SCHOOL OF ECONOMICS AND MANAGEMENT
FACULTY OF SOCIAL SCIENCES
AARHUS UNIVERSITY

CREATES
Center for Research in Econometric Analysis of Time Series

CREATES Research Paper 2009-37

Nearly Efficient Likelihood Ratio Tests of the Unit Root Hypothesis

Michael Jansson and Morten Ørregaard Nielsen

School of Economics and Management
Aarhus University
Bartholins Allé 10, Building 1322, DK-8000 Aarhus C
Denmark

Nearly Efficient Likelihood Ratio Tests of the Unit Root Hypothesis*

Michael Jansson† — UC Berkeley and CREATES
Morten Ørregaard Nielsen — Queen's University and CREATES

August 31, 2009

Abstract. Seemingly absent from the arsenal of currently available "nearly efficient" testing procedures for the unit root hypothesis, i.e. tests whose local asymptotic power functions are indistinguishable from the Gaussian power envelope, is a test admitting a (quasi-)likelihood ratio interpretation. We show that the likelihood ratio unit root test derived in a Gaussian AR(1) model with standard normal innovations is nearly efficient in that model. Moreover, these desirable properties carry over to more complicated models allowing for serially correlated and/or non-Gaussian innovations.

KEYWORDS: 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 local asymptotic power functions are indistinguishable from the Gaussian power envelope, can be constructed. Subsequent research (e.g., Ng and Perron (2001)) has enlarged the class of tests whose local asymptotic power functions are indistinguishable from the Gaussian power envelope, but seemingly absent from the arsenal of currently available nearly efficient testing

*We are grateful to Niels Haldrup, Søren Johansen, Tom Rothenberg, and participants at 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), Phillips and Xiao (1998), and Haldrup and Jansson (2006).

procedures is a test admitting a (quasi-)likelihood ratio interpretation. The purpose of this paper is to propose such a test.

In models with an unknown mean and/or a linear trend, the class of tests attaining the power envelopes does not contain the Dickey and Fuller (1979, 1981, henceforth DF) tests (or their modifications, such as Phillips (1987a) and Phillips and Perron (1988)). 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. 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.

Specifically, we consider the leading special case of the model in ERS, namely a Gaussian AR(1) model with standard normal innovations and with presample observations assumed to be equal to their expected values. We examine the local power properties of the likelihood ratio unit root test derived in this model and show that it is nearly efficient. Moreover, these desirable properties are found to be shared by computationally simple likelihood ratio-type tests in models allowing for serially correlated non-Gaussian innovations and presamble observations that deviate from their expected values. More details will be provided below. We remark at the outset that the new tests are related to, but distinct from, the DF-GLS tests of ERS, even asymptotically. Again, more details will be provided below.

The remainder of the paper is organized as follows. Section 2 contains our results on the likelihood ratio test for a unit root, with additional discussion and conclusions in Section 3. Proofs of our results are provided in Section 4.

2. The Likelihood Ratio Test for a Unit Root

Section 2.1 studies likelihood ratio tests of the unit root hypothesis in the simplest possible setting, namely the zero-mean Gaussian AR(1) model. Section 2.2 accommodates a deterministic component, while extensions to models with serially correlated and/or non-Gaussian errors are considered in Section 2.3.

2.1. No Deterministic Component.

Suppose $\{y_t : 1 \leq t \leq T\}$ is an observed univariate time series generated by the zero-mean Gaussian AR(1) model

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

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

The likelihood ratio test associated with the unit root testing problem

$$H_0 : \rho = 1 \qquad \text{vs.} \qquad H_1 : \rho < 1 \tag{2}$$

rejects for large values of

$$LR_T = \max_{\bar{\rho} \leq 1} L_T(\bar{\rho}) - L_T(1),$$

where $L_T(\rho) = -\sum_{t=1}^{T}(y_t - \rho y_{t-1})^2/2$ is the log likelihood function.

Defining $(S_T, H_T) = \left(T^{-1}\sum_{t=2}^{T} y_{t-1}\Delta y_t, T^{-2}\sum_{t=2}^{T} y_{t-1}^2\right)$, 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

$$LR_T = \max_{\bar{c} \leq 0}\left[\bar{c} S_T - \frac{1}{2}\bar{c}^2 H_T\right]. \tag{3}$$

The large sample behavior of the pair (S_T, H_T) is well understood. Under local-to-unity asymptotics, with $c = T(\rho - 1)$ held fixed as $T \to \infty$,

$$(S_T, H_T) \to_d (\mathcal{S}_c, \mathcal{H}_c) = \left[\int_0^1 W_c(r)\, dW_c(r), \int_0^1 W_c(r)^2\, dr\right], \tag{4}$$

where $W_c(r) = \int_0^r \exp[c(r-s)]\, dW(s)$ and $W(\cdot)$ is a standard Wiener process, e.g., Chan and Wei (1987) and Phillips (1987b). Theorem 1 gives the corresponding result about the local-to-unity asymptotic behavior of the likelihood ratio statistic LR_T.

Theorem 1. *If $\{y_t\}$ is generated by* (1) *and $c = T(\rho - 1)$ is held fixed as $T \to \infty$, then $LR_T \to_d \max_{\bar{c} \leq 0} \Lambda_c(\bar{c})$, where*

$$\Lambda_c(\bar{c}) = \bar{c}\mathcal{S}_c - \frac{1}{2}\bar{c}^2\mathcal{H}_c.$$

Theorem 1 follows from (3), (4), and the continuous mapping theorem (CMT) applied to the functional $f(s,h)=\min(0,s)^2/h$. Specifically,

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

where the second and third equalities use simple facts about quadratic functions. The implicit characterization of the weak limit of LR_T as $\max_{\bar{c}\leq 0}\Lambda_c\left(\bar{c}\right)$ is employed in anticipation of Theorem 2(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 $\max_{\bar{c}\leq 0}\Lambda_c\left(\bar{c}\right)$ enables us to address the asymptotic optimality properties of the likelihood ratio test. For any α less than $\Pr\left[\mathcal{S}_0\leq 0\right]\approx 0.6827$, the (asymptotic) size α likelihood ratio test rejects when LR_T exceeds $k_{LR}\left(\alpha\right)$, where $k_{LR}\left(\alpha\right)$ satisfies $\Pr\left[\max_{\bar{c}\leq 0}\Lambda_0\left(\bar{c}\right)>k_{LR}\left(\alpha\right)\right]=\alpha$. For any such α, the local asymptotic power function (with argument $c\leq 0$) associated with the size α likelihood ratio test is given by $\Pr\left[\max_{\bar{c}\leq 0}\Lambda_c\left(\bar{c}\right)>k_{LR}\left(\alpha\right)\right]$, and 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 assumptions of Theorem 1. 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 local asymptotic power function is indistinguishable from the Gaussian power envelope, which (with argument $c\leq 0$) for tests with asymptotic level α is given by $\Pr\left[\Lambda_c\left(\bar{c}\right)>k_{\bar{c}}\left(\alpha\right)\right]|_{\bar{c}=c}$, where $k_{\bar{c}}\left(\alpha\right)$ satisfies $\Pr\left[\Lambda_0\left(\bar{c}\right)>k_{\bar{c}}\left(\alpha\right)\right]=\alpha$.

2.2. Deterministics. 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). It is therefore of interest to explore the local asymptotic power properties of likelihood ratio tests in such models. Accordingly, suppose $\{y_t:1\leq t\leq T\}$ is generated by the Gaussian AR(1) model

$$y_t=\beta' d_t+u_t,\qquad u_t=\rho u_{t-1}+\varepsilon_t, \tag{5}$$

where $d_t=1$ or $d_t=(1,t)'$, β is an unknown parameter, $u_0=0$, and $\varepsilon_t\sim i.i.d.\mathcal{N}\left(0,1\right)$.

In this case, the log likelihood function $L_T^d\left(\cdot\right)$ is conveniently expressed as

$$L_T^d\left(\rho,\beta\right)=-\frac{1}{2}\left(Y_\rho-D_\rho\beta\right)'\left(Y_\rho-D_\rho\beta\right),$$

where $Y_\rho=\left(y_1,y_2-\rho y_1,\ldots,y_T-\rho y_{T-1}\right)'$ and $D_\rho=\left(d_1,d_2-\rho d_1,\ldots,d_T-\rho d_{T-1}\right)'$.

The likelihood ratio test associated with the unit root testing problem (2) rejects for large values of

$$\begin{aligned} LR_T^d &= \max_{\bar{\rho}\leq 1,\beta} L_T^d(\bar{\rho},\beta) - \max_\beta L_T^d(1,\beta) \\ &= \max_{\bar{\rho}\leq 1} \mathcal{L}_T^d(\bar{\rho}) - \mathcal{L}_T^d(1), \end{aligned}$$

where

$$\mathcal{L}_T^d(\rho) = \max_\beta L_T^d(\rho,\beta) = -\frac{1}{2}Y_\rho' Y_\rho + \frac{1}{2}\left(Y_\rho' D_\rho\right)\left(D_\rho' D_\rho\right)^{-1}\left(D_\rho' Y_\rho\right)$$

is the profile log likelihood function obtained by maximizing $L_T^d(\rho,\beta)$ with respect to the nuisance parameter β. Unlike $L_T(\rho) = -Y_\rho' Y_\rho/2$, the profile log likelihood function $\mathcal{L}_T^d(\rho)$ depends on ρ in a complicated way and no closed form expression for LR_T^d will be available in general. This feature complicates, but does not prohibit, the derivation of the asymptotic distribution of LR_T^d under local-to-unity asymptotics.

Theorem 2. *Suppose $\{y_t\}$ is generated by (5) and suppose $c = T(\rho - 1)$ is held fixed as $T \to \infty$. Then:*

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

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

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

The proof of Theorem 2 proceeds by showing that the likelihood ratio statistic can be written as $LR_T^d = \max_{\bar{c}\leq 0} F(\bar{c}, X_T)$ for some function $F(\cdot)$ and some random vector X_T, where the latter satisfies a convergence property of the form $X_T \to_d \mathcal{X}_c$ under the assumptions of Theorem 2 and the functional $\max_{\bar{c}\leq 0} F(\bar{c},\cdot)$ is continuous on a set $\mathbb{X}$ satisfying $\Pr[\mathcal{X}_c \in \mathbb{X}] = 1$ (for every $c \leq 0$).

Because the profile log likelihood function $\mathcal{L}_T^d(\cdot)$ is invariant under transformations of the form $y_t \to y_t + b'd_t$, so is LR_T^d (and any other test statistic that can expressed as a functional of $\mathcal{L}_T^d(\cdot)$). It therefore makes sense to compare the local asymptotic power properties of the tests based on 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 without deterministics. Similarly, it follows from Theorem 2(a) that the local asymptotic power of the constant mean likelihood ratio test coincides with the local asymptotic 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 ABOUT HERE

The local asymptotic power function (with argument $c \leq 0$) of the size α linear trend likelihood ratio test is given by $\Pr\left[\max_{\bar{c}\leq 0} \Lambda_c^\tau(\bar{c}) > k_{LR}^\tau(\alpha)\right],$ where $k_{LR}^\tau(\alpha)$ satisfies $\Pr\left[\max_{\bar{c}\leq 0} \Lambda_0^\tau(\bar{c}) > k_{LR}^\tau(\alpha)\right] = \alpha$. Figure 1 plots this function for $\alpha = 0.05$. Also plotted in Figure 1 is the Gaussian power envelope, which (for size α tests) is given by $\Pr\left[\Lambda_c^\tau(\bar{c}) > k_{\bar{c}}^\tau(\alpha)\right]\big|_{\bar{c}=c}$, where $k_{\bar{c}}^\tau(\alpha)$ satisfies $\Pr\left[\Lambda_0^\tau(\bar{c}) > k_{\bar{c}}^\tau(\alpha)\right] = \alpha$. As in the no deterministics and constant mean cases, the local asymptotic 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.

The near optimality properties of the likelihood ratio test are shared by two related classes of tests proposed by ERS, namely the point optimal tests and DF-GLS tests. The point optimal test statistics are of the form

$$P_T^d(\bar{c}_{ERS}) = \mathcal{L}_T^d\left(1 + T^{-1}\bar{c}_{ERS}\right) - \mathcal{L}_T^d(1),$$

where $\bar{c}_{ERS}$ is a negative constant. By construction, these tests attain 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 $\hat{c}_{LR} = \arg\max_{\bar{c}\leq 0} \mathcal{L}_T^d(1 + T^{-1}\bar{c})$, the likelihood ratio test statistic can be expressed as $LR_T^d = P_T^d(\hat{c}_{LR})$. Because $\hat{c}_{LR}$ is random even in the limit, the likelihood ratio test cannot be interpreted as an (asymptotically) point optimal test.

In the model considered in this subsection, the DF-GLS test is asymptotically equivalent to the test based on the test statistic

$$\hat{\tau}_T^{DF-GLS}(\bar{c}_{ERS}) = \max_{\rho\leq 1} L_T^d\left(\rho, \hat{\beta}_T(\bar{c}_{ERS})\right) - L_T^d\left(1, \hat{\beta}_T(\bar{c}_{ERS})\right),$$

where $\bar{c}_{ERS}$ is a negative constant and $\hat{\beta}_T(\bar{c}_{ERS})$ is a plug-in estimator of β given by

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

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 2(a), the likelihood ratio test is asymptotically equivalent to the DF-GLS test in the constant mean case, 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 2(b), the asymptotic properties of $\hat{\tau}_T^{DF-GLS}(\bar{c}_{ERS})$ depend on $\bar{c}_{ERS}$ and the likelihood ratio test cannot be interpreted as being asymptotically equivalent to a DF-GLS test in the linear trend case.

2.3. Serial Correlation and Unknown Error Distribution. The results of the preceding sections are mainly of theoretical interest, the reason being that the assumptions made about the errors $\{u_t\}$ are implausible in many applications. Specifically, the AR(1) specification and the assumption that the innovations $\{\varepsilon_t\}$ are generated by a known distribution are too restrictive for comfort and it is of interest to relax them. To that end, suppose $\{y_t : 1 \leq t \leq T\}$ is generated by the model

$$y_t = \beta' d_t + u_t, \qquad (1-\rho L)\,\gamma(L)\,u_t = \varepsilon_t, \tag{6}$$

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) order p satisfying $\min_{|z|\leq 1} |\gamma(z)| > 0$, the initial conditions are $u_0 = \ldots = u_{-p} = 0$, and the ε_t are *i.i.d.* errors from a distribution with mean zero and unknown variance σ^2.

It follows from ERS that if the errors are Gaussian, then the power envelopes computed under the "as if" assumption that σ^2 and the coefficients of $\gamma(L)$ are known are attainable and coincide with the envelopes computed under the assumption that $\sigma^2 = 1$ and $\gamma(L) = 1$. Moreover, the Gaussian power envelopes can be attained by means of procedures based on the Gaussian quasi-likelihood also when the error distribution is non-Gaussian.[2]

We now discuss how to modify the test statistic LR_T^d in such a way that the modified tests attain the Gaussian power envelope in the more general model considered in this subsection. The Gaussian quasi-log likelihood function can be expressed as

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

where, setting $y_0 = \ldots = y_{-p} = 0$ and $d_0 = \ldots = d_{-p} = 0$, $Y_{\rho,\gamma}$ and $D_{\rho,\gamma}$ are matrices

[2] Relaxing the assumption of normality of the error distribution affects the shape of the power envelope in the model considered here, see Jansson (2008). To conserve space we make no attempt to achieve full efficiency also under departures from Gaussianity.

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 likelihood ratio-type test statistic of the form

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

where

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

while $\hat{\sigma}_T^2$ and $\hat{\gamma}_T$ are plug-in estimators of σ^2 and $\gamma = \left(\gamma_1, \ldots, \gamma_p\right)'$, respectively.

Being based on a plug-in version of $\mathcal{L}_T^d\left(\rho; \sigma^2, \gamma\right)$, the statistic $\widehat{LR}_T^d$ is straightforward to construct, requiring only maximization of $\mathcal{L}_T^d\left(\rho; \hat{\sigma}_T^2, \hat{\gamma}_T\right)$ with respect to the scalar parameter ρ. Based on the findings of ERS summarized in the second paragraph of this subsection, the functional form of $\widehat{LR}_T^d$ is motivated by two conjectures to be proved below. First, we base our formulation on a Gaussian quasi-likelihood because we wish to attain the Gaussian power envelopes. Second, because the unknown parameters σ^2 and $\gamma(L)$ can be treated as if they are known when deriving Gaussian power envelopes, it should be possible to replace σ^2 and $\gamma(L)$ by reasonable estimators in $L_T^d\left(\rho, \beta; \sigma^2, \gamma\right)$ without sacrificing asymptotic efficiency.

In particular, the test statistic $\widehat{LR}_T^d$ is asymptotically equivalent to LR_T^d under the assumptions of Theorem 2 for consistent estimators of σ^2 and γ. Moreover, because the Gaussian power envelopes are invariant with respect to σ^2 and $\gamma(L)$, and can be attained by procedures based on the Gaussian quasi-likelihood also when the error distribution is non-Gaussian, $\widehat{LR}_T^d$ is asymptotically pivotal in the sense that if $\{y_t\}$ is generated by (6) then its local-to-unity asymptotic distribution depends only on the local-to-unity parameter $c = T(\rho - 1)$. The next result confirms these conjectures under fairly minimal assumptions about the estimators $\hat{\sigma}_T^2$ and $\hat{\gamma}_T$.

Theorem 3. *Suppose $\{y_t\}$ is generated by (6), $c = T(\rho - 1)$ is held fixed as $T \to \infty$, and $\left(\hat{\sigma}_T^2, \hat{\gamma}_T\right) \to_p \left(\sigma^2, \gamma\right)$. Then:*

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

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

Simulated critical values $k_{LR}^{\tau}(\alpha)$ associated with $\widehat{LR}_T^d$ are reported in Table 1.

TABLE 1 ABOUT HERE

The consistency requirement on the estimators $\hat{\sigma}_T^2$ and $\hat{\gamma}_T$ is mild. For instance, it is met by

$$\hat{\sigma}_T^2 = \frac{1}{T-p-1}\sum_{t=p+2}^{T}\left(\Delta y_t - \hat{\kappa}_T' Z_t\right)^2, \qquad \hat{\gamma}_T = (0, I_p)\,\hat{\kappa}_T,$$

where

$$\hat{\kappa}_T = \left(\sum_{t=p+2}^{T} Z_t Z_t'\right)^{-1}\left(\sum_{t=p+2}^{T} Z_t \Delta y_t\right), \qquad Z_t = (1, \Delta y_{t-1}, \ldots, \Delta y_{t-p})'.$$

The function $L_T^d(\rho, \beta; \sigma^2, \gamma)$ upon which the statistic $\widehat{LR}_T^d$ is based is a quasi-likelihood in the usual sense that it is constructed under an "as if" assumption of normality on the part of the innovations ε_t. In addition, it is a quasi-likelihood in the sense that the assumption $u_0 = \ldots = u_{-p} = 0$ about the initial values can be interpreted as an "as if" assumption that can be relaxed without invalidating the asymptotic results of this section. Specifically, Theorem 3 remains valid (as does the method of its proof) under the somewhat weaker assumption that $\max(|u_0|, \ldots, |u_{-p}|) = o_p(\sqrt{T})$. On the other hand, different distributional results and hence different local power properties will generally be obtained if $\max(|u_0|, \ldots, |u_{-p}|) \neq o_p(\sqrt{T})$, a feature that is shared by the DF-GLS tests of ERS.[3]

3. Discussion and Conclusions

In this paper we have shown that the likelihood ratio unit root test derived in a Gaussian AR(1) model with standard normal innovations is nearly efficient in that model. Moreover, these desirable properties are found to be shared by computationally simple likelihood ratio-type tests in models allowing for serially correlated and/or non-Gaussian innovations.

Although the tests based on $\widehat{LR}_T^d$ are virtually identical to the DF-GLS tests of ERS in terms of local asymptotic 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

[3]For a more elaborate discussion of the role played by the initial values assumptions, see e.g. ERS (p. 819), Müller and Elliott (2003), and Harvey, Leybourne, and Taylor (2009).

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 local asymptotic power envelopes. In other words, $\widehat{LR}_T^d$ is obtained by plugging in those nuisance parameters which do not affect local asymptotic power, σ^2 and γ, and maximizing the likelihood fully over those nuisance parameters which do influence local asymptotic 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 LeCam) associated with maximal invariants for the model (5), see e.g. van der Vaart (1998). 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.

It would be of interest to conduct a thorough investigation of the finite sample properties of $\widehat{LR}_T^d$, including its sensitivity with respect to the choice of estimators $\hat{\sigma}_T^2$ and $\hat{\gamma}_T$. To conserve space we do not carry out such an investigation here, but it seems worth remarking that preliminary Monte Carlo evidence suggests that $\widehat{LR}_T^d$ performs well compared to DF-GLS when implemented with the abovementioned choice of $\left(\hat{\sigma}_T^2, \hat{\gamma}_T\right)$. Also left for future work is an extension of our theoretical results to tests of cointegration. Like the DF tests for unit roots, the cointegration tests due to Johansen (1988, 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. Finally, we remark that a possible advantage of the likelihood ratio tests analyzed here is that extensions of our results to, e.g., alternative distributional assumptions or seasonal unit roots, appear conceptually much simpler than a corresponding generalization of the DF-GLS tests.

4. Proofs

4.1. Proof of Theorem 2. Because $\mathcal{L}_T^d(\cdot)$ is invariant under transformations of the form $y_t \to y_t + b'd_t$, we can assume without loss of generality that $\beta = 0$. The proofs of parts (a) and (b) are very similar, the latter being slightly more involved, so to conserve space we omit the details for part (a).

Setting $d_0 = 0$ and $y_0 = 0$ and defining $\tilde{d}_{Tt} = diag(1, 1/\sqrt{T})d_t$, the linear trend likelihood ratio statistic can be written as $LR_T^d = \max_{\bar{c}\leq 0} F(\bar{c}, X_T)$, where

$$X_T = (S_T, H_T, A_T, B_T),$$

$$A_T = [A_T(0), A_T(1), A_T(2)],$$

$$B_T = [B_T(0), B_T(1), B_T(2)],$$

for $A_T(0) = \sum_{t=1}^T \Delta\tilde{d}_{Tt}\Delta y_t$, $A_T(1) = \frac{1}{T}\sum_{t=1}^T(\Delta\tilde{d}_{Tt}y_{t-1} + \tilde{d}_{T,t-1}\Delta y_t)$, $A_T(2) = \frac{1}{T^2}\sum_{t=1}^T \tilde{d}_{T,t-1}y_{t-1}$, $B_T(0) = \sum_{t=1}^T \Delta\tilde{d}_{Tt}\Delta\tilde{d}'_{Tt}$, $B_T(1) = \frac{1}{T}\sum_{t=1}^T(\Delta\tilde{d}_{Tt}\tilde{d}'_{T,t-1}+\tilde{d}_{T,t-1}\Delta\tilde{d}'_{T,t})$, $B_T(2) = \frac{1}{T^2}\sum_{t=1}^T \tilde{d}_{T,t-1}\tilde{d}'_{T,t-1}$, and

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

with

$$N(\bar{c}, a) = N[\bar{c}, a(0), a(1), a(2)] = a(0) - \bar{c}a(1) + \bar{c}^2 a(2),$$

$$D(\bar{c}, b) = D[\bar{c}, b(0), b(1), b(2)] = b(0) - \bar{c}b(1) + \bar{c}^2 b(2).$$

It follows from standard results, e.g., Chan and Wei (1987) and Phillips (1987b), that $X_T \to_d \mathcal{X}_c = (\mathcal{S}_c, \mathcal{H}_c, \mathcal{A}_c, \mathcal{B})$ under the assumptions of Theorem 2, where

$$\mathcal{A}_c = \left[\begin{pmatrix} \mathcal{Y} \\ W_c(1) \end{pmatrix}, \begin{pmatrix} 0 \\ W_c(1) \end{pmatrix}, \begin{pmatrix} 0 \\ \int_0^1 rW_c(r)\,dr \end{pmatrix}\right],$$

$$\mathcal{B} = \left[\begin{pmatrix} K & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1/3 \end{pmatrix}\right],$$

$\mathcal{Y} \sim \varepsilon_1$ is a random variable independent of $W_c(\cdot)$, and $K = 1$ is a positive constant. This convergence result implies in particular that $F(\bar{c}, X_T) \to_d F(\bar{c}, \mathcal{X}_c) = \Lambda_c^\tau(\bar{c})$ for every $\bar{c} \leq 0$ (under the assumptions of Theorem 2). 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 desired result 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 $\{p_i(\cdot) : 1 \leq i \leq 6\}$ and $\{q_i(\cdot) : 0 \leq i \leq 4\}$ 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

$$F(\bar{c}, x) = \frac{\sum_{i=1}^{6} p_i(x)\,\bar{c}^i}{\sum_{i=0}^{4} q_i(x)\,\bar{c}^i},$$

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 is not hard to show 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

$$\max_{\bar{c} \leq 0} F(\bar{c}, x) = \max_{-M \leq \bar{c} \leq 0} F(\bar{c}, x). \tag{7}$$

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 (7).

4.2. Proof of Theorem 3. As in the proof of Theorem 2, we assume without loss of generality that $\beta = 0$ and give only the proof of part (b).

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 linear trend likelihood ratio statistic can be written as $\widehat{LR}_T^d = \max_{\bar{c} \leq 0} F(\bar{c}, \hat{X}_T)$, where

$$\hat{X}_T = \left(\hat{S}_T, \hat{H}_T, \hat{A}_T, \hat{B}_T\right),$$

$$(\hat{S}_T, \hat{H}_T) = \left(\frac{1}{T\hat{\sigma}_T^2} \sum_{t=2}^{T} \hat{y}_{T,t-1}\Delta\hat{y}_{Tt}, \frac{1}{T^2\hat{\sigma}_T^2} \sum_{t=2}^{T} \hat{y}_{T,t-1}^2 \right),$$

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

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

for $\hat{A}_T(0) = \sum_{t=1}^T \Delta\hat{d}_{Tt}\Delta\hat{y}_{Tt}$, $\hat{A}_T(1) = \frac{1}{T}\sum_{t=1}^T(\Delta\hat{d}_{Tt}y_{T,t-1} + \hat{d}_{T,t-1}\Delta\hat{y}_{Tt})$, $\hat{A}_T(2) = \frac{1}{T^2}\sum_{t=1}^T \hat{d}_{T,t-1}\hat{y}_{T,t-1}$, $\hat{B}_T(0) = \sum_{t=1}^T \Delta\hat{d}_{Tt}\Delta\hat{d}_{Tt}'$, $\hat{B}_T(1) = \frac{1}{T}\sum_{t=1}^T(\Delta\hat{d}_{Tt}\hat{d}_{T,t-1}' + \hat{d}_{T,t-1}\Delta\hat{d}_{T,t}')$, $\hat{B}_T(2) = \frac{1}{T^2}\sum_{t=1}^T \hat{d}_{T,t-1}\hat{d}_{T,t-1}'$, and $F(\cdot)$ is defined as in the proof of Theorem 2.

Theorem 3 now follows from the proof of Theorem 2 because routine calculations can be used to show that $\hat{X}_T \to_d \hat{\mathcal{X}}_c$, where $\hat{\mathcal{X}}_c$ is defined as $\mathcal{X}_c$ in the proof of Theorem 2, except that in this case the distribution of $\mathcal{Y}$ is that of a linear combination of $\varepsilon_1, \ldots, \varepsilon_{p+1}$ (with coefficients depending on $\gamma_1, \ldots, \gamma_p$) and $K = \left(1 + \gamma_1^2 + \ldots + \gamma_p^2\right) / \left(1 + \gamma_1 + \ldots + \gamma_p\right)^2$.

References

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.

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

Elliott, G., T. J. Rothenberg, and J. H. Stock (1996): "Efficient Tests for an Autoregressive Unit Root," *Econometrica*, 64, 813–836.

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.

Harvey, D. I., S. J. Leybourne, and 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. (2008): "Semiparametric Power Envelopes for Tests of the Unit Root Hypothesis," *Econometrica*, 76, 1103–1142.

Johansen, S. (1988): "Statistical Analysis of Cointegration Vectors," *Journal of Economic Dynamics and Control*, 12, 231–254.

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

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. (1987a): "Time Series Regression with a Unit Root," *Econometrica*, 55, 277–301.

——— (1987b): "Towards a Unified Asymptotic Theory for Autoregression," *Biometrika*, 74, 535–547.

Phillips, P. C. B., and P. Perron (1988): "Testing for a Unit Root in Time Series Regression," *Biometrika*, 75, 335–346.

Phillips, P. C. B., and Z. Xiao (1998): "A Primer on Unit Root Testing," *Journal of Economic Surveys*, 12, 423–469.

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.

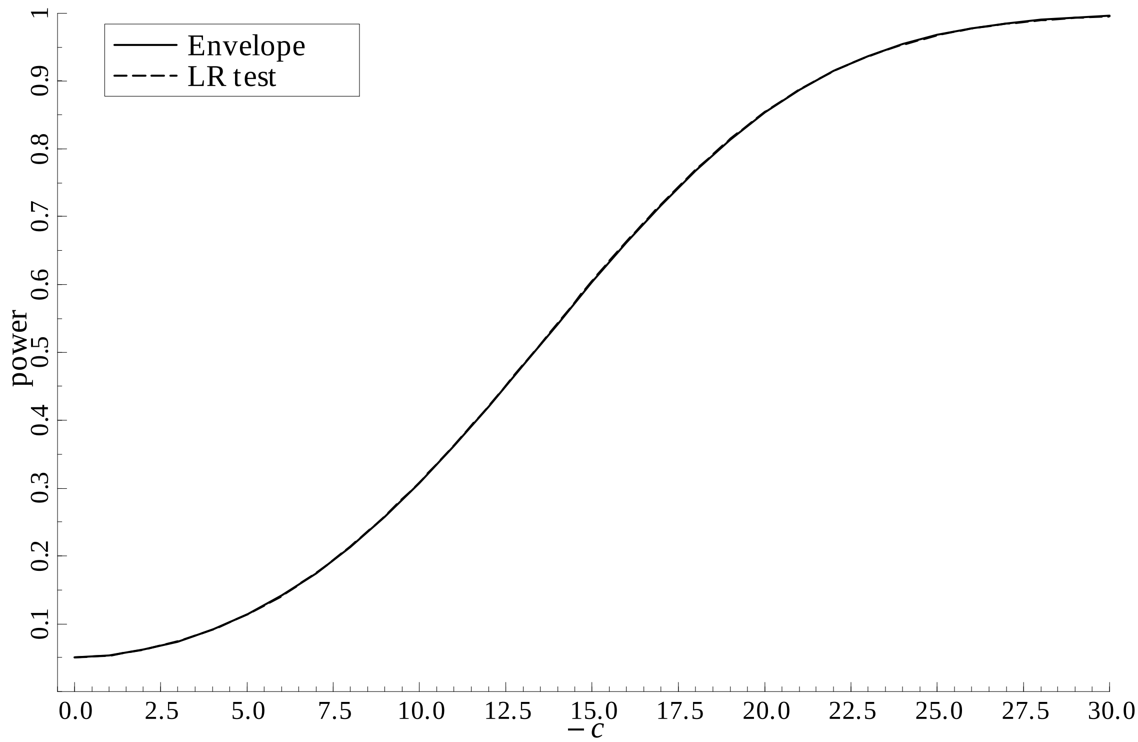
van der Vaart, A. W. (1998): *Asymptotic Statistics*. New York: Cambridge University Press.

Table 1: Quantiles of the distribution $\max_{\bar{c}\leq 0} \Lambda_0^\tau(\bar{c})$

T	80%	85%	90%	95%	97.5%	99%	99.5%	99.9%
100	2.5041	2.8564	3.3411	4.1401	4.9067	5.8850	6.5997	8.1734
250	2.4702	2.8155	3.2928	4.0947	4.8783	5.8927	6.6522	8.3838
500	2.4595	2.8014	3.2750	4.0678	4.8479	5.8649	6.6254	8.3574
1000	2.4550	2.7951	3.2660	4.0546	4.8289	5.8352	6.5867	8.3097
∞	2.4524	2.7925	3.2616	4.0481	4.8192	5.8232	6.5716	8.2905

Note: Entries for finite T are simulated quantiles of LR_T^d with $\varepsilon_t \sim i.i.d.\mathcal{N}(0,1)$, $t = 1,\ldots,T$. Entries for $T = \infty$ are simulated quantiles of $\max_{\bar{c}\leq 0} \Lambda_c^\tau(\bar{c})$, where Wiener processes are approximated by $10,000$ discrete steps with standard Gaussian white noise innovations. All entries are based on ten million Monte Carlo replications.

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 one million Monte Carlo replications, where Wiener processes were approximated by $T = 10,000$ discrete steps with standard Gaussian white noise innovations.

Research Papers 2009

2009-23: Michael Frömmel and Robinson Kruse: Interest rate convergence in the EMS prior to European Monetary Union

2009-24: Dominique Guégan: A Meta-Distribution for Non-Stationary Samples

2009-25: Ole E. Barndorff-Nielsen and Almut E. D. Veraart: Stochastic volatility of volatility in continuous time

2009-26: Tim Bollerslev and Viktor Todorov: Tails, Fears and Risk Premia

2009-27: Kim Christensen, Roel Oomen and Mark Podolskij: Realised Quantile-Based Estimation of the Integrated Variance

2009-28: Takamitsu Kurita, Heino Bohn Nielsen and Anders Rahbek: An I(2) Cointegration Model with Piecewise Linear Trends: Likelihood Analysis and Application

2009-29: Martin M. Andreasen: Stochastic Volatility and DSGE Models

2009-30: Eduardo Rossi and Paolo Santucci de Magistris: Long Memory and Tail dependence in Trading Volume and Volatility

2009-31: Eduardo Rossi and Paolo Santucci de Magistris: A No Arbitrage Fractional Cointegration Analysis Of The Range Based Volatility

2009-32: Alessandro Palandri: The Effects of Interest Rate Movements on Assets' Conditional Second Moments

2009-33: Peter Christoffersen, Redouane Elkamhi, Bruno Feunou and Kris Jacobs: Option Valuation with Conditional Heteroskedasticity and Non-Normality

2009-34: Peter Christoffersen, Steven Heston and Kris Jacobs: The Shape and Term Structure of the Index Option Smirk: Why Multifactor Stochastic Volatility Models Work so Well

2009-35: Peter Christoffersen, Jeremy Berkowitz and Denis Pelletier: Evaluating Value-at-Risk Models with Desk-Level Data

2009-36: Tom Engsted and Thomas Q. Pedersen: The dividend-price ratio does predict dividend growth: International evidence

2009-37: Michael Jansson and Morten Ørregaard Nielsen: Nearly Efficient Likelihood Ratio Tests of the Unit Root Hypothesis