The Restricted Likelihood Ratio Test at the Boundary in Autoregressive Series

Willa W. Chen [*]   Rohit S. Deo [†]

July 6, 2009

Abstract. The restricted likelihood ratio test, RLRT, for the autoregressive coefficient in autoregressive models has recently been shown to be second order pivotal when the autoregressive coefficient is in the interior of the parameter space and so is very well approximated by the χ_1^2 distribution. In this paper, the non-standard asymptotic distribution of the RLRT for the unit root boundary value is obtained and is found to be almost identical to that of the χ_1^2 in the right tail. Together, the above two results imply that the χ_1^2 distribution approximates the RLRT distribution very well even for near unit root series and transitions smoothly to the unit root distribution.

Keywords. Boundary value; confidence interval; curvature; restricted likelihood, unit root.

1 Introduction:

Assume that the $n \times 1$ vector $\mathbf{Z}$ follows the linear model

$$\mathbf{Z} = \mathbf{W}\boldsymbol{\theta} + \boldsymbol{\varepsilon}, \tag{1}$$

where $\mathbf{W}$ is an $n \times k$ design matrix, $\boldsymbol{\theta}$ is a vector of regression coefficients , $\boldsymbol{\varepsilon} \sim N\left(\mathbf{0}, \boldsymbol{\Sigma}\left(\boldsymbol{\delta}\right)\right)$ and $\boldsymbol{\delta}$ is a parameter vector that characterises the error covariance matrix. The Restricted Likelihood (Kalbfleisch and Sprott, 1970) is useful for inference in this model when interest centers on the parameter vector $\boldsymbol{\delta}$ and $\boldsymbol{\theta}$ is a nuisance parameter vector. The Restricted Likelihood, RL, is defined as the exact likelihood of the linearly transformed data $\mathbf{TZ}$, where $\mathbf{T}$ is a full row rank matrix chosen such that $\mathbf{TW} = \mathbf{0}$, so that the nuisance parameters $\boldsymbol{\theta}$ are eliminated. The RL

---

[*]Department of Statistics, Texas A&M University, College Station, Texas 77843, USA

[†]Stern School of Business, New York University, 8-57 KMEC, 44 W. 4'th Street, New York NY 10012, USA

has been very popular in mixed linear models and the behaviour of RLRT when parameters are on the boundary has been recently studied (Claeskens 2004). A similar boundary problem in a linear model arises in an autoregressive model of order 1, AR(1), with intercept

$$\mathbf{X} = \mathbf{1}\mu + \boldsymbol{\varepsilon} \tag{2}$$

where $\mathbf{X} = (X_1, ..., X_n)'$, $\boldsymbol{\varepsilon} = (\varepsilon_1, ..., \varepsilon_n)'$ and ε follows a zero mean AR(1) given by $\varepsilon_t = \alpha\varepsilon_{t-1} + v_t$ for $t \geq 2$, where $v_t \sim i.i.d.N(0, \sigma_v^2)$. The initial condition ε_1 is assumed to follow $N\left(0, \left(1 - \alpha^2\right)^{-1}\sigma_v^2\right)$ when $|\alpha| < 1$ and $N\left(0, \sigma_v^2\right)$ when $\alpha = 1$. When the AR coefficient satisfies $|\alpha| < 1$, the series $\{X_t\}$ is stationary and the t-statistic and Likelihood Ratio Test, LRT, for α have standard asymptotic distributions. However, the series $\{X_t\}$ is non-stationary at the boundary when $\alpha = 1$ and the asymptotic distribution of both the t-statistic and the LRT for α in model (2) is very different from the standard normal and the χ_1^2 respectively in this case (Fuller, 1996). As a result, in finite samples the distribution of the t-statistic and the full LRT for α deviate substantially from the standard normal and the χ_1^2 respectively when α is close to the unit boundary. Indeed, using results by Hayakawa (1977) on the expansion of the distribution of likelihood ratio tests, Chen and Deo (2009a) obtained the formal expansion of the distribution of the full LRT for α in model (2) and showed that when $|\alpha| < 1$,

$$P\left(LRT \leq x\right) = P\left(\chi_1^2 \leq x\right) + \frac{1 + 7\alpha}{1 - \alpha}\frac{0.25}{n}\left[P\left(\chi_3^2 \leq x\right) - P\left(\chi_1^2 \leq x\right)\right] + O\left(n^{-2}\right). \tag{3}$$

The fact that the leading error term blows up as α approaches unity implies that the χ_1^2 distribution will be a poor approximation to the distribution of the full LRT for values of α close to unity. This is seen very clearly in the plot on the left side of Figure 1, where we plot the empirical densities of the full LRT for various values of α and samples of size $n = 100$, along with that of the χ_1^2. (Since the χ_1^2 distribution is very right skewed, we plot the distribution of the cube root of the LRT to reduce skewness and make comparisons in the right tail clearer).

Chen and Deo (2009a) have shown that the prime cause of this problem for likelihood based inference near the unit root is the nuisance intercept parameter μ in (2). Since the RL is free of such nuisance location parameters, one might thus expect the RLRT to have better finite sample performance. Furthermore, Chen and Deo (2009a) also showed that the RL has small Efron (1975) curvature, which additionally supports the notion of a well behaved RLRT. That this is indeed the case can be seen from the formal expansion of the distribution of the RLRT for α in model (2) when $|\alpha| < 1$,

$$P\left(RLRT \leq x\right) = P\left(\chi_1^2 \leq x\right) - \frac{0.25}{n}\left[P\left(\chi_3^2 \leq x\right) - P\left(\chi_1^2 \leq x\right)\right] + O\left(n^{-2}\right), \tag{4}$$
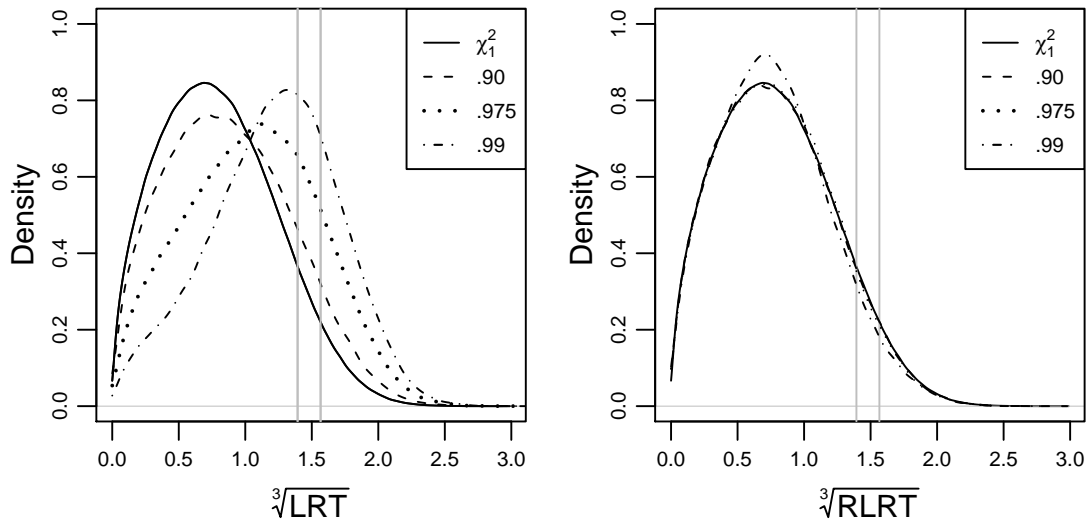
2

Figure 1: Empirical densities of cubic root transformed LRT statistics and Retricted LRT statistics of AR(1) processes with unknown mean. The vertical lines are 90th and 95th percentiles. Both plots are based on 100,000 repetitions of an AR(1) with sample size $n = 100$ and AR coefficient $\alpha = .9, .975$, and $.99$.

which Chen and Deo (2009a) obtained. Thus, the RLRT is second order pivotal, in sharp contrast to the strong second order dependence on α, particularly near the unit root, of the distribution of the full LRT seen in (3). This second order pivotal property of the RLRT suggests that the χ_1^2 distribution should provide a good approximation even near the unit root. This can be seen in the plot on the right side of Figure 1, where we plot the empirical density of RLRT for various values of α with $n = 100$. In this paper, we obtain the asymptotic distribution of the RLRT at the unit root boundary. The method of proof required is novel since regularity conditions, such as asymptotic normality of the score function, that are assumed in the related boundary value literature (eg. Self and Liang, 1987 and Vu and Zhou, 1997) do not hold here. It is found that the asymptotic distribution of the RLRT at the boundary has a right tail that is almost identical to that of the χ_1^2. Thus, inference based on the RLRT in conjunction with the χ_1^2 critical values will provide almost exactly sized tests, no matter what the value of the AR coefficient. Some related work by Francke and de Vos (2006) suggests the RLRT intervals may also be close to uniformly most accurate invariant, while simulation results in Chen and Deo (2009b) show that RLRT based confidence intervals uniformly dominate competing bootstrap based intervals in terms of average length, power against the unit root and ease of computation.

3

In the next Section we describe the RL and present our main result.

2 The Restricted Likelihood at the Unit Root:

The RL has been considered for time series models by, among others, Tunnicliffe Wilson (1989), Rahman and King (1997), Cheang and Reinsel (2000) and most recently by Francke and de Vos (2006), who actually study it in the context of unit root tests. The RL has appealing properties in that the restricted maximum likelihood (REML) estimates do not lose efficiency (Harville 1977) and are also less biased than full maximum likelihood estimates in nearly integrated AR models with intercept (Cheang and Reinsel, 2000) and with trend (Kang, Shin and Lee, 2003). Furthermore, in an AR(1) model the RL is equivalent to the approximate conditional likelihood (Cruddas, Cox and Reid, 1989).

In the case of the AR(1) model in (2), the restricted log-likelihood (Appendix A of Tunnicliffe-Wilson, 1989) simplifies to

$$L\left(\mathbf{X}, \alpha, \sigma_v^2\right) = -\frac{n-1}{2}\log \sigma_v^2 + \frac{1}{2}\log\left\{\frac{1+\alpha}{(n-2)(1-\alpha)+2}\right\} - \frac{1}{2\sigma_v^2}Q\left(\alpha\right), \tag{5}$$

where

$$Q\left(\alpha\right) = \left(1-\alpha^2\right)X_1^2 + \sum_{t=2}^{n}\left(X_t - \alpha X_{t-1}\right)^2 - \frac{1-\alpha}{(n-2)(1-\alpha)+2}\left\{X_1 + X_n + (1-\alpha)\sum_{t=2}^{n-1}X_t\right\}^2. \tag{6}$$

It is worth pointing out, as noted in Francke and de Vos (2006), that the restricted log-likelihood for the AR(1) is well defined at the boundary $\alpha = 1$. Indeed, since the restricted likelihood for a time series with an intercept is the exact likelihood of the first difference (which is the linear transformation that eliminates an intercept parameter in a linear model), it follows that the restricted likelihood will be well defined and finite at the unit root since the first difference of a unit root AR(1) is a stationary white noise. The RLRT for testing the null hypothesis $H_0 : \alpha = \alpha_0$ against $H_1 : \alpha \neq \alpha_0$ for any α_0 is

$$R_T\left(\alpha_0\right) = 2L\left(\mathbf{X}, \hat{\alpha}, \hat{\sigma}_v^2\right) - 2L\left(\mathbf{X}, \alpha_0, \hat{\sigma}_{v,0}^2\right),$$

where $L\left(\cdot\right)$ is the restricted log-likelihood in (5) and $\left(\hat{\alpha}, \hat{\sigma}_v^2\right)$ and $\left(\alpha_0, \hat{\sigma}_{v0}^2\right)$ are the unconstrained and constrained (under H_0) REML estimates respectively. The unconstrained estimate $\hat{\alpha}$ can be obtained through a univariate minimisation as

$$\hat{\alpha} = \arg\min_{\alpha \in (-1,1]}\left\{(n-1)\log Q\left(\alpha\right) - \log\left(\frac{1+\alpha}{(n-2)(1-\alpha)+2}\right)\right\} \tag{7}$$

4

after concentrating out σ_v^2 from $L(\cdot)$, while the innovation variance estimate is obtained as $\hat{\sigma}_v^2 = (n-1)^{-1} Q(\hat{\alpha})$. The constrained innovation variance estimate is $\hat{\sigma}_{v,0}^2 = (n-1)^{-1} Q(\alpha_0)$.

The use of the RLRT for unit root tests has already been proposed in the literature by Francke and de Vos (2006). They obtained the limiting distribution of the RLRT for testing $H_0 : \alpha = 1$ versus $H_a : \alpha = 1 - n^{-1}\gamma$, where $\gamma > 0$ is a fixed constant, given by

$$2L\left(\mathbf{X}, 1 - n^{-1}\gamma, (n-1)^{-1} Q\left(1 - n^{-1}\gamma\right)\right) - 2L\left(\mathbf{X}, 1, (n-1)^{-1} Q(1)\right).$$

Thus, they obtained the limiting distribution of the RLRT in a test of the unit root against a sequence of local-to-unity alternatives. However, they did not establish the consistency of the REML estimate $\hat{\alpha}$ under the boundary unit root, nor did they obtain its limiting distribution or the limiting distribution of the RLRT for the general alternative $H_a : \alpha < 1$ given by

$$R_T(1) = 2L\left(\mathbf{X}, \hat{\alpha}, (n-1)^{-1} Q(\hat{\alpha})\right) - 2L\left(\mathbf{X}, 1, (n-1)^{-1} Q(1)\right), \tag{8}$$

the proof of which involves non-standard arguments. Francke and de Vos did provide critical values based on simulations of the distribution of both $n(\hat{\alpha} - 1)$ and the RLRT when $\alpha = 1$. In their simulation study of the power of the RLRT , they found that the RLRT tends to be more powerful than other standard unit root tests and that its power almost coincides with the power envelope for Gaussian AR(1) series, yielding tests that are almost uniformly most powerful invariant. Francke and de Vos (2006) also provided an intriguing heuristic explanation of this finding, showing that the restricted likelihood is almost monotonic in the AR(1) case. In Theorem 1 below we provide the limiting distribution of the REML estimate $\hat{\alpha}$ as well as that of the RLRT when $\alpha = 1$. As stated earlier, it should be noted that earlier literature on likelihood ratio tests for boundary values (eg. Self and Liang, 1987 and Vu and Zhou, 1997) generally requires the score function to be asymptotically normal. From Phillips (1987), it is easy to show that the score function of the RL at $\alpha = 1$ is not asymptotically normal. Hence, a new proof needs to be constructed to obtain the limit theory for the RLRT in this case.

Before stating our main theorem, we define some quantities that appear in the limiting distribution of $\hat{\alpha}$ as well as a useful Lemma. Let $W(\cdot)$ be the standard Wiener process and

5

define the random variables (see Fuller 1996, Ch. 10)

$$G \equiv \int_0^1 W^2(t)\, dt = \sum_{i=1}^\infty \gamma_i^2 U_i^2$$

$$H \equiv \int_0^1 W(t)\, dt = \sqrt{2} \sum_{i=1}^\infty \gamma_i^2 U_i$$

$$T \equiv W(1) = \sqrt{2} \sum_{i=1}^\infty \gamma_i U_i, \tag{9}$$

where $\gamma_i = (-1)^{i+1} 2\left[(2i-1)\pi\right]^{-1}$ and $U_i \sim i.i.d.N(0,1)$ and define the random cubic polynomial

$$\begin{aligned}
f(z) &= 2\left(G - H^2\right) z^3 + \left\{ 8\left(G - H^2\right) + H^2 - 1 + (T-H)^2 \right\} z^2 \\
&\quad + \left\{ 4G + 4(T-H)^2 + 4\left(G - H^2\right) - 3 \right\} z + 2\left(T^2 - 1\right) \\
&\equiv a_3 z^3 + a_2 z^2 + a_1 z + a_0
\end{aligned} \tag{10}$$

with associated discriminant function

$$\Delta = a_1^2 a_2^2 - 4a_0 a_2^3 - 4a_1^3 a_3 + 18 a_0 a_1 a_2 a_3 - 27 a_0^2 a_3^2 \tag{11}$$

and the random function

$$g(z) = z^2 G + z\left(T^2 - 1\right) - \frac{z}{z+2}(T + zH)^2 + \log\left(\frac{z+2}{2}\right).$$

The following Lemma describes the nature of the roots of the polynomial $f(z)$.

Lemma 1 *The random cubic polynomial $f(z)$ in (10) has exactly one positive real root γ_0 when $T^2 \leq 1$ and at least one real negative root, γ_{01}, when $T^2 > 1$. Furthermore, when $T^2 > 1$, all three roots, $\gamma_{01} < \gamma_{02} < \gamma_{03}$, are real if $\Delta > 0$, in which case either $\gamma_{01} < \gamma_{02} < \gamma_{03} < 0$ or $\gamma_{01} < 0 < \gamma_{02} < \gamma_{03}$, while γ_{02} and γ_{03} are complex if $\Delta < 0$.*

We are now in a position to state our result.

Theorem 1 *Assume that $\alpha = 1$ in model (2) and let $\hat{\alpha}$ denote the REML estimate obtained as*

$$\hat{\alpha} = \arg \min_{[-1+\delta, 1]} \left\{ (n-1)\log Q(\alpha) - \log\left(\frac{1+\alpha}{(n-2)(1-\alpha)+2} \right) \right\},$$

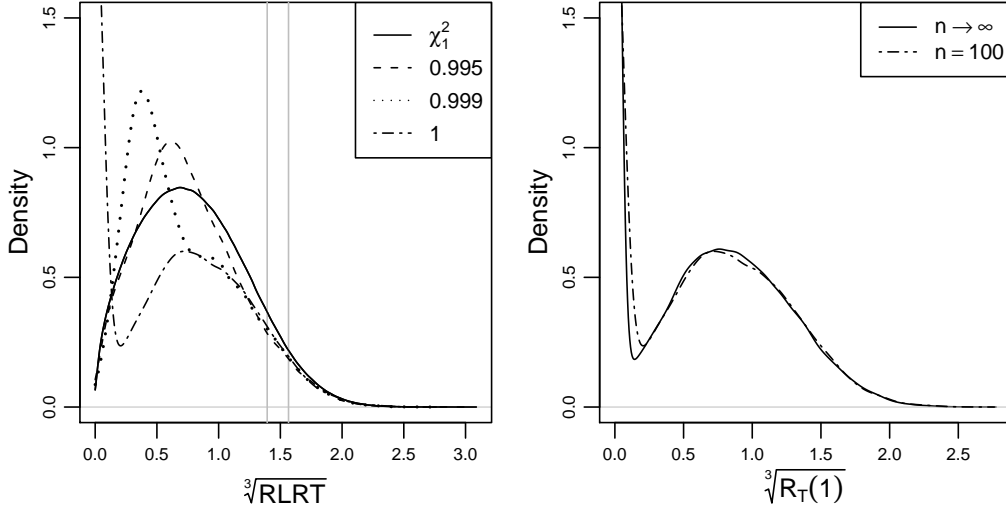*where $Q(\alpha)$ is as defined in (6) and $\delta > 0$ is fixed. Then*

Figure 2: Empirical densities of cubic root transformed RLRT statistics of AR(1) processes when $\alpha \to 1$ and the limiting distribution of cubic root transformed RLRT at unit root, $R_T(1)$. The vertical lines are 90th and 95th percentiles of χ_1^2.

(i) $\hat{\alpha} \xrightarrow{P} 1$

(ii) $n(\hat{\alpha} - 1) \xrightarrow{D} \Gamma \equiv -\gamma_0 I(T^2 \leq 1) - \gamma_{03} I(A)$ where $A = (T^2 > 1) \cap (\gamma_{03} > 0) \cap (g(\gamma_{03}) < 0)$, where γ_0 and γ_{03} are as defined in Lemma 1

(iii)   For the RLRT to test $H_0 : \alpha = 1$ versus $H_a : \alpha < 1$ given in (8), we have

$$R_T(1) \xrightarrow{D} R \equiv -g(\gamma_0) I(T^2 \leq 1) - g(\gamma_{03}) I(A)$$

The expression for the limiting distribution of the RLRT in Theorem 1 looks awkward at first glance. However, the random variables G, H and T are very easy to simulate using the infinite series representation in (9) and hence the limiting distributions of both $n(\hat{\alpha} - 1)$ and $R_T(1)$ can be simulated very easily. In Table I we present the quantiles of these two limiting distributions based on 200,000 replications, using the series representations in (9) truncated at 500,000. These values are very close to the finite sample values of the same distributions given in Table 2 of Francke and de Vos (Note: They report critical values for 0.5×RLRT instead of RLRT). The truncation value of 500,000 was found to be more than adequate to get stable results since we got almost identical results by using a truncation value of 100,000 and 10,000. In the 200,000 replications, the set $(T^2 \leq 1)$ occurred 68.383 % times, while the set A of Theorem 1 occurred

7

only 0.132% times. The convergence of the RLRT distribution at the unit root to its limit given in Theorem 1 occurs very quickly for samples as small as $n = 100$. This can be seen in the plot on the right in Figure 2, where we plot the densities of the cube root of the limit distribution R as well as that of the RLRT for $n = 100$.

As we had postulated earlier, we expect the limiting distribution of $R_T(1)$ to be very close to that of a χ_1^2 in the right tail. To check this, we also present in Table I the values of $P(R_T(1) > q_{\chi_1^2})$ for various values of $q_{\chi_1^2}$, where $q_{\chi_1^2}$ are the quantiles of the χ_1^2 distribution and the probability is computed as the proportion of exceedances in the 200,000 replications of $R_T(1)$. It is seen that these probabilities are very close to, but a little smaller, than the nominal values for quantiles from 90% onwards. As a result, inference for the unit root based on $R_T(1)$ with the χ_1^2 distribution used as a reference will result in tests that have size just a little less than nominal. This close yet conservative approximation in the right tail by the χ_1^2 distribution continues to hold for the RLRT distribution for values of α that are less than but very close to 1. This can be seen in the plot on the left in Figure 2, where we plot the densities for $\alpha = 0.995, 0.999$ and 1. Correspondingly, confidence intervals for α obtained by inverting the acceptance region of R_T using the χ_1^2 distribution will result in intervals that will have almost exact coverage in the neighbourhood of the unit root, with slight over-coverage at the unit root.

The theory in Theorem 1 has been established for an AR(1) process with intercept. In this paper, we do not attempt to extend the theory to higher order AR(p) processes and to processes with trend, though we conjecture that the result of Theorem 1 will continue to hold in intercept models for any $p \geq 1$. Based on the simulation results of Francke and de Vos (2006), it is apparent that the limiting distribution of the RLRT will be slightly different in the case of trend though the difference does not seem to be large enough to compromise the approximation by the χ_1^2 distribution in the right tail.

Table I. Critical Values from Limiting Distributions of $n(\hat{\alpha} - 1)$ and R_T

	1%	2.5%	5%	10%	20%	30%	40%	50%
Γ	17.5868	13.8811	10.9997	8.1232	5.2684	3.5554	2.3412	1.3834
R	6.4937	4.8544	3.6218	2.4564	1.3725	0.7933	0.4393	0.2084
$P(R > q_{\chi_1^2})$	0.0093	0.0227	0.0442	0.0857	0.1667	0.2446	0.3207	0.3945

Appendix:

Proof of lemma 1: Proof: We note that $G - H^2 > 0$ by the Cauchy Schwarz inequality.

(i) Consider $T^2 < 1$. We can write $f(z)$ as

$$f(z) = p_1(z) - p_2(z),$$

where

$$p_1(z) = 2\left(G - H^2\right)z^3 + \left\{8\left(G - H^2\right) + \left[H^2 + (T - H)^2\right]\right\}z^2$$
$$+ 4\left\{4G + 4(T - H)^2\right\}z + 2T^2$$

and

$$p_2(z) = \left(z^2 + 3z + 2\right).$$

Since $G - H^2 > 0$, we have $\lim_{z \to \infty} f(z) = \infty$ while $f(0) = 2\left(T^2 - 1\right) < 0$, implying that f must have at least one positive root. If $f(\gamma_0) = 0$ for some $\gamma_0 > 0$, then we must have $p_1(\gamma_0) = p_2(\gamma_0)$. Thus, the number of positive real roots of $f(z) = 0$ are the same as the number of values of z for which $p_1(z) = p_2(z)$. Note that $p_1(0) < p_2(0)$ since $T^2 < 1$ and $\lim_{z \to \infty} p_1(z) > \lim_{z \to \infty} p_2(z)$ (since p_1 is a cubic with $a_3 > 0$ whereas p_2 is merely a quadratic). Hence, p_1 must cross p_2 at least for one value of $z > 0$. However, because all the coefficients of p_1 and p_2 are positive, it implies that both p_1 and p_2 are convex for $z > 0$. Hence, p_1 can cross p_2 exactly for one $z > 0$, which is equivalent to $f(z) = 0$ having exactly one positive real root.

(ii) Consider $T^2 > 1$. We can write $f(z)$ as $f(z) = a_3(z - r_1)(z - r_2)(z - r_3)$, where r_i are the roots of $f(z)$. This implies that $-a_3 r_1 r_2 r_3 = f(0) = 2\left(T^2 - 1\right) > 0$. Since $\alpha_3 > 0$, this implies that at least one root has to be real negative. The existence of real/complex roots when $\Delta < 0$ or $\Delta > 0$ follows from standard theory on the roots of cubic polynomials.

Proof of Theorem 1: As noted in (7), the concentrated likelihood is

$$l(\alpha) = (n - 1)\log Q(\alpha) - \log\left\{\frac{1 + \alpha}{(n - 2)(1 - \alpha) + 2}\right\}$$

and

$$\hat{\alpha} = \arg\min_{\alpha \in [-1 + \delta, 1]} l(\alpha)$$

9

for some positive δ. Some simple algebra establishes that $l(\alpha)$ is a function of the random quantities X_1, X_n and

$$S_n = \sum_{t=2}^{n} v_t^2, \qquad G_n = \sum_{t=2}^{n-1} X_t^2, \quad H_n = \sum_{t=2}^{n-1} X_t, \quad U_n = \sum_{t=2}^{n} v_t X_{t-1},$$

where (see Fuller 1996),

$$
\begin{aligned}
n^{-2} G_n &\xrightarrow{D} \sigma_v^2 G = \sigma_v^2 \int_0^1 W^2(s)\, ds, \\
n^{-3/2} H_n &\xrightarrow{D} \sigma_v H = \sigma_v \int_0^1 W(s)\, ds, \\
n^{-1/2} X_n &\xrightarrow{D} \sigma_v T = \sigma_v W(1), \qquad n^{-1/2} X_1 \xrightarrow{D} 0 \\
n^{-1} U_n &\xrightarrow{D} \frac{1}{2}\sigma_v^2 \left(T^2 - 1 \right) = (1/2)\,\sigma_v^2 \left\{ W^2(1) - 1 \right\}, \\
n^{-1} S_n &\xrightarrow{P} \sigma_v^2,
\end{aligned}
\tag{12}
$$

and the random variables G, H and T are as defined in (9). By Skorohod's device (See Billingsley, 1991), there exists a probability space on which one can define random sequences $(\tilde{S}_n, \tilde{G}_n, \tilde{H}_n, \tilde{U}_n, \tilde{X}_n, \tilde{X}_1)'$ which are identically distributed as $(S_n, G_n, H_n, U_n, X_n, X_1)'$ respectively but for which the convergence results of (12) hold almost surely. It follows that the minimiser $\tilde{\alpha} = \arg\min_\alpha \tilde{l}(\alpha)$, where $\tilde{l}(\alpha)$ is defined analogously to $l(\alpha)$ but using $(\tilde{S}_n, \tilde{G}_n, \tilde{H}_n, \tilde{U}_n, \tilde{X}_n, \tilde{X}_1)'$ instead of $(S_n, G_n, H_n, U_n, X_n, X_1)'$ will be identically distributed as $\hat{\alpha}$. Hence, any strong convergence result that is obtained for $n(1-\tilde{\alpha})$ will hold in a "convergence in distribution" sense for $n(1-\hat{\alpha})$. We will now proceed to obtain a limiting strong convergence result for $n(1-\tilde{\alpha})$ and for the rest of the proof we will work on the part of the sample space on which the convergence of $(\tilde{S}_n, \tilde{G}_n, \tilde{H}_n, \tilde{U}_n, \tilde{X}_n, \tilde{X}_1)'$ holds with probability 1. To establish the limiting behaviour of $\tilde{\alpha}$, we start by investigating the nature of the sign changes of $\tilde{l}'(z)$ as well as the positive real roots of $\tilde{l}'(z) = \frac{d}{dz}\tilde{l}'(z) = 0$, where $z = 1 - \alpha$. Some tedious algebra shows that

$$\tilde{l}'(z) = -\frac{(n-1)\,K(z)}{(n-2)\,z+2}\left\{ A(z) + \frac{B(z)}{(n-2)\,z+2} + \frac{C(z)}{[(n-2)\,z+2]^2} \right\},$$

where $K(z)$ is a strictly positive function for $z \in [0,2)$,

$$A(z) = 2\tilde{G}_n z^2 + (2\tilde{U}_n - 4\tilde{G}_n)z - 4\tilde{U}_n,$$

$$
\begin{aligned}
B(z) = {}&-3\tilde{H}_n z^3 + \left\{ 6\tilde{H}_n - 2\tilde{G}_n - 4(\tilde{X}_1 + \tilde{X}_n)\tilde{H}_n \right\} z^2 \\
&+ \left\{ 8(\tilde{X}_1 + \tilde{X}_n)\tilde{H}_n - (\tilde{X}_1 + \tilde{X}_n)^2 - 4\tilde{U}_n \right\} z + 4\tilde{U}_n + 2\tilde{X}_1 \tilde{X}_n
\end{aligned}
$$

10

and

$$C(z) = (n-2)\tilde{H}_n z^4 + 2\left\{(n-2)(\tilde{X}_1 + \tilde{X}_n)\tilde{H}_n - (n-3)\tilde{H}_n\right\}z^3$$
$$+ \left\{(n-2)(\tilde{X}_1 + \tilde{X}_n)^2 - 4(n-3)(\tilde{X}_1 + \tilde{X}_n)\tilde{H}_n\right\}z^2 - 2(n-3)(\tilde{X}_1 + \tilde{X}_n)^2 z.$$

Thus, the order of sign changes and positive real roots of $\tilde{l}'(z)$ are the same as those of $\tilde{g}(z)$, where

$$\tilde{g}(z) = -\frac{1}{4}\left\{[(n-2)z + 2]^2 A(z) + [(n-2)z + 2]B(z) + C(z)\right\}$$
$$\equiv c_{4n}z^4 + c_{3n}z^3 + c_{2n}z^2 + c_{1n}z + c_{0n}, \tag{13}$$

where

$$c_{4n} = -\frac{1}{2}(n-2)^2\tilde{G}_n + \frac{1}{2}(n-2)\tilde{H}_n$$

$$c_{3n} = (n-2)(n-\frac{7}{2})\tilde{G}_n - (n-3)\tilde{H}_n - \frac{1}{2}(n-2)^2\tilde{U}_n + \frac{1}{2}(n-2)\left(\tilde{X}_1 + \tilde{X}_n\right)\tilde{H}_n,$$

$$c_{2n} = 4\left(n-\frac{9}{4}\right)\tilde{G}_n - 3\tilde{H}_n + (n-2)(n-3)\tilde{U}_n - (n-3)\left(\tilde{X}_1 + \tilde{X}_n\right)\tilde{H}_n,$$

$$c_{1n} = 4\tilde{G}_n + 3(n-2)\tilde{U}_n - 4\left(\tilde{X}_1 + \tilde{X}_n\right)\tilde{H}_n + \frac{1}{2}(n-2)\left(\tilde{X}_1 + \tilde{X}_n\right)^2,$$

and

$$c_{0n} = 2\tilde{U}_n - 2\tilde{X}_1\tilde{X}_n.$$

The order of sign changes and roots of the polynomial in (13) are the same as those of

$$P_n(z) = -z^4 + b_{3n}z^3 + b_{2n}z^2 + b_{1n}z + b_{0n}, \quad b_{kn} = -c_{kn}/c_{4n}, \ k = 0,\ldots,3. \tag{14}$$

From (12) we have

$$n^3 b_{0n} \to b_0 \equiv 2\left(T^2 - 1\right)\left(G - H^2\right)^{-1},$$
$$n^2 b_{1n} \to b_1 \equiv \frac{4T^2 + 8G - 8HT - 3}{G - H^2}, \tag{15}$$
$$n b_{2n} \longrightarrow b_2 \equiv \frac{8G - 6H^2 - 2HT + T^2 - 1}{G - H^2}$$

and

$$b_{3n} \longrightarrow b_3 \equiv 2.$$

Thus, $b_{3n} = 2 + O\left(n^{-1}\right)$, $b_{2n} = O\left(n^{-1}\right)$, $b_{1n} = O\left(n^{-2}\right)$, $b_{0n} = O\left(n^{-3}\right)$ and it follows that the roots of $P_n(z)$, say z_{1n}, z_{2n}, z_{3n} and z_{4n}, converge to 0, 0, 0 and 2 respectively. Since the polynomial $P_n(z)$ has real coefficients, any complex roots that it may possess must occur in

11

complex-conjugate pairs. Hence, z_{4n} must be real, since otherwise its complex-conjugate would also have to be a root with an imaginary coefficient converging to zero. This would imply the existence of two roots of $P_n(z)$ that converge to 2, contradicting the fact that only one root converges to 2. Armed with the observation that at least one root, viz. z_{4n}, is real we now proceed to study the sign changes and real roots of $P_n(z)$ lying in $[0, 2 - \varepsilon)$ for any $\varepsilon > 0$. We consider two separate cases, $T^2 < 1$ and $T^2 > 1$.

Case (i) $T^2 < 1$: Since $G - H^2 > 0$, it follows from (15) that $b_{0n} = P_n(0) < 0$ for large n. This fact in conjunction with the observations that the coefficient of z^4 in $P_n(z)$ is -1 and that z_{4n} is real implies that $P_n(z)$ must have either exactly two positive real roots or all four positive real roots, i.e. the positive real roots must occur in pairs. Since $b_{0n} = P_n(0) < 0$ for large n, it also follows that one of these positive roots, and not the boundary value $z = 0$, is the minimiser of $\tilde{l}(z)$. We now consider the limiting behaviour of the positive real roots in $[0, 2)$. Let $\gamma_0 > 0$ be the positive root of $f(z)$ in Lemma 1. From (15) and Lemma 1 it follows that for any $\varepsilon > 0$,

$$n^3 P_n\left(\frac{\gamma_0 + \varepsilon}{n}\right) = -\frac{(\gamma_0 + \varepsilon)^4}{n} + b_{3n}(\gamma_0 + \varepsilon)^3 + n b_{2n}(\gamma_0 + \varepsilon)^2 + n^2 b_{1n}(\gamma_0 + \varepsilon) + n^3 b_{0n} \quad (16)$$

$$\to f(\gamma_0 + \varepsilon) > 0.$$

A similar argument shows that

$$n^3 P_n\left(\frac{\gamma_0 - \varepsilon}{n}\right) \to f(\gamma_0 - \varepsilon) < 0. \tag{17}$$

This implies that at least one of the remaining real positive roots, say z_{1n}, lies in $(n^{-1}(\gamma_0 - \varepsilon), \ n^{-1}(\gamma_0 + \varepsilon))$. Since $\varepsilon > 0$ is arbitrary, this implies that $n z_{1n} \to \gamma_0 > 0$. If $P_n(z)$ has exactly two positive real roots, then the preceding argument has shown that one of the roots is converging to 2 while the other, z_{1n}, lies asymptotically in $(0, 2)$ and satisfies $n z_{1n} \to \gamma_0$. We now consider the possibility that $P_n(z)$ has all four positive roots. Rewriting $P_n(z) = -(z - z_{1n})(z - z_{2n})(z - z_{3n})(z - z_{4n})$, expanding this product and equating the coefficient of z^2 with that in (14), we have

$$z_{3n}z_{4n} + z_{2n}z_{4n} + z_{2n}z_{3n} + z_{1n}z_{4n} + z_{1n}z_{3n} + z_{1n}z_{2n} = -b_{2n}$$

and hence

$$z_{4n}(n z_{3n} + n z_{2n} + n z_{2n}z_{3n}) + n z_{1n}(z_{4n} + z_{3n} + z_{2n}) = -n b_{2n}. \tag{18}$$

Since $n z_{1n} \to \gamma_0$, $z_{4n} \to 2$, $z_{2n} \to 0$ and $z_{3n} \to 0$ and since by assumption $z_{2n} \geq 0$, $z_{3n} \geq 0$, it follows from (15) that $n z_{2n}$ and $n z_{3n}$ are bounded. Consider first $n z_{2n}$. Since this is a bounded

sequence, it must have at least once convergent subsequence $n_k z_{2n_k}$ with non-negative limit, say, γ_1. Using the fact that z_{1n} is a solution of $P_n(z)$, we then get

$$0 = n_k^3 P_{n_k}(n_k z_{2n_k}) = -n_k^{-1} z_{2n_k}^4 + b_{3n_k} z_{2n_k}^3 + n_k b_{2n_k} z_{2n_k}^2 + n_k^2 b_{1n_k} z_{2n_k} + n_k^3 b_{0n_k} \rightarrow f(\gamma_1). \quad (19)$$

From Lemma 1 it follows that $\gamma_1 = \gamma_0$ and hence $n_k z_{2n_k} \rightarrow \gamma_0$. This argument shows that every convergent subsequence of $n z_{2n}$ must converge to γ_0 and hence we have $n z_{2n} \rightarrow \gamma_0$. A similar argument establishes that $n z_{3n} \rightarrow \gamma_0$. Thus, if all four roots of $P_n(z)$ are real positive, then $z_{4n} \rightarrow 2$, while the other three roots z_{in}, $i = 1, 2, 3$ lie in $(0, 2 - \varepsilon)$ for any $\varepsilon > 0$ and satisfy $n z_{in} \rightarrow \gamma_0$ for $i = 1, 2, 3$. All of the above arguments taken together show that when $T^2 < 1$, there exists an $\tilde{\alpha}$ in $(-1, 1)$ that minimises $\tilde{l}(\alpha)$ on the set $(-1, 1]$ and satisfies $n(1 - \tilde{\alpha}) \rightarrow \gamma_0$.

Case (ii) $T^2 > 1$: It follows from (15) that $b_{0n} = P_n(0) > 0$ for large n. Hence, $P_n(z)$ must have either

(I) exactly two real roots, one positive and the other negative or

(II) all four real roots with one positive and three negative or

(III) all four real roots with one negative and three positive.

We now study the behaviour of $n(\tilde{\alpha} - 1)$ for each of these three cases on the sets $(\gamma_{03} > 0) \cap (g(\gamma_{03}) < 0)$, $(\gamma_{03} > 0) \cap (g(\gamma_{03}) < 0)^c$, $(\gamma_{03} > 0)^c \cap (g(\gamma_{03}) < 0)$ and $(\gamma_{03} > 0)^c \cap (g(\gamma_{03}) < 0)^c$.

Consider first $(\gamma_{03} > 0) \cap (g(\gamma_{03}) < 0)$. Since it has been shown above that $z_{4n} \rightarrow 2$ is real, both cases (I) and (II) imply that asymptotically there is no real positive root in $(0, 2 - \varepsilon)$ for any positive $\varepsilon > 0$. However, an argument similar to that used in (16) and (17) above establishes that there exists a positive real root, say z_{3n}, such that $n z_{3n} \rightarrow \gamma_{03}$, where $\gamma_{03} > 0$ by assumption. Hence, cases (I) and (II) cannot happen when $(\gamma_{03} > 0) \cap (g(\gamma_{03}) < 0)$. For case (III), an argument similar to that in (16) and (17) as well as (18) and (19) implies that $n z_{1n} \rightarrow \gamma_{01}$ and $n z_{2n} \rightarrow \gamma_{02}$, where $z_{1n} < 0 < z_{2n} < z_{3n} < z_{4n} \rightarrow 2$. Since $b_{0n} = P_n(0) > 0$ for large n, it implies that $\tilde{l}(z)$ is increasing in the neighbourhood of 0 with a local maxima at z_{2n} and hence $\arg\min_{z \in [0,2)} \tilde{l}(z)$ is either the boundary value $z = 0$ or $z = z_{3n}$. Since for any $\gamma > 0$ we have $\lim_n \tilde{l}(1 - n^{-1}\gamma) - \tilde{l}(1) = g(\gamma)$, it follows that on the set $(T^2 > 1) \cap (\gamma_{03} > 0) \cap (g(\gamma_{03}) < 0)$ we have $g(0) = 0 > g(\gamma_{03}) = \lim_n \tilde{l}(1 - z_{3n}) - \tilde{l}(1)$ and hence $\arg\min_{z \in [0,2)} \tilde{l}(z) = z_{3n}$ from whence we get $n(1 - \tilde{\alpha}) \rightarrow \gamma_{03}$. Similar arguments show that on the remaining three sets, $n(\tilde{\alpha} - 1) = 0$.

The above arguments establish part (ii) of Theorem 1, from whence the consistency result of part (i) follows immediately. The result in part (iii) of Theorem 1 follows from part (ii) and by noting that $\lim_n \tilde{l}(1 - n^{-1}\gamma_n) - \tilde{l}(1) = g(\gamma)$ for any $\gamma_n \rightarrow \gamma > 0$.

13

REFERENCES

BILLINGSLEY, P. (1991) Probability and Measure. 2nd ed., John Wiley and Sons, New York.

CHEANG, W.K. AND REINSEL, G. (2000) Bias Reduction of Autoregressive Estimates in Time Series Regression Model through Restricted Maximum Likelihood. *Journal of the American Statistical Association* 95, 1173-1184.

CHEN, W. AND DEO, R. (2009a) Bias Reduction and Likelihood Based Almost-Exactly Sized Hypothesis Testing in Predictive Regressions using the Restricted Likelihood. *Econometric Theory*, forthcoming

CHEN, W. AND DEO, R. (2009b) The Chi-Square Approximation of the Restricted Likelihood Ratio Test for the Sum of Autoregressive Coefficients with Interval Estimation. Working paper.

CLAESKENS, G. (2004) Restricted Likelihood Ratio Lack-of-fit Tests using Mixed Spline Models. *Journal of the Royal Statistical Society B* 66, 909-926.

CRUDDAS, A. M., REID, N., COX, D. R. (1989) A time series illustration of approximate conditional likelihood. *Biometrika* 76, 231–237.

FRANCKE, M, AND DE VOS, A. (2007) Marginal Likelihood and Unit Roots. *Journal of Econometrics* 137, 708-728

FULLER, W. A. (1996) Introduction to Statistical Time Series. 2nd ed., John Wiley and Sons, New York.

HARVILLE, D. (1977) Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems. *Journal of the American Statistical Association* 72, 320-338.

HAYAKAWA, T. (1977) The likelihood ratio criterion and the asymptotic expansion of its distribution. *Annals of the Institute of Statistical Mathematics* 29 , 359-378.

KALBFLEISCH, J. AND SPROTT, D. (1970) Application of Likelihood Methods to Models Involving Large Numbers of Parameters. *Journal of the Royal Statistical Society B* 32, 175-194.

KANG, W., SHIN, D. AND LEE, Y. (2003) Baises of the Restricted Maximum Likelihood Estimators for ARMA Processes with Polynomial Time Trend. *Journal of Statistical Planning and Inference* 116, 163-176.

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

RAHMAN, S. AND KING, M. (1997) Marginal-Likelihood Score-Based Tests of Regression Disturbances in the Presence of Nuisance Parameters. *Journal of Econometrics* 82, 81-106.

SELF, S. AND LIANG, K-Y. (1987) Asymptotic Properties of Maximum Likelihood Estimators and Likelihood Ratio Tests under Nonstandard Conditions. *Journal of the American Statistical Association* 82, 605-610.

TUNNICLIFFE WILSON, G. (1989) On the Use of Marginal Likelihood in Time Series Model Estimation. *Journal of the Royal Statistical Society B* 51, 15-27.

VU, H. AND ZHOU, S. (1997) Generalization of Likelihood Ratio Tests under Nonstandard Conditions. *Annals of Statistics* 25, 897-916.