Ann. Inst. Statist. Math.
Vol. 50, No. 1, 29–48 (1998)

THE ORDER OF THE ERROR TERM FOR MOMENTS OF THE LOG LIKELIHOOD RATIO UNIT ROOT TEST IN AN AUTOREGRESSIVE PROCESS

ROLF LARSSON

Department of Statistics, Stockholm University, S-106 91 Stockholm, Sweden

(Received October 25, 1996; revised March 24, 1997)

Abstract. This paper investigates the asymptotics of the log likelihood ratio test for a unit root in an autoregressive (AR) process of general order. The main result is that the expectation and variance (in fact, all moments) of the test statistic may, to the order of T^{-1}, where T is the number of observations, be approximated by the expectation and variance of the corresponding test in an AR(1) process. This result has obvious implications for the asymptotics of unit root tests for panels. An explicit formula for the approximation error of a test in an AR(2) process is also given.

Key words and phrases: Approximation error, unit root test.

1. Introduction

Consider the AR$(p+1)$ model

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_{p+1} X_{t-p-1} + \varepsilon_t, \tag{1.1}$$

where we assume that the ε_t's are independent and normally distributed with mean zero and variance σ^2. We may rewrite (1.1) in error correction form as

$$\Delta X_t = \beta X_{t-1} + \sum_{j=1}^{p} \rho_j \Delta X_{t-j} + \varepsilon_t, \tag{1.2}$$

where $\Delta X_t = X_t - X_{t-1}$ for all t. Now, assume that all the roots of the equation $1 - \rho_1 r - \cdots - \rho_p r^p = 0$ have modulus larger than one. Under this assumption, the only possibility for the characteristic equation of the $\{X_t\}$ process to have a unit root is that $\beta = 0$, and the initial values of the process may be given a distribution such that the process becomes stationary if.f. $\beta < 0$. Hence, it is interesting to test the hypothesis $H_0 : \beta = 0$ against the alternative $H_1 : \beta < 0$ (a unit root test).

Im *et al.* (1996) study a situation with N independent series following the model (1.2), with a constant term added on the r.h.s. They consider the log likelihood ratio test of the null hypothesis that the β's for all series are zero against the alternative that all the β's are negative. The log likelihood ratio statistic for this test becomes a sum of the N independent log likelihood ratio test statistics for the single series, and as N tends to infinity, the distribution of this sum tends to normal by the central limit theorem (CLT). Hence, to explicitly find this asymptotic distribution, it is enough to calculate the expectation and variance of the log likelihood ratio test, $-2\log Q_T(p,\underline{\rho})$ say ($\underline{\rho}' \overset{\text{def}}{=} (\rho_1,\ldots,\rho_p)$), of H_0 against H_1 for a single series (such as the one given in (1.2) with a constant term added on the r.h.s.). Moreover, Im *et al.* suggest to approximate the expectation and variance of $-2\log Q_T(p,\underline{\rho})$ by the expectation and variance of the corresponding test statistics in an AR(1) process, $-2\log Q_T(0,\underline{0})$. (This has the advantage of giving an approximating distribution of the test statistic free of nuisance parameters.) They conjecture that, for fixed initial values of the process, the errors of these approximations are of order T^{-1}, where T is the number of observations, i.e.

$$E(-2\log Q_T(p,\underline{\rho})) = E(-2\log Q_T(0,\underline{0})) + O(T^{-1}) \tag{1.3}$$

and

$$\operatorname{Var}(-2\log Q_T(p,\underline{\rho})) = \operatorname{Var}(-2\log Q_T(0,\underline{0})) + O(T^{-1}). \tag{1.4}$$

(As is seen below, $-2\log Q_T(p,\underline{\rho})$ and $-2\log Q_T(0,\underline{0})$ have the same limiting distributions. The convergence of moments to the corresponding moments of the limit distribution may be proved using uniform integrability arguments as in Larsson (1997).) Hence, if T observations are taken from each series, Im *et al.* get a CLT saying that their test statistic is asymptotically normal with the same expectation and variance as $-2\log Q_T(0,\underline{0})$ has, under the conditions $N\to\infty$, $T\to\infty$ and $\sqrt{N}/T\to 0$. (This is an improvement from tests proposed earlier in the literature, which required $N/T\to 0$.) In the present paper, we aim to prove the conjecture of Im *et al.*, in the special case where the constant in the regression equation is zero.

As a preparation, consider now the special case of an AR(1) process, i.e. $p=0$ in (1.1) and (1.2), with fixed initial value X_0. Introducing O_p notation, $Y_T = O_p(T^\alpha)$ meaning that for each $\delta>0$, there exists a constant $A_\delta>0$ such that $P(|Y_T|\le A_\delta T^\alpha)\ge 1-\delta$ for all T (cf. Mann and Wald (1943)), it follows that if $\beta=0$, the log likelihood ratio test of H_0 against H_1 fulfills

$$-2\log Q_T(0,\underline{0}) = -T\log\left(1-\frac{1}{T}Z_T\right) = Z_T + O_p(T^{-1}) \tag{1.5}$$

(the last equality is obtained via Taylor expansion), with

$$Z_T \overset{\text{def}}{=} T\frac{(\sum S_{t-1}\varepsilon_t)^2}{\sum {\varepsilon_t}^2\sum S_{t-1}^2},$$

where the summation runs over $\{1\le t\le T\}$ and $S_t \overset{\text{def}}{=} \sum_{i=1}^t \varepsilon_i$ for all t. Moreover (cf. Phillips (1987)), as $T\to\infty$ (in the rest of Section 1, we put $\sigma^2=1$),

$$\frac{1}{T}\sum S_{t-1}\varepsilon_t \xrightarrow{d} \int_0^1 W_\tau dW_\tau = \frac{1}{2}(W_1^2-1), \tag{1.6}$$

(1.7) $$\frac{1}{T^2}\sum S_{t-1}^2 \xrightarrow{\text{d}} \int_0^1 W_\tau^2 d\tau,$$

(1.8) $$\frac{1}{T}\sum {\varepsilon_t}^2 \xrightarrow{\text{p}} 1$$

($\xrightarrow{\text{d}}$ and $\xrightarrow{\text{p}}$ mean convergence in distribution and conv. in probability, respectively), where $\{W_\tau\}$ is a standard Wiener process. Hence,

$$Z_T \xrightarrow{\text{d}} \frac{(\int_0^1 W_\tau dW_\tau)^2}{\int_0^1 W_\tau^2 d\tau} = \frac{(W_1^2 - 1)^2}{4\int_0^1 W_\tau^2 d\tau} \quad \text{as} \quad T \to \infty,$$

and because of (1.5), the same limit result holds for $-2\log Q_T(0, \underline{0})$. The limit distribution of Z_T and related quantities has been studied by e.g. Rao (1978), Evans and Savin (1981), Larsson (1995*a*) and Abadir (1995).

In this context, it is also interesting to note that, as was pointed out by Nielsen (1995), $\sum {\varepsilon_t}^2$ and Z_T are independent, and as a consequence, defining

$$\tilde{Z}_T \stackrel{\text{def}}{=} \frac{(\sum S_{t-1}\varepsilon_t)^2}{\sum S_{t-1}^2},$$

we get

(1.9) $$\begin{aligned} E(\tilde{Z}_T^n) &= E\left(\left(\frac{\sum {\varepsilon_t}^2}{T}\right)^n\right) E(Z_T^n) \\ &= \left(\frac{2}{T}\right)^n \frac{\Gamma\left(\frac{T}{2} + n\right)}{\Gamma\left(\frac{T}{2}\right)} E(Z_T^n), \qquad n = 1, 2, \ldots. \end{aligned}$$

The paper is organized as follows: in Section 2, the required result for the general (AR($p+1$)) model is proved. In Section 3, an expression for the first order (T^{-1}) error term, in the special case of an AR(2) ($p = 1$) process, is derived, and in Section 4, some concluding thoughts are given.

2. The general model

Consider an AR($p + 1$) model, written in error correction form as

(2.1) $$\Delta X_t = \beta X_{t-1} + \sum_{j=1}^{p} \rho_j \Delta X_{t-j} + \varepsilon_t,$$

with $\Delta X_t \stackrel{\text{def}}{=} X_t - X_{t-1}$, where the initial values $X_0, X_{-1}, \ldots, X_{-p}$ are fixed and the ε_t's ($t \geq 1$) are independent and normally distributed with mean 0 and variance σ^2. Assume that we have observations $X_1, \ldots, X_T$, and that we want to test $H_0 : \beta = 0$ against $H_1 : \beta < 0$. Moreover, assume that the ρ_j's fulfill

ASSUMPTION A. The roots of $1 - \rho_1 r - \cdots - \rho_p r^p = 0$ all have modulus larger than one.

Guaranteeing the stationarity of the $\{\Delta X_t\}$ process when $\beta = 0$, if the initial ΔX_t's are given the stationary distribution. Furthermore, let $Q_T(p, \underline{\rho})$ be the likelihood ratio test of H_0 against H_1. The theorem to be proved is

THEOREM 2.1. *Assume that $\beta = 0$ and that Assumption A holds. Then,*

(a) $$E(-2\log Q_T(p,\rho)) = E(-2\log Q_T(0,\underline{0})) + O(T^{-1}),$$

(b) $$\mathrm{Var}(-2\log Q_T(p,\rho)) = \mathrm{Var}(-2\log Q_T(0,\underline{0})) + O(T^{-1}).$$

In words, the theorem means that the expectation and variance of the log likelihood ratio test of H_0 against H_1 for the AR$(p+1)$ model in (2.1) may be approximated by the expectation and variance of the corresponding test in an AR(1) model, with an approximation error of order T^{-1}, proving the conjecture of Im *et al.* (1996) in the special case of no constant term in the regression equation.

Before proving the theorem, we start with some derivations and lemmas. In a standard manner it follows that,

$$-2\log Q_T = -T\log\left(\frac{\hat\sigma^2}{\hat\sigma_0^2}\right) = -T\log\left(1 - \frac{M_T}{T}\right), \qquad M_T \overset{\text{def}}{=} T\frac{\hat\sigma_0^2 - \hat\sigma^2}{\hat\sigma_0^2}, \tag{2.2}$$

where $\hat\sigma^2$, $\hat\sigma_0^2$ are the ML estimates of σ^2 under H_1 and H_0, respectively, and via Taylor expansion (M_T will turn out to be $O_p(1)$),

$$-2\log Q_T = M_T + O_p\left(\frac{M_T^2}{T}\right). \tag{2.3}$$

Furthermore, denoting the ML estimates of β and ρ_k, $k = 1, \ldots, p$ under H_1 by $\hat\beta$ and $\hat\rho_k$, respectively, we get

$$\begin{cases} \hat\beta \sum_t X_{t-1}^2 = \sum_t X_{t-1}\Delta X_t - \sum_j \hat\rho_j \sum_t X_{t-1}\Delta X_{t-j}, \\ \sum_j \hat\rho_j \sum_t \Delta X_{t-k}\Delta X_{t-j} = \sum_t \Delta X_{t-k}\Delta X_t \\ \qquad\qquad - \hat\beta \sum_t X_{t-1}\Delta X_{t-k}, \qquad k = 1, \ldots, p. \end{cases} \tag{2.4}$$

(Here, we denote the sum over $\{1 \le t \le T\}$ by $\sum_t$ and the sum over $\{1 \le j \le p\}$ by $\sum_j$.) Under H_0, the ML estimates of ρ_k, $k = 1, \ldots, p$, denoted $\hat\rho_{k0}$, satisfy

$$\sum_j \hat\rho_{j0} \sum_t \Delta X_{t-k}\Delta X_{t-j} = \sum_t \Delta X_{t-k}\Delta X_t, \qquad k = 1, \ldots, p. \tag{2.5}$$

Now, if $\beta = 0$ which we henceforth assume, (2.4) implies

$$(2.6)\qquad \begin{cases} \hat{\beta}\sum_t X_{t-1}^2 + \sum_j (\hat{\rho}_j - \rho_j)\sum_t X_{t-1}\Delta X_{t-j} = \sum_t X_{t-1}\varepsilon_t, \\ \hat{\beta}\sum_t X_{t-1}\Delta X_{t-k} + \sum_j (\hat{\rho}_j - \rho_j)\sum_t \Delta X_{t-j}\Delta X_{t-k} \\ \qquad = \sum_t \Delta X_{t-k}\varepsilon_t, \qquad k = 1, \ldots, p, \end{cases}$$

and by defining the p-dimensional vectors $\underline{\Delta X}_t \stackrel{\text{def}}{=} (\Delta X_{t-1}, \ldots, \Delta X_{t-p})'$, $\underline{Y} \stackrel{\text{def}}{=} T^{-1}\sum_t X_{t-1}\underline{\Delta X}_t$, $\underline{Z} \stackrel{\text{def}}{=} T^{-1}\sum_t \underline{\Delta X}_t\varepsilon_t$, $(\underline{\hat{\rho}} - \underline{\rho})' = (\hat{\rho}_1 - \rho_1, \ldots, \hat{\rho}_p - \rho_p)$, the $p \times p$ matrix $\Gamma_T \stackrel{\text{def}}{=} T^{-1}\sum_t \underline{\Delta X}_t\underline{\Delta X}_t'$ and the scalars $U_T \stackrel{\text{def}}{=} T^{-2}\sum_t X_{t-1}^2$ and $V_T \stackrel{\text{def}}{=} T^{-1}\sum_t X_{t-1}\varepsilon_t$, we may rewrite (2.6) as

$$(2.7)\qquad \begin{cases} T\hat{\beta}U_T + \underline{Y}'(\underline{\hat{\rho}} - \underline{\rho}) = V_T, \\ \hat{\beta}\underline{Y} + \Gamma_T(\underline{\hat{\rho}} - \underline{\rho}) = \underline{Z}, \end{cases}$$

and solving (2.7) for $\hat{\beta}$,

$$(2.8)\qquad T\hat{\beta} = \frac{V_T - \underline{Y}'\Gamma_T^{-1}\underline{Z}}{U_T - \frac{1}{T}\underline{Y}'\Gamma_T^{-1}\underline{Y}}.$$

But by (2.4) and (2.5), we have

$$\sum_j (\hat{\rho}_{j0} - \hat{\rho}_j)\sum_t \Delta X_{t-j}\Delta X_{t-k} = \hat{\beta}\sum_t X_{t-1}\Delta X_{t-k}, \qquad k = 1, \ldots, p,$$

or in matrix form, letting $(\underline{\hat{\rho}}_0 - \underline{\hat{\rho}})' \stackrel{\text{def}}{=} (\hat{\rho}_{10} - \hat{\rho}_1, \ldots, \hat{\rho}_{p0} - \hat{\rho}_p)$,

$$(2.9)\qquad \Gamma_T(\underline{\hat{\rho}}_0 - \underline{\hat{\rho}}) = \hat{\beta}\underline{Y}.$$

To find an expression for M_T (cf. (2.2)) in terms of these quantities, note that

$$(2.10)\qquad \begin{aligned} T(\hat{\sigma}_0^2 - \hat{\sigma}^2) &= \sum_t \left(\Delta X_t - \sum_k \hat{\rho}_{k0}\Delta X_{t-k}\right)^2 \\ &\quad - \sum_t \left(\Delta X_t - \hat{\beta}X_{t-1} - \sum_k \hat{\rho}_k\Delta X_{t-k}\right)^2 \\ &= 2\sum_t \left(\Delta X_t - \sum_k \hat{\rho}_{k0}\Delta X_{t-k}\right) \\ &\quad \cdot \left(\hat{\beta}X_{t-1} - \sum_k (\hat{\rho}_{k0} - \hat{\rho}_k)\Delta X_{t-k}\right) \\ &\quad - \sum_t \left(\hat{\beta}X_{t-1} - \sum_k (\hat{\rho}_{k0} - \hat{\rho}_k)\Delta X_{t-k}\right)^2. \end{aligned}$$

Here, via (2.9) and (2.8),

$$
\begin{aligned}
&\sum_t \left(\hat{\beta} X_{t-1} - \sum_k (\hat{\rho}_{k0} - \hat{\rho}_k) \Delta X_{t-k} \right)^2 \\
&\quad = \sum_t \left(\hat{\beta} X_{t-1} - \sum_k (\underline{\hat{\rho}}_0 - \underline{\hat{\rho}})' \underline{\Delta X}_t \right)^2 \\
&\quad = \hat{\beta}^2 \sum_t (X_{t-1} - \underline{Y}'\Gamma_T^{-1}\underline{\Delta X}_t)(X_{t-1} - \underline{Y}'\Gamma_T^{-1}\underline{\Delta X}_t)' \\
&\quad = T^2 \hat{\beta}^2 \left(U_T - \frac{1}{T}\underline{Y}'\Gamma_T^{-1}\underline{Y} \right) = \frac{(V_T - \underline{Y}'\Gamma_T^{-1}\underline{Z})^2}{U_T - \frac{1}{T}\underline{Y}'\Gamma_T^{-1}\underline{Y}},
\end{aligned}
$$

and similarly, since by (2.5) and arguments as above,

$$
(\underline{\hat{\rho}}_0 - \underline{\rho}) = \Gamma_T^{-1}\underline{Z}, \tag{2.11}
$$

we find

$$
\begin{aligned}
&\sum_t \left(\Delta X_t - \sum_k \hat{\rho}_{k0} \Delta X_{t-k} \right) \left(\hat{\beta} X_{t-1} - \sum_k (\hat{\rho}_{k0} - \hat{\rho}_k) \Delta X_{t-k} \right) \\
&\quad = \sum_t \left(\varepsilon_t - \sum_k (\hat{\rho}_{k0} - \rho_k) \Delta X_{t-k} \right) \left(\hat{\beta} X_{t-1} - \sum_k (\hat{\rho}_{k0} - \hat{\rho}_k) \Delta X_{t-k} \right) \\
&\quad = \hat{\beta} \sum_t (\varepsilon_t - \underline{Z}'\Gamma_T^{-1}\underline{\Delta X}_t)(X_{t-1} - \underline{Y}'\Gamma_T^{-1}\underline{\Delta X}_t)' \\
&\quad = T\hat{\beta}(V_T - \underline{Y}'\Gamma_T^{-1}\underline{Z}) = \frac{(V_T - \underline{Y}'\Gamma_T^{-1}\underline{Z})^2}{U_T - \frac{1}{T}\underline{Y}'\Gamma_T^{-1}\underline{Y}}.
\end{aligned}
$$

Thus, by (2.2) and (2.10),

$$
M_T = \frac{1}{\hat{\sigma}_0^2} \frac{(V_T - \underline{Y}'\Gamma_T^{-1}\underline{Z})^2}{U_T - \frac{1}{T}\underline{Y}'\Gamma_T^{-1}\underline{Y}}. \tag{2.12}
$$

In the following, the plan is to study the convergence of different components of M_T in (2.12), starting with

LEMMA 2.1. *If $\beta = 0$ and Assumption* A *holds,*

$$
\hat{\sigma}_0^2 = \frac{1}{T} \sum_t {\varepsilon_t}^2 + O_p(T^{-1}).
$$

PROOF. Using (2.1) with $\beta = 0$, it follows that, via (2.11),

$$(2.13)\quad T\hat{\sigma}_0^2 = \sum_t \left(\Delta X_t - \sum_k \hat{\rho}_{k0}\Delta X_{t-k}\right)^2$$
$$= \sum_t \left(\varepsilon_t - \sum_k (\hat{\rho}_{k0} - \rho_k)\Delta X_{t-k}\right)^2$$
$$= \sum_t (\varepsilon_t - \underline{Z}'\Gamma_T^{-1}\underline{\Delta X}_t)(\varepsilon_t - \underline{Z}'\Gamma_T^{-1}\underline{\Delta X}_t)' = \sum_t {\varepsilon_t}^2 - T\underline{Z}'\Gamma_T^{-1}\underline{Z}$$
$$= \sum_t {\varepsilon_t}^2 - T(\underline{\hat{\rho}}_0 - \underline{\rho})'\Gamma_T(\underline{\hat{\rho}}_0 - \underline{\rho}).$$

Now, by the results of Anderson (1971), Chapter 5, we have that if $\beta = 0$ and Assumption A holds, $\underline{\hat{\rho}}_0 - \underline{\rho} = O_p(T^{-1/2})$ and as $T \to \infty$, Γ_T converges in probability to Γ (say), the covariance matrix of $\{\Delta X_1, \ldots, \Delta X_p\}$. Hence, the lemma follows. □

Now, let $S_t \overset{\text{def}}{=} \sum_{i=1}^t \varepsilon_i$, where the ε_i's are independent and normally distributed with mean 0 and variance σ^2. In the following, we may without loss of generality assume that $\sigma^2 = 1$. It will turn out to be useful to consider the moving average representations of X_t and ΔX_t given in

LEMMA 2.2. *If $\beta = 0$ and Assumption* A *holds*,

(a) $$\Delta X_t = C_t(L)\varepsilon_t + a_t,$$

(b) $$X_t = C_t(L)S_t + b_t,$$

where

$$a_t \overset{\text{def}}{=} \sum_{s=1}^p c_{t-s}d_s, \qquad d_s \overset{\text{def}}{=} (\rho_s\Delta X_0 + \cdots + \rho_p\Delta X_{s-p}), \qquad b_t \overset{\text{def}}{=} \sum_{s=1}^p c_{t-s}e_s,$$

$$e_s \overset{\text{def}}{=} (\rho_s X_0 + \cdots + \rho_p X_{s-p}), \qquad C_t(L) \overset{\text{def}}{=} \sum_{i=0}^{t-1} c_i L^i,$$

where L is the lag operator, $c_0 = 1$ and c_n is defined recursively through

$$c_n = \sum_{j=1}^{\min(p,n)} c_{n-j}\rho_j, \qquad n = 1, 2, \ldots.$$

Moreover, we have the representations

(c) $$C_t(L) = C_t(1) + (1-L)C_t^{(1)}(L), \qquad C_t^{(1)}(L) = \sum_{i=0}^{t-1} c_i^{(1)} L^i,$$

(d) $$C_t^{(1)}(L) = C_t^{(1)}(1) + (1-L)C_t^{(2)}(L), \qquad C_t^{(2)}(L) = \sum_{i=0}^{t-1} c_i^{(2)} L^i,$$

where $c_i^{(1)} \stackrel{\text{def}}{=} -\sum_{j=i+1}^{t-1} c_j$, $c_i^{(2)} \stackrel{\text{def}}{=} -\sum_{j=i+1}^{t-1} c_j^{(1)}$, $i = 0, \ldots, t-1$. *Furthermore, for some* $\delta > 0$, *the sums*

$$C_\infty(L) \stackrel{\text{def}}{=} \lim_{t \to \infty} C_t(L) = \sum_{i=0}^{\infty} c_i L^i,$$

$$C_\infty^{(m)}(L) \stackrel{\text{def}}{=} \lim_{t \to \infty} C_t^{(m)}(L) = \sum_{i=0}^{\infty} c_i^{(m)} L^i, \qquad m = 1, 2$$

are absolutely convergent for $|L| < 1 + \delta$, *and* c_i, $c_i^{(1)}$ *and* $c_i^{(2)}$ *tend to zero exponentially fast as* $i \to \infty$.

Proof. Theorem 2.1 of Johansen (1995) yields (a), and (b) is obtained by summation in (a). For (c), just note that

$$C_t(L) - C_t(1) = \sum_{i=0}^{t-1} c_i (L^i - 1) = -(1 - L) \sum_{i=0}^{t-1} c_i (1 + L + \cdots + L^{i-1}).$$

Now, let

$$\sum_{i=0}^{t-1} c_i^{(1)} L^i = C_t^{(1)}(L) = -\sum_{i=0}^{t-1} c_i (1 + L + \cdots + L^{i-1}),$$

and by equating coefficients,

$$c_i^{(1)} = -\sum_{j=i+1}^{t-1} c_j, \qquad i = 0, 1, \ldots, t-1.$$

By repeating the same argument, (d) follows.

The convergence of $C_\infty(L)$, $C_\infty^{(1)}(L)$ and $C_\infty^{(2)}(L)$ follows from Theorem 2.2 of Johansen (1995), and the exponential decay of c_i, $c_i^{(1)}$ and $c_i^{(2)}$ is an immediate consequence of this convergence. □

We will also need

Lemma 2.3.

(a) $$\sum_t (C_t(1) S_{t-1}) \varepsilon_t = C_T(1) \sum_t S_{t-1} \varepsilon_t + O_p(1),$$

(b) $$\sum_{t=1}^{T-1} (C_t(1) S_t)^2 = C_{T-1}(1)^2 \sum_{t=1}^{T-1} S_{t-1}^2 + O_p(1).$$

Observe that, by (1.6) and (1.7), $\sum_t S_{t-1} \varepsilon_t = O_p(T)$ and $\sum_t S_{t-1}^2 = O_p(T^2)$, respectively.

PROOF. To prove (a), we have for a start

$$\sum_t (C_t(1)S_{t-1})\varepsilon_t = \sum_{i=0}^{T-1} c_i \sum_{t=i+1}^{T} S_{t-1}\varepsilon_t = C_T(1)\sum_t S_{t-1}\varepsilon_t - v,$$

where

$$v \stackrel{\text{def}}{=} \sum_{i=0}^{T-1} c_i \sum_{t=1}^{i} S_{t-1}\varepsilon_t = \sum_{t=1}^{T-1} S_{t-1}\varepsilon_t \sum_{i=t}^{T-1} c_i.$$

We want to prove that v is $O_p(1)$. But because the c_i's decay exponentially fast as $i \to \infty$, we may write $c_i \sim \gamma^i$ for some $|\gamma| < 1$, where $f_n \sim g_n$ means that $f_n/g_n \to 1$ as $n \to \infty$, and so

$$v \sim \sum_{t=1}^{T-1} S_{t-1}\varepsilon_t \sum_{i=t}^{T-1} \gamma^i \sim \frac{1}{1-\gamma} \sum_{t=1}^{T-1} \gamma^t S_{t-1}\varepsilon_t.$$

Hence, v is asymptotically normal with mean zero and variance (assuming $\sigma^2 = 1$ for simplicity)

$$E(v^2) \sim \frac{1}{(1-\gamma)^2} \sum_{t=1}^{T-1} \gamma^{2t} E(S_{t-1}^2 {\varepsilon_t}^2) = \frac{1}{(1-\gamma)^2} \sum_{t=1}^{T-1} \gamma^{2t}(t-1),$$

which is finite as $T \to \infty$, and we conclude that v is $O_p(1)$.

Looking at (b), we get

$$\begin{aligned}\sum_{t=1}^{T-1} (C_t(1)S_t)^2 &= \sum_{t=1}^{T-1} \left(\sum_{i=0}^{t-1} c_i\right)^2 S_t^2 \\ &= \sum_{i=0}^{T-2}\sum_{j=0}^{T-2} c_i c_j \sum_{t=i\vee j+1}^{T-1} S_t^2 = C_{T-1}(1)^2 \sum_t S_t^2 - u,\end{aligned}$$

($i \vee j$ means the maximum of i and j) where

$$u \stackrel{\text{def}}{=} \sum_{i=0}^{T-2}\sum_{j=0}^{T-2} c_i c_j \sum_{t=1}^{i\vee j} S_t^2 = \sum_{t=1}^{T-2} S_t^2 \left(\sum_{i=t}^{T-2} c_i^2 + 2\sum_{i=t}^{T-2}\sum_{j=0}^{i-1} c_i c_j\right).$$

But arguing as above, it is easily seen that all moments of u remain finite as $T \to \infty$, hence u is $O_p(1)$ which was to be shown. □

The next thing to show is

LEMMA 2.4. *If $\beta = 0$ and Assumption* A *holds,*

$$\hat{\sigma}_0^2 M_T = \frac{(\sum_t S_{t-1}\varepsilon_t + C_\infty(1)^{-1}(R - T^{-1}y'\Gamma^{-1}\underline{z}))^2}{\sum_t S_{t-1}^2} + O_p(T^{-1}),$$

where Γ *is the covariance matrix of* $\{\Delta X_t\}_{t=1}^p$,

$$R \stackrel{\text{def}}{=} \sum_{i=0}^{T-1} c_i^{(1)} \sum_{t=i+1}^{T} \varepsilon_{t-1-i}\varepsilon_t = O_p(T^{1/2}),$$

and $\underline{y}$ *and* $\underline{z}$ *are p-dimensional vectors with components*

$$y_k \stackrel{\text{def}}{=} \sum_{i=0}^{T-1}\sum_{j=0}^{T-1} c_i c_j \sum_{i\vee j+1}^{T} S_{t-1-i}\varepsilon_{t-k-j} = O_p(T)$$

and

$$z_k \stackrel{\text{def}}{=} \sum_{i=0}^{T-1} c_i \sum_{t=i+1}^{T} \varepsilon_{t-k-i}\varepsilon_t = O_p(T^{1/2})$$

for $k = 1, \ldots, p$. *(We regard all* ε_t, $t \leq 0$, *as fixed.)*

PROOF. We will start by showing

$$U_T = \frac{1}{T^2} C_{T-1}(1)^2 \sum_t S_{t-1}^2 + O_p(T^{-1}). \tag{2.14}$$

(Note that, by (1.7), $T^{-2}\sum_t S_{t-1}^2$ is $O_p(1)$.) To this end, using Lemma 2.2, we get

$$\begin{aligned} T^2 U_T &= \sum_t X_{t-1}^2 = \sum_{t=1}^{T-1} (C_t(L)S_t + b_t)^2 \\ &= \sum_{t=1}^{T-1} ((C_t(1) + (1-L)C_t^{(1)}(L))S_t)^2 + r_1 \\ &= \sum_{t=1}^{T-1} (C_t(1)S_t)^2 + 2\sum_{t=1}^{T-1} (C_t(1)S_t)(C_t^{(1)}(L)\varepsilon_t) \\ &\quad + \sum_{t=1}^{T-1} C_t^{(1)}(L)^2 {\varepsilon_t}^2 + r_1, \end{aligned} \tag{2.15}$$

where

$$r_1 \stackrel{\text{def}}{=} 2R_1 + \sum_{t=1}^{T-1} b_t^2, \qquad R_1 \stackrel{\text{def}}{=} \sum_{t=1}^{T-1} (C_t(L)S_t)b_t.$$

The first term on the r.h.s. of (2.15) is treated in Lemma 2.3(b). As for the second term, Lemma 2.2 implies

$$\begin{aligned} \sum_{t=1}^{T-1} (C_t(1)S_t)(C_t^{(1)}(L)\varepsilon_t) &= \sum_{i=0}^{T-2}\sum_{j=0}^{T-2} c_i c_j^{(1)} \sum_{t=i\vee j+1}^{T-1} S_t\varepsilon_t \\ &\quad + \sum_{i=0}^{T-2} c_i \sum_{t=i+1}^{T-1} S_t(C_t^{(2)}(L)\Delta\varepsilon_t), \end{aligned} \tag{2.16}$$

where

$$\sum_{i=0}^{T-2}\sum_{j=0}^{T-2} c_i c_j^{(1)} \sum_{t=i\vee j+1}^{T-1} S_t\varepsilon_t = \sum_{i=0}^{T-2}\sum_{j=0}^{T-2} c_i c_j^{(1)} \sum_{t=1}^{T-1} S_t\varepsilon_t - \sum_{i=0}^{T-2}\sum_{j=0}^{T-2} c_i c_j^{(1)} \sum_{t=1}^{i\vee j} S_t\varepsilon_t, \tag{2.17}$$

which is $O_p(T)$, since via (1.6) and (1.8),

$$\sum_{t=1}^{T-1} S_t\varepsilon_t = \sum_{t=1}^{T-1} S_{t-1}\varepsilon_t + \sum_{t=1}^{T-1} {\varepsilon_t}^2 = O_p(T),$$

and because the second term on the r.h.s. of (2.17) may be proved to be $O_p(1)$ in the same fashion as the rest terms v and u in the proof of Lemma 2.3.

Furthermore,

$$\sum_{t=1}^{T-1} S_t(C_t^{(2)}(L)\Delta\varepsilon_t) = \sum_{i=0}^{T-2} c_i^{(2)} \sum_{t=i+1}^{T-1} S_t\Delta\varepsilon_{t-i}, \tag{2.18}$$

and by partial summation,

$$\sum_{t=i+1}^{T-1} S_t\Delta\varepsilon_{t-i} = S_T\varepsilon_{T-1-i} - \sum_{t=i+1}^{T-1} \varepsilon_{t+1}\varepsilon_{t-i}. \tag{2.19}$$

The r.h.s. terms are $\leq O_p(T^{1/2})$, the first one since S_T is normally distributed with mean zero and variance T, and the second one because of the central limit theorem for m-dependent sequences (cf. Chung (1974)). Hence, via (2.18),

$$\left|\sum_{t=1}^{T-1} S_t(C_t^{(2)}(L)\Delta\varepsilon_t)\right| \leq \sum_{i=0}^{T-2} |c_i^{(2)}| \left|\sum_{t=i+1}^{T-1} S_t\Delta\varepsilon_{t-i}\right| \leq O_p(T^{1/2}),$$

since $\sum_{i=0}^{\infty} |c_i^{(2)}|$ is convergent by Lemma 2.2, so by (2.16) and (2.17), the second term on the r.h.s. of (2.15) is $O_p(T)$. For the third term, letting $C_t^{(1)}(L)^2 = (\sum_{i=0}^{t-1} c_i^{(1)} L^i)^2 \stackrel{\text{def}}{=} \sum_{i=0}^{2(t-1)} c_i^{(1,2)} L^i$, we simply note that

$$\sum_{t=1}^{T-1} C_t^{(1)}(L)^2 {\varepsilon_t}^2 = \sum_{t=1}^{T-1}\sum_{i=0}^{2(t-1)} c_i^{(1,2)}\varepsilon_{t-i}^2 = \sum_{t=1}^{T-1} {\varepsilon_t}^2 \sum_{i=0}^{2(T-2)} c_i^{(1,2)} = O_p(T),$$

from (1.8) and the convergence of $C_t^{(1)}(L)$. Moreover,

$$R_1 = \sum_{t=1}^{T-1}\sum_{i=0}^{t-1} c_i S_{t-i} b_t = \sum_{t=1}^{T-1} S_t \sum_{i=0}^{T-2} c_i b_{t+i},$$

and as above, $c_i \sim \gamma^i$ for $|\gamma| < 1$, implying

$$b_t \sim \sum_{s=1}^{p} \gamma^{t-s} e_s \sim \gamma^{t-p} e_p \quad \text{as} \quad t \to \infty.$$

Hence,

$$R_1 \sim \gamma^{-p} e_p \sum_{t=1}^{T-1} \gamma^t S_t \sum_{i=0}^{T-2} \gamma^{2i} = O_p(1),$$

arguing as in the proof of Lemma 2.3. Similarly, as for the second term in the r_1 expression,

$$\sum_{t=1}^{T-1} b_t^2 \sim \gamma^{-2p} e_p^2 \sum_{t=1}^{T-1} \gamma^{2t} = O(1),$$

completing the proof of (2.14).

Looking at V_T, we have via Lemma 2.2,

$$\begin{aligned} TV_T &= \sum_t X_{t-1}\varepsilon_t = \sum_t (C_t(L)S_{t-1} + b_t)\varepsilon_t \\ &= \sum_t (C_t(1)S_{t-1})\varepsilon_t + R + \sum_t b_t \varepsilon_t, \end{aligned} \tag{2.20}$$

where the first term on the r.h.s. is dealt with in Lemma 2.3(a), and where

$$R = \sum_t (C_t^{(1)}(L)\varepsilon_{t-1})\varepsilon_t = \sum_{i=0}^{T-1} c_i^{(1)} \sum_{t=i+1}^{T} \varepsilon_{t-1-i}\varepsilon_t, \tag{2.21}$$

(the "initial" ε_t's, $\{\varepsilon_t\}_{t \le 0}$, may be regarded as fixed), which is $O_p(T^{1/2})$ by the CLT for m-dependent sequences and the convergence of $\sum_{i=0}^{\infty} |c_i^{(1)}|$. The quantity $\sum_t b_t \varepsilon_t$ is normally distributed with mean zero and variance $\sum_t b_t^2 = O(1)$, hence $O_p(1)$.

Moreover, by Lemma 2.2 again we have that, for the k-th component, Y_k say, of $\underline{Y}$, $k = 1, \ldots, p$,

$$\begin{aligned} TY_k &= \sum_t (C_t(L)S_{t-1} + b_{t-1})(C_t(L)\varepsilon_{t-k} + a_{t-k}) \\ &= y_k + \sum_t C_t(L)S_{t-1}a_{t-k} + \sum_t b_{t-1}(C_t(L)\varepsilon_{t-k} + a_{t-k}), \end{aligned} \tag{2.22}$$

where, in the same manner as above, the second and third r.h.s. terms are seen to be $O_p(1)$. Moreover, via Lemma 2.2,

$$y_k \stackrel{\text{def}}{=} \sum_{i=0}^{T-1} \sum_{j=0}^{T-1} c_i c_j \sum_{t=i \vee j+1}^{T} S_{t-1-i}\varepsilon_{t-k-j} = \sum_t (C_t(L)S_{t-1})(C_t(L)\varepsilon_{t-k})$$

$$= \sum_t (C_t(1)S_{t-1})(C_t(1)\varepsilon_{t-k}) + \sum_t (C_t^{(1)}(L)\varepsilon_{t-1})(C_t(1)\varepsilon_{t-k})$$
$$+ \sum_t (C_t(1)S_{t-1})(C_t^{(1)}(L)\Delta\varepsilon_{t-k}) + \sum_t (C_t^{(1)}(L)\varepsilon_{t-1})(C_t^{(1)}(L)\Delta\varepsilon_{t-k})$$
$$= \sum_{i=0}^{T-1}\sum_{j=0}^{T-1} c_i c_j \sum_{t=i\vee j+1}^{T} S_{t-1}\varepsilon_{t-k} + \sum_{i=0}^{T-1}\sum_{j=0}^{T-1} c_i^{(1)} c_j \sum_{t=i\vee j+1}^{T} \varepsilon_{t-1-i}\varepsilon_{t-k}$$
$$+ \sum_{i=0}^{T-1}\sum_{j=0}^{T-1} c_i c_j^{(1)} \sum_{t=i\vee j+1}^{T} S_{t-1}\Delta\varepsilon_{t-k-j}$$
$$+ \sum_{i=0}^{T-1}\sum_{j=0}^{T-1} c_i^{(1)} c_j^{(1)} \sum_{t=i\vee j+1}^{T} \varepsilon_{t-1-i}\Delta\varepsilon_{t-k-j} \le O_p(T),$$

because of the convergence of $\sum_{i=0}^{\infty}|c_i|$, etc., and since

$$\sum_t S_{t-1}\varepsilon_{t-k} = \sum_t S_{t-k-1}\varepsilon_{t-k} + \sum_{l=1}^{k}\sum_t \varepsilon_{t-l}\varepsilon_{t-k} \le O_p(T)$$

(all S_t, $t \le 0$ are interpreted as zero), $\sum_t \varepsilon_{t-1-i}\varepsilon_{t-k} \le O_p(T)$ (indeed, the order is T if.f. $k = i+1$, otherwise it is $T^{1/2}$), $\sum_t S_{t-1}\Delta\varepsilon_{t-k-j} \le O_p(T^{1/2})$ by partial summation as in (2.19), and finally, by the Hölder inequality,

$$\left|\sum_t \varepsilon_{t-1-i}\Delta\varepsilon_{t-k-j}\right| \le \left(\sum_t \varepsilon_{t-1-i}^2 \sum_t (\Delta\varepsilon_{t-k-j})^2\right)^{1/2} \le O_p(T).$$

Moreover, from Lemma 2.2, the k-th component of $\underline{Z}$, Z_k, satisfies

$$TZ_k = \sum_t (C_t(L)\varepsilon_{t-k} + a_{t-k})\varepsilon_t = z_k + \sum_t a_{t-k}\varepsilon_t, \tag{2.23}$$

where

$$z_k \stackrel{\text{def}}{=} \sum_{i=0}^{T-1} c_i \sum_{t=i+1}^{T} \varepsilon_{t-k-i}\varepsilon_t,$$

which, for the same reasons as R (cf. (2.21)), is $O_p(T^{1/2})$, and where $\sum_t a_{t-k}\varepsilon_t$ is normally distributed with mean zero and variance of order 1, hence $O_p(1)$.

Furthermore, as before, Γ_T converges to the covariance matrix Γ in probability as $T \to \infty$, so via Lemma 2.3, insertion of (2.14) and (2.20)–(2.23) into (2.12) and from the convergence of $C_T(1)$ (cf. Lemma 2.2), the proof is completed. □

Looking at the magnitudes of the terms in the numerator of the main term of Lemma 2.4, (1.6) yields $\sum_t S_{t-1}\varepsilon_t = O_p(T)$, and from Lemma 2.4, we know that $R = O_p(T^{1/2})$, $\underline{y} = O_p(T)$ and $\underline{z} = O_p(T^{1/2})$, hence the term $R^* \stackrel{\text{def}}{=} C_\infty(1)^{-1}(R - T^{-1}\underline{y}'\Gamma^{-1}\underline{z})$ is $O_p(T^{1/2})$. However, when taking moments, R^* gives a slighter impact than what would be expected from this, because of the following lemma.

LEMMA 2.5. *The correlation of R with $\sum_t S_{t-1}\varepsilon_t$, $\sum_t {\varepsilon_t}^2$ and $\sum_t S_{t-1}^2$ as well as the correlations of the components of $\underline{z}$ with $\sum_t S_{t-1}\varepsilon_t$, $\sum_t {\varepsilon_t}^2$, $\sum_t S_{t-1}^2$ and the components of $\underline{y}$ are all at most of order $T^{-1/2}$.*

PROOF. We will confine ourselves to handle $\mathrm{Corr}(R, \sum_t S_{t-1}\varepsilon_t)$, the rest of the proof being similar. (In fact, the correlations of the components of $\underline{z}$ with $\sum_t {\varepsilon_t}^2$ equal zero.) As is easily seen, $ER = 0 = E(\sum_t S_{t-1}\varepsilon_t)$, $\mathrm{Var}\, R = O(T)$ and $\mathrm{Var}(\sum_t S_{t-1}\varepsilon_t) = O(T^2)$ (cf. (1.6)). Hence, we need to deduce that $E(R\sum_t S_{t-1}\varepsilon_t) \leq O(T)$. But this follows from the calculation

$$
\begin{aligned}
E\left(R\sum_t S_{t-1}\varepsilon_t\right) &= \sum_s\sum_t E\left(\left(\sum_{i=0}^{t-1} c_i^{(1)}\varepsilon_{s-1-i}\right)\varepsilon_s S_{t-1}\varepsilon_t\right) \\
&= \sum_t\sum_{i=0}^{t-1} c_i^{(1)} E(\varepsilon_{t-1-i}^2)E({\varepsilon_t}^2) \\
&= \sum_t\sum_{i=0}^{t-1} c_i^{(1)} = O(T),
\end{aligned}
$$

the last equality following from the convergence of $\sum_{i=0}^{\infty} |c_i^{(1)}|$. □

As a simple consequence of the lemma, the correlation of R with any almost surely smooth function of $\sum_t S_{t-1}$, $\sum_t {\varepsilon_t}^2$ and $\sum_t S_{t-1}^2$ is at most of order $T^{-1/2}$, and similarly for the components of $\underline{z}$. Using this argument, we may now readily prove

THEOREM 2.2. *If $\beta = 0$, Assumption* A *holds and n is an arbitrary positive integer,*

$$E(M_T^n) = E(Z_T^n) + O(T^{-1}),$$

where

$$Z_T = T\frac{(\sum_t S_{t-1}\varepsilon_t)^2}{\sum_t {\varepsilon_t}^2 \sum_t S_{t-1}^2}.$$

PROOF. Lemma 2.1 and Lemma 2.4 imply

$$M_T = T\frac{(\sum_t S_{t-1}\varepsilon_t + R^*)^2}{\sum_t {\varepsilon_t}^2 \sum_t S_{t-1}^2} + O_p(T^{-1}),$$

and since $\sum_t S_{t-1}^2$ is $O_p(T^2)$ (by (1.7)), $\sum_t {\varepsilon_t}^2$ and $\sum_t S_{t-1}\varepsilon_t$ are $O_p(T)$ (via (1.8) and (1.6), respectively) and R^* is $O_p(T^{1/2})$, a binomial expansion yields

$$E(M_T^n) = E(Z_T^n) + 2nE(A_n R^*) + O(T^{-1}), \tag{2.24}$$

letting

$$A_n \stackrel{\text{def}}{=} T^n \frac{(\sum_t S_{t-1}\varepsilon_t)^{2n-1}}{(\sum_t {\varepsilon_t}^2 \sum_t S_{t-1}^2)^n} = O_p(T^{-1}).$$

However,

$$E(A_nR^*) = E(A_nR) - T^{-1}E(A_n\underline{y}'\Gamma^{-1}\underline{z}),$$

and, as is implied by Lemma 2.5, $\mathrm{Corr}(A_n, R) = O_p(T^{-1/2})$, so because $ER = 0$, $\mathrm{Var}\, R = O(T)$ and $\mathrm{Var}\, A_n = O(T^{-2})$,

$$E(A_nR) = \mathrm{Corr}(A_n, R)\sqrt{\mathrm{Var}\, A_n \,\mathrm{Var}\, R} = O(T^{-1}).$$

Similarly, by applying Lemma 2.4 componentwise, it follows that

$$T^{-1}E(A_n\underline{y}'\Gamma^{-1}\underline{z}) = O(T^{-1}).$$

Hence, (2.24) yields the result of the theorem. □

Finally, via (2.3) (observe that $M_T^n = O_p(1)$), we may translate the result for M_T to corresponding result for the log likelihood test statistic $-2\log Q_T$:

COROLLARY 2.1. *If* $\beta = 0$ *and Assumption* A *holds*,

(a) $$E(-2\log Q_T) = E(Z_T) + O(T^{-1}),$$

(b) $$\mathrm{Var}(-2\log Q_T) = \mathrm{Var}(Z_T) + O(T^{-1})$$

In conjunction with (1.5), the corollary immediately implies Theorem 2.1.

Remark 1. Because of (1.9), it is clear that the results of Theorem 2.2 and Corollary 2.1 may be rephrased as

$$E(M_T^n) = \left(\frac{T}{2}\right)^n \frac{\Gamma\left(\frac{T}{2}\right)}{\Gamma\left(\frac{T}{2}+n\right)} E(\tilde{Z}_T^n) + O(T^{-1}), \tag{2.25}$$

$$E(-2\log Q_T) = E(\tilde{Z}_T) + O(T^{-1}), \tag{2.26}$$

$$\mathrm{Var}(-2\log Q_T) = \mathrm{Var}(\tilde{Z}_T) + O(T^{-1}), \tag{2.27}$$

where

$$\tilde{Z}_T = \frac{(\sum_t S_{t-1}\varepsilon_t)^2}{\sum_t S_{t-1}^2}.$$

Remark 2. In fact, it is evident that all moments of $-2\log Q_T$ may be approximated by the corresponding moments of Z_T up to an error of order T^{-1}. This fact strongly suggests that also the distribution function of $-2\log Q_T$ (and M_T) may be approximated by the distribution function of Z_T up to order T^{-1} (cf. Larsson (1995*b*) for this kind of result for unit root testing in AR(1) processes).

3. The AR(2) case: The size of the error term

Naturally, to get a better feeling for the size of the $O(T^{-1})$ approximation errors of Section 2 (e.g. in connection with Bartlett correction, cf. Larsson (1994)), one could try to compute these terms numerically. However, a problem is that these error terms will depend on the nuisance parameters $\rho_1, \ldots, \rho_p$ of the model (2.1).

To simplify matters, let us study the special case of an AR(2) process ($p = 1$), with fixed initial values X_0 and X_{-1}. Under $H_0 : \beta = 0$, $\rho = \rho_1$ is the only parameter, and Assumption A translates to the condition $|\rho| < 1$. We are able to deduce the result

THEOREM 3.1. *If $\beta = 0$ and $|\rho| < 1$, we have for a process $\{X_t\}$ satisfying* (2.1) *with $p = 1$, $\rho = \rho_1$,*

$$EM_T = EZ_T + \frac{R(\rho)}{T} + O(T^{-2}),$$

where

$$R(\rho) = R(0) + \rho R'(0) + O(\rho^2)$$

with

$$\begin{aligned} R(0) = \lim_{T\to\infty}\Bigg(& T^2 E\left(\frac{1}{\sum_t S_{t-1}^2}\right) + 2TE\left(\frac{\sum_t S_{t-1}\varepsilon_t}{\sum_t S_{t-1}^2}\right) + 2E\left(\frac{(\sum_t S_{t-1}\varepsilon_t)^2}{\sum_t S_{t-1}^2}\right) \\ & + T^2 E\left(\frac{(\sum_t S_{t-1}\varepsilon_t)^2}{(\sum_t S_{t-1}^2)^2}\right) + 2TE\left(\frac{(\sum_t S_{t-1}\varepsilon_t)^3}{(\sum_t S_{t-1}^2)^2}\right) \\ & + E\left(\frac{(\sum_t S_{t-1}\varepsilon_t)^4}{(\sum_t S_{t-1}^2)^2}\right) \\ & - 2TE\left(\frac{\sum_t S_{t-1}\varepsilon_t}{\sum_t S_{t-1}^2}\sum_t \varepsilon_{t-1}\varepsilon_t\right) - 2E\left(\frac{(\sum_t S_{t-1}\varepsilon_t)^2}{\sum_t S_{t-1}^2}\sum_t \varepsilon_{t-1}\varepsilon_t\right)\Bigg) \end{aligned}$$

and

$$\begin{aligned} R'(0) = 2\lim_{T\to\infty}\Bigg(& T^2 E\left(\frac{1}{\sum_t S_{t-1}^2}\right) + TE\left(\frac{\sum_t S_{t-1}\varepsilon_t}{\sum_t S_{t-1}^2}\right) + E\left(\frac{(\sum_t S_{t-1}\varepsilon_t)^2}{\sum_t S_{t-1}^2}\right) \\ & + T^2 E\left(\frac{(\sum_t S_{t-1}\varepsilon_t)^2}{(\sum_t S_{t-1}^2)^2}\right) + 2TE\left(\frac{(\sum_t S_{t-1}\varepsilon_t)^3}{(\sum_t S_{t-1}^2)^2}\right) \\ & + E\left(\frac{(\sum_t S_{t-1}\varepsilon_t)^4}{(\sum_t S_{t-1}^2)^2}\right) \\ & - TE\left(\frac{\sum_t S_{t-1}\varepsilon_t}{\sum_t S_{t-1}^2}\sum_t \varepsilon_{t-1}\varepsilon_t\right) - E\left(\frac{(\sum_t S_{t-1}\varepsilon_t)^2}{\sum_t S_{t-1}^2}\sum_t \varepsilon_{t-1}\varepsilon_t\right) \\ & - TE\left(\frac{\sum_t S_{t-1}\varepsilon_t}{\sum_t S_{t-1}^2}\sum_t \varepsilon_{t-2}\varepsilon_t\right) \\ & - E\left(\frac{(\sum_t S_{t-1}\varepsilon_t)^2}{\sum_t S_{t-1}^2}\sum_t \varepsilon_{t-2}\varepsilon_t\right)\Bigg). \end{aligned}$$

Remark 3. Observe that the initial values enter none of the terms $R(0)$ or $R'(0)$.

Remark 4. Via the Taylor expansion (cf. (2.3))

$$-2\log Q_T = -T\log\left(1-\frac{M_T}{T}\right) = M_T + \frac{M_T^2}{2T} + O_p(T^{-2}),$$

a similar result for $-2\log Q_T$ is easily obtained.

Remark 5. In Larsson (1994), Laplace transform methods were employed to calculate $R(0)$ to be ≈ 1.241. Moreover, from the numerical results of Larsson (1994), giving all but the last three terms in the expression for $R'(0)$ of Theorem 3.1, and the simulation results (using $T = 800$ and 1.000.000 replications)

$$\lim_{T\to\infty} TE\left(\frac{\sum_t S_{t-1}\varepsilon_t}{\sum_t S_{t-1}^2}\sum_t \varepsilon_{t-2}\varepsilon_t\right) \approx 5.6$$

and

$$\lim_{T\to\infty} E\left(\frac{(\sum_t S_{t-1}\varepsilon_t)^2}{\sum_t S_{t-1}^2}\sum_t \varepsilon_{t-2}\varepsilon_t\right) \approx -1.3$$

(as an alternative, these terms may be calculated more accurately by generalizing the method of Larsson (1994)), we arrive at the approximation $R'(0) \approx 3.7$.

PROOF OF THEOREM 3.1. The proof consists of Taylor expansions of the components of (2.12) in ρ, which are then inserted into (2.12), giving the result by a final Taylor expansion.

For a start, we see that, from the proof of (2.14) (somewhat generalized),

$$T^2 U_T = C_T(1)^2 \sum_t S_{t-1}^2 + 2C_{T-1}(1)C_{T-1}^{(1)}(1)\sum_t S_t\varepsilon_t + C_T^{(1)}(1)^2 \sum_t {\varepsilon_t}^2 + O_p(1). \tag{3.1}$$

Here, because $c_t = \rho^t$, we have by Lemma 2.2,

$$c_t^{(1)} = -\sum_{j=t+1}^{T-1} \rho^j \approx -\frac{\rho^{t+1}}{1-\rho} \tag{3.2}$$

($\approx$ signifying equality modulo exponentially small errors in T),

$$C_T(1) = \sum_{t=0}^{T-1} \rho^t \approx \frac{1}{1-\rho} \tag{3.3}$$

and

$$C_T^{(1)}(1) = -\sum_{t=0}^{T-1} \frac{\rho^{t+1}}{1-\rho} \approx -\frac{\rho}{(1-\rho)^2}. \tag{3.4}$$

Now, insertion of (3.3) and (3.4) into (3.1) yields

$$\begin{aligned} T^2 U_T &= \frac{1}{(1-\rho)^2} \sum_t S_{t-1}^2 - \frac{2\rho}{(1-\rho)^3} \sum_t S_t \varepsilon_t + \frac{\rho^2}{(1-\rho)^4} \sum_t {\varepsilon_t}^2 + O_p(1) \\ &= \frac{1}{(1-\rho)^2} \left(\sum_t S_{t-1}^2 - 2\rho \sum_t S_t \varepsilon_t + O(\rho^2) \right) + O_p(1). \end{aligned} \tag{3.5}$$

As for V_T, we have in view of (2.20),

$$\sum_t (C_t(1) S_{t-1}) \varepsilon_t = \sum_{i=0}^{T-1} \rho^i \sum_{t=i+1}^{T} S_{t-1} \varepsilon_t \approx \frac{1}{1-\rho} \sum_t S_{t-1} \varepsilon_t,$$

and by (2.21) and (3.2),

$$R \approx -\frac{1}{1-\rho} \sum_{i=0}^{T-1} \rho^{i+1} \sum_{t=i+1}^{T} \varepsilon_{t-1-i} \varepsilon_t = \frac{1}{1-\rho} \left(\rho \sum_t \varepsilon_{t-1} \varepsilon_t + O(\rho^2) \right),$$

so (2.20) implies

$$T V_T = \frac{1}{1-\rho} \left(\sum_t S_{t-1} \varepsilon_t - \rho \sum_t \varepsilon_{t-1} \varepsilon_t + O(\rho^2) \right) + O_p(1). \tag{3.6}$$

As for the covariance matrix Γ, it consists, in this special case, of one element only, which is easily shown to be

$$\Gamma = \frac{1}{1-\rho^2}, \tag{3.7}$$

and moreover, (2.22) implies

$$\begin{aligned} T Y_1 &= \sum_{i=0}^{T-1} \sum_{j=0}^{T-1} \rho^{i+j} \sum_{t=i \vee j+1}^{T} S_{t-1-i} \varepsilon_{t-1-j} + O_p(1) \\ &= \sum_t S_t \varepsilon_t + \rho \left(\sum_t S_{t-1} \varepsilon_t + \sum_t S_t \varepsilon_{t-1} \right) + O(\rho^2) + O_p(1), \end{aligned} \tag{3.8}$$

and via (2.23),

$$\begin{aligned} T Z_1 &= \sum_{i=0}^{T-1} \rho^i \sum_{t=i+1}^{T} \varepsilon_{t-1-i} \varepsilon_t + O_p(1) \\ &= \sum_t \varepsilon_{t-1} \varepsilon_t + \rho \sum_t \varepsilon_{t-2} \varepsilon_t + O(\rho^2) + O_p(1). \end{aligned} \tag{3.9}$$

Finally, by inserting (3.7) and (3.9) into (2.13) and using that $T^{-1}(\sum \varepsilon_{t-1}\varepsilon_t)^2$ tends to one in probability as $T \to \infty$,

$$\begin{aligned}(3.10)\qquad \hat{\sigma}_0^2 &= \frac{1}{T}\sum_t {\varepsilon_t}^2 - \frac{1}{T^2}\frac{(TZ_1)^2}{\Gamma} \\ &= \frac{1}{T}\sum_t {\varepsilon_t}^2 - \frac{1}{T}\left(1 + \frac{2\rho}{T}\sum_t \varepsilon_{t-1}\varepsilon_t \sum_t \varepsilon_{t-2}\varepsilon_t + O(\rho^2)\right) \\ &\quad + O_p(T^{-2}),\end{aligned}$$

and the result of the theorem is obtained by inserting (3.5)–(3.10) into (2.12), Taylor expanding in ρ, taking expectation and neglecting terms of the type

$$\frac{1}{T}E\left(\frac{(\sum_t S_{t-1}\varepsilon_t)^2}{\sum_t S_{t-1}^2}\sum_t \varepsilon_{t-1}\varepsilon_t \sum_t \varepsilon_{t-2}\varepsilon_t\right)$$

for reasons as in Lemma 2.5. (Details may be provided from the author at request.) □

4. Concluding remarks

The results of the present paper may be generalized in many different directions. The obvious one is to study the contribution of nuisance parameters (as in Section 3) further, perhaps using transform methods as in Larsson (1994, 1997). One may also try to include a constant term, a linear term and so on in the model (cf. Larsson (1997) and Nielsen (1996) for related results for AR(1) processes). Yet another line of generalization is to study vector-valued autoregressive (VAR) processes, where the unit root test carries over to a test for cointegration (cf. Johansen (1995)).

Acknowledgements

I am grateful to Hashem Pesaran, who inspired me to study this problem, and to the two anonymous referees, whose comments helped to improve the paper.

References

Abadir, K. M. (1995). The limiting distribution of the t ratio under a unit root, *Econometric Theory*, **11**, 775–793.

Anderson, T. W. (1971). *The Statistical Analysis of Time Series*, Wiley, New York.

Chung, K. L. (1974). *A Course in Probability Theory*, 2nd ed., Academic Press, London.

Evans, G. B. A. and Savin, N. E. (1981). The calculation of the limiting distribution of the least squares estimator of the parameter in a random walk model, *Ann. Statist.*, **9**, 1114–1118.

Im, K. S., Pesaran, M. H. and Shin, Y. (1996). Testing for unit roots in heterogeneous panels, Unpublished manuscript, University of Cambridge.

Johansen, S. (1995). *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*, Oxford University Press, Oxford.

Larsson, R. (1994). Bartlett corrections for unit root test statistics, Preprint no. 2, Institute of Mathematical Statistics, University of Copenhagen.

Larsson, R. (1995*a*). The asymptotic distributions of some test statistics in near-integrated AR processes, *Econometric Theory*, **11**, 306–330.

Larsson, R. (1995*b*). Small sample corrections for unit root test statistics, Report no. 13, Department of Mathematics, Uppsala University.

Larsson, R. (1997). On the asymptotic expectations of some unit root tests in a first order autoregressive process in the presence of trend, *Ann. Inst. Statist. Math.*, **49**, 585–599.

Mann, H. B. and Wald, A. (1943). On stochastic limit and order relationships, *Ann. Math. Statist.*, **14**, 217–226.

Nielsen, B. (1995). Bartlett correction of the unit root test in autoregressive models, Discussion Paper no. 98, Nuffield College.

Nielsen, B. (1996). Moments of unit root tests in autoregressive models with drifts, Unpublished manuscript, Institute of Mathematical Statistics, University of Copenhagen.

Phillips, P. C. B. (1987). Time series regression with a unit root, *Econometrica*, **55**, 277–301.

Rao, M. M. (1978). Asymptotic distribution of an estimator of the boundary parameter of an unstable process, *Ann. Statist.*, **6**, 185–190.