Nearly Efficient Likelihood Ratio Tests of the Unit Root Hypothesis*

MICHAEL JANSSON†
UC BERKELEY AND CREATES

MORTEN ØRREGAARD NIELSEN
QUEEN'S UNIVERSITY AND CREATES

January 7, 2011

ABSTRACT. Seemingly absent from the arsenal of currently available "nearly efficient" testing procedures for the unit root hypothesis, i.e. tests whose asymptotic local power functions are virtually indistinguishable from the Gaussian power envelope, is a test admitting a (quasi-)likelihood ratio interpretation. We study the large sample properties of a quasi-likelihood ratio unit root test based on a Gaussian likelihood and show that this test is nearly efficient.

KEYWORDS: Efficiency, likelihood ratio test, unit root hypothesis
JEL CODES: C12, C22

1. INTRODUCTION

The unit root testing problem has been and continues to be a testing problem of great theoretical interest in time series econometrics.[1] In a seminal paper, Elliott, Rothenberg, and Stock (1996, henceforth ERS) derived Gaussian power envelopes for unit root tests and demonstrated by example that these envelopes are sharp in the sense that "nearly efficient" tests, i.e., tests whose asymptotic local power functions are virtually indistinguishable from the Gaussian power envelope, can be constructed. Subsequent research (e.g., Ng and Perron (2001)) has enlarged the class of tests whose asymptotic local power functions are indistinguishable from the Gaussian power envelope, but seemingly absent from the arsenal of currently available nearly efficient testing procedures is a test admitting a (quasi-)likelihood ratio interpretation. The purpose of this note is to propose and analyze such a test.

In models with an unknown mean and/or a linear trend, the class of nearly efficient unit root tests does not contain the Dickey and Fuller (1979, 1981, henceforth DF) tests. Therefore, although the DF tests can be given a likelihood ratio interpretation it is perhaps not ex ante obvious that nearly efficient likelihood ratio tests even exist. In other words, it would appear to be an open question whether the unit root testing problem can be added to list of testing problems for which likelihood ratio tests perform poorly (e.g., Lehmann (2006) and the references therein).

*We are grateful to Jim Stock, three referees, Niels Haldrup, Søren Johansen, Tom Rothenberg, and seminar participants at the University of Aarhus, the Econometric Society World Congress in Shanghai, the HEC Montréal-CIRPÉE Applied Financial Time Series Workshop, and the CREATES conference on Periodicity, Non-stationarity, and Forecasting of Economic and Financial Time Series for comments and discussion, and to the Danish Social Sciences Research Council (FSE grant no. 275-05-0220), the Social Sciences and Humanities Research Council of Canada (SSHRC grant no. 410-2009-0183), and the Center for Research in Econometric Analysis of Time Series (CREATES, funded by the Danish National Research Foundation) 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]For reviews, see Stock (1994) and Haldrup and Jansson (2006).

The DF tests can be derived from a conditional likelihood, conditioning being with respect to the initial observation. In the model considered by ERS the initial observation is very informative about the parameters governing the deterministic component, so it seems plausible that a likelihood ratio test derived from the full likelihood implied by an ERS-type model would have superior power properties to those of the DF tests in models with deterministic components and this is exactly what we find. Indeed, we find that a likelihood ratio test constructed in this way does belong to the class of nearly efficient tests. Moreover, we show that the new tests are related to, but distinct from, the point optimal and DF-GLS tests of ERS, even asymptotically.

Section 2 contains our results on the likelihood ratio test for a unit root, with additional discussion in Section 3. The proof of our main result is provided in Section 4.

2. The Likelihood Ratio Test for a Unit Root

We initially consider unit root testing in a model devoid of nuisance parameters, namely the zero-mean Gaussian AR(1) model where $\{y_t : 1 \leq t \leq T\}$ is generated as

$$y_t = \rho y_{t-1} + \varepsilon_t, \tag{1}$$

where $y_0 = 0$ and $\varepsilon_t \sim i.i.d.\ \mathcal{N}(0,1)$.

In this model, the likelihood ratio test associated with the unit root testing problem $H_0 : \rho = 1$ vs. $H_1 : \rho < 1$ rejects for large values of $LR_T = \max_{\bar{\rho} \leq 1} L_T(\bar{\rho}) - L_T(1)$, where $L_T(\rho) = -\frac{1}{2}\sum_{t=1}^{T}(y_t - \rho y_{t-1})^2$ is, up to a constant, the log likelihood function. In terms of the sufficient statistics $S_T = T^{-1}\sum_{t=2}^{T} y_{t-1}\Delta y_t$ and $H_T = T^{-2}\sum_{t=2}^{T} y_{t-1}^2$, the log likelihood function can be expressed as $L_T(\rho) = L_T(1) + T(\rho - 1) S_T - \frac{1}{2}[T(\rho-1)]^2 H_T$. As a consequence, defining $\bar{c} = T(\bar{\rho} - 1)$ to obtain non-degenerate asymptotic behavior, LR_T admits the representation

$$LR_T = \max_{\bar{c} \leq 0}[\bar{c} S_T - \frac{1}{2}\bar{c}^2 H_T]. \tag{2}$$

The large sample behavior of (S_T, H_T) is well understood (e.g., Chan and Wei (1987) and Phillips (1987)): Under local-to-unity asymptotics, with $c = T(\rho - 1)$ held fixed as $T \to \infty$,

$$(S_T, H_T) \to_d (\mathcal{S}_c, \mathcal{H}_c) = \left(\int_0^1 W_c(r)\, dW_c(r),\ \int_0^1 W_c(r)^2\, dr\right), \tag{3}$$

where $W_c(r) = \int_0^r \exp(c(r-s))\, dW(s)$ and $W(\cdot)$ is a standard Wiener process.

The corresponding result about the local-to-unity asymptotic behavior of the likelihood ratio statistic LR_T follows from (2), (3), and the continuous mapping theorem (CMT) applied to the functional $f(s,h) = \min(0,s)^2/h$. Specifically, using simple facts about quadratic functions and defining $\Lambda_c(\bar{c}) = \bar{c}\mathcal{S}_c - \frac{1}{2}\bar{c}^2\mathcal{H}_c$,

$$LR_T = \max_{\bar{c} \leq 0}[\bar{c} S_T - \frac{1}{2}\bar{c}^2 H_T] = \frac{\min(S_T, 0)^2}{2H_T} \to_d \frac{\min(\mathcal{S}_c, 0)^2}{2\mathcal{H}_c} = \max_{\bar{c} \leq 0}\Lambda_c(\bar{c}).$$

The implicit characterization of the weak limit of LR_T as $\max_{\bar{c} \leq 0}\Lambda_c(\bar{c})$ is employed in anticipation of Theorem 1(b) below, which covers a case where no closed form expression for the limiting random variable seems to be available.

In addition to facilitating the verification of the continuity property required to invoke the CMT, the closed form expression for $\max_{\bar{c}\leq 0}\Lambda_c(\bar{c})$ enables us to address the asymptotic optimality properties of the likelihood ratio test. For any α less than $\Pr[\mathcal{S}_0 \leq 0] \approx 0.6827$, the (asymptotic) size α likelihood ratio test rejects when LR_T exceeds $k_{LR}(\alpha)$, where $k_{LR}(\alpha)$ satisfies $\Pr[\max_{\bar{c}\leq 0}\Lambda_0(\bar{c}) > k_{LR}(\alpha)] = \alpha$. For any such α the asymptotic local power function associated with the size α likelihood ratio test coincides with that of the size α test based on the DF t-statistic $\hat{\tau}_T^{DF}$, the reason being that $\hat{\tau}_T^{DF} \to_d \mathcal{S}_c/\sqrt{\mathcal{H}_c}$ under the above assumptions. It therefore follows from ERS's results about the DF t-test that the likelihood ratio test is nearly efficient in the sense that its asymptotic local power function is virtually indistinguishable from the Gaussian power envelope.

The near-efficiency result for the test based on the DF t-statistic does not extend to models with a constant mean or a linear trend (e.g., ERS). Moreover, the assumptions that the quasi-differences $\{y_t - \rho y_{t-1}\}$ are $i.i.d.$ with a known distribution are implausible in many applications. It is therefore of theoretical and practical interest to explore the asymptotic local power properties of quasi-likelihood ratio tests in models with nuisance parameters governing deterministics and/or serial correlation. To that end, suppose $\{y_t : 1 \leq t \leq T\}$ is generated by the model

$$y_t = \beta' d_t + u_t, \qquad (1-\rho L)\gamma(L) u_t = \varepsilon_t, \tag{4}$$

where $d_t = 1$ or $d_t = (1,t)'$, β is an unknown parameter, $\gamma(L) = 1 - \gamma_1 L - \ldots - \gamma_p L^p$ is a lag polynomial of (known, finite[2]) order p satisfying $\min_{|z|\leq 1}|\gamma(z)| > 0$, the initial condition is $\max(|u_0|,\ldots,|u_{-p}|) = o_p(\sqrt{T})$, and the ε_t form a conditionally homoskedastic martingale difference sequence with (unknown) variance σ^2 and $\sup_t E|\varepsilon_t|^r < \infty$ for some $r > 2$.

The Gaussian quasi-log likelihood function corresponding to the model with $u_0 = \ldots = u_{-p} = 0$ can be expressed, up to a constant, as

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

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.

Consider a quasi-likelihood ratio-type test statistic of the form

$$\begin{aligned}\widehat{LR}_T^d &= \max_{\bar{\rho}\leq 1,\beta} L_T^d(\bar{\rho},\beta;\hat{\sigma}_T^2,\hat{\gamma}_T) - \max_\beta L_T^d(1,\beta;\hat{\sigma}_T^2,\hat{\gamma}_T) \\ &= \max_{\bar{\rho}\leq 1}\mathcal{L}_T^d(\bar{\rho};\hat{\sigma}_T^2,\hat{\gamma}_T) - \mathcal{L}_T^d(1;\hat{\sigma}_T^2,\hat{\gamma}_T),\end{aligned}$$

where $\hat{\sigma}_T^2$ and $\hat{\gamma}_T$ are estimators of σ^2 and $\gamma = (\gamma_1,\ldots,\gamma_p)'$, respectively, and

$$\mathcal{L}_T^d(\rho;\sigma^2,\gamma) = -\frac{T}{2}\log\sigma^2 - \frac{1}{2\sigma^2}Y_{\rho,\gamma}'Y_{\rho,\gamma} + \frac{1}{2\sigma^2}\left(Y_{\rho,\gamma}'D_{\rho,\gamma}\right)\left(D_{\rho,\gamma}'D_{\rho,\gamma}\right)^{-1}\left(D_{\rho,\gamma}'Y_{\rho,\gamma}\right)$$

is the profile log likelihood function obtained by maximizing $L_T^d(\rho,\beta;\sigma^2,\gamma)$ with respect to the nuisance parameter β governing the deterministic component. Being based on a plug-in version of $\mathcal{L}_T^d(\rho;\sigma^2,\gamma)$, the statistic $\widehat{LR}_T^d$ is straightforward to compute, requiring only

[2]If p is allowed to diverge slowly to infinity, it seems plausible that results analogous to Theorem 1 can be obtained when $(1-\rho L)u_t$ is generated by a linear process satisfying mild summability conditions (e.g., Chang and Park (2002)). Monte Carlo results reported in the supplementary material are consistent with this conjecture, but for simplicity our theoretical developments proceed under the assumption that p is fixed.

maximization of $\mathcal{L}_T^d(\rho;\hat{\sigma}_T^2,\hat{\gamma}_T)$ with respect to the scalar parameter ρ. Unlike $L_T(\rho)$, however, the profile log likelihood function $\mathcal{L}_T^d(\rho;\sigma^2,\gamma)$ depends on ρ in a complicated way and no closed form expression for $\widehat{LR}_T^d$ will be available in general; a feature which complicates, but does not prohibit, the derivation of its local-to-unity asymptotic distribution.

The proof of the following result proceeds by showing that the likelihood ratio statistic can be written as $\widehat{LR}_T^d = \max_{\bar{c}\le 0} F(\bar{c},\hat{X}_T)$ for some function $F(\cdot)$ and some random vector $\hat{X}_T$, where the latter enjoys a convergence property of the form $\hat{X}_T \to_d \mathcal{X}_c$ and the functional $\max_{\bar{c}\le 0} F(\bar{c},\cdot)$ is continuous on a set $\mathbb{X}$ satisfying $\Pr[\mathcal{X}_c \in \mathbb{X}] = 1$ (for every $c \le 0$).[3]

Theorem 1. *Suppose $\{y_t\}$ is generated by (4), $c = T(\rho-1)$ is held fixed as $T\to\infty$, and $(\hat{\sigma}_T^2,\hat{\gamma}_T)\to_p(\sigma^2,\gamma)$. Then:*

(a) If $d_t = 1$ then $\widehat{LR}_T^d \to_d \max_{\bar{c}\le 0}\Lambda_c(\bar{c})$.

(b) If $d_t = (1,t)'$ then $\widehat{LR}_T^d \to_d \max_{\bar{c}\le 0}\Lambda_c^\tau(\bar{c})$, where

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

The consistency requirement on the estimators $\hat{\sigma}_T^2$ and $\hat{\gamma}_T$ in Theorem 1 is mild. For instance, it is met by $\hat{\sigma}_T^2 = (T-p-1)^{-1}\sum_{t=p+2}^T(\Delta y_t - \hat{\kappa}_T' Z_t)^2$ and $\hat{\gamma}_T = (0, I_p)\,\hat{\kappa}_T$, where $\hat{\kappa}_T = (\sum_{t=p+2}^T Z_t Z_t')^{-1}(\sum_{t=p+2}^T Z_t\Delta y_t)$ and $Z_t = (\Delta d_t', \Delta y_{t-1},\ldots,\Delta y_{t-p})'$. In Monte Carlo simulations reported in the supplementary material this choice of estimators was found to deliver tests with good small sample properties. Critical values associated with $\widehat{LR}_T^d$ are reported in Table 1.

Because the profile log likelihood function $\mathcal{L}_T^d(\cdot;\sigma^2,\gamma)$ is invariant under transformations of the form $y_t \to y_t + b'd_t$, so is $\widehat{LR}_T^d$ (and any other test statistic that can expressed as a functional of $\mathcal{L}_T^d(\cdot;\hat{\sigma}_T^2,\hat{\gamma}_T)$) provided $(\hat{\sigma}_T^2,\hat{\gamma}_T)$ is invariant.[4] It therefore makes sense to compare the asymptotic local power properties of the tests based on $\widehat{LR}_T^d$ with ERS's Gaussian power envelopes for invariant tests. In the constant mean case, the envelope for invariant tests coincides with the envelope for the model (1) without deterministics. Similarly, it follows from Theorem 1(a) that the asymptotic local power of the constant mean likelihood ratio test coincides with the asymptotic local power of the no deterministics likelihood ratio test. The constant mean likelihood ratio test therefore inherits the near optimality property of its no deterministics counterpart. Figure 1 plots the asymptotic local power function (with argument $c \le 0$) of the size $\alpha = 0.05$ linear trend likelihood ratio test along with the Gaussian power envelope. As in the no deterministics and constant mean cases, the asymptotic local power function of the likelihood ratio test is indistinguishable from the Gaussian power envelope, so near optimality claims can be made on the part of the likelihood ratio test also in the linear trend case.

---

[3] An alternative method of proof, more heavily reliant on empirical process methods, has been outlined for a closely related test statistic by Boswijk (1998). We are grateful to Peter Boswijk for bringing that manuscript to our attention.

[4] The latter invariance property is enjoyed by the estimators of σ^2 and γ described in the preceding paragraph.

Table 1: Quantiles of the distribution of $\widehat{LR}_T^d$

T	80%	85%	90%	95%	97.5%	99%	99.5%	99.9%
Panel A: constant mean case								
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								
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

Note: Entries for finite T are simulated quantiles of $\widehat{LR}_T^d$ with $(\hat{\sigma}_T^2, \hat{\gamma}_T) = (\sigma^2, \gamma)$ and $\varepsilon_t \sim i.i.d.\ \mathcal{N}(0,1)$. 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.

3. Discussion

The near optimality properties of the likelihood ratio test are shared by two related, but distinct, classes of tests proposed by ERS, namely the point optimal tests and DF-GLS tests. To clarify the differences between the three classes of tests, consider the model (4) with $p = 0$ (so that $\gamma(L) = 1$) and $\sigma^2 = 1$, in which case the log likelihood and profile log likelihood functions are $L_T^*(\rho, \beta) = L_T^d(\rho, \beta; 1, 0)$ and $\mathcal{L}_T^*(\rho) = \mathcal{L}_T^d(\rho; 1, 0)$, respectively, and a version of the likelihood ratio test statistic is given by $LR_T^d = \max_{\bar{\rho}\leq 1} \mathcal{L}_T^*(\bar{\rho}) - \mathcal{L}_T^*(1)$.

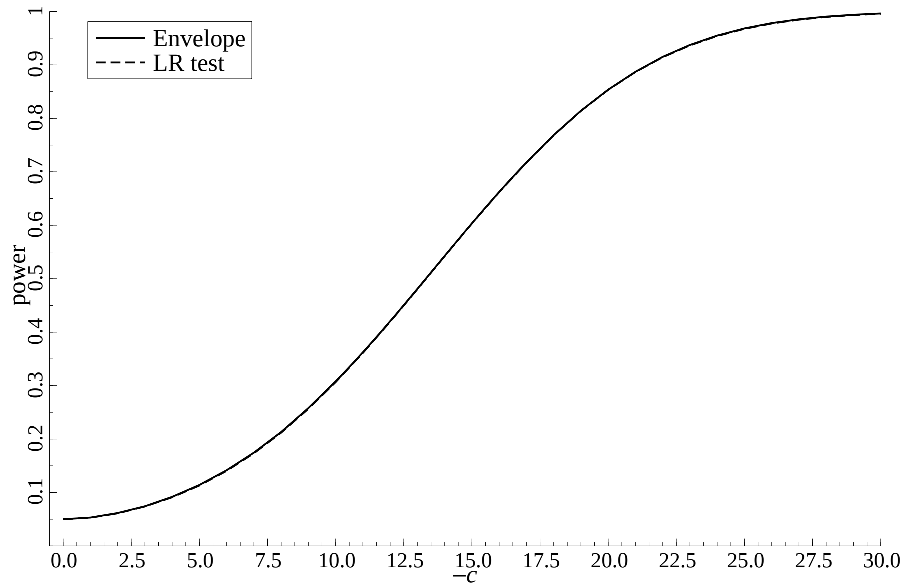
The point optimal test statistics are of the form $P_T(\bar{c}_{ERS}) = \mathcal{L}_T^*(1 + T^{-1}\bar{c}_{ERS}) - \mathcal{L}_T^*(1)$, where $\bar{c}_{ERS}$ is a negative constant. By construction, these tests are tangent to the Gaussian power envelope at $c = \bar{c}_{ERS}$. It was found by ERS that the choices $\bar{c}_{ERS} = -7$ and $\bar{c}_{ERS} = -13.5$ produce nearly efficient tests in the constant mean and linear trend cases, respectively. Defining $\hat{c}_{LR} = \arg\max_{\bar{c}\leq 0} \mathcal{L}_T^*(1 + T^{-1}\bar{c})$, the likelihood ratio test statistic can be expressed as $LR_T^d = P_T(\hat{c}_{LR})$. Because $\hat{c}_{LR}$ is random even in the limit, the likelihood ratio test cannot be interpreted as an (asymptotically) point optimal test.

The DF-GLS test is asymptotically equivalent to the test based on the test statistic $\hat{\tau}_T^{DF-GLS}(\bar{c}_{ERS}) = \max_{\bar{\rho}\leq 1} L_T^*(\bar{\rho}, \hat{\beta}_T(\bar{c}_{ERS})) - L_T^*(1, \hat{\beta}_T(\bar{c}_{ERS}))$, where $\bar{c}_{ERS}$ is a negative constant and $\hat{\beta}_T(\bar{c}_{ERS})$ is a plug-in estimator of β given by

$$\hat{\beta}_T(\bar{c}_{ERS}) = \arg\max_b L_T^d\left(1 + T^{-1}\bar{c}_{ERS}, b\right) = \left(D_{\bar{\rho},0}' D_{\bar{\rho},0}\right)^{-1} \left(D_{\bar{\rho},0} Y_{\bar{\rho},0}\right)\Big|_{\bar{\rho}=1+T^{-1}\bar{c}_{ERS}}.$$

As with the point optimal tests, ERS recommend setting $\bar{c}_{ERS}$ equal to -7 and -13.5 in the constant mean and linear trend cases, respectively. Under the assumptions of Theorem 1(a), the likelihood ratio test is asymptotically equivalent to the DF-GLS test since, for any $\bar{c}_{ERS} \leq 0$, $\hat{\tau}_T^{DF-GLS}(\bar{c}_{ERS}) \to_d \max_{\bar{c}\leq 0} \Lambda_c(\bar{c})$. In contrast, under the assumptions of Theorem 1(b), the asymptotic properties of $\hat{\tau}_T^{DF-GLS}(\bar{c}_{ERS})$ depend on $\bar{c}_{ERS}$ and the likelihood ratio

Figure 1: Power envelope and asymptotic local power of LR test with a linear trend



Note: Simulated power envelope and asymptotic local power function based on 10^6 Monte Carlo replications, where Wiener processes were approximated by 10^4 discrete steps with standard Gaussian innovations.

test cannot be interpreted as being asymptotically equivalent to a DF-GLS test in the linear trend case.

Thus, although the tests based on $\widehat{LR}_T^d$ are virtually identical to the DF-GLS tests of ERS in terms of asymptotic local power properties, the LR-type tests introduced herein are conceptually distinct from the DF-GLS tests. Specifically, while both tests achieve nuisance parameter elimination by first plugging in estimators of one subset of the nuisance parameters and then profiling out the remaining nuisance parameters, the tests differ markedly with respect to the choice of nuisance parameters that are being eliminated by plug-in and profiling, respectively. In the case of the DF-GLS tests, the parameter β governing the deterministic component is eliminated using a plug-in approach whereas the parameters (σ^2, γ) governing the scale and serial correlation of the errors are eliminated by profiling. The statistic $\widehat{LR}_T^d$, in contrast, is obtained by plugging in estimators of σ^2 and γ and then profiling out β. Removing (σ^2, γ) by plug-in is computationally convenient and can be motivated by statistical considerations, as σ^2 and γ are nuisance parameters that (unlike β) can treated "as if" they are known when deriving asymptotic local power envelopes. In other words, $\widehat{LR}_T^d$ is obtained by plugging in those nuisance parameters which do not affect asymptotic local power, σ^2 and γ, and maximizing the likelihood fully over the parameter that does influence asymptotic local power, namely β.

In addition to characterizing the asymptotic behavior of the likelihood ratio statistics, the functionals $\max_{\bar{c}\leq 0} \Lambda_c(\bar{c})$ and $\max_{\bar{c}\leq 0} \Lambda_c^\tau(\bar{c})$ can be interpreted as likelihood ratio test statistics in the limiting experiments (in the sense of Le Cam; see e.g. van der Vaart (1998)) associated with maximal invariants for the model (4) when the errors are *i.i.d.* Gaussian. As a consequence, our results shed light on the properties of these limiting experiments by demonstrating that likelihood ratio tests (of $H_0 : c = 0$ vs. $H_1 : c < 0$) are nearly efficient in these experiments, a result which may seem surprising in view of Ploberger (2004, 2008).

In the constant mean case, our model admits locally asymptotically quadratic (LAQ) log likelihood ratios, so the asymptotic optimality of the likelihood ratio tests can be viewed as a testing analog of the efficiency results for maximum likelihood estimators established by Gushchin (1995) and Ploberger and Phillips (2010). We are not aware of any optimality results for estimators in non-LAQ models such as (4) in the linear trend case, but our results suggest that also in some of these situations it may be possible to establish efficiency results on the part of maximum likelihood estimators.

As is well understood from the work of Elliott (1999) and Müller and Elliott (2003), the validity of Theorem 1 (for $c < 0$) and the near efficiency claims made about $\widehat{LR}_T^d$ depend crucially on the assumption that the initial conditions are asymptotically negligible in the sense that $\max(|u_0|, \ldots, |u_{-p}|) = o_p(\sqrt{T})$. Employing a model similar to that of Elliott (1999), Chen and Deo (2009) developed a likelihood ratio test statistic and derived its asymptotic null distribution. It would be of interest to investigate whether that likelihood ratio test enjoys near efficiency properties similar to those obtained herein. We are not aware of any likelihood ratio statistics developed for the more general model of Müller and Elliott (2003). In that model, the unit root testing problem is further complicated by the presence of an unidentified nuisance parameter under the null hypothesis and it would be of interest to explore the possibility of constructing likelihood ratio tests with optimality properties such as an "admissibility at ∞" property reminiscent of Andrews and Ploberger (1995).

A limitation of Theorem 1 is the fact that optimality claims cannot necessarily be made without the assumption of normality, the reason being that relaxing the assumption of normality of the error distribution affects the shape of the power envelope when the errors ε_t are *i.i.d.* (e.g., Rothenberg and Stock (1997) and Jansson (2008)).[5] By basing inference on a Gaussian quasi-likelihood we have made no attempt to achieve full efficiency also under departures from Gaussianity, but it seems plausible that likelihood ratio-type tests with more global optimality properties can be constructed by proceeding as in Jansson (2008). On the other hand, by enlarging the class of models under consideration to contain all error processes for which the weak convergence result $\hat{X}_T \to_d \mathcal{X}_c$ exploited in the proof of Theorem 1 is valid, it should be possible to use the methods of Müller (2010) to establish a semiparametric near efficiency result for the tests developed herein.

Left for future work is an extension of our theoretical results to tests of cointegration.[6] Like the DF tests for unit roots, the cointegration tests due to Johansen (1991) are derived from a conditional likelihood and it would be of interest to know if our qualitative finding about the relative merits of likelihood ratio tests derived from conditional and full likelihoods extends to tests of cointegration.

4. Proof of Theorem 1

Because $\mathcal{L}_T^d(\cdot; \sigma^2, \gamma)$ is invariant under transformations of the form $y_t \to y_t + b'd_t$, we can assume without loss of generality that $\beta = 0$. The proofs of parts (a) and (b) are very similar,

[5] In addition to investigating the effects of non-normality, Rothenberg and Stock (1997, Section 4) obtain large-sample representations of (signed and unsigned) likelihood ratio test statistics in a model without deterministics.

[6] An extension to seasonal unit roots is considered in Jansson and Nielsen (2010).

the latter being slightly more involved, so we omit the details for part (a).

Defining $\hat{d}_{Tt} = \hat{\gamma}_T(1)^{-1} diag(1, 1/\sqrt{T})\hat{\gamma}_T(L) d_t$ and $\hat{y}_{Tt} = \hat{\sigma}_T^{-1}\hat{\gamma}_T(L) y_t$, the test statistic can be written as $\widehat{LR}_T^d = \max_{\bar{c}\leq 0} F(\bar{c}, \hat{X}_T)$, where $\hat{X}_T = (\hat{S}_T, \hat{H}_T, \hat{A}_T, \hat{B}_T)$, $\hat{S}_T = \hat{\sigma}_T^{-2}T^{-1}\sum_{t=2}^T \hat{y}_{T,t-1}\Delta\hat{y}_{Tt}$, $\hat{H}_T = \hat{\sigma}_T^{-2}T^{-2}\sum_{t=2}^T \hat{y}_{T,t-1}^2$,

$$\hat{A}_T = \left(\hat{A}_T(0), \hat{A}_T(1), \hat{A}_T(2)\right), \ \hat{B}_T = \left(\hat{B}_T(0), \hat{B}_T(1), \hat{B}_T(2)\right),$$

$$\hat{A}_T(0) = \sum_{t=1}^T \Delta\hat{d}_{Tt}\Delta\hat{y}_{Tt}, \ \hat{A}_T(1) = \frac{1}{T}\sum_{t=1}^T(\Delta\hat{d}_{Tt}y_{T,t-1} + \hat{d}_{T,t-1}\Delta\hat{y}_{Tt}), \ \hat{A}_T(2) = \frac{1}{T^2}\sum_{t=1}^T \hat{d}_{T,t-1}\hat{y}_{T,t-1},$$

$$\hat{B}_T(0) = \sum_{t=1}^T \Delta\hat{d}_{Tt}\Delta\hat{d}_{Tt}', \ \hat{B}_T(1) = \frac{1}{T}\sum_{t=1}^T(\Delta\hat{d}_{Tt}\hat{d}_{T,t-1}' + \hat{d}_{T,t-1}\Delta\hat{d}_{T,t}'), \ \hat{B}_T(2) = \frac{1}{T^2}\sum_{t=1}^T \hat{d}_{T,t-1}\hat{d}_{T,t-1}',$$

and, with $x = (s, h, a, b)$,

$$F(\bar{c}, x) = \bar{c}s - \frac{1}{2}\bar{c}^2 h + \frac{1}{2}N(\bar{c}, a)' D(\bar{c}, b)^{-1} N(\bar{c}, a) - \frac{1}{2}N(0, a)' D(0, b)^{-1} N(0, a),$$
$$N(\bar{c}, a) = a(0) - \bar{c}a(1) + \bar{c}^2 a(2), \ D(\bar{c}, b) = b(0) - \bar{c}b(1) + \bar{c}^2 b(2).$$

Under the assumptions of Theorem 1 it follows from standard results (e.g., Chan and Wei (1987) and Phillips (1987)), that $\hat{X}_T \to_d \mathcal{X}_c = (\mathcal{S}_c, \mathcal{H}_c, \mathcal{A}_c, \mathcal{B})$, where $(\mathcal{S}_c, \mathcal{H}_c)$ is given in (3),

$$\mathcal{A}_c = \left(\begin{pmatrix} \mathcal{Y} \\ W_c(1) \end{pmatrix}, \begin{pmatrix} 0 \\ W_c(1) \end{pmatrix}, \begin{pmatrix} 0 \\ \int_0^1 rW_c(r)\,dr \end{pmatrix}\right),$$
$$\mathcal{B} = \left(\begin{pmatrix} K & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1/3 \end{pmatrix}\right),$$

$\mathcal{Y}$ is a linear combination of $\varepsilon_1, \ldots, \varepsilon_{p+1}$ (with coefficients depending on γ) independent of $W_c(\cdot)$, and $K = (1 + \sum_{i=1}^p \gamma_i^2) / (1 + \sum_{i=1}^p \gamma_i)^2$. This convergence result implies in particular that $F(\bar{c}, \hat{X}_T) \to_d F(\bar{c}, \mathcal{X}_c) = \Lambda_c^\tau(\bar{c})$ for every $\bar{c} \leq 0$ (under the assumptions of Theorem 1). Moreover, $\Pr(\mathcal{X}_c \in \mathbb{X}) = 1$ for every $c \leq 0$, where $\mathbb{X}$ is the set of all quadruplets (s, h, a, b) satisfying $s > -1/2$, $h > 0$, $b = \mathcal{B}$, and

$$a = \left(\begin{pmatrix} r_1 \\ r_2\sqrt{2(s+1)} \end{pmatrix}, \begin{pmatrix} 0 \\ r_2\sqrt{2(s+1)} \end{pmatrix}, \begin{pmatrix} 0 \\ r_3\sqrt{h/3} \end{pmatrix}\right)$$

for some $(r_1, r_2, r_3) \in \mathbb{R} \times \{-1, 1\} \times (0, 1)$. The result $\widehat{LR}_T^d \to_d \max_{\bar{c}\leq 0} F(\bar{c}, \mathcal{X}_c)$ therefore follows from the CMT if $\max_{\bar{c}\leq 0} F(\bar{c}, \cdot)$ is continuous at every $x_0 \in \mathbb{X}$.

There exists an open set $\tilde{\mathbb{X}} \supseteq \mathbb{X}$ and continuous functions $\{p_i(\cdot)\}$ and $\{q_i(\cdot)\}$ defined on $\tilde{\mathbb{X}}$ such that if $x \in \tilde{\mathbb{X}}$, then $F(\bar{c}, x)$ is a rational polynomial function of $\bar{c}$ of the form $F(\bar{c}, x) = \sum_{i=1}^6 p_i(x)\bar{c}^i / \sum_{i=0}^4 q_i(x)\bar{c}^i$, where $p_6(x) < 0$ and $\sum_{i=0}^4 q_i(x)\bar{c}^i = \det[D(\bar{c}, b)]$ is positive for every $\bar{c} \leq 0$.

Using these facts it follows that for every $x_0 \in \mathbb{X}$ there is a finite constant M and an open set $\tilde{\mathbb{X}}_0 \subseteq \tilde{\mathbb{X}}$ containing x_0 such that $F(\bar{c}, x)$ is negative whenever $(\bar{c}, x) \in (-\infty, -M) \times \tilde{\mathbb{X}}_0$. Because $F(0, x) = 0$, this fact implies that if $x \in \tilde{\mathbb{X}}_0$, then

$$\max_{\bar{c}\leq 0} F(\bar{c}, x) = \max_{-M\leq\bar{c}\leq 0} F(\bar{c}, x). \tag{5}$$

Because $F(\cdot)$ is continuous on $[-M,0]\times\tilde{\mathbb{X}}_0$ and $[-M,0]$ is compact, it follows from the theorem of the maximum (e.g., Stokey and Lucas (1989, Theorem 3.6)) that $\max_{-M\leq\bar{c}\leq 0} F(\bar{c},\cdot)$ is continuous on $\tilde{\mathbb{X}}_0$. The desired continuity property of $\max_{\bar{c}\leq 0} F(\bar{c},\cdot)$ follows from this result and the representation (5).

References

ANDREWS, D. W. K., AND 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.

BOSWIJK, H. P. (1998): "Likelihood Ratio Tests for a Unit Root and Cointegration with a Linear Trend," Working paper, University of Amsterdam.

CHAN, N. H., AND C. Z. WEI (1987): "Asymptotic Inference for Nearly Nonstationary AR(1) Processes," *Annals of Statistics*, 15, 1050–1063.

CHANG, Y., AND J. Y. PARK (2002): "On the Asymptotics of ADF Tests for Unit Roots," *Econometric Reviews*, 21, 431–447.

CHEN, W. W., AND R. S. DEO (2009): "The Restricted Likelihood Ratio Test at the Boundary in Autoregressive Series," *Journal of Time Series Analysis*, 30, 618–630.

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.

——— (1981): "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root," *Econometrica*, 49, 1057–1072.

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, AND J. H. STOCK (1996): "Efficient Tests for an Autoregressive Unit Root," *Econometrica*, 64, 813–836.

GUSHCHIN, A. A. (1995): "Asymptotic Optimality of Parameter Estimators under the LAQ Condition," *Theory of Probability and its Applications*, 40, 261–272.

HALDRUP, N., AND M. JANSSON (2006): "Improving Size and Power in Unit Root Testing," in *Palgrave Handbook of Econometrics, Volume 1: Econometric Theory*, ed. by T. C. Mills, and K. Patterson. New York: Palgrave Macmillan, 252-277.

JANSSON, M. (2008): "Semiparametric Power Envelopes for Tests of the Unit Root Hypothesis," *Econometrica*, 76, 1103–1142.

JANSSON, M., AND M. Ø. NIELSEN (2010): "Nearly efficient likelihood ratio tests for seasonal unit roots," Forthcoming in *Journal of Time Series Econometrics.*

JOHANSEN, S. (1991): "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models," *Econometrica*, 59, 1551–1580.

Lehmann, E. L. (2006): "On Likelihood Ratio Tests," in *Optimality: The Second Erich L. Lehmann Symposium*, ed. by J. Rojo. Beachwood, OH: Institute of Mathematical Statistics, 1-8.

Müller, U. K. (2010): "Efficient Tests under a Weak Convergence Assumption," Forthcoming in *Econometrica*.

Müller, U. K., and G. Elliott (2003): "Tests for Unit Root and the Initial Condition," *Econometrica*, 71, 1269–1286.

Ng, S., and P. Perron (2001): "Lag Length Selection and the Construction of Unit Root Tests with Good Size and Power," *Econometrica*, 69, 1519–1554.

Phillips, P. C. B. (1987): "Towards a Unified Asymptotic Theory for Autoregression," *Biometrika*, 74, 535–547.

Ploberger, W. (2004): "A Complete Class of Tests When the Likelihood is Locally Asymptotically Quadratic," *Journal of Econometrics*, 118, 67–94.

——— (2008): "Admissible and Nonadmissible Tests in Unit-Root-Like Situations," *Econometric Theory*, 24, 15–42.

Ploberger, W., and P. C. B. Phillips (2010): "Optimal Estimation under Nonstandard Conditions," Cowles Foundation Discussion Paper No. 1748.

Rothenberg, T. J., and J. H. Stock (1997): "Inference in a Nearly Integrated Autoregressive Model with Nonnormal Innovations," *Journal of Econometrics*, 80, 269–286.

Stock, J. H. (1994): "Unit Roots, Structural Breaks and Trends," in *Handbook of Econometrics, Volume 4*, ed. by R. F. Engle, and D. L. McFadden. New York: North Holland, 2739-2841.

Stokey, N. L., and R. E. Lucas (1989): *Recursive Methods in Economic Dynamics*. Cambridge: Harvard University Press.

van der Vaart, A. W. (1998): *Asymptotic Statistics*. New York: Cambridge University Press.