*Econometric Theory*, **40**, 2024, 1159–1183.
doi:10.1017/S0266466622000652

# NEARLY EFFICIENT LIKELIHOOD RATIO TESTS OF A UNIT ROOT IN AN AUTOREGRESSIVE MODEL OF ARBITRARY ORDER

SAMUEL BRIEN
*Queen's University*

MICHAEL JANSSON
*UC Berkeley and CREATES*

MORTEN ØRREGAARD NIELSEN
*Aarhus University*

We study large sample properties of likelihood ratio tests of the unit-root hypothesis in an autoregressive model of arbitrary order. Earlier research on this testing problem has developed likelihood ratio tests in the autoregressive model of order 1, but resorted to a plug-in approach when dealing with higher-order models. In contrast, we consider the full model and derive the relevant large sample properties of likelihood ratio tests under a local-to-unity asymptotic framework. As in the simpler model, we show that the full likelihood ratio tests are nearly efficient, in the sense that their asymptotic local power functions are virtually indistinguishable from the Gaussian power envelopes. Extensions to sieve-type approximations and different classes of alternatives are also considered.

## 1. INTRODUCTION

In their seminal contribution, Elliott, Rothenberg, and Stock (1996; henceforth ERS) derived Gaussian power envelopes for the unit-root testing problem in autoregressive models and demonstrated how to construct tests that are "nearly efficient" in the sense that their asymptotic local power functions are virtually indistinguishable from the Gaussian power envelopes. In particular, they showed that generalized least squares (GLS) detrended versions of the well-known augmented Dickey–Fuller (ADF) tests (Dickey and Fuller, 1979, 1981) are nearly efficient. More recently, Jansson and Nielsen (2012; henceforth JN) developed a class of tests that are also nearly efficient, yet distinct from the tests proposed by ERS. In the autoregressive model of order 1, the tests proposed by JN admit a

We are grateful to Peter Phillips, two anonymous referees, and seminar participants at Queen's University for comments and discussion, and to the Danish National Research Foundation for financial support (DNRF Chair Grant No. DNRF154). Address correspondence to Michael Jansson, Department of Economics, University of California, Berkeley, 530 Evans Hall #3880, Berkeley, CA 94720, USA; e-mail: mjansson@econ.berkeley.edu.

quasi-likelihood ratio (QLR) interpretation, but for higher-order autoregressive models, the method of proof employed by JN forces them to use a "two-step"/"plug-in" approach, where the nuisance parameters arising from the lag augmentation are replaced with consistent estimators when defining the criterion function used to construct the test.

Although nearly efficient, the tests of JN therefore do not admit a QLR interpretation in the higher-order case. In fact, even after several decades of intense research into this testing problem, it would appear that a nearly efficient QLR test of the unit-root hypothesis in an autoregressive model of arbitrary order has still not been developed and investigated. In this paper, we fill this apparent hole in the literature. Our analysis is motivated partly by a desire to make the theory of univariate unit-root testing more complete by developing QLR tests in the workhorse model of the literature, and showing that these tests belong to the class of nearly efficient tests. Moreover, and perhaps just as importantly, with an eye toward other nonstandard testing problems, it is of interest to understand the consequences of (and demonstrate the feasibility of) handling all nuisance parameters in a unified way in this canonical nonstandard testing problem.

The remainder of the paper is organized as follows. In the next section, we present the model, derive the test statistics, and characterize their large sample properties. In Section 3, we analyze a sieve version of our QLR test. In Section 4, we present the results of a small simulation study of the finite-sample properties of the new test and compare with some existing tests. Section 5 discusses different classes of alternatives, and Section 6 offers some concluding remarks. Finally, the proofs of our main results are given in the Appendix.

## 2. MODEL AND QLR TEST STATISTIC

Our goal is to develop unit-root tests that are of QLR type, are easy to implement, and enjoy good size and power properties in a model of the type considered in ERS. To this end, suppose that the observed time series $\{y_t : 1 \leq t \leq T\}$ is generated as

$$y_t = \beta' d_t + u_t, \tag{1}$$

where $d_t = 1$ or $d_t = (1,t)'$, $\beta$ is an unknown parameter, and the error term $u_t$ is generated by the AR$(p+1)$ model

$$(1-\rho L)\gamma(L)u_t = \varepsilon_t, \tag{2}$$

with $\rho \leq 1$ a scalar parameter of interest and $\gamma(z) = 1-\gamma_1 z - \cdots - \gamma_p z^p$ a lag polynomial of order $p$ satisfying the stability condition $\gamma = (\gamma_1,\ldots,\gamma_p)' \in \left\{\gamma \in \mathbb{R}^p : \min_{|z|\leq 1}|\gamma(z)| > 0\right\} = \Gamma$. When developing formal results, we will complete the specification of the model by assuming that $\max\left\{|u_0|,\ldots,|u_{-p}|\right\} = o_p(T^{1/2})$ and that the $\varepsilon_t$ form a conditionally homoskedastic martingale difference sequence with (unknown) variance $\sigma^2$ and $\sup_t E|\varepsilon_t|^r < \infty$ for some $r > 2$.

In the model characterized by (1) and (2), an implication of assuming $\gamma \in \Gamma$ is that the order of integration of $u_t$ is governed solely by $\rho$. In particular, the unit-root testing problem is the problem of testing

$$H_0 : \rho = 1 \qquad \text{versus} \qquad H_1 : \rho < 1.$$

The Gaussian quasi-log-likelihood function corresponding to the model given by (1) and (2) with initial conditions $u_0 = \cdots = u_{-p} = 0$ depends on the parameter of interest, $\rho$, and the nuisance parameters $\beta$, $\gamma$, and $\sigma^2$. To be specific, setting $y_0 = \cdots = y_{-p} = 0$ and $d_0 = \cdots = d_{-p} = 0$, the Gaussian quasi-log-likelihood function can be expressed, up to an additive constant, as

$$L_T(\rho,\beta,\gamma,\sigma^2) = -\frac{T}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{t=1}^{T}((1-\rho L)\gamma(L)(y_t - \beta' d_t))^2.$$

The QLR test statistic associated with the problem of testing $H_0$ versus $H_1$ is

$$LR_T = \max_{\rho\leq 1,\,\beta,\,\gamma\in\Gamma,\,\sigma^2>0} L_T(\rho,\beta,\gamma,\sigma^2) - \max_{\beta,\,\gamma\in\Gamma,\,\sigma^2>0} L_T(1,\beta,\gamma,\sigma^2).$$

In general, the problem of maximizing $L_T(\rho,\beta,\gamma,\sigma^2)$ with respect to $\gamma \in \Gamma$ does not have a closed-form solution. For this reason, $LR_T$ can be tedious to compute unless $p$ is small, which reduces the practical usefulness of $LR_T$. A separate concern of a more technical nature is that the lack of a closed-form expression for $\max_{\gamma\in\Gamma} L_T(\rho,\beta,\gamma,\sigma^2)$ makes the development of large sample theory for $LR_T$ quite challenging when $p$ is allowed to grow with the sample size (see Section 3). For these reasons, it is natural to ask whether the model (2) can be embedded in a model whose associated QLR unit-root test statistic is analytically tractable even when $p$ is large, yet enjoys attractive large sample (power) properties.

Analytical tractability could be restored by simply dropping the constraint $\gamma \in \Gamma$, as the problem of maximizing $L_T(\rho,\beta,\gamma,\sigma^2)$ with respect to $\gamma \in \mathbb{R}^p$ has a well-known solution. Because the constraint $\gamma \in \Gamma$ implies that the order of integration of $u_t$ is governed solely by $\rho$, dropping the constraint could have consequences for power, however. To anticipate those consequences, notice that when $p = 1$, $\rho$ and $\gamma$ are not separately identified in (2), implying that maximizing the quasi-likelihood with a restriction on $\rho$, but not on $\gamma$, is equivalent to maximizing it with a restriction on $\gamma$, but not on $\rho$. The associated test statistic is therefore equivalent to a statistic associated with the two-sided problem of testing $H_0 : \rho = 1$ versus $H_2 : \rho \neq 1$. In fact, it can be shown that the QLR test statistic implemented without restrictions on $\gamma$ behaves like a "two-sided" test statistic in the sense that its limiting distribution is that of $\max_{\bar{c}\in\mathbb{R}} \Lambda_c(\bar{c})$ ($\max_{\bar{c}\in\mathbb{R}} \Lambda_c^\tau(\bar{c})$) under the assumptions of part (a) (part (b)) of Theorem 1. In other words, although dropping the constraint $\gamma \in \Gamma$ is computationally convenient, the resulting model is not asymptotically equivalent (in the appropriate sense) to the model imposing $\gamma \in \Gamma$.

Fortunately, the model (2) can be embedded in a model that is locally equivalent to it in a suitable sense, yet generates a QLR test statistic that is relatively easy to compute. To be specific, the model (2) can be embedded in a model of ADF type, namely

$$\eta(L)\Delta u_t = \pi u_{t-1} + \varepsilon_t, \tag{3}$$

where $\{u_t\}$ and $\{\varepsilon_t\}$ are as before, $\pi \leq 0$, and $\eta(z) = 1 - \eta_1 z - \cdots - \eta_p z^p$ is an unrestricted lag polynomial of order $p$; that is, $\eta = (\eta_1, \ldots, \eta_p)' \in \mathbb{R}^p$.

When $(\rho, \gamma)$ and $(\pi, \eta)$ are unrestricted, the models (2) and (3) are equivalent in the sense that one is a reparameterization of the other. In particular, as pointed out by a referee, (3) can be obtained from (2) by setting $\pi = (\rho - 1)\gamma(1)$ and

$$\eta(z) = (1-\rho)\gamma(1) + \frac{(1-\rho z)\gamma(z) - (1-\rho)\gamma(1)}{1-z}.$$

However, the equivalence between (2) and (3) breaks down once the parameter restrictions mentioned in the above text are imposed. On the one hand, when $\rho \leq 1$ and $\gamma \in \Gamma$ in (2), the parameter $\pi$ in (3) satisfies $\pi = (\rho - 1)\gamma(1) \leq 0$, so the (single) restriction $\pi \leq 0$ imposed in (3) is implied by restrictions $\rho \leq 1$ and $\gamma \in \Gamma$ imposed in (2). On the other hand, not all models of the form (3) with $\pi \leq 0$ can be written in the form (2) with $\rho \leq 1$ and $\gamma \in \Gamma$. Perhaps the easiest way to see this is to observe that, whereas the model (3) can generate I($d$) processes for any $d = 0, 1, \ldots, p+1$, the model (2) can only generate I(0) and I(1) processes (when $-1 < \rho \leq 1$). There is therefore a meaningful "global" sense in which the models (2) and (3) differ (once the parameter restrictions mentioned in the above text are imposed), with the latter being a strict generalization of the former.

Precisely for this reason, one might reasonably worry that QLR tests developed for the model (3) would suffer from power losses when applied to data generated by the model of interest in this paper, namely (2). Fortunately, and perhaps surprisingly, that turns out not to happen. Indeed, Theorem 1 implies, among other things, that QLR tests developed for the model (3) are nearly efficient in the model (2). Loosely speaking, the model (3) (with $\pi \leq 0$) therefore enjoys the "Goldilocks" property of being just flexible enough relative to the model (2) (with $\rho \leq 1$ and $\gamma \in \Gamma$) to achieve tractability on the part of QLR statistics, yet restrictive enough to ensure that no loss of power is suffered by the resulting QLR test. More precisely, our findings demonstrate by example that the models (2) and (3) are "locally" equivalent in a neighborhood of the null hypothesis $\rho = 1$ in a meaningful sense. In particular, although (2) imposes multiple restrictions on the parameters $\rho \leq 1$ and $\gamma \in \Gamma$ that are not imposed by (3), an implication of our results is that the only asymptotically relevant restriction imposed in (2) is captured by the single restriction $\pi \leq 0$ imposed in (3).

To summarize, following ERS (and many others), we are interested in developing tests that are powerful in the model (2). To this end, it turns out to be attractive to use the model (3) because it gives rise to QLR test statistics that are not only (computationally and analytically) tractable, but also powerful when applied to data generated by the model of interest, namely (2). In large part, the fact that working with (3) enables us to achieve the dual objectives of tractability and power

https://doi.org/10.1017/S0266466622000652 Published online by Cambridge University Press

is attributable to the property of the model (3) that, with or without restrictions on $\eta$, the restriction $\pi \leq 0$ turns out to incorporate the main statistical content of the restrictions $\rho \leq 1$ and $\gamma(1) > 0$ imposed in (2).

The problem of testing $H_0$ versus $H_1$ in the model characterized by (1) and (2) is subsumed in the problem of testing

$$H_0^{ADF}: \pi = 0 \qquad \text{versus} \qquad H_1^{ADF}: \pi < 0$$

in the model characterized by (1) and (3). In terms of the parameter of interest $\pi$ and the nuisance parameters $\beta$, $\eta$, and $\sigma^2$, the Gaussian quasi-log-likelihood function associated with the model given by (1) and (3) with initial conditions $u_0 = \cdots = u_{-p} = 0$ can be expressed, up to an additive constant, as

$$L_T^{ADF}(\pi,\beta,\eta,\sigma^2) = -\frac{T}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{t=1}^{T}((\eta(L)(1-L) - \pi L)(y_t - \beta' d_t))^2,$$

and the QLR test statistic associated with the problem of testing $H_0^{ADF}$ versus $H_1^{ADF}$ is

$$LR_T^{ADF} = \max_{\pi\leq 0,\beta,\eta,\sigma^2>0} L_T^{ADF}(\pi,\beta,\eta,\sigma^2) - \max_{\beta,\eta,\sigma^2>0} L_T^{ADF}(0,\beta,\eta,\sigma^2).$$

The statistic $LR_T^{ADF}$ is relatively easy to compute and analyze. The main reason is that, because $\eta$ is unrestricted, the profile quasi-log likelihood for $(\pi,\beta)$ obtained by maximizing $L_T^{ADF}(\pi,\beta,\eta,\sigma^2)$ with respect to $(\eta,\sigma^2)$ is available in closed form. For any $(\pi,\beta)$, define $V_{\pi,\beta}$ and $Z_\beta$ as the matrices with row $t = 1,\ldots,T$ given by $(1-L-\pi L)(y_t - \beta' d_t)$ and $(1-L)(y_{t-1} - \beta' d_{t-1},\ldots,y_{t-p} - \beta' d_{t-p})$, respectively. Employing this notation, $L_T^{ADF}(\pi,\beta,\eta,\sigma^2)$ can be written as

$$L_T^{ADF}(\pi,\beta,\eta,\sigma^2) = -\frac{T}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\left.(V-Z\eta)'(V-Z\eta)\right|_{V=V_{\pi,\beta},Z=Z_\beta}.$$

It follows from standard least-squares arguments that

$$\text{argmax}_\eta L_T^{ADF}(\pi,\beta,\eta,\sigma^2) = \left.(Z'Z)^{-1}Z'V\right|_{V=V_{\pi,\beta},Z=Z_\beta}$$

and

$$\text{argmax}_{\sigma^2} L_T^{ADF}(\pi,\beta,\eta,\sigma^2) = \frac{1}{T}\left.(V-Z\eta)'(V-Z\eta)\right|_{V=V_{\pi,\beta},Z=Z_\beta}.$$

As a consequence, up to an additive constant, the profile quasi-log likelihood for $(\pi,\beta)$ is given by

$$\mathcal{L}_T^{ADF}(\pi,\beta) = -\frac{T}{2}\log\left.(V'V - V'Z(Z'Z)^{-1}Z'V)\right|_{V=V_{\pi,\beta},Z=Z_\beta}.$$

Therefore, the statistic $LR_T^{ADF}$ admits the representation

$$LR_T^{ADF} = \max_{\pi\leq 0,\beta}\mathcal{L}_T^{ADF}(\pi,\beta) - \max_{\beta}\mathcal{L}_T^{ADF}(0,\beta), \tag{4}$$

where both terms on the right-hand side are relatively easy to evaluate numerically.

https://doi.org/10.1017/S0266466622000652 Published online by Cambridge University Press

In addition, and perhaps more importantly, $LR_T^{ADF}$ turns out to have attractive large sample power properties. For any $c$, let $W_c$ denote the Ornstein–Uhlenbeck process given by

$$W_c(r) = \int_0^r \exp(c(r-s))dW(s), \tag{5}$$

where $W$ is a standard Wiener process.

THEOREM 1. *Suppose that* $\{y_t\}$ *is generated by (1) and (2) and that* $c = T(\rho - 1)$ *is held fixed as* $T \to \infty$.

(a) *If* $d_t = 1$, *then* $LR_T^{ADF} \to_d \max_{\bar{c} \le 0} \Lambda_c(\bar{c})$, *where*

$$\Lambda_c(\bar{c}) = \bar{c} \int_0^1 W_c(r) dW_c(r) - \frac{1}{2}\bar{c}^2 \int_0^1 W_c(r)^2 dr.$$

(b) *If* $d_t = (1,t)'$, *then* $LR_T^{ADF} \to_d \max_{\bar{c} \le 0} \Lambda_c^{\tau}(\bar{c})$, *where*

$$\Lambda_c^{\tau}(\bar{c}) = \Lambda_c(\bar{c}) + \frac{1}{2} \frac{\left((1-\bar{c})W_c(1) + \bar{c}^2 \int_0^1 rW_c(r)dr\right)^2}{1 - \bar{c} + \bar{c}^2/3} - \frac{1}{2} W_c(1)^2.$$

A proof of Theorem 1 is provided in the Appendix. The asymptotic distributions obtained in the theorem coincide with those obtained by JN for their statistic $\widehat{LR}_T^d$. As a consequence, $LR_T^{ADF}$ shares with $\widehat{LR}_T^d$ the property that a test based on it is nearly efficient in the sense that its asymptotic local power function is indistinguishable from the Gaussian power envelope. Moreover, the critical values obtained by JN are applicable to $LR_T^{ADF}$ as well. For completeness, we reproduce these in Table 1.

In the spirit of JN, one can obtain statistics that are asymptotically equivalent to $LR_T^{ADF}$ by replacing judiciously chosen nuisance parameters with estimators and then maximizing the resulting plug-in version of the quasi-likelihood under $H_0^{ADF}$ and $H_1^{ADF}$. To be specific, a natural ADF version of the statistic $\widehat{LR}_T^d$ of JN is given by

$$\max_{\pi \le 0, \beta} L_T^{ADF}(\pi, \beta, \tilde{\eta}_T, \tilde{\sigma}_T^2) - \max_{\beta} L_T^{ADF}(0, \beta, \tilde{\eta}_T, \tilde{\sigma}_T^2),$$

where $\tilde{\eta}_T$ and $\tilde{\sigma}_T^2$ are estimators of $\eta$ and $\sigma^2$, respectively. It can be shown that, under the assumptions of Theorem 1, this statistic is asymptotically equivalent to $LR_T^{ADF}$ if $\tilde{\eta}_T$ and $\tilde{\sigma}_T^2$ are consistent.

The ADF test and the DF-GLS test of ERS are both asymptotically equivalent to tests based on a statistic of the form

$$\max_{\pi \le 0, \eta, \sigma^2 > 0} L_T^{ADF}(\pi, \tilde{\beta}_T, \eta, \sigma^2) - \max_{\eta, \sigma^2 > 0} L_T^{ADF}(0, \tilde{\beta}_T, \eta, \sigma^2),$$

https://doi.org/10.1017/S0266466622000652 Published online by Cambridge University Press

**Table 1.** Quantiles of the distribution of $LR_T^{ADF}$.

| $T$ | 80% | 85% | 90% | 95% | 97.5% | 99% | 99.5% | 99.9% |
|---|---|---|---|---|---|---|---|---|
| Panel A: constant mean case, $d_t = 1$ | | | | | | | | |
| 100 | 0.81 | 1.07 | 1.45 | 2.14 | 2.84 | 3.74 | 4.42 | 5.93 |
| 250 | 0.78 | 1.02 | 1.36 | 1.99 | 2.65 | 3.56 | 4.25 | 5.86 |
| 500 | 0.77 | 1.00 | 1.33 | 1.93 | 2.56 | 3.44 | 4.11 | 5.70 |
| 1,000 | 0.77 | 0.99 | 1.32 | 1.91 | 2.52 | 3.36 | 4.01 | 5.57 |
| $\infty$ | 0.76 | 0.98 | 1.31 | 1.88 | 2.48 | 3.29 | 3.92 | 5.40 |
| Panel B: linear trend case, $d_t = (1,t)'$ | | | | | | | | |
| 100 | 2.50 | 2.86 | 3.34 | 4.14 | 4.91 | 5.89 | 6.60 | 8.17 |
| 250 | 2.47 | 2.82 | 3.29 | 4.09 | 4.88 | 5.89 | 6.65 | 8.38 |
| 500 | 2.46 | 2.80 | 3.28 | 4.07 | 4.85 | 5.86 | 6.63 | 8.36 |
| 1,000 | 2.46 | 2.80 | 3.27 | 4.05 | 4.83 | 5.84 | 6.59 | 8.31 |
| $\infty$ | 2.45 | 2.79 | 3.26 | 4.05 | 4.82 | 5.82 | 6.57 | 8.29 |

*Notes:* Entries for finite $T$ are simulated quantiles of $LR_T^{ADF}$ with known $(\gamma,\sigma^2)$ and with $\varepsilon_t \sim$ *i.i.d.* $\mathcal{N}(0,1)$. Entries for $T=\infty$ are simulated quantiles of $\max_{\bar{c}\leq 0}\Lambda_0(\bar{c})$ and $\max_{\bar{c}\leq 0}\Lambda_0^{\tau}(\bar{c})$, respectively, where Wiener processes are approximated by $10^4$ discrete steps with standard Gaussian innovations. All entries are based on $10^7$ Monte Carlo replications.
*Source:* This table is taken from Table 1 of JN.

where $\tilde{\beta}_T$ is an estimator of $\beta$. This statistic differs from $LR_T^{ADF}$ (only) because the nuisance parameter $\beta$ has been replaced by the estimator $\tilde{\beta}_T$. The ADF test employs an OLS estimator of $\beta$, whereas the DF-GLS test employs a GLS-type estimator, but irrespective of the choice of $\tilde{\beta}_T$, the displayed statistic turns out to be asymptotically distinct from $LR_T^{ADF}$ when $d_t = (1,t)'$. In other words, although $\eta$ and/or $\sigma^2$ can be replaced with well-behaved estimators without any asymptotic consequences, a plug-in version of $LR_T^{ADF}$ in which $\beta$ has been replaced by an estimator turns out to be distinct from $LR_T^{ADF}$, even in the limit. Similarly, the point optimal test statistic of ERS is asymptotically distinct from $LR_T^{ADF}$, being of the form

$$\max_{\beta} L_T^{ADF}(T^{-1}\bar{c}_{ERS},\beta,0,\tilde{\omega}_T^2) - \max_{\beta} L_T^{ADF}(0,\beta,0,\tilde{\omega}_T^2),$$

where $\bar{c}_{ERS}$ is a negative constant and $\tilde{\omega}_T^2$ is an estimator of $\gamma(1)^{-2}\sigma^2$, the long-run variance of $(1-\rho L)u_t$. For additional details and further discussion, see Section 3 of JN.

## 3. SIEVE QLR TEST STATISTIC

It would be of interest, both practically and theoretically, to allow for more general short-run dynamics than the AR($p$) model considered in (2). In particular, as

https://doi.org/10.1017/S0266466622000652 Published online by Cambridge University Press

alluded to earlier, the fact that the dimension of $(\pi, \beta)$ does not depend on $p$ suggests that $LR_T^{ADF}$ will be well behaved also when $p$ is allowed to grow with $T$, in which case the model (2) can be interpreted as a sieve-type approximation to a more general model (e.g., Berk, 1974; Said and Dickey, 1984). Following Chang and Park (2002), these heuristics can be made precise by replacing (2) with the ARMA$(1,\infty)$ model

$$(1-\rho L)u_t = \psi(L)\varepsilon_t, \tag{6}$$

where $u_0 = O_p(1)$, the $\varepsilon_t$ form a conditionally homoskedastic martingale difference sequence with (unknown) variance $\sigma^2$ and $\sup_t E\varepsilon_t^4 < \infty$, and $\psi(z) = \sum_{j=0}^{\infty} \psi_j z^j$ satisfies $\sum_{j=0}^{\infty} j|\psi_j| < \infty$ and $\min_{|z|\leq 1} |\psi(z)| > 0$.

In this more general situation, the model (2) with a finite lag order $p$ is a sieve approximation to the process (6) (cf. Remark 2.1 of Chang and Park, 2002). The quality of the approximation improves as $p$ increases, so we let $p$ increase with the sample size, and to make this explicit, we sometimes write $p = p_T$. Letting $LR_T^{ADF}$ be calculated as in (4), the method of proof of Theorem 1 can be adapted with the help of results from Chang and Park (2002) and Phillips and Solo (1992) to show the following result (the proof of which is given in the Appendix).

THEOREM 2. *Suppose that $\{y_t\}$ is generated by (1) and (6) and that $c = T(\rho-1)$ is held fixed as $T \to \infty$. Suppose also that $p = p_T$ satisfies $p_T \to \infty$ and $p_T = o(T^{1/3})$ as $T \to \infty$. Then the results of parts (a) and (b) of Theorem 1 continue to hold.*

## 4. MONTE CARLO SIMULATIONS

To assess the finite-sample properties of $LR_T^{ADF}$ and some of its rivals, we conduct a small Monte Carlo simulation experiment. For specificity, we consider data generating processes (DGPs) of the form (1) and (2) with $\beta = 0, p = 3, u_0 = u_{-1} = \cdots = u_{-3} = 0$, and $\varepsilon_t \sim i.i.d.\,\mathcal{N}(0,1)$. For each of $10^5$ replications, we simulate data from the model with sample size $T \in \{300, 1{,}000\}$ and the parameter of interest $\rho$ either equal to 1, mildly explosive, or belonging to a grid chosen to ensure that the rejection frequencies of the various tests are around 0.4, 0.7, and 0.9, respectively. Regarding the nuisance parameter $\gamma$, we employ a parameterization of the form $\gamma(z) = \prod_{i=1}^{3}(1-\phi_i z)$, where $\phi_i$ are the inverse roots of the polynomial $\gamma(z)$. A range of values of $\phi = (\phi_1, \phi_2, \phi_3)$ was considered, but to conserve space, we only report results for some representative cases, where $\phi$ equals $(0,0,0)$, $(0.2, 0.4, 0.6)$, $(0.4, 0.4, 0.4)$, and $(0.6, 0.6, 0.6)$, respectively. These all correspond to roots that are outside the unit circle.

For each DGP, we implement three tests. The first of these is the test based on $LR_T^{ADF}$ using a lag length selected by applying the modified Akaike information criterion (MAIC) of Perron and Qu (2007) (see also Ng and Perron, 2001) to the ADF model characterized by (1) and (3). The other two are the tests based on the statistic $\widehat{LR}_T^d$ of JN and the DF-GLS statistic of ERS, each using the lag length

**TABLE 2.** Rejection frequencies of unit-root tests, constant mean case.

| DGP | $T = 300$ | | | | $T = 1,000$ | | | |
|---|---|---|---|---|---|---|---|---|
| $\phi_1, \phi_2, \phi_3$ | $\rho$ | $LR_T^{ADF}$ | $\widehat{LR}_T^d$ | DF-GLS | $\rho$ | $LR_T^{ADF}$ | $\widehat{LR}_T^d$ | DF-GLS |
| 0, 0, 0 | 1.020 | 0.001 | 0.835 | 0.001 | 1.006 | 0.001 | 0.775 | 0.001 |
| | 1.000 | 0.043 | 0.040 | 0.048 | 1.000 | 0.048 | 0.047 | 0.050 |
| | 0.980 | 0.338 | 0.321 | 0.367 | 0.994 | 0.384 | 0.376 | 0.391 |
| | 0.960 | 0.747 | 0.725 | 0.770 | 0.988 | 0.840 | 0.831 | 0.843 |
| | 0.940 | 0.911 | 0.899 | 0.916 | 0.982 | 0.980 | 0.977 | 0.980 |
| 0.2, 0.4, 0.6 | 1.020 | 0.001 | 0.811 | 0.002 | 1.006 | 0.001 | 0.805 | 0.002 |
| | 1.000 | 0.034 | 0.041 | 0.100 | 1.000 | 0.044 | 0.075 | 0.095 |
| | 0.980 | 0.268 | 0.265 | 0.448 | 0.994 | 0.351 | 0.430 | 0.493 |
| | 0.960 | 0.632 | 0.552 | 0.708 | 0.988 | 0.798 | 0.787 | 0.826 |
| | 0.940 | 0.848 | 0.781 | 0.877 | 0.982 | 0.967 | 0.958 | 0.970 |
| 0.4, 0.4, 0.4 | 1.020 | 0.001 | 0.818 | 0.002 | 1.006 | 0.001 | 0.801 | 0.001 |
| | 1.000 | 0.033 | 0.035 | 0.084 | 1.000 | 0.045 | 0.060 | 0.078 |
| | 0.980 | 0.267 | 0.235 | 0.393 | 0.994 | 0.356 | 0.370 | 0.427 |
| | 0.960 | 0.637 | 0.549 | 0.697 | 0.988 | 0.802 | 0.769 | 0.808 |
| | 0.940 | 0.854 | 0.794 | 0.883 | 0.982 | 0.969 | 0.958 | 0.969 |
| 0.6, 0.6, 0.6 | 1.020 | 0.001 | 0.769 | 0.003 | 1.006 | 0.001 | 0.824 | 0.003 |
| | 1.000 | 0.038 | 0.056 | 0.160 | 1.000 | 0.046 | 0.135 | 0.153 |
| | 0.980 | 0.265 | 0.329 | 0.550 | 0.994 | 0.359 | 0.535 | 0.569 |
| | 0.960 | 0.576 | 0.521 | 0.692 | 0.988 | 0.787 | 0.775 | 0.813 |
| | 0.940 | 0.774 | 0.663 | 0.805 | 0.982 | 0.959 | 0.919 | 0.938 |

*Notes:* Rejection frequencies for the QLR test ($LR_T^{ADF}$), the plug-in likelihood ratio test of JN ($\widehat{LR}_T^d$), and the DF-GLS test of ERS. Simulations are based on $10^5$ replications of the autoregressive DGP, allowing for a constant mean only in the regression model. The lag orders are chosen by minimization of the MAIC of Perron and Qu (2007) applied to the ADF model (3) for the $LR_T^{ADF}$ test, and to the DF-GLS regression for the other two tests.

chosen by the MAIC applied to the DF-GLS regression. In all cases, the maximum lag order is $p_{\max} = \lfloor 12(T/100)^{1/4} \rfloor$. Table 2 reports rejection frequencies of tests with nominal size 5% for the constant mean case, whereas the corresponding results for the linear trend case are reported in Table 3.

The $LR_T^{ADF}$ test exhibits excellent size and power properties across all cases considered in Tables 2 and 3. The other tests also have good power properties, but tend to exhibit size distortions, especially so in the model with the largest degree of persistence, namely when $\phi = (0.6, 0.6, 0.6)$.

Because the testing problem is one-sided, one might expect good tests to have low power against those alternatives that are on the "wrong" side of the null (as happens when testing hypotheses about the mean of a normal distribution, for instance). Interestingly, the $\widehat{LR}_T^d$ test seems to reject against $\rho > 1$. This

https://doi.org/10.1017/S0266466622000652 Published online by Cambridge University Press

**TABLE 3.** Rejection frequencies of unit-root tests, linear trend case.

| DGP | $T=300$ | | | | $T=1{,}000$ | | | |
|---|---|---|---|---|---|---|---|---|
| $\phi_1,\phi_2,\phi_3$ | $\rho$ | $LR_T^{ADF}$ | $\widehat{LR}_T^d$ | DF-GLS | $\rho$ | $LR_T^{ADF}$ | $\widehat{LR}_T^d$ | DF-GLS |
| 0, 0, 0 | 1.020 | 0.001 | 0.801 | 0.000 | 1.006 | 0.001 | 0.774 | 0.000 |
| | 1.000 | 0.039 | 0.034 | 0.029 | 1.000 | 0.045 | 0.042 | 0.032 |
| | 0.950 | 0.471 | 0.429 | 0.394 | 0.985 | 0.552 | 0.529 | 0.451 |
| | 0.900 | 0.861 | 0.832 | 0.820 | 0.970 | 0.972 | 0.963 | 0.939 |
| | 0.850 | 0.913 | 0.886 | 0.884 | 0.955 | 0.996 | 0.992 | 0.988 |
| 0.2, 0.4, 0.6 | 1.020 | 0.000 | 0.626 | 0.001 | 1.006 | 0.000 | 0.797 | 0.001 |
| | 1.000 | 0.028 | 0.021 | 0.110 | 1.000 | 0.039 | 0.096 | 0.109 |
| | 0.950 | 0.303 | 0.193 | 0.355 | 0.985 | 0.472 | 0.490 | 0.485 |
| | 0.900 | 0.712 | 0.541 | 0.652 | 0.970 | 0.943 | 0.913 | 0.893 |
| | 0.850 | 0.851 | 0.761 | 0.811 | 0.955 | 0.993 | 0.986 | 0.980 |
| 0.4, 0.4, 0.4 | 1.020 | 0.000 | 0.664 | 0.001 | 1.006 | 0.000 | 0.793 | 0.001 |
| | 1.000 | 0.025 | 0.020 | 0.089 | 1.000 | 0.040 | 0.075 | 0.084 |
| | 0.950 | 0.301 | 0.184 | 0.312 | 0.985 | 0.476 | 0.444 | 0.432 |
| | 0.900 | 0.726 | 0.575 | 0.668 | 0.970 | 0.945 | 0.914 | 0.892 |
| | 0.850 | 0.864 | 0.788 | 0.826 | 0.955 | 0.994 | 0.987 | 0.982 |
| 0.6, 0.6, 0.6 | 1.020 | 0.000 | 0.374 | 0.003 | 1.006 | 0.001 | 0.769 | 0.002 |
| | 1.000 | 0.028 | 0.003 | 0.233 | 1.000 | 0.043 | 0.163 | 0.203 |
| | 0.950 | 0.231 | 0.059 | 0.466 | 0.985 | 0.460 | 0.532 | 0.551 |
| | 0.900 | 0.512 | 0.288 | 0.538 | 0.970 | 0.918 | 0.833 | 0.826 |
| | 0.850 | 0.678 | 0.508 | 0.655 | 0.955 | 0.988 | 0.962 | 0.952 |

*Notes:* Rejection frequencies for the QLR test ($LR_T^{ADF}$), the plug-in likelihood ratio test of JN ($\widehat{LR}_T^d$), and the DF-GLS test of ERS. Simulations are based on $10^5$ replications of the autoregressive DGP, allowing for a constant mean and linear trend in the regression model. The lag orders are chosen by minimization of the MAIC of Perron and Qu (2007) applied to the ADF model (3) for the $LR_T^{ADF}$ test, and to the DF-GLS regression for the other two tests.

happens because the plug-in estimates of the autoregressive parameters capture the explosive root, which leaves only stationary roots and hence cause rejection. This phenomenon does not occur for the $LR_T^{ADF}$ test.

We next report some results for the notoriously difficult case of a moving average process with a negative root. The DGP is similar to that above, except we replace (2) with (6), for which we simulate from the moving average model $\psi(z) = 1+\psi z$ with $\psi \in \{-0.25, -0.50, -0.75\}$.

The simulation results for the moving average DGP are reported in Table 4. These show a clear size-power trade-off. For the smaller sample size and the largest negative root, the QLR test is somewhat oversized, but has much higher power than the DF-GLS test. The size distortion is reduced for the larger sample size.

https://doi.org/10.1017/S0266466622000652 Published online by Cambridge University Press

**TABLE 4.** Rejection frequencies of unit-root tests, linear trend case, moving average DGP.

| DGP | $T=300$ | | | | $T=1,000$ | | | |
|---|---|---|---|---|---|---|---|---|
| $\psi$ | $\rho$ | $LR_T^{ADF}$ | $\widehat{LR}_T$ | DF-GLS | $\rho$ | $LR_T^{ADF}$ | $\widehat{LR}_T$ | DF-GLS |
| −0.25 | 1.030 | 0.000 | 0.954 | 0.000 | 1.010 | 0.000 | 0.997 | 0.797 |
| | 1.000 | 0.047 | 0.040 | 0.035 | 1.000 | 0.049 | 0.045 | 0.035 |
| | 0.940 | 0.592 | 0.536 | 0.503 | 0.985 | 0.562 | 0.534 | 0.459 |
| | 0.920 | 0.754 | 0.700 | 0.670 | 0.980 | 0.791 | 0.765 | 0.693 |
| | 0.860 | 0.896 | 0.852 | 0.828 | 0.970 | 0.966 | 0.954 | 0.927 |
| −0.50 | 1.030 | 0.000 | 0.972 | 0.000 | 1.010 | 0.000 | 0.997 | 0.779 |
| | 1.000 | 0.060 | 0.046 | 0.039 | 1.000 | 0.058 | 0.052 | 0.039 |
| | 0.940 | 0.578 | 0.495 | 0.436 | 0.985 | 0.576 | 0.537 | 0.454 |
| | 0.920 | 0.724 | 0.641 | 0.572 | 0.980 | 0.790 | 0.755 | 0.671 |
| | 0.860 | 0.895 | 0.821 | 0.722 | 0.970 | 0.960 | 0.944 | 0.897 |
| −0.75 | 1.030 | 0.000 | 0.989 | 0.000 | 1.010 | 0.000 | 0.996 | 0.714 |
| | 1.000 | 0.092 | 0.063 | 0.045 | 1.000 | 0.076 | 0.064 | 0.044 |
| | 0.940 | 0.611 | 0.475 | 0.325 | 0.985 | 0.602 | 0.541 | 0.403 |
| | 0.920 | 0.763 | 0.623 | 0.419 | 0.980 | 0.790 | 0.735 | 0.564 |
| | 0.860 | 0.952 | 0.872 | 0.582 | 0.970 | 0.951 | 0.923 | 0.756 |

*Notes:* Rejection frequencies for the QLR test ($LR_T^{ADF}$), the plug-in likelihood ratio test of JN ($\widehat{LR}_T^d$), and the DF-GLS test of ERS. Simulations are based on $10^5$ replications of the moving average DGP, allowing for a constant mean and linear trend in the regression model. The lag orders are chosen by minimization of the MAIC of Perron and Qu (2007) applied to the ADF model (3) for the $LR_T^{ADF}$ test, and to the DF-GLS regression for the other two tests.

Results for other values of the autoregressive parameter $\phi$ and the moving average parameter $\psi$ are qualitatively similar and are omitted to conserve space. Overall, the simulation results are consistent with the theory developed in this paper, suggesting in particular that the test based on $LR_T^{ADF}$ is competitive with (if not superior to) its natural rivals also in samples of moderate size.

## 5. DIFFERENT CLASSES OF ALTERNATIVES

The models considered so far all have the feature that, under local departures from the unit-root hypothesis, the weak limit of the process $T^{1/2}u_{\lfloor T\cdot \rfloor}$ is the Ornstein–Uhlenbeck process (5). This section briefly explores the properties of QLR tests in two distinct types of models giving rise to different weak limits on the part of $T^{1/2}u_{\lfloor T\cdot \rfloor}$. More specifically, Section 5.1 allows for a nonnegligible initial condition in the model for $\{u_t\}$, whereas Section 5.2 is concerned with functional local-to-unity models.

To conserve space and focus on the main issues, we abstract from the presence of short-run dynamics and consider models for $\{u_t\}$ that can be interpreted as

https://doi.org/10.1017/S0266466622000652 Published online by Cambridge University Press

generalizations of the AR(1) version of the model (2). In other words, our point of departure in both subsections of this section is the model characterized by (1) and (2) with $\gamma(z) = 1$ and $u_0 = o_p(T^{1/2})$. We would expect, though, that very similar results could be obtained for the model with $p > 1$ and for the model characterized by (3) instead of (2).

### 5.1. (Possibly) Nonnegligible Initial Condition

As noted by Elliott (1999), if $u_0$ is drawn from the stationary distribution of the model (2) under local departures from the unit-root hypothesis, then the initial condition is nonnegligible in the sense that $T^{-1/2}u_0 = O_p(1)$, but $T^{-1/2}u_0 \neq o_p(1)$. Following Müller and Elliott (2003), an interesting way of accommodating a (possibly) nonnegligible initial condition is to model $u_0$ as a (possibly diverging) parameter. Doing so, the initial condition acts as an unidentified nuisance parameter under the null hypothesis of a unit root. Even in otherwise standard testing problems, the presence of nuisance parameters that are unidentified under the null renders these testing problems nonstandard (e.g., Andrews and Ploberger, 1994) and affects the optimality properties of likelihood ratio tests (e.g., Andrews and Ploberger, 1995).

As shown by Müller and Elliott (2003), similar phenomena occur in a unit-root testing context. We have no new results regarding general optimality theory for testing problems of this sort. Instead, the purpose of this subsection is to document the consequences of treating $u_0$ as an unknown/unrestricted nuisance parameter when developing QLR tests associated with the problem of testing $H_0$ versus $H_1$ in the model characterized by (1) and (2). As already indicated, we simplify the exposition by setting $\gamma(z) = 1$ in (2).

Setting $\gamma(z) = 1$ and letting $\xi$ denote the value of $u_0$, the Gaussian quasi-log-likelihood function corresponding to the model given by (1) and (2) can be expressed, up to an additive constant, as

$$L_T^{IC}(\rho,\beta,\sigma^2,\xi) = -\frac{T}{2}\log\sigma^2 - \frac{1}{2\sigma^2}(Y_\rho^{IC} - D_\rho^{IC}\beta)'(Y_\rho^{IC} - D_\rho^{IC}\beta) - \frac{1}{2\sigma^2}(y_1 - \beta' d_1 - \rho\xi)^2,$$

where $Y_\rho^{IC}$ and $D_\rho^{IC}$ are defined as the matrices with row $t = 1,\ldots,T-1$ given by $y_{t+1} - \rho y_t$ and $d_{t+1} - \rho d_t$, respectively. Under $H_0$ (i.e., when $\rho = 1$), the first column of $D_\rho^{IC}$ is zero and $\xi$ is not separately identified, although the sum of $\xi$ and the first element of $\beta$ is identified. More importantly (for the present purposes at least), for any $\rho \neq 0$, it is remarkably straightforward to eliminate $\xi$ by profiling it out. Thus, setting $\xi = \rho^{-1}(y_1 - \beta' d_1)$, the last term in $L_T^{IC}(\rho,\beta,\xi,\sigma^2)$ drops out and we obtain

$$\max_\xi L_T^{IC}(\rho,\beta,\sigma^2,\xi) = -\frac{T}{2}\log\sigma^2 - \frac{1}{2\sigma^2}(Y_\rho^{IC} - D_\rho^{IC}\beta)'(Y_\rho^{IC} - D_\rho^{IC}\beta).$$

https://doi.org/10.1017/S0266466622000652 Published online by Cambridge University Press

We therefore find that the QLR test statistic associated with the problem of testing $H_0$ versus $H_1$ satisfies

$$LR_T^{IC} = \max_{\rho \leq 1, \beta, \xi, \sigma^2 > 0} L_T^{IC}(\rho, \beta, \sigma^2, \xi) - \max_{\beta, \xi, \sigma^2 > 0} L_T^{IC}(1, \beta, \sigma^2, \xi)$$
$$= \max_{\rho \leq 1} \mathcal{L}_T^{IC}(\rho) - \mathcal{L}_T^{IC}(1),$$

where, up to an additive constant,

$$\mathcal{L}_T^{IC}(\rho) = -\frac{T}{2} \log \left(Y'Y - Y'D(D'D)^{-}D'Y\right)\Big|_{Y=Y_\rho^{IC}, D=D_\rho^{IC}}$$

is the profile quasi-log-likelihood obtained by maximizing $\max_\xi L_T^{IC}(\rho, \beta, \sigma^2, \xi)$ with respect to $(\beta, \sigma^2)$. Note that $\mathcal{L}_T^{IC}(\rho)$ involves the Moore–Penrose inverse, denoted $(\cdot)^-$, because $D'D\big|_{D=D_\rho^{IC}}$ is singular when $\rho = 1$.

To interpret $LR_T^{IC}$, we rewrite it as

$$LR_T^{IC} = \frac{T}{T-1} LR_T^{CL},$$

where

$$LR_T^{CL} = \max_{\rho \leq 1, \beta, \sigma^2 > 0} L_T^{CL}(\rho, \beta, \sigma^2) - \max_{\beta, \sigma^2 > 0} L_T^{CL}(1, \beta, \sigma^2)$$

is the QLR test statistic based on

$$L_T^{CL}(\rho, \beta, \sigma^2) = -\frac{T-1}{2} \log \sigma^2 - \frac{1}{2\sigma^2}(Y_\rho^{IC} - D_\rho^{IC}\beta)'(Y_\rho^{IC} - D_\rho^{IC}\beta).$$

Here, $L_T^{CL}(\rho, \beta, \sigma^2)$ is the conditional (on $y_1$) quasi-likelihood function corresponding to the model given by (1) and (2). The latter is precisely the quasi-likelihood upon which the Dickey and Fuller (1979, 1981) tests are based. Thus, apart from the factor $T/(T-1)$, the QLR statistic $LR_T^{IC}$ is numerically identical to the conditional QLR statistic $LR_T^{CL}$, and consequently $LR_T^{IC}$ is asymptotically equivalent to the one-sided Dickey–Fuller $t$-statistic.

In other words, treating the initial condition as a nuisance parameter and profiling it out of the likelihood in the ERS-type model results in the same QLR test as when conditioning on the first observation as in the work of Dickey and Fuller (1979, 1981). Although expected in hindsight, this is an interesting and, to the best of our knowledge, new insight.

The advantages and disadvantages of the Dickey–Fuller $t$-statistic are well understood in models with and without a nonnegligible initial condition. As a practical matter, Harvey, Leybourne, and Taylor (2009) recommend combining a test of DF-GLS type with the Dickey–Fuller $t$-test (using a union of rejections decision rule) when there is uncertainty about whether the initial condition is asymptotically negligible or not. A (purely) likelihood-based version of their proposal could employ a union of rejections decision rule based on $LR_T^{ADF}$ and a version of $LR_T^{IC}$ (or $LR_T^{CL}$) adapted to a model with $\gamma(z) \neq 1$.

https://doi.org/10.1017/S0266466622000652 Published online by Cambridge University Press

### 5.2. Functional Local-to-Unity Models

Setting $\gamma(z) = 1$ and assuming that $c = T(\rho - 1)$ is held fixed, the model (2) can be written as

$$\Delta u_t = \frac{c}{T} u_{t-1} + \varepsilon_t.$$

As an interesting generalization of this model, Bykhovskaya and Phillips (2020) proposed the functional local-to-unity model

$$\Delta u_t = \frac{C(t/T)}{T} u_{t-1} + \varepsilon_t, \tag{7}$$

where $C(\cdot)$ is some (possibly) nonconstant function (see also Bykhovskaya and Phillips, 2018). Letting $\mathcal{C}$ denote a set of functions containing the zero function, we assume that the maintained hypothesis is of the form $C(\cdot) \in \mathcal{C}$, in which case the unit-root testing problem is the problem of testing

$$H_0^{FLU} : C(\cdot) = 0 \qquad \text{versus} \qquad H_1^{FLU} : C(\cdot) \in \mathcal{C} \backslash \{0\}.$$

Specializing to the case where $H_1^{FLU}$ is simple, Bykhovskaya and Phillips (2020) characterized the large sample properties of QLR tests in the case where $\beta$ in (1) is known (and normalized to zero), $\varepsilon_t \sim i.i.d.\ \mathcal{N}(0, \sigma^2)$ with $\sigma^2$ known (and normalized to one) in (7), and the initial condition is $u_0 = 0$. By the Neyman–Pearson lemma, these tests are point optimal and the power of the tests can therefore be used to obtain Gaussian power envelopes. In this subsection, our goal is to describe (some of) the consequences of allowing $H_1^{FLU}$ to be composite. To facilitate comparison with our earlier results, we once again treat $\beta$ and $\sigma^2$ as unknown nuisance parameters.

The Gaussian quasi-log-likelihood function corresponding to the model given by (1) and (7) with initial condition $u_0 = 0$ can be expressed, up to an additive constant, as

$$L_T^{FLU}(\bar{C}, \beta, \sigma^2) = -\frac{T}{2} \log \sigma^2 - \frac{1}{2\sigma^2} (Y_{\bar{C}}^{FLU} - D_{\bar{C}}^{FLU} \beta)' (Y_{\bar{C}}^{FLU} - D_{\bar{C}}^{FLU} \beta),$$

where, setting $y_0 = 0$ and $d_0 = 0$, $Y_{\bar{C}}^{FLU}$ and $D_{\bar{C}}^{FLU}$ are defined as the matrices with row $t = 1, \ldots, T$ given by $\Delta y_t - T^{-1} \bar{C}(t/T) y_{t-1}$ and $\Delta d_t - T^{-1} \bar{C}(t/T) d_{t-1}$, respectively. The corresponding QLR test statistic associated with the problem of testing $H_0^{FLU}$ versus $H_1^{FLU}$ is therefore given by

$$\begin{aligned} LR_T^{FLU} &= \max_{\bar{C}(\cdot) \in \mathcal{C}, \beta, \sigma^2 > 0} L_T^{FLU}(\bar{C}, \beta, \sigma^2) - \max_{\beta, \sigma^2 > 0} L_T^{FLU}(0, \beta, \sigma^2) \\ &= \max_{\bar{C}(\cdot) \in \mathcal{C}} \mathcal{L}_T^{FLU}(\bar{C}) - \mathcal{L}_T^{FLU}(0), \end{aligned}$$

https://doi.org/10.1017/S0266466622000652 Published online by Cambridge University Press

where, up to an additive constant,

$$\mathcal{L}_T^{FLU}(\bar{C}) = -\frac{T}{2}\log\left(Y'Y - Y'D(D'D)^{-1}D'Y\right)\Big|_{Y=Y_{\bar{C}}^{FLU},D=D_{\bar{C}}^{FLU}}$$

is the profile quasi-log-likelihood obtained by maximizing $L_T^{FLU}(\bar{C},\beta,\sigma^2)$ with respect to $(\beta,\sigma^2)$.

In important respects, the distributional properties of $LR_T^{FLU}$ are similar to those obtained for $LR_T^{ADF}$ in Theorem 1. To state the results, let

$$W_C(r) = \int_0^r \exp\left(\int_s^r C(\tau)d\tau\right)dW(s),$$

where $W$ is a standard Wiener process. If $d_t = 1$, then it follows as in Lemma 2 of Bykhovskaya and Phillips (2020) that, under mild conditions on $C(\cdot)$ and $\bar{C}(\cdot)$, we have

$$\mathcal{L}_T^{FLU}(\bar{C}) - \mathcal{L}_T^{FLU}(0) \to_d \Lambda_C^{FLU}(\bar{C}),$$

where

$$\Lambda_C^{FLU}(\bar{C}) = \int_0^1 \bar{C}(r)W_C(r)dW_C(r) - \frac{1}{2}\int_0^1 \bar{C}(r)^2 W_C(r)^2 dr.$$

Under mild conditions on $C(\cdot)$ and $\mathcal{C}$, we therefore obtain the following generalization of the result reported in Theorem 1(a):

$$LR_T^{FLU} \to_d \max_{\bar{C}\in\mathcal{C}} \Lambda_C^{FLU}(\bar{C}).$$

Similarly, if $d_t = (1,t)'$ and under mild conditions on $C(\cdot)$ and $\mathcal{C}$, we obtain the following generalization of the result reported in Theorem 1(b):

$$LR_T^{FLU} \to_d \max_{\bar{C}\in\mathcal{C}} \Lambda_C^{FLU,\tau}(\bar{C}),$$

where

$$\begin{aligned}\Lambda_C^{FLU,\tau}(\bar{C}) = \Lambda_C^{FLU}(\bar{C}) \\ &+ \frac{1}{2}\frac{\left(W_C(1) - \int_0^1 \bar{C}(r)(rdW_C(r) + W_C(r)dr) + \int_0^1 \bar{C}(r)^2 rW_C(r)dr\right)^2}{1 - 2\int_0^1 \bar{C}(r)rdr + \int_0^1 \bar{C}(r)^2 r^2 dr} \\ &- \frac{1}{2}W_C(1)^2\end{aligned}$$

is the pointwise (in $\bar{C}(\cdot)$) weak limit of $\mathcal{L}_T^{FLU}(\bar{C}) - \mathcal{L}_T^{FLU}(0)$.

As observed by JN, the fact that $\Lambda_c(\bar{c})$ is quadratic in $\bar{c}$ implies that the functional $\max_{\bar{c}\le 0}\Lambda_c(\bar{c})$ in Theorem 1(a) admits the closed-form representation

https://doi.org/10.1017/S0266466622000652 Published online by Cambridge University Press

$$\max_{\bar{c}\leq 0} \Lambda_c(\bar{c}) = \frac{1}{2}\frac{\min\left\{\int_0^1 W_c(r)dW_c(r),0\right\}^2}{\int_0^1 W_c(r)^2dr}.$$

A similar representation is available for $\max_{\bar{C}\in\mathcal{C}} \Lambda_C^{FLU}(\bar{C})$ in many cases. To be specific, suppose that $\mathcal{C}$ is a cone of the form

$$\mathcal{C} = \left\{\bar{c}\bar{C}_\lambda(\cdot) : \bar{c}\leq 0, \lambda\in\Lambda\right\}, \tag{8}$$

where, for each $\lambda$ in some set $\Lambda$, $\bar{C}_\lambda(\cdot)$ is a known function. In this case, the parameter of interest is $\bar{c}$, whereas $\lambda$ is a nuisance parameter that is unidentified under the null, and we have

$$\max_{\bar{C}\in\mathcal{C}} \Lambda_C^{FLU}(\bar{C}) = \max_{\lambda\in\Lambda}\frac{1}{2}\frac{\min\left\{\int_0^1 \bar{C}_\lambda(r)W_C(r)dW_C(r),0\right\}^2}{\int_0^1 \bar{C}_\lambda(r)^2W_C(r)^2dr}.$$

In particular, if $\Lambda = \{1/2\}$ (implying in particular that there are no unidentified nuisance parameters under $H_0^{FLU}$), then

$$\max_{\bar{C}\in\mathcal{C}} \Lambda_C^{FLU}(\bar{C}) = \frac{1}{2}\frac{\min\left\{\int_0^1 \bar{C}_{1/2}(r)W_C(r)dW_C(r),0\right\}^2}{\int_0^1 \bar{C}_{1/2}(r)^2W_C(r)^2dr}.$$

On the other hand, $\max_{\bar{C}\in\mathcal{C}} \Lambda_C^{FLU,\tau}(\bar{C})$ shares with $\max_{\bar{c}\leq 0} \Lambda_c^{\tau}(\bar{c})$ in Theorem 1(b) the property that it does not seem to admit a closed-form representation.

The list of classes $\mathcal{C}$ satisfying (8) is long, but for specificity, we mention some prominent examples here. A particularly simple class is the one where $\Lambda = \{1/2\}$ and $\bar{C}_{1/2}(\cdot)$ is given by the triangular function,

$$\bar{C}_{1/2}(\tau) = \min\{\tau, 1-\tau\} = \lambda\min\left\{\frac{\tau}{\lambda},\frac{1-\tau}{1-\lambda}\right\}\Bigg|_{\lambda=1/2}.$$

This class is motivated by Section 2.2.2 of Bykhovskaya and Phillips (2020), wherein the function $\bar{c}\bar{C}_{1/2}(\cdot)$ is denoted $c^*_{\bar{c}/2}(\cdot)$. Similarly, the following classes exhibiting a (possibly) nontrivial dependence on $\lambda$, are inspired by Section 2.2.1 of Bykhovskaya and Phillips (2020):

$$\bar{C}_\lambda(\tau) = \mathbb{I}(\tau<\lambda), \qquad \Lambda\subset(0,1),$$
$$\bar{C}_\lambda(\tau) = \mathbb{I}(\tau>\lambda), \qquad \Lambda\subset(0,1),$$
$$\bar{C}_\lambda(\tau) = \mathbb{I}(|\tau-1/2|<\lambda), \qquad \Lambda\subset(0,1/2),$$
$$\bar{C}_\lambda(\tau) = \mathbb{I}(|\tau-1/2|>\lambda), \qquad \Lambda\subset(0,1/2),$$

where $\mathbb{I}(\cdot)$ denotes the indicator function.

It would be of interest to explore the power properties of $LR_T^{FLU}$ for various choices of $\mathcal{C}$. Among other things, it may be useful to isolate a class $\mathcal{C}$ for which $LR_T^{FLU}$ is nearly efficient in the sense that its local asymptotic power function is indistinguishable from the Gaussian power envelope. For instance, it seems natural

**TABLE 5.** Rejection frequencies of unit-root tests, functional case.

| $\rho$ | $LR_T^{ADF}$ | $\widehat{LR}_T^d$ | DF-GLS | $LR_T^{FLU}$ | | |
|---|---|---|---|---|---|---|
| | | | | $\lambda = 0.25$ | $\lambda = 0.5$ | $\lambda = 0.75$ |
| Panel A: constant mean case | | | | | | |
| 1.000 | 0.048 | 0.047 | 0.049 | 0.047 | 0.048 | 0.050 |
| 0.993 | 0.472 | 0.462 | 0.479 | 0.352 | 0.366 | 0.374 |
| 0.990 | 0.721 | 0.710 | 0.726 | 0.547 | 0.557 | 0.565 |
| 0.985 | 0.940 | 0.934 | 0.940 | 0.801 | 0.808 | 0.807 |
| $1-0.03\bar{C}_{1/2}(\cdot)$ | 0.737 | 0.728 | 0.735 | 0.876 | 0.947 | 0.828 |
| Panel B: linear trend case | | | | | | |
| 1.000 | 0.045 | 0.042 | 0.031 | 0.035 | 0.044 | 0.054 |
| 0.990 | 0.277 | 0.263 | 0.212 | 0.179 | 0.207 | 0.233 |
| 0.980 | 0.795 | 0.773 | 0.700 | 0.547 | 0.588 | 0.616 |
| 0.970 | 0.972 | 0.963 | 0.940 | 0.834 | 0.858 | 0.869 |
| $1-0.06\bar{C}_{1/2}(\cdot)$ | 0.763 | 0.744 | 0.641 | 0.885 | 0.976 | 0.906 |

*Notes:* Rejection frequencies for the QLR test ($LR_T^{ADF}$), the plug-in likelihood ratio test of JN ($\widehat{LR}_T^d$), the DF-GLS test of ERS, and the QLR test with triangular functional alternatives ($LR_T^{FLU}$ with $\lambda \in \{0.25, 0.5, 0.75\}$). Simulations are based on $10^5$ replications of the autoregressive DGP with $\gamma(z) = 1$ and $T = 1,000$, allowing for either a constant mean only (Panel A) or a linear trend (Panel B) in the regression model. The lag orders are chosen by the minimization of the MAIC of Perron and Qu (2007) applied to the relevant model.

to ask whether $LR_T^{FLU}$ is nearly efficient when $\mathcal{C}$ is of the form (8) with $\Lambda = \{1/2\}$ and $\bar{C}_{1/2}(\cdot)$ equal to the triangular function or some other plausible alternative to the conventional constant function. In addition, and perhaps even more so, it would be interesting to know whether it is possible to isolate a class $\mathcal{C}$ for which $LR_T^{FLU}$ has good power properties even when $\mathcal{C}$ is misspecified. Bykhovskaya and Phillips (2020) argue convincingly that this property fails when $\mathcal{C}$ consists only of constant functions and present evidence to suggest that setting $\Lambda = \{1/2\}$ and letting $\bar{C}_{1/2}(\cdot)$ be given by the triangular function might be an attractive alternative to the more conventional approach of (implicitly) letting $\mathcal{C}$ consist only of constant functions.

Although it is beyond the scope of this paper to do so, we hope that future work will answer these questions and more generally shed additional light on the properties of functional local-to-unity models. To motivate such analysis, we report in Table 5 rejection frequencies for the $LR_T^{ADF}$ test, the $\widehat{LR}_T^d$ test of JN, the DF-GLS test, and the QLR test with triangular functional alternatives, i.e., $LR_T^{FLU}$ with $\lambda \in \{0.25, 0.5, 0.75\}$. The results in Table 5 show that the $LR_T^{FLU}$ tests suffer some power loss relative to the other tests under the "usual" alternatives with $\rho < 1$ being constant. On the other hand, the $LR_T^{FLU}$ tests are more powerful against the functional alternative. Moreover, the $LR_T^{FLU}$ test with correctly specified $\lambda = 0.5$ is substantially more powerful than the $LR_T^{FLU}$ tests with $\lambda = 0.25$ and $\lambda = 0.75$.

https://doi.org/10.1017/S0266466622000652 Published online by Cambridge University Press

## 6. CONCLUDING REMARKS

This paper has developed and analyzed QLR test statistics in an autoregressive model of arbitrary order, whose deterministic components and short-run dynamics are governed by unknown nuisance parameters. Previous work, notably that of ERS and JN, has developed tests that can be interpreted as "plug-in" versions of QLR test statistics, developed under the counterfactual assumption that nuisance parameters governing either deterministic components or short-run dynamics are known. In particular, our work generalizes that of JN by allowing the nuisance parameters that are "profiled out" to include those of a finite-order autoregressive process governing short-run dynamics. Our main theoretical result shows that this generalization can be achieved without sacrificing analytical tractability or statistical efficiency. In addition, the resulting test is attractive from a practical point of view, being simple to compute and enjoying good properties in a simulation experiment. We have also considered extensions of the finite-order autoregressive model to sieve-type approximations and to different classes of alternatives.

Although doing so is beyond the scope of this paper, it would be of both theoretical and practical interest to explore whether the generalizations of JN developed in this paper could be extended to other unit-root-type models. For example, an interesting model related to the functional local-to-unity model considered in Section 5.2 is the hybrid stochastic local unit-root model of Lieberman and Phillips (2020), which leads to nonlinear diffusions that match certain financial models with high kurtosis (see also Lieberman and Phillips (2014, 2017)). It would be of interest to develop and study QLR tests of the unit-root hypothesis also in models of that type. Furthermore, it seems likely that the methods could be developed to cover seasonally integrated models as in Jansson and Nielsen (2011) or cointegrated vector autoregressive models as in Boswijk, Jansson, and Nielsen (2015). It would also be of interest to develop and analyze QLR tests for unit roots in more complicated settings such as panel data models. Important progress on understanding optimal unit-root testing in such models has been made by, among others, Moon, Perron, and Phillips (2007, 2014) and Becheri, Drost, and van den Akker (2015), but to the best of our knowledge, it is still an open question whether optimality can be achieved by tests admitting a QLR interpretation.

## APPENDIX. Proofs of Main Results

### A.1. Proof of Theorem 1

Because $\mathcal{L}_T^{ADF}(\cdot)$ is invariant under transformations of the form $y_t \to y_t + b'd_t$, we can assume without loss of generality that $\beta = 0$. Furthermore, the proof of part (a) is a special case of the proof of part (b), so we only give the proof of part (b).

In what follows, we employ a local (around $\pi = 0$, $\beta = 0$, and $\eta = \gamma$) reparameterization of the form $\pi = \pi_T(\bar{c}) = \gamma(1)\bar{c}/T$, $\beta = \beta_T(\bar{b}) = \gamma(1)^{-1}\bar{b}\,\mathrm{diag}(1, T^{-1/2})$, and $\eta = \eta_T(\bar{h}) = \gamma + \bar{h}T^{-1/2}$, where $\bar{c} \in \mathbb{R}$, $\bar{b} \in \mathbb{R}^2$, and $\bar{h} \in \mathbb{R}^p$. To reiterate, the model of interest is that

https://doi.org/10.1017/S0266466622000652 Published online by Cambridge University Press

given by (1) and (2) with $\gamma$ and $c = T(\rho - 1)$ kept fixed, but (solely) for the purpose of analyzing $\mathcal{L}_T^{ADF}$, it is natural to work with (and reparameterize) $\eta$ and $\pi$.

For any $(\pi, \beta)$, define $V_{\pi,\beta}^{\gamma}$ as the vector with row $t = 1, \ldots, T$ given by $(\gamma(L)(1-L) - \pi L)(y_t - \beta' d_t)$. Moreover, define

$$\tilde{\sigma}_T^2 = \frac{1}{T}\left(V'V - V'Z(Z'Z)^{-1}Z'V\right)\Big|_{V=V_{0,0}^{\gamma}, Z=Z_0} = T^{-1}\sum_{t=1}^{T}\varepsilon_t^2 + o_p(1) \to_p \sigma^2,$$

and for any $(\bar{c}, \bar{b})$, let

$$\lambda_T^{ADF}(\bar{c},\bar{b}) = \lambda_T^{DF}(\bar{c},\bar{b}) + \frac{1}{2\tilde{\sigma}_T^2}\left(V'Z(Z'Z)^{-1}Z'V\right)\Big|_{V=V_{\pi_T(\bar{c}),\beta_T(\bar{b})}^{\gamma}, Z=Z_{\beta_T(\bar{b})}}$$
$$- \frac{1}{2\tilde{\sigma}_T^2}\left(V'Z(Z'Z)^{-1}Z'V\right)\Big|_{V=V_{0,0}^{\gamma}, Z=Z_0},$$

where

$$\lambda_T^{DF}(\bar{c},\bar{b}) = \frac{1}{2\tilde{\sigma}_T^2}\, V'V\big|_{V=V_{0,0}^{\gamma}} - \frac{1}{2\tilde{\sigma}_T^2}\, V'V\big|_{V=V_{\pi_T(\bar{c}),\beta_T(\bar{b})}^{\gamma}}.$$

Because

$$\mathcal{L}_T^{ADF}(\pi_T(\bar{c}),\beta_T(\bar{b})) - \mathcal{L}_T^{ADF}(0,0) = G_T(\lambda_T^{ADF}(\bar{c},\bar{b})),$$

where

$$G_T(x) = -\frac{T}{2}\log\left(1 - \frac{2}{T}x\right), \qquad x < \frac{T}{2},$$

is monotonically increasing in $x$, the statistic $LR_T^{ADF}$ admits the representation

$$LR_T^{ADF} = G_T\left(\max_{\bar{c}\le 0,\bar{b}} \lambda_T^{ADF}(\bar{c},\bar{b})\right) - G_T\left(\max_{\bar{b}} \lambda_T^{ADF}(0,\bar{b})\right).$$

Suppose

$$\left(\max_{\bar{c}\le 0,\bar{b}} \lambda_T^{ADF}(\bar{c},\bar{b}), \max_{\bar{b}} \lambda_T^{ADF}(0,\bar{b})\right) \to_d \left(\max_{\bar{c}\le 0,\bar{b}} \Lambda_c^{ADF}(\bar{c},\bar{b}), \max_{\bar{b}} \Lambda_c^{ADF}(0,\bar{b})\right), \qquad \textbf{(A.1)}$$

where the process $\Lambda_c^{ADF}$ is of the form

$$\Lambda_c^{ADF}(\bar{c},\bar{b}) = \Lambda_c(\bar{c}) + \bar{b}'\begin{pmatrix} \mathcal{E} \\ (1-\bar{c})W_c(1) + \bar{c}^2\int_0^1 rW_c(r)dr \end{pmatrix} - \frac{1}{2}\bar{b}'\begin{pmatrix} \mathcal{K} & 0 \\ 0 & 1-\bar{c}+\bar{c}^2/3 \end{pmatrix}\bar{b},$$

with $\mathcal{K}$ a positive constant (possibly depending on $\gamma$) and $\mathcal{E}$ a random variable independent of $W_c$ (but possibly depending on $\gamma$). Then

$$LR_T^{ADF} = \max_{\bar{c}\le 0,\bar{b}} \lambda_T^{ADF}(\bar{c},\bar{b}) - \max_{\bar{b}} \lambda_T^{ADF}(0,\bar{b}) + o_p(1)$$
$$\to_d \max_{\bar{c}\le 0,\bar{b}} \Lambda_c^{ADF}(\bar{c},\bar{b}) - \max_{\bar{b}} \Lambda_c^{ADF}(0,\bar{b})$$

https://doi.org/10.1017/S0266466622000652 Published online by Cambridge University Press

$$= \max_{\bar{c}\le 0}\left(\Lambda_c(\bar{c}) + \frac{1}{2}\frac{\left((1-\bar{c})W_c(1)+\bar{c}^2\int_0^1 rW_c(r)dr\right)^2}{(1-\bar{c}+\bar{c}^2/3)} + \frac{1}{2}\frac{\mathcal{E}^2}{\mathcal{K}}\right) - \left(\frac{1}{2}W_c(1)^2 + \frac{1}{2}\frac{\mathcal{E}^2}{\mathcal{K}}\right)$$
$$= \max_{\bar{c}\le 0}\Lambda_c^\tau(\bar{c}),$$

where the first equality follows from the facts that (i) the left-hand side of (A.1) is $O_p(1)$ and (ii) $\lim_{T\to\infty}\sup_{|x|\le M}|G_T(x)-x| = 0$ for any $0\le M<\infty$. The proof can therefore be completed by verifying (A.1). We shall do so by showing that

$$(\hat{c}_T,\hat{b}_T) = \arg\max_{\bar{c}\le 0,\bar{b}}\lambda_T^{ADF}(\bar{c},\bar{b}) = O_p(1) \quad \text{and} \quad \tilde{b}_T = \arg\max_{\bar{b}}\lambda_T^{ADF}(0,\bar{b}) = O_p(1), \tag{A.2}$$

and that $\lambda_T^{ADF}$ converges to $\Lambda_c^{ADF}$ in the topology of uniform convergence on compacta.

We first show (A.2). To this end, let

$$e_t = \gamma(L)\Delta y_t, \quad Q_{Tt} = (q_{Tt}', r_{Tt}', s_{Tt}')', \quad \theta_T = (\theta_{q,T}', \theta_{r,T}', \theta_{s,T}')',$$

with

$$q_{Tt} = \begin{pmatrix} T^{-1}\gamma(1)y_{t-1} \\ \mathbb{I}(t=1) - \sum_{j=1}^p \gamma_j \mathbb{I}(t=j+1) \\ T^{-1/2} \\ T^{-3/2}(t-1)\end{pmatrix}, \quad r_{Tt} = \begin{pmatrix}\mathbb{I}(t=2)\\ \vdots \\ \mathbb{I}(t=p+1)\end{pmatrix}, \quad s_{Tt} = \begin{pmatrix} T^{-1/2}\Delta y_{t-1} \\ \vdots \\ T^{-1/2}\Delta y_{t-p}\end{pmatrix},$$

and

$$\theta_{q,T}(\bar{c},\bar{b},\bar{h}) = \begin{pmatrix} \bar{c} \\ \gamma(1)^{-1}(1+T^{-1}\bar{c}\gamma(1))\bar{b}_1 + T^{-1/2}\gamma(1)^{-1}\bar{b}_2\iota'\eta_T(\bar{h}) \\ \gamma(1)^{-1}(1-T^{-1/2}\iota'\bar{h})\bar{b}_2 - T^{-1/2}\bar{c}\bar{b}_1 \\ -\bar{c}\bar{b}_2 \end{pmatrix},$$
$$\theta_{r,T}(\bar{c},\bar{b},\bar{h}) = T^{-1}\bar{c}\bar{b}_1\gamma + T^{-1/2}\gamma(1)^{-1}\bar{b}_2\iota'\eta_T(\bar{h})\gamma - T^{-1/2}\gamma(1)^{-1}\bar{b}_1\bar{h} + T^{-1/2}\gamma(1)^{-1}\bar{b}_2\mathcal{U}\eta_T(\bar{h}),$$
$$\theta_{s,T}(\bar{c},\bar{b},\bar{h}) = \bar{h},$$

where $\mathbb{I}(\cdot)$ is the indicator function, $\iota$ is a $p$-vector of ones, and $\mathcal{U}$ is a $p\times p$ strictly upper triangular matrix with ones above the main diagonal.

The pair $(\hat{c}_T,\hat{b}_T)$ satisfies

$$(\hat{c}_T,\hat{b}_T,\hat{h}_T) = \arg\min_{\bar{c}\le 0,\bar{b},\bar{h}}\lambda_T(\bar{c},\bar{b},\bar{h}),$$

where

$$\lambda_T(\bar{c},\bar{b},\bar{h}) = \sum_{t=1}^T (e_t - Q_{Tt}'\theta_T(\bar{c},\bar{b},\bar{h}))^2 - \sum_{t=1}^T e_t^2 = -2G_T'\theta + \theta' H_T\theta\big|_{\theta=\theta_T(\bar{c},\bar{b},\bar{h})},$$

https://doi.org/10.1017/S0266466622000652 Published online by Cambridge University Press

with

$$G_T = \begin{pmatrix} G_{q,T} \\ G_{r,T} \\ G_{s,T} \end{pmatrix} = \sum_{t=1}^{T} \begin{pmatrix} q_{Tt} \\ r_{Tt} \\ s_{Tt} \end{pmatrix} e_t = \sum_{t=1}^{T} Q_{Tt} e_t$$

and

$$H_T = \begin{pmatrix} H_{qq,T} & H_{qr,T} & H_{qs,T} \\ H_{rq,T} & H_{rr,T} & H_{rs,T} \\ H_{sq,T} & H_{sr,T} & H_{ss,T} \end{pmatrix} = \sum_{t=1}^{T} \begin{pmatrix} q_{Tt} \\ r_{Tt} \\ s_{Tt} \end{pmatrix} \begin{pmatrix} q_{Tt} \\ r_{Tt} \\ s_{Tt} \end{pmatrix}' = \sum_{t=1}^{T} Q_{Tt} Q_{Tt}'.$$

Defining $\hat{\theta}_T = \theta_T(\hat{c}_T, \hat{b}_T, \hat{h}_T)$, we therefore have

$$\begin{aligned} 0 &\geq \lambda_T(\hat{c}_T, \hat{b}_T, \hat{h}_T) = -2G_T'\hat{\theta}_T + \hat{\theta}_T' H_T \hat{\theta}_T \\ &\geq -2\|G_T\|\|\hat{\theta}_T\| + \lambda_{\min}(H_T)\|\hat{\theta}_T\|^2 \end{aligned} \quad \textbf{(A.3)}$$

by the Rayleigh–Ritz theorem, where $\|\cdot\|$ denotes the Euclidean norm and where $\lambda_{\min}(\cdot)$ denotes the smallest eigenvalue of the argument. This rearranges straightforwardly as

$$\|\hat{\theta}_T\| \leq 2\|G_T\|/\lambda_{\min}(H_T).$$

It follows from standard results (e.g., Chan and Wei, 1987; Phillips, 1987) that $(G_T, H_T) \to_d (\mathcal{G}_c, \mathcal{H}_c)$ for some $(\mathcal{G}_c, \mathcal{H}_c)$ depending on $c$ (and $\gamma$) with $\mathcal{H}_c$ positive definite. In particular, $G_T = O_p(1)$ and $1/\lambda_{\min}(H_T) = O_p(1)$, and therefore $\hat{\theta}_T = O_p(1)$.

Consequently, $\hat{h}_T = O_p(1)$, $\hat{c}_T = O_p(1)$, and

$$\begin{pmatrix} (1+o_p(1))\hat{b}_{1,T} + \hat{b}_{2,T} o_p(1) \\ (1+o_p 1))\hat{b}_{2,T} - \hat{c}_T \hat{b}_{1,T} o(1) \\ -\hat{c}_T \hat{b}_{2,T} \end{pmatrix} = O_p(1),$$

implying in turn that also $\hat{b}_T = O_p(1)$. This proves the first statement of (A.2).

Similarly, the representation

$$(\tilde{b}_T, \tilde{h}_T) = \arg\min_{\bar{b}, \bar{h}} \lambda_T(0, \bar{b}, \bar{h})$$

can be used to show that $\tilde{h}_T = O_p(1)$ and

$$\begin{pmatrix} \tilde{b}_{1,T} + \tilde{b}_{2,T} o_p(1) \\ (1+o_p(1))\tilde{b}_{2,T} \end{pmatrix} = O_p(1),$$

which implies that $\tilde{b}_T = O_p(1)$. This proves the second statement of (A.2).

Next, we prove that $\lambda_T^{ADF}$ converges to $\Lambda_c^{ADF}$ in the topology of uniform convergence on compacta. For any compact set $K$, it can be shown that

$$\sup_{(\bar{c}, \bar{b}')' \in K} \left| \lambda_T^{ADF}(\bar{c}, \bar{b}) - \lambda_T^{DF}(\bar{c}, \bar{b}) \right| = O_p(T^{-1/2}) = o_p(1).$$

It therefore follows from Prohorov's theorem (e.g., Kallenberg, 2002, Thm. 16.5) that $\lambda_T^{ADF}$ converges to $\Lambda_c^{ADF}$ in the topology of uniform convergence on compacta if $\lambda_T^{DF}$ converges

https://doi.org/10.1017/S0266466622000652 Published online by Cambridge University Press

to $\Lambda_c^{ADF}$ in the sense of weak convergence of finite-dimensional projections and if the process $\left\{\lambda_T^{DF}(\bar{c},\bar{b}) : (\bar{c},\bar{b}')' \in K\right\}$ is tight for any compact set $K$.

For any fixed $(\bar{c},\bar{b})$, it follows from standard results (e.g., Chan and Wei, 1987; Phillips, 1987) that

$$\lambda_T^{DF}(\bar{c},\bar{b}) \to_d \Lambda_c^{ADF}(\bar{c},\bar{b}).$$

Moreover, the Cramér–Wold device can be used to show that $\lambda_T^{DF}$ converges to $\Lambda_c^{ADF}$ in the sense of weak convergence of finite-dimensional projections. Finally, letting $\mathcal{V}_T$, $\dot{\mathcal{V}}_T$, $\mathcal{D}_T$, and $\dot{\mathcal{D}}_T$ be matrices with row $t = 1,\ldots,T$ given by $\gamma(L)\Delta y_t$, $T^{-1}\gamma(1)y_{t-1}$, $\gamma(1)^{-1}\gamma(L)\Delta d_t' \operatorname{diag}(1,T^{-1/2})$, and $d_{t-1}' \operatorname{diag}(T^{-3/2},T^2)$, respectively, $\lambda_T^{DF}$ admits a representation of the form

$$\lambda_T^{DF}(\bar{c},\bar{b}) = F(\bar{c},\bar{b},\mathcal{S}_T),$$

where $F$ is continuous and where

$$\mathcal{S}_T = \left(\mathcal{V}_T'\dot{\mathcal{V}}_T, \dot{\mathcal{V}}_T'\dot{\mathcal{V}}_T, \mathcal{V}_T'\mathcal{D}_T, \mathcal{V}_T'\dot{\mathcal{D}}_T, \dot{\mathcal{V}}_T'\mathcal{D}_T, \dot{\mathcal{V}}_T'\dot{\mathcal{D}}_T, \mathcal{D}_T'\mathcal{D}_T, \mathcal{D}_T'\dot{\mathcal{D}}_T, \dot{\mathcal{D}}_T'\dot{\mathcal{D}}_T\right) = O_p(1)$$

by standard results (e.g., Chan and Wei, 1987; Phillips, 1987). Because $F$ is continuous, it follows from the Arzelà–Ascoli theorem (Dudley, 2002, Thm. 2.4.7) that for any compact sets $K$ and $\mathcal{K}$, the set $\{F(\cdot,\mathcal{S})|_K : \mathcal{S} \in \mathcal{K}\}$ is relatively compact (i.e., has compact closure), where $F(\cdot,\mathcal{S})|_K$ is the restriction of $F(\cdot,\mathcal{S})$ to $K$. As a consequence, the fact that $\mathcal{S}_T = O_p(1)$ implies that, for any compact set $K$, the process $\left\{\lambda_T^{ADF}(\bar{c},h) : (\bar{c},h')' \in K\right\} = \left\{F(\bar{c},h,\mathcal{S}_T) : (\bar{c},h')' \in K\right\}$ is tight.

## A.2. Proof of Theorem 2

Define $\bar{\gamma}(z) = 1 - \sum_{j=1}^{\infty} \bar{\gamma}_j z^j = \psi(z)^{-1}$, which exists under our conditions on $\psi(z)$. For any $p$, letting $\gamma = (\gamma_1,\ldots,\gamma_p)' = (\bar{\gamma}_1,\ldots,\bar{\gamma}_p)'$ and proceeding as in the proof of Theorem 1, it suffices to show that (A.2) holds and that $\lambda_T^{ADF}$ converges to $\Lambda_c^{ADF}$ in the topology of uniform convergence on compacta.

The fact that $\lambda_T^{ADF}$ converges to $\Lambda_c^{ADF}$ in the topology of uniform convergence on compacta can be shown by adapting the proof of Theorem 1 with the help of results and ideas from Chang and Park (2002) and Phillips and Solo (1992). Specifically, when $p = p_T \to \infty$, it holds that (i) for any compact set $K$,

$$\sup_{(\bar{c},\bar{b}')' \in K} \left|\lambda_T^{ADF}(\bar{c},\bar{b}) - \lambda_T^{DF}(\bar{c},\bar{b})\right| = O_p(T^{-1/2}p_T),$$

which is $o_p(1)$ because $p_T = o(T^{1/3})$, (ii) $\lambda_T^{DF}$ converges to $\Lambda_c^{ADF}$ in the sense of weak convergence of finite-dimensional projections, and (iii) $\mathcal{S}_T = O_p(1)$. To conserve space, we do not report the details of those derivations.

Relative to the proof of Theorem 1, the most difficult part of the proof of Theorem 2 is to show that (A.2) holds. Proceeding as in the proof of Theorem 1, and using in particular the argument in (A.3), we have that

$$\|\hat{\theta}_T\| \leq 2\|G_T\|/\lambda_{\min}(H_T) = O_p(p_T^{1/2}).$$

https://doi.org/10.1017/S0266466622000652 Published online by Cambridge University Press

Here, the equality uses the facts that, when $p = p_T \to \infty$,

$$\|G_{q,T}\| + p_T^{-1/2}\|G_{r,T}\| + p_T^{-1/2}\|G_{s,T}\| = O_p(1)$$

and

$$\lambda_{\max}(H_T) + 1/\lambda_{\min}(H_T) = O_p(1),$$

where $\lambda_{\max}(\cdot)$ denotes the largest eigenvalue of the argument, the first display follows from the Chebyshev inequality and the fact that the elements of $G_T$ have bounded second moments, and the second display uses $\sum_{j=1}^{\infty} \bar{\gamma}_j^2 < \infty$ and the fact (also noted by Berk, 1974) that

$$\lambda_{\max}(\Sigma) + 1/\lambda_{\min}(\Sigma) = O(1),$$

where

$$\Sigma = E(\breve{v}_t \breve{v}_t'), \qquad \breve{v}_t = \gamma(L)^{-1}(\varepsilon_t, \ldots, \varepsilon_{t-p+1})'.$$

In particular, this implies that $\hat{h}_T = O_p(p_T^{1/2})$, $\hat{c}_T = O_p(p_T^{1/2})$, and, using $p_T = o(T)$,

$$\begin{pmatrix} (1+o_p(1))\hat{b}_{1,T} + \hat{b}_{2,T} o_p(1) \\ (1+o_p(1))\hat{b}_{2,T} - \hat{c}_T \hat{b}_{1,T} O(T^{-1/2}) \\ -\hat{c}_T \hat{b}_{2,T} \end{pmatrix} = O_p(p_T^{1/2}),$$

implying in turn that also $\hat{b}_T = O_p(p_T^{1/2})$.

To sharpen these rates and prove that (A.2) holds, we first note that $\hat{h}_T = O_p(p_T^{1/2})$, $\hat{c}_T = O_p(p_T^{1/2})$, and $\hat{b}_T = O_p(p_T^{1/2})$ together with $p_T = o(T^{1/3})$ imply that

$$G_T'\hat{\theta}_T = G_{q,T}'\hat{\theta}_{q,T} + G_{s,T}'\hat{\theta}_{s,T} + o_p(1)$$

and

$$\hat{\theta}_T' H_T \hat{\theta}_T = \hat{\theta}_{q,T}' H_{qq,T} \hat{\theta}_{q,T} + \hat{\theta}_{s,T}' H_{ss,T} \hat{\theta}_{s,T} + o_p(1).$$

Consequently,

$$\begin{aligned} \min_{\bar{c} \le 0, \bar{b}, \bar{h}} \lambda_T(\bar{c}, \bar{b}, \bar{h}) &= -2G_T'\hat{\theta}_T + \hat{\theta}_T' H_T \hat{\theta}_T \\ &= -2G_{q,T}'\hat{\theta}_{q,T} + \hat{\theta}_{q,T}' H_{qq,T}\hat{\theta}_{q,T} - 2G_{s,T}'\hat{\theta}_{s,T} + \hat{\theta}_{s,T}' H_{ss,T}\hat{\theta}_{s,T} + o_p(1), \end{aligned}$$

where

$$-2G_{s,T}'\hat{\theta}_{s,T} + \hat{\theta}_{s,T}' H_{ss,T}\hat{\theta}_{s,T} = \lambda_T(0,0,\bar{h}) \ge \min_{\bar{h}} \lambda_T(0,0,\bar{h}).$$

Therefore,

$$\begin{aligned} 0 &\ge \min_{\bar{c} \le 0, \bar{b}, \bar{h}} \lambda_T(\bar{c}, \bar{b}, \bar{h}) - \min_{\bar{h}} \lambda_T(0,0,\bar{h}) \\ &\ge -2G_{q,T}'\hat{\theta}_{q,T} + \hat{\theta}_{q,T}' H_{qq,T}\hat{\theta}_{q,T} + o_p(1), \end{aligned}$$

which implies that, with probability converging to 1,

$$\|\hat{\theta}_{q,T}\| \le 2\|G_{q,T}\|/\lambda_{\min}(H_{qq,T}) + o_p(1)/\|\hat{\theta}_{q,T}\| \le 2\|G_{q,T}\|/\lambda_{\min}(H_{qq,T}) + 1 = O_p(1). \qquad \textbf{(A.4)}$$

Here, the first inequality follows by the Rayleigh–Ritz theorem as in (A.3), the second inequality follows by noting that the $o_p(1)$ term is smaller than one in absolute value with probability converging to 1 and then considering separately the cases $\|\hat{\theta}_{q,T}\| \le 1$ and $\|\hat{\theta}_{q,T}\| \ge 1$, and the equality follows because $G_{q,T} = O_p(1)$ and $1/\lambda_{\min}(H_{qq,T}) = O_p(1)$.

Using (A.4) and the fact that

$$\hat{\theta}_{q,T} = \begin{pmatrix} \hat{c}_T \\ \gamma(1)^{-1}\hat{b}_{1,T} \\ \gamma(1)^{-1}\hat{b}_{2,T} \\ -\hat{c}_T\hat{b}_{2,T} \end{pmatrix} + o_p(1),$$

we have

$$\left\| \begin{pmatrix} \hat{c}_T \\ \gamma(1)^{-1}\hat{b}_{1,T} \\ \gamma(1)^{-1}\hat{b}_{2,T} \\ -\hat{c}_T\hat{b}_{2,T} \end{pmatrix} \right\| \le \|\hat{\theta}_{q,T}\| + o_p(1) = O_p(1),$$

implying in particular that $\hat{c}_T = O_p(1)$ and $\hat{b}_T = O_p(1)$. By a very similar proof, it can be shown that $\tilde{b}_T = O_p(1)$.

## *REFERENCES*


Andrews, D.W.K. & W. Ploberger (1994) Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica* 62, 1383–1414.

Andrews, D.W.K. & W. Ploberger (1995) Admissibility of the likelihood ratio test when a nuisance parameter is present only under the alternative. *Annals of Statistics* 23, 1609–1629.

Becheri, I.G., F.C. Drost, & R. van den Akker (2015) Asymptotically UMP panel unit root tests—The effect of heterogeneity in the alternatives. *Econometric Theory* 31, 539–559.

Berk, K.N. (1974) Consistent autoregressive spectral estimates. *Annals of Statistics* 2, 489–502.

Boswijk, H.P., M. Jansson, & M.Ø. Nielsen (2015) Improved likelihood ratio tests for cointegration rank in the VAR model. *Journal of Econometrics* 184, 97–110.

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

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

Chan, N.H. & C.Z. Wei (1987) Asymptotic inference for nearly nonstationary AR(1) processes. *Annals of Statistics* 15, 1050–1063.

Chang, Y. & J.Y. Park (2002) On the asymptotics of ADF tests for unit roots. *Econometric Reviews* 21, 431–447.

Dickey, D.A. & W.A. Fuller (1979) Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association* 74, 427–431.

Dickey, D.A. & W.A. Fuller (1981) Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica* 49, 1057–1072.

Dudley, R.M. (2002) *Real Analysis and Probability*, *2nd Edition*. Cambridge University Press.

Elliott, G. (1999) Efficient tests for a unit root when the initial observation is drawn from its unconditional distribution. *International Economic Review* 40, 767–783.

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

Harvey, D.I., S.J. Leybourne, & A.M.R. Taylor (2009) Unit root testing in practice: Dealing with uncertainty over the trend and initial condition. *Econometric Theory* 25, 587–636.

Jansson, M. & M.Ø. Nielsen (2011) Nearly efficient likelihood ratio tests for seasonal unit roots. *Journal of Time Series Econometrics* 3(1), Article no. 5.

Jansson, M. & M.Ø. Nielsen (2012) Nearly efficient likelihood ratio tests of the unit root hypothesis. *Econometrica* 80, 2321–2332.

Kallenberg, O. (2002) *Foundations of Modern Probability*, *2nd Edition*. Springer.

Lieberman, O. & P.C.B. Phillips (2014) Norming rates and limit theory for some time-varying coefficient autoregressions. *Journal of Time Series Analysis* 35, 592–623.

Lieberman, O. & P.C.B. Phillips (2017) A multivariate stochastic unit root model with an application to derivative pricing. *Journal of Econometrics* 196, 99–110.

Lieberman, O. & P.C.B. Phillips (2020) Hybrid stochastic local unit roots. *Journal of Econometrics* 215, 257–285.

Moon, H.R., B. Perron, & P.C.B. Phillips (2007) Incidental trends and the power of panel unit root tests. *Journal of Econometrics* 141, 416–459.

Moon, H.R., B. Perron, & P.C.B. Phillips (2014) Point-optimal panel unit root tests with serially correlated errors. *The Econometrics Journal* 17, 338–372.

Müller, U.K. & G. Elliott (2003) Tests for unit roots and the initial condition. *Econometrica* 71, 1269–1286.

Ng, S. & P. Perron (2001) LAG length selection and the construction of unit root tests with good size and power. *Econometrica* 69, 1519–1554.

Perron, P. & Z. Qu (2007) A simple modification to improve the finite sample properties of Ng and Perron's unit root tests. *Economics Letters* 94, 12–19.

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

Phillips, P.C.B. & V. Solo (1992) Asymptotics for linear processes. *Annals of Statistics* 20, 971–1001.

Said, S.E. & D.A. Dickey (1984) Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika* 71, 599–607.