THE INSTITUTE OF STATISTICS

UNIVERSITY OF NORTH CAROLINA SYSTEM

AN UNCONDITIONAL MAXIMUM LIKELIHOOD TEST FOR A UNIT ROOT

By

Graciela Gonzalez-Farias and David A. Dickey

Institute of Statistics

Mimeo Series #~~2245~~ 2245

January, 1992

NORTH CAROLINA STATE UNIVERSITY

Raleigh, North Carolina

AN UNCONDITIONAL MAXIMUM LIKELIHOOD TEST FOR A UNIT ROOT

Graciela Gonzalez-Farias and David A. Dickey

North Carolina State University Raleigh, North Carolina

Key Words: ***Time Series, Nonstationary***

Abstract

We investigate a test for unit roots in autoregressive time series based on maximization of the unconditional likelihood. This is the likelihood function appropriate for stationary time series. While this function is the true likelihood only under the stationary alternative, it can nevertheless be maximized for any data including data from a unit root process. It thus gives a way to test for unit roots, provided percentiles can be calculated. For models with estimated means, the power of the new test is better than that of some currently popular tests.

1. INTRODUCTION

Time series modeling often involves the selection and fitting of an ARIMA (autoregressive integrated moving average) model. The order of integration is defined as the degree of differencing required to make the series stationary where stationarity implies constant mean and variance over time and a covariance which depends only on the time separating two observations. The fitting of a series traditionally involves differencing the data if necessary, until they appear stationary then fitting autoregressive and moving average parameters to the, possibly differenced, data. We investigate statistical ways to check whether differencing is necessary.

Appropriate differencing renders a series stationary and thus makes the resulting estimation theory easier to work out. The results tend to be classical in nature, for example normal limit distributions of estimators. Classic methods of estimation, such as least squares and maximum likelihood are not necessarily poor estimation methods for the parameters of nonstationary series, however the distributions are not standard even in the limit. If percentiles of the distributions can be obtained, then these can be used for hypothesis testing.

For ARIMA models, stationarity can be characterized by a condition on the roots of a polynomial involving the autoregressive coefficients, called the characteristic polynomial. If all the roots are larger than 1 in magnitude, the series is stationary. Therefore we can base a test for stationarity on the coefficients or roots of the characteristic polynomial. These in turn must be estimated in some way. Such tests are often called unit root tests (unit roots being the null hypothesis) but could arguably also be called tests for stationarity when that is taken as the alternative. The most important motivation for developing a test for stationarity is that certain economic hypotheses are mathematically equivalent to stating that there are unit roots in the corresponding data series.

Tests based on least squares estimation are reasonably well known. Least squares maximizes the likelihood conditional on the initial observation(s). This is in contrast to the unconditional likelihood function for a stationary model that is used by computer programs written to do maximum likelihood analysis. This unconditional likelihood function can be maximized regardless of the true nature of the data and thus might be thought of as an objective function rather than a likelihood. Under the alternative hypothesis of stationarity, such an estimator should do well. However, it is not clear how well it would perform under the null hypothesis of a unit root nor is it clear what the distribution of the unconditional maximum likelihood estimator would be in this case.

We show that the distribution in question is nonstandard and differs, even in the limit, from that of the least squares estimator. Further, this new estimator has superior power in some instances of practical interest.

2. TEST CRITERIA, MEAN KNOWN ($\mu = 0$)

While the case of a known mean is of little practical value, the algebra is simple and the ideas of the proof carry over to the more practical cases. In what follows, we outline the main steps of the development of the estimator. The interested reader can refer to Gonzalez-Farias (1992) for technical details.

Hasza (1980) and Anderson (1971, pg. 354) study the AR(1) case with known mean using the stationary likelihood function. They show that the maximum likelihood estimator of ρ in the model $Y_t = \rho Y_{t-1} + e_t$, $e_t \sim N(0,\sigma^2)$ can be written as the solution $\hat{\rho}$ to the cubic equation $g(\rho) = 0$ where

$$g(\rho) = \left(\frac{n-1}{n} \sum_{t=2}^{n-1} Y_t^2\right)\rho^3 - \left(\frac{n-2}{n} \sum_{t=2}^{n-2} Y_t Y_{t-1}\right)\rho^2 - \left(\sum_{t=2}^{n-1} Y_t^2 + \frac{1}{n} \sum_{t=1}^{n} Y_t^2\right)\rho + \left(\sum_{t=2}^{n} Y_t Y_{t-1}\right)$$

Using the formula for the roots of a cubic equation, a closed form solution can be given (see Hasza, 1980).

Although this gives a neat solution in the first order case, we want to look at higher order processes. Let $X = n(\rho - 1)$ and divide $g(\rho)$ by $\frac{n-1}{n^3} \sum_{t=2}^{n-1} Y_t^2$. The result is

$$q_n(X) = \frac{1}{n} X^3 + \left(3 - \frac{n-2}{n-1}\hat{\rho}\right) X^2 - 2\left(n(\hat{\rho} - 1) - n(\bar{\rho} - 1)/(n-1)\right) X + 2\left(\frac{n}{n-1}\right) n(\bar{\rho} - 1)$$

where $\hat{\rho} = \left(\sum_{t=2}^{n-1} Y_t^2\right)^{-1}\left(\sum_{t=2}^{n} Y_t Y_{t-1}\right)$ which is almost exactly the ordinary least squares estimator and

$\bar{\rho} = \left[\frac{1}{2}\left(Y_1^2 + Y_n^2\right) + \sum_{t=2}^{n-1} Y_t^2\right]^{-1} \left[\sum_{t=2}^{n} Y_t Y_{t-1}\right]$ which is the symmetric estimator as defined in Dickey, Hasza, and Fuller (1984). Both $n(\hat{\rho} - 1)$ and $n(\bar{\rho} - 1)$ are $0_p(1)$. Since the leading coefficient in $q_n(X)$ is $1/n$, we see that the probability limit of the polynomial $q_n(X)$ over any closed X interval is the same as the probability limit of the quadratic

$$Q(X) = 2X^2 - 2\left[n(\hat{\rho} - 1)\right] X + 2\left[n(\bar{\rho} - 1)\right]. \tag{2.1}$$

Because of the $0_p(1)$ order of the random coefficients and the fact that $n(\bar{\rho} - 1)$ is strictly negative, we can find a closed X interval such that with arbitrarily high probability for a given $\delta > 0$, the roots of $q_n(X)$ are real, the largest two are within the closed X interval, and these two are within δ of the roots of $Q(X)$. That is, the negative root of this quadratic polynomial will have the same limit distribution as the maximum likelihood estimator.

Using the notation of Dickey and Fuller (1979) we define

$$(\Gamma,\zeta) = \lim_{n\to\infty} (n^2 \sum_{t=2}^{n} W_{t-1}^2,\ n^{-1} \sum_{t=2}^{n} W_{t-1} Z_t)$$

where W_t is the random walk defined by $W_t = W_{t-1} + Z_t$ and $Z_t \sim N(0, 1)$. Alternatively, Chan and Wei define (Γ, ζ) as $\left(\int_0^1 W^2(t)dt, \frac{1}{2}(W^2(1) - 1)\right)$ where W(t) is Brownian Motion. The limit of expression (2.1) in terms of (Γ, ζ) becomes

$$2\, X^2 - 2(\zeta/\Gamma)\, X + 2\left(-1/(2\Gamma)\right) \tag{2.2}$$

Techniques described in Dickey (1976) can be used to simulate the random vector (Γ, ζ) from which the roots of (2.2) can be found, and percentiles tabulated. Table 2.1 contains percentiles of the unconditional maximum likelihood estimator's distribution for finite samples and the limit. These are labelled $n(\widetilde{\rho}_{ml} - 1)$. Also listed for comparison are the least squares estimate $n(\widehat{\rho}_{ols} - 1)$ as in Fuller (1976) and the symmetric estimator $n(\overline{\rho} - 1)$ as in Dickey, Hasza, and Fuller (1984). Gonzalez-Farias (1992) also gives tables of the corresponding studentized statistics.

Table 2.1. Empirical cumulative distribution of the estimators of ρ, ρ=1.

Sample Size	Probability of a Smaller Value.							
	$n(\tilde{\rho}_{ml} - 1)$							
n	0.01	0.025	0.05	0.10	0.90	0.95	0.975	0.99
25	−12.19	−9.73	−7.75	−5.80	−0.38	−0.27	−0.21	−0.16
50	−13.18	−10.41	−8.29	−6.20	−0.39	−0.27	−0.21	−0.16
100	−13.81	−10.83	−8.59	−6.42	−0.39	−0.28	−0.21	−0.16
250	−14.29	−11.14	−8.81	−6.57	−0.39	−0.28	−0.21	−0.16
500	−14.50	−11.26	−8.89	−6.62	−0.39	−0.28	−0.21	−0.16
∞	−14.86	−11.46	−9.00	−6.69	−0.39	−0.28	−0.21	−0.16
	$n(\hat{\rho}_{ols} - 1)$							
25	−11.90	−9.30	−7.30	−5.30	1.01	1.40	1.79	2.28
50	−12.90	−9.90	−7.70	−5.50	0.97	1.35	1.70	2.16
100	−13.30	−10.20	−7.90	−5.60	0.95	1.31	1.65	2.09
250	−13.60	−10.30	−8.00	−5.70	0.93	1.28	1.62	2.04
500	−13.70	−10.40	−8.00	−5.70	0.93	1.28	1.61	2.04
∞	−13.80	−10.50	−8.10	−5.70	0.93	1.28	1.60	2.03
	$n(\bar{\rho} - 1)$							
25	−12.90	−9.67	−7.75	−5.83	−0.40	−0.29	−0.23	−0.18
50	−13.27	−10.41	−8.26	−6.16	−0.41	−0.30	−0.24	−0.18
100	−13.87	−10.82	−8.55	−6.34	−0.41	−0.30	−0.23	−0.18
250	−14.25	−11.08	−8.73	−6.45	−0.42	−0.30	−0.23	−0.18
500	−14.38	−11.17	−8.79	−6.49	−0.42	−0.30	−0.23	−0.18
∞	−14.51	−11.26	−8.86	−6.53	−0.42	−0.30	−0.23	−0.18

The limit distribution does not change if additional lags are included in the model. This will also be true in the case where the mean is estimated. To illustrate this result, consider a second order, AR(2), model. The AR(2) model with mean 0 can be written as

$$Y_t = (m_1 + m_2)\, Y_{t-1} - m_1 m_2\, Y_{t-2} + e_t$$

and the logarithm of the stationary likelihood for e_t normal is

$$\ln(\mathcal{L}) = -\tfrac{n}{2}\ell n(2\Pi) - \tfrac{n}{2}\ell n(\sigma^2) + \tfrac{1}{2}\ell n(1-m_1) + \tfrac{1}{2}\ell n\Big[(1+m_1)(1-m_2^2)(1-m_1 m_2)^2\Big] -$$
$$\frac{1}{2}\left[\frac{Y_1^2}{\sigma^2} + \frac{Y_2^2}{\sigma^2}\right](1 - m_1^2\, m_2^2) + \frac{Y_1 Y_2}{\sigma^2}\,(m_1 + m_2)\,(1 - m_1 m_2) - \frac{1}{2}\,(SSQ/\sigma^2)$$

where the sum of squares SSQ is

$$SSQ = \sum_{t=3}^{n}\Big(Y_t - (m_1 + m_2)Y_{t-1} + m_1 m_2 Y_{t-2}\Big)^2$$

Suppose the true value of (m_1, m_2) is $(1, \alpha)$ with $|\alpha| < 1$. Now let $X = n(m_1 - 1)$ and $S = \sqrt{n}(m_2 - \alpha)$ and consider the function $\ell n(\mathcal{L})$ for (X, S) in an arbitrary closed rectangular region R. Let $\tilde{\sigma}^2$ denote the maximum likelihood estimator of σ^2 for any given (X, S) and note that for (X, S) in R we have

$$\tilde{\sigma}^2 = SSQ/n + O_p(1/n) = \sigma^2 + O_p(1/\sqrt{n}\,)$$

since, in R, $m_1 = 1 + O(1/n)$ and $m_2 = \alpha + O(1/\sqrt{n}\,)$. Substituting $\tilde{\sigma}^2$ into $\ell n(\mathcal{L})$ we have the concentrated likelihood

$$\ell n(\mathcal{L}) = -(n/2)\ell n(\tilde{\sigma}^2) + (1/2)\ell n(-X/n) - (n/2)\ell n(2\Pi)$$
$$+ (1/2)\ell n\Big[(1+m_1)(1-m_2^2)(1-m_1 m_2)^2\Big] - (n/2)$$

Only the first two terms affect the limit distribution. In fact the $X{=}n(m_1 - 1)$ and $S{=}\sqrt{n}(m_2 - \alpha)$ which maximize $\ell n(\mathcal{L})$ are asymptotically the same as the (X, S) which maximize

$$F_n(X, S) = \Big[-(n/2)\ell n\Big((SSQ)/n\Big)\Big] + (1/2)\ell n(-X) + C \qquad (2.3)$$

where C is constant with respect to X and S. See Gonzalez-Farias (1992 appendix B). For (X, S) in R we have

$$SSQ = \sum_{t=3}^{n}\Big[e_t - X\,(Y_{t-1} - \alpha Y_{t-2})/n - S(Y_{t-1} - Y_{t-2})/\sqrt{n} + XS(Y_{t-2}/n^{3/2})\Big]^2$$

from which we see that

$$SSQ/n \rightarrow \sigma^2.$$

Notice that XS in SSQ is multiplied by $n^{-\frac{3}{2}} Y_{t-2}$, a term whose sum of squares converges to 0. Furthermore $\Sigma\left[(Y_{t-1} - \alpha Y_{t-2})/n \; (Y_{t-1} - Y_{t-2})/\sqrt{n}\right]$ converges to 0. In the limit, then, the log likelihood is the sum of two functions, one involving only S and one involving only X. Specifically, $SSQ/\sigma^2 - \sum_{t=3}^{n} e_t^2/\sigma^2$ is a polynomial in (X, S) and converges uniformly in R to

$$\left[-2X\zeta + X^2\Gamma\right] + \left[S^2/(1-\alpha^2) - 2\,SV\right] \tag{2.4}$$

where $V \sim N\left(0,\ (1-\alpha^2)^{-1}\right)$. The derivatives of SSQ/σ^2 also converge to the derivatives of (2.4). This result follows from the well known rates of convergence of sums of squares and cross products for stationary and unit root processes. Taking the derivative, with respect to X, of $F_n(X, S)$, we see that the limit maximum likelihood estimator satisfies

$$\zeta - X\,\Gamma + \frac{1}{2X} = 0$$

or

$$2X^2 - 2(\zeta/\Gamma)X + 2\left(\frac{-1}{2\Gamma}\right) = 0$$

which is the same as (2.2). Note also that taking the derivative of $F_n(X, S)$ with respect to S gives the same limit normal distribution for $\sqrt{n}(\widehat{m}_2 - \alpha)$ that would be obtained from applying least squares or maximum likelihood to the model

$$Y_t - Y_{t-1} = \alpha_2(Y_{t-1} - Y_{t-2}) + e_t$$

Since the region R can be any closed rectangular region with (X, S)=(0, 0) an interior point, the orders of sums of squares and cross products in (2.3) imply that, given any $\delta > 0$, the (X, S) which maximizes F_n is eventually within this region with probability at least $(1-\delta)$. Over R, the function $\ell n(\mathcal{L}) - F_n(X, S)$ converges uniformly to 0, so the normalized maximum likelihood estimates of m_1 and m_2 converge to the maximizing (X, S).

We have outlined the case of the known mean since the algebra is relatively easy to follow and involves only a few terms. In the next section, we look at the more practical case in which the mean μ is estimated. The main ideas are similar to the case just presented, but the algebra is more tedious and, for example, the limit estimator will now be the solution of a fifth degree polynomial (3.10).

3. TEST CRITERIA, MEAN ESTIMATED

Consider the model for $Y_1, \cdots, Y_n$

$$Y_t = \mu(1-\rho)+\rho Y_{t-1}+e_t,\ t=2,\ 3,\ \cdots;|\rho| < 1 \tag{3.1}$$

where e_t is a sequence of iid $N(0,\ \sigma^2)$ random variables. We will study the model (3.1) under two scenarios, namely, Y_1 being fixed and

$$Y_1 \sim N\left(\mu,\ \sigma^2 \big/ (1-\rho^2)\right).$$

3.1. Case 1: Y_1 fixed.

When Y_1 is considered fixed, the estimator of ρ is obtained by maximizing the log likelihood function conditioned on Y_1, namely

$$\mathcal{L}(\mu,\ \rho,\ \sigma^2) = -\frac{n}{2}\log(2\Pi) - \frac{n}{2}\log\sigma^2 -$$

$$\frac{1}{2\sigma^2} \quad \sum_{t=2}^{n} \{(Y_t - \mu) - \rho(Y_{t-1} - \mu)\}^2 . \tag{3.2}$$

This is commonly called the conditional maximum likelihood estimator. If μ is estimated, the conditional maximum likelihood estimator is also asymptotically equivalent to the ordinary least squares estimator $\hat{\rho}_{\mu,\text{ols}}$ of ρ, obtained by regressing Y_t on 1 and Y_{t-1}. Asymptotic properties of $\hat{\rho}_{\mu,\text{ols}}$ when $\rho = 1$ have been very well established in the literature, see for example, Dickey and Fuller (1979). The estimator's distribution does not depend on the value of μ in (3.1).

Dickey and Fuller (1979) show that the pivotal statistic

$$\hat{t}_{\mu,\text{ols}} = \frac{\hat{\rho}_{\text{ols}}\text{-}1}{\sqrt{s^2\left[\sum_{2}^{n}(Y_{t-1} - \overline{Y}_{(-1)})^2\right]^{-1}}} \xrightarrow{\mathfrak{D}} \hat{\tau}_\mu$$

where $S^2 = \frac{1}{n-2}\sum_{t\text{-}2}^{n}\hat{e}_t^2$ and $\overline{Y}_{(-1)} = \frac{1}{n-1}\sum_{2}^{n} Y_{t-1}$ and $\hat{\tau}_\mu$ can be expressed as a function of standard normal variates. Again this statistic does not depend on the true value of μ in (3.1). In fact, adding a constant c to every observation has no effect on either $\hat{\rho}_{\mu,\text{ols}}$ or $\hat{\tau}_{\mu}$. Table 8.5.2 of Fuller (1976) gives the percentiles of the $\hat{\tau}_{\mu}$ distribution.

3.2. Case 2: $Y_1 \sim N\left(\mu, \sigma^2 \big/ (1-\rho^2)\right)$.

Consider n observations $Y_1, \cdots, Y_n$ from the model (2.1). Then, the log likelihood function is given by

$$\mathcal{L}(\mu, \rho, \sigma^2) = \mathcal{L}^*(\mu, \rho, \sigma^2) + \tfrac{1}{2}\log(1+\rho) \tag{3.3}$$

where $\mathcal{L}^*(\mu, \rho, \sigma^2) = -\frac{n}{2}\log 2\Pi - \frac{n}{2}\log\sigma^2 + \frac{1}{2}\log(1-\rho)$

$$-\frac{1}{2\sigma^2}\left[(Y_1-\mu)^2(1-\rho^2) + \sum_{t=2}^{n}\{(Y_t-\mu) - \rho(Y_{t-1}-\mu)\}^2\right].$$

For any given ρ and σ^2, both $\mathcal{L}$ and $\mathcal{L}^*$ are maximized at

$$\hat{\mu}(\rho) = \frac{Y_1 + Y_n - (\rho-1)\sum_{2}^{n-1} Y_t}{2-(n-2)(\rho-1)} = \overline{Y}_1 + \frac{Y_1 + Y_n - 2\overline{Y}_1}{2-(n-2)(\rho-1)} \tag{3.4}$$

where $\overline{Y}_1 = \frac{1}{n-2}\sum_{t=2}^{n} Y_t$.

Notice that $\hat{\mu}(\rho)$ is a weighted average of the data so that any statistic defined as a function of $Y_t - \hat{\mu}(\rho)$ will be unchanged by the addition of a constant c to every observation. This shows that such a statistic is independent of the value of μ in (3.1).

Let $\hat{\mu}^*_m$, $\hat{\rho}^*_m$ and $\sigma^{*\,2}_m$ be the values of μ, ρ and σ^2 that maximize $\mathcal{L}^*$. It is shown in Gonzalez-Farias (1992) that the asymptotic distribution of $(\hat{\mu}^*_m, \hat{\rho}^*_m, \hat{\sigma}^{*2}_m)$ is the same as $(\tilde{\mu}_{ml}, \tilde{\rho}_{m,ml}, \tilde{\sigma}^2_{m,ml})$ the maximum likelihood estimator of (μ, ρ, σ^2), so it is sufficient to use $\mathcal{L}^*$ as the objective to maximize.

THEOREM. Suppose $Y_t = Y_{t-1} + e_t$ where e_t is a sequence of iid $(0, \sigma_0^2)$ random variables with $E(e_t^4) = \eta\sigma_0^4$. Without loss of generality, assume $\sigma_0^2 = 1$. Let $\hat{\mu}^*_m$, $\hat{\rho}^*_m$ and $\hat{\sigma}^{*2}_m$ be the values of μ, ρ, σ^2

that maximize $\mathcal{L}^*(\mu, \rho, \sigma^2)$. Then

$$\begin{pmatrix} n^{-\frac{1}{2}}\hat{\mu}^*_m \\ n(\hat{\rho}^*_m - 1) \end{pmatrix} \xrightarrow{\mathfrak{D}} \begin{pmatrix} H^* \\ A^* \end{pmatrix} \tag{3.5}$$

and

$$\hat{\sigma}^{*2}_m = \frac{1}{n}\sum_{t=2}^{n} e_t^2 + 0_p(\tfrac{1}{n})$$

where $H^* = H + \frac{T-2H}{2-A^*}$ and A^* is the unique negative solution to

$$a_4X^4 + a_3X^3 + a_2X^2 + a_1X + a_0 = 0 \tag{3.6}$$

The a_i's are functionals of a standard Brownian motion W(t). Following Dickey and Fuller (1979), and Chan and Wei (1988), $(\Gamma, H, T) = \left(\int_0^1 W^2(t)db, \int_0^1 W(t)dt, W(1)\right)$.

Let $\Gamma_\mu = \Gamma - H^2$, $\zeta = \frac{1}{2}(T^2 - 1)$, and

$\zeta_\mu = \zeta - TH$. Then

$a_4 = 2\Gamma_\mu$, $a_3 = -2(\zeta_\mu + H^2) - 8\Gamma_\mu$,

$a_2 = 8\Gamma_\mu + 8(\zeta_\mu + H^2) - 1$,

$a_1 = -8(\zeta_\mu + H^2) + 2(T-2H)^2 + 4$, $a_0 = -4$.

Proof: We present here a sketch of a proof. A detailed proof appears in Gonzalez-Farias (1992).

Taking partial derivatives of $\mathcal{L}^*$ with respect to μ, and σ^2 and setting them equal to zero we get

$$\hat{\mu}^*_m = \hat{\mu}(\hat{\rho}^*_m) \tag{3.7}$$

$$\hat{\sigma}^{*2}_m = \frac{1}{n}\left\{(Y_1 - \overline{\mu}^*_m)^2(1 - \hat{\rho}^{*2}_m) + \sum_{t=2}^{n}\left[(Y_t - \hat{\mu}^*_m) - \hat{\rho}^*_m(Y_{t-1} - \hat{\mu}^*_m)\right]^2\right\} \tag{3.8}$$

and $\hat{\beta}^*_m$ is a solution to

$$a^*_{4m}\hat{\beta}^{*4}_m + a^*_{3m}\hat{\beta}^{*3}_m + a^*_{2m}\hat{\beta}^{*2}_m + a^*_{1m}\widehat{\beta}^*_m + a^*_{0m} = 0 \tag{3.9}$$

where $\hat{\beta}^*_m = n(\hat{\rho}^*_m - 1)$,

$$a^*_{4,n} = 2\Gamma_{n,\mu} + 0_p(\tfrac{1}{n}),$$

$$a^*_{3,n} = -8\Gamma_{n,\mu} - 2\zeta_{n,\mu} - (2/n)\overline{Y}_1^2 + 0_p\Big(\frac{1}{\sqrt{n}}\Big),$$

$$a^*_{2,n} = 8\Gamma_{n,\mu} + 8\zeta_{n,\mu} + \frac{8}{n}\overline{Y}_1^2 - 1 + 0_p\Big(\frac{1}{\sqrt{n}}\Big),$$

$$a^*_{1,n} = -8\zeta_{n,\mu} - \frac{8}{n}\overline{Y}_1^2 + \frac{2}{n}(Y_n - 2\overline{Y}_1)^2 + 4 + 0_p\Big(\frac{1}{\sqrt{n}}\Big),$$

$$a^*_{0,n} = -4 + 0_p\Big(\frac{1}{\sqrt{n}}\Big),$$

$$\Gamma_{n,\mu} = \frac{1}{n^2}\sum_{2}^{n-1}(Y_t - \overline{Y}_1)^2$$

and

$$\zeta_{n,\mu} = \frac{1}{n}\sum_{t=2}^{n-1}(Y_{t-1} - \overline{Y}_1)(Y_t - Y_{t-1}).$$

The coefficients $a^*_{i,n}$ in (3.9) converge jointly in distribution to the coefficients a_i in (3.6).

By Theorem 13.8 of Breiman (1968), we can change to a new probability space on which

$$\{a^*_{i,n}\,,\ i = 0,\ \cdots,\ 4\} \xrightarrow{P} \{a_i,\ i = 0,\ \cdots,\ 4\}.$$

The limit polynomial given in (3.6) can be factored giving

$$(X-2)^2[2X^2\Gamma_\mu - 2X(\zeta_\mu + H^2) - 1] + 2X(T - 2H)^2 = 0$$

and hence, we can show that it has a unique root $X = A^*$ in $(-\infty, 0)$.

Also, since $a^*_{i,n} = 0_p(1)$, $i=0, \cdots, 4$ and $a^*_{4,n}$ and a_4 are positive random variables, it is possible to show that with a very high probability the polynomial in (3.9) has a solution, $\hat{\beta}^*_m$ in the interval $(-n, 0)$.

Then, using the Implicit Function Theorem we get that with high probability, there exists a unique $\hat{\rho}^*_m$ that satisfies (3.9) and is a continuous function of $a^*_{i,n}, i=0,\cdots,4$ and hence $\hat{\beta}^*_m \xrightarrow{\mathfrak{D}} A^*$.

Note that, the partial derivative of $\mathfrak{L}(\mu, \rho, \sigma^2)$ is $\dfrac{\partial \mathfrak{L}}{\partial \rho} = \dfrac{1}{2(1+\rho)} + \dfrac{\partial \mathfrak{L}^*(\mu, \rho, \sigma^2))}{\partial \rho}$.

After some algebra we get that the maximum likelihood estimator of ρ, or equivalently $\tilde{\beta}^*_m = n(\tilde{\rho}^*_{m,ml} - 1)$ is a solution to a fifth degree polynomial,

$$\sum_{i=0}^{5} a_{i,n}\tilde{\beta}^i_m = 0 \qquad (3.10)$$

where $a_{5,n} \xrightarrow{P} 0$, $a_{i,n} = a^*_{i,n} + 0_p(\frac{1}{n})$, i=1, 2, 3, 4 and $a_{0,n} = a^*_{0,n}$.

Note that the polynomial in (3.10) behaves similarly to the fourth degree polynomial we obtained for $\mathcal{L}^*$ and hence the asymptotic distribution for the maximum likelihood estimator is the same as that of $(\hat{\mu}^*_m, \hat{\rho}^*_m, \hat{\sigma}^{*2}_m)$.

4. Remarks

1. The maximum likelihood estimators may also be obtained iteratively. For a given $\hat{\rho}_{i-1}$ and $\hat{\sigma}^2_{i-1}$, $\hat{\mu}_i$ may be obtained from (3.4). For a given $\hat{\mu}_i$, we then obtain $\hat{\rho}_i$ and $\hat{\sigma}^2_i$ by maximizing $\mathcal{L}(\hat{\mu}_i, \rho, \sigma^2)$. It can be shown that $\hat{\rho}_i$ is a solution to a cubic equation where the observations are centered at $\hat{\mu}_i$ and a closed form expression is given in Hasza (1980).

2. When we assume that μ=0 in the model (3.1) then the MLE of ρ that maximizes $\mathcal{L}(0, \rho, \sigma^2)$ is obtained by solving (2.2), namely

$$n(\hat{\rho}_{ml} - 1) \xrightarrow{\mathfrak{D}} A = \frac{1}{2}\left\{ \frac{\zeta}{\Gamma} - \left\{ \frac{\zeta^2}{\Gamma^2} + \frac{2}{\Gamma} \right\}^{1/2} \right\}.$$

An empirical study that compares the powers of $\hat{\rho}_{ols}$ and $\tilde{\rho}_{ml}$ indicates that there are essentially no differences in the power among these test criteria in the μ=0 case.

3. A Wald type pivotal t-statistic can be constructed for testing ρ=1,

$$\hat{t} = \frac{(\hat{\rho} - 1)}{\sqrt{\hat{V}(\hat{\rho})}}$$

where $\hat{V}(\hat{\rho})$ is determined from the negative matrix of the second derivatives of the corresponding objective function.

4. For the higher order processes, we consider the model

$$(Y_t - \mu) = \rho(Y_{t-1} - \mu) + Z_t \tag{3.11}$$

where Z_t is a stationary AR(p − 1) model, $1 > \rho \geq \max |m_i|$ and m_i, i = 2, ⋯, p are the roots of the characteristic equation of the Z_t process. Using methods similar to those in section 2, Gonzalez-Farias (1992) has shown that the limiting distribution of $n(\tilde{\rho}_{m,ml} - 1)$ where $\tilde{\rho}_{m,ml}$ is the MLE of ρ for the model (3.11), is the same as that of $n(\tilde{\rho}_{m,ml})$ derived for the AR(1) case.

Tables of percentiles for the different estimators mentioned above are given in Table 4.1 – 4.2. The top panel of each table is the maximizer of $\mathcal{L}(\mu, \rho, \sigma^2)$ and the second the maximizer of $\mathcal{L}^*(\mu, \rho, \sigma^2)$ as in (3.3). The last two panels are the least squares and symmetric estimators as in Table 2.1 but with estimated means.

Table 4.1 Empirical distribution of the normalized bias estimators.

Sample Size	Probability of a Smaller Value.							
				$n(\widehat{\rho}_{ml,\ m} - 1)$				
n	0.01	0.025	0.05	0.10	0.90	0.95	0.975	0.99
25	− 16.96	− 14.23	− 12.01	− 9.69	− 1.05	− 0.74	− 0.56	− 0.42
50	− 18.05	− 14.92	− 12.45	− 9.93	− 1.03	− 0.72	− 0.53	− 0.39
100	− 18.70	− 15.34	− 12.72	− 10.07	− 1.01	− 0.70	− 0.52	− 0.38
250	− 19.19	− 15.66	− 12.93	− 10.18	− 1.00	− 0.70	− 0.51	− 0.37
500	− 19.40	− 15.80	− 13.02	− 10.22	− 0.99	− 0.68	− 0.50	− 0.36
∞	− 19.72	− 16.01	− 13.16	− 10.27	− 0.95	− 0.66	− 0.50	− 0.35
				$n(\widehat{\rho}^*_m - 1)$				
25	− 17.31	− 14.50	− 12.26	− 9.90	− 0.99	− 0.71	− 0.55	− 0.42
50	− 18.33	− 15.12	− 12.61	− 10.04	− 0.99	− 0.70	− 0.55	− 0.42
100	− 18.91	− 15.49	− 12.82	− 10.13	− 0.99	− 0.70	− 0.55	− 0.42
250	− 19.31	− 15.74	− 12.97	− 10.19	− 0.99	− 0.70	− 0.55	− 0.42
500	− 19.46	− 15.84	− 13.04	− 10.21	− 0.99	− 0.70	− 0.55	0.42
∞	− 19.66	− 15.98	− 13.13	− 10.25	− 0.95	− 0.70	− 0.49	− 0.35
				$n(\widehat{\rho}_{\mu,\ ols} - 1)$				
25	− 17.2	− 14.6	− 12.5	− 10.2	− 0.76	0.01	0.65	1.40
50	− 18.9	− 15.7	− 13.3	− 10.7	− 0.81	− 0.07	0.53	1.22
100	− 19.8	− 16.3	− 13.7	− 11.0	− 0.83	− 0.10	0.47	1.14
250	− 20.3	− 16.6	− 14.0	− 11.2	− 0.84	− 0.12	0.43	1.09
500	− 20.05	− 16.8	− 14.0	− 11.2	− 0.84	− 0.13	0.42	1.06
∞	− 20.7	− 16.9	− 14.1	− 11.3	− 0.85	− 0.13	0.41	1.04
				$n(\widehat{\rho}_{\mu,\ sym} - 1)$				
25	− 17.9	− 14.62	− 12.49	− 10.17	− 1.52	− 1.17	− 0.95	− 0.77
50	− 18.64	− 15.54	− 13.09	− 10.52	− 1.48	− 1.13	− 0.90	− 0.72
100	− 19.39	− 16.00	− 13.39	− 10.70	− 1.46	− 1.10	− 0.88	− 0.69
250	− 19.85	− 16.29	− 13.56	− 10.80	− 1.45	− 1.09	− 0.87	− 0.68
500	− 20.01	− 16.38	− 13.62	− 10.83	− 1.44	− 1.09	− 0.87	− 0.68
∞	− 20.16	− 16.47	− 13.68	− 10.87	− 1.44	− 1.08	− 0.86	− 0.67

Table 4.2 Empirical distribution of pivotal statistics

Sample Size	Probability of a Smaller Value.							
				$\tilde{t}_{ml,m}$				
n	0.01	0.025	0.05	0.10	0.90	0.95	0.975	0.99
25	−3.49	−3.08	−2.76	−2.42	−0.90	−0.83	−0.79	−0.76
50	−3.31	−2.96	−2.68	−2.38	−0.91	−0.83	−0.79	−0.76
100	−3.24	−2.92	−2.66	−2.36	−0.91	−0.83	−0.79	−0.76
250	−3.21	−2.90	−2.65	−2.36	−0.91	−0.83	−0.79	−0.76
500	−3.20	−2.90	−2.64	−2.36	−0.91	−0.83	−0.79	−0.76
∞	−3.20	−2.90	−2.64	−2.36	−0.91	−0.83	−0.79	−0.76
				$\hat{t}^*_m$				
25	−3.53	−3.10	−2.78	−2.44	−0.88	−0.81	−0.78	−0.75
50	−3.35	−2.99	−2.70	−2.39	−0.90	−0.83	−0.79	−0.76
100	−3.36	−2.94	−2.68	−2.37	−0.90	−0.83	−0.80	−0.77
250	−3.21	−2.91	−2.65	−2.36	−0.91	−0.84	−0.80	−0.77
500	−3.21	−2.90	−2.64	−2.36	−0.91	−0.84	−0.81	−0.78
∞	−3.20	−2.90	−2.64	−2.35	−0.91	−0.84	−0.81	−0.78
				$\hat{\tau}_\mu$				
25	−3.75	−3.33	−3.00	−2.63	−0.37	0.00	0.34	0.72
50	−3.58	−3.22	−2.93	−2.60	−0.40	−0.03	0.29	0.66
100	−3.51	−3.17	−2.89	−2.58	−0.42	−0.05	0.26	0.63
250	−3.46	−3.14	−2.88	−2.57	−0.43	−0.07	0.24	0.62
500	−3.44	−3.13	−2.87	−2.57	−0.44	−0.07	0.24	0.61
∞	−3.43	−3.12	−2.86	−2.57	−0.44	−0.07	0.23	0.60
				$\hat{\tau}_{\mu,sym.}$				
25	−3.40	−3.02	−2.71	−2.37	−0.83	−0.73	−0.65	−0.59
50	−3.28	−2.94	−2.66	−2.35	−0.84	−0.73	−0.65	−0.58
100	−3.23	−2.90	−2.64	−2.34	−0.84	−0.73	−0.65	−0.58
250	−3.20	−2.88	−2.62	−2.34	−0.85	−0.73	−0.66	−0.58
500	−3.19	−2.88	−2.62	−2.33	−0.85	−0.73	−0.66	−0.58
∞	−3.17	−2.87	−2.62	−2.33	−0.85	−0.73	−0.66	−0.58

5. Power Study

We generate 50 observations from the model

$$Y_t = \mu(1-\rho)+\rho Y_{t-1} + e_t$$

with $Y_0 = 0$, and $e_t \sim$ NID (0, 1). We consider the values μ=0 and ρ=.98, .95, .90, .85, .80, and .70. For each parameter combination 5,000 data sets were generated and the percentage of runs for which the test criteria reject the unit root null hypothesis were recorded.

In Fig. 1, we give the empirical powers of the test criteria of the form $n(\hat{\rho}-1)$, often called normalized bias. We have

$\hat{\rho}_{\mu,\mathrm{OLS}}$: ordinary least squares estimator obtained by regressing Y_t on 1 and Y_{t-1}. This is the solid line.

$\hat{\rho}_{\mu,\mathrm{sym}}$: symmetric estimator obtained by regressing Y_t on 1, Y_{t-1} and on 1 and Y_{t+1} as in Dickey, Hasza, and Fuller (1984). This is the middle dashed line.

$\tilde{\rho}_{\mathrm{m,\,ml}}$: maximum likelihood estimator obtained as a solution to the fifth degree polynomial (3.10). This is one of the top, nearly coincident, dashed lines.

$\hat{\rho}^*_{\mathrm{m}}$: approximate maximum likelihood estimator obtained as a solution to the fourth degree polynomial (3.9). This is the other dashed line at the top.

Figure 2, shows the empirical powers of the corresponding pivotal statistics, $\hat{\tau}_{\mu}$, $\hat{\tau}_{\mu,\mathrm{sym}}$, $\tilde{t}_{\mathrm{m,\,ml}}$, and $\hat{t}^*_{\mathrm{m}}$, for testing $\rho = 1$.

We observe that:

1. The test criteria based on $\tilde{\rho}_{\mathrm{m,\,ml}}$ and its approximation $\tilde{\rho}^*_{\mathrm{m}}$ have much higher power than the criteria based on the OLS estimate.
2. The test criteria based on the pivotal statistics have marginally higher power than the criteria based on the corresponding normalized bias statistics. Also, for some ρ values, the empirical power of $\tilde{t}_{\mathrm{m,\,ml}}$ is almost twice that of Dickey and Fuller (1979) statistic $\hat{\tau}_{\mu}$.
3. The test criteria based on the symmetric estimator have powers between that of the OLS and $\tilde{\rho}_{\mathrm{m,\,ml}}$.

A more extensive study may be found in Chapter 4 in G-F (1992).

4. It is not surprising to see the best power associated with the statistic whose 5*th* percentile, under the null hypothesis, is closest to 0. After all, under the stationary alternative, the estimators in each table converge to the same distribution.

Figure 1: Normalized Biases Power, n=50

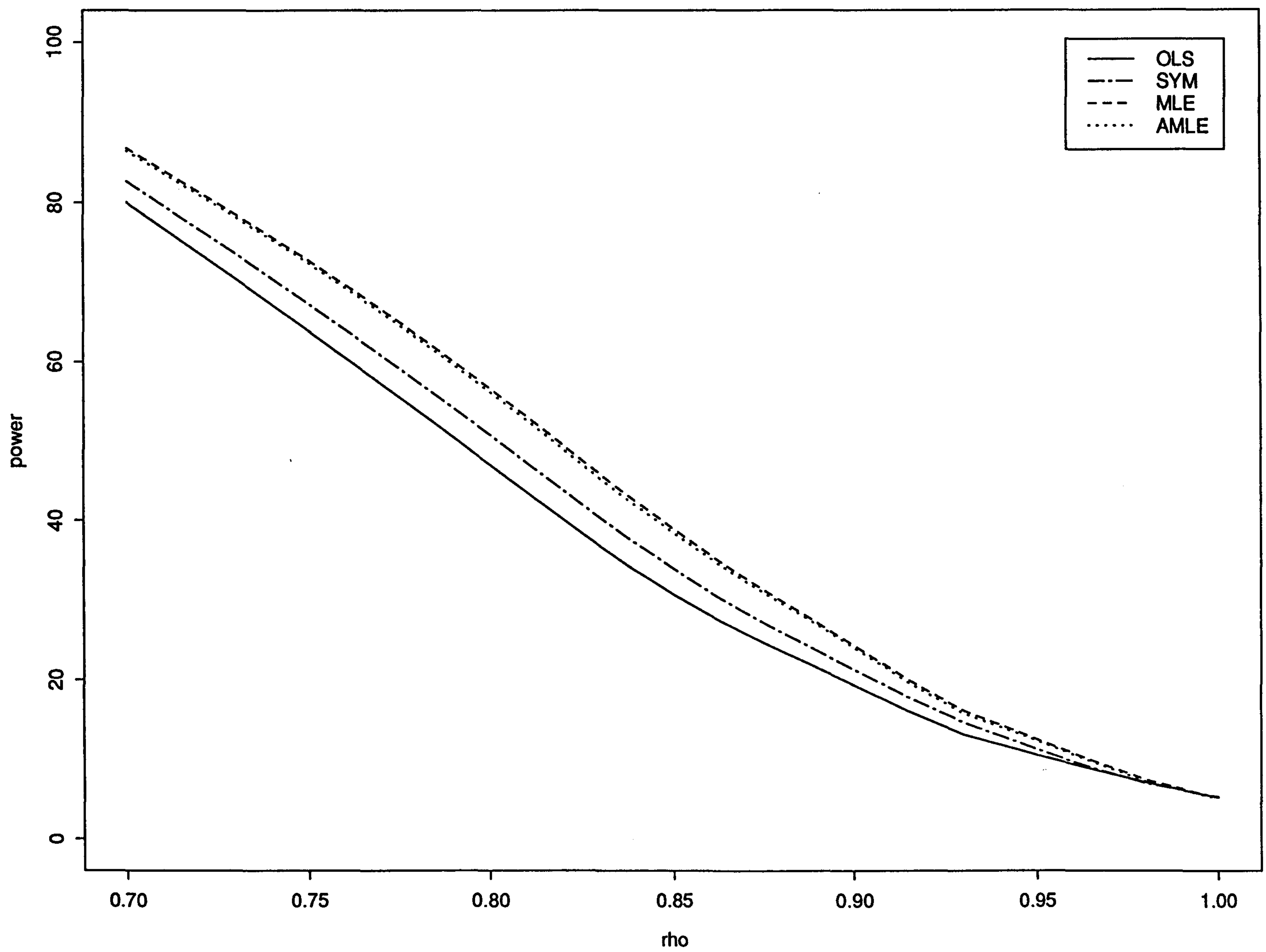
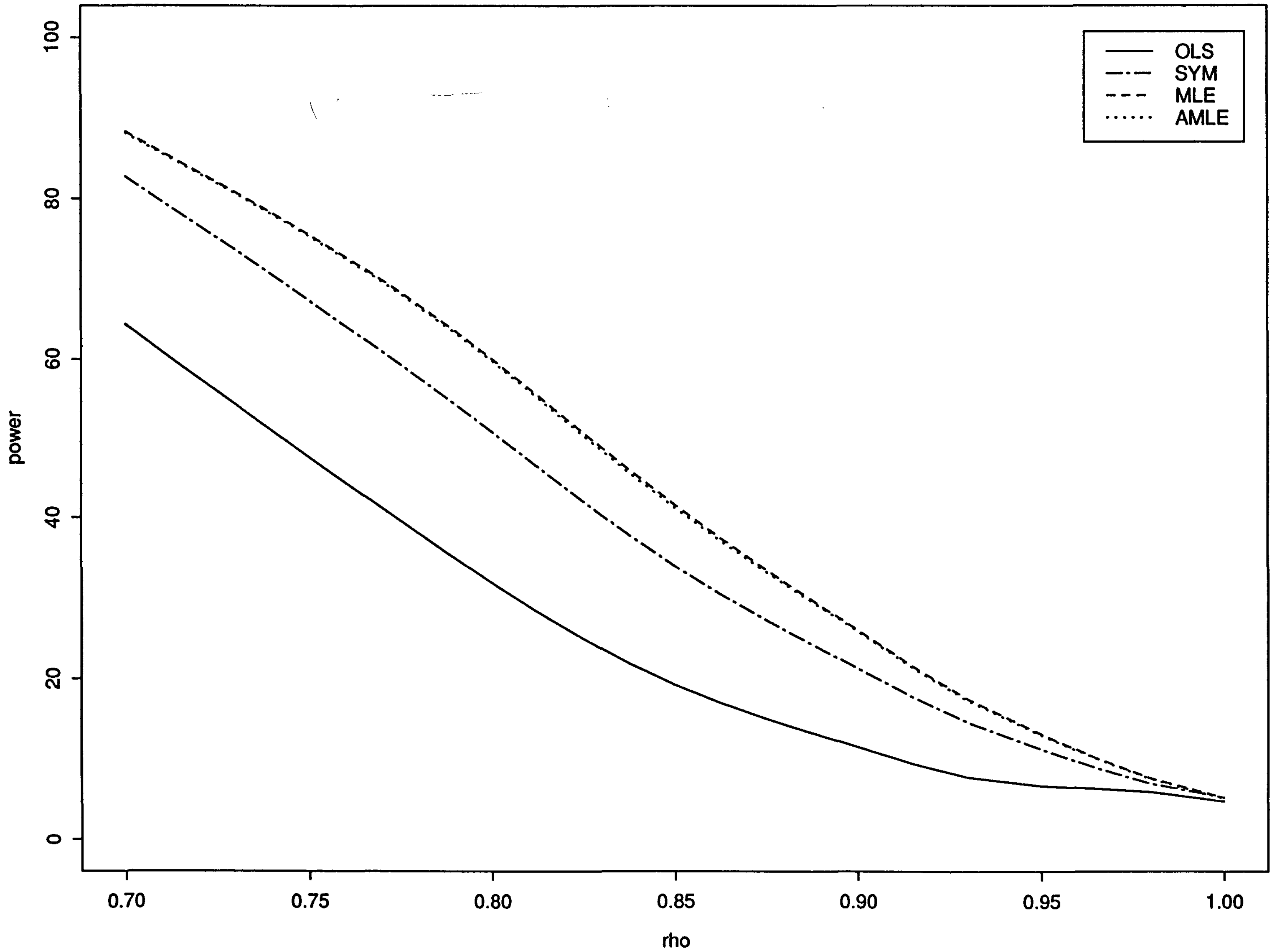


Figure 2: Pivotal Statistics Power, n=50



REFERENCES

(1) Breiman, L. (1968). Probability, Addison-Wesley, Reading, Mass.

(2) Chan, N. H. and Wei, C. Z. (1988). Limiting distribution of least squares estimates of unstable autoregressive processes. The Annals of Statistics, **16**, 367-401.

(3) Dickey, D. A. (1976). Estimation and Hypothesis Testing in Nonstationary Time Series, PhD thesis, Iowa State University.

(3) Dickey, D. A. and 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.

(4) Dickey, D. A., Hasza, D. P., and Fuller, W. A. (1984). Testing for Unit Roots in Seasonal Time Series. Journal of the American Statistical Association, **79**, 355-367.

(5) Fuller, W. A. (1976). Introduction to Statistical Time Series. Wiley, New York.

(6) Gonzalez-Farias, G. (1992). A new unit root test for autoregressive time series. Unpublished Dissertation, North Carolina State University, Raleigh, N. C.

(7) Hasza, D. P. (1980). A note or maximum likelihood estimation for the first-order autoregressive process. Communications in Statistics - Theory and Methods, A9(13), 1411-1415.