrument problem, when the true ρ is unity or very near unity, confidence intervals obtained by inverting this $\mathbb{M}$ statistic may be less informative in the sense that they may be relatively wide in finite samples and, asymptotically, their width shrinks toward zero at the slower rate of $T^{-1/2}$ even when both the cross section (N) and

[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.

[2] For readers interested in the asymptotic properties of the Anderson-Hsiao IV estimator, we would like to refer them to Theorem SA-1 in the Technical Supplement to this paper. There, we present a very extensive set of results on the large sample behavior of this estimator under various parameter sequences both near and far away from unity. In addition, the proof of Theorem SA-1 is provided in Appendix SB of the Technical Supplement.

the 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 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.[3] 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 to implement. Simulations confirm that the proposed method performs well in finite samples.

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

[3] 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 the Appendix to this paper. Proofs of additional supporting lemmas as well as additional Monte Carlo results are reported in an online supplement to this paper (Chao and Phillips, 2019). The supplement provides an extensive set of results for panel estimation limit theory in unit root and near unit root cases that help deliver the main results in the paper but are of wider interest regarding asymptotic behavior of panel IV estimators, particularly in weak instrument settings. The supplement includes additional simulation results concerning the performance of the estimation procedures considered in the paper.

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, $\mathcal{Z}$ denotes the standard normal random variable, and $W_i(r)$ is the standard Brownian motion on the unit interval $[0,1]$ for each i. 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 $T \to \infty$. Similarly, for random variables X_T and Y_T, we take $X_T \overset{p}{\sim} Y_T$ to mean that $X_T/Y_T = O_p(1)$ and $Y_T/X_T = O_p(1)$, as $T \to \infty$. In addition, the notations $\Pr(\cdot|\rho)$ and $\Pr(\cdot|\rho_T)$ denote, respectively, a probability measure indexed by the fixed parameter ρ and one indexed by the local-to-unity parameter ρ_T. Finally, we use wid($\mathbb{C}$) to denote the width of the confidence interval $\mathbb{C}$.

2 Model and Assumptions

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

$$y_{it} = a_i + w_{it}, \tag{1}$$

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

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, a_i denotes an (unobserved) individual effect, and ρ denotes the panel autoregressive parameter, which is assumed to belong in the parameter space $\Theta_\rho = (-1, 1]$. In this paper, we will show that certain properties of our procedure holds uniformly for $\rho \in \Theta_\rho = (-1, 1]$. We do this by making use of a result (Lemma 2.6.2) from Lehmann (1999) which establishes the equivalence of uniform convergence and convergence for every parameter sequence belonging to a given parameter space. For this purpose, it is convenient for us to consider a general class of local-to-unity parameter sequences of the form $\rho = \rho_T = \exp\{-1/q(T)\}$, where $q(T)$ is a non-negative function of T such that $q(T) \to \infty$ as

$T \to \infty$[4]. Moreover, parameter sequences for stable AR processes can also be written in this general form by considering parameter sequences $\{\rho_T\}$ which belong to the collection

$$\mathcal{G}_{\text{St}} = \left\{ \{\rho_T\} : |\rho_T| = \exp\left\{ -\frac{1}{q(T)} \right\}, \, q(T) \geq 0, \text{ and } q(T) = O(1) \text{ as } T \to \infty \right\}. \tag{3}$$

Note also that, in the case where such parameter sequences are considered, the AR process given in expression (2) will depend on an indexed parameter ρ_T, as opposed to a fixed autoregressive parameter, so that, strictly speaking, the observed data and the latent process, in this case, will be strictly triangular indexed arrays $\{y_{it,T}, w_{it,T}\}$, which depend additionally on T. However, for notational ease, we shall suppress this additional dependence and simply write $\{y_{it}, w_{it}\}$ in what follows.

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.,

$$y_{it} = a_i(1-\rho) + \rho y_{it-1} + \varepsilon_{it} = \eta_i + \rho y_{it-1} + \varepsilon_{it}, \tag{4}$$

where $\eta_i = a_i(1-\rho)$. 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) $E[a_i^4] < \infty$; (c) ε_{it} *and* a_j, *are mutually independent for all* $i, j = 1, 2, ..., N$ *and for all* $t = 1, 2, ..., T$.

Assumption 3 (Initialization): *Let* $y_{i0} = a_i + w_{i0}$. *Suppose that* $\{w_{i0}\}$ *is independent across* i. *Suppose also that there exists a positive constant* C *such that* $\sup_i E[w_{i0}^2] \leq C < \infty$, *and that* w_{i0} *and* ε_{jt} *are independent for all* $i, j = 1, 2, ..., N$ *and for all* $t = 1, 2, ..., T$.

Note that Assumption 3 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 3 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 so that 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.

In lieu of Assumption 2, we also consider in this paper a fixed-effects specification given by the following assumption.

[4]The reason we consider indexed parameter ρ_T which depends on T only, and not on both N and T, is because our main results are obtained under a general pathwise asymptotic scheme where N can grow as an arbitrary positive real-valued power of T. In such a framework, the asymptotics is effectively single-indexed. Hence, it suffices that we only consider parameter sequences that depend only on T.

Assumption 2* (Fixed Effects): Let $\{a_i\}$ be a nonrandom sequence. Suppose that there exists a positive constant C such that $\sup_{1\leq i\leq N}|a_i|\leq C<\infty$ for all N.

All our theoretical results hold under either the random-effects specification given by Assumption 2 or the fixed-effects specification given by Assumption 2*.

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 $\widehat{\omega}^2=\widetilde{\sigma}^2\left[N^{-1}T^{-1}\sum_{i=1}^{N}\sum_{t=4}^{T}\left(y_{it-3}-y_{it-2}\right)^2+N^{-1}T^{-1}\sum_{i=1}^{N}y_{iT-2}^2\right]$ and $\widetilde{\sigma}^2=N^{-1}T_1^{-1}\sum_{i=1}^{N}\sum_{t=2}^{T}\left(y_{it}-\overline{y}_i-\widetilde{\rho}\left[y_{it-1}-\overline{y}_{i,-1}\right]\right)$ and where $\overline{y}_i=T_1^{-1}\sum_{t=2}^{T}y_{it}$ and $\overline{y}_{i,-1}=T_1^{-1}\sum_{t=2}^{T}y_{it-1}$ with $T_1=T-1$. In addition, we let $\widetilde{\rho}$ denote any preliminary estimator of ρ that satisfies the following conditions

Assumption 4: Let $\widetilde{\rho}$ be an estimator of ρ. Suppose that the following conditions hold for this estimator as $N,T\to\infty$ such that $N^{\kappa}/T=\tau$, for $\kappa\in(0,\infty)$ and $\tau\in(0,\infty)$.

(a) $\widetilde{\rho}-\rho_T=o_p\left(T^{-1/2}\right)$, if $\rho_T=1$ for all T sufficient large or if $\rho_T=\exp\{-1/q(T)\}$ such that $T/q(T)=O(1)$;

(b) $\widetilde{\rho}-\rho_T=o_p\left(q(T)^{-1/2}\right)$, if $\rho_T=\exp\{-1/q(T)\}$ such that $q(T)\to\infty$ but $q(T)/T\to 0$;

(c) $\widetilde{\rho}-\rho_T=o_p(1)$ if $\{\rho_T\}\in\mathcal{G}_{\mathrm{St}}$, where $\mathcal{G}_{\mathrm{St}}$ is given in expression (3) above.

The asymptotic properties of $\mathbb{M}(\rho)$ under different parameter sequences $\{\rho_T\}$ are given by the following result.

Theorem 3.1:

Let Assumptions 1, 3, 4 and either 2 or 2 hold. The following statements hold as* $N, T \to \infty$ *such that* $N^{\kappa}/T = \tau$*, for* $\kappa \in (0, \infty)$ *and* $\tau \in (0, \infty)$*.*

(a) *Suppose that* $\{\rho_T\} \in \mathcal{G}_1^{\mathbb{M}}$*, where* $\mathcal{G}_1^{\mathbb{M}} = \{\{\rho_T\} : \rho_T = 1$ *for all* T *sufficiently large*$\}$*. Then,*

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

(b) *Suppose that* $\{\rho_T\} \in \mathcal{G}_2^{\mathbb{M}}$*, where* $\mathcal{G}_2^{\mathbb{M}} = \{\{\rho_T\} : \rho_T = \exp\{-1/q(T)\}$ *and* $T/q(T) \to 0\}$*. Then,*

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

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

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

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

(d) *Suppose that* $\{\rho_T\} \in \mathcal{G}_4^{\mathbb{M}}$*, where*

$\mathcal{G}_4^{\mathbb{M}} = \{\{\rho_T\} : \rho_T = \exp\{-1/q(T)\}$ *and* $q(T) \to \infty$ *such that* $q(T)/T \to 0\}$*. Then,*

$$\mathbb{M}(\rho_T) = -\frac{1}{\sigma^2\sqrt{NT}} \sum_{i=1}^{N}\sum_{t=4}^{T} \varepsilon_{it-2}\varepsilon_{it-1} + o_p(1) \Rightarrow N(0,1).$$

(e) *Suppose that* $\{\rho_T\} \in \mathcal{G}_5^{\mathbb{M}}$*, where*

$\mathcal{G}_5^{\mathbb{M}} = \{\{\rho_T\} : |\rho_T| = \exp\{-1/q(T)\},\ q(T) \geq 0,\ and\ q(T) = O(1)\ \ as\ T \to \infty\}$. *Then,*

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

To provide some intuition about the $\mathbb{M}(\rho)$ statistic and about the conditions placed on the preliminary estimator $\widetilde{\rho}$ (i.e., Assumption 4), we set $\mathbb{M}^*(\rho) = \widehat{\omega}\mathbb{M}(\rho)$ to be the unstandardized version of $\mathbb{M}(\rho)$. From the proof of Theorem 3.1, given in the Appendix, it is evident 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 is 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 a scale estimator that will adapt to variation in the normalization factor under alternative parameter sequences. The estimator

$$\widehat{\omega} = \widetilde{\sigma}^2\left[N^{-1}T^{-1}\sum\nolimits_{i=1}^{N}\sum\nolimits_{t=4}^{T}(y_{it-3}-y_{it-2})^2 + N^{-1}T^{-1}\sum\nolimits_{i=1}^{N}y_{iT-2}^2\right]$$

turns out to have these adaptive properties, as shown in Lemma SC-13 and its proof (given in the supplement). An important component in the construction of a proper normalization factor is to have a preliminary estimator $\widetilde{\rho}$ with a fast enough rate of convergence, so that the resulting estimator of σ^2 is consistent under every possible parameter sequence $\{\rho_T\}$ in the parameter space $\Theta_\rho = (-1, 1]$. Examination of the proof of Lemma SC-12 reveals that the conditions needed on $\widetilde{\rho}$ are precisely those given in Assumption 4[5].

It should be noted that the Anderson-Hsiao IV estimator, which we will denote by $\widehat{\rho}_{\mathrm{IVD}}$ in this paper, does not satisfy the conditions of Assumption 4[6]. This is because, as shown in Theorem SA-1 of the supplement to this paper, $\widehat{\rho}_{\mathrm{IVD}} - \rho_T \overset{p}{\sim} T^{-1/2}$ when $\rho_T = 1$ for all T sufficient large or when $\rho_T = \exp\{-1/q(T)\}$ such that $T/q(T) = O(1)$, so that its rate of

[5] The proof of Lemma SC-12 is also given in Appendix SC of the technical supplement.

[6] We use the notation $\widehat{\rho}_{\mathrm{IVD}}$ 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 $\widehat{\rho}_{\mathrm{IVL}}$ to denote the IV estimator introduced by Arellano and Bover (1995) since, in that procedure, IV is performed on the panel autoregression in levels.

convergence is not fast enough in the unit root and near unit root regions of the parameter space. Furthermore, the pooled OLS (POLS) estimator, which we will denote by $\widehat{\rho}_{\text{pols}}$, also does not satisfy these conditions since it is inconsistent in the stationary region as shown in Theorem SA-2 of the supplement. Hence, in Appendix SA of the supplement to our paper, we introduce a new point estimator, $\widehat{\rho}_{\text{AIP}}$, which is an average of $\widehat{\rho}_{\text{IVD}}$ and $\widehat{\rho}_{\text{pols}}$ where the average is taken using a data-dependent weight function that, in turn, depends on a unit root statistic. This estimator turns out to satisfy the conditions of Assumption 4 because it exploits the differential strengths of $\widehat{\rho}_{\text{IVD}}$ and $\widehat{\rho}_{\text{pols}}$ in different parts of the parameter space and can place more or less weight on one or the other of these two estimators, depending on the information provided by a preliminary unit root test on the true value of ρ. We use $\widehat{\rho}_{\text{AIP}}$ in constructing the scale estimator $\widehat{\omega}$ for the Monte Carlo results reported section 4 of this paper, but we note that $\widehat{\rho}_{\text{AIP}}$ is not the only estimator which satisfies Assumption 4, as both the within-group OLS estimator and the bias-corrected within-group estimator proposed by Hahn and Kuersteiner (2002) could also be used to obtain a scale estimator $\widehat{\omega}$ with the desired properties, although $\widehat{\rho}_{\text{AIP}}$ does have a faster rate of convergence than the uncorrected within-group estimator in the unit root and near unit root cases. Because the focus of this paper is on confidence procedures, and not on point estimation, we will not give technical details of $\widehat{\rho}_{\text{AIP}}$ in the body of this paper, but will instead refer interested readers to Appendix SA of the supplement for more details as well as for formal results on the rate of convergence of $\widehat{\rho}_{\text{AIP}}$ under alternative parameter sequences.

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

Theorem 3.2:

Let $\Phi(x)$ denote the cdf of a standard normal random variable. Suppose that Assumptions 1, 3, 4 and either 2 or 2 hold. Then, for each $x \in \mathbb{R}$,*

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

as $N, T \to \infty$ such that $N^\kappa / T = \tau$ for constants $\kappa \in (0, \infty)$ and $\tau \in (0, \infty)$.[7]

Remarks 3.1:

(i) Let z_α denote the $1 - \alpha$ quantile of the standard normal distribution. A level $1 - \alpha$

[7]Note that we use the notation $\Pr(\cdot|\rho)$ instead of perhaps the more familiar notation $P_\rho(\cdot)$ to denote a probability measure indexed by the parameter ρ because, in this paper, we will often consider somewhat complicated local-to-unity parameters and subsequences of such parameters, which are less conveniently expressed in terms of subscripts.

confidence interval based on the statistic $\mathbb{M}(\rho)$ can be taken to be

$$\mathbb{C}_{\alpha}^{\mathbb{M}} = \left\{\rho \in (-1,1] : -z_{\alpha/2} \leq \mathbb{M}(\rho) \leq z_{\alpha/2}\right\} \tag{5}$$

It is immediate from Theorem 3.2 that the confidence procedure defined by (5) is asymptotically valid in the sense that its coverage probability is equal to the nominal level $1-\alpha$ in large samples, uniformly over $\rho \in (-1,1]$.

(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 $\kappa \in (0,\infty)$ 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 for a whole range of possible paths indexed by $\kappa \in (0,\infty)$; and, hence, our results do not require the kind of restrictions on the relative magnitudes of N and T that are often imposed in other asymptotic analysis of panel data models. Indeed, by allowing T to grow as any positive (real-valued) power of N, our framework can accommodate a wide variety of settings where T may be of smaller, larger, or similar order of magnitude as N.

(iii) As noted earlier and as is evident from the proof of Theorem 3.2 given in the Appendix, uniform convergence here is established by showing convergence to the same distribution under every parameter sequence in the parameter space. To the best of our knowledge, the use of this approach in statistics originated with the book on large sample theory by Lehmann (1999). Important extensions as well as applications of this approach to a variety of econometric models and inferential procedures have also been made more recently in the papers by Andrews and Guggenberger (2009) and Andrews, Cheng, and Guggenberger (2011).

(iv) As we noted earlier in the Introduction, 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. As also noted in the Introduction, 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, SC-1, and SC-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 $\mathbb{C}_\alpha^{\mathbb{M}}$ can be a very substantial fraction of the width of the entire parameter space when $\rho = 1$. 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 the informativeness of the confidence 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

$$\begin{aligned}\mathbb{C}_{\gamma,\alpha,N,T} &= \mathbb{I}\left\{\mathbb{T}_1 \leq -z_{\gamma_1}\right\}\mathbb{I}\left\{\mathbb{T}_2 \leq -z_{\gamma_2}\right\}\mathbb{C}_{\alpha_1}^{\mathbb{M}} + \mathbb{I}\left\{\mathbb{T}_1 > -z_{\gamma_1}\right\}\mathbb{C}_{\gamma_1,\alpha_2}^{\text{UR1}} \\ &\quad + \mathbb{I}\left\{\mathbb{T}_1 \leq -z_{\gamma_1}\right\}\mathbb{I}\left\{\mathbb{T}_2 > -z_{\gamma_2}\right\}\mathbb{C}_{\gamma_2,\alpha_2}^{\text{UR2}}, \end{aligned} \tag{6}$$

where $\mathbb{C}_{\alpha_1}^{\mathbb{M}}$ is as defined in (5) above,

$$\mathbb{C}_{\gamma_1,\alpha_2}^{\text{UR1}} = \left\{\rho \in (-1,1] : 1 - \frac{\sqrt{2}\left(z_{\gamma_1} + z_{\alpha_2}\right)}{T\sqrt{N}} \leq \rho \leq 1\right\}, \text{ and} \tag{7}$$

$$\mathbb{C}_{\gamma_2,\alpha_2}^{\text{UR2}} = \left\{\rho \in (-1,1] : 1 - \frac{2\left(z_{\gamma_2} + z_{\alpha_2}\right)}{\sqrt{NT}} \leq \rho \leq 1\right\} \tag{8}$$

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

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

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 $\widehat{\omega}_{\mathrm{IVL}} = \widehat{\sigma}^{-2}N^{-1/2}T^{-1/2}\sum_{i=1}^{N}\sum_{t=3}^{T}\Delta y_{it-1}y_{it-1}$, 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 (6), 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 $\mathbb{C}^{\mathrm{UR1}}_{\gamma_1,\alpha_2}$. 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 $\mathbb{C}^{\mathrm{UR2}}_{\gamma_2,\alpha_2}$. 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 $\kappa \in (0,\infty)$ and $\tau \in (0,\infty)$. Set $N = N(T) = (\tau T)^{1/\kappa}$ and $C_{\gamma,\alpha,N,T} = C_{\gamma,\alpha,N(T),T} = C_{\alpha,T}$. Suppose that Assumptions 1, 3, 4 and either 2 or 2 hold. Then,*

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

Remarks 3.2:

(i) The pre-test based confidence procedure proposed here is inspired by the 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 that differs from the one studied in Lepski (1999) and because we use a dual pre-test framework, the construction and analysis of our procedure also differ,

even though we use the same idea to improve set estimation accuracy.

(ii) Since

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

it follows that the result obtained in Theorem 3.3, i.e.,

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

$\liminf_{T\to\infty}\inf_{\rho\in(-1,1]}\Pr\left(\rho\in\mathbb{C}_{\gamma,\alpha,T}|\rho\right)\geq 1-\alpha$, so that the proposed confidence interval has asymptotic coverage probability that is at least the nominal level $1-\alpha$ uniformly over $\rho\in(-1,1]$.

(iii) In the procedure given by (6), 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 the Introduction and 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 of 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 $H_0:\rho=1$, 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 $\rho_T=\exp\left\{-1/q\left\{T\right\}\right\}$ such that $\max\left\{T,\sqrt{NT}\right\}\ll q\left(T\right)\ll\sqrt{N}T$, 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, $\mathrm{wid}\left(\mathbb{C}^{\mathbb{M}}_{\alpha_1}\right)=O_p\left(N^{-1/2}T^{-3/2}q\left(T\right)\right)$. The result is that we can make further improvement by using the interval $\mathbb{C}^{\mathrm{UR2}}_{\gamma_2,\alpha_2}$ which has width $\mathrm{wid}\left(\mathbb{C}^{\mathrm{UR2}}_{\gamma_2,\alpha_2}\right)=O_p\left(N^{-1/2}T^{-1/2}\right)=o_p\left(N^{-1/2}T^{-3/2}q\left(T\right)\right)$. 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 $\mathbb{C}^{\mathbb{M}}_{\alpha_1}$. 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.[8]

(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 $\mathbb{C}^{\text{UR1}}_{\gamma_1,\alpha_2}$. 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 $\mathbb{C}^{\text{UR1}}_{\gamma_1,\alpha_2}$ but could also lead to a greater chance that $\mathbb{T}_1$ will falsely reject the null hypothesis and switch to either $\mathbb{C}^{\text{UR2}}_{\gamma_2,\alpha_2}$ or $\mathbb{C}^{\mathbb{M}}_{\alpha_1}$, both of which are wider than $\mathbb{C}^{\text{UR1}}_{\gamma_1,\alpha_2}$ 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 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 $\mathbb{C}^{\mathbb{M}}_{\alpha_1}$, we will still end up with asymptotic coverage probability greater than the nominal level $1-\alpha$ since $\mathbb{C}^{\mathbb{M}}_{\alpha_1}$ 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 $\mathbb{C}^{\mathbb{M}}_{\alpha_1}$ which controls the asymptotic coverage probability properly.

[8] A 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. The focus of the Bun and Kleibergen (2014) paper differs substantially from that of the present 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 inferential 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.

(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 $\mathbb{C}^{\text{UR1}}_{\gamma_1,\alpha_2}$, for which $\mathbb{T}_1$ may fail to reject $H_0 : \rho = 1$ even in large samples. In addition, there may be parameter sequences which lie just outside of $\mathbb{C}^{\text{UR2}}_{\gamma_2,\alpha_2}$, 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[9]. Hence, by constructing $\mathbb{C}^{\text{UR1}}_{\gamma_1,\alpha_2}$ and $\mathbb{C}^{\text{UR2}}_{\gamma_2,\alpha_2}$ in the manner suggested above, we can properly control the probability of not switching to $\mathbb{C}^{\mathbb{M}}_{\alpha_1}$ 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 $\mathbb{C}^{\mathbb{M}}_{\alpha_1}$ 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 $\mathbb{C}^{\mathbb{M}}_{\alpha_1}$, 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 the initial condition. Hence, the design of our procedure has taken into consideration certain trade-offs on the competing goals of interval accuracy, computational simplicity, and the 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

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

[9]The statement of Lemma A1 is given in the Appendix of the paper. Its proof is lengthy and, it is therefore placed in the technical supplement.

where $\{\varepsilon_{it}\} \equiv i.i.d.N(0,1)$ and $\{a_i\} \equiv i.i.d.\ N(2,1)$. We vary $\rho = 1.00$, 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 (6) 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	0.049	0.001
$\mathbb{C}_{\text{PCI3}}$	0.05	0.05	0.025	0.025
$\mathbb{C}_{\text{PCI4}}$	0.05	0.05	0.049	0.001

Tables 1-12 below provide simulation results comparing the four PCI procedures described above with the $\mathbb{C}^{\mathbb{M}}_{0.05}$ procedure given in (5) 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, 7, 9, and 11 report the average width of the confidence intervals under each of forty-eight experimental settings, obtained by varying ρ, N, T, and w_{i0}. In addition, in Tables 6, 8, 10, and 12; we report of number of instances out of 10,000 simulation repetitions that a particular confidence procedure leads to an empty interval, which occurs when the intersection of the (unrestricted) interval and the parameter space is the null set. For example, in the case of the $\mathbb{C}^{\mathbb{M}}_{0.05}$ procedure, an empty interval would arise if

$$\begin{aligned}\left\{\rho \in (-1,1] : -z_{\alpha/2} \leq \mathbb{M}(\rho) \leq z_{\alpha/2}\right\} &= (-1,1] \cap \left\{\rho \in \mathbb{R} : -z_{\alpha/2} \leq \mathbb{M}(\rho) \leq z_{\alpha/2}\right\} \\ &= \emptyset.\end{aligned}$$

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. On the other hand, 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 $\mathbb{C}^{\text{UR1}}_{\gamma_1,\alpha_2}$ or $\mathbb{C}^{\text{UR2}}_{\gamma_2,\alpha_2}$, 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 $\rho = 1$ 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.

A possible deficiency of the $\mathbb{C}^{\mathbb{M}}_{0.05}$ procedure as shown in Tables 5, 7, 9, and 11 is that the average widths of intervals obtained from this procedure are substantially wider than that of the other procedures when $\rho = 1$. Moreover, in the $\rho = 1$ case, the use of the $\mathbb{C}^{\mathbb{M}}_{0.05}$ procedure results in empty intervals in roughly 2.61% of the times, ranging from a low of 215 empty intervals (out of 10,000 repetitions) in the case with $N = 100, T = 50$, and $w_{i0} = 0$ to a high of 295 empty intervals (out of 10,000 repetitions) in the case where $N = 100, T = 100$, and $w_{i0} = 2$.[10] In contrast, no empty interval is observed for any of the pre-test procedures in any of the 48 experiments, including experiments where $\rho = 1$. It should also be noted that, outside of the unit root case, the results of Tables 5-12 do show $\mathbb{C}^{\mathbb{M}}_{0.05}$ to provide informative intervals with average widths that are much smaller than those in the $\rho = 1$ case. In addition, as the true value of ρ moves significantly away from unity, such as in the cases where $\rho = 0.95$, 0.9, 0.8, and 0.6; empty intervals were no longer observed for $\mathbb{C}^{\mathbb{M}}_{0.05}$.

For the four alternative specifications of PCI, there does not seem to be a great deal of differences in their performance across the different experiments, 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 $\rho = 1$. 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 0.1446, 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

[10] It might initially seem strange in Tables 6 and 8 that in the cases where $\rho = 1$ and $N = 100$, the number of empty intervals for $\mathbb{C}^{\mathbb{M}}_{0.05}$ actually increased as the sample size in the time dimension increased from $T = 50$ to $T = 100$. However, there is an intuitive explanation for this result. As noted earlier, in the unit root case, the rate of concentration of the width of the $\mathbb{C}^{\mathbb{M}}_{0.05}$ interval is $O_p\left(T^{-1/2}\right)$, so that intervals obtained under this procedure are wider in the $T = 50$ case than in the $T = 100$ case, leading to a higher chance of a non-null intersection with the parameter space.

strictly less than 0.1 in 38 of the experiments. On the other hand, PCI2 sets $\alpha_1 = 0.049$ 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 downside 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. More specifically, the coverage probabilities of the IV-based confidence intervals are especially poor when ρ is unity or near-unity, whereas the coverage probabilities for the POLS-based confidence intervals begin to deviate dramatically from the nominal level when $\rho = 0.95$ or less. Moreover, from the results reported in Tables 6, 8, 10, and 12, we note that CI_{IVD}, the confidence procedure based on inverting the Studentized statistic associated with $\widehat{\rho}_{\text{IVD}}$, leads to an empty interval in more than 40% of the simulation runs when $\rho = 1$. This is perhaps not surprising since, as shown in Theorem SA-1 in Appendix SA of the supplement to this paper, $\widehat{\rho}_{\text{IVD}}$ is not uniformly convergent over the parameter space and does not have an asymptotic normal distribution when the true ρ equals unity. Hence, when $\rho = 1$, the CI_{IVD} procedure, which is designed to achieve the correct asymptotic coverage in the stationary case, will not only exhibit poor coverage probabilities but will often deliver intervals that lie entirely outside the parameter space. Interestingly, even though the CI_{POLS} procedure is based on the correct asymptotics when $\rho = 1$, it nevertheless produces some empty intervals in the unit root case as shown in Tables 6, 8, 10, and 12. This suggests a need to modify the usual t-ratio based confidence procedure in cases where there is interest in a point on the boundary of a bounded parameter space, such as $\rho = 1$.

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

ρ	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	0.9996
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	0.9934	0.9891	0.9828
0.99	100	0.9443	0.6239	0.9448	0.9918	0.9850	0.9872	0.9752
0.95	50	0.7995	0.8046	0.9369	0.9839	0.9678	0.9839	0.9678
0.95	100	0.6816	0.8911	0.9445	0.9874	0.9749	0.9874	0.9743
0.90	50	0.2384	0.8738	0.9376	0.9833	0.9649	0.9758	0.9491
0.90	100	0.0507	0.9223	0.9465	0.9715	0.9489	0.9715	0.9476
0.80	50	0.0002	0.9055	0.9378	0.9677	0.9386	0.9677	0.9386
0.80	100	0.0000	0.9254	0.9421	0.9705	0.9432	0.9705	0.9432
0.60	50	0.0000	0.9162	0.9351	0.9650	0.9361	0.9650	0.9361
0.60	100	0.0000	0.9335	0.9425	0.9700	0.9435	0.9700	0.9435

Results based on 10,000 simulations

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

ρ	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	0.9929	0.9878	0.9792
0.99	100	0.9449	0.6648	0.9370	0.9910	0.9859	0.9856	0.9750
0.95	50	0.8056	0.8720	0.9164	0.9752	0.9595	0.9748	0.9578
0.95	100	0.6864	0.9172	0.9326	0.9842	0.9706	0.9839	0.9682
0.90	50	0.2498	0.9227	0.9150	0.9715	0.9490	0.9624	0.9322
0.90	100	0.0546	0.9376	0.9359	0.9632	0.9397	0.9634	0.9372
0.80	50	0.0002	0.9353	0.9198	0.9567	0.9213	0.9567	0.9213
0.80	100	0.0000	0.9393	0.9313	0.9644	0.9331	0.9644	0.9331
0.60	50	0.0000	0.9357	0.9225	0.9563	0.9250	0.9563	0.9250
0.60	100	0.0000	0.9414	0.9358	0.9657	0.9376	0.9657	0.9376

Results based on 10,000 simulations

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

ρ	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	0.9786	0.9856	0.9742
0.99	200	0.9409	0.8101	0.9483	0.9867	0.9759	0.9863	0.9748
0.95	100	0.4377	0.8949	0.9436	0.9844	0.9696	0.9782	0.9558
0.95	200	0.1796	0.9243	0.9444	0.9736	0.9467	0.9736	0.9461
0.90	100	0.0010	0.9186	0.9451	0.9705	0.9462	0.9705	0.9462
0.90	200	0.0000	0.9349	0.9475	0.9742	0.9481	0.9742	0.9481
0.80	100	0.0000	0.9320	0.9447	0.9740	0.9457	0.9740	0.9457
0.80	200	0.0000	0.9353	0.9422	0.9715	0.9433	0.9715	0.9433
0.60	100	0.0000	0.9368	0.9452	0.9707	0.9466	0.9707	0.9466
0.60	200	0.0000	0.9439	0.9482	0.9732	0.9490	0.9732	0.9490

Results based on 10,000 simulations

Table 4: Coverage Probabilities (nominal level=0.95)
$N = 200$, $w_{i0} = 2$

ρ	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	0.9996
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	0.9773	0.9824	0.9712
0.99	200	0.9421	0.8297	0.9441	0.9857	0.9755	0.9847	0.9728
0.95	100	0.4475	0.9167	0.9327	0.9802	0.9614	0.9725	0.9473
0.95	200	0.1850	0.9325	0.9401	0.9702	0.9423	0.9702	0.9413
0.90	100	0.0009	0.9351	0.9341	0.9622	0.9354	0.9622	0.9354
0.90	200	0.0000	0.9420	0.9411	0.9703	0.9431	0.9703	0.9431
0.80	100	0.0000	0.9436	0.9364	0.9683	0.9372	0.9683	0.9372
0.80	200	0.0000	0.9433	0.9398	0.9678	0.9407	0.9678	0.9407
0.60	100	0.0000	0.9419	0.9374	0.9669	0.9392	0.9669	0.9392
0.60	200	0.0000	0.9469	0.9447	0.9710	0.9456	0.9710	0.9456

Results based on 10,000 simulations

Table 5: Average Width of Confidence Intervals
$N = 100$, $w_{i0} = 0$

ρ	T	CI_{POLS}	CI_{IVD}	$CI_{\mathbb{M}}$	CI_{PCI1}	CI_{PCI2}	CI_{PCI3}	CI_{PCI4}
1.00	50	0.0059	0.0493	0.9810	0.0133	0.0168	0.0161	0.0204
1.00	100	0.0030	0.0361	0.7866	0.0070	0.0088	0.0090	0.0114
0.99	50	0.0126	0.0576	0.2362	0.1001	0.1233	0.1235	0.1456
0.99	100	0.0073	0.0452	0.0947	0.0898	0.1105	0.0866	0.1029
0.95	50	0.0184	0.0902	0.1215	0.1385	0.1504	0.1357	0.1367
0.95	100	0.0123	0.0721	0.0838	0.0963	0.0967	0.0944	0.0889
0.90	50	0.0234	0.1063	0.1273	0.1446	0.1311	0.1442	0.1285
0.90	100	0.0161	0.0764	0.0838	0.0958	0.0842	0.0958	0.0842
0.80	50	0.0297	0.1052	0.1171	0.1339	0.1177	0.1339	0.1177
0.80	100	0.0207	0.0744	0.0784	0.0897	0.0788	0.0897	0.0788
0.60	50	0.0369	0.0992	0.1057	0.1209	0.1062	0.1209	0.1062
0.60	100	0.0259	0.0701	0.0724	0.0828	0.0727	0.0828	0.0727

Results based on 10,000 simulations

Table 6: Number of Empty Intervals (out of 10,000)
$N = 100$, $w_{i0} = 0$

ρ	T	CI_{POLS}	CI_{IVD}	$CI_{\mathbb{M}}$	CI_{PCI1}	CI_{PCI2}	CI_{PCI3}	CI_{PCI4}
1.00	50	185	4400	215	0	0	0	0
1.00	100	169	4339	263	0	0	0	0
0.99	50	0	2768	243	0	0	0	0
0.99	100	0	1347	164	0	0	0	0
0.95	50	0	62	9	0	0	0	0
0.95	100	0	0	0	0	0	0	0

There are no empty intervals for any of the procedures in the cases $\rho = 0.9, 0.8$, and 0.6.

Table 7: Average Width of Confidence Intervals
$N = 100$, $w_{i0} = 2$

ρ	T	CI_{POLS}	CI_{IVD}	$CI_{\mathbb{M}}$	CI_{PCI1}	CI_{PCI2}	CI_{PCI3}	CI_{PCI4}
1.00	50	0.0059	0.0491	0.9114	0.0132	0.0166	0.0152	0.0195
1.00	100	0.0030	0.0360	0.7521	0.0069	0.0087	0.0091	0.0112
0.99	50	0.0126	0.0576	0.1821	0.1005	0.1242	0.1172	0.1402
0.99	100	0.0073	0.0457	0.0844	0.0888	0.1096	0.0831	0.0999
0.95	50	0.0182	0.0925	0.1025	0.1260	0.1409	0.1182	0.1220
0.95	100	0.0122	0.0728	0.0764	0.0896	0.0914	0.0867	0.0823
0.90	50	0.0230	0.1071	0.1054	0.1210	0.1109	0.1200	0.1069
0.90	100	0.0160	0.0764	0.0757	0.0866	0.0761	0.0866	0.0761
0.80	50	0.0292	0.1052	0.0995	0.1137	0.0999	0.1137	0.0999
0.80	100	0.0206	0.0744	0.0722	0.0826	0.0726	0.0826	0.0726
0.60	50	0.0366	0.0992	0.0954	0.1091	0.0958	0.1091	0.0958
0.60	100	0.0258	0.0701	0.0688	0.0786	0.0691	0.0786	0.0691

Results based on 10,000 simulations

Table 8: Number of Empty Intervals (out of 10,000)
$N = 100$, $w_{i0} = 2$

ρ	T	CI_{POLS}	CI_{IVD}	$CI_{\mathbb{M}}$	CI_{PCI1}	CI_{PCI2}	CI_{PCI3}	CI_{PCI4}
1.00	50	185	4360	270	0	0	0	0
1.00	100	169	4363	295	0	0	0	0
0.99	50	0	2411	275	0	0	0	0
0.99	100	0	1144	184	0	0	0	0
0.95	50	0	12	6	0	0	0	0
0.95	100	0	0	0	0	0	0	0

There are no empty intervals for any of the procedures in the cases $\rho = 0.9, 0.8$, and 0.6.

Table 9: Average Width of Confidence Intervals
$N = 200$, $w_{i0} = 0$

ρ	T	CI_{POLS}	CI_{IVD}	$CI_{\mathbb{M}}$	CI_{PCI1}	CI_{PCI2}	CI_{PCI3}	CI_{PCI4}
1.00	100	0.0021	0.0253	0.7838	0.0048	0.0060	0.0066	0.0083
1.00	200	0.0010	0.0178	0.6377	0.0027	0.0033	0.0037	0.0047
0.99	100	0.0051	0.0340	0.0696	0.0662	0.0796	0.0657	0.0751
0.99	200	0.0032	0.0272	0.0384	0.0454	0.0544	0.0429	0.0487
0.95	100	0.0087	0.0540	0.0635	0.0722	0.0654	0.0720	0.0641
0.95	200	0.0060	0.0387	0.0421	0.0481	0.0423	0.0481	0.0423
0.90	100	0.0114	0.0540	0.0592	0.0677	0.0595	0.0677	0.0595
0.90	200	0.0080	0.0382	0.0400	0.0457	0.0402	0.0457	0.0402
0.80	100	0.0147	0.0526	0.0554	0.0634	0.0557	0.0634	0.0557
0.80	200	0.0103	0.0372	0.0382	0.0437	0.0384	0.0437	0.0384
0.60	100	0.0183	0.0496	0.0512	0.0585	0.0514	0.0585	0.0514
0.60	200	0.0129	0.0351	0.0356	0.0407	0.0358	0.0407	0.0358

Results based on 10,000 simulations

Table 10: Number of Empty Intervals (out of 10,000)
$N = 200$, $w_{i0} = 0$

ρ	T	CI_{POLS}	CI_{IVD}	$CI_{\mathbb{M}}$	CI_{PCI1}	CI_{PCI2}	CI_{PCI3}	CI_{PCI4}
1.00	100	208	4538	256	0	0	0	0
1.00	200	241	4685	247	0	0	0	0
0.99	100	0	1108	114	0	0	0	0
0.99	200	0	252	51	0	0	0	0
0.95	100	0	0	0	0	0	0	0
0.95	200	0	0	0	0	0	0	0

There are no empty intervals for any of the procedures in the cases $\rho = 0.9, 0.8$, and 0.6.

Table 11: Average Width of Confidence Intervals
$N = 200$, $w_{i0} = 2$

ρ	T	CI_{POLS}	CI_{IVD}	$CI_{\mathbb{M}}$	CI_{PCI1}	CI_{PCI2}	CI_{PCI3}	CI_{PCI4}
1.00	100	0.0021	0.0252	0.7507	0.0048	0.0060	0.0064	0.0081
1.00	200	0.0010	0.0179	0.6201	0.0027	0.0033	0.0037	0.0047
0.99	100	0.0051	0.0344	0.0623	0.0647	0.0783	0.0622	0.0720
0.99	200	0.0031	0.0274	0.0365	0.0448	0.0539	0.0417	0.0477
0.95	100	0.0086	0.0542	0.0571	0.0653	0.0594	0.0650	0.0578
0.95	200	0.0060	0.0387	0.0398	0.0455	0.0400	0.0455	0.0400
0.90	100	0.0113	0.0540	0.0535	0.0612	0.0538	0.0612	0.0538
0.90	200	0.0079	0.0382	0.0380	0.0435	0.0382	0.0435	0.0382
0.80	100	0.0145	0.0526	0.0511	0.0584	0.0513	0.0584	0.0513
0.80	200	0.0103	0.0372	0.0366	0.0419	0.0368	0.0419	0.0368
0.60	100	0.0182	0.0496	0.0486	0.0556	0.0488	0.0556	0.0488
0.60	200	0.0129	0.0351	0.0347	0.0397	0.0349	0.0397	0.0349

Results based on 10,000 simulations

Table 12: Number of Empty Intervals (out of 10,000)
$N = 200$, $w_{i0} = 2$

ρ	T	CI_{POLS}	CI_{IVD}	$CI_{\mathbb{M}}$	CI_{PCI1}	CI_{PCI2}	CI_{PCI3}	CI_{PCI4}
1.00	100	208	4552	280	0	0	0	0
1.00	200	241	4608	261	0	0	0	0
0.99	100	0	906	127	0	0	0	0
0.99	200	0	209	52	0	0	0	0
0.95	100	0	0	0	0	0	0	0
0.95	200	0	0	0	0	0	0	0

There are no empty intervals for any of the procedures in the cases $\rho = 0.9, 0.8$, and 0.6.

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 a wide parameter space that includes unity. 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. The panel AR model considered here is a simple model. But it is the kernel of all dynamic panel models and embodies all the characteristics that make uniform inference and confidence interval construction difficult. Even in the time series case these problems are well known to be challenging. In the panel case, the challenges are accentuated by additional issues arising from the presence of incidental effects and multi-index limit theory. The pre-test confidence interval solution proposed here addresses these challenges and has potential for application in more complex models.

Acknowledgments:

We thank Chris Sims for a thought-provoking comment which led us to a substantial improvement in some of our results. We also thank participants at the 2017 Greater New York Metropolitan Area Econometrics Colloquium, 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 Boston College, Indiana University Bloomington, Pennsylvania State University, and the University of Pennsylvania for comments. Phillips acknowledges support from the NSF under Grant SES 1258258, the Kelly Fund at the University of Auckland, and an LKC Fellowship at Singapore Management University.

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] Chang, Y. 2002. Nonlinear IV Unit Root Tests in Panel with Cross-sectional Dependency. *Journal of Econometrics* 110: 261-292.

[13] Chang, Y. 2004. Bootstrap Unit Root Tests in Panels with Cross-sectional Dependency. *Journal of Econometrics* 120: 263-293.

[14] Chao, J., and P. C. B. Phillips. 2019. Technical Supplement to 'Uniform Inference in Panel Autoregression. Working Paper, University of Maryland.

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

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

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

[18] 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.

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

[20] Han, C., P. C. B. Phillips, and D. Sul. 2014. X-Differencing and Dynamic Panel Model Estimation. *Econometric Theory* 30: 201-251.

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

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

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

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

[25] Mikusheva, A. 2007. Uniform Inference in Autoregressive Models. *Econometrica* 75: 1411-1452.

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

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

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

[29] Moon, H. R., B. Perron, and P. C. B. Phillips. 2015. Incidental Parameters and Dynamic Panel Modeling. In *The Oxford Handbook of Panel Data*. Edited by B. Baltagi. New York: Oxford University Press.

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

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

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

[33] Phillips, P. C. B. 1987. Time Series Regression with a Unit Root. *Econometrica* 55: 277-301.

[34] Phillips, P. C. B. 2014. On Confidence Intervals for Autoregressive Roots and Predictive Regression. *Econometrica* 82: 1177-1195.

[35] 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.

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

[37] Phillips, P. C. B., and D. Sul. 2003. Dynamic Panel Estimation and Homogeneity Testing under Cross Sectional Dependence", *Econometrics Journal* 6: 217-259.

[38] 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, 2019). These results are designated in the derivations that follow by use of the prefix S. Lemmas A1 and A2 are stated in this Appendix and their proofs are given in the Online Supplement.

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

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

so that

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

We turn first to part (a). In this case, by assumption, $\rho_T = 1$ for all T sufficiently large. Under the random-effects specification given by Assumption 2, we can apply parts (g) and

(i) of Lemma SD-11, part (a) of Lemma SD-25, and part (a) of Lemma SC-13 to obtain

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

It follows from applying Lemma SD-24 that $\mathbb{M}(\rho_T) \Rightarrow N(0,1)$, as required. Moreover, it is easily seen that, by applying part (b) of Lemma SE-1 in lieu of part (g) of Lemma SD-11 in the argument given above, the same result can be obtained under the fixed-effects specification given by Assumption 2*.

Next consider part (b), where we take $\rho_T = \exp\{-1/q(T)\}$ such that $T/q(T) \to 0$. In the case of the random-effects specification given by Assumption 2, we can use the results in parts (g) and (i) of Lemma SD-11, part (b) of Lemma SD-25, and part (b) of Lemma SC-13 to deduce that

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

It follows from part (a) of Lemma SD-22 that $\mathbb{M}(\rho_T) \Rightarrow N(0,1)$. Moreover, it is easily seen that, by applying part (b) of Lemma SE-1 in lieu of part (g) of Lemma SD-11 in the argument given above, the same result can be obtained under the fixed-effects specification given by Assumption 2*.

Consider part (c), where we take $\rho_T = \exp\{-1/q(T)\}$ such that $q(T) \sim T$. Under the random-effects specification given by Assumption 2, we can apply parts (g) and (i) of Lemma

SD-11, part (c) of Lemma SD-25, and part (c) of Lemma SC-13 to deduce that

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

where $\overline{\omega}_T=\sigma^2\left\{1+\left[q(T)/2T\right]\left[1-\exp\left\{-2T/q(T)\right\}\right]\right\}^{1/2}$. It follows from part (b) of Lemma SD-22 that $\mathbb{M}(\rho_T)\Rightarrow N(0,1)$. Moreover, it is easily seen that, by applying part (b) of Lemma SE-1 in lieu of part (g) of Lemma SD-11 in the argument given above, the same result can be obtained under the fixed-effects specification given by Assumption 2*.

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

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

Under the random-effects specification given by Assumption 2, we further apply parts (g) and (i) of Lemma SD-11 and part (d) of Lemma SD-25 to obtain

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

By part (c) of Lemma SD-22, we then deduce that $\mathbb{M}(\rho_T)\Rightarrow N(0,1)$, as required for (d). Moreover, it is easily seen that, by applying part (b) of Lemma SE-1 in lieu of part (g) of Lemma SD-11 in the argument given above, the same result can be obtained under the fixed-effects specification given by Assumption 2*.

Finally, to show part (e), we first consider the random-effects specification given by Assumption 2. In this case, note that, by applying parts (g) and (i) of Lemma SD-11, part

(e) of Lemma SD-21, and part (e) of Lemma SC-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 SD-23 that $\mathbb{M}\left(\rho_T\right) = \sqrt{\left(1+\rho_T\right)/\left(2\sigma^4\right)}N^{-1/2}\sum_{i=1}^{N}\left(X_{i,T}+Y_{i,T}\right)+o_p\left(1\right) \Rightarrow N\left(0,1\right)$. Moreover, it is easily seen that, by applying part (b) of Lemma SE-1 in lieu of part (g) of Lemma SD-11 in the argument given above, the same result can be obtained under the fixed-effects specification given by Assumption 2*. □

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 $\mathcal{G}_{s_k}^{\mathbb{M}}$.[11] 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 $\{\rho_T\} = \bigcup_{j=1}^{d}\left\{\rho_{j,T_j}\right\}$, 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 $\Pr\left(\rho_{k,T} \notin \mathbb{C}_{\alpha_1,T}^{\mathbb{M}}|\rho_{k,T}\right) = \Pr\left(\left|\mathbb{M}_T\left(\rho_{k,T}\right)\right| > z_{\alpha_1/2}|\rho_{k,T}\right)$. 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 $T \geq M_k$,

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

where $\mathcal{Z} \sim N\left(0,1\right)$. 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 SD-33 (given in Appendix SD in the technical supplement to this paper),

[11] 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.

from which we further deduce that

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

Next, let $\overline{M} = \max\left\{f_1\left(M_1\right), ..., f_k\left(M_d\right)\right\}$. Consider any positive integer $T \geq \overline{M}$; we must have $T = f_k\left(T^*\right)$ for some $k = 1, ..., d$ and for some $T^* \in \mathbb{N}$. Given that $T = f_k\left(T^*\right) \geq \overline{M} \geq f_k\left(M_k\right)$, we also deduce that $T \geq T^* \geq M_k$ by Lemma SD-33 since $f_k\left(\cdot\right)$ is an increasing function of its argument. It follows that for every sequence $\{\rho_T\}$ and for all $T \geq \overline{M}$

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

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

Lemma A1:

Let $\rho_T = \exp\left\{-1/q\left(T\right)\right\}$, and suppose that Assumptions 1, 3, 4, and either 2 or 2* hold. Then, the following statements are true as $N, T \to \infty$ such that $N^\kappa/T = \tau$, for constants $\kappa \in (0, \infty)$ and $\tau \in (0, \infty)$.

(a) Let $\mathcal{G}_1^{\mathrm{P}} = \left\{\{\rho_T\} : \rho_T = 1 \text{ for all } T \text{ sufficiently large}\right\}$, and set $N = N\left(T\right) = \left(\tau T\right)^{1/\kappa}$ and $\mathbb{C}_{\gamma,\alpha,N,T} = \mathbb{C}_{\gamma,\alpha,N(T),T} = \mathbb{C}_{\gamma,\alpha,T}$. Then, for $\{\rho_T\} \in \mathcal{G}_1^{\mathrm{P}}$,

$$\limsup_{T\to\infty} \Pr\left(\rho_T \notin \mathbb{C}_{\gamma,\alpha,T}|\rho_T\right) \leq \alpha_1 < \alpha.$$

(b) Let $\mathcal{G}_2^{\mathrm{P}} = \left\{\{\rho_T\} : \rho_T = \exp\left\{\frac{1}{q(T)}\right\}, \sqrt{N}T \ll q\left(T\right)\right\}$, and set $N = N\left(T\right) = \left(\tau T\right)^{1/\kappa}$ and $\mathbb{C}_{\gamma,\alpha,N,T} = \mathbb{C}_{\gamma,\alpha,N(T),T} = \mathbb{C}_{\gamma,\alpha,T}$. Then, for $\{\rho_T\} \in \mathcal{G}_2^{\mathrm{P}}$,

$$\limsup_{T\to\infty} \Pr\left(\rho_T \notin \mathbb{C}_{\gamma,\alpha,T}|\rho_T\right) \leq \alpha_1 < \alpha.$$

(c) Let

$$\begin{aligned} & \mathcal{G}_3^{\mathrm{P}} \\ = \; & \left\{\{\rho_T\} : \; \rho_T = \exp\left\{\frac{1}{q\left(T\right)}\right\}, \sqrt{N}T \sim q\left(T\right), \text{ and } \rho_T \geq 1 - \frac{\left(z_{\gamma_1} + z_{\alpha_2}\right)\sqrt{2}}{\sqrt{N}T} \text{ eventually}\right\}, \end{aligned}$$

and set $N = N\left(T\right) = \left(\tau T\right)^{1/\kappa}$ and $\mathbb{C}_{\gamma,\alpha,N,T} = \mathbb{C}_{\gamma,\alpha,N(T),T} = \mathbb{C}_{\gamma,\alpha,T}$. Then, for $\{\rho_T\} \in$

$\mathcal{G}_3^{\mathrm{P}}$,

$$\limsup_{T\to\infty} \Pr\left(\rho_T \notin \mathbb{C}_{\gamma,\alpha,T} | \rho_T\right) \leq \alpha_1 < \alpha.$$

(d) Let

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

and set $N = N\left(T\right) = \left(\tau T\right)^{1/\kappa}$ and $\mathbb{C}_{\gamma,\alpha,N,T} = \mathbb{C}_{\gamma,\alpha,N(T),T} = \mathbb{C}_{\gamma,\alpha,T}$. Then, for $\{\rho_T\} \in \mathcal{G}_4^{\mathrm{P}}$,

$$\limsup_{T\to\infty} \Pr\left(\rho_T \notin \mathbb{C}_{\gamma,\alpha,T} | \rho_T\right) \leq \alpha_1 + \alpha_2 = \alpha.$$

(e) Let $\mathcal{G}_5^{\mathrm{P}} = \left\{\{\rho_T\} : \rho_T = \exp\left\{1/q\left(T\right)\right\}, \sqrt{NT} \ll q\left(T\right), \text{ and } T \ll q\left(T\right) \ll \sqrt{N}T\right\}$, and set $N = N\left(T\right) = \left(\tau T\right)^{1/\kappa}$ and $\mathbb{C}_{\gamma,\alpha,N,T} = \mathbb{C}_{\gamma,\alpha,N(T),T} = \mathbb{C}_{\gamma,\alpha,T}$. Then, for $\{\rho_T\} \in \mathcal{G}_5^{\mathrm{P}}$,

$$\limsup_{T\to\infty} \Pr\left(\rho_T \notin \mathbb{C}_{\gamma,\alpha,T} | \rho_T\right) \leq \alpha_1 < \alpha.$$

(f) Let

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

and set $N = N\left(T\right) = \left(\tau T\right)^{1/\kappa}$ and $\mathbb{C}_{\gamma,\alpha,N,T} = \mathbb{C}_{\gamma,\alpha,N(T),T} = \mathbb{C}_{\gamma,\alpha,T}$. Then, for $\{\rho_T\} \in \mathcal{G}_6^{\mathrm{P}}$,

$$\limsup_{T\to\infty} \Pr\left(\rho_T \notin \mathbb{C}_{\gamma,\alpha,T} | \rho_T\right) \leq \alpha_1 < \alpha.$$

(g) Let

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

and set $N = N\left(T\right) = \left(\tau T\right)^{1/\kappa}$ and $\mathbb{C}_{\gamma,\alpha,N,T} = \mathbb{C}_{\gamma,\alpha,N(T),T} = \mathbb{C}_{\gamma,\alpha,T}$. Then, for $\{\rho_T\} \in$

$\mathcal{G}_7^{\mathrm{P}}$,

$$\limsup_{T\to\infty} \Pr\left(\rho_T \notin \mathbb{C}_{\gamma,\alpha,T} | \rho_T\right) \leq \alpha_1 + \alpha_2 = \alpha.$$

(h) Let $\mathcal{G}_8^{\mathrm{P}} = \left\{\{\rho_T\} : \rho_T = \exp\left\{\frac{1}{q(T)}\right\}, \sqrt{NT} \ll q(T) \sim T\right\}$, and 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, for $\{\rho_T\} \in \mathcal{G}_8^{\mathrm{P}}$,

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

(i) Let $\mathcal{G}_9^{\mathrm{P}} = \left\{\{\rho_T\} : \rho_T = \exp\left\{1/q(T)\right\}, \sqrt{NT} \ll q(T) \ll T\right\}$, and 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, for $\{\rho_T\} \in \mathcal{G}_9^{\mathrm{P}}$,

$$\limsup_{T\to\infty} \Pr\left(\rho_T \notin \mathbb{C}_{\gamma,\alpha,T} | \rho_T\right) \leq \alpha_1 < \alpha.$$

(j) Let

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

and set $N = N(T) = \tau T$ and $\mathbb{C}_{\gamma,\alpha,N,T} = \mathbb{C}_{\gamma,\alpha,N(T),T} = \mathbb{C}_{\gamma,\alpha,T}$. Then, for $\{\rho_T\} \in \mathcal{G}_{10}^{\mathrm{P}}$,

$$\limsup_{T\to\infty} \Pr\left(\rho_T \notin \mathbb{C}_{\gamma,\alpha,T} | \rho_T\right) \leq \alpha_1 < \alpha.$$

(k) Let

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

and set $N = N(T) = \tau T$ and $\mathbb{C}_{\gamma,\alpha,N,T} = \mathbb{C}_{\gamma,\alpha,N(T),T} = \mathbb{C}_{\gamma,\alpha,T}$. Then, for $\{\rho_T\} \in \mathcal{G}_{11}^{\mathrm{P}}$,

$$\limsup_{T\to\infty} \Pr\left(\rho_T \notin \mathbb{C}_{\gamma,\alpha,T} | \rho_T\right) \leq \alpha_1 + \alpha_2 = \alpha.$$

(l) Let

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

and 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, for $\{\rho_T\} \in \mathcal{G}_{12}^{\mathrm{P}}$,

$$\limsup_{T\to\infty} \Pr\left(\rho_T \notin \mathbb{C}_{\gamma,\alpha,T} | \rho_T\right) \leq \alpha_1 < \alpha.$$

(m) Let

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

and 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, for $\rho_T \in \mathcal{G}_{13}^{\mathrm{P}}$,

$$\limsup_{T\to\infty} \Pr\left(\rho_T \notin \mathbb{C}_{\gamma,\alpha,T} | \rho_T\right) \leq \alpha_1 + \alpha_2 = \alpha.$$

(n) Let

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

and 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, for $\{\rho_T\} \in \mathcal{G}_{14}^{\mathrm{P}}$,

$$\limsup_{T\to\infty} \Pr\left(\rho_T \notin \mathbb{C}_{\gamma,\alpha,T} | \rho_T\right) \leq \alpha_1 < \alpha.$$

(o) Let

$$\mathcal{G}_{15}^{\mathrm{P}} = \left\{ \{\rho_T\} : \rho_T = \exp\left\{ \frac{1}{q(T)} \right\},\ T \ll q(T),\ N^{1/4}T^{1/4} \ll q(T), \text{ and } q(T)/\sqrt{N} = O(1) \right\},$$

and 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, for $\{\rho_T\} \in \mathcal{G}_{15}^{\mathrm{P}}$,

$$\limsup_{T\to\infty} \Pr\left(\rho_T \notin \mathbb{C}_{\gamma,\alpha,T} | \rho_T\right) \leq \alpha_1 < \alpha.$$

(p) Let $\mathcal{G}_{16}^{\mathrm{P}} = \left\{\{\rho_T\} : (T \ll q(T)) \cap \left(q(T)/N^{1/4}T^{1/4} = O(1)\right)\right\}$, and 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, for $\{\rho_T\} \in \mathcal{G}_{16}^{\mathrm{P}}$,

$$\limsup_{T\to\infty} \Pr\left(\rho_T \notin \mathbb{C}_{\gamma,\alpha,T} | \rho_T\right) \leq \alpha_1 < \alpha.$$

(q) Let $\mathcal{G}_{17}^{\mathrm{P}} = \left\{\{\rho_T\} : \rho_T = \exp\{1/q(T)\},\ N^{1/3}T^{1/3} \ll q(T) \sim T \ll \sqrt{NT}\right\}$, and 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, for $\{\rho_T\} \in \mathcal{G}_{17}^{\mathrm{P}}$,

$$\limsup_{T\to\infty} \Pr\left(\rho_T \notin \mathbb{C}_{\gamma,\alpha,T} | \rho_T\right) \leq \alpha_1 < \alpha.$$

(r) Let

$$\mathcal{G}_{18}^{\mathrm{P}} = \left\{\{\rho_T\} : \rho_T = \exp\left\{\frac{1}{q(T)}\right\},\ N^{1/4}T^{1/4} \ll q(T) \sim T, \text{ and } \frac{q(T)}{N^{1/3}T^{1/3}} = O(1)\right\},$$

and 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, for $\{\rho_T\} \in \mathcal{G}_{18}^{\mathrm{P}}$,

$$\limsup_{T\to\infty} \Pr\left(\rho_T \notin \mathbb{C}_{\gamma,\alpha,T} | \rho_T\right) \leq \alpha_1 < \alpha.$$

(s) Let $\mathcal{G}_{19}^{\mathrm{P}} = \left\{\{\rho_T\} : \rho_T = \exp\{1/q(T)\},\ q(T) \sim T, \text{ and } q(T)/\left(N^{1/4}T^{1/4}\right) = O(1)\right\}$, and 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, for $\{\rho_T\} \in \mathcal{G}_{19}^{\mathrm{P}}$,

$$\limsup_{T\to\infty} \Pr\left(\rho_T \notin \mathbb{C}_{\gamma,\alpha,T} | \rho_T\right) \leq \alpha_1 < \alpha.$$

(t) Let

$$\mathcal{G}_{20}^{\mathrm{P}} = \left\{\{\rho_T\} : \rho_T = \exp\left\{\frac{1}{q(T)}\right\},\ N^{1/3}T^{1/3} \ll q(T) \ll \sqrt{NT}, \text{ and } (q(T) \ll T)\right\},$$

and 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, for $\{\rho_T\} \in \mathcal{G}_{20}^{\mathrm{P}}$,

$$\limsup_{T\to\infty} \Pr\left(\rho_T \notin \mathbb{C}_{\gamma,\alpha,T} | \rho_T\right) \leq \alpha_1 < \alpha.$$

(u) Let $\mathcal{G}_{21}^{\mathrm{P}} = \left\{\{\rho_T\} : \rho_T = \exp\{1/q(T)\},\ q(T) \ll T, \text{ and } q(T) \sim N^{1/3}T^{1/3}\right\}$, and 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, for $\{\rho_T\} \in \mathcal{G}_{17}^{\mathrm{P}}$,

$$\limsup_{T\to\infty} \Pr\left(\rho_T \notin \mathbb{C}_{\gamma,\alpha,T} | \rho_T\right) \leq \alpha_1 < \alpha.$$

(v) Let

$$\mathcal{G}_{22}^{\mathrm{P}}=\left\{\{\rho_T\}:\rho_T=\exp\left\{\frac{1}{q\left(T\right)}\right\},q\left(T\right)\to\infty,q\left(T\right)\ll T,\text{ and }q\left(T\right)/N^{1/3}T^{1/3}\to 0\right\},$$

and set $N=N\left(T\right)=\left(\tau T\right)^{1/\kappa}$ and $\mathbb{C}_{\gamma,\alpha,N,T}=\mathbb{C}_{\gamma,\alpha,N(T),T}=\mathbb{C}_{\gamma,\alpha,T}$. Then, for $\{\rho_T\}\in\mathcal{G}_{22}^{\mathrm{P}}$,

$$\limsup_{T\to\infty}\Pr\left(\rho_T\notin\mathbb{C}_{\gamma,\alpha,T}|\rho_T\right)\leq\alpha_1<\alpha.$$

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

Lemma A2:

Suppose that Assumptions 1, 3, 4, and either 2 or 2* hold, and let

$$\mathcal{G}_{23}^{\mathrm{P}}=\left\{\{\rho_T\}:|\rho_T|=\exp\left\{-\frac{1}{q\left(T\right)}\right\},q\left(T\right)\geq 0,\text{ and }q\left(T\right)=O\left(1\right)\text{ as }T\to\infty\right\},$$

Also, let $N,T\to\infty$ such that $N^{\kappa}/T=\tau$, for constants $\kappa\in(0,\infty)$ and $\tau\in(0,\infty)$, so that we can set $N=N\left(T\right)=\left(\tau T\right)^{1/\kappa}$ and $\mathbb{C}_{\gamma,\alpha,N,T}=\mathbb{C}_{\gamma,\alpha,N(T),T}=\mathbb{C}_{\gamma,\alpha,T}$. Then,

$$\limsup_{T\to\infty}\Pr\left(\rho_T\notin\mathbb{C}_{\gamma,\alpha,T}|\rho_T\right)\leq\alpha_1<\alpha,$$

for $\{\rho_T\}\in\mathcal{G}_{19}^{\mathrm{P}}$.

The proof of Lemma A2 is also 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

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

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

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

it suffices to show that

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

for every sequence $\{\rho_T \in (-1,1] : T \geq 1\}$. To proceed, let $\left\{\mathcal{G}_j^{\mathrm{P}} : j = 1, 2, ..., 23\right\}$ be the collections of parameter sequences defined in Lemmas A1 and A2 given above. Moreover, let $\left\{\rho_{k,T}\right\} \in \mathcal{G}_{s_k}^{\mathrm{P}}$ (for $k = 1, ..., 23$), 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 $d \leq 23$, 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

$$\mathbb{N} = \bigcup_{k=1}^{d} \{T_k = f_k(T) : T \in \mathbb{N}\} \tag{9}$$

with $\mathbb{N}$ denoting the set of natural numbers $\{1, 2, ...\}$. Moreover, define $\upsilon_{k,T} = \sup_{m\geq T} \Pr\left(\rho_{k,m} \notin \mathbb{C}_{\gamma,\alpha,m}|\rho_{k,m} \in \mathcal{G}_{s_k}^{\mathrm{P}}\right)$ and $\overline{p}_k = \limsup_{T\to\infty} \Pr\left(\rho_{k,T} \notin \mathbb{C}_{\gamma,\alpha,T}|\rho_{k,T} \in \mathcal{G}_{s_k}^{\mathrm{P}}\right)$. 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 SD-33 given in Appendix SD 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

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

Next, let $L^{\max} = \max\left\{f_1\left(L_1\right), ..., f_d\left(L_d\right)\right\}$. Consider any positive integer $T \geq L^{\max}$; then, (9) implies that $T = f_k\left(T^*\right)$ for some $k = 1, ..., d$ and for some $T^* \in \mathbb{N}$. By the fact that $f_k\left(\cdot\right)$ is an increasing function of its argument, we have that $T = f_k\left(T^*\right) \geq L^{\max} \geq f_k\left(L_k\right) \geq L_k$, from which it follows that for every positive integer $T \geq L^{\max}$

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

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

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

Since ε is arbitrary, we deduce that

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

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