Queen's Economics Department Working Paper No. 1429

Nearly Efficient Likelihood Ratio Tests of a Unit Root in an Autoregressive Model of Arbitrary Order

Samuel Brien

Michael Jansson
UC Berkeley and CREATES

Morten Ørregaard Nielsen
Queen's University and CREATES

Department of Economics
Queen's University
94 University Avenue
Kingston, Ontario, Canada
K7L 3N6

5-2020

Nearly Efficient Likelihood Ratio Tests of a Unit Root in an Autoregressive Model of Arbitrary Order*

Samuel Brien
Queen's University
samuel.brien@queensu.ca

Michael Jansson†
UC Berkeley and CREATES
mjansson@econ.berkeley.edu

Morten Ørregaard Nielsen
Queen's University and CREATES
mon@econ.queensu.ca

May 7, 2020

Abstract

We study large-sample properties of likelihood ratio tests of the unit root hypothesis in an autoregressive model of arbitrary, finite order. Earlier research on this testing problem has developed likelihood ratio tests in the autoregressive model of order one, 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.

Keywords: Efficiency, likelihood ratio test, nuisance parameters, unit root hypothesis

JEL Codes: C12, C22

*We are grateful to seminar participants at Queen's University for comments and discussion, and to the Canada Research Chairs program, the Social Sciences and Humanities Research Council of Canada (SSHRC), and the Center for Research in Econometric Analysis of Time Series (CREATES, funded by the Danish National Research Foundation, DNRF78) for financial support.

†Corresponding author. Address: University of California, Berkeley, Department of Economics, 508-1 Evans Hall #3880, Berkeley, CA 94720. Email: mjansson@econ.berkeley.edu

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 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) demonstrate that (quasi-) likelihood ratio tests in the ERS model are also nearly efficient in the autoregressive model of order one. 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 (quasi-) likelihood ratio interpretation in the higher-order case. In fact, even after several decades of intense research into this testing problem, it would appear that a likelihood ratio 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 (quasi-) likelihood ratio 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 towards other non-standard 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 non-standard 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 present the results of a small simulation study of the finite-sample properties of the likelihood ratio test and compare with some existing tests. Section 4 offers some concluding remarks. Finally, the proof of our main result in given in the appendix.

2 Model and Quasi-Likelihood Ratio Test Statistic

Our goal is to develop unit root tests that are of quasi-likelihood ratio 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 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 $d_t = (1,t)'$, β is an unknown parameter, and where 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, $\gamma(L) = 1 - \gamma_1 L - \ldots - \gamma_p L^p$ a lag polynomial of order p satisfying $\min_{|z|\leq 1} |\gamma(z)| > 0$. When developing formal results, we will complete the specification of the model by assuming that $\max(|u_0|, \ldots, |u_{-p}|) = o_p(\sqrt{T})$ and that the ε_t form a conditionally homoskedastic martingale difference sequence with (unknown) variance σ^2 and $\sup_t E(|\varepsilon_t|^r |\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots) < C < \infty$ for some $r > 2$.

In the model characterized by (1) and (2), the unit root testing problem is the problem of testing

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

A quasi-likelihood ratio test statistic associated with this testing problem can be based on the Gaussian quasi-likelihood corresponding to the model with $u_0 = \ldots = u_{-p} = 0$. In terms of the parameter of interest ρ and the nuisance parameters β, $\gamma = (\gamma_1, \ldots, \gamma_p)'$, and σ^2, the Gaussian quasi-log likelihood function corresponding to the model given by (1) and (2) with initial conditions $u_0 = \ldots = u_{-p} = 0$ can be expressed, up to a constant, as

$$L_T\left(\rho, \beta, \gamma, \sigma^2\right) = -\frac{T}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\left(Y_{\rho,\gamma} - D_{\rho,\gamma}\beta\right)'\left(Y_{\rho,\gamma} - D_{\rho,\gamma}\beta\right),$$

where, setting $y_0 = \ldots = y_{-p} = 0$ and $d_0 = \ldots = d_{-p} = 0$, $Y_{\rho,\gamma}$ and $D_{\rho,\gamma}$ are matrices with row $t = 1, \ldots, T$ given by $(1-\rho L)\,\gamma(L)\,y_t$ and $(1-\rho L)\,\gamma(L)\,d_t'$, respectively. The corresponding quasi-likelihood ratio test statistic associated with the problem of testing H_0 versus H_1 is

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

where $\Gamma = \left\{\gamma : \min_{|z|\leq 1} |\gamma(z)| > 0\right\}$.

Up to a constant, the profile quasi-log likelihood obtained by maximizing $L_T(\rho, \beta, \gamma, \sigma^2)$ with respect to (β, σ^2) is given by

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

As a consequence, the statistic LR_T admits the representation

$$LR_T = \max_{\rho\leq 1, \gamma\in\Gamma} \mathcal{L}_T(\rho, \gamma) - \max_{\gamma\in\Gamma} \mathcal{L}_T(1, \gamma). \tag{3}$$

Incorporating the constraint $\gamma \in \Gamma$ on the nuisance parameter γ is computationally cumbersome. Fortunately, the model (2) can be embedded in a model that is locally equivalent to it in a suitable sense, yet makes a quasi-likelihood ratio test statistic easier to compute. To be specific, the model (2) can be embedded in a model of ADF type, namely

$$\eta(L)\,\Delta u_t = \pi u_{t-1} + \varepsilon_t, \tag{4}$$

where $\pi \leq 0$ and $\eta(L) = 1 - \eta_1 L - \ldots - \eta_p L^p$ is an unrestricted lag polynomial of order p, and where $\{u_t\}$ and $\{\varepsilon_t\}$ are as before.

An important advantage of working with the parameterization (4) is that, even without restricting $\eta = (\eta_1, \ldots, \eta_p)'$, the key implication of the constraint $\min_{|z|\leq 1} |\gamma(z)| > 0$, namely $\gamma(1) > 0$, is incorporated in the sign restriction $\pi \leq 0$ on π. To be specific, when (2) is satisfied, the parameter π in (4) is given by $(\rho - 1)\,\gamma(1) \leq 0$, where the inequality is a consequence of the facts that $\rho \leq 1$ and $\gamma(1) > 0$. The problem of testing H_0 versus H_1 in the model characterized by (1) and (2) is therefore 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 (4). As it turns out, the quasi-likelihood ratio test statistic associated with the problem of testing H_0^{ADF} versus H_1^{ADF} in the more general model (4) is asymptotically equivalent to the quasi-likelihood ratio test statistic (3) associated with the problem of testing H_0 versus H_1 in the model (2).

Following the derivation of (3), we next derive the quasi-likelihood ratio test statistic associated with the problem of testing H_0^{ADF} versus H_1^{ADF}. The parameter of interest is now π, while (β, η, σ^2) is a nuisance parameter. The Gaussian quasi-log likelihood function corresponding to the model given by (1) and (4) with initial conditions $u_0 = \ldots = u_{-p} = 0$ can be expressed, up to a constant, as

$$L_T^{ADF}\left(\pi, \beta, \eta, \sigma^2\right) = -\frac{T}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\left(Y_{\pi,\eta}^{ADF} - D_{\pi,\eta}^{ADF}\beta\right)'\left(Y_{\pi,\eta}^{ADF} - D_{\pi,\eta}^{ADF}\beta\right),$$

where, setting again $y_0 = \ldots = y_{-p} = 0$ and $d_0 = \ldots = d_{-p} = 0$, we now define $Y_{\pi,\eta}^{ADF}$ and $D_{\pi,\eta}^{ADF}$ as the matrices with row $t = 1, \ldots, T$ given by $(\eta(L)\,(1 - L) - \pi L)\, y_t$ and $(\eta(L)\,(1 - L) - \pi L)\, d_t'$, respectively. The corresponding quasi-likelihood ratio test statistic for testing H_0^{ADF} versus H_1^{ADF} is

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

As before, the profile quasi-log likelihood for (π, η) is given, up to a constant, by

$$\mathcal{L}_T^{ADF}(\pi, \eta) = -\frac{T}{2} \log \left(Y'Y - Y'D(D'D)^{-1}D'Y \right) \big|_{Y = Y_{\pi,\eta}^{ADF}, D = D_{\pi,\eta}^{ADF}},$$

so that the statistic LR_T^{ADF} admits the representation

$$LR_T^{ADF} = \max_{\pi \leq 0, \eta} \mathcal{L}_T^{ADF}(\pi, \eta) - \max_{\eta} \mathcal{L}_T^{ADF}(0, \eta). \tag{5}$$

Unlike (3), both terms on the right-hand side of (5) are relatively easy to evaluate numerically. In addition, LR_T^{ADF} turns out to have attractive large sample properties.

As alluded to above, the main difference between (2) and (4) lies in the restrictions imposed on the parameters. Without any parameter restrictions, these two models are equivalent in the sense that one is a reparametrization of the other. However, when the parameter restrictions mentioned in the text are imposed, there is a substantive "global" difference between the two models. In particular, (2) can only generate I(0) and I(1) processes, whereas (4) can in fact generate I(d) models for $d = 0, 1, 2, \ldots$. However, the models (2) and (4) are fortunately "locally" equivalent in a neighborhood of the null hypothesis $\rho = 1$. Specifically, (2) imposes many restrictions on the parameters, but the only asymptotically relevant restriction can be more elegantly formulated in (4).

In other words, without restrictions on γ, the test statistic in (3) behaves like a "two-sided" test in the sense that its limiting distribution is $\max_{\bar{c}} \Lambda_c(\bar{c})$, i.e. the maximum without a sign-restriction on $\bar{c}$; c.f. Theorem 1. For example, when $p = 1$, ρ and γ are not separately identified in (2), so that maximizing the likelihood with a restriction on ρ but not on γ is equivalent to maximizing with a restriction on γ but not on ρ. The latter statistic corresponds to a two-sided test. An important property of the model (4) is that, although (2) implies restrictions on $\eta(L)$, there is a sense in which those restrictions do not bind asymptotically. We can therefore achieve asymptotic optimality without imposing restrictions on $\eta(L)$, which is computationally advantageous. In contrast, of course, the restriction on π does bind asymptotically and it is important to impose it.

Because of the advantages of LR_T^{ADF} relative to LR_T described above, we analyze the limiting behavior of the former in the following theorem.

Theorem 1 *Suppose $\{y_t\}$ is generated by (1) and (2) and that $c = T(\rho - 1)$ is held fixed as $T \to \infty$. Let $W_c(r) = \int_0^r \exp(c(r-s))\, dW(s)$, where W is a standard Wiener process.*

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

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

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
1000	0.77	0.99	1.32	1.91	2.52	3.36	4.01	5.57
∞	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
1000	2.46	2.80	3.27	4.05	4.83	5.84	6.59	8.31
∞	2.45	2.79	3.26	4.05	4.82	5.82	6.57	8.29

Notes: This table is taken from JN, Table 1. Entries for finite T are simulated quantiles of LR_T^{ADF} with known (γ, σ^2) and with $\varepsilon_t \sim i.i.d.N(0,1)$. 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.

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

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

A proof of the theorem is provided in the appendix. As further discussed below, the same method of proof can be used to analyze LR_T.

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 upon 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 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

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

where $\tilde{\eta}_T$ and $\tilde{\sigma}_T^2$ are estimators of η and σ^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. Likewise, the statistic

$$\max_{\pi\leq 0}\mathcal{L}_T^{ADF}\left(\pi,\tilde{\eta}_T\right)-\mathcal{L}_T^{ADF}\left(0,\tilde{\eta}_T\right)$$

is asymptotically equivalent to LR_T^{ADF} if $\tilde{\eta}_T$ is consistent.

Conversely, it stands to reason that a counterpart of JN's statistic, $\widehat{LR}_T^d$, that does not replace any nuisance parameters with estimators will be asymptotically equivalent to $\widehat{LR}_T^d$. In the notation of this paper, $\widehat{LR}_T^d$ can be represented as

$$\widehat{LR}_T^d=\max_{\rho\leq 1,\beta}L_T\left(\rho,\beta,\tilde{\gamma}_T,\tilde{\sigma}_T^2\right)-\max_{\beta}L_T\left(1,\beta,\tilde{\gamma}_T,\tilde{\sigma}_T^2\right),$$

where $\tilde{\gamma}_T$ and $\tilde{\sigma}_T^2$ are (consistent) estimators of γ and σ^2. In other words, $\widehat{LR}_T^d$ is a plug-in version of LR_T and as one would expect it can be shown that the two statistics are asymptotically equivalent under the assumptions of Theorem 1.

A direct proof of the fact that LR_T has the same limiting distribution as $\widehat{LR}_T^d$ and LR_T^{ADF} can be obtained by slightly modifying the proof of Theorem 1. The proof of Theorem 1 first shows that $\hat{\pi}_T=O_p(T^{-1})$, $\hat{\eta}_T=\eta+O_p(T^{-1/2})$, and $\tilde{\eta}_T=\eta+O_p(T^{-1/2})$, where

$$\left(\hat{\pi}_T,\hat{\eta}_T\right)=\arg\max_{\pi\leq 0,\eta}\mathcal{L}_T^{ADF}\left(\pi,\eta\right)\quad\text{and}\quad\tilde{\eta}_T=\arg\max_{\eta}\mathcal{L}_T^{ADF}\left(0,\eta\right),$$

and then obtains the limiting distribution of a centered and localized empirical process. Similarly, LR_T can be analyzed by first showing that $\hat{\rho}_T=1+O_p(T^{-1})$, $\hat{\gamma}_T=\gamma+O_p(T^{-1/2})$, and $\tilde{\gamma}_T=\gamma+O_p(T^{-1/2})$, where

$$\left(\hat{\rho}_T,\hat{\gamma}_T\right)=\arg\max_{\rho\leq 1,\gamma\in\Gamma}\mathcal{L}_T\left(\rho,\gamma\right)\quad\text{and}\quad\tilde{\gamma}_T=\arg\max_{\gamma\in\Gamma}\mathcal{L}_T\left(1,\gamma\right),$$

and then obtaining the limiting distribution of a centered and localized empirical process. Indeed, the empirical processes in question are the same in both cases, and incorporating the restriction $\gamma\in\Gamma$ does not affect the argument we use to obtain convergence rates, so analyzing LR_T involves no additional conceptual difficulties.

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

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

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

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

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

3 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 $\beta = 0$, $p = 3$, $u_0 = u_{-1} = \ldots = u_{-3} = 0$, and $\varepsilon_t \sim i.i.d.\,\mathcal{N}\left(0,1\right)$. For each of 10^5 replications, we simulate data from the model with sample size $T \in \{300, 1000\}$ and the parameter of interest ρ either equal to one or belonging to a grid chosen to ensure that the rejection rates of the various tests are around 0.5, 0.75, and 0.9, respectively. Regarding the nuisance parameter γ, we employ a parameterization of the form $\gamma\left(L\right) = \prod_{i=1}^{3}\left(1-\phi_i L\right)$, where ϕ_i are the inverse roots of the polynomial $\gamma(L)$. 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 ϕ 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 all well outside the unit circle.

For each DGP, we implement four 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 (4). The remaining three are the test based on LR_T^{ADF}, the test based on the statistic $\widehat{LR}_T^d$ of JN, and the DF-GLS test of ERS, respectively, each using the lag length chosen by the MAIC applied to the DF-GLS regression. Table 2 reports rejection rates of tests with nominal size 5% for the constant mean case, while the corresponding results for the linear trend case are reported in Table 3.

When LR_T^{ADF} is implemented using a lag length selected by applying the MAIC to the ADF model (4), the test exhibits excellent size and power properties across all cases considered. 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 $\phi_1 = \phi_2 = \phi_3 = 0.6$. Also, even in cases where it exhibits size inflation, the DF-GLS test does not dominate the LR_T^{ADF} test in terms of power.

Table 2: Rejection frequencies of unit root tests, constant mean case

DGP	$T=300$					$T=1000$				
ϕ_1, ϕ_2, ϕ_3	ρ	LR_T^{ADF}	$LR_{T,gls}^{ADF}$	$\widehat{LR}_T^d$	DF-GLS	ρ	LR_T^{ADF}	$LR_{T,gls}^{ADF}$	$\widehat{LR}_T^d$	DF-GLS
0, 0, 0	1.000	0.043	0.043	0.040	0.048	1.000	0.048	0.048	0.047	0.049
	0.970	0.568	0.566	0.544	0.599	0.993	0.475	0.471	0.464	0.480
	0.950	0.862	0.861	0.842	0.869	0.990	0.719	0.713	0.706	0.722
	0.930	0.944	0.944	0.928	0.938	0.985	0.943	0.939	0.935	0.941
0.2, 0.4, 0.6	1.000	0.035	0.077	0.042	0.101	1.000	0.045	0.088	0.076	0.097
	0.970	0.461	0.515	0.408	0.581	0.993	0.438	0.534	0.498	0.557
	0.950	0.767	0.762	0.679	0.809	0.990	0.676	0.713	0.682	0.730
	0.930	0.905	0.895	0.846	0.917	0.985	0.918	0.915	0.900	0.923
0.4, 0.4, 0.4	1.000	0.034	0.061	0.035	0.086	1.000	0.045	0.067	0.058	0.076
	0.970	0.459	0.477	0.386	0.545	0.993	0.442	0.470	0.439	0.493
	0.950	0.773	0.763	0.690	0.810	0.990	0.681	0.675	0.646	0.695
	0.930	0.914	0.905	0.862	0.925	0.985	0.919	0.910	0.895	0.918
0.6, 0.6, 0.6	1.000	0.038	0.156	0.056	0.159	1.000	0.047	0.149	0.133	0.150
	0.970	0.435	0.604	0.446	0.635	0.993	0.442	0.615	0.586	0.622
	0.950	0.698	0.712	0.593	0.752	0.990	0.673	0.738	0.707	0.747
	0.930	0.839	0.815	0.726	0.848	0.985	0.906	0.879	0.857	0.886

Notes: Rejection frequencies for the likelihood ratio test (LR_T^{ADF}, and $LR_{T,gls}^{ADF}$), the plug-in likelihood ratio test of JN, 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 (4) for the first test (LR_T^{ADF}), and to the DF-GLS regression for the other three tests.

Results for other values of ϕ are qualitatively similar and are omitted to conserve space. Also omitted are results for models outside the AR class, notably the notoriously difficult case of an MA(1) process with a negative root, where all tests exhibit size distortions. 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 weakly superior to) its natural rivals also in samples of moderate size.

4 Concluding Remarks

This paper has developed and analyzed quasi-likelihood ratio test statistics in an autoregressive model of arbitrary, finite 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 quasi-likelihood ratio 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 an-

Table 3: Rejection frequencies of unit root tests, linear trend case

DGP	T = 300					T = 1000				
ϕ_1, ϕ_2, ϕ_3	ρ	LR_T^{ADF}	$LR_{T,gls}^{ADF}$	$\widehat{LR}_T^d$	DF-GLS	ρ	LR_T^{ADF}	$LR_{T,gls}^{ADF}$	$\widehat{LR}_T^d$	DF-GLS
0, 0, 0	1.000	0.038	0.038	0.033	0.028	1.000	0.046	0.045	0.043	0.032
	0.940	0.612	0.612	0.563	0.527	0.985	0.554	0.547	0.528	0.452
	0.920	0.795	0.798	0.750	0.725	0.980	0.802	0.793	0.775	0.703
	0.860	0.928	0.921	0.880	0.876	0.970	0.980	0.973	0.962	0.938
0.2, 0.4, 0.6	1.000	0.028	0.129	0.021	0.111	1.000	0.040	0.137	0.096	0.109
	0.940	0.407	0.468	0.257	0.411	0.985	0.474	0.569	0.489	0.483
	0.920	0.590	0.603	0.398	0.536	0.980	0.714	0.741	0.667	0.650
	0.860	0.854	0.844	0.733	0.793	0.970	0.953	0.945	0.913	0.893
0.4, 0.4, 0.4	1.000	0.027	0.102	0.020	0.091	1.000	0.040	0.107	0.075	0.084
	0.940	0.409	0.440	0.255	0.382	0.985	0.481	0.514	0.442	0.429
	0.920	0.600	0.603	0.416	0.536	0.980	0.718	0.714	0.645	0.621
	0.860	0.870	0.861	0.765	0.810	0.970	0.953	0.944	0.915	0.893
0.6, 0.6, 0.6	1.000	0.027	0.251	0.003	0.231	1.000	0.043	0.223	0.163	0.205
	0.940	0.303	0.519	0.092	0.472	0.985	0.470	0.610	0.535	0.554
	0.920	0.422	0.553	0.181	0.493	0.980	0.689	0.725	0.643	0.656
	0.860	0.669	0.702	0.474	0.635	0.970	0.929	0.886	0.834	0.827

Notes: Rejection frequencies for the likelihood ratio test (LR_T^{ADF}, and $LR_{T,gls}^{ADF}$), the plug-in likelihood ratio test of JN, 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 (4) for the first test (LR_T^{ADF}), and to the DF-GLS regression for the other three tests.

alytical 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.

Although doing so is beyond the scope of this paper, it would be of both theoretical and practical interest to allow for more general short-run dynamics than those considered here. In particular, it would be of interest to explore whether one can accommodate either sieve-type autoregressive approximations involving a growing value of p or more complicated finite-dimensional parametric models than the AR(p) model.

It would also be of interest to develop and analyze quasi-likelihood ratio 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 quasi-likelihood ratio interpretation.

A 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 $\beta = 0$. Also, the proof of part (a) is a special case of the proof of part (b), so we only give the proof of part (b).

Let the process Λ_c^{ADF} be given by

$$\Lambda_c^{ADF}(\bar{c}, h) = \Lambda_c^{\tau}(\bar{c}) + h'\mathcal{Z} - \frac{1}{2}h'\Xi h,$$

where

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

and where $\mathcal{Z} \sim \mathcal{N}(0, \Xi)$ is independent of W_c. Also, define

$$\tilde{\sigma}_T^2 = \frac{1}{T}\left(Y'Y - Y'D(D'D)^{-1}D'Y\right)\Big|_{Y=Y_{0,\gamma}^{ADF}, D=D_{0,\gamma}^{ADF}} = T^{-1}\sum_{t=1}^{T}\varepsilon_t^2 + o_p(1) \to_p \sigma^2,$$

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

$$\lambda_T^{ADF}(\bar{c}, h) = \lambda_T^{ADF,0}(\bar{c}, h) + \lambda_T^{ADF,d}(\bar{c}, h),$$

where

$$\lambda_T^{ADF,0}(\bar{c}, h) = \frac{1}{2\tilde{\sigma}_T^2}\, Y'Y\big|_{Y=Y_{0,\gamma}^{ADF}} - \frac{1}{2\tilde{\sigma}_T^2}\, Y'Y\big|_{Y=Y_{\gamma(1)\bar{c}/T, \gamma+h/\sqrt{T}}^{ADF}}$$

and

$$\lambda_T^{ADF,d}(\bar{c}, h) = \frac{1}{2\tilde{\sigma}_T^2}\left(Y'D(D'D)^{-1}D'Y\right)\Big|_{Y=Y_{\gamma(1)\bar{c}/T, \gamma+h/\sqrt{T}}^{ADF}, D=D_{\gamma(1)\bar{c}/T, \gamma+h/\sqrt{T}}^{ADF}} - \frac{1}{2\tilde{\sigma}_T^2}\left(Y'D(D'D)^{-1}D'Y\right)\Big|_{Y=Y_{0,\gamma}^{ADF}, D=D_{0,\gamma}^{ADF}}.$$

Because

$$\mathcal{L}_T^{ADF}(\gamma(1)\bar{c}/T, \gamma + h/\sqrt{T}) - \mathcal{L}_T^{ADF}(0, \gamma) = G_T(\lambda_T^{ADF}(\bar{c}, h)),$$

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

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

Suppose

$$\left(\max_{\bar{c}\leq 0,h}\lambda_T^{ADF}(\bar{c},h),\max_h\lambda_T^{ADF}(0,h)\right)\to_d\left(\max_{\bar{c}\leq 0,h}\Lambda_c^{ADF}(\bar{c},h),\max_h\Lambda_c^{ADF}(0,h)\right). \quad (6)$$

Then

$$\begin{aligned}LR_T^{ADF}&=\max_{\bar{c}\leq 0,h}\lambda_T^{ADF}(\bar{c},h)-\max_h\lambda_T^{ADF}(0,h)+o_p(1)\\&\to_d\max_{\bar{c}\leq 0,h}\Lambda_c^{ADF}(\bar{c},h)-\max_h\Lambda_c^{ADF}(0,h)\\&=\max_{\bar{c}\leq 0}\Lambda_c^{\tau}(\bar{c})+\frac{1}{2}\mathcal{Z}'\Xi^{-1}\mathcal{Z}-\frac{1}{2}\mathcal{Z}'\Xi^{-1}\mathcal{Z}=\max_{\bar{c}\leq 0}\Lambda_c^{\tau}(\bar{c}),\end{aligned}$$

where the first equality uses the facts that (i) the left-hand side of (6) is $O_p(1)$ and (ii) $\lim_{T\to\infty}\sup_{|x|\leq M}|G_T(x)-x|=0$ for any $0\leq M<\infty$. The proof can therefore be completed by verifying (6). We shall do so by showing that

$$(\hat{c}_T,\hat{h}_T)=\arg\max_{\bar{c}\leq 0,h}\lambda_T^{ADF}(\bar{c},h)=O_p(1)\quad\text{and}\quad\tilde{h}_T=\arg\max_h\lambda_T^{ADF}(0,h)=O_p(1), \quad (7)$$

and that λ_T^{ADF} converges to Λ_c^{ADF} in the topology of uniform convergence on compacta.

Define $\mathcal{Y}_T=Y_{0,\gamma}^{ADF}$ and $\mathcal{D}_T=D_{0,\gamma}^{ADF}\,\mathrm{diag}(1,1/\sqrt{T})$, and let $\dot{\mathcal{Y}}_T$ and $\dot{\mathcal{D}}_T$ be matrices with row $t=1,\ldots,T$ given by

$$\left(T^{-1}\gamma(1)\,y_{t-1},T^{-1/2}\Delta y_{t-1},\ldots,T^{-1/2}\Delta y_{t-p}\right)$$

and

$$\left(T^{-1}\gamma(1)\,d_{t-1}',T^{-1/2}\Delta d_{t-1}',\ldots,T^{-1/2}\Delta d_{t-p}'\right)\left(I_{p+1}\otimes\mathrm{diag}(1,1/\sqrt{T})\right),$$

respectively. Because

$$Y_{\gamma(1)\bar{c}/T,\gamma+h/\sqrt{T}}^{ADF}=\mathcal{Y}_T+\dot{\mathcal{Y}}_T(\bar{c},h')'$$

and

$$D_{\gamma(1)\bar{c}/T,\gamma+h/\sqrt{T}}^{ADF}\,\mathrm{diag}(1,1/\sqrt{T})=\mathcal{D}_T+\dot{\mathcal{D}}_T\left((\bar{c},h')'\otimes I_2\right),$$

the function λ_T^{ADF} admits a representation of the form

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

where

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

by standard results (e.g., Chan and Wei, 1987; Phillips, 1987) and where F is a continuous function. Moreover, the function F satisfies (with probability one, for every T) that

$$\lim_{\|(\bar{c},h')'\|\to\infty} \frac{1}{\|(\bar{c},h')'\|^2} \left|F(\bar{c},h,\mathcal{S}_T) - F^*(\bar{c},h,\mathcal{S}_T)\right| = 0,$$

where, letting $\dot{\mu}_T$ denote the smallest eigenvalue of $\dot{\mathcal{Y}}_T'\dot{\mathcal{Y}}_T - \dot{\mathcal{Y}}_T'\dot{\mathcal{D}}_T(\dot{\mathcal{D}}_T'\dot{\mathcal{D}}_T)^{-1}\dot{\mathcal{D}}_T'\dot{\mathcal{Y}}_T$,

$$\begin{aligned}
F^*(\bar{c},h,\mathcal{S}_T) &= -\frac{1}{2\tilde{\sigma}_T^2}(\bar{c},h')\left(\dot{\mathcal{Y}}_T'\dot{\mathcal{Y}}_T - \dot{\mathcal{Y}}_T'\dot{\mathcal{D}}_T\theta(\theta'\dot{\mathcal{D}}_T'\dot{\mathcal{D}}_T\theta)^{-1}\theta'\dot{\mathcal{D}}_T'\dot{\mathcal{Y}}_T\right)(\bar{c},h')'\Big|_{\theta=(\bar{c},h')'\otimes I_2} \\
&= -\frac{1}{2\tilde{\sigma}_T^2}\min_{B\in\mathbb{R}^2}\left\|\dot{\mathcal{Y}}_T(\bar{c},h')' - \dot{\mathcal{D}}_T\theta B\right\|^2\Big|_{\theta=(\bar{c},h')'\otimes I_2} \\
&\le -\frac{1}{2\tilde{\sigma}_T^2}\min_{B\in\mathbb{R}^{2(p+1)}}\left\|\dot{\mathcal{Y}}_T(\bar{c},h')' - \dot{\mathcal{D}}_T B\right\|^2 \\
&= -\frac{1}{2\tilde{\sigma}_T^2}(\bar{c},h')\left(\dot{\mathcal{Y}}_T'\dot{\mathcal{Y}}_T - \dot{\mathcal{Y}}_T'\dot{\mathcal{D}}_T(\dot{\mathcal{D}}_T'\dot{\mathcal{D}}_T)^{-1}\dot{\mathcal{D}}_T'\dot{\mathcal{Y}}_T\right)(\bar{c},h')' \\
&\le -\frac{\dot{\mu}_T}{2\tilde{\sigma}_T^2}\|(\bar{c},h')'\|^2.
\end{aligned}$$

Using $\tilde{\sigma}_T^2 \to_p \sigma^2 > 0$, $\mathcal{S}_T = O_p(1)$, and the fact that $\dot{\mu}_T > 0$ with probability approaching one, it can be shown that every $\delta > 0$ admits compact sets $K_0 \subseteq \{0\}\times\mathbb{R}^p$, $K_1 \subseteq \mathbb{R}_- \times \mathbb{R}^p$, and $\mathcal{K}$ satisfying

$$\limsup_{T\to\infty} \Pr[\mathcal{S}_T \notin \mathcal{K}] \le \delta,$$

$$\max_{(\bar{c},h)\in\{0\}\times\mathbb{R}^p\setminus K_0,\mathcal{S}\in\mathcal{K}} F(\bar{c},h,\mathcal{S}) \le 0 = F(0,0,\mathcal{S}) \le \max_{(\bar{c},h)\in K_0,\mathcal{S}\in\mathcal{K}} F(\bar{c},h,\mathcal{S}),$$

and

$$\max_{(\bar{c},h)\in\mathbb{R}_-\times\mathbb{R}^p\setminus K_1,\mathcal{S}\in\mathcal{K}} F(\bar{c},h,\mathcal{S}) \le 0 = F(0,0,\mathcal{S}) \le \max_{(\bar{c},h)\in K_1,\mathcal{S}\in\mathcal{K}} F(\bar{c},h,\mathcal{S}).$$

As a consequence,

$$\limsup_{T\to\infty} \Pr\left(\max_{(\bar{c},h)\in K_0} \lambda_T^{ADF}(\bar{c},h) < \max_h \lambda_T^{ADF}(0,h)\right) \le \limsup_{T\to\infty} \Pr(\mathcal{S}_T \notin \mathcal{K}) \le \delta$$

and

$$\limsup_{T\to\infty} \Pr\left(\max_{(\bar{c},h)\in K_1} \lambda_T^{ADF}(\bar{c},h) < \max_{\bar{c}\le 0,h} \lambda_T^{ADF}(\bar{c},h)\right) \le \limsup_{T\to\infty} \Pr(\mathcal{S}_T \notin \mathcal{K}) \le \delta,$$

implying in particular that (7) holds.

Next, it follows from Prohorov's Theorem (e.g., Kallenberg, 2002, Theorem 16.5) that λ_T^{ADF} converges to Λ_c^{ADF} in the topology of uniform convergence on compacta if λ_T^{ADF} converges to Λ_c^{ADF} in the sense of weak convergence of finite-dimensional projections and if the process $\left\{\lambda_T^{ADF}(\bar{c},h) : (\bar{c},h')' \in K\right\}$ is tight for any compact set K. For any fixed $(\bar{c},h)$, it follows from standard results (e.g., Chan and Wei, 1987; Phillips, 1987) that

$$\lambda_T^{ADF,0}(\bar{c},h) = \bar{c}T^{-1}\sum_{t=1}^{T}\breve{y}_{t-1}\left(\breve{\varepsilon}_t + cT^{-1}\breve{y}_{t-1}\right) - \frac{1}{2}\bar{c}^2T^{-2}\sum_{t=1}^{T}\breve{y}_{t-1}^2$$
$$+ h'T^{-1/2}\sum_{t=1}^{T}\breve{v}_{t-1}\breve{\varepsilon}_t - \frac{1}{2}h'T^{-1}\sum_{t=1}^{T}\breve{v}_{t-1}\breve{v}_{t-1}'h + o_p(1)$$
$$\to_d \Lambda_c(\bar{c}) + h'\mathcal{Z} - \frac{1}{2}h'\Xi h,$$

and

$$\lambda_T^{ADF,d}(\bar{c},h) = \frac{1}{2}\frac{\left(T^{-1/2}\sum_{t=1}^{T}\left(1-\bar{c}T^{-1}t\right)\left(\breve{\varepsilon}_t + cT^{-1}\breve{y}_{t-1} - \bar{c}T^{-1}\breve{y}_{t-1}\right)\right)^2}{1-\bar{c}+\bar{c}^2/3}$$
$$- \frac{1}{2}\frac{\left(T^{-1/2}\sum_{t=1}^{T}\left(\breve{\varepsilon}_t + cT^{-1}\breve{y}_{t-1}\right)\right)^2}{1} + o_p(1)$$
$$\to_d \frac{1}{2}\frac{\left(\left(1-\bar{c}\right)W_c(1) + \bar{c}^2\int_0^1 rW_c(r)\,dr\right)^2}{1-\bar{c}+\bar{c}^2/3} - \frac{1}{2}W_c(1)^2$$

jointly, where $\breve{\varepsilon}_t = \sigma^{-1}\varepsilon_t$ and $\breve{y}_t = \sigma^{-1}\gamma(1)\,y_t$. Therefore,

$$\lambda_T^{ADF}(\bar{c},h) \to_d \Lambda_c^{ADF}(\bar{c},h)$$

for fixed $(\bar{c},h)$ and the Cramér-Wold device can be used to show that λ_T^{ADF} converges to Λ_c^{ADF} in the sense of weak convergence of finite-dimensional projections. Finally, because F is continuous, it follows from the Arzelà-Ascoli Theorem (Dudley, 2002, Theorem 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 $\mathcal{S}_T$ is tight implies that, for any compact set K, the process $\left\{\lambda_T^{ADF}(\bar{c},h) : (\bar{c},h')'\in K\right\} = \{F(\bar{c},h,\mathcal{S}_T) : (\bar{c},h')'\in K\}$ is tight.

References

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

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

Dickey, D. A. and 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. and 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* (Second ed.). Cambridge, UK: Cambridge University Press.

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

Jansson, M. and 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* (Second ed.). New York, NY: Springer.

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

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

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

Perron, P. and 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.