Econometric Theory, **42**, 2026, 443–469.
doi:10.1017/S0266466625000015

TESTING FOR COEFFICIENT RANDOMNESS IN LOCAL-TO-UNITY AUTOREGRESSIONS

MIKIHITO NISHI
Hitotsubashi University

This study proposes a test for coefficient randomness in autoregressive models where the autoregressive coefficient is local to unity, which is empirically relevant given earlier work. Under this specification, we analyze the effect of the correlation between the random coefficient and disturbance on the properties of tests, a matter that remains largely unexplored in the literature. Our analysis reveals that tests proposed in earlier studies can have poor power when the correlation is moderate to large. The test proposed here is designed to have power functions robust to the correlation. A modified version of the test is suggested that can be applied when the disturbance is serially correlated and conditionally heteroskedastic. The test is shown to have better power properties than existing ones in large and finite samples.

1. INTRODUCTION

We consider the random-coefficient autoregressive (RCA) model,

$$y_t = (\rho + \omega v_t) y_{t-1} + \varepsilon_t, \quad t = 1, 2, \ldots, T, \tag{1}$$

where $(\varepsilon_t, v_t)'$ is a random vector with $\mathbb{E}[\varepsilon_t] = \mathbb{E}[v_t] = 0$ and $\mathbb{V}[v_t] = 1$. Model (1) is a generalization of the usual AR(1) model, in that the autoregressive coefficient fluctuates over time around its mean ρ with variance ω^2. Much attention has been paid to estimation and inference in model (1) in the literature; see, e.g., Nicholls and Quinn (1982) Hwang and Basawa (2005), Aue and Horváth (2011), Horváth and Trapani (2019) and references therein.

One aspect of the inferential theory for model (1) that has attracted attention is how to test the hypothesis of $\omega^2 = 0$. This is because uniform estimation theory that is applicable to both the standard AR ($\omega^2 = 0$) and RCA ($\omega^2 > 0$) models is scarce: Estimators constructed under one model are often not robust to the other.[1] Therefore, to know whether $\omega^2 = 0$ plays an important role in choosing the way to estimate parameters.

I am grateful to the editor Peter Phillips, the co-editor Giuseppe Cavaliere, and two anonymous referees for their constructive comments and suggestions. I am greatly indebted to Eiji Kurozumi, my advisor, for discussions and his help, support and encouragement. I also thank the participants in a conference at Osaka University for their helpful comments and suggestions. All errors are mine. Address correspondence to Mikihito Nishi, Graduate School of Economics, Hitotsubashi University, Tokyo, Japan, e-mail: ed225007@g.hit-u.ac.jp.

[1] An exception is Hill and Peng's (2014) estimator, which is applicable when both $\omega^2 = 0$ and $\omega^2 > 0$.

© The Author(s), 2025. Published by Cambridge University Press.

https://doi.org/10.1017/S0266466625000015 Published online by Cambridge University Press

Nicholls and Quinn (1982), Lee (1998) and Nagakura (2009) proposed tests for $H_0 : \omega^2 = 0$ under the condition $\rho^2 + \omega^2 < 1$ or $(\rho, \omega^2) = (1, 0)$. These conditions, however, are restrictive and limit the applicability of those tests in practice, in view of the analyses by Hill and Peng (2014) and Horváth and Trapani (2019). They applied model (1) to several macroeconomic variables and estimated ρ to be near 1 and possibly greater than 1.[2] Their results indicate that RCA models where $\rho^2 + \omega^2$ and ρ may be greater than 1 with ρ near unity are more empirically relevant. Horváth and Trapani (2019) proposed a test for $H_0 : \omega^2 = 0$ valid under such conditions.

Another insight earlier studies provide is that the variation of the autoregressive root is smaller (if it exists) than assumed in the literature. For example, Horváth and Trapani (2019) estimated that 4 out of 7 variables have a positive variance in their coefficient ranging from 7.2×10^{-5} to 8.2×10^{-3}, but they studied the finite-sample power of their test only for $\omega^2 \geq 2.5 \times 10^{-1}$, which are of larger magnitude than observed in practice (as long as macroeconomic variables are concerned).[3] In fact, their test has almost no power when ω^2 is of small magnitude that is typical in empirical studies, as our simulation studies will reveal.

Given these observations, we consider model (1) with ρ near unity and ω^2 near zero and model them as $\rho = \rho_T = 1 + a/T$ ($a \in \mathbb{R}$ is fixed as $T \to \infty$) and $\omega = \omega_T = c/T^{3/4}$.[4] Letting $\omega_T^2 = 0$ yields the (conventional) local-to-unity AR(1) process (e.g., Phillips, 1987b; Elliott, Rothenberg, and Stock, 1996; Campbell and Yogo, 2006; Phillips, 2014). The local-to-zero variance $\omega_T^2 = c^2/T^{3/2}$ has been employed under $\rho_T = 1$ by McCabe and Smith (1998) and Nishi and Kurozumi (2024). Introducing the local-to-zero variance provides us with a convenient framework in which we can evaluate the power of several tests against the alternatives that are close to the null of $\omega^2 = 0$ and relevant for empirical applications.[5,6] The joint specification $\rho_T = 1 + a/T$ and $\omega_T = c/T^{3/4}$ yields a more empirically relevant model than the conventional AR, local-to-unity AR and RCA models. This specification is similar in spirit to the functional local unit root (FLUR) model studied by Bykhovskaya and Phillips (2018, 2020), who model the AR root as $1 + f(t,T)/T$. Our model may be written in the FLUR form as $f(\cdot) = f_1(\cdot) + f_2(\cdot)$ with $f_1(t,T) = a$ and $f_2(t,T) = cv_t T^{1/4}$.

In the literature, tests for $H_0 : \omega^2 = 0$ have often been studied in the stochastic unit root (STUR) model, a special case of model (1) with $\rho = 1$; see McCabe and Tremayne (1995), Leybourne, McCabe, and Tremayne (1996), McCabe and Smith (1998), Distaso (2008), and Nishi and Kurozumi (2024). On the one hand, the STUR model is a generalization of the pure unit root model. But on the other hand,

[2] For instance, Horváth and Trapani (2019) obtained ρ estimates ranging between 0.9884 and 1.0021.

[3] Horváth and Trapani (2023) estimated that a cryptocurrency index has ω^2 around 10^{-1}.

[4] If a depends on T, ρ_T may equivalently be formulated as a mildly integrated or mildly explosive root, which causes problems of identification as discussed in Phillips (2023).

[5] As demonstrated by Nishi and Kurozumi (2024), if we let $\omega_T = cT^{-\kappa}$ for $\kappa < 3/4$, test statistics diverge as $T \to \infty$, in which situation asymptotic power comparison is infeasible.

[6] For example, when $T = 200$ and $0 < c^2 \leq 50$, the variance ω_T^2 takes values between 0 and 1.8×10^{-2}.

https://doi.org/10.1017/S0266466625000015 Published online by Cambridge University Press

the assumption of ρ exactly being unity is a restrictive condition that is unlikely to hold in practice. Our model generalizes the STUR model by allowing $\rho = \rho_T$ to take a value different from unity (but near it).

We will extend the result of Nishi and Kurozumi (2024) and show that the test for $H_0 : \omega^2 = 0$ proposed by Lee (1998) and Nagakura (2009) (hereafter, LN) has a higher local asymptotic power function than other tests when $\mathrm{Corr}(\varepsilon_t, v_t) = 0$ under $\rho_T = 1 + a/T$, but also show that the LN test can perform poorly when $\mathrm{Corr}(\varepsilon_t, v_t) \neq 0$. We will therefore propose several tests for $H_0 : \omega^2 = 0$ whose power properties are robust to the value of $\mathrm{Corr}(\varepsilon_t, v_t)$. One of those tests turns out to be more powerful than the LN test for moderate to large values of $\mathrm{Corr}(\varepsilon_t, v_t)$. To the best of our knowledge, this study is the first to investigate the effect of $\mathrm{Corr}(\varepsilon_t, v_t)$ on the power of tests for $H_0 : \omega^2 = 0$, with the only exception of Su and Roca (2012), who analyzed through simulation this effect in finite samples under $\rho = 1$.

This article also discusses how to remove a nuisance parameter from the null distributions of the test statistics mentioned above. As pointed out by Nagakura (2009), the limiting null distribution of the LN test statistic depends on the correlation between ε_t and $\varepsilon_t^2 - \sigma_\varepsilon^2$, and so do the test statistics proposed in this article. To eliminate this nuisance parameter, we base the proposed tests on the Bonferroni approach as Campbell and Yogo (2006) did in the context of predictive regressions. In fact, we use a modified version of their approach following Phillips (2014) because the original version is known to suffer from serious size distortion. We verify that the resulting Bonferroni test is as powerful as the infeasible version calculated using the true value of ρ_T after a refinement is made.

The rest of the article is organized as follows. In Section 2, we consider the model with the true value of ρ_T (or equivalently a) known and propose new tests. In Section 3, we consider the model with unknown ρ_T and discuss how to implement the proposed tests in this setting. Section 4 discusses the extension of the baseline model to allow for deterministic terms and serially dependent and/or heteroskedastic disturbances. Section 5 reports the finite-sample properties of the proposed tests. In Section 6, we apply our tests to real data. Section 7 concludes. In the Supplementary Material, all the mathematical proofs are given in Appendices B–D and Appendix E contains additional results related to Sections 2.5 and 5.

2. THE CASE OF KNOWN ρ_T

2.1. Model and Assumptions

Our model is given by

$$y_t = (\rho_T + \omega_T v_t) y_{t-1} + \varepsilon_t, \tag{2}$$

where $\rho_T = 1 + a/T$ ($a \in \mathbb{R}$) and $\omega_T = c/T^{3/4}$. Under $\omega_T = 0$, model (2) is the near unit root process and lies in the stationary, unit root, and explosive regions when $a < 0$, $a = 0$, and $a > 0$, respectively. When ρ and $\omega \neq 0$ are fixed, a sufficient

https://doi.org/10.1017/S0266466625000015 Published online by Cambridge University Press

condition for strict stationarity is $\mathbb{E}[\ln|\rho+\omega v_t|]<0$, and the condition $\rho^2+\omega^2>1$ is referred to as the explosiveness condition (Hwang and Basawa, 2005).[7] When $\omega\neq 0$, there is no counterpart to the unit root condition. In the literature, the RCA models are studied under either of the aforementioned conditions (Hwang and Basawa, 2005; Horváth and Trapani, 2019). Our model (2), however, cannot be nested in the usual RCA model because ρ_T and ω_T are dependent on T, which makes it impossible to use an approach relying on the (non)stationarity conditions of RCA processes. Consequently, our approach will be more of one used in the classical local unit root literature.

We begin our analysis assuming that the true value of ρ_T is known, which will be relaxed in Section 3, and imposing the following assumption.

Assumption 1. $(\varepsilon_t, v_t)' \sim$ i.i.d $(0,\Omega)$, where

$$\Omega := \begin{pmatrix} \sigma_\varepsilon^2 & \sigma_{\varepsilon v} \\ \sigma_{\varepsilon v} & 1 \end{pmatrix}.$$

Also, $\mathbb{E}[\varepsilon_t^4]<\infty$ and $\mathbb{E}[v_t^8]<\infty$. Moreover, $y_0 = o_p(T^{1/2})$.

The finite 4th moment assumption for disturbance ε_t is standard in the RCA literature (Nagakura, 2009; Horváth and Trapani, 2019). We believe that the assumption of a finite 8th moment for v_t is not restrictive because the AR root, $\rho_T+\omega_T v_t$, is expected to fluctuate within a close neighborhood of unity. This reasoning makes it natural to restrict the tail probabilities of v_t. In Assumption 1, we also impose the i.i.d. condition on disturbance ε_t. This will be relaxed in Section 4 to allow for serial correlation and conditional heteroskedasticity.

Define $\eta_t := \varepsilon_t^2 - \sigma_\varepsilon^2$ and $\sigma_\eta^2 := \mathbb{E}[\eta_t^2]$. Define also the partial sum process $(W_{\varepsilon,T}, W_{\eta,T})'$ on $[0,1]$ by $W_{\varepsilon,T}(r) := T^{-1/2}\sigma_\varepsilon^{-1}\sum_{t=1}^{\lfloor Tr\rfloor}\varepsilon_t$ and $W_{\eta,T}(r) := T^{-1/2}\sigma_\eta^{-1}\sum_{t=1}^{\lfloor Tr\rfloor}\eta_t$. Then, it follows from the functional central limit theorem (FCLT) that under Assumption 1, as $T\to\infty$, $(W_{\varepsilon,T}, W_{\eta,T})' \Rightarrow (W_\varepsilon, W_\eta)'$ in the Skorokhod space $D_2[0,1]$, where $(W_\varepsilon, W_\eta)'$ is a vector Brownian motion with the covariance coefficient

$$\mathbb{V}\begin{pmatrix} W_\varepsilon \\ W_\eta \end{pmatrix} = \begin{pmatrix} 1 & \psi \\ \psi & 1 \end{pmatrix}$$

with $\psi := \mathrm{Corr}(\varepsilon_t,\eta_t) = \mathbb{E}[\varepsilon_t\eta_t]/(\sigma_\varepsilon\sigma_\eta)$. Note that W_ε and W_η are not necessarily independent because of the correlation ψ, and the Brownian motion W_η satisfies the following equality in distribution:

$$W_\eta \overset{d}{=} \psi W_\varepsilon + \sqrt{1-\psi^2}W_1, \tag{3}$$

where W_1 is a standard Brownian motion independent of W_ε. Lemma B.3 in Appendix B of the Supplementary Material shows that under Assumption 1, the standardized process $T^{-1/2}y_{\lfloor T\cdot\rfloor}$ on $[0,1]$ constructed from (2) weakly

[7] Note that $\rho^2+\omega^2<1$ also suffices for stationarity (Nicholls and Quinn, 1982). When $\rho^2+\omega^2<1$, the RCA model is a strictly stationary process with finite variance.

https://doi.org/10.1017/S0266466625000015 Published online by Cambridge University Press

converges to the Ornstein–Uhlenbeck (OU) process $\sigma_\varepsilon J_a(\cdot)$, where J_a solves $dJ_a(r) = aJ_a(r)dr + dW_\varepsilon(r)$. We can also construct consistent estimators of variances, namely, σ_ε^2 and σ_η^2. Define $z_t(\rho_T) \coloneqq y_t - \rho_T y_{t-1} (= \omega_T v_t y_{t-1} + \varepsilon_t)$. Then, the estimators $\hat{\sigma}^2_{\varepsilon,T}(\rho_T) \coloneqq T^{-1}\sum_{t=1}^T z_t^2(\rho_T)$ and $\hat{\sigma}^2_{\eta,T}(\rho_T) \coloneqq T^{-1}\sum_{t=1}^T \{z_t^2(\rho_T) - \hat{\sigma}^2_{\varepsilon,T}(\rho_T)\}^2$ are consistent for σ_ε^2 and σ_η^2, respectively as proven in Appendix B of the Supplementary Material.

In subsequent sections, we will analyze the effect of $\mathrm{Corr}(\varepsilon_t, v_t) = \sigma_{\varepsilon v}/\sigma_\varepsilon$ on the behavior of tests for $\mathrm{H}_0 : \omega_T^2 = 0$. Given the localized model (2), we also localize the correlation between ε_t and v_t to derive relevant local asymptotic distributions:

Assumption 2. $\mathrm{Corr}(\varepsilon_t, v_t) = \sigma_{\varepsilon v}/\sigma_\varepsilon = q/T^{1/4}$.

The localizing rate $T^{-1/4}$ is necessary because a stochastic component of test statistics that is a multiple of $\omega \times \mathrm{Corr}(\varepsilon_t, v_t)$ is $O_p(T)$ if ω and $\mathrm{Corr}(\varepsilon_t, v_t)$ are constant; since we set $\omega = \omega_T = cT^{-3/4}$, $\mathrm{Corr}(\varepsilon_t, v_t) = qT^{-1/4}$ is necessary to prevent test statistics from diverging. Here, q is interpreted as the correlation coefficient in the limit as $T \to \infty$.

2.2. Effect of $\mathrm{Corr}(\varepsilon_t, v_t)$ on the LN Test

Several tests of $\mathrm{H}_0 : \omega_T^2 = 0$ for model (2) have been proposed by earlier work such as LN. The LN test statistic is defined by

$$\mathrm{LN}_T(\rho_T) \coloneqq \frac{\sum_{t=1}^T (y_{t-1}^2 - T^{-1}\sum_{t=1}^T y_{t-1}^2) z_t^2(\rho_T)}{\hat{\sigma}_{\eta,T}(\rho_T)\{\sum_{t=1}^T (y_{t-1}^2 - T^{-1}\sum_{t=1}^T y_{t-1}^2)^2\}^{1/2}}. \tag{4}$$

Nishi and Kurozumi (2024) found that the LN test has a high power function under the assumption that $\mathrm{Corr}(\varepsilon_t, v_t) = 0$ and $\rho_T = 1$.

As pointed out by Nishi and Kurozumi (2024), the LN test, which was originally derived as a locally best invariant test, can also be obtained as the t-test for $\mathrm{H}_0 : \omega_T^2 = 0$ under $\mathbb{E}[\varepsilon_t v_t] = \sigma_{\varepsilon v} = 0$. To see this, note that from model (2), a simple calculation gives

$$z_t^2(\rho_T) = \sigma_\varepsilon^2 + \omega_T^2 y_{t-1}^2 + \xi_t, \tag{5}$$

where $\xi_t \coloneqq \omega_T^2 y_{t-1}^2(v_t^2 - 1) + 2\omega_T y_{t-1}\varepsilon_t v_t + (\varepsilon_t^2 - \sigma_\varepsilon^2)$. Because $\mathbb{E}[\xi_t] = 0$ and $\mathbb{E}[y_{t-1}^2 \xi_t] = 0$ under Assumption 1 with $\sigma_{\varepsilon v} = 0$, model (5) may be viewed as the linear regression model with ξ_t playing the role of the disturbance, and the LN test statistic is obtained as the t-test for $\mathrm{H}_0 : \omega_T^2 = 0$ (with the variance estimator under the null used).

The above observation, however, does not hold when $\sigma_{\varepsilon v} \neq 0$, because it results in $\mathbb{E}[y_{t-1}^2 \xi_t] = 2\omega_T \sigma_{\varepsilon v} \mathbb{E}[y_{t-1}^3] \neq 0$, that is, endogeneity. As we will shortly show, the power of the LN test is crucially affected by the value of $\sigma_{\varepsilon v}$, or equivalently, $\mathrm{Corr}(\varepsilon_t, v_t) = \sigma_{\varepsilon v}/\sigma_\varepsilon$ (for a given value of σ_ε). If $\mathrm{Corr}(\varepsilon_t, v_t)$ may not be zero, the LN test statistic has the following asymptotic distribution.

https://doi.org/10.1017/S0266466625000015 Published online by Cambridge University Press

THEOREM 1. *Consider model (2). Under Assumptions 1 and 2, we have*

$$\mathrm{LN}_T(\rho_T) \Rightarrow \frac{\int_0^1 \tilde{J}_{a,2}(r)dW_\eta(r)}{\left\{\int_0^1 \left(\tilde{J}_{a,2}\right)^2(r)dr\right\}^{1/2}} + \frac{\sigma_\varepsilon^2}{\sigma_\eta}\left[\frac{c^2\int_0^1 \left(\tilde{J}_{a,2}\right)^2(r)dr + 2cq\int_0^1 \tilde{J}_{a,2}(r)\tilde{J}_{a,1}(r)dr}{\left\{\int_0^1 \left(\tilde{J}_{a,2}\right)^2(r)dr\right\}^{1/2}}\right], \tag{6}$$

where $\tilde{J}_{a,1}(r) \coloneqq J_a(r) - \int_0^1 J_a(s)ds$ *and* $\tilde{J}_{a,2}(r) \coloneqq J_a^2(r) - \int_0^1 J_a^2(s)ds$ *with* J_a *satisfying* $dJ_a(r) = aJ_a(r)dr + dW_\varepsilon(r)$. *a and c are defined in (2), and q is defined in Assumption 2.*

Note that when $q = 0$ and $\rho_T = 1$ (or $a = 0$), the limit distribution reduces to the one derived by Nishi and Kurozumi (2024), who found that the LN test performs well in such a case. To see how the value of q alters the power properties of the LN test, we simulate the asymptotic distribution in (6) by 100,000 replications. The replications are based on $\varepsilon_t \sim$ i.i.d $N(0,1)$, so that $\psi = \mathrm{Corr}(\varepsilon_t, \eta_t) = 0$. Note that the effect of q on the limit distribution is symmetric when $\psi = 0$; that is, $q = \bar{q}$ and $q = -\bar{q}$ produce the identical distribution. This is because $(W_\varepsilon, W_\eta)' \overset{d}{=} (-W_\varepsilon, W_\eta)'$ when $\psi = 0$, and hence $(J_a, W_\eta)' \overset{d}{=} (-J_a, W_\eta)'$ in this case. Thus, in this simulation, and in the replications conducted later where $\psi = 0$ holds, we only consider positive values of q. Figure 1 shows the power functions of the LN test for $q = 0, 1, 2, 3$ under $a = 0$. It is easily observed that as q gets larger, the power function gets lower and flatter. From this observation, we conclude that the LN test performs poorly when q is large (in absolute value).[8]

2.3. Augmented Regression and New Tests

The poor performance by the LN test for large q could be attributed to the endogeneity $\mathbb{E}[y_{t-1}^2\xi_t] = 2c\sigma_\varepsilon qT^{-1}\mathbb{E}[y_{t-1}^3] \neq 0$ in (5). One solution for this endogeneity is to augment model (5) by adding y_{t-1} as a regressor,

$$z_t^2(\rho_T) = \sigma_\varepsilon^2 + \delta_T y_{t-1} + \omega_T^2 y_{t-1}^2 + \xi_t^*, \tag{7}$$

where $\delta_T \coloneqq 2c\sigma_\varepsilon qT^{-1}$ and $\xi_t^* \coloneqq \omega_T^2 y_{t-1}^2(v_t^2 - 1) + 2\omega_T y_{t-1}(\varepsilon_t v_t - \sigma_{\varepsilon v}) + (\varepsilon_t^2 - \sigma_\varepsilon^2)$. Because $\mathbb{E}[\xi_t^*] = \mathbb{E}[y_{t-1}\xi_t^*] = \mathbb{E}[y_{t-1}^2\xi_t^*] = 0$ under Assumption 1, we propose using the t-test for $\mathrm{H}_0 : \omega_T^2 = 0$ in model (7). We also propose using the Wald test for $(\delta_T, \omega_T^2) = (0,0)$ because $\omega_T^2 = 0$ if and only if $(\delta_T, \omega_T^2) = (0,0)$. To express the t- and Wald test statistics, first regress $z_t^2(\rho_T)$, y_{t-1} and y_{t-1}^2 on a constant:

$$\widetilde{z}_t^2(\rho_T) = \delta_T x_{1,t-1} + \omega_T^2 x_{2,t-1} + \widetilde{\xi}_t^*, \tag{8}$$

[8] In our unreported simulations, we found a similar tendency in the power function of Distaso's (2008) LM test, which is based on the assumption of ε_t and v_t being independent. The results are omitted to save space.

https://doi.org/10.1017/S0266466625000015 Published online by Cambridge University Press

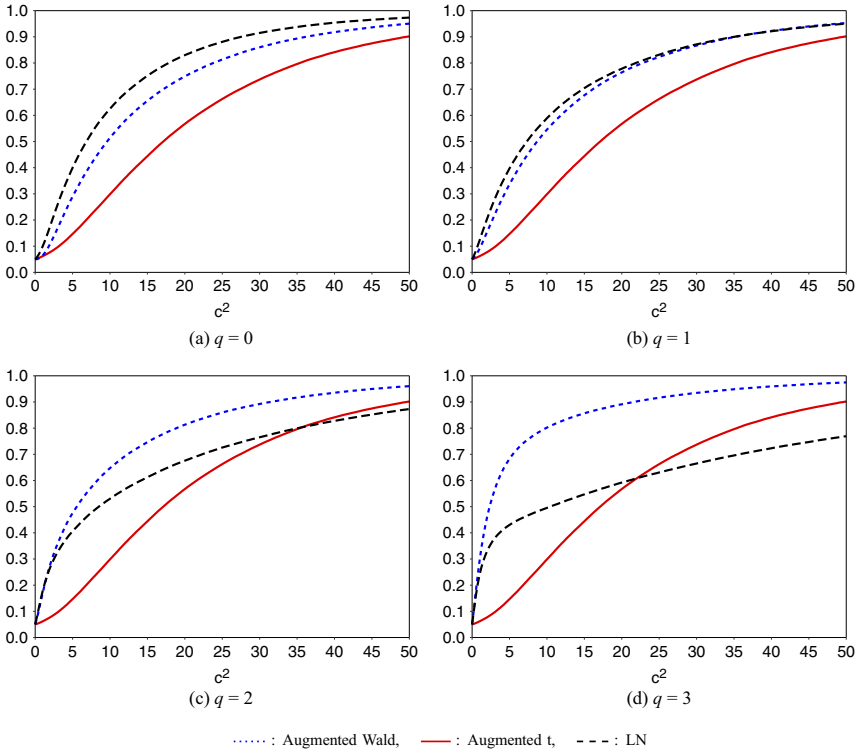


FIGURE 1. Asymptotic power for $a = 0$.

where $\widetilde{z_t^2}(\rho_T) := z_t^2(\rho_T) - T^{-1}\sum_{t=1}^{T} z_t^2(\rho_T)$, $x_{1,t-1} := y_{t-1} - T^{-1}\sum_{t=1}^{T} y_{t-1}$, $x_{2,t-1} := y_{t-1}^2 - T^{-1}\sum_{t=1}^{T} y_{t-1}^2$ and $\widetilde{\xi_t^*} := \xi_t^* - T^{-1}\sum_{t=1}^{T}\xi_t^*$. Let $\tilde{Z}_2(\rho_T) := (\widetilde{z_1^2}(\rho_T), \widetilde{z_2^2}(\rho_T), \ldots, \widetilde{z_T^2}(\rho_T))'$, $X_1 := (x_{1,0}, x_{1,1}, \ldots, x_{1,T-1})'$, $X_2 := (x_{2,0}, x_{2,1}, \ldots, x_{2,T-1})'$, and $\widetilde{\Xi}^* := (\widetilde{\xi_1^*}, \widetilde{\xi_2^*}, \ldots, \widetilde{\xi_T^*})'$. Then, model (8) can be expressed in matrix notation as

$$\tilde{Z}_2(\rho_T) = X\theta_T + \widetilde{\Xi}^*, \tag{9}$$

where $X := (X_1, X_2)$ and $\theta_T := (\delta_T, \omega_T^2)'$. Then, the t- and Wald test statistics are given by

$$t_{\hat{\omega}_T^2}(\rho_T) := \hat{\sigma}_{\widetilde{\xi^*}}^{-1}(\rho_T)(X_2'M_1X_2)^{-1/2}X_2'M_1\tilde{Z}_2(\rho_T) \tag{10}$$

and

$$W_T(\rho_T) := \frac{\hat{\theta}_T(\rho_T)'(X'X)\hat{\theta}_T(\rho_T)}{\hat{\sigma}_{\widetilde{\xi^*}}^2(\rho_T)}, \tag{11}$$

https://doi.org/10.1017/S0266466625000015 Published online by Cambridge University Press

where $M_1 \coloneqq I_T - X_1(X_1'X_1)^{-1}X_1'$, and $\hat{\sigma}^2_{\tilde{\xi}^*}(\rho_T)$ and $\hat{\theta}_T(\rho_T)$ are the OLS variance and coefficient estimators of (9), respectively. We shall call the t- and Wald tests in model (7) augmented t- and augmented Wald tests, respectively.

THEOREM 2. *Consider model (2). Under Assumptions 1 and 2, we have*

$$t_{\hat{\omega}^2_T}(\rho_T) \Rightarrow \frac{\int_0^1 Q_a(r)dW_\eta(r)}{\left[\int_0^1 Q_a^2(r)dr\right]^{1/2}} + \frac{c^2\sigma_\varepsilon^2}{\sigma_\eta}\left[\int_0^1 Q_a^2(r)dr\right]^{1/2},$$

where

$$Q_a(r) \coloneqq \tilde{J}_{a,2}(r) - \frac{\int_0^1 \tilde{J}_{a,1}(s)\tilde{J}_{a,2}(s)ds}{\int_0^1 \left(\tilde{J}_{a,1}\right)^2(s)ds}\tilde{J}_{a,1}(r),$$

and

$$\begin{aligned} &W_T(\rho_T) \\ &\Rightarrow \left\{ \begin{pmatrix} \int_0^1 \tilde{J}_{a,1}(r)dW_\eta(r) \\ \int_0^1 \tilde{J}_{a,2}(r)dW_\eta(r) \end{pmatrix} + \frac{\sigma_\varepsilon^2}{\sigma_\eta} \begin{pmatrix} c^2\int_0^1 \tilde{J}_{a,1}(r)\tilde{J}_{a,2}(r)dr + 2cq\int_0^1 \left(\tilde{J}_{a,1}\right)^2(r)dr \\ c^2\int_0^1 \left(\tilde{J}_{a,2}\right)^2(r)dr + 2cq\int_0^1 \tilde{J}_{a,1}(r)\tilde{J}_{a,2}(r)dr \end{pmatrix} \right\}' \\ &\quad \times \begin{pmatrix} \int_0^1 \left(\tilde{J}_{a,1}\right)^2(r)dr & \int_0^1 \tilde{J}_{a,1}(r)\tilde{J}_{a,2}(r)dr \\ \int_0^1 \tilde{J}_{a,2}(r)\tilde{J}_{a,1}(r)dr & \int_0^1 \left(\tilde{J}_{a,2}\right)^2(r)dr \end{pmatrix}^{-1} \\ &\quad \times \left\{ \begin{pmatrix} \int_0^1 \tilde{J}_{a,1}(r)dW_\eta(r) \\ \int_0^1 \tilde{J}_{a,2}(r)dW_\eta(r) \end{pmatrix} + \frac{\sigma_\varepsilon^2}{\sigma_\eta} \begin{pmatrix} c^2\int_0^1 \tilde{J}_{a,1}(r)\tilde{J}_{a,2}(r)dr + 2cq\int_0^1 \left(\tilde{J}_{a,1}\right)^2(r)dr \\ c^2\int_0^1 \left(\tilde{J}_{a,2}\right)^2(r)dr + 2cq\int_0^1 \tilde{J}_{a,1}(r)\tilde{J}_{a,2}(r)dr \end{pmatrix} \right\}. \end{aligned}$$

In particular, the asymptotic null distributions ($c^2 = 0$) are

$$t_{\hat{\omega}^2_T}(\rho_T) \Rightarrow \frac{\int_0^1 Q_a(r)dW_\eta(r)}{\left[\int_0^1 Q_a^2(r)dr\right]^{1/2}}$$

and

$$\begin{aligned} W_T(\rho_T) \Rightarrow &\begin{pmatrix} \int_0^1 \tilde{J}_{a,1}(r)dW_\eta(r) \\ \int_0^1 \tilde{J}_{a,2}(r)dW_\eta(r) \end{pmatrix}' \begin{pmatrix} \int_0^1 \left(\tilde{J}_{a,1}\right)^2(r)dr & \int_0^1 \tilde{J}_{a,1}(r)\tilde{J}_{a,2}(r)dr \\ \int_0^1 \tilde{J}_{a,2}(r)\tilde{J}_{a,1}(r)dr & \int_0^1 \left(\tilde{J}_{a,2}\right)^2(r)dr \end{pmatrix}^{-1} \\ &\times \begin{pmatrix} \int_0^1 \tilde{J}_{a,1}(r)dW_\eta(r) \\ \int_0^1 \tilde{J}_{a,2}(r)dW_\eta(r) \end{pmatrix}, \end{aligned}$$

which are standard normal and chi-square with 2 degrees of freedom, respectively, when $\psi = \mathrm{Corr}(\varepsilon_t, \eta_t) = 0$.

There are two points worth mentioning. First, $t_{\hat{\omega}^2_T}(\rho_T)$ has the asymptotic (alternative) distribution independent of q. This is a direct result of the augmentation of model (5). When performing the augmented t-test, we regress the endogenous regressor $x_{2,t-1}$ (and $\widetilde{z_t^2}(\rho_T)$) on $x_{1,t-1}$. In the limit, this projection amounts to

https://doi.org/10.1017/S0266466625000015 Published online by Cambridge University Press

replacing $\tilde{J}_{a,2}$ in the limiting distribution of LN_T with Q_a, the residual from the linear projection of $\tilde{J}_{a,2}$ on $\tilde{J}_{a,1}$ in the Hilbert space.

The second point is that the augmented t- and Wald test statistics have pivotal, standard limiting null distributions when $\psi = \mathrm{Corr}(\varepsilon_t, \eta_t) = 0$. This will be seen immediately upon noting J_a and W_η are independent when $\psi = 0$ (due to independence between W_ε and W_η), and the limit distributions under $c^2 = 0$ conditional on W_ε are standard normal and chi-square with 2 degrees of freedom (and so are they unconditionally). Along the lines of this argument, the asymptotic null distribution of the LN test statistic is seen to be standard normal when $\psi = 0$.

2.4. Removing ψ from the Limiting Null Distributions

Unless $\psi = 0$, the test statistics discussed thus far have asymptotic null distributions dependent on ψ. This dependence stems from the strong persistence of the regressors y_{t-1} and y_{t-1}^2 and the long run endogeneity present in the linearized models (5) and (7). To illustrate this, consider model (5) and the LN test. Under the null of $\omega_T^2 = 0$, the model reduces to

$$z_t^2(\rho_T) = \sigma_\varepsilon^2 + \omega_T^2 y_{t-1}^2 + \eta_t, \tag{12}$$

where $\eta_t = \varepsilon_t^2 - \sigma_\varepsilon^2$ and $y_t = \rho_T y_{t-1} + \varepsilon_t$. Although the regressor, y_{t-1}^2, satisfies $\mathbb{E}[y_{t-1}^2 \eta_t] = 0$, it is"endogenous" in the limit as $T \to \infty$, because of the correlation ψ between its innovation ε_t and the disturbance η_t. Indeed, the LN test statistic becomes

$$\mathrm{LN}_T(\rho_T) = \frac{\sum_{t=1}^T x_{2,t-1}\eta_t}{\hat{\sigma}_{\eta,T}(\rho_T)\{\sum_{t=1}^T (x_{2,t-1})^2\}^{1/2}} \Rightarrow \frac{\int_0^1 \tilde{J}_{a,2}(r) dW_\eta(r)}{\left\{\int_0^1 \left(\tilde{J}_{a,2}\right)^2(r) dr\right\}^{1/2}},$$

where $\tilde{J}_{a,2}(r)$ ($\approx T^{-1} x_{2,\lfloor Tr \rfloor}$) is correlated with the differential $dW_\eta(r)$ ($\approx T^{-1/2}\eta_{\lfloor Tr \rfloor}$). This correlation originates from that between W_ε and W_η, which is denoted by ψ.

Interestingly, the situation we are in is analogous to that considered in the literature on predictive regressions. In predictive regressions, stock returns (r_t) are regressed on a constant and the lag of another economic or financial variable (p_t):

$$r_t = \mu + \beta p_{t-1} + u_{r,t}.$$

In the literature on predictive regressions, the usual t-test for the hypothesis $\beta = 0$ is known to be problematic when regressor p_t has a generating mechanism of the form $p_t = \rho_{p,T} p_{t-1} + u_{p,t}$ with $\rho_{p,T} = 1 + a_p/T$, and its innovation $u_{p,t}$ is correlated with regression disturbance $u_{r,t}$. The problem arising in such a case is that the asymptotic null distribution of the t-statistic depends on $\mathrm{Corr}(u_{p,t}, u_{r,t})$ and is not standard normal unless $\mathrm{Corr}(u_{p,t}, u_{r,t}) = 0$ (see, for example, Campbell and Yogo, 2006). Our situation here is essentially the same, where y_{t-1}^2, ε_t, and η_t in (12) play the same roles as p_{t-1}, $u_{p,t}$, and $u_{r,t}$, respectively.

https://doi.org/10.1017/S0266466625000015 Published online by Cambridge University Press

For the predictive regression model, there is an extensive literature on this problem, and numerous solutions have been proposed (Campbell and Yogo, 2006; Phillips and Lee, 2013; Phillips, 2014; Kostakis, Magdalinos, and Stamatogiannis, 2015). We follow Campbell and Yogo (2006) to modify the test statistics (LN, augmented t- and augmented Wald) so that their asymptotic null distributions are standard and pivotal. To explain the idea, take the LN test as an example. Note that under the null of $c^2 = 0$, $z_t^2(\rho_T)$ in the numerator of the LN test statistic may be asymptotically expressed as

$$z_{\lfloor Tr \rfloor}^2(\rho_T) = \sigma_\varepsilon^2 + (\varepsilon_{\lfloor Tr \rfloor}^2 - \sigma_\varepsilon^2) \approx \sigma_\varepsilon^2 + T^{1/2} \times \sigma_\eta dW_\eta(r)$$
$$\stackrel{d}{=} \sigma_\varepsilon^2 + T^{1/2} \times \sigma_\eta \psi dW_\varepsilon(r) + T^{1/2} \times \sigma_\eta \sqrt{1-\psi^2} dW_1(r),$$

given the distributional equivalence (3). This observation leads us to propose the following modification to remove the effect of ψ:

$$z_t^2(\rho_T) \to z_t^{2*}(\rho_T) := \frac{z_t^2(\rho_T) - \hat{\sigma}_{\eta,T}(\rho_T)\hat{\psi}_T(\rho_T)(z_t(\rho_T)/\hat{\sigma}_{\varepsilon,T}(\rho_T))}{\sqrt{1-\hat{\psi}_T^2(\rho_T)}}, \tag{13}$$

where $\hat{\psi}_T(\rho_T) := T^{-1}\sum_{t=1}^T z_t(\rho_T)\{z_t^2(\rho_T) - \hat{\sigma}_{\varepsilon,T}^2(\rho_T)\}/(\hat{\sigma}_{\varepsilon,T}(\rho_T)\hat{\sigma}_{\eta,T}(\rho_T))$ is a consistent estimator of ψ as proven in Appendix B of the Supplementary Material.[9] With the replacement given in (13), the modified LN test statistic, denoted by $\mathrm{LN}_T^*(\rho_T)$, is defined by (4) except that $z_t^2(\rho_T)$ is replaced with $z_t^{2*}(\rho_T)$. The modified augmented t- and Wald tests are based on the following regression model:

$$\widetilde{Z}_2^*(\rho_T) = X\theta_T + \widetilde{\Xi^{**}}, \tag{14}$$

where $\widetilde{Z}_2^*(\rho_T) := (\widetilde{z_1^{2*}}(\rho_T), \widetilde{z_2^{2*}}(\rho_T), \ldots, \widetilde{z_T^{2*}}(\rho_T))'$ with $\widetilde{z_t^{2*}}(\rho_T) := z_t^{2*}(\rho_T) - T^{-1}\sum_{t=1}^T z_t^{2*}(\rho_T)$. The modified augmented t- and Wald statistics, denoted by $t_{\hat{\omega}_T^2}^*(\rho_T)$ and $W_T^*(\rho_T)$, respectively, are (10) and (11) calculated using $\hat{\sigma}_{\widetilde{\xi^{**}}}^2(\rho_T)$, $\widetilde{Z}_2^*(\rho_T)$, and $\hat{\theta}_T^*(\rho_T)$ in place of $\hat{\sigma}_{\xi^*}^2(\rho_T)$, $\tilde{Z}_2(\rho_T)$, and $\hat{\theta}_T(\rho_T)$, where $\hat{\sigma}_{\widetilde{\xi^{**}}}^2(\rho_T)$ and $\hat{\theta}_T^*(\rho_T)$ are the OLS variance and coefficient estimators of (14), respectively.

THEOREM 3. *Consider model (2). Under Assumptions 1 and 2, we have*

$$\mathrm{LN}_T^*(\rho_T) \Rightarrow \frac{\int_0^1 \tilde{J}_{a,2}(r)dW_1(r)}{\left\{\int_0^1 \left(\tilde{J}_{a,2}\right)^2(r)dr\right\}^{1/2}}$$
$$+ \frac{\sigma_\varepsilon^2}{\sigma_\eta\sqrt{1-\psi^2}}\left[\frac{c^2\int_0^1 \left(\tilde{J}_{a,2}\right)^2(r)dr + 2cq\int_0^1 \tilde{J}_{a,2}(r)\tilde{J}_{a,1}(r)dr}{\left\{\int_0^1 \left(\tilde{J}_{a,2}\right)^2(r)dr\right\}^{1/2}}\right],$$
$$t_{\hat{\omega}_T^2}^*(\rho_T) \Rightarrow \frac{\int_0^1 Q_a(r)dW_1(r)}{\left[\int_0^1 Q_a^2(r)dr\right]^{1/2}} + \frac{c^2\sigma_\varepsilon^2}{\sigma_\eta\sqrt{1-\psi^2}}\left[\int_0^1 Q_a^2(r)dr\right]^{1/2},$$

[9] In fact, Campbell and Yogo (2006) proposed the modification based on the optimality argument.

https://doi.org/10.1017/S0266466625000015 Published online by Cambridge University Press

and

$$
\begin{aligned}
W_T^*(\rho_T) \Rightarrow & \left\{ \begin{pmatrix} \int_0^1 \tilde{J}_{a,1}(r)dW_1(r) \\ \int_0^1 \tilde{J}_{a,2}(r)dW_1(r) \end{pmatrix} \right. \\
& \left. + \frac{\sigma_\varepsilon^2}{\sigma_\eta \sqrt{1-\psi^2}} \begin{pmatrix} c^2 \int_0^1 \tilde{J}_{a,1}(r)\tilde{J}_{a,2}(r)dr + 2cq \int_0^1 \left(\tilde{J}_{a,1}\right)^2(r)dr \\ c^2 \int_0^1 \left(\tilde{J}_{a,2}\right)^2(r)dr + 2cq \int_0^1 \tilde{J}_{a,1}(r)\tilde{J}_{a,2}(r)dr \end{pmatrix} \right\}' \\
& \times \begin{pmatrix} \int_0^1 \left(\tilde{J}_{a,1}\right)^2(r)dr & \int_0^1 \tilde{J}_{a,1}(r)\tilde{J}_{a,2}(r)dr \\ \int_0^1 \tilde{J}_{a,2}(r)\tilde{J}_{a,1}(r)dr & \int_0^1 \left(\tilde{J}_{a,2}\right)^2(r)dr \end{pmatrix}^{-1} \\
& \times \left\{ \begin{pmatrix} \int_0^1 \tilde{J}_{a,1}(r)dW_1(r) \\ \int_0^1 \tilde{J}_{a,2}(r)dW_1(r) \end{pmatrix} \right. \\
& \left. + \frac{\sigma_\varepsilon^2}{\sigma_\eta \sqrt{1-\psi^2}} \begin{pmatrix} c^2 \int_0^1 \tilde{J}_{a,1}(r)\tilde{J}_{a,2}(r)dr + 2cq \int_0^1 \left(\tilde{J}_{a,1}\right)^2(r)dr \\ c^2 \int_0^1 \left(\tilde{J}_{a,2}\right)^2(r)dr + 2cq \int_0^1 \tilde{J}_{a,1}(r)\tilde{J}_{a,2}(r)dr \end{pmatrix} \right\},
\end{aligned}
$$

where W_1 is a standard Brownian motion independent of W_ε.

It should be noticed that the modified test statistics have pivotal asymptotic null distributions (standard normal and chi-square with 2 degrees of freedom), thanks to independence between J_a and W_1. Also note that the limiting distributions are unaffected by our modification when $\psi = 0$ (cf. Theorems 1 and 2).

2.5. Comparison of Asymptotic Power Functions

Figure 1 compares the asymptotic power functions of the (modified) LN, augmented t-, and augmented Wald tests under $q = 0, 1, 2, 3$ and $a = 0$ with $\varepsilon_t \sim$ i.i.d $N(0, 1)$. When $q = 0, 1$, the LN test performs best, closely followed by the Wald test. The Wald test has better power properties than the augmented t-test. When $q = 2, 3$, the Wald test outperforms the LN test (and the augmented t-test) with greater dominance by the Wald test for a larger q. We also compute the asymptotic power functions under $a \in \{-5, -10\}$, which are relegated to Appendix E of the Supplementary Material to save space (Figures E.1 and E.2). The conclusion that the Wald test is better than the LN test except for small q continues to hold. This renders the Wald test preferable in empirical applications, where in general the degree of the correlation between the random coefficient and disturbance is unknown to practitioners.

Although all the above results are based on the assumption that the true ρ_T is known and therefore the tests discussed so far are infeasible, they suggest the potential of the augmentation to enhance the ability to detect the nonzero variance of the autoregressive root.

https://doi.org/10.1017/S0266466625000015 Published online by Cambridge University Press

3. THE CASE OF UNKNOWN ρ_T

In this section, we consider model (2) with the true value of ρ_T unknown. To deal with the uncertainty about ρ_T, we use the OLS estimator $\hat{\rho}_T$ of ρ_T: $\hat{\rho}_T = \sum_{t=1}^{T} y_{t-1}y_t / \sum_{t=1}^{T} y_{t-1}^2$. In Appendix C of the Supplementary Material, we show that $\hat{\rho}_T$ is T-consistent, and that other estimators defined in the preceding section such as $\hat{\sigma}^2_{\varepsilon,T}(\rho_T)$ remain consistent if they are computed using $\hat{\rho}_T$ instead of the true ρ_T. Despite these consistency results, the asymptotic null distributions of $\mathrm{LN}^*_T(\hat{\rho}_T)$, $t^*_{\hat{\omega}^2_T}(\hat{\rho}_T)$ and $W^*_T(\hat{\rho}_T)$ are no longer normal or chi-square. This is because $a = T(\rho_T - 1)$ is not consistently estimable, as is well known in the literature (Campbell and Yogo, 2006).

To deal with this problem, following Cavanagh, Elliott, and Stock (1995) and Campbell and Yogo (2006), we base our tests on the Bonferroni approach by using a confidence interval for ρ_T instead of its point estimate $\hat{\rho}_T$. The Bonferroni-based test consists of two steps: first, construct a confidence interval for ρ_T by inverting the t-statistic, and then repeat either of the tests considered above using all the hypothetical ρ_T values belonging to the confidence interval. Specifically, letting $S_T(\rho_T)$ denote either of the modified LN, augmented t, or augmented Wald test statistic (calculated using ρ_T), the testing procedure based on the Bonferroni approach is described as follows:

Algorithm 1.

Step 1. Given the data $\{y_t\}_{t=0}^{T}$, calculate for each hypothetical $\bar{\rho}_T = 1 + \bar{a}/T$ (on some grid) the t-statistic $t_{\bar{\rho}_T} := (\hat{\rho}_T - \bar{\rho}_T)(\hat{\sigma}_T^{-2}(\hat{\rho}_T)\sum_{t=1}^{T} y_{t-1}^2)^{1/2}$, to construct the (equal-tailed) $100(1-\alpha_1)\%$ confidence interval for ρ_T, denoted by $\mathrm{CI}(\alpha_1)$:

$$\mathrm{CI}(\alpha_1) = \{\bar{\rho}_T = 1 + \bar{a}/T : \mathrm{cv}(\bar{a})_{\alpha_1/2} \le t_{\bar{\rho}_T} \le \mathrm{cv}(\bar{a})_{1-\alpha_1/2}\},$$

where $\mathrm{cv}(\bar{a})_{\alpha_1/2}$ and $\mathrm{cv}(\bar{a})_{1-\alpha_1/2}$ denote the lower and upper $\alpha_1/2$ quantiles of the asymptotic distribution of $t_{\bar{\rho}_T}$ derived under the assumption that $\rho_T = \bar{\rho}_T$. $\mathrm{cv}(\bar{a})_{\alpha_1/2}$ and $\mathrm{cv}(\bar{a})_{1-\alpha_1/2}$ for selected values of $\bar{a}$ with $\alpha_1 = 0.05$ are shown in Table 1.[10]

Step 2. For each $\bar{\rho}_T \in \mathrm{CI}(\alpha_1)$, calculate $S_T(\bar{\rho}_T)$ and compare its value with the critical value cv_{α_2} with significance level α_2 (based on the standard normal or chi-square distributions). Reject the null only if all the calculated $S_T(\bar{\rho}_T)$ ($\bar{\rho}_T \in \mathrm{CI}(\alpha_1)$) exceed the critical value cv_{α_2}. The eventual significance level is $\alpha = \alpha_1 + \alpha_2$ by the Bonferroni inequality.

We shall call tests following these two steps Bonferroni tests.

[10]The quantiles are simulated via 30,000 Monte Carlo iterations where y_t is generated according to $y_t = (1 + \bar{a}/T)y_{t-1} + \varepsilon_t$, $t = 1, \ldots, T$ with $T = 5{,}000$ and $\varepsilon_t \sim$ i.i.d. $N(0,1)$.

https://doi.org/10.1017/S0266466625000015 Published online by Cambridge University Press

TABLE 1. Asymptotic 0.025 and 0.975 quantiles for $t_{\bar{\rho}_T}$

(a) No deterministics			(b) Intercept only			(c) Intercept and trend		
$\bar{a}$	0.025	0.975	$\bar{a}$	0.025	0.975	$\bar{a}$	0.025	0.975
−500	−1.99	1.91	−500	−2.05	1.86	−500	−2.12	1.79
−200	−2.03	1.91	−200	−2.12	1.81	−200	−2.22	1.71
−100	−2.04	1.90	−100	−2.18	1.77	−100	−2.32	1.63
−80	−2.04	1.90	−80	−2.19	1.75	−80	−2.35	1.59
−60	−2.04	1.89	−60	−2.20	1.71	−60	−2.40	1.53
−40	−2.05	1.86	−40	−2.25	1.63	−40	−2.49	1.41
−30	−2.06	1.84	−30	−2.30	1.57	−30	−2.56	1.33
−20	−2.08	1.82	−20	−2.38	1.46	−20	−2.69	1.15
−10	−2.11	1.76	−10	−2.53	1.23	−10	−2.94	0.80
−5	−2.15	1.67	−5	−2.69	0.91	−5	−3.17	0.35
−4	−2.15	1.64	−4	−2.74	0.80	−4	−3.25	0.22
−3	−2.16	1.61	−3	−2.80	0.66	−3	−3.34	0.05
−2	−2.18	1.59	−2	−2.87	0.48	−2	−3.43	−0.14
−1	−2.19	1.60	−1	−2.97	0.31	−1	−3.53	−0.35
0	−2.21	1.62	0	−3.12	0.21	0	−3.65	−0.66
1	−2.23	1.66	1	−3.24	0.24	1	−3.80	−1.03
2	−2.31	1.72	2	−3.15	0.43	2	−3.84	−0.33
3	−2.51	1.76	3	−2.99	0.74	3	−3.60	1.09
4	−2.62	1.82	4	−2.86	1.05	4	−3.35	1.73
5	−2.35	1.88	5	−2.60	1.29	5	−2.91	1.97
10	−1.96	1.97	10	−2.10	1.77	10	−1.69	2.17
15	−1.94	2.00	15	−2.03	1.87	15	−1.74	2.15

Remark 1.

1. The asymptotic distribution of $t_{\bar{\rho}_T}$ under $\rho_T = \bar{\rho}_T$ is

$$\frac{\int_0^1 J_{\bar{a}}(r) dW_\varepsilon(r)}{\{\int_0^1 J_{\bar{a}}^2(r) dr\}^{1/2}}. \tag{15}$$

2. Campbell and Yogo (2006) constructed the confidence interval for ρ_T by inverting the Dickey–Fuller type t-statistic, which is calculated by centering $\hat{\rho}_T$ by unity rather than $\bar{\rho}_T$. However, it has been well documented that the use of the Dickey–Fuller type t ratio leads to severe size distortion when the regressor is stationary (Phillips, 2014). Our unreported simulations showed that our Bonferroni tests also suffer from the same problem when they are based on the Dickey–Fuller t ratio. Following Phillips (2014), we here construct the confidence interval for ρ_T using the t-statistic centered by hypothetical $\bar{\rho}_T$'s.

https://doi.org/10.1017/S0266466625000015 Published online by Cambridge University Press

TABLE 2. Significance levels of the equal-tailed confidence interval for a

$\lvert\psi\rvert \in$	α_1[a]	$\lvert\psi\rvert \in$	α_1[a]
$[0, 0.05)$	0.09	$(0.55, 0.6]$	0.42
$[0.05, 0.1)$	0.17	$(0.6, 0.65]$	0.38
$[0.1, 0.15)$	0.23	$(0.65, 0.7]$	0.35
$[0.15, 0.2)$	0.31	$(0.7, 0.75]$	0.31
$[0.2, 0.25)$	0.38	$(0.75, 0.8]$	0.26
$[0.25, 0.3)$	0.45	$(0.8, 0.85]$	0.22
$[0.3, 0.4]$	0.5	$(0.85, 0.9]$	0.17
$(0.4, 0.45]$	0.48	$(0.9, 0.95]$	0.11
$(0.45, 0.5]$	0.46	$(0.95, 1)$[b]	0.05
$(0.5, 0.55]$	0.44		

[a.] Entries in the second and fourth columns are the significance level of the equal-tailed confidence interval for ρ_T when the significance levels of the Bonferroni–Wald test ($\tilde{\alpha}$) and individual modified Wald test (α_2) are 0.05.
[b.] To determine the α_1 value for the interval $(0.95, 1)$, we actually computed type 1 errors over $(0.95, 0.99]$.

Empirical sizes of our Bonferroni tests when y_t follows a stationary AR process are investigated via simulation in Section 5.

As pointed out by Cavanagh et al. (1995) and Campbell and Yogo (2006), the type 1 errors of the Bonferroni test defined above (dependent on (a, ψ)) can be quite smaller than $\alpha = \alpha_1 + \alpha_2$. To solve the problem of conservativeness, they proposed using a pair (α_1, α_2) with which the type 1 errors of the Bonferroni test are close to the desired nominal level $\tilde{\alpha}$ (say). To simplify the searching process for such a pair, we fix α_2 and set $\alpha_2 = \tilde{\alpha}$, following Campbell and Yogo (2006). In this study, we consider the case $\tilde{\alpha} = \alpha_2 = 0.05$, giving the Bonferroni test with significance level 0.05. Then, for each $\psi \in (-1, 1)$, we numerically find the value of $\alpha_1(\psi)$ such that under the null $P_{a,\psi}\left(\bigcap_{\tilde{\rho}_T \in \mathrm{CI}(\alpha_1(\psi))} \left\{S_T(\tilde{\rho}_T) > \mathrm{cv}_{0.05}\right\}\right) \leq \tilde{\alpha} = 0.05$, and this probability is as close to 0.05 as possible, for all a on some grid. The simulation process to find the $\alpha_1(\psi)$ values is described in Appendix A of the Supplementary Material.

Table 2 displays the significance level $\alpha_1(\psi)$ of the confidence interval for ρ_T along with corresponding intervals of $|\psi|$ values. Given the good asymptotic performance by the Wald test, we report in Table 2 α_1 values for the case of $S_T(\rho_T) = W_T^*(\rho_T)$. When performing the Bonferroni–Wald test, first estimate ψ by its consistent estimator $\hat{\psi}(\hat{\rho}_T)$, and then select the value of α_1 based on Table 2. For example, if $|\hat{\psi}(\hat{\rho}_T)| = 0.17$, then $\alpha_1 = 0.31$ is selected. The whole procedure of the Bonferroni–Wald test with this refinement is summarized as follows.

https://doi.org/10.1017/S0266466625000015 Published online by Cambridge University Press

Algorithm 2.

Step 1. Given the data $\{y_t\}_{t=0}^{T}$, calculate $\hat{\psi}(\hat{\rho}_T)$ and pick the significance level α_1 for the confidence interval for ρ_T from Table 2. Then, construct the $100(1-\alpha_1)\%$ confidence interval $\mathrm{CI}(\alpha_1)$ for ρ_T using this α_1 as in Step 1 of Algorithm 1. Critical values $\mathrm{cv}(\bar{a})_{\alpha_1/2}$ and $\mathrm{cv}(\bar{a})_{1-\alpha_1/2}$ have to be simulated using the given α_1.

Step 2. For each $\bar{\rho}_T \in \mathrm{CI}(\alpha_1)$, calculate $W_T^*(\bar{\rho}_T)$ and compare its value with the critical value $\mathrm{cv}_{0.05} = 5.99$ with significance level $\alpha_2 = 0.05$ (based on the chi-square distribution with 2 degrees of freedom). Reject the null only if all the calculated $W_T^*(\bar{\rho}_T)$ ($\bar{\rho}_T \in \mathrm{CI}(\alpha_1)$) exceed the critical value $\mathrm{cv}_{0.05}$. The eventual significance level is $\tilde{\alpha} = \alpha_2 = 0.05$.

Ideally, the local asymptotic power of the Bonferroni test should be derived and compared with that of the infeasible test calculated using the true value of ρ_T. This will enable us to evaluate the power losses originating from the use of the Bonferroni approach, but it seems a difficult task. Instead, we investigate the finite-sample power of the Bonferroni test and compare it with the power of its infeasible counterpart in Section 5.

4. EXTENSIONS

4.1. Model with Deterministic Terms

Our analysis thus far is based on the model without intercept or linear trend. When the data contains deterministic components, our model becomes

$$y_t = \mu' d_t + u_t,$$
$$u_t = (\rho_T + \omega_T v_t) u_{t-1} + \varepsilon_t.$$

In this case, the Bonferroni tests can be performed after data is demeaned or detrended. For instance, if the model contains an intercept ($d_t = 1$), the test statistics are constructed using demeaned data. Then, the asymptotic distributions remain the same as when $d_t = 0$ except that the OU process J_a is replaced with the demeaned OU process $J_a - \int_0^1 J_a$. Because the null distribution remains standard normal or chi-square, the testing procedures explained in Section 3 (Algorithms 1 and 2) can be applied. However, notice that demeaning or detrending affects the asymptotic distribution of $t_{\bar{\rho}_T}$, the t-statistic for the hypothesis $\rho_T = \bar{\rho}_T$, in which case the appropriate one is given by (15) with $J_{\bar{a}}(r)$ replaced by the demeaned or detrended $J_{\bar{a}}(r)$. This changes the referenced critical values for $t_{\bar{\rho}_T}$, and we show them for the demeaned and detrended cases in Tables 1b and 1c, respectively (with $\alpha_1 = 0.05$).

4.2. Weaker Assumptions on the Disturbance

4.2.1. *Modification of Test Statistics.* We have assumed that disturbance ε_t is i.i.d, but the test statistics considered above can be modified to accommodate serial

https://doi.org/10.1017/S0266466625000015 Published online by Cambridge University Press

correlation and conditional heteroskedasticity in the disturbance by applying the technique used by Nishi and Kurozumi (2024). Before presenting the modified test statistics, we have to note that serial correlation and conditional heteroskedasticity in ε_t have the same effect under $\mathrm{H}_0 : \omega_T = 0$. This is because the LN and Wald tests are based on linearized model (5) and (7), respectively, where ξ_t and ξ_t^* play the role of the disturbance. Because $\xi_t = \xi_t^* = \eta_t$ under the null, serial correlation and conditional heteroskedasticity in ε_t affects the behavior of the test statistics through $\eta_t = \varepsilon_t^2 - \sigma_\varepsilon^2$. If ε_t is serially dependent, η_t is also serially dependent, albeit with a different dependence structure from that of ε_t. However, if ε_t is conditionally heteroskedastic (e.g., GARCH), η_t becomes a serially correlated sequence. To facilitate the discussion, the following regularity condition is imposed on η_t as well as ε_t.

Assumption 3.

(a) ε_t is stationary and α-mixing with mixing coefficients $\alpha(m) = O(m^{-\gamma})$ for some $\gamma > 6$. Moreover, $\mathbb{E}[|\varepsilon_t|^{12}] < \infty$.

(b) The long-run variance–covariance matrix of $(\varepsilon_t, \eta_t)'$ is positive definite:

$$\begin{pmatrix} \sigma_{\varepsilon,lr}^2 & \sigma_{\varepsilon\eta,lr} \\ \sigma_{\varepsilon\eta,lr} & \sigma_{\eta,lr}^2 \end{pmatrix} > 0,$$

where $\sigma_{\varepsilon,lr}^2 := \mathbb{E}[\varepsilon_1^2] + 2\sum_{t=1}^{\infty} \mathbb{E}[\varepsilon_1 \varepsilon_{t+1}]$, $\sigma_{\eta,lr}^2 := \mathbb{E}[\eta_1^2] + 2\sum_{t=1}^{\infty} \mathbb{E}[\eta_1 \eta_{t+1}]$, and $\sigma_{\varepsilon\eta,lr} := \mathbb{E}[\varepsilon_1 \eta_1] + \sum_{t=1}^{\infty} \mathbb{E}[\varepsilon_1 \eta_{t+1}] + \sum_{t=1}^{\infty} \mathbb{E}[\eta_1 \varepsilon_{t+1}]$.

Assumption 3(a) allows for weak serial correlation in ε_t. Note that $\eta_t = \varepsilon_t^2 - \sigma_\varepsilon^2$ is also stationary and α-mixing with the same mixing coefficients as ε_t. Conditional heteroskedasticity is also permitted under Assumption 3(a), but some conditionally heteroskedastic processes may not satisfy this assumption. For example, if ε_t is a GARCH process, then η_t is an ARMA process but may not be α-mixing, because there are ARMA processes that are not mixing. In Assumption 3, we also assume that ε_t has a finite 12th moment. This condition is a technical one to guarantee that Theorem 3.1 of Liang et al. (2016) holds. This assumption might be weakened if a more sophisticated asymptotic theory is available.

Under Assumption 3, we have $\left(T^{-1/2}\sigma_{\varepsilon,lr}^{-1}\sum_{t=1}^{\lfloor T\cdot\rfloor}\varepsilon_t, T^{-1/2}\sigma_{\eta,lr}^{-1}\sum_{t=1}^{\lfloor T\cdot\rfloor}\eta_t\right)' \Rightarrow \left(W_{\varepsilon,lr}(\cdot), W_{\eta,lr}(\cdot)\right)'$, where $(W_{\varepsilon,lr}, W_{\eta,lr})'$ is a vector Brownian motion satisfying

$$\mathbb{V}\begin{pmatrix} W_{\varepsilon,lr}(r) \\ W_{\eta,lr}(r) \end{pmatrix} = \begin{pmatrix} 1 & \psi_{lr} \\ \psi_{lr} & 1 \end{pmatrix} r,$$

where $\psi_{lr} := \sigma_{\varepsilon\eta,lr}/(\sigma_{\varepsilon,lr}\sigma_{\eta,lr})$. $y_{\lfloor Tr\rfloor}/\sqrt{T}$ also weakly converges to $\sigma_{\varepsilon,lr}J_a(r)$ under the null of $\omega_T = 0$, where J_a solves $dJ_a(r) = aJ_a(r)dr + dW_{\varepsilon,lr}(r)$.[11] Furthermore, the one-sided long-run covariance between ε_t and η_t, $\Lambda_{\varepsilon\eta} := \sum_{t=1}^{\infty}\mathbb{E}[\varepsilon_1\eta_{t+1}]$, exists.

[11] With abuse of notation, we use J_a to denote the OU process driven by $W_{\varepsilon,lr}$.

https://doi.org/10.1017/S0266466625000015 Published online by Cambridge University Press

To save space, we will focus on the modification of the Wald test. The LN and augmented t-tests can be modified similarly. Under Assumption 3 and $H_0 : \omega_T = 0$, it can be shown that

$$W_T(\rho_T) \Rightarrow \sigma_\eta^{-2} \begin{pmatrix} \sigma_{\eta,lr} \int_0^1 \tilde{J}_{a,1}(r) dW_\eta(r) + \Lambda_{\varepsilon\eta}/\sigma_{\varepsilon,lr} \\ \sigma_{\eta,lr} \int_0^1 \tilde{J}_{a,2}(r) dW_\eta(r) + 2(\Lambda_{\varepsilon\eta}/\sigma_{\varepsilon,lr}) \int_0^1 J_a(r) dr \end{pmatrix}' \times \begin{pmatrix} \int_0^1 (\tilde{J}_{a,1})^2(r) dr & \int_0^1 \tilde{J}_{a,1}(r)\tilde{J}_{a,2}(r) dr \\ \int_0^1 \tilde{J}_{a,2}(r)\tilde{J}_{a,1}(r) dr & \int_0^1 (\tilde{J}_{a,2})^2(r) dr \end{pmatrix}^{-1} \times \begin{pmatrix} \sigma_{\eta,lr} \int_0^1 \tilde{J}_{a,1}(r) dW_\eta(r) + \Lambda_{\varepsilon\eta}/\sigma_{\varepsilon,lr} \\ \sigma_{\eta,lr} \int_0^1 \tilde{J}_{a,2}(r) dW_\eta(r) + 2(\Lambda_{\varepsilon\eta}/\sigma_{\varepsilon,lr}) \int_0^1 J_a(r) dr \end{pmatrix}; \quad \textbf{(16)}$$

see Appendix D of the Supplementay Material for the proof. Therefore, not only do we need to replace $\hat{\sigma}^2_{\tilde{\xi}^*}(\rho_T)$ with its long-run counterpart but also we need to eliminate $\Lambda_{\varepsilon\eta}$ from the limiting null distribution. To do this, define the following nonparametric estimators for long-run (co)variances:

$$\hat{\sigma}^2_{\tilde{\xi}^*,lr}(\rho_T) := T^{-1} \sum_{t=1}^{T} (\hat{\tilde{\xi}}^*_t)^2 + \sum_{i=1}^{l} w\Big(\frac{i}{l+1}\Big)\Big(2T^{-1} \sum_{t=i+1}^{T} \hat{\tilde{\xi}}^*_{t-i} \hat{\tilde{\xi}}^*_t\Big),$$

($\hat{\tilde{\xi}}^*_t$ is the residual from (8))

$$\hat{\sigma}^2_{\varepsilon,lr}(\rho_T) := T^{-1} \sum_{t=1}^{T} z_t^2(\rho_T) + \sum_{i=1}^{l} w\Big(\frac{i}{l+1}\Big)\Big(2T^{-1} \sum_{t=i+1}^{T} z_{t-i}(\rho_T) z_t(\rho_T)\Big),$$

$$\hat{\psi}_{lr}(\rho_T) := \frac{\hat{\sigma}_{\varepsilon\eta,lr}(\rho_T)}{\hat{\sigma}_{\varepsilon,lr}(\rho_T)\hat{\sigma}_{\tilde{\xi}^*,lr}(\rho_T)},$$

where

$$\hat{\sigma}_{\varepsilon\eta,lr}(\rho_T) := T^{-1} \sum_{t=1}^{T} z_t(\rho_T) \hat{\tilde{\xi}}^*_t + \sum_{i=1}^{l} w\Big(\frac{i}{l+1}\Big) \times \Big(T^{-1} \sum_{t=1+i}^{T} z_{t-i}(\rho_T) \hat{\tilde{\xi}}^*_t + T^{-1} \sum_{t=i+1}^{T} \hat{\tilde{\xi}}^*_{t-i} z_t(\rho_T)\Big),$$

and

$$\hat{\Lambda}_{\varepsilon\eta}(\rho_T) := \sum_{i=1}^{l} w\Big(\frac{i}{l+1}\Big)\Big(T^{-1} \sum_{t=i+1}^{T} z_{t-i}(\rho_T) \hat{\tilde{\xi}}^*_t\Big).$$

Here, $w(\cdot)$ is a kernel function and l is the truncation number. We use the Bartlett kernel $w(i/(l+1)) = 1 - i/(l+1)$ and set $l = \lfloor 4(T/100)^{1/4} \rfloor$ following Nishi and Kurozumi (2024). Under the null $H_0 : \omega_T = 0$, in which case y_t is a near unit root process, a standard argument will show that these estimators are consistent because $\hat{\rho}_T - \rho_T = O_p(T^{-1})$; see, e.g., Phillips (1987a). It will be tedious and much involved

https://doi.org/10.1017/S0266466625000015 Published online by Cambridge University Press

to establish the consistency of these estimators under the alternative hypothesis. However, we expect that they are consistent under the sequence of local alternatives given by $\omega_T = cT^{-3/4}$ because the $T^{-3/4}$-rate localization is sufficiently fast that the presence of random coefficient does not affect the consistency of $\hat{\rho}_T$ and short-run variance estimators (see Lemmas B.4, B.5, and C.1).

From the expression given in (16) and the need to eliminate ψ_{lr}, the long-run correlation between ε_t and η_t, we modify the Wald statistic in the following way (cf. (11)):

$$W_T^{lr}(\rho_T) := \frac{\left(\hat{\theta}_T(\rho_T) - (X'X)^{-1}U_T(\rho_T)\right)'(X'X)\left(\hat{\theta}_T(\rho_T) - (X'X)^{-1}U_T(\rho_T)\right)}{\hat{\sigma}^2_{\tilde{\xi}^*,lr}(\rho_T)(1-\hat{\psi}^2_{lr,T}(\rho_T))},$$

where

$$U_T(\rho_T) := \begin{pmatrix} T\hat{\Lambda}_{\varepsilon\eta}(\rho_T) + \hat{\psi}_{lr,T}(\rho_T)\hat{\sigma}_{\tilde{\xi}^*,lr}(\rho_T)\hat{\sigma}^{-1}_{\varepsilon,lr}(\rho_T)\sum_{t=1}^T x_{1,t-1}z_t(\rho_T) \\ 2\hat{\Lambda}_{\varepsilon\eta}(\rho_T)\sum_{t=1}^T y_{t-1} + \hat{\psi}_{lr,T}(\rho_T)\hat{\sigma}_{\tilde{\xi}^*,lr}(\rho_T)\hat{\sigma}^{-1}_{\varepsilon,lr}(\rho_T)\sum_{t=1}^T x_{2,t-1}z_t(\rho_T) \end{pmatrix}.$$

The following theorem establishes the asymptotic null distribution of $W_T^{lr}(\rho_T)$.

THEOREM 4. *Consider model (2). Under Assumption 3 and under the null* H_0 : $\omega_T = 0$, $W_T^{lr}(\rho_T)$ *converges in distribution to the chi-square distribution with 2 degrees of freedom.*

Remark 2. The above asymptotics may break down when $\rho_T = \rho$ is fixed and $\rho \neq 1$. For example, if $|\rho| < 1$ (corresponding to $a = T(\rho_T - 1) \to -\infty$) and ε_t is GARCH, then, under the null of $\omega_T^2 = 0$, regression model (7) becomes

$$z_t^2(\rho_T) = \sigma_\varepsilon^2 + \delta_T y_{t-1} + \omega_T^2 y_{t-1}^2 + \eta_t,$$

where η_t is serially correlated. In this case, stationary regressors y_{t-1} and y_{t-1}^2 can be correlated with η_t, because η_t is correlated with lagged ε_t^2, and lagged ε_t^2 may be correlated with y_{t-1} and y_{t-1}^2. This causes (δ_T, ω_T) to be unidentified in (7), and thus all tests discussed thus far are not employable. Indeed, our simulation shows that the Bonferroni–Wald test is oversized when $-a = T(1-\rho)$ is large and ε_t is a GARCH process (see Appendix E of the Supplementary Material).

Remark 3. Because ω^2 is not identified when $|\rho| < 1$ is distant from unity and ε_t is conditionally heteroskedastic (see Remark 2), the long-run variance estimators based on $\hat{\tilde{\xi}}^*_t$, the residual from (8), should perform poorly as $a = T(\rho_T - 1) \to -\infty$. In this case, $\hat{\tilde{\xi}}^*_t$ should be replaced with $\hat{\eta}_t := z_t^2(\rho_T) - T^{-1}\sum_{t=1}^T z_t^2(\rho_T)$. Since $z_t(\rho_T) = \varepsilon_t$ under the null of $\omega^2 = 0$, $\hat{\eta}_t$ is essentially the estimate of η_t under the null, while $\hat{\tilde{\xi}}^*_t$ is the estimate of η_t under both the null and alternative (as long as ω^2 is identified). This implies that the Bonferroni test based on long-run variance estimators calculated using $\hat{\eta}_t$ is relatively conservative. This replacement

https://doi.org/10.1017/S0266466625000015 Published online by Cambridge University Press

alleviates size distortions to some extent, but the Bonferroni test is still invalid when $a = T(\rho - 1) \to -\infty$ and ε_t is heteroskedastic.

4.2.2. *Uniform Confidence Interval for* ρ_T. As argued in Section 3, the validity of our Bonferroni test crucially depends on the uniform inference for the persistence parameter ρ_T. Because the confidence interval advocated by Phillips (2014) is derived under the assumption of ε_t being i.i.d., we follow Andrews and Guggenberger (2014) to adapt the confidence interval for ρ_T to allow for serial correlation and heteroskedasticity.

To fix ideas, we suppose that serial correlation in ε_t can be eliminated by a lag augmentation for model (2). Specifically, following Leybourne et al. (1996) and Andrews and Guggenberger (2014), suppose y_t has the following AR(p) representation:

$$\delta(L)\phi(L)y_t = u_t, \tag{17}$$

where $\delta(L) = 1 - (\rho_T + \omega_T v_t)L$, $\phi(L) = 1 - \phi_1 L - \ldots - \phi_{p-1}L^{p-1}$, and u_t is possibly conditionally heteroskedastic but serially uncorrelated. Furthermore, the lag polynomial $\phi(L)$ has all roots outside the unit circle. For simplicity, we consider the case where $H_0 : \omega_T = 0$ holds. We expect that the consistency argument stated below will continue to hold under the local alternatives given by $\omega_T = cT^{-3/4}$. Define $b(L) \coloneqq \delta(L)\phi(L) = (1 - \rho_T L)\phi(L)$, and write $b(L) = 1 - b_1 L - \ldots - b_p L^p$. Andrews and Guggenberger (2014) proposed a uniform confidence interval for the sum of the AR coefficients, denoted by $\gamma_T \coloneqq \sum_{i=1}^{p} b_i = 1 - b(1)$. Noting that $b(1) = \delta(1)\phi(1) = (1 - \rho_T)\phi(1) = -a\phi(1)/T$, γ_T can be further expressed as $\gamma_T = 1 + a\phi(1)/T$. This means that the confidence interval developed by Andrews and Guggenberger (2014) is for the sum of the p AR coefficients $\gamma_T = 1 + a\phi(1)/T$, not for the largest AR root $\rho_T = 1 + a/T$ (unless $p = 1$). But we can recover the confidence interval for ρ_T from that for γ_T.

To do this, we need to obtain estimates $(\hat{\phi}_1, \ldots, \hat{\phi}_{p-1})$ for $(\phi_1, \ldots, \phi_{p-1})$. First, notice that under $\rho_T = \bar{\rho}_T$, model (17) can be transformed into

$$z_t(\bar{\rho}_T) = \phi_1 z_{t-1}(\bar{\rho}_T) + \ldots + \phi_{p-1} z_{t-p+1}(\bar{\rho}_T) + u_t, \tag{18}$$

where $z_t(\bar{\rho}_T) = \delta(L)y_t = y_t - \bar{\rho}_T y_{t-1}$. Since the polynomial $\phi(L)$ has all roots outside the unit circle, the OLS estimates $(\hat{\phi}_1(\rho_T), \ldots, \hat{\phi}_{p-1}(\rho_T))$ from (18) are consistent for $(\phi_1, \ldots, \phi_{p-1})$. Therefore, $\phi \equiv \phi(1)$ can be consistently estimated by $\hat{\phi}(\bar{\rho}_T) \coloneqq 1 - \hat{\phi}_1(\bar{\rho}_T) - \ldots - \hat{\phi}_{p-1}(\bar{\rho}_T)$ given that $\rho_T = \bar{\rho}_T$ is true. Define $\bar{a} = T(\bar{\rho}_T - 1)$. If the hypothesis $\rho_T = \bar{\rho}_T$ is conformable to the data, then $\hat{\gamma}_T(\bar{\rho}_T) \coloneqq 1 + \bar{a}\hat{\phi}(\bar{\rho}_T)/T$ should belong to the $100(1 - \alpha_1)\%$ confidence interval for γ_T, denoted by $\mathrm{CI}_\gamma(\alpha_1)$, with probability $1 - \alpha_1$ asymptotically. This reasoning gives the $100(1 - \alpha_1)\%$ confidence interval for ρ_T of the form

https://doi.org/10.1017/S0266466625000015 Published online by Cambridge University Press

$$\mathrm{CI}(\alpha_1) = \Big\{ \bar{\rho}_T = 1 + \frac{\bar{a}}{T} : \hat{\gamma}_T(\bar{\rho}_T) = 1 + \frac{\bar{a}\hat{\phi}(\bar{\rho}_T)}{T} \in [\gamma_T^{\min}, \gamma_T^{\max}], \\ \text{where } \gamma_T^{\min} := \min \mathrm{CI}_\gamma(\alpha_1) \text{ and } \gamma_T^{\max} := \max \mathrm{CI}_\gamma(\alpha_1) \Big\}. \quad \textbf{(19)}$$

The whole procedure of the Bonferroni–Wald test, adapted to the above setting, is summarized as follows.

Algorithm 3.

Step 1. For model (17) with $\omega_T = 0$, following Andrews and Guggenberger (2014), compute the $100(1-\alpha_1)\%$ confidence interval for the sum of p AR coefficients, γ_T. The referenced critical values for $\alpha_1 = 0.05$ are presented in Table 1.[12] The resulting confidence interval is denoted by $\mathrm{CI}_\gamma(\alpha_1)$.

Step 2. For each $\bar{\rho}_T = 1 + \bar{a}/T$, obtain the OLS estimates $(\hat{\phi}_1(\bar{\rho}_T), \ldots, \hat{\phi}_{p-1}(\bar{\rho}_T))$ from (18). Calculate $\hat{\gamma}_T(\bar{\rho}_T) = 1 + \bar{a}\hat{\phi}(\bar{\rho}_T)/T$, where $\hat{\phi}(\bar{\rho}_T) = 1 - \hat{\phi}_1(\bar{\rho}_T) - \ldots - \hat{\phi}_{p-1}(\bar{\rho}_T)$. Then, the $100(1-\alpha_1)\%$ confidence interval $\mathrm{CI}(\alpha_1)$ for ρ_T is defined by (19).

Step 3. Conduct the procedure described in Step 2 of Algorithm 1 using $S_T(\rho_T) = W_T^{lr}(\rho_T)$ and the chi-square distribution with 2 degrees of freedom.

5. FINITE-SAMPLE PERFORMANCE

In this section, we verify whether the Bonferroni test we have proposed performs well in finite samples. We also verify whether the good performance of the infeasible Wald test (based on the true value of ρ_T) carries over to the Bonferroni–Wald test. To do this, we consider the Bonferroni–Wald test with refinement (following Algorithm 2). This is a 0.05-level test by construction.

5.1. Empirical Size

Following Nagakura (2009), we employ three data generating processes (DGP) to evaluate empirical sizes: (a) $\varepsilon_t \sim$ i.i.d $N(0,1)$, so that $\psi = 0$, (b) $\varepsilon_t \sim$ i.i.d $(\chi^2(10) - 10)/\sqrt{20}$, so that $\psi = 0.5$, and (c) $\varepsilon_t \sim$ i.i.d $(\chi^2(1) - 1)/\sqrt{2}$, so that $\psi = 0.756$. For each case, we set $y_0 = 0$ and $T \in \{200, 500, 1,000\}$. The number of replications is 5,000. The $\rho_T = \rho$ values are fixed across T, and we consider $\rho \in [0.7, 1.01]$. The simulation results are collected in Figure 2. For the normal case (Figure 2(a)), rejection rates are stable around 0.05 over $\rho \in [0.7, 1.01]$, with them approaching 0.05 as T increases. As for the chi-square cases (Figures 2(b) and (c)), rejection rates stay around the nominal level, but they get farther away from 0.05 when $|\psi|$ is larger, T is smaller and ρ is closer to 1. In particular, in case (c) with $T = 200$, the rejection rates can be as large as 0.09 (around $\rho = 1$), although they approach 0.05 as T increases. A similar tendency can be observed in the finite-sample behavior

[12] The asymptotic distribution of the t-statistic developed by Andrews and Guggenberger (2014) is (15).

https://doi.org/10.1017/S0266466625000015 Published online by Cambridge University Press

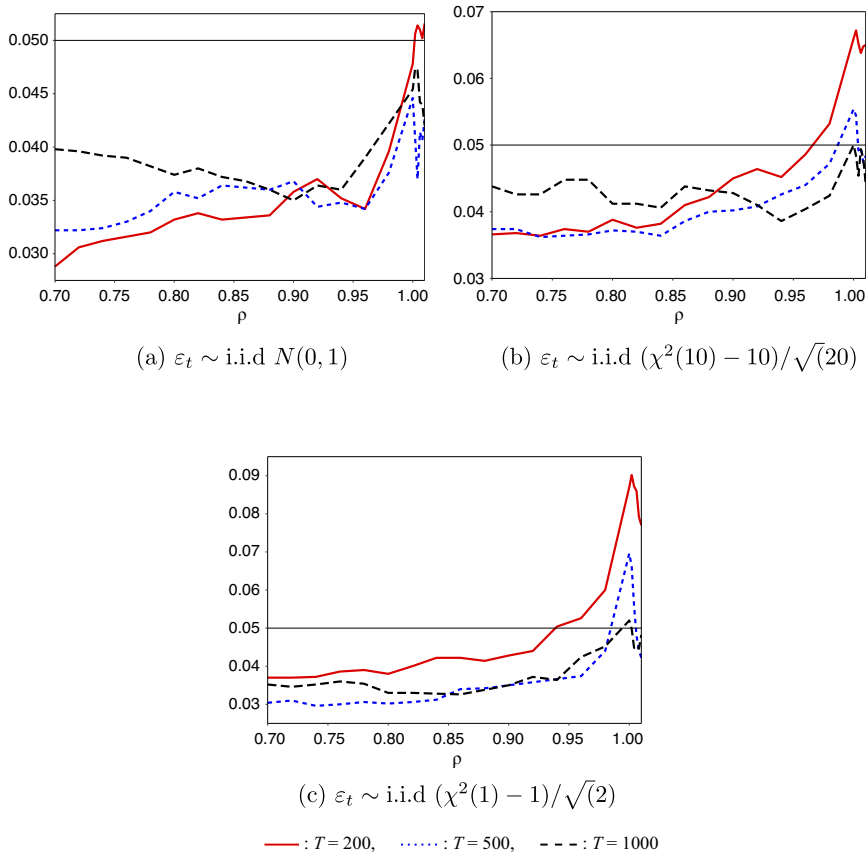


FIGURE 2. Finite-sample rejection rates of the Bonferroni–Wald test under the null with significance level 0.05.

of modified LN tests proposed by Nagakura (2009) according to his simulation results. One possible cause of this phenomenon would be the finite-sample bias involved in the estimation of ψ when the ψ value is large. One may be advised to use a more conservative Bonferroni–Wald test (with smaller α_1 values) when the $|\psi|$ estimate is large and the sample size is not so large.

As remarked in Section 3, our Bonferroni test follows the steps advocated by Phillips (2014) to avoid size distortion when y_t follows stationary AR processes.[13] Because the type 1 errors of our Bonferroni test are stable as ρ deviates from 1, the size distortion problem associated with the Campbell–Yogo approach is absent.

[13]Note that this stationarity statement is used in the usual sense ($|\rho| < 1$) since we consider $\omega = 0$.

https://doi.org/10.1017/S0266466625000015 Published online by Cambridge University Press

We also study the empirical size of the Bonferroni–Wald test modified as explained in Section 4.2 under serial correlation and heteroskedasticity. See Appendix E of the Supplementay Material for details.

5.2. Power Comparison

Next, we investigate power properties. The simulation design is as follows: $T = 200$, $\varepsilon_t \sim$ i.i.d $N(0,1)$ and $v_t \sim$ i.i.d $N(0,1)$ with $\mathrm{Corr}(\varepsilon_t, v_t) \in \{0, 0.25, 0.5, 0.75\}$, $\rho \in \{1.01, 1, 0.98, 0.95\}$, and $\omega^2 \in (0, 0.01]$ (corresponding to $c^2 \in (0, 28.28]$). The tests we consider here are the Bonferroni–Wald test, the infeasible modified Wald test (calculated using the true ρ), and one of the modified LN tests proposed by Nagakura (2009), which is denoted by $\widetilde{G}_{T,1}$ in his notation. Because Nagakura (2009) did not derive the asymptotic null distribution of $\widetilde{G}_{T,1}$ under $\rho > 1$, we do not perform it for the case $\rho = 1.01$. We also consider the test proposed recently by Horváth and Trapani (2019) (hereafter HT), which is denoted by $\Theta_{T,R}$.[14]

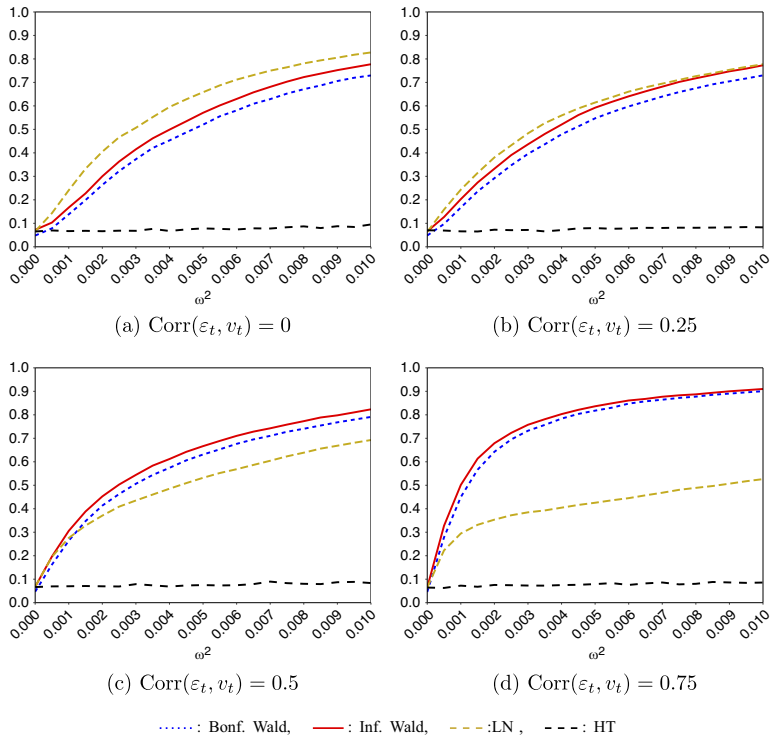


FIGURE 3. Finite-sample power functions under $\rho = 1$.

[14]The HT test needs a tuning parameter, $x \in (0, 0.5)$, to be performed. They suggested setting $x = 0.1$, but this choice somehow caused size distortions in our simulation. Thus, we set $x = 0.38$, to obtain simulation results similar to those of HT.

https://doi.org/10.1017/S0266466625000015 Published online by Cambridge University Press

For space reasons, we show the simulation results for $\rho = 1$ only (Figure 3). It is noticeable that the LN test performs well for small values of $\text{Corr}(\varepsilon_t, v_t)$, but Wald type tests outperform the LN test otherwise. The discrepancy between the infeasible and Bonferroni–Wald tests is small, thanks to the refinement of the construction of the confidence interval for ρ. In contrast, the power function of the HT test stays around the nominal level 0.05 over $\omega^2 \in (0, 0.01]$, which implies it has almost no distinguishing power for these alternatives. Indeed, HT conjectured (based on some theoretical analysis) that their test would have no power when $1 - \rho_T = O(T^{-1})$ and $\omega_T^2 = O(T^{-1})$, which is of larger magnitude than our local-to-zero variance $\omega_T^2 = O(T^{-3/2})$. Our simulation results corroborate their statement. Overall, the Bonferroni–Wald test performs better than the LN and HT tests for moderate to large values of $\text{Corr}(\varepsilon_t, v_t)$. Moreover, it performs almost as efficiently as its ideal, infeasible counterpart. The theoretical evidence proven in Section 2 that the (infeasible) Wald test has good power properties carries over to the Bonferroni–Wald test (with refinement). The same comment applies to the cases $\rho = 1.01, 0.98, 0.95$ (see Appendix E of the Supplementary Material).

6. EMPIRICAL APPLICATION

In this section, we apply the tests considered above to several U.S. macroeconomic and financial time series, following Hill and Peng (2014) and HT. The dataset includes CPI, real GDP, industrial production, M2, S&P500, the 3 month Treasury bill rate, and unemployment rate. We take the logarithm of the first five series. All data have been extracted from Federal Reserve Economic Data. The data description is provided in Table 3.

As an anonymous referee pointed out, macroeconomic and financial variables often experience structural shifts in their means and/or trends. Because our framework does not allow for structural shifts in deterministic terms, we should exclude variables that experience them from our analysis. We do this by graphical inspection since to the best of our knowledge, little is known about the properties of structural break tests under the RCA models. The time series are plotted in

TABLE 3. Data description of the series used in Section 6

Series	Frequency	Sample period	T
CPI (1982–1984 = 100)	Monthly	Jan. 1913–Dec. 2019	1,284
Real GDP (2012 chained)	Quarterly	Q1 1947–Q4 2019	292
Industrial production (2017 = 100)	Monthly	Jan. 1919–Dec. 2019	1,212
M2	Weekly	Nov. 3, 1980–Dec. 30, 2019	2,044
S&P 500	Daily	Dec. 3, 2012–Dec. 31, 2019	1,782
3-month T-bill rate	Daily	Dec. 3, 2012–Dec. 31, 2019	1,770
Unemployment rate	Monthly	Jan. 1948–Dec. 2019	864

https://doi.org/10.1017/S0266466625000015 Published online by Cambridge University Press

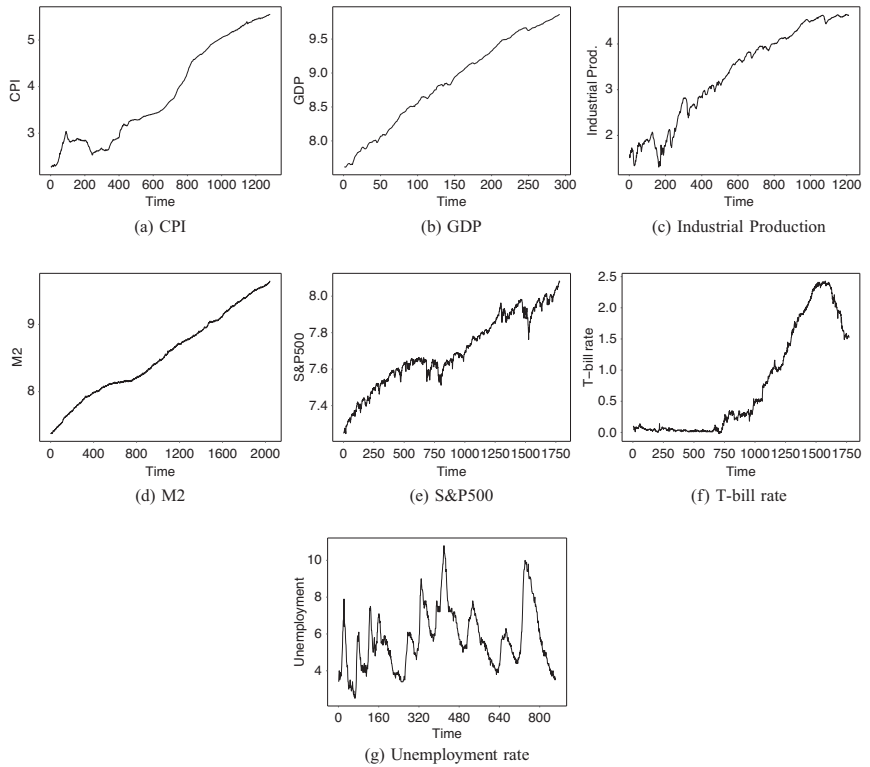


FIGURE 4. Plots of the time series used in Section 6.

Figure 4. Industrial production experienced a mean shift after the Great Crash of 1929 and a trend break in the late 1940s, and the T-bill rate exhibited mean and trend breaks in 2015 and 2016, respectively. Therefore we will not use the two variables, and detrend the remaining five variables before applying tests for coefficient randomness.

We use the modified Bonferroni–Wald test following Algorithm 3 (without refinement) with $\alpha_1 = \alpha_2 = 0.05$, and the HT test, in case ε_t is serially dependent and/or heteroskedastic.[15,16] We do not apply the LN test because its construction does not accommodate serial dependence or heteroskedasticity in the disturbance. Table 4 reports the result of 10%-level tests along with lag length p used in Step 2 of Algorithm 3 for the Bonferroni–Wald test.

[15]For the Bonferroni–Wald test, we calculate the long-run variances using $\hat{\eta}_t = z_t^2(\rho_T) - T^{-1}\sum_{t=1}^{T} z_t^2(\rho_T)$ instead of $\tilde{\tilde{\xi}}^*{}_t$; see Remark 3.

[16]Our simulation shows that the modified test is reasonably sized if $-a = T(1-\rho)$ is not large or $\rho > 0.9$, which seem to hold in our case; see the column labeled $\hat{\rho}_T$ in Table 4.

https://doi.org/10.1017/S0266466625000015 Published online by Cambridge University Press

TABLE 4. Estimation and testing results from the empirical analysis in Section 6.

Series	T	$\hat{\rho}_T$	$\hat{\omega}^2_{\mathrm{HT}}$	p	Bonferroni–Wald	HT ($\Theta_{T,R}$)
CPI	1,284	0.999	4.12×10^{-4}	8	Yes	–
GDP	292	0.994	1.16×10^{-3}	4	–	–
M2	2,044	0.995	3.12×10^{-4}	8	Yes	–
S&P 500	1,782	0.984	9.38×10^{-3}	8	Yes	–
Unemployment rate	864	0.993	4.03×10^{-4}	8	–	–

For entries in the last two columns, "Yes" ("–") signifies the rejection (nonrejection) of the null $\mathrm{H}_0 : \omega^2 = 0$ by the corresponding test with the 10% significance level.

For GDP and unemployment rate, we have obtained consistent results from both tests: The null of $\mathrm{H}_0 : \omega^2 = 0$ is not rejected for these series. For CPI, M2 and S&P500, the Bonferroni–Wald test rejects the null while the HT test does not. This result will be attributed to the fact that the former test is much more powerful than the latter.

Table 4 also presents the ω^2 estimates proposed by Horváth and Trapani (2019) (denoted by $\hat{\omega}^2_{\mathrm{HT}}$). The values of $\hat{\omega}^2_{\mathrm{HT}}$ seem to support our testing results because $\hat{\omega}^2_{\mathrm{HT}}$ takes values around 10^{-4} to 10^{-3}, magnitudes of coefficient randomness with which the HT test tends to be powerless under the sample size given in Table 4. For example, for M2, the sample size is $T = 2044$, and $\hat{\omega}^2_{\mathrm{HT}}$ is 3.12×10^{-4}, translating into $\hat{c}^2 = \hat{\omega}^2_{\mathrm{HT}} \times T^{3/2} = 28.83$ estimate of the localizing coefficient c^2. According to our simulation results, the HT test has almost no power against the alternative of this magnitude, hence the result reported in Table 4.

7. CONCLUSION AND DISCUSSION

Given the results of empirical analyses conducted by earlier studies, the local-to-unity RCA model is empirically relevant. Under this setting, we analyzed the effect of the correlation between the random coefficient and disturbance on the power properties of tests for coefficient randomness. Theoretical and simulation analyses revealed that tests proposed by earlier studies can perform poorly when the correlation is moderate to large and the coefficient randomness is local to zero, while the augmented-Wald test we proposed performs well even in such cases. Our test is operational due to its independence from nuisance parameter ψ, the correlation between the disturbance and its square. To deal with the uncertainty about the mean ρ of the autoregressive root, we proposed using a confidence interval for ρ, leading to the Bonferroni–Wald test. Our Bonferroni method succeeds in avoiding the size distortion problem encountered with earlier approaches. Simulations corroborated that good performance of the infeasible Wald test (based on the true value of ρ) carries over to the Bonferroni–Wald test after a refinement is made. A modification of the Bonferroni–Wald test was

https://doi.org/10.1017/S0266466625000015 Published online by Cambridge University Press

suggested to allow for nonzero mean and linear trend in the data and serial dependence and conditional heteroskedasticity in the disturbance.

Several directions for future research are possible. First, from the similarity in the construction of test statistics between our model and predictive regressions, it is expected that the theory developed in numerous studies on predictive regressions can be applied to testing for coefficient randomness. For example Phillips and Lee (2013, 2016) proposed the use of the so-called IVX procedure for predictability testing (see also Kostakis et al., 2015). This procedure shows good size and power, and also requires less computational burden for implementation than the Bonferroni approach. The use of the IVX approach may facilitate testing for coefficient randomness in local-to-unity autoregressions. The analysis of test performance when ρ is distant from unity will also be of interest. In such a case, more preferable tests might be available than the Bonferroni–Wald test based on the local-to-unity asymptotics.

SUPPLEMENTARY MATERIAL

Nishi, Mikihito (2025): Supplement to "Testing for Coefficient Randomness in Local-to-Unity Autoregressions", Econometric Theory Supplementary Material. To view, please visit https://doi.org/10.1017/S0266466625000015.

REFERENCES

Andrews, D. W. K., & Guggenberger, P. (2014). A conditional-heteroskedasticity-robust confidence interval for the autoregressive parameter. *The Review of Economics and Statistics*, 96(2), 376–381.

Aue, A., & Horváth, L. (2011). Quasi-likelihood estimation in stationary and nonstationary autoregressive models with random coefficients. *Statistica Sinica*, 21(3), 973–999.

Bykhovskaya, A., & Phillips, P. C. B. (2018). Boundary limit theory for functional local to unity regression. *Journal of Time Series Analysis*, 39(4), 523–562.

Bykhovskaya, A., & Phillips, P. C. B. (2020). Point optimal testing with roots that are functionally local to unity. *Journal of Econometrics*, 219(2), 231–259.

Campbell, J. Y., & Yogo, M. (2006). Efficient tests of stock return predictability. *Journal of Financial Economics*, 81(1), 27–60.

Cavanagh, C. L., Elliott, G., & Stock, J. H. (1995). Inference in models with nearly integrated regressors. *Econometric Theory*, 11(5), 1131–1147.

Distaso, W. (2008). Testing for unit root processes in random coefficient autoregressive models. *Journal of Econometrics*, 142(1), 581–609.

Elliott, G., Rothenberg, T. J., & Stock, J. H. (1996). Efficient tests for an autoregressive unit root. *Econometrica*, 64(4), 813–836.

Hill, J., & Peng, L. (2014). Unified interval estimation for random coefficient autoregressive models. *Journal of Time Series Analysis*, 35(3), 282–297.

Horváth, L., & Trapani, L. (2019). Testing for randomness in a random coefficient autoregression model. *Journal of Econometrics*, 209(2), 338–352.

Horváth, L., & Trapani, L. (2023). Changepoint detection in heteroscedastic random coefficient autoregressive models. *Journal of Business & Economic Statistics*, 41(4), 1300–1314.

Hwang, S. Y., & Basawa, I. V. (2005). Explosive random-coefficient AR(1) processes and related asymptotics for least-squares estimation. *Journal of Time Series Analysis*, 26(6), 807–824.

https://doi.org/10.1017/S0266466625000015 Published online by Cambridge University Press

Kostakis, A., Magdalinos, T., & Stamatogiannis, M. P. (2015). Robust econometric inference for stock return predictability. *The Review of Financial Studies*, 28(5), 1506–1553.

Lee, S. (1998). Coefficient constancy test in a random coefficient autoregressive model. *Journal of Statistical Planning and Inference*, 74(1), 93–101.

Leybourne, S. J., McCabe, B. P. M., & Tremayne, A. R. (1996). Can economic time series be differenced to stationarity? *Journal of Business & Economic Statistics*, 14(4), 435.

Liang, H., Phillips, P. C. B., Wang, H., & Wang, Q. (2016). Weak convergence to stochastic integrals for econometric applications. *Econometric Theory*, 32(6), 1349–1375.

McCabe, B. P. M., & Smith, R. J. (1998). The power of some tests for difference stationarity under local heteroscedastic integration. *Journal of the American Statistical Association*, 93(442), 751–761.

McCabe, B. P. M., & Tremayne, A. R. (1995). Testing a time series for difference stationarity. *The Annals of Statistics*, 23(3), 1015–1028.

Nagakura, D. (2009). Testing for coefficient stability of AR(1) model when the null is an integrated or a stationary process. *Journal of Statistical Planning and Inference*, 139(8), 2731–2745.

Nicholls, D. F., & Quinn, B. G. (1982). *Random coefficient autoregressive models: An introduction*, 11 of Lecture Notes in Statistics, New York, Springer. https://doi.org/10.1007/978-1-4684-6273-9.

Nishi, M., & Kurozumi, E. (2024). Stochastic local and moderate departures from a unit root and its application to unit root testing. *Journal of Time Series Analysis*, 45(1), 133–157.

Phillips, P. C. B. (1987a). Time series regression with a unit root. *Econometrica*, 55(2), 277–301.

Phillips, P. C. B. (1987b). Towards a unified asymptotic theory for autoregression. *Biometrika*, 74(3), 535–547.

Phillips, P. C. B. (2014). On confidence intervals for autoregressive roots and predictive regression. *Econometrica*, 82(3), 1177–1195.

Phillips, P. C. B. (2023). Estimation and inference with near unit roots. *Econometric Theory*, 39(2), 221–263.

Phillips, P. C. B., & Lee, J. H. (2013). Predictive regression under various degrees of persistence and robust long-horizon regression. *Journal of Econometrics*, 177(2), 250–264.

Phillips, P. C. B., & Lee, J. H. (2016). Robust econometric inference with mixed integrated and mildly explosive regressors. *Journal of Econometrics*, 192(2), 433–450.

Su, J.-J., & Roca, E. (2012). Examining the power of stochastic unit root tests without assuming independence in the error processes of the underlying time series. *Applied Economics Letters*, 19(4), 373–377.

https://doi.org/10.1017/S0266466625000015 Published online by Cambridge University Press