Communications in Statistics—Simulation and Computation®, 00: 1–16, 2012

ISSN: 0361-0918 print / 1532-4141 online
DOI: 10.1080/03610918.2012.655828

Developments in Maximum Likelihood Unit Root Tests

YING ZHANG,[1] HAO YU,[2] AND A. IAN McLEOD[2]

[1]Department of Mathematics and Statistics, Acadia University, Wolfville, Nova Scotia, Canada
[2]Department of Statistical and Actuarial Sciences, University of Western Ontario, London, Ontario, Canada

The exact maximum likelihood estimate provides a test statistic for the unit root test that is more powerful than the usual least-squares approach. In this article, a new derivation is given for the asymptotic distribution of this test statistic that is simpler and more direct than the previous method. The response surface regression method is used to obtain a fast algorithm that computes accurate finite-sample critical values. This algorithm is available in the R package `mleur` *that is available on CRAN. The empirical power of the new test is shown to be much better than the usual test not only in the normal case but also for innovations generated from an infinite variance stable distribution as well as for innovations generated from a GARCH(1,1) process.*

Keywords Exact maximum likelihood estimator; Response surface regression; Robust unit root test; Symbolic computation.

Mathematics Subject Classification Primary 62M10; Secondary 91B84.

1. Introduction

The AR(1) model is widely used in many applications as well as in unit root testing. Modern approaches to the unit root testing problem emphasize the importance of model selection (Enders, 2010; Patterson, 2010; Pfaff, 2006). This article focuses on testing the null model known as random walk,

$$\nabla z_t = a_t, \quad t = 1, 2, \ldots, \tag{1}$$

where $\nabla z_t = z_t - z_{t-1}$ and a_t are independent and normally distributed with mean zero and variance σ_a^2. The alternative is assumed to be the stationary AR(1) model with intercept term β,

$$z_t = \beta + \rho z_{t-1} + a_t, \quad t = 1, 2, \ldots, \tag{2}$$

where $|\rho| < 1$.

Sometimes it is assumed that $\beta = 0$ is known. This case corresponds to the zero-mean AR(1) processes. Both of these models were discussed in the original formulation of the

Received December 19, 2010; Accepted December 28, 2011
Address correspondence to Ying Zhan, Department of Mathematics and Statistics, Acadia University, Wolfville, NS, B4P 2R6, Canada; E-mail: ying.zhang@acadiau.ca

unit root testing problem by Dickey and Fuller (1979) but using the least-squares estimates (LSE) instead of the maximum likelihood estimate (MLE). The random walk model and the stationary AR(1) alternative provide a suitable family of models for many financial and economic time series. However, as is discussed in Sec. 6, other methods are needed if the diagnostic checks reveal that further lagged values need to be included in the model.

Fuller (1996, p. 577) indicates that if the objective is to test the hypothesis of a unit root against the alternative of a stationary process with an unknown mean, the test statistics associated with the exact MLE are more powerful than that with the LSE. The exact MLE referred to is the MLE in the stationary case that corresponds to the alternative hypothesis in the unit root test. Empirical power comparisons among various unit root tests showed that the MLE-based tests had much higher power than the Dickey–Fuller (DF) tests (Pantula et al., 1994). Extensions of the MLE method to the ARMA(1, 1) and other autoregressive moving average (ARMA) processes were discussed by Shin and Fuller (1998).

Fuller (1996, Sec. 10.1.3) and Gonzalez-Farias and Dickey (1999) derive the limiting distributions of normalized statistics associated with the exact MLE unit root test under Eqs. (1) and (2). This approach is indirect whereas our new derivation in Sec. 3 is essentially simpler and more direct. Our method using the Taylor series linearization of the test statistic is carried out through symbolic computer algebra. The exact MLE itself is also derived symbolically through the solution of a cubic equation in Sec. 2. The usual approach to the exact MLE using a numerical optimization technique can occasionally have convergence problems. This more direct approach using a symbolic Taylor series linearization is easier to generalize to other problems as well. It is known that computer algebra may handle complicated statistical inference problems (Andrews and Stafford, 2000). There are several examples in time series analysis. Smith and Field (2001) show how a symbolic operator can be used to calculate the joint cumulants of the linear combinations of products of discrete Fourier transforms. Zhang and McLeod (2006) discuss a computer algebra approach to the asymptotic bias and variance coefficients to order $O(1/n)$ for linear estimators in stationary time series. Computer algebra no doubt has many more applications in statistics and time series analysis.

In Sec. 4, using response surface curves, we show that the critical values for the MLE test may be efficiently computed. With our fast algorithm, in Sec. 5, we demonstrate that the exact MLE test provides not only a sizeable increase in power but also the robustness against alternative specifications for the innovations such as an infinite variance stable distribution and a GARCH(1, 1) process. We illustrate how to implement the exact MLE unit root test with two real world examples in Sec. 6.

2. Exact MLE

The AR(1) model (2) may also be written as

$$z_t - \mu = \rho(z_{t-1} - \mu) + a_t, t = 1, 2, \ldots, \tag{3}$$

where $E(z_t) = \mu$ and $\beta = \mu(1 - \rho)$. When μ is known, without loss of generality, it is assumed that $\mu = 0$. The time series process is stationary if $|\rho| < 1$. In the random walk case, $\rho = 1$ and the process is said to be unit root nonstationary. If $\rho > 1$, the process is explosively nonstationary.

Most of unit root tests have been derived under the data generation model,

$$z_t - \mu = \rho(z_{t-1} - \mu) + a_t, t = 1, 2, \ldots, \tag{4}$$

where $z_0 = \mu$ is a fixed value. The only difference between model (3) and (4) is the initial value. The time series represented by (4) is mostly same as that by (3), except that under (4) the process is asymptotically stationary when $|\rho| < 1$ and the LSE is the maximum likelihood estimator of ρ conditionally on the initial value.

First consider the zero-mean stationary time series under (3). Its initial value follows a normally distributed random variable with zero mean and a variance of $\sigma^2/(1-\rho^2)$. The exact log-likelihood function of n consecutive observations, $z_t, t = 1, \ldots, n$, may be written as (Minozzo and Azzalini, 1993)

$$l(\sigma^2, \rho) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) + \frac{1}{2}\log(1-\rho^2) - \frac{1}{2\sigma^2}(a - 2\rho b + \rho^2 c), \qquad (5)$$

where

$$a = \sum_{t=1}^{n} {z_t}^2, \quad b = \sum_{t=2}^{n} z_t z_{t-1}, \quad c = \sum_{t=2}^{n-1} {z_t}^2. \qquad (6)$$

Maximizing $l(\sigma^2, \rho)$ in Eq. (5), White (1961) and Minozzo and Azzalini (1993) show that the exact MLE of ρ is the unique real root of the following equation, whose absolute value is less than one:

$$\frac{n-1}{n}c\rho^3 - \frac{n-2}{n}b\rho^2 - \left(c + \frac{a}{n}\right)\rho + b = 0. \qquad (7)$$

Dent and Min (1978), Hasza (1980), and Minozzo and Azzalini (1993) point out that the exact MLE may be written as

$$\hat{\rho} = 2\left(\frac{{d_2}^2 - 3d_1}{9}\right)^{1/2} \cos\left(\frac{\theta}{3} + \frac{4\pi}{3}\right) - \frac{d_2}{3}, \qquad (8)$$

where

$$\theta = \cos^{-1}\left\{\frac{9d_2d_1 - 27d_0 - 2{d_2}^3}{2({d_2}^2 - 3d_1)^{3/2}}\right\}$$

and

$$d_2 = -\frac{n-2}{n-1}\frac{b}{c}, \quad d_1 = -\frac{n}{n-1}\left(1 + \frac{a}{nc}\right), \quad d_0 = \frac{n}{n-1}\frac{b}{c}.$$

Using *Mathematica* (Wolfram, 1999), the cubic Eq. (7) is easily solved and the exact MLE $\hat{\rho}$ may be expressed as the ratio of complex polynomials,

$$\begin{aligned}\hat{\rho} = {} & (n-2)b/3c(-1+n) + ((1-i\sqrt{3}) \\ & (-b^2(-2+n)^2 + 3c(-1+n)(-a-cn)))/(32^{2/3}c(1-n) \\ & (16b^3 - 18abc - 24b^3n + 27abcn + 9bc^2n + 12b^3n^2 - 9abcn^2 \\ & - 27bc^2n^2 - 2b^3n^3 + 18bc^2n^3 + (((16b^3 - 18abc - 24b^3n + 27abcn \\ & + 9bc^2n + 12b^3n^2 - 9abcn^2 - 27bc^2n^2 - 2b^3n^3 + 18bc^2n^3)^2 + 4(-b^2(-2+n)^2 \\ & + 3c(-1+n)(-a-cn))^3))^{1/2})^{1/3}) - 1/(62^{1/3}c(1-n)) \\ & ((1+i\sqrt{3})(16b^3 - 18abc - 24b^3n + 27abcn + 9bc^2n + 12b^3n^2\end{aligned}$$

$$- 9abcn^2 - 27bc^2n^2 - 2b^3n^3 + 18bc^2n^3 + (((16b^3 - 18abc -$$
$$- 24b^3n + 27abcn + 9bc^2n + 12b^3n^2 - 9abcn^2 - 27bc^2n^2 - 2b^3n^3$$
$$+ 18bc^2n^3)^2 + 4(-b^2(-2+n)^2 + 3c(-1+n)(-a-cn))^3))^{1/2})^{1/3}),$$

where $i = \sqrt{-1}$, and a, b, and c are defined in (6).

For a stationary AR(1) process with an unknown mean under (3), there are two mean correction methods: sample mean correction and the maximum likelihood mean estimation. It is well known that for ARMA(p, q) model, the sample mean is asymptotically efficient (Brockwell and Davis, 1987, Sec. 7.1). The exact MLE for the μ may be obtained iteratively as in McLeod and Zhang (2008) but in the AR(1) case the sample mean has close to 100% efficiency in finite samples (McLeod and Zhang, 2008, Table 3). For speed and convenience, we may just consider the sample mean estimator in Eq. (3). That is, the exact MLE is the $\hat{\rho}$ described above with $z_t - \overline{z}_n$ $(t = 1, \ldots, n)$ replacing z_t where $\overline{z}_n$ is the sample mean, which is denoted as $\hat{\rho}_\mu$.

Under the stationary alternative, the exact MLE and the LSE have the same limiting distribution (Brockwell and Davis, 1987, Sec. 8) but this is not the case under the nonstationary null hypothesis Eq. (1). The next section provides a new derivation of this distribution.

3. Computer Algebra Derivations to Limiting Distributions

In the unit root case, $\rho = 1$, we consider the random walk

$$z_t = z_{t-1} + a_t, \quad t = 1, 2, \ldots, \tag{9}$$

where $\{a_t\}$ is a sequence of i.i.d. random variables with mean 0 and finite variance $\sigma_a^2 > 0$.

Fixing $z_0 = 0$, the random walk process may be generated by

$$z_t = \sum_{j=1}^{t} a_j. \tag{10}$$

For the zero-mean case, the normalized and pivotal type statistics may be written as

$$\hat{\delta} = n(\hat{\rho} - 1), \tag{11}$$

where $\hat{\rho}$ is described in Sec. 2, and

$$\hat{\tau} = \frac{1}{\hat{\sigma}}\left(\sum_{t=2}^{n} z_{t-1}^2\right)^{1/2}(\hat{\rho} - 1), \tag{12}$$

where

$$\hat{\sigma}^2 = (n-2)^{-1}\sum_{t=2}^{n}(z_t - \hat{\rho} z_{t-1})^2.$$

For the unknown mean case, the normalized statistic may be written as

$$\hat{\delta}_\mu = n(\hat{\rho}_\mu - 1), \tag{13}$$

where $\hat{\rho}_\mu$ is described in Sec. 2, and the corresponding pivotal statistic may be written by

$$\hat{\tau}_\mu = \hat{\sigma}_\mu^{-1}\left(\sum_{t=2}^{n}(z_{t-1}-\overline{z}_n)^2\right)^{1/2}(\hat{\rho}_\mu - 1), \tag{14}$$

where

$$\hat{\sigma}_\mu^2 = (n-3)^{-1}\sum_{t=2}^{n}(z_t - \overline{z}_n - \hat{\rho}_\mu(z_{t-1}-\overline{z}_n))^2.$$

The limiting distributions of statistics in Eqs. (11)–(14) are given in Theorems 1 and 2 below.

Theorem 3.1. *Under a random walk* (9),

$$n(\hat{\rho}_\mu - 1) \xrightarrow{D} \frac{1}{2}\left(\mathfrak{C}_\mu - \sqrt{\mathfrak{C}_\mu^2 - 4\mathfrak{C}_\mu + 2\mathfrak{B}_\mu}\right), \tag{15}$$

$$\hat{\tau}_\mu \xrightarrow{D} \frac{\sqrt{\mathfrak{A}_\mu}}{2}\left(\mathfrak{C}_\mu - \sqrt{\mathfrak{C}_\mu^2 - 4\mathfrak{C}_\mu + 2\mathfrak{B}_\mu}\right), \tag{16}$$

where

$$\begin{aligned}
\mathfrak{A}_\mu &= \int_0^1 W^2(t)\,dt - \left(\int_0^1 W(t)\,dt\right)^2, \\
\mathfrak{B}_\mu &= \mathfrak{A}_\mu^{-1}\left(\left(\int_0^1 W(t)\,dt\right)^2 + \left(W(1) - \int_0^1 W(t)\,dt\right)^2\right), \\
\mathfrak{C}_\mu &= \mathfrak{A}_\mu^{-1}\left(\frac{1}{2}\left(W^2(1)-1\right) - W(1)\int_0^1 W(t)\,dt + \left(\int_0^1 W(t)\,dt\right)^2\right),
\end{aligned}$$

and $\{W(t),\ 0 \le t \le 1\}$ *is a standard Wiener process.*

Theorem 3.2. *Under a random walk* (9),

$$n(\hat{\rho} - 1) \xrightarrow{D} \frac{1}{2}\left(\mathfrak{C} - \sqrt{\mathfrak{C}^2 - 4\mathfrak{C} + 2\mathfrak{B}}\right), \tag{17}$$

$$\hat{\tau} \xrightarrow{D} \frac{\sqrt{\mathfrak{A}}}{2}\left(\mathfrak{C} - \sqrt{\mathfrak{C}^2 - 4\mathfrak{C} + 2\mathfrak{B}}\right), \tag{18}$$

where

$$\begin{aligned}
\mathfrak{A} &= \int_0^1 W(t)^2\,dt, \\
\mathfrak{B} &= \mathfrak{A}^{-1}W(1)^2, \\
\mathfrak{C} &= \mathfrak{A}^{-1}\frac{1}{2}\left(W^2(1)-1\right)
\end{aligned}$$

and $\{W(t),\ 0 \le t \le 1\}$ *is a standard Wiener process.*

To fix ideas, below is demonstrated how *Mathematica* helping to prove the limiting distribution of $n(\hat{\rho}_\mu - 1)$, Eq. (15) in Theorem 3.1.

$\hat{\rho}_\mu$ may be further simplified as follows:

$$\hat{\rho}_\mu = -\frac{(n-2)G}{3(1-n)} + \frac{(1-i\sqrt{3})u}{2^{\frac{2}{3}}3(1-n)(v+\sqrt{v^2+4u^3})^{\frac{1}{3}}} - \frac{(1+i\sqrt{3})(v+\sqrt{v^2+4u^3})^{\frac{1}{3}}}{2^{\frac{1}{3}}6(1-n)},$$

where

$$\begin{aligned} u =& \ -(n-2)^2G^2 + 3(1-n)(H+n), \\ v =& \ 16G^3 - 18GH - 24G^3n + 27GHn + 9Gn \\ & +12G^3n^2 - 9GHn^2 - 27Gn^2 - 2G^3n^3 + 18Gn^3, \end{aligned}$$

where

$$G = b/c, \quad H = a/c, \tag{19}$$

where a, b, and c are defined in (6) with $z_t - \bar{z}_n$ $(t = 1, \ldots, n)$ replacing z_t. The limiting distributions of G and H are given in the following lemma.

Lemma 3.1 *Under a random walk* (9),

$$(c/(\sigma^2 n^2), n(H-1), n(G-1)) \xrightarrow{D} (\mathfrak{A}_\mu, \mathfrak{B}_\mu, \mathfrak{C}_\mu), \tag{20}$$

where $\mathfrak{A}_\mu$, $\mathfrak{B}_\mu$, $\mathfrak{C}_\mu$ *are defined in Theorem* 3.1.

A detailed proof of Lemma 3.1 can be found in the Appendix.

Proof of Eq. (15) *in Theorem* 3.1. Let $W = n(G-1)$ and $X = n(H-1)$. Lemma 3.1 implies that $W = O_p(1)$ and $X = O_p(1)$. $\hat{\rho}_\mu$ can be considered as a function of $1/n$ with $1 + W/n$ and $1 + X/n$ replacing G and H. In order to obtain the limit distribution of $\hat{\rho}_\mu$, taking $\hat{\rho}_\mu$ with one-term Taylor expansion with respect to $1/n$ at zero,

$$\hat{\rho}_\mu = 1 + \frac{1}{2n}\left(W - \sqrt{W^2 - 4W + 2X}\right) + \frac{1}{n^2}R_n(W, X), \tag{21}$$

where $\sup_{n\geq 1}|R_n(W, X)| \leq C(W, X)$ that is a continuous function of W and X. Below is a *Mathematica* script and its output for deriving Eq. (21).

```
In[1]:= u = -(n-2)^2G^2+3(1-n)(H+n);
In[2]:= v = 16 G^3-18 G H - 24 G^3 n+27 G H n + 9G n + 12 G^3 n^2
          -9 G H n^2-27 G n^2-2 G^3 n^3+18 G n^3;
In[3]:= rho = -(-2+n)G/(3(1-n))
        +((1-i Sqrt[3])(u))/(3 2^(2/3)(1-n)(v+Sqrt[v^2+4u^3])^(1/3))
        -1/(6 2^(1/3)(1-n))((1+i Sqrt[3])(v+Sqrt[v^2+4u^3])^(1/3));
In[4]:= G = 1 + W/n; H = 1 + X/n; n = 1/z;
        Simplify[Series[rho, {z, 0, 1}]]
```

and the output of the final input is

```
Out[4]= 1+1/2(W+i(4W-W^2-2X)^(1/2))z+O[z]^2
```

which leads to Eq. (21). Following the fact that $W = O_p(1)$, $X = O_p(1)$, and the continuity of $C(W, X)$,

$$\frac{1}{n} R_n(W, X) \xrightarrow{P} 0.$$

By Lemma 1,

$$(W, X) \xrightarrow{D} (\mathfrak{C}_\mu, \mathfrak{B}_\mu).$$

Thus, applying the continuous mapping theorem described in the Appendix and the Slutsky's theorem to Eq. (21), Eq. (15) is obtained.

The limiting distributions of $\hat{\tau}_\mu$, $\hat{\rho}$, and $\hat{\tau}$ in Eqs. (16)–(18) can be very similarly derived.

Fuller (1996, Theorem 10.1.10 and Corollary 10.1.10) shows that Eq. (15) and Eq. (17) hold, which indicates that the computer algebra derivations implemented here are appropriate. Other than the normalized statistics, Eqs. (16) and (18) show the limiting distributions on the unit root boundary of pivotal statistics for both zero-mean and unknown mean cases.

4. Methods of Implementing the Test

The asymptotic distribution may be evaluated by computer simulation methods for Brownian motion. Such methods are discussed in the book by Iacus (2008). Then this asymptotic distribution could be used to obtain critical values and/or p-values for the test. As we will show below, this method will not work unless the series length is very long.

The simplest approach is to use a Monte Carlo test. Under general conditions, this approach provides an accurate test that can be efficiently computed using parallel processing capabilities found on many modern computer environments. For example, the necessary steps are outlined below for the normalized test:

(1) Simulate M random walks under (1) with the length of n and compute the simulated testing statistic sample, $n(\hat{\rho}_\mu^1 - 1), n(\hat{\rho}_\mu^2 - 1), \ldots, n(\hat{\rho}_\mu^M - 1)$.
(2) Compute the observed testing statistic value for the given time series $\{z_t\}, n(\hat{\rho}_\mu^0 - 1)$.
(3) Count the number of times k that the simulated test statistic $n(\hat{\rho}_\mu^i - 1)$ $(i = 1, \ldots, M)$ is less than or equal to the observed test statistic $n(\hat{\rho}_\mu^0 - 1)$.
(4) Compute the p-value as $(k + 1)/(M + 1)$.

Instead of using independent normal random variables to generate the random walks in Step (1), we could use a bootstrap sample of the residuals. This test has been implemented in the function `mctest` in our R package for MLE unit root tests (McLeod et al., 2011).

An even more computationally efficient approach is to use response surface regression (MacKinnon, 2000) to estimate the quantile functions for the exact distribution. The response surface regressions are of the form

$$Q^\alpha(n) = \theta_\infty + \theta_1/n + \theta_2/n^2 + \theta_3/n^3 + \epsilon,$$

where $Q^\alpha(n)$ is an α percentile of the finite-sample distribution that is estimated using N replications and ϵ is an error term. The curve was fit with the cluster computer SHARCNET (Shared Hierarchical Academic Research Computing Network) utilizing 221 compute nodes for about 1 hour. Thirty-six series lengths n used were 20, (5), 100, (20), 300, (50), 500, (100), 1000. For each series length, $N = 200{,}000$ replications were done and this was repeated $M = 100$ times. From this the mean and variance of each percentile were estimated and used in a weighted least-squares regression to obtain the final fitted regression. The weighted least-squares approach is needed to account for heteroscedasticity in the error terms.

In the case of the model specified in Eqs. (1) and (2), the critical values for the test statistic $\hat{\tau}_\mu$ given in Eq. (14) are

$$\hat{Q}^\alpha(n) = \begin{cases} -3.110 - 4.652/n - 51.466/n^2 \text{ 1\% point} \\ -2.531 - 2.062/n - 17.529/n^2 \text{ 5\% point} \\ -2.233 - 1.219/n - 8.178/n^2 \text{ 10\% point.} \end{cases} \quad (22)$$

Fig. 1 illustrates these curves for series lengths up to 500. The dashed line shows the critical point from the asymptotic distribution. It is seen that a reasonably large sample is needed to obtain accurate critical values using the asymptotic distribution. The y-axis in each panel is scaled so scaling unit is the same. This scaling reveals the critical values corresponding 10% converge more quickly while the 1% critical values converge slowly.

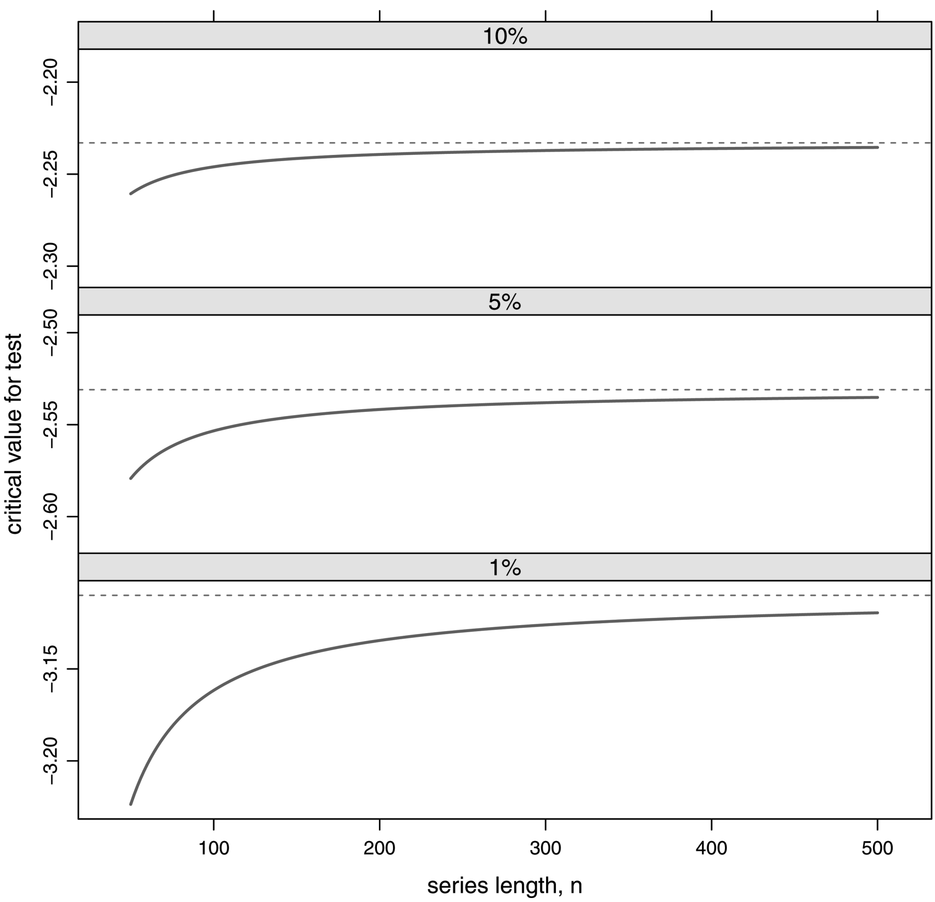

Figure 1. The 1%, 5%, and 10% critical values for the MLE test statistic $\hat{\tau}_\mu$.

Extensive simulation experiments were performed for a variety of series lengths, n, and parameters, ρ, to check that the p-values produced by the Monte Carlo method agreed with that produced by the critical values from Eq. (22).

Implementing the explicit expression of the exact MLE derived in Sec. 2 and the critical value equations such as Eq. (22), our R function `mleur` for the MLE unit root tests is available in our R package `mleur` (McLeod et al., 2011).

5. Power Comparisons

We investigated the power of the MLE unit root tests under various types of innovations in comparison with that of the standard DF test. Under our null model (1), and alternative model (2) or (3), the unknown mean case is more realistic than the known mean case. Thus, the MLE unit root test was implemented with the sample mean correction in the normalized form $n(\hat{\rho}_\mu - 1)$ or the pivotal form $\hat{\tau}_\mu$, denoted by MLEn or MLEp, respectively. In R, the standard DF test is implemented in several packages and usually the pivotal form of test statistic is used. We used the implementation of the DF pivotal test for the same model as (1) and (2) with an unknown mean or interpret in the R package `urca` by Pfaff (2010), represented by DF in this article. The function `GetPower` for making such power comparisons is given in our package (McLeod et al., 2011).

Figure 2. Time series plot for money velocity.

Table 1

Empirical power of 5% unit root tests based on 25,000 simulations using innovations from various distributions. Tests were Dickey–Fuller (DF), MLE normalized (MLEn), and MLE pivotal (MLEp). The distributions used were standard normal, stable distribution with index parameter 1.5, and a GARCH(1, 1) process. The 0.95 level MOE for the table percentage is 0.62

		Normal			Stable			GARCH(1, 1)		
n	ρ	DF	MLEn	MLEp	DF	MLEn	MLEp	DF	MLEn	MLEp
30	0.65	39.8	56.5	59.6	36.5	55.7	59.4	42.0	56.2	59.0
70	0.65	97.6	99.8	99.7	97.7	98.6	98.1	95.6	98.9	98.8
100	0.65	100.0	100.0	100.0	99.7	99.3	99.1	99.7	100.0	99.9
200	0.65	100.0	100.0	100.0	99.9	99.7	99.6	100.0	100.0	100.0
30	0.85	12.1	16.6	18.3	11.6	12.5	13.6	14.1	18.3	20.0
70	0.85	37.4	55.1	57.4	33.6	53.9	57.0	39.7	55.9	57.8
100	0.85	63.2	83.2	84.2	65.2	84.3	84.6	64.5	81.4	82.1
200	0.85	99.6	100.0	100.0	99.4	98.9	98.4	98.7	99.7	99.6
30	0.90	8.4	10.7	11.9	8.9	8.3	8.9	9.8	11.8	13.1
70	0.90	19.4	29.8	31.4	17.4	24.6	26.8	22.0	32.0	33.7
100	0.90	33.3	51.0	52.8	29.7	49.4	52.8	36.3	51.9	53.5
200	0.90	86.8	97.2	97.0	89.3	95.7	94.8	84.3	94.7	94.7
30	0.95	6.7	7.6	8.3	6.5	5.5	5.7	7.8	8.1	9.1
70	0.95	9.2	12.5	13.3	9.0	9.5	9.9	11.0	14.7	15.4
100	0.95	12.5	19.0	19.8	11.8	14.2	15.1	14.6	21.2	22.0
200	0.95	32.5	51.1	52.5	28.9	47.7	50.7	36.0	52.6	53.9
30	1.00	5.5	4.9	5.5	6.3	4.3	4.4	7.0	6.1	6.7
70	1.00	5.2	5.1	5.3	6.0	3.7	3.7	7.0	6.1	6.4
100	1.00	5.0	5.3	5.6	6.0	3.9	3.8	6.7	6.3	6.3
200	1.00	4.9	4.8	4.9	5.9	3.8	3.8	6.2	6.2	6.3

In constructing critical value (Eq. (22)), the simulated series were assumed to be Gaussian. But since the asymptotic distribution only relies on the assumption that the innovations are independent with mean zero and finite variance σ_a^2, it is plausible that the critical values given in Eq. (22) may also be applicable for other nonnormal distributions with finite variance. In fact, using our R function `GetPower`, we found no difference from the normal distribution results with Student's t on five degrees of freedom. A more challenging question is how well these results continue to hold when these assumptions are not met as in the case of infinite variance distributions, or series exhibiting conditional heteroscedasticity and nonlinear dependence. To answer this question, a portion of our simulation results is shown in Table 1. A total of 25,000 replications were done for a series of lengths $n = 30, 70, 100, 200$ and parameters $\rho = 0.65, 0.85, 0.9, 0.95, 1.0$ for the innovations generated by a stable distribution and a Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model described in the following. With so many replications the 95% margin of error (MOE) was about 0.0062 or 0.62 in percentage terms. These computations took less than 3 hours on a multicore PC.

The random variable Z has a stable distribution with index α, scale $\sigma > 0$, skewness $|\beta| < 1$, and location $\mu \in \mathcal{R}$, if its characteristic function is given by

$$\mathrm{E}\,(e^{itZ}) = \begin{cases} \exp\left\{-\sigma|t|^{\alpha}\ \left(1 - i\beta\ \ \mathrm{sgn}(t)\ \tan\frac{\pi\alpha}{2}\right) + i\mu t\right\} & \text{if } \alpha \neq 1 \\ \exp\left\{-\sigma|t|\ \left(1 + i\beta\frac{2}{\pi}\ \ \mathrm{sgn}(t)\ \log|t|\right) + i\mu t\right\} & \text{if } \alpha = 1, \end{cases}$$

where

$$\mathrm{sgn}(t) = \begin{cases} 1 \text{ if} & t > 0 \\ 0 \text{ if} & t = 0 \\ -1 \text{ if} & t < 0. \end{cases}$$

Since it has been suggested that many financial time series appear to have a stable distribution with α in the range (1.35, 1.75), α was set to 1.5 for our simulations. Also, $\sigma = 1$, $\beta = 0$, and $\mu = 0$.

A GARCH(1, 1) sequence a_t, $t = \ldots, -1, 0, 1, \ldots$ is of the form

$$a_t = \sigma_t \epsilon_t$$

and

$$\sigma_t^2 = \omega + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2,$$

where we took ϵ_t to be independent standard normal, $\omega = 10^{-6}$, $\alpha_1 = 0.2$, and $\beta_1 = 0.7$. The parameters were chosen to approximate models that have been used in actual applications.

Table 1 shows that there can be substantial difference in power between the MLE unit root test and the DF tests not only in the normal case but also for innovations generated from an infinite variance stable distribution as well as for innovations generated from a GARCH(1,1) process. It is observed that the size of the test is slightly inflated for the nonnormal case, so this needs to be taken into account in the power comparison. In general, it appears that the pivotal form of the test statistic, MLEp, is preferable to the normalized form, MLEn. MLEp is just as robust as MLEn and has slightly better power.

Further empirical power analysis may easily be carried out similarly with our R function `GetPower`.

6. Illustrative Applications

In actual applications, it is recommended that diagnostic checks should be done for residual autocorrelation. If there is significant autocorrelation in the residuals of the fitted AR(1) model, then other methods such as the augmented DF test must be used. The model building procedure needed for this DF test family is discussed by Pfaff (2006) and is available in the R package `urca` (Pfaff, 2010). Our R package `mleur` (McLeod et al., 2011) provides suitable model diagnostic checks for applying the MLE root test and is used in the applications discussed below. R scripts to generate the analyses reported below are available in our package documentation.

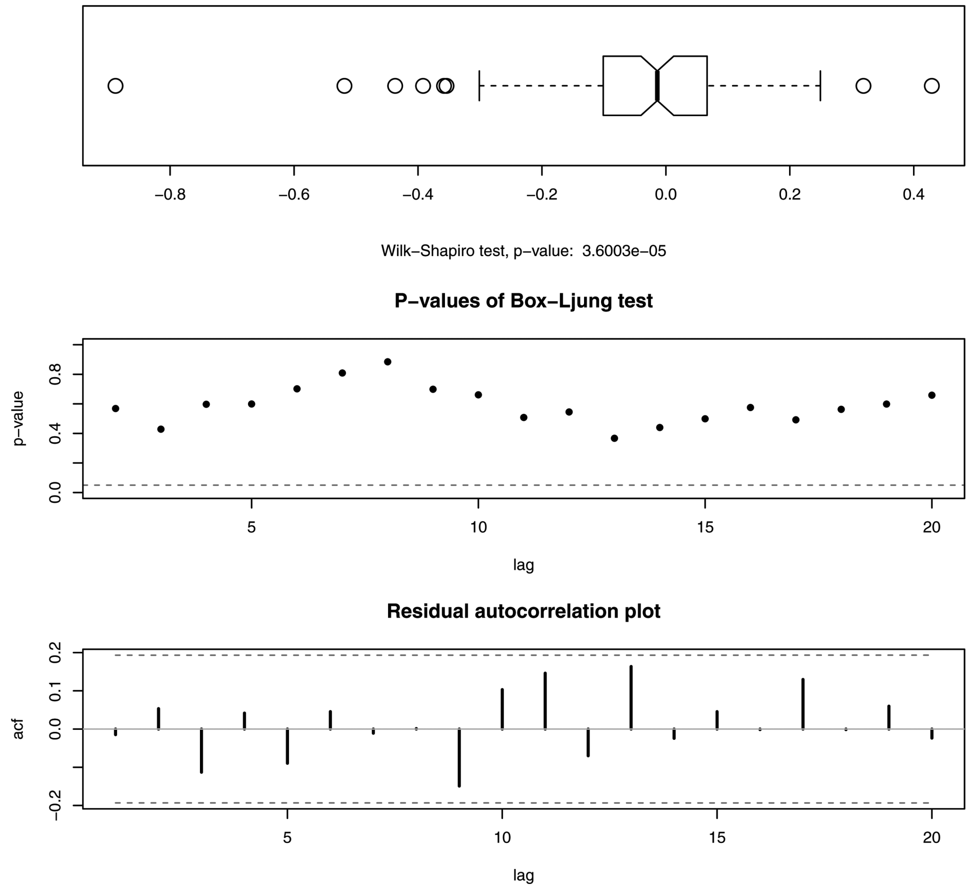

Figure 3. Diagnostic plots for money velocity time series.

6.1. *Velocity of Money*

The time series plot for the velocity of money in the United States (1869–1970) is shown in Fig. 2. From the plot, we see that the series has historically exhibited a strong stochastic trends characteristic of random walk behavior. No doubt with modern emphasis on fiscal policies to control inflation the series has stabilized. But just for a numerical illustration of the difference in the unit root tests, we will compare the maximum likelihood and least squares or DF tests. The first step in the analysis is the check that the fitted model is adequate and that no additional lags are required. Fig. 3 shows the diagnostic checks for this data. The residuals appear nonnormal but in view of the simulation results this is not a concern. Most importantly no evidence of residual autocorrelation is found in the fitted AR(1) model. Applying the unit root tests, the pivotal test statistics for the MLE and DF tests were, respectively, -0.26 and -3.28. The MLE test is not even close to being significant at the 10% level while the DF test has a p-value between 5% and 1%. The MLE unit root test gives a result that appears to be more in line with the overall impression of strong stochastic trends exhibited in Fig. 2. Even though the length of the series was 102, there is a considerable difference in the conclusion between the two methods.

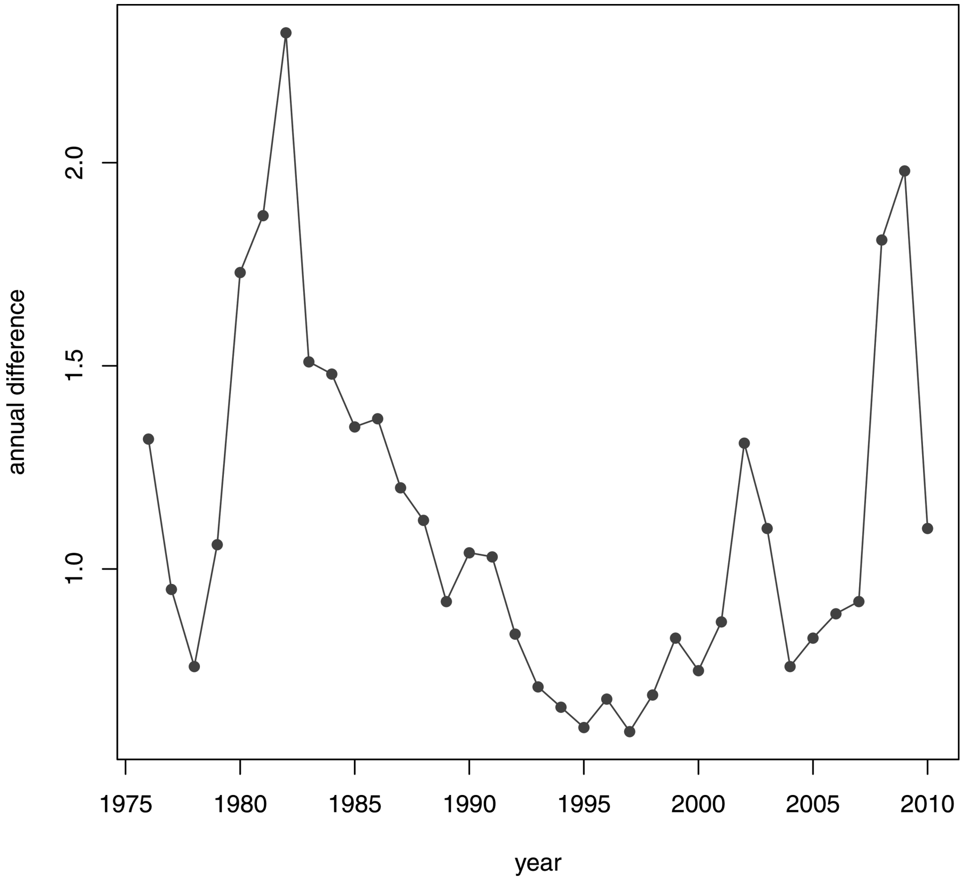

Figure 4. Time series plot of difference in bond yields.

6.2. *Bond Yield Differences*

The annual difference in Moody's BAA and AAA corporate bond yields from 1976 to 2010 is shown in Fig. 4. From the diagnostic check plots, we conclude that there is no significant autocorrelation in the residuals and so the AR(1) may be fit. The DF test is not significant at 10% whereas the MLE test does reject at the 10% level. This is not surprising in view of the empirical power computations.

7. Summary

In this article, we presented a new derivation of the asymptotic distribution for the MLE unit root test utilizing computer algebra to obtain an explicit expression for the MLE and a Taylor series linearization for the test statistic. This technique is no doubt applicable in other situations where the manual derivation is difficult.

An efficient computational method based on the response surface curves has been implemented to obtain critical values of the MLE test statistics. An empirical power study has demonstrated that not only does the MLE procedure outperform the LSE in the Gaussian case but also for fat-tailed distributions, infinite variance distributions, and for weak dependence as exhibited in a GARCH(1, 1) process. The R package `mleur` based on the developments in this article is available on CRAN.

Two illustrative applications of the test demonstrate that unit root testing also requires diagnostic checking. It is important for proper applications that there should be no residual autocorrelation present in the fitted AR(1) model.

Appendix

First, we state the Donsker's theorem (Billingsley, 1999). Let $\{a_t\}$ be a sequence of i.i.d. random variables with mean 0 and finite variance $\sigma^2 > 0$, $z_t = \sum_{j=1}^{t} a_j$, and $z_0 = 0$. Then

$$\left\{ \frac{z_{[nt]}}{\sqrt{n}\sigma},\ 0 \le t \le 1 \right\} \xrightarrow{D} \{W(t),\ 0 \le t \le 1\}$$

in the Skorokhod space $D[0,\ 1]$ with J_1 topology, where $[x]$ denotes the integer part of x. One of the important applications of the Donker's theorem is the following continuous mapping theorem. If $f(\cdot)$ is a continuous functional on $[0,\ 1]$, then

$$f\left(\frac{z_{[nt]}}{\sqrt{n}\sigma} \right) \xrightarrow{D} f(W(t)).$$

Proof of Lemma 3.1. We have

$$\frac{\bar{z}_n}{\sqrt{n}} = \int_0^1 \frac{z_{[nt]}}{\sqrt{n}}\, dt + \frac{z_n}{n\sqrt{n}}.$$

By the Donsker's theorem

$$\frac{\bar{z}_n}{\sqrt{n}} \xrightarrow{D} \sigma \int_0^1 W(t)\, dt.$$

Similarly,

$$\frac{1}{n^2} \sum_{t=1}^{n-1} z_t^2 = \int_0^1 \left(\frac{z_{[nt]}}{\sqrt{n}} \right)^2 dt \xrightarrow{D} \sigma^2 \int_0^1 W^2(t)\, dt.$$

It can be shown that

$$c = \sum_{t=2}^{n-1} (z_t - \bar{z}_n)^2 = \sum_{t=1}^{n-1} z_t^2 - z_1^2 - (n+2)\bar{z}_n^2 + 2\bar{z}_n(z_1 + z_n). \tag{A.1}$$

By some simple algebra steps,

$$\begin{aligned} b - c &= \sum_{t=1}^{n-1} (z_{t+1} - z_t)(z_t - \bar{z}_n) + (z_1 - \bar{z}_n)^2 \\ &= \frac{1}{2} z_n^2 - \frac{1}{2} \sum_{t=1}^{n} a_t^2 - \bar{z}_n \sum_{t=2}^{n} a_t + (z_1 - \bar{z}_n)^2, \end{aligned}$$

that is,

$$\frac{b-c}{n} = \frac{1}{2}\left(\frac{z_n}{\sqrt{n}}\right)^2 - \frac{1}{2n}\sum_{t=1}^{n} a_t^2 - \frac{\bar{z}_n}{\sqrt{n}}\frac{z_n}{\sqrt{n}} + \frac{\bar{z}_n}{\sqrt{n}}\frac{a_1}{\sqrt{n}} + \left(\frac{z_1}{\sqrt{n}} - \frac{\bar{z}_n}{\sqrt{n}}\right)^2. \tag{A.2}$$

Moreover,

$$\frac{a-c}{n} = \left(\frac{z_1}{\sqrt{n}} - \frac{\bar{z}_n}{\sqrt{n}}\right)^2 + \left(\frac{z_n}{\sqrt{n}} - \frac{\bar{z}_n}{\sqrt{n}}\right)^2. \tag{A.3}$$

Since

$$\frac{b-c}{n} = n(G-1)\left(\frac{c}{n^2}\right)$$

and

$$\frac{a-c}{n} = n(H-1)\left(\frac{c}{n^2}\right),$$

where G and H are defined in Eq. (19). Eqs. (A.1), (A.2), and (A.3) imply that

$$\left(\frac{c}{\sigma^2 n^2}, n(H-1), n(G-1)\right) = f\left(\frac{z_{[nt]}}{\sqrt{n}\sigma}\right) + o_p(1),$$

where f is a functional. Hence, by the continuous mapping theorem described above and Slutsky's theorem,

$$\left(\frac{c}{\sigma^2 n^2}, n(H-1), n(G-1)\right) \xrightarrow{D} f(W(t)).$$

Applying the Cramer–Wold device, we find the marginal distributions as

$$\frac{c}{\sigma^2 n^2} \xrightarrow{D} \int_0^1 W^2(t)\,dt - \left(\int_0^1 W(t)\,dt\right)^2 = \mathfrak{A}_\mu,$$

$$n(G-1) \xrightarrow{D} \mathfrak{A}_\mu^{-1}\left(\frac{1}{2}\left(W^2(1)-1\right) - W(1)\int_0^1 W(t)\,dt + \left(\int_0^1 W(t)\,dt\right)^2\right) = \mathfrak{C}_\mu,$$

and

$$n(H-1) \xrightarrow{D} \mathfrak{A}_\mu^{-1}\left(\left(\int_0^1 W(t)\,dt\right)^2 + \left(W(1) - \int_0^1 W(t)\,dt\right)^2\right) = \mathfrak{B}_\mu.$$

Acknowledgements

The authors would like to thank the Editor for his encouragement and also two referees for their helpful suggestions and comments. The authors' research was supported by Natural Sciences and Engineering Research Council of Canada (NSERC) and SHARCNET at the University of Western Ontario.

References

Andrews, D. F., Stafford, J. E. (2000). *Symbolic Computation for Statistical Inference*. Oxford: Oxford University Press.

Billingsley, P. (1999). *Convergence of Probability Measures*. New York: Wiley.

Brockwell, P. J., Davis, R. A. (1987). *Time Series: Theory and Methods*. New York: Springer-Verlag.

Dent, W., Min, A. S. (1978). A Monte Carlo study of autoregressive integrated moving average processes. *Journal of Econometrics* 7:23–55.

Dickey, A. D., Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association* 74:427–431.

Enders, W. (2010). *Applied Econometric Time Series*. 3rd ed. New York: Wiley.

Fuller, W. A. (1996). *Introduction to Statistical Time Series*. New York: Wiley.

Gonzalez-Farias, G. M., Dickey, D. A. (1999). Unit root test: An unconditional maximum likelihood approach. *Boletin de la Sociedad Matematica Mexicana* 5:199–221.

Hasza, D. P. (1980). A note on the maximum likelihood estimation for the first-order autoregressive processes. *Communications in Statistics: Theory and Methods* 13:1411–1415.

Iacus, S. M. (2008). *Simulation and Inference for Stochastic Differential Equations: With R Examples*. New York: Springer Science+Business Media, LLC.

MacKinnon, J. G. (2000). Computing Numerical Distribution Functions in Econometrics. In: Pollard, A., Mewhort, D., and Weaver, D. (eds.). *High Performance Computing Systems*. Amsterdam: Luwer. Ch 45, pp. 455–470.

McLeod, A. I., Yu, H., Zhang, Y. (2011). *mleur: Maximum likelihood unit root test*. R package version 1.0-4. Available at http://CRAN.R-project.org/package=mleur.

McLeod, A. I., Zhang, Y. (2008). Faster ARMA maximum likelihood estimation. *Computational Statistics and Data Analysis* 52(4):2166–2176.

Minozzo, M., Azzalini, A. (1993). On the unimodality of the exact likelihood function for normal ar(2) series. *Journal of Time Series Analysis* 14:497–510.

Pantula, S. G., Gonzalez-Farias, G., Fuller, W. A. (1994). A comparison of unit-root test criteria. *Journal of Business & Economic Statistics* 12(4):449–459.

Patterson, K. (2010). *A Primer for Unit Root Testing*. London: Palgrave.

Pfaff, B. (2006). *Analysis of Integrated and Cointegrated Time Series with R*. New York: Springer.

Pfaff, B. (2010). *urca: Unit root and cointegration tests for time series data*. R package version 1.2-5. Available at http://cran.r-project.org/web/packages/urca/index.html.

Shin, D. W., Fuller, W. (1998). Unit root tests based on unconditional maximum likelihood estimation for the autoregressive moving average. *Journal of Time Series Analysis* 19(5):591–599.

Smith, B., Field, C. (2001). Symbolic cumulant calculations for frequency domain time series. *Statistics and Computing* 11:75–82.

White, J. S. (1961). Asymptotic expansions for the mean and variance of the serial correlation coefficient. *Biometrika* 48:85–94.

Wolfram, S. (1999). *The Mathematica Book*. Cambridge: Cambridge University Press.

Zhang, Y., McLeod, A. I. (2006). Computer algebra derivation of the bias of linear estimators of autoregressive models. *Journal of Time Series Analysis* 27:157–165.