HYPOTHESIS TESTS FOR REGRESSION MODELS
WITH AUTOCORRELATED ERRORS

by

Steve P. Millard

&

Peter Guttorp

Tech. Report # 87

August 1985

Prepared in part under support to SIMS from the United States Environmental Protection Agency
Department of Statistics, University of British Columbia, 2021 West Mall, Vancouver, B.C. V6T 1W5
Hypothesis Tests for Regression Models With Autocorrelated Errors

Steven P. Millard

Center for Quantitative Science in Forestry, Fisheries and Wildlife, HR-20, University of Washington, Seattle, WA 98195

and

Peter Guttorp

Department of Statistics, GN-22, University of Washington, Seattle, WA 98195

SUMMARY

Three new methods of conducting tests of hypotheses for regression models with autocorrelated errors for small sample sizes are proposed. The methods are introduced in the context of the one-sample location problem, and their extension to the general case discussed. The new methods are based upon: 1) a jackknifed estimate of standard error, 2) a bootstrapped estimate of standard error, and 3) a saddlepoint-type approximation to the density of the estimate of location. The performance of each method was judged according to how closely the level of type I error was maintained. Of these three methods, only method 1 was judged to offer a substantial improvement over the usual asymptotically correct likelihood-based tests. An application of method 1 to sunspot data is presented.

KEY WORDS: Autocorrelated errors; Bootstrap; Jackknife; Maximum likelihood; Saddlepoint approximation; Type I error
1. INTRODUCTION

Consider the usual regression model

\[ y = X\beta + \epsilon \]  

(1.1)

where \( y \) is the \( n \times 1 \) vector of observations, \( X \) is the \( n \times p \) matrix of explanatory variables of rank \( p \), \( \beta \) is the \( p \times 1 \) vector of regression coefficients, and \( \epsilon \) is the \( n \times 1 \) vector of errors. Suppose we wish to test the hypothesis \( H_0: L\beta = c \), where \( L \) is a known \( q \times p \) matrix of rank \( q \), and \( c \) is a known \( q \times 1 \) vector. A common assumption is that \( \epsilon \) has a multivariate normal distribution with mean zero and covariance matrix \( \sigma^2 I_n \), where \( I_n \) denotes the \( n \times n \) identity matrix. Under this assumption, the least squares estimator of \( \beta \), given by

\[ \hat{\beta} = (X'X)^{-1}X'y, \]  

(1.2)

is distributed as multivariate normal with mean \( \beta \) and variance-covariance matrix \( \sigma^2(X'X)^{-1} \). The statistic

\[ F = \frac{(L\hat{\beta} - c)'[L(X'X)^{-1}L']^{-1}(L\hat{\beta} - c)}{q\hat{\sigma}^2} \]  

(1.3)

has an F-distribution with \( q \) and \( n-p \) degrees of freedom under the null hypothesis (Seber 1977), where

\[ \hat{\sigma}^2 = (y - X\hat{\beta})'(y - X\hat{\beta})/(n-p), \]  

(1.4)
and $A'$ denotes the transpose of a matrix $A$.

Suppose now that the vector $\mathbf{y}$ represents $n$ observations of a time series taken between equally spaced time intervals. Such data often arise in biological, environmental, economic, and industrial applications (e.g., Turk 1978, Tiao 1983, Harvey 1981, Box and Jenkins 1976). In such a situation, it is usually desirable to assume some sort of correlation structure in the errors since the observations cannot be randomized with respect to time. A common, simple model for temporal correlation is the first-order autoregressive, or AR(1), process (Box and Jenkins 1976)

$$
\epsilon_t = \rho \epsilon_{t-1} + \eta_t, \quad (1.5)
$$

where $\epsilon_t$ is the error at time $t$ (the $t$'th element of $\epsilon$), and $\eta$ is the $n \times 1$ vector of random innovations that has a multivariate normal distribution with mean zero and covariance matrix $\sigma^2 \mathbf{I}_n$. In the stationary case (the only case considered in this paper), $|\rho| < 1$ and $\rho$ is the lag-one correlation in the errors. Thus $\epsilon$ has a multivariate normal distribution with mean zero and covariance matrix $\sigma^2 \mathbf{V}(\rho)$, where $\mathbf{V}(\rho)$ is an $n \times n$ symmetric matrix of the form

$$
\mathbf{V}(\rho) = \frac{1}{(1-\rho^2)} \begin{bmatrix}
1 & \rho & \ldots & \rho^{n-1} \\
\rho & 1 & \ldots & \\
\vdots & \ddots & \ddots & \\
\rho^{n-1} & \ldots & 1
\end{bmatrix} \quad (1.6)
$$

If $\rho=0$, $\mathbf{V}(\rho)$ reduces to $\mathbf{I}_n$. 
If the errors follow the structure given by (1.5), the F-statistic of (1.3) can no longer be used to test \( H_0: L\beta = c \), since its distribution under \( H_0 \) is unknown. The consequences of using this F-statistic when the errors are temporally correlated have been explored by Cochran (1947), Box (1954), Scheffé (1959), Hotelling (1961), Lissitz and Chardos (1975), Praetz (1981), and Millard et al. (1985). The effect of correlated errors on the nonparametric analogues of the F-statistic was studied by Gastwirth and Rubin (1971), and Pettitt and Siskind (1981). The general conclusion from all of these studies is that when the errors are correlated, the probability of a type I error is increased above or decreased below the nominal \( \alpha \)-level, depending upon whether the correlation is positive or negative, respectively. The magnitude of the deviation of the type I error probability away from \( \alpha \) depends upon the magnitude of the correlation, but even slight correlation can grossly affect the type I error level. In other words, a confidence interval (region) for \( L\beta \) based upon the usual assumption of independent errors will be too small or too large, depending upon whether the correlation is positive or negative, respectively.

If the value of \( \rho \) is known, the generalized least squares (GLS) estimator of \( \beta \) is

\[
\hat{\beta}(\rho) = (X'V(\rho)^{-1}X)^{-1}X'V(\rho)^{-1}y,
\]

(1.7)
distributed as multivariate normal with mean \( \beta \) and variance-covariance matrix \( \sigma^2(X'V(\rho)^{-1}X)^{-1} \). The statistic
\[ F(\rho) = \frac{(L\hat{\beta}(\rho) - c)'(L[X'V(\rho)^{-1}X]^{-1}L')^{-1}(L\hat{\beta}(\rho) - c)}{q\hat{\sigma}^2(\rho)} \] (1.8)

has an F-distribution with q and n-p degrees of freedom under $H_0$ (Seber 1977), where

\[ \hat{\sigma}^2(\rho) = (y - X\hat{\beta}(\rho))'(y - X\hat{\beta}(\rho))/ (n-p). \] (1.9)

A similar problem arises in two-way analysis of variance (ANOVA) with observations within a row serially correlated (Larsen 1969, Ljung and Box 1980, Andersen et al. 1981). If the ANOVA involves r rows and n columns, we can write this model in terms of the expressions of (1.1), where now $y$ and $\varepsilon$ are both $rn \times 1$ vectors, and the covariance matrix of $\varepsilon$ is given by $\sigma^2 V$, where $V$ is an $rn \times rn$ block diagonal matrix of the form

\[ V = \begin{bmatrix} V(\rho) & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & V(\rho) \end{bmatrix} \] (1.10)

and $V(\rho)$ is the $n \times n$ matrix defined in (1.6).

In many applications, a researcher may have time series data and no a priori knowledge of the value of $\rho$. The question then arises of how to properly conduct a test of $H_0: L\beta = c$. As Anderson (1971, p.258) notes, when $\rho$ is unknown, the minimal sufficient statistic is of dimension greater than the number of unknown parameters, hence it is impossible to formulate inference problems in a way that leads to
relatively simple optimal procedures. It is the purpose of this paper to review the techniques that have been proposed to deal with this problem, and to present some new methods for conducting such tests. The new methods are introduced in the context of the one-sample location problem, and their extension to the general case discussed. The most successful method is illustrated with sunspot data.

2. PREVIOUSLY PROPOSED METHODS

One obvious solution to the problem posed above is to appeal to the asymptotic behavior of maximum likelihood estimates (MLE's). If we define $\theta = (\sigma^2, \rho, \beta')'$ to be the $g \times 1$ vector of parameters, where $g = p+2$, and $\hat{\theta}$ is the MLE of $\theta$, then $\hat{\theta}$ is asymptotically distributed as a multivariate normal random vector with mean $\theta$ and covariance matrix equal to the inverse of the Fisher information matrix (e.g., Anderson 1971). If $\mathcal{J}(\theta)$ denotes the $g \times g$ Fisher information matrix, then it can be shown (Beach and MacKinnon 1978)

$$\mathcal{J}(\theta) = \begin{bmatrix} A & C & O \\ C & B & O \\ O & O & D \end{bmatrix}$$  \hspace{1cm} (2.1)

where

$$A = n/2\sigma^4$$, \hspace{1cm} (2.2)

$$B = [(n-1)-(n-3)p^2]/(1-p^2)^2 = n/(1-p^2),$$  \hspace{1cm} (2.3)
\[ C = \rho / \sigma^2 (1 - \rho^2), \quad (2.4) \]

and the \( jk \)'th element of the \( p \times p \) matrix \( D \) is given by

\[ D_{jk} = (1/\sigma^2) [(1 - \rho^2) x_{1j} x_{1k} + \sum_{t=2}^{n} (x_{tj} - \rho x_{t-1,j})(x_{tk} - \rho x_{t-1,k})]. \quad (2.5) \]

The expression for \( B \) given in Beach and MacKinnon (1978), unlike the one given in (2.3), is based upon the observed, rather than expected, information.

The MLE of \( \beta \) when \( \rho \) is unknown is given by \( \hat{\beta}(\hat{\rho}) \), where \( \hat{\beta}(\rho) \) is defined by (1.7) and \( \hat{\rho} \) is the MLE of \( \rho \). Thus, to test \( H_0: \beta = c \), we can apply the usual Wald statistic (Silvey 1975)

\[ W = [L \hat{\beta}(\hat{\rho}) - c]' (L D(\hat{\theta})^{-1} L')^{-1} [L \hat{\beta}(\hat{\rho}) - c] \quad (2.6) \]

which is asymptotically distributed as a chi-square random variable with \( q \) degrees of freedom under \( H_0 \). Now \( D(\theta)^{-1} = \sigma^2 [X' V(\rho)^{-1} X]^{-1} \), so the Wald statistic of (2.6) is equivalent to \( q F(\hat{\rho}) \), where \( F(\rho) \) is defined in (1.8).

The major drawback of using likelihood related tests is that the goodness of the normal approximation will depend on the magnitudes of \( n, \rho, \) and \( p \). The results of this paper show that, in general, relatively large sample sizes are required even if \( \rho \) is only moderately large (e.g., \( n \geq 50 \) for \( p = 1 \) and \( \rho = 0.3 \)).

Albers (1978) constructed an asymptotically correct test for the one-sample location problem. His statistic is asymptotically
equivalent to the Wald statistic (2.6). Eriksson (1981) investigated small sample properties of confidence intervals based on Albers' modified t-statistic.

Vinod (1976) calculated critical values for the distribution of the F-statistic of (1.3) when the errors follow an AR(1) or MA(1) process. He also calculated critical values for the distribution of the F-statistic of (1.8) when an estimate is used for the unknown value of $\rho$. In both cases the critical values depend upon the true value of $\rho$.

Brillinger (1980) is a general overview of problems involving temporal dependence. Box and Tiao (1965, 1975) discussed the two-sample problem, i.e., testing for a step change intervention. They analyzed the case of unknown correlation in a Bayesian fashion. Lettenmaier (1976) studied the one- and two-sample location problem for known correlation. Anderson et al. (1981) approximated the usual F-statistic for two-way ANOVA when the errors within rows constitute a stationary Gaussian process with known covariance function. Larsen (1969) and Ljung and Box (1980) studied the ANOVA problem from a Bayesian point of view.

In the next section, results of a Monte Carlo study of the performance of the $W$-statistic of (2.6) for small sample sizes are given. Subsequently, three new methods for dealing with autocorrelated errors in regression models are presented.

3. PERFORMANCE OF THE USUAL LIKELIHOOD-BASED STATISTICS FOR SMALL SAMPLE SIZES

Consider the one-sample location problem, where the expressions of (1.1) reduce to $X = 1_n$ and $\theta = \mu$, where $1_n$ denotes the $n \times 1$
vector of 1's, and μ is the mean of the population from which the sample was drawn. When ρ is known, the GLS estimates of μ and σ² are given by

\[ \hat{\mu} = \hat{\mu}(\rho) = \frac{\sum_{t=2}^{n-1} y_t + y_n}{[2 + (n-2)(1-\rho)]}, \]

(3.1)

and

\[ \hat{\sigma}^2 = \hat{\sigma}^2(\rho) = \frac{\text{RSS}(\hat{\mu}, \rho)}{(n-1)}, \]

(3.2)

where

\[ \text{RSS}(\mu, \rho) = (1-\rho^2)(y_1-\mu)^2 + \sum_{t=2}^{n} [(y_t-\mu) - \rho(y_{t-1}-\mu)]^2. \]

(3.3)

Under the null hypothesis H₀: μ = c, the quantity

\[ t(\rho) = \frac{(\hat{\mu}(\rho) - c)/\hat{\sigma}^2(\rho)/[(1-\rho^2) + (n-1)(1-\rho)^2]}{1/2} \]

(3.4)

is distributed as Student's t with n-1 degrees of freedom. Note that \([t(\rho)]^2\) corresponds to the F-statistic given in (1.8) for L=1 and q=1.

When ρ is unknown, \([t(\hat{\rho})]^2\) corresponds to the W-statistic of (2.6) and is asymptotically distributed as a chi-square random variable with 1 degree of freedom; \(t(\hat{\rho})\) is asymptotically \(N(0,1)\). Table 1 summarizes the performance of both \(t(\hat{\rho})\) and the usual
t-statistic (under the assumption of independent errors) for various values of \( n \) and \( \rho \). For a given value of \( n \) and \( \rho \), observations were generated according to (1.1) and (1.5), with \( \mu = 0 \) and \( \sigma = 1 \), and the values of the two t-statistics computed with \( c = 0 \) (i.e., \( H_0 \) was true). This was done for 500 Monte Carlo trials. The proportion of values greater than the critical t-percentile associated with \( n-1 \) degrees of freedom was recorded for the upper one-, five-, and ten-percentiles. If \( \hat{\alpha} \) denotes the Monte Carlo estimate of \( \alpha \) (the true probability of a type I error), then the standard deviation (s.d.) of \( \hat{\alpha} \) is approximately \([\hat{\alpha}(1-\hat{\alpha})/r]^{1/2}\), where \( r \) denotes the number of Monte Carlo trials. Hence, for Table 1, the s.d. of \( \hat{\alpha} \) is approximately \( .04[\hat{\alpha}(1-\hat{\alpha})]^{1/2} \).

The results for the usual t-statistic conform to the findings cited previously. As \( \rho \) increases, the type I error increases drastically above the nominal (assumed) \( \alpha \)-level. The results for \( t(\hat{\rho}) \) show that when \( \rho \) reaches the value 0.4, the type I error can be inflated for sample sizes of 30 or smaller.

One might conjecture that the results in Table 1 for \( t(\hat{\rho}) \) are due to the possibility that \( \text{Var}[\hat{\rho}] \) is much larger than \( \text{Var}[\hat{\rho}(\rho)] \), and hence the denominator of \( t(\hat{\rho}) \), as given in (3.4), underestimates s.d.\([\hat{\rho}] \). Table 2 shows that this is not the case. The standard deviation of \( \hat{\rho}(\rho) \) is given by

\[
\text{s.d.}[\hat{\rho}(\rho)] = \frac{\sigma^2}{[(1-\rho^2) + (n-1)(1-\rho)^2]^{1/2}} \quad (3.5)
\]

which is recorded in column 3 of Table 2 for \( \sigma = 1 \). The estimate of s.d.\([\hat{\rho}] \), based on 500 Monte Carlo trials, is given in column 4 of
Table 2. It is evident that replacing \( \rho \) with \( \hat{\rho} \) induces little added variability in the estimate of \( \mu \), i.e., \( \text{s.d.}[\hat{\mu}(\hat{\rho})] \approx \text{s.d.}[\hat{\mu}(\rho)] \). A Taylor series expansion given in Appendix A confirms this finding analytically.

The real problem with approximating the distribution of \( t(\hat{\rho}) \) with a \( t \)- or normal distribution lies in the bias of \( \hat{\rho} \). Any estimate of \( \rho \) that is constrained to lie within the stationarity bounds of \((-1, +1)\) is biased towards 0, and the bias increases with the magnitude of \( \rho \) (Marriott and Pope 1954, De Gooijer 1980, Taibah and Kassab 1981, Tjøstheim and Paulsen 1983). Figures 1 and 2 show the bias of \( \hat{\rho} \) for three different estimators of \( \rho \) for various values of \( n \) and \( \rho \). \( \hat{\rho}_{\text{ML}} \) denotes the MLE of \( \rho \), \( \hat{\rho}_{\text{YW}} \) denotes the Yule-Walker estimate of \( \rho \), given by

\[
\hat{\rho}_{\text{YW}} = C_1(\bar{y})/C_0(\bar{y}) \quad (3.6)
\]

with

\[
C_j(\mu) = (1/n) \sum_{t=1+j}^{n} (y_t - \mu)(y_{t-j} - \mu), \quad (3.7)
\]

and \( \hat{\rho}_{\text{BG}} \) denotes Berg's maximum entropy estimate (Berg 1967), given by

\[
\hat{\rho}_{\text{BG}} = C_1(\bar{y})/C_0^*(\bar{y}) \quad (3.8)
\]

with
\[ C_0^*(\mu) = C_0(\mu) - S(\mu) \]  
\( (3.9) \)

and

\[ S(\mu) = (1/n)\left\{\left[(y_1-\mu)^2 + (y_n-\mu)^2\right]/2\right\}. \]  
\( (3.10) \)

The "S" notation denotes a correction factor for the starting and stopping points. Figure 1 shows the bias of all three estimators of \( \rho \) as a function of \( \rho \) for \( n=20 \), based upon 500 Monte Carlo trials. It is apparent that \( \hat{\rho}_{ML} \) is the least biased estimator of \( \rho \), and \( \hat{\rho}_{YM} \) the most biased of these three. Figure 2 shows the bias of \( \hat{\rho}_{ML} \) as a function of \( \rho \) for \( n=10, 20, \) and 30.

Because \( \hat{\rho} \) underestimates \( \rho \) for \( \rho > 0 \), the denominator of \( t(\hat{\rho}) \) underestimates s.d.\( [\hat{\mu}(\hat{\rho})] \) (see (3.4) and (3.5)). The following three sections discuss various ways to circumvent this bias problem in the case of small samples in order to conduct tests of hypotheses on \( \mu \).

4. APPROXIMATING THE DENSITY OF THE ESTIMATE OF LOCATION

The first method proposed for conducting tests of hypothesis on regression models with autocorrelated errors is based upon the work of Durbin (1980a,b). Durbin (1980a) developed a saddlepoint-type approximation for the densities of sufficient estimators that are not necessarily standardized sums of independent and identically distributed vectors. The error term of this approximation is of order \( n^{-1} \) (and can usually be reduced further to order \( n^{-3/2} \) by
renormalization), compared to the error term of order $n^{-1/2}$ of the usual asymptotic normal density approximation.

The approximation can be described as follows. If $I_n$ is an $m$-dimensional sufficient estimator of (an $m$-dimensional parameter) $\theta$ which has bias of order $n^{-1}$ at most, then the approximation for the density of $I_n$ at $I_n = \theta$ is

$$g(t, \theta) = (2\pi)^{-m/2} |G(t)|^{1/2} \left[ f(y, \theta)/f(y, t) \right] \left[ 1 + O(n^{-1}) \right]$$

(4.1)

where $f(y, \theta)$ is the density of the underlying observations $y$. In the context of a regression model with autocorrelated errors, for the one-sample location problem we have $\theta = (\sigma^2, \rho, \mu)^t$ and

$$f(y, \theta) = (2\pi \sigma^2)^{-n/2} (1-\rho^2)^{1/2} \exp[-\text{RSS}(\mu, \rho)/2\sigma^2]$$

(4.2)

where $\text{RSS}(\mu, \rho)$ is defined in (3.3).

A problem arises in applying Durbin's approximation in that, as previously mentioned, the minimal sufficient statistic is of dimension greater than 3 (in this case it is of dimension 5). To circumnavigate this problem, Durbin (1980b), working with estimators of partial correlation coefficients, used a modified density for $y$, which for the one-sample location problem with AR(1) errors is given by

$$f^*(y, \theta) = (2\pi \sigma^2)^{-n/2} [(1-\rho^2)(1-\rho)/(1+\rho)]$$

$$ \exp[-\text{RSS}^*(\mu, \rho)/2\sigma^2],$$

(4.3)
where

\[ RSS^*(\mu, \rho) = (1+\rho^2) n C_0(\mu) - 2 \rho n C_1^*(\mu), \]  
\[ C_1^*(\mu) = C_1(\mu) + S(\mu), \]

and \( C_0(\mu), C_1(\mu), \) and \( S(\mu) \) are defined in (3.7) and (3.10). Note that

\[ RSS^*(\mu, \rho) = RSS(\mu, \rho) - 2 \rho (1-\rho) S(\mu), \]

where \( RSS(\mu, \rho) \) is defined in (3.3). For the modified density of (4.3), a sufficient estimate of \( \Theta \) is given by \( \hat{\Theta}=(\hat{\sigma}^2, \hat{\rho}, \hat{\mu})' \) where

\[ \hat{\sigma}^2 = RSS^*(\hat{\mu}, \hat{\rho})/n, \]

\[ \hat{\rho} = C_1^*(\hat{\mu})/C_0(\hat{\mu}), \]

and

\[ \hat{\mu} = \bar{y}. \]

Ignoring the exponentially small term \( \rho^{2n} \) in \( f^*(\chi, \Theta) \), the saddlepoint approximation to the density of \( \hat{\Theta} \) is given by

\[ \hat{g}(\hat{\Theta}, \Theta) = \left( n/2\pi \right)^{3/2} [e^{n(1-\rho)/2(1+\rho)}]^{1/2} (\sigma_{n-3}^n)^{1/2} \exp[-RSS^*(\mu, \rho)/2\sigma^2] \]
Rewriting $RSS^*(\mu, \rho)$ as

$$RSS^*(\mu, \rho) = n\sigma^2 + nC_0(\hat{\mu})(\hat{\rho} - \rho)^2 + n(1-\rho)^2(\hat{\mu} - \mu)^2$$  \hspace{1cm} (4.11)$$

and integrating out $\sigma^2$ and $\rho$ yields

$$\hat{g}(\hat{\mu}, \Theta) = \left[ \frac{n}{2\pi} \right]^{1/2} \left[ \frac{(1-\rho)}{(1+\rho)} \right]^{1/2} \left[ \frac{(n-1)/n}{1} \right]^{(n/2)-1} \left[ \frac{e}{C_0(\hat{\mu})} \right]^{1/2}
\{ \Phi[(1-\rho)(nC_0(\hat{\mu}))/\sigma^2]^{1/2} - \Phi[(-1-\rho)(nC_0(\mu))/\sigma^2]^{1/2} \}
\exp[-n(1-\rho)^2(\hat{\mu} - \mu)^2/2\sigma^2]$$  \hspace{1cm} (4.12)$$

where $\Phi$ denotes the c.d.f. of a N(0,1) random variable. In this case the true density of $\hat{\mu}$ is normal and given by

$$g(\hat{\mu}, \Theta) = \left[ \frac{n}{2\pi} \right]^{1/2} \left[ \frac{(1-\rho)}{\sigma} \right]^{1/2} \exp[-n(1-\rho)^2\xi_n^2(\hat{\mu} - \mu)^2/2\sigma^2],$$  \hspace{1cm} (4.13)$$

where

$$\xi_n = \frac{(1-\rho^2)}{[(1-\rho^2) + \frac{2\rho}{n}(\rho-1)]},$$  \hspace{1cm} (4.14)$$

as given in Lettenmaier (1976), for example. Comparing (4.12) and (4.13), we see that they are asymptotically equivalent, since the quantity $[C_0(\hat{\mu})/n]$ approaches $\sigma^2/(1-\rho^2)$ in probability and $\xi_n$ approaches 1 as $n$ tends to infinity.

The approximation to the density $g(\hat{\mu}, \Theta)$ of (4.12) can be used to conduct tests of the form $H_0: \mu=c$. For a one-sided test against the
alternative $H_A$: $\mu > c$, one would reject $H_0$ at level-$\alpha$ if the quantity $p_D < \alpha$, where $p_D$ is defined by

$$p_D = \int_{\hat{\mu}}^{\infty} \frac{\hat{g}(t, \theta_0)dt}{\int_{-\infty}^{\infty} \hat{g}(t, \theta_0)dt} \quad (4.15)$$

and $\theta_0 = (\sigma^2, \rho, c)'$. Note, however, that the density approximation depends on the true and unknown values of $\sigma^2$ and $\rho$. This is hardly surprising since the distribution of $\hat{\mu}$ depends on $\sigma^2$ and $\rho$ (see (4.13) and (4.14)). This difficulty was not encountered by Durbin (1980b), since the asymptotic distribution of a partial correlation coefficient depends only on one parameter: the true partial correlation. The true values of $\sigma^2$ and $\rho$ in (4.12) could be replaced by their estimates $\hat{\sigma}^2$ and $\hat{\rho}$ given in (4.7) and (4.8).

At this point, one may wonder why anyone would use Durbin's approximation to the density of $\hat{\mu}$ (4.12) when the exact density is known ((4.13) and (4.14)). The problem is that the exact density of $\hat{\mu}$ also depends upon the true (and unknown) values of $\sigma^2$ and $\rho$. Hence, it is hoped that Durbin's approximation, with $\sigma^2$ and $\rho$ replaced by their estimates, performs better than the usual asymptotic normal approximation to the density of $\hat{\mu}(\hat{\rho})$, and better than the approximation that $g(\hat{\mu}, \hat{\rho})$ is normal.

Table 3 summarizes the performance of the density approximation in the case when $\sigma^2$ and $\rho$ are known, and in the case when $\sigma^2$ and $\rho$ are replaced by their estimates. In the former case, it is seen that the approximation works well for small and moderate values of $\rho$, but
for $\rho = 0.8$, the approximation leads to conservative tests (i.e., the true type I error probability is below $\alpha$). In the case when $\sigma^2$ and $\rho$ are replaced by their estimates, the approximation is poor for all positive values of $\rho$; the tests are too liberal. Comparison of Tables 1 and 3 shows that Durbin's approximation with $\sigma^2$ and $\rho$ unknown offers no improvement over the usual $W$-statistic $t(\hat{\rho})$ of (2.6) and (3.4).

5. A MODIFIED $t$-STATISTIC BASED ON A BOOTSTRAPPED ESTIMATE OF STANDARD ERROR

The second method proposed for conducting tests of hypothesis on regression models with autocorrelated errors uses the bootstrap (Efron 1982). In the context of GLS with estimated covariance matrices, this method has already been outlined by Freedman and Peters (1984). We will now present the basic idea in the context of the one-sample location problem.

The goal is to estimate $\text{sd}[^{\hat{\mu}(\hat{\rho})}]$, where $\hat{\mu}(\rho)$ is defined in (3.1). We examine the behavior of $\hat{\mu}(\hat{\rho})$ computed from "pseudo-data" generated by resampling the residuals from the fit to the original data. More explicitly,

1. From the original data, compute $\hat{\sigma}^2$, $\hat{\rho}$, and $\hat{\mu}$.

2. Compute the residuals as $\hat{\varepsilon}_t = y_t - \hat{\mu}$, $t=1,...,n$.

3. Compute the estimated random innovations as $\hat{\eta}_1 = (1-\hat{\rho}^2)^{1/2}\hat{\varepsilon}_1$, and $\hat{\eta}_t = \hat{\varepsilon}_t - \hat{\rho}\hat{\varepsilon}_{t-1}$, $t=2,...,n$.

4. For bootstrap trial $i$,

   a) Create the pseudo-errors: $\varepsilon^*_1 = [1/(1-\hat{\rho}^2)^{1/2}]\eta^*_1$, and $\varepsilon^*_t = \hat{\rho}\varepsilon^*_{t-1} + \eta^*_t$, $t=2,...,n$, where the pseudo-random innovations $\eta^*_1,...,\eta^*_n$ are generated.
by choosing from the set \( \hat{\eta}_1, \ldots, \hat{\eta}_n \) in a multinomial fashion, with equal probability of choosing each \( \hat{\eta}_t \).

b) Create the pseudo-observations: \( y^*_t = \hat{\mu} + \epsilon^*_t \), \( t=1, \ldots, n \).

c) Estimate the parameters based upon the pseudo-observations: \( \theta^*, \hat{\rho}^* \), and \( \hat{\mu}^* \).

d) Compute the pseudo-error in the estimate of location: \( \Delta_i = \hat{\mu}^* - \hat{\mu} \).

5. Repeat step 4 \( N \) times. \( N \) is the number of bootstraps.

The quantity s.d.\( [\hat{\mu}(\hat{\rho})] \) is estimated by

\[
\theta^*_\mu = \left\{ \frac{1}{N-1} \sum_{i=1}^{N} (\Delta_i - \bar{\Delta})^2 \right\}^{1/2}
\]

(5.1)

It is then hoped that the statistic

\[
t_B = (\hat{\mu}(\hat{\rho}) - c)/\theta^*_\mu
\]

(5.2)

follows a \( t \)- or normal distribution (under \( H_0: \mu = c \)) more closely than the usual \( W \)-statistic \( t(\hat{\rho}) \) of (3.4) for small or moderate values of \( n \).

Table 4 summarizes the performance of \( t_B \) for various values of \( n \) and \( \rho \), when \( N=100 \), \( \mu=0 \), and \( \sigma=1 \). Two different forms of the modified \( t \)-statistic were considered: \( t_{BM} \), based on the MLE of \( \rho \), and \( t_{BY} \), based on the Yule-Walker estimate. This second statistic was
considered due to its ease of computation compared to the first. In fact, the Yule-Walker estimate of \( \rho \) can be considered a one-step MLE. The results of Table 4 show that there is little difference in type I error between \( t_{BM} \) and \( t_{BY} \). Comparison of Tables 1 and 4 shows that the modified t-statistics based on the bootstrap perform no better (in terms of being close to the nominal \( \alpha \)-level) than \( t(\hat{\rho}) \).

An alternative form of the bootstrap method was also investigated, in which an attempt was made to derive the density of \( t(\hat{\rho}) \), and thus the upper one-, five-, and ten-percentiles of its distribution. The procedure followed was exactly the same as that described above for the modified t-statistics, except that in step 4.d), rather than compute \( \Delta_i \) (the pseudo-error in \( \hat{\mu}(\hat{\rho}) \)), a pseudo-t-statistic \( t^*(\hat{\rho}) \) was computed according to (3.4) with \( c=\hat{\mu} \) (i.e., \( H_0: \mu=\hat{\mu} \) was true for the pseudo-observations). The distribution of the \( N t^*(\hat{\rho}) \)'s was then used as the reference distribution for \( t(\hat{\rho}) \) assuming that \( H_0: \mu=0 \) was true. For 100 Monte Carlo trials, the null hypothesis \( H_0: \mu=0 \) was rejected at (assumed) level-\( \alpha \) if \( t(\hat{\rho}) \) was larger than the estimated \( (1-\alpha)100^{th} \) percentile of its distribution. The results of this experiment for \( N=1000 \) bootstraps, \( \mu=0 \), and \( \sigma=1 \) are given in Table 5. Comparison of Tables 1 and 5 shows that again there is no substantial improvement over using \( t(\hat{\rho}) \) and following the usual procedure of comparing it to the critical value from Student's t-distribution.

6. A MODIFIED t-STATISTIC BASED ON A JACKKNIFED ESTIMATE OF STANDARD ERROR

The third method proposed for conducting tests of hypotheses on regression models with autocorrelated errors relies on the jackknife
(Quenouille 1956). As in the case of the first bootstrap method introduced in the previous section, the object of the method is to estimate \( \text{s.d.}[\hat{\rho}(\hat{\rho})] \). As noted in section 3, for small sample sizes, \( \hat{\rho} \) tends to underestimate \( \rho \) for \( \rho > 0 \), and hence the tails of the distribution of \( t(\hat{\rho}) \) are heavier than those of the standard Student's t-distribution. One possible way to circumvent the increased type I error associated with comparing \( t(\hat{\rho}) \) to Student's t-distribution, therefore, is to decrease the bias in \( \hat{\rho} \). Quenouille (1949, 1956) suggested a way to reduce the order \( n^{-1} \) bias of \( \hat{\rho} \) by using

\[
\hat{\rho} = 2\hat{\rho} - [(\hat{\rho}_1 + \hat{\rho}_2)/2],
\]

where \( \hat{\rho}_1 \) and \( \hat{\rho}_2 \) are the estimates of \( \rho \) based upon the first and second halves of the data, respectively. Unfortunately, in an experiment to compare the performance of \( t(\hat{\rho}) \) with \( t(\hat{\rho}) \), this method of reducing bias proved unsuccessful due to the frequency with which \( \hat{\rho} \) lay outside the stationarity boundaries (-1, +1). This result is not surprising since Mosteller and Tukey (1977, p.138) warn that the jackknife has not proved successful in cases where the possible values of the quantity jackknifed are restricted to a small interval. They suggest instead to jackknife a transformed estimate (thus reducing bias in the transformed space), and then transform back to the original interval. This method was tried with various transformations, including the variance stabilizing \( \sin^{-1} \). In all cases, the bias was reduced or remained the same for \( \rho \) in some small interval, but was increased for \( \rho \) outside that interval.
Rather than attempt to reduce the bias in $\hat{\beta}$, it would appear more logical to attempt to reduce the bias in the estimate of $s.d. [\hat{\mu}(\hat{\beta})]$ or some function of it. Unfortunately, the estimate of $s.d. [\hat{\mu}(\hat{\beta})]$ cannot be jackknifed because it explicitly involves $n$ (see (3.5)). It can be seen from (3.4), however, that $t(\hat{\beta})$ can be rewritten as

$$t(\hat{\beta}) = (n-1)^{1/2}(\hat{\mu}(\hat{\beta}) - c)(\hat{\gamma}_n^2 + \hat{\gamma}^2)^{1/2}$$  \hspace{1cm} (6.2)

where

$$\lambda_n = \{(1-\rho^2)/[(n-1)\sigma^2]\}^{1/2},$$  \hspace{1cm} (6.3)

$$\gamma = (1-\rho)/\sigma,$$  \hspace{1cm} (6.4)

and thus

$$t(\hat{\beta}) \approx t_a(\hat{\beta}) = (n-1)^{1/2}(\hat{\mu}(\hat{\beta}) - c)\hat{\gamma}$$  \hspace{1cm} (6.5)

for large $n$. Note that the ratio $t_a(\hat{\beta})/t(\hat{\beta})$ is given by $\hat{\gamma}^2/(\hat{\gamma}_n^2 + \hat{\gamma}^2)$, and thus $|t_a(\hat{\beta})| < |t(\hat{\beta})|$. This ratio is plotted for various values of $n$ and $\rho$ in Figure 3.

Now the quantity $\hat{\gamma}$ can be jackknifed to produce a less biased estimate of $\gamma$. The jackknifed estimate is given by

$$\hat{\gamma} = 2\hat{\gamma} - [(\hat{\gamma}_1 + \hat{\gamma}_2)/2]$$  \hspace{1cm} (6.6)
where \( \hat{\gamma}_1 \) and \( \hat{\gamma}_2 \) are the estimates of \( \gamma \) based upon the first and second halves of the data, respectively. Thus, the modified t-statistic based on the jackknifed estimate of s.d.\( [\hat{\gamma}(\hat{\rho})] \) can be written as

\[
t_J = (n-1)^{1/2}(\hat{\gamma}(\hat{\rho}) - c)\hat{\gamma}.
\]

(6.7)

Although \( \gamma \) is greater than 0 by definition, it is possible to have \( \hat{\gamma} \leq 0 \). Thus, in practice, instead of using \( t_J \), we must use

\[
t_J^* = \begin{cases} 
  t_J & \text{if } \hat{\gamma} > 0, \\
  t_a(\hat{\rho}) & \text{otherwise}.
\end{cases}
\]

(6.8)

It is worth noting that \( t_J^* \) is asymptotically distributed as a \( N(0,1) \) random variable (see Appendix B).

Table 6 summarizes the performance of \( t_a(\hat{\rho}) \) and \( t_J^* \) for various values of \( n \) and \( \rho \) when \( \mu=0 \) and \( \sigma=1 \). As in the case of the modified t-statistics based on the bootstrap, two different t-statistics were computed: \( t_J^{*M} \), based upon the MLE of \( \rho \), and \( t_J^{*Y} \), based upon the Yule-Walker estimate of \( \rho \). Comparison of Tables 1 and 6 shows that the modified t-statistics based on the jackknife method are superior to \( t(\hat{\rho}) \) in terms of approximating the assumed 1%, 5%, and 10% critical regions when \( \rho \) is positive. For \( \rho \) between 0.2 and 0.4, the \( t_J^* \) statistics behave well for \( n \geq 20 \). At \( \rho=0.6 \), the 5% and 10% critical regions are well approximated for \( n \geq 20 \), and for \( \rho=0.8 \), the 10% critical region is well approximated for \( n \geq 30 \). There appears to be little difference between \( t_J^{*M} \) and \( t_J^{*Y} \) in terms of goodness of
approximation to the critical regions. Also, although the type I error associated with \( t_a(\hat{\rho}) \) appears to be slightly less than that associated with \( t(\hat{\rho}) \) due to the fact that \(|t_a(\hat{\rho})|<|t(\hat{\rho})| \) (see Fig. 3), \( t^*_j \) offers a substantial improvement even over \( t_a(\hat{\rho}) \).

The third and fourth columns of Table 6 show the proportion of times \( t^*_j = t_a(\hat{\rho}) \), i.e., the proportion of times \( \hat{\gamma} \) was outside the permitted interval \((0, \infty)\). For a fixed value of \( \rho \), this proportion decreases with \( n \), and for a fixed \( n \), increases with \( \rho \). It might be thought that this latter phenomenon may partially account for the poorer performance of \( t^*_j \) for large values of \( \rho \). Table 7, however, shows the type I error associated with each component of \( t^*_j \) (6.8) for various values of \( n \) and \( \rho \), based on the MLE of \( \rho \). The third column of Table 7 shows the proportion of times \( \hat{\gamma} \leq 0 \), i.e., the proportion of times \( t^*_j = t_a(\hat{\rho}) \). The fourth, sixth, and eighth columns of Table 7 show the estimated type I error of \( t^*_j \), given that \( \hat{\gamma} \leq 0 \). The fifth, seventh, and ninth columns show the estimated type I error of \( t^*_j \), given \( \hat{\gamma} > 0 \) (i.e., given \( t^*_j = t_j \)). It appears that the type I error associated with the \( t_a(\hat{\rho}) \) component of \( t^*_j \) is in fact smaller than that associated with the \( t_j \) component. Hence, the poorer performance of \( t^*_j \) for larger values of \( \rho \) is not due to the larger probability that \( \hat{\gamma} \) cannot be jackknifed (and hence that \( t^*_j = t_a(\hat{\rho}) \)), but rather to the poorer performance of \( t_j \).

7. APPLICATION OF THE JACKKNIFE METHOD TO SUNSPOT DATA

Figure 4 shows the time series of annual average Wölf er sunspot numbers for the years 1770-1869, as reported in Box and Jenkins (1976). Similar data have been analyzed by Schuster (1906), Yule
(1927), McNish and Lincoln (1949), and Moran (1954). Box and Jenkins (1976) tentatively modeled this series as a mean with errors generated by an AR(2) process. For the purposes of the following example, it will be assumed that the series can be modeled as a mean with errors generated by an AR(1) process. The goal is to compute confidence intervals for \( \mu \), the mean number of sunspots per year.

Two types of (1-\( \alpha \))100 percent confidence intervals can be computed: the "usual" one can be written as

\[
\hat{\mu}(\hat{\phi}) + t_{n-1}^{\alpha} \sqrt{\frac{\hat{\sigma}^2(\hat{\phi})}{((1-\hat{\phi})^2 + (n-1)(1-\hat{\phi})^2)}}^{1/2},
\]  

(7.1)

or equivalently as

\[
\hat{\mu}(\hat{\phi}) + t_{n-1}^{\alpha} \sqrt{\frac{\hat{\sigma}^2}{((n-1)(\hat{\sigma}_n^2 + \hat{\gamma}^2)}}^{1/2},
\]

(7.2)

while the "jackknifed" confidence interval is given by

\[
\hat{\mu}(\hat{\phi}) + t_{n-1}^{\alpha} \sqrt{\frac{\hat{\sigma}_n^2}{((n-1)^{1/2})}}.
\]

(7.3)

Figure 5 shows a plot of the estimated random innovations vs. the observed values, and Figure 6 shows the corresponding normal probability (Q-Q) plot. The formulas for computing the estimated random innovations is given in step 3 of Section 5. If the AR(1) model is appropriate, these estimated innovations should be distributed approximately as independent \( N(0, \sigma^2) \) random variables. Figure 5 indicates heterogeneity of variance: the innovations are more variable for larger observed values. Figure 6 indicates a
positively skewed distribution, which is not surprising since the observations are bounded below by zero.

Figure 7 shows estimated innovations vs. observed values for the log-transformed sunspot data (actually, \( \log(x+1) \) was used since a 0 is recorded for the year 1810). Figure 8 shows the Q-Q plot of these innovations. The assumptions of homoscedasticity and normality appear to hold to a better degree for the log-transformed data than the original observations.

Table 8 reports the nominally 95% confidence intervals for \( \mu^* \), the mean of the log-transformed observations, and \( \mu \). The confidence interval for \( \mu \) can be derived by

\[
\mu_L = \exp(\mu_L^*),
\]

\[
\mu_U = \exp(\mu_U^*),
\]

where \( \mu_L^* \) and \( \mu_U^* \) denote the lower and upper confidence bounds on \( \mu^* \), respectively, and \( \mu_L \) and \( \mu_U \) denote the lower and upper confidence bounds on \( \mu \). All confidence intervals are based on \( \hat{\rho}_{HL} \); those based on \( \hat{\rho}_{YW} \) are quite similar. For the smaller samples the confidence intervals for \( \mu \) based upon the jackknife are much wider than the usual confidence intervals.

8. DISCUSSION

For the one-sample location problem, it appears that the modified t-statistic based upon the jackknifed estimate of standard error (section 6) gives the closest approximation to the nominal
\( \alpha \)-level when carrying out the test \( H_0: \mu = c \). For a fixed sample size \( n \), the true \( \alpha \)-level increases as \( p \) increases, but even for \( p = 0.8 \), the approximation to the 10% critical region is satisfactory for \( n > 30 \). Ironically, the other methods described, based upon the saddlepoint approximation and the bootstrap, are much more computationally involved (the integrals in (4.15) must be evaluated numerically), yet perform no better than the usual Wald-statistic. The jackknife \( t \)-statistic using the Yule-Walker estimate of \( p \) is easily computed.

Both the bootstrap method of Section 5 and the jackknife method of Section 6 are easily extended to the case of the general regression model of (1.1) and (1.5) for testing \( H_0: \lambda = c \). The bootstrap method involves the same process of resampling the residuals as described in Section 5, the only difference being that \( \mu \) and \( \mu^* \) are replaced with \( \hat{\theta} \) and \( \hat{\theta}^* \) in steps 1.-5. \( \hat{\theta}_{\mu} \) of (5.1) is replaced with

\[
\hat{\theta}_{\hat{\theta}} = \frac{1}{1/(N-1)} \sum_{i=1}^{N} (A_i - \hat{A})(A_i - \hat{A})'
\]  

It is hoped that the statistic

\[
W_B = [L(\hat{\rho} - c)]'[L(\hat{\theta})^{-1}L']^{-1}[L(\hat{\rho} - c)]
\]

(8.2)

follows a chi-square distribution with \( q \) degrees of freedom more closely than the \( W \)-statistic of (2.6). Based upon the results of this paper, however, it is doubtful that \( W_B \) will perform any better than the usual \( W \)-statistic.
The extension of the jackknife method involves the approximation

\[(LD(\hat{\theta})^{-1}L')^{-1} \approx h(n)\bar{\Gamma}(\theta)\]  \hspace{2cm} (8.3)

where \(\bar{\Gamma}(\theta)\) does not explicitly involve \(n\). The jackknifed estimate of \(\bar{\Gamma}(\theta)\) is computed to produce the modified \(W\)-statistic

\[W_J = h(n)[L\hat{\theta}(\hat{\theta})-c]'\hat{\bar{\Gamma}}[L\hat{\theta}(\hat{\theta})-c].\]  \hspace{2cm} (8.4)

It is hoped that \(W_J\) follows a chi-square distribution with \(q\) degrees of freedom more closely than the \(W\)-statistic of (2.6). In practice, one would use \(W_J^*\) rather than \(W_J\), where

\[W_J^* = W_J \quad \text{if} \quad \hat{\bar{\Gamma}} \text{ is positive definite,} \]  \hspace{2cm} (8.5)
\[h(n)[L\hat{\theta}(\hat{\theta})-c]'\bar{\Gamma}(\hat{\theta})[L\hat{\theta}(\hat{\theta})-c] \quad \text{otherwise,}\]

since \(\hat{\bar{\Gamma}}\) is not guaranteed to be positive definite.

At least two further questions arise from the problems considered in this paper. The first is the question of approximating the power of the tests discussed. For example, can the power of the test of \(H_0: \mu = c\) based on \(t_J^*\) be approximated well by a non-central \(t\)-percentile? The second question concerns how to model the autocorrelation. Suppose that a researcher is willing to model possible autocorrelation in the errors as either an AR(0), AR(1), or AR(2) process. What are the effects on the type I and II error for testing \(H_0: \text{La}=c\) if the researcher first chooses the model of autocorrelation based on the same data that will be used to test \(H_0?\)
(The sunspot data given in Box and Jenkins (1976) was actually identified by them as an AR(2) process, not the AR(1) process assumed for the example of Section 7.) It would seem reasonable to postulate that the test statistic will be affected by how successful the researcher is in choosing the correct model each time.
Appendix A

A TAYLOR SERIES EXPANSION OF $\hat{\mu}(\hat{\rho})$

**Theorem.** Let $\hat{\mu}(\hat{\rho})$ be the statistic defined in (3.1), where $\hat{\rho}$ is the Yule-Walker estimate of $\rho$ defined in (3.6). Assume that $\rho$ is bounded away from 1, i.e., there exists $\epsilon > 0$ such that $\rho < (1-\epsilon)$. Then

$$\text{Var}[\hat{\mu}(\hat{\rho})] = \text{Var}[\hat{\mu}(\rho)] + O(n^{-2}) \quad (A.1)$$

**Proof.** $E[\hat{\mu}(\hat{\rho})] = E[E[\hat{\mu}(\hat{\rho}) | \hat{\rho}]] = E(\hat{\mu}) = \mu,$

(A.2)

since

$$E[\hat{\mu}(\rho)] = \mu, \quad (A.3)$$

for $\rho$ constant, thus

$$\text{Var}[\hat{\mu}(\hat{\rho})] - \text{Var}[\hat{\mu}(\rho)] = E[\hat{\mu}(\rho)^2] - E[\hat{\mu}(\rho)^2]. \quad (A.4)$$

It remains to be shown that

$$E[\hat{\mu}(\hat{\rho})^2] = E[\hat{\mu}(\rho)^2] + O(n^{-2}). \quad (A.5)$$

The outline of the proof will now be given, and the details will follow.
By a Taylor series expansion, we have

\[ \hat{\mu}(\hat{\rho}) = \hat{\mu}(\rho) + \hat{\mu}'(\rho^*)(\hat{\rho} - \rho) \]  

(A.6)

where \(|\rho^* - \rho| < |\rho - \rho|\) and the ' notation denotes the first derivative. Thus,

\[ \hat{\mu}(\hat{\rho})^2 = \hat{\mu}(\rho)^2 + 2\hat{\mu}(\rho)\hat{\mu}'(\rho^*)(\hat{\rho} - \rho) + \hat{\mu}'(\rho^*)^2(\hat{\rho} - \rho)^2. \]  

(A.7)

By Lemma 1,

\[ |E[\hat{\mu}(\rho)\hat{\mu}'(\rho^*)(\hat{\rho} - \rho)]| = O(n^{-2}) \]  

(A.8)

and by Lemma 2,

\[ E[\hat{\mu}'(\rho^*)^2(\hat{\rho} - \rho)^2] = O(n^{-3}). \]  

(A.9)

Q. E. D.

Lemma 1. \[ |E[\hat{\mu}(\rho)\hat{\mu}'(\rho^*)(\hat{\rho} - \rho)]| = O(n^{-2}) \]

Proof. \[ \hat{\mu}(\rho)\hat{\mu}'(\rho^*)(\hat{\rho} - \rho) = (\hat{\mu}(\rho) - \mu)\hat{\mu}'(\rho^*)(\hat{\rho} - \rho) \]  

(A.1.1)

+ \mu\hat{\mu}'(\rho^*)(\hat{\rho} - \rho),

hence, by the triangle inequality
\[ |E(\hat{\varphi}(\rho)\hat{\varpi}(\rho^*)\hat{(\beta-\rho})| \leq |E(\hat{\varphi}(\rho-\mu)\hat{\varpi}(\rho^*)\hat{(\beta-\rho})| + |\mu||E(\hat{\varpi}(\rho^*)\hat{(\beta-\rho})|. \] (A.1.2)

Now

\[ E(\hat{\varpi}(\rho^*)\hat{(\beta-\rho}) = 0 \] (A.1.3)

since \( E(\hat{\varphi}(\beta)) = E(\hat{\varphi}(\rho)) = \mu \) (see A.2, A.3, and A.6). Also,

\[ |E(\hat{\varphi}(\rho-\mu)\hat{\varpi}(\rho^*)\hat{(\beta-\rho})| \leq E(\hat{\varphi}(\rho-\mu)||\hat{\varpi}(\rho^*)||\hat{(\beta-\rho})| \]
\[ \leq \{E((\hat{\varphi}(\rho-\mu))^4E(\hat{\varpi}(\rho^*)^4E(\hat{(\beta-\rho})^4)\}^{1/4} \] (A.1.4)

by the generalization of Hölder's inequality (e.g., Rao 1973, p.55).

Now

\[ E((\hat{\varphi}(\rho-\mu))^4] = 3\{Var(\hat{\varphi}(\rho))\}^2 = O(n^{-2}) \] (A.1.5)

since \( \hat{\varphi}(\rho) \) is normally distributed with mean \( \mu \), and \( Var(\hat{\varphi}(\rho)) = O(n^{-1}) \) by (3.5). Now

\[ \varphi'(\rho) = n[2 + (n-2)(1-\rho)]^{-2}(y_1 + y_n - 2\bar{y}), \] (A.1.6)

so

\[ E(\varphi'(\rho)^*4] = E(n^4[2 + (n-2)(1-\rho^*)]^{-8}(y_1 + y_n - 2\bar{y})^4). \] (A.1.7)
Since $\hat{\rho} \overset{a.s.}{\rightarrow} \rho$ by Lemma 3, let $n$ be large enough such that $|\hat{\rho} - \rho| < (\varepsilon/2)$ with probability 1. Then $|\rho^* - \rho| < |\hat{\rho} - \rho| < (\varepsilon/2)$ and $\rho < (1-\varepsilon)$ implies that $(1-\rho^*) > (\varepsilon/2)$ with probability 1. Hence

$$E[\hat{\rho}'(\rho^*)^4] \leq n^4[2 + (n-2)(\varepsilon/2)]^{-8}E[(y_1 + y_n - 2\bar{y})^4]$$

$$= O(n^{-4}) E[(y_1 + y_n - 2\bar{y})^4].$$

But

$$E[(y_1 + y_n - 2\bar{y})^4] = O(1),$$

so

$$E[\hat{\rho}'(\rho^*)^4] = O(n^{-4}).$$

Finally, Taibah and Kassab (1981) show

$$E[(\hat{\rho} - E(\hat{\rho}))^4] = O(n^{-2}).$$

But

$$E(\hat{\rho}) = \rho + O(n^{-1})$$

(e.g., Marriott and Pope 1954, De Goijer 1981, Taibah and Kassab 1981), so

$$E[(\hat{\rho} - \rho)^4] = O(n^{-2}).$$
since \( E[(\delta-\rho)^2] = O(n^{-1}) \) and \( E[(\delta-\rho)^3] = O(n^{-2}) \) (Taibah and Kassab 1981). Hence,

\[
|E[\hat{\mu}(\rho)\hat{\mu}'(\rho^*)(\hat{\delta}-\rho)]| \leq (O(n^{-2})O(n^{-4})O(n^{-2}))^{1/4} = O(n^{-2}).
\] 

(A.1.14)

Q. E. D.

**Lemma 2.** \( E[\hat{\mu}'(\rho^*)^2(\hat{\delta}-\rho)^2] = O(n^{-3}) \).

**Proof.** By the Cauchy-Schwarz inequality (e.g., Rao 1973, p.54),

\[
E[\hat{\mu}'(\rho^*)^2(\hat{\delta}-\rho)^2] \leq \{E[\hat{\mu}'(\rho^*)^4]E[(\hat{\delta}-\rho)^4]\}^{1/2}.
\] 

(A.2.1)

But \( E[\hat{\mu}'(\rho^*)^4] = O(n^{-4}) \) by (A.1.10), and \( E[(\hat{\delta}-\rho)^4] = O(n^{-2}) \) by (A.1.13). Hence

\[
E[\hat{\mu}'(\rho^*)^2(\hat{\delta}-\rho)^2] \leq (O(n^{-4})O(n^{-2}))^{1/2} = O(n^{-3}).
\]

(A.2.2)

Q. E. D.

**Lemma 3.** \( \hat{\rho}_{YW} \xrightarrow{d.s.} \rho \).

**Proof.** By definition,

\[
\hat{\rho}_{YW} = C_1(\bar{y})/C_0(\bar{y})
\]

(A.3.1)
where \( C_0(\nu) \) and \( C_1(\nu) \) are defined by (3.7). Now

\[
C_0(\bar{y}) \overset{a.s.}{\rightarrow} \sigma(0) = \frac{\sigma^2}{1-\rho^2} \tag{A.3.2}
\]

and

\[
C_1(\bar{y}) \overset{a.s.}{\rightarrow} \sigma(1) = \frac{\rho \sigma^2}{1-\rho^2} \tag{A.3.3}
\]

(see Hannan 1970, Chap. IV). Hence,

\[
\hat{\beta}_{YW} \overset{a.s.}{\rightarrow} \frac{\sigma(1)}{\sigma(0)} = \rho \tag{A.3.4}
\]

Q. E. D.
APPENDIX B: ASYMPTOTIC NORMALITY OF $t^*_j$

Theorem. Let $t^*_j$ be the statistic defined in (6.8). Then

$$t^*_j \overset{D}{\rightarrow} N(0,1) \quad (B.1)$$

Proof. Consider the four estimates of $\gamma$:

$$\hat{\gamma} \overset{P}{\rightarrow} \gamma \quad (B.2)$$

$$\hat{\gamma}_1 \overset{P}{\rightarrow} \gamma \quad (B.3)$$

$$\hat{\gamma}_2 \overset{P}{\rightarrow} \gamma \quad (B.4)$$

hence

$$\hat{\gamma} \overset{P}{\rightarrow} \gamma \quad (B.5)$$

Thus, $t^*_j = t_j$ with probability 1 as $n \rightarrow \infty$, where $t_j$ is defined in (6.7). It remains to be shown that $t_j \overset{D}{\rightarrow} N(0,1)$.

Now

$$n^{1/2}(\hat{\mu}(\hat{\rho}) - \mu) \overset{D}{\rightarrow} N(0, \sigma^2/(1-\rho)^2) \quad (B.6)$$

hence

$$n^{1/2}(\hat{\mu}(\hat{\rho}) - \mu) \gamma \overset{D}{\rightarrow} N(0,1) \quad (B.7)$$
and so

\[ t_j \overset{D}{\to} N(0,1) \quad (B.8) \]

under \( H_0: \mu = c \), by Slutsky's theorem.
REFERENCES


De Gooijer, J.G. (1980). Exact moments of the sample autocovariances from series generated by general ARIMA processes of order \( (p,d,q) \), \( d=0 \) or \( 1 \). J. Econometrics 14, 365-379.


Table 1. Estimated Type I Error for Two t-Statistics

<table>
<thead>
<tr>
<th>ρ</th>
<th>n</th>
<th>P(t ≥ t̂_\text{.01}_n-1)</th>
<th>P(t ≥ t̂_\text{.05}_n-1)</th>
<th>P(t ≥ t̂_\text{.10}_n-1)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>t</td>
<td>t(\hat{β})</td>
<td>t</td>
</tr>
<tr>
<td>0.0</td>
<td>10</td>
<td>.02</td>
<td>.03</td>
<td>.05</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>.01</td>
<td>.02</td>
<td>.05</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>.01</td>
<td>.02</td>
<td>.06</td>
</tr>
<tr>
<td>0.2</td>
<td>10</td>
<td>.02</td>
<td>.04</td>
<td>.10</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>.03</td>
<td>.03</td>
<td>.09</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>.02</td>
<td>.01</td>
<td>.10</td>
</tr>
<tr>
<td>0.4</td>
<td>10</td>
<td>.05</td>
<td>.04</td>
<td>.16</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>.06</td>
<td>.04</td>
<td>.14</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>.05</td>
<td>.03</td>
<td>.14</td>
</tr>
<tr>
<td>0.5</td>
<td>10</td>
<td>.07</td>
<td>.05</td>
<td>.15</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>.09</td>
<td>.04</td>
<td>.18</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>.09</td>
<td>.03</td>
<td>.19</td>
</tr>
<tr>
<td>0.6</td>
<td>10</td>
<td>.09</td>
<td>.06</td>
<td>.19</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>.12</td>
<td>.03</td>
<td>.20</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>.09</td>
<td>.03</td>
<td>.17</td>
</tr>
<tr>
<td>0.8</td>
<td>10</td>
<td>.23</td>
<td>.14</td>
<td>.31</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>.24</td>
<td>.08</td>
<td>.32</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>.22</td>
<td>.06</td>
<td>.29</td>
</tr>
</tbody>
</table>

NOTE: Based on 500 Monte Carlo trials. See text for explanation.
Table 2. The Standard Deviation of Two Different Estimators of Location for $\sigma = 1$

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>$n$</th>
<th>s.d.[$\hat{\theta}(\rho)$]</th>
<th>s.d.[$\hat{\theta}(\hat{\rho})$]</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>10</td>
<td>0.39</td>
<td>0.37</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>0.28</td>
<td>0.28</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>0.23</td>
<td>0.23</td>
</tr>
<tr>
<td>0.5</td>
<td>10</td>
<td>0.58</td>
<td>0.61</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>0.43</td>
<td>0.44</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>0.35</td>
<td>0.38</td>
</tr>
<tr>
<td>0.8</td>
<td>10</td>
<td>1.18</td>
<td>1.27</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>0.94</td>
<td>0.97</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>0.81</td>
<td>0.83</td>
</tr>
</tbody>
</table>

NOTE: s.d.[$\hat{\theta}(\hat{\rho})$] based on 500 Monte Carlo trials. See text for explanation.
Table 3. Estimated Type I Error of Tests Based on Durbin's Density Approximation

<table>
<thead>
<tr>
<th>η</th>
<th>n</th>
<th>P(D &lt; .01) P(D &lt; .05) P(D &lt; .10)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P(D1) P(D2)</td>
<td>P(D1) P(D2)</td>
</tr>
<tr>
<td>-------</td>
<td>-------------</td>
<td>-------------</td>
</tr>
<tr>
<td>0.2</td>
<td>10</td>
<td>.01 .06</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>.01 .03</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>.01 .02</td>
</tr>
<tr>
<td>0.5</td>
<td>10</td>
<td>.00 .07</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>.01 .03</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>.01 .04</td>
</tr>
<tr>
<td>0.8</td>
<td>10</td>
<td>.00 .13</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>.00 .07</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>.00 .06</td>
</tr>
</tbody>
</table>

NOTE: Based on 500 Monte Carlo trials. See text for explanation. 1 refers to the approximation with ν2 and η known, 2 refers to the approximation with these quantities replaced by their estimates.
Table 4. Estimated Type I Error for Two Modified t-Statistics, Based on the Bootstrap Method

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>$n$</th>
<th>$P(t_B \geq t_{0.01}^{BM})$</th>
<th>$P(t_B \geq t_{0.05}^{BM})$</th>
<th>$P(t_B \geq t_{0.10}^{BM})$</th>
<th>$P(t_B \geq t_{0.01}^{BY})$</th>
<th>$P(t_B \geq t_{0.05}^{BY})$</th>
<th>$P(t_B \geq t_{0.10}^{BY})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>10</td>
<td>.02</td>
<td>.07</td>
<td>.13</td>
<td>.03</td>
<td>.10</td>
<td>.15</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>.01</td>
<td>.06</td>
<td>.12</td>
<td>.01</td>
<td>.06</td>
<td>.11</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>.02</td>
<td>.07</td>
<td>.11</td>
<td>.02</td>
<td>.07</td>
<td>.11</td>
</tr>
<tr>
<td>0.2</td>
<td>10</td>
<td>.04</td>
<td>.10</td>
<td>.15</td>
<td>.03</td>
<td>.09</td>
<td>.16</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>.03</td>
<td>.07</td>
<td>.13</td>
<td>.03</td>
<td>.09</td>
<td>.14</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>.01</td>
<td>.07</td>
<td>.12</td>
<td>.01</td>
<td>.07</td>
<td>.12</td>
</tr>
<tr>
<td>0.4</td>
<td>10</td>
<td>.05</td>
<td>.12</td>
<td>.18</td>
<td>.03</td>
<td>.15</td>
<td>.20</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>.03</td>
<td>.09</td>
<td>.14</td>
<td>.03</td>
<td>.11</td>
<td>.14</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>.03</td>
<td>.09</td>
<td>.16</td>
<td>.02</td>
<td>.08</td>
<td>.13</td>
</tr>
<tr>
<td>0.5</td>
<td>10</td>
<td>.07</td>
<td>.15</td>
<td>.19</td>
<td>.07</td>
<td>.13</td>
<td>.19</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>.05</td>
<td>.10</td>
<td>.16</td>
<td>.06</td>
<td>.12</td>
<td>.18</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>.02</td>
<td>.06</td>
<td>.11</td>
<td>.03</td>
<td>.09</td>
<td>.14</td>
</tr>
<tr>
<td>0.6</td>
<td>10</td>
<td>.10</td>
<td>.18</td>
<td>.22</td>
<td>.06</td>
<td>.19</td>
<td>.26</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>.06</td>
<td>.12</td>
<td>.16</td>
<td>.03</td>
<td>.12</td>
<td>.18</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>.03</td>
<td>.09</td>
<td>.15</td>
<td>.05</td>
<td>.10</td>
<td>.18</td>
</tr>
<tr>
<td>0.8</td>
<td>10</td>
<td>.15</td>
<td>.24</td>
<td>.27</td>
<td>.14</td>
<td>.26</td>
<td>.31</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>.08</td>
<td>.14</td>
<td>.20</td>
<td>.08</td>
<td>.21</td>
<td>.25</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>.08</td>
<td>.14</td>
<td>.20</td>
<td>.08</td>
<td>.17</td>
<td>.22</td>
</tr>
</tbody>
</table>

NOTE: Based on 500 Monte Carlo trials. See text for explanation. $N$ refers to the quantity computed based on the ML estimate of $\rho$, $Y$ refers to that computed based on the Yule-Walker estimate.
Table 5. Estimated Type I Error of $t(\hat{p})$ Using the Bootstrap Approximation to its Density

<table>
<thead>
<tr>
<th>$p$</th>
<th>$n$</th>
<th>$\text{P}(t(\hat{p}) \geq t^*_\alpha,01)$</th>
<th>$\text{P}(t(\hat{p}) \geq t^*_\alpha,05)$</th>
<th>$\text{P}(t(\hat{p}) \geq t^*_\alpha,10)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>10</td>
<td>.03</td>
<td>.10</td>
<td>.14</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>.02</td>
<td>.07</td>
<td>.12</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>.02</td>
<td>.06</td>
<td>.07</td>
</tr>
<tr>
<td>0.5</td>
<td>10</td>
<td>.03</td>
<td>.11</td>
<td>.15</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>.02</td>
<td>.08</td>
<td>.14</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>.05</td>
<td>.09</td>
<td>.14</td>
</tr>
<tr>
<td>0.8</td>
<td>10</td>
<td>.08</td>
<td>.16</td>
<td>.20</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>.04</td>
<td>.13</td>
<td>.23</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>.03</td>
<td>.11</td>
<td>.16</td>
</tr>
</tbody>
</table>

NOTE: Based on 100 Monte Carlo trials and N=1000 Bootstraps. See text for explanation.

$t^*_\alpha$ refers to the estimated (1-$\alpha$)100'th percentile of the distribution of $t(\hat{p})$, based upon the bootstrap approximation.
Table 6. Estimated Type I Error for $t_a(\hat{\rho})$ and Two Modified $t$-Statistics Based on the Jackknife Method

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>n</th>
<th>$\hat{\gamma}_M$</th>
<th>$\hat{\gamma}_Y$</th>
<th>$t_a(\hat{\rho})$</th>
<th>$t_{J_M}^*$</th>
<th>$t_{J_Y}^*$</th>
<th>$t_a(\hat{\rho})$</th>
<th>$t_{J_M}^*$</th>
<th>$t_{J_Y}^*$</th>
<th>$t_a(\hat{\rho})$</th>
<th>$t_{J_M}^*$</th>
<th>$t_{J_Y}^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>10</td>
<td>.19</td>
<td>.14</td>
<td>.02</td>
<td>.02</td>
<td>.01</td>
<td>.09</td>
<td>.07</td>
<td>.06</td>
<td>.13</td>
<td>.08</td>
<td>.08</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>.01</td>
<td>.01</td>
<td>.02</td>
<td>.01</td>
<td>.01</td>
<td>.05</td>
<td>.03</td>
<td>.03</td>
<td>.09</td>
<td>.06</td>
<td>.06</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>.00</td>
<td>.00</td>
<td>.02</td>
<td>.01</td>
<td>.01</td>
<td>.05</td>
<td>.04</td>
<td>.04</td>
<td>.10</td>
<td>.09</td>
<td>.09</td>
</tr>
<tr>
<td>0.2</td>
<td>10</td>
<td>.28</td>
<td>.19</td>
<td>.04</td>
<td>.02</td>
<td>.02</td>
<td>.09</td>
<td>.07</td>
<td>.05</td>
<td>.16</td>
<td>.11</td>
<td>.10</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>.03</td>
<td>.02</td>
<td>.02</td>
<td>.02</td>
<td>.02</td>
<td>.07</td>
<td>.05</td>
<td>.05</td>
<td>.11</td>
<td>.08</td>
<td>.08</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.07</td>
<td>.05</td>
<td>.05</td>
<td>.13</td>
<td>.08</td>
<td>.09</td>
</tr>
<tr>
<td>0.4</td>
<td>10</td>
<td>.34</td>
<td>.30</td>
<td>.04</td>
<td>.03</td>
<td>.02</td>
<td>.12</td>
<td>.07</td>
<td>.07</td>
<td>.16</td>
<td>.11</td>
<td>.10</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>.10</td>
<td>.08</td>
<td>.03</td>
<td>.02</td>
<td>.02</td>
<td>.07</td>
<td>.06</td>
<td>.06</td>
<td>.13</td>
<td>.09</td>
<td>.09</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>.03</td>
<td>.03</td>
<td>.03</td>
<td>.02</td>
<td>.01</td>
<td>.08</td>
<td>.05</td>
<td>.05</td>
<td>.13</td>
<td>.09</td>
<td>.09</td>
</tr>
<tr>
<td>0.5</td>
<td>10</td>
<td>.38</td>
<td>.31</td>
<td>.05</td>
<td>.03</td>
<td>.03</td>
<td>.09</td>
<td>.07</td>
<td>.07</td>
<td>.15</td>
<td>.11</td>
<td>.11</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>.14</td>
<td>.12</td>
<td>.04</td>
<td>.03</td>
<td>.04</td>
<td>.09</td>
<td>.06</td>
<td>.07</td>
<td>.14</td>
<td>.10</td>
<td>.10</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>.02</td>
<td>.03</td>
<td>.02</td>
<td>.01</td>
<td>.01</td>
<td>.06</td>
<td>.05</td>
<td>.05</td>
<td>.14</td>
<td>.09</td>
<td>.10</td>
</tr>
<tr>
<td>0.6</td>
<td>10</td>
<td>.46</td>
<td>.39</td>
<td>.05</td>
<td>.03</td>
<td>.04</td>
<td>.11</td>
<td>.08</td>
<td>.08</td>
<td>.16</td>
<td>.12</td>
<td>.12</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>.18</td>
<td>.18</td>
<td>.03</td>
<td>.02</td>
<td>.02</td>
<td>.08</td>
<td>.05</td>
<td>.05</td>
<td>.12</td>
<td>.08</td>
<td>.08</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>.07</td>
<td>.09</td>
<td>.02</td>
<td>.02</td>
<td>.02</td>
<td>.06</td>
<td>.03</td>
<td>.05</td>
<td>.10</td>
<td>.07</td>
<td>.08</td>
</tr>
<tr>
<td>0.8</td>
<td>10</td>
<td>.53</td>
<td>.46</td>
<td>.12</td>
<td>.08</td>
<td>.09</td>
<td>.18</td>
<td>.14</td>
<td>.16</td>
<td>.22</td>
<td>.18</td>
<td>.22</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>.33</td>
<td>.32</td>
<td>.07</td>
<td>.04</td>
<td>.04</td>
<td>.12</td>
<td>.09</td>
<td>.09</td>
<td>.18</td>
<td>.13</td>
<td>.15</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>.22</td>
<td>.23</td>
<td>.04</td>
<td>.02</td>
<td>.02</td>
<td>.10</td>
<td>.06</td>
<td>.07</td>
<td>.16</td>
<td>.09</td>
<td>.11</td>
</tr>
</tbody>
</table>

NOTE: Based on 500 Monte Carlo trials. See text for explanation. See Table 4 for explanation of $M$ and $Y$ notation.
Table 7. Estimated Type I Error Associated with Each Component of $t_J$

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>$n$</th>
<th>$P(\hat{\gamma} \leq 0)$</th>
<th>$P(t \geq t_{n-1}^{.01})$</th>
<th>$P(t \geq t_{n-1}^{.05})$</th>
<th>$P(t \geq t_{n-1}^{.10})$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$t_a(\hat{\rho})$</td>
<td>$t_J$</td>
<td>$t_a(\hat{\rho})$</td>
<td>$t_J$</td>
</tr>
<tr>
<td>0.2</td>
<td>10</td>
<td>.24</td>
<td>.01</td>
<td>.06</td>
<td>.09</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>.03</td>
<td>.05</td>
<td>.07</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>.04</td>
<td>0</td>
<td>0</td>
<td>.11</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>.01</td>
<td>.05</td>
<td>.08</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>.00</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>.01</td>
<td>.05</td>
<td>.08</td>
</tr>
<tr>
<td>0.5</td>
<td>10</td>
<td>.42</td>
<td>.04</td>
<td>.08</td>
<td>.14</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>.06</td>
<td>.12</td>
<td>.16</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>.13</td>
<td>.02</td>
<td>.03</td>
<td>.05</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>.02</td>
<td>.05</td>
<td>.09</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>.03</td>
<td>0</td>
<td>.07</td>
<td>.07</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>.03</td>
<td>.12</td>
<td>.12</td>
</tr>
<tr>
<td>0.8</td>
<td>10</td>
<td>.52</td>
<td>.05</td>
<td>.12</td>
<td>.17</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>.12</td>
<td>.17</td>
<td>.20</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>.34</td>
<td>.02</td>
<td>.05</td>
<td>.09</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>.06</td>
<td>.12</td>
<td>.14</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>.21</td>
<td>.03</td>
<td>.06</td>
<td>.07</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>.06</td>
<td>.09</td>
<td>.11</td>
</tr>
</tbody>
</table>

NOTE: Based on 500 Monte Carlo trials. See text for explanation.
Table 8. Confidence Intervals for $\mu^*$ and $\mu$ Based on Log-Transformed Sunspot Series

<table>
<thead>
<tr>
<th>Years</th>
<th>n</th>
<th>$\hat{\mu}(\hat{\phi})$</th>
<th>$\hat{\phi}$</th>
<th>$\hat{\gamma}$</th>
<th>$\hat{\gamma}$</th>
<th>$\mu^*$ _Usual</th>
<th>$\mu^*$ _Jackknifed</th>
<th>$\mu$ _Usual</th>
<th>$\mu$ _Jackknifed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1784-1793</td>
<td>10</td>
<td>3.78</td>
<td>0.26</td>
<td>0.81</td>
<td>0.35</td>
<td>0.16</td>
<td>2.35</td>
<td>5.20</td>
<td>-0.82</td>
</tr>
<tr>
<td>1784-1803</td>
<td>20</td>
<td>3.34</td>
<td>0.26</td>
<td>0.84</td>
<td>0.30</td>
<td>0.18</td>
<td>2.11</td>
<td>4.57</td>
<td>0.67</td>
</tr>
<tr>
<td>1784-1813</td>
<td>30</td>
<td>2.98</td>
<td>0.36</td>
<td>0.86</td>
<td>0.23</td>
<td>0.22</td>
<td>1.65</td>
<td>4.31</td>
<td>1.25</td>
</tr>
<tr>
<td>1784-1823</td>
<td>40</td>
<td>2.93</td>
<td>0.33</td>
<td>0.88</td>
<td>0.21</td>
<td>0.15</td>
<td>1.63</td>
<td>4.23</td>
<td>0.74</td>
</tr>
<tr>
<td>1784-1833</td>
<td>50</td>
<td>3.04</td>
<td>0.33</td>
<td>0.85</td>
<td>0.26</td>
<td>0.26</td>
<td>2.05</td>
<td>4.02</td>
<td>1.91</td>
</tr>
<tr>
<td>1770-1869</td>
<td>100</td>
<td>3.51</td>
<td>0.37</td>
<td>0.82</td>
<td>0.30</td>
<td>0.28</td>
<td>2.88</td>
<td>4.13</td>
<td>2.80</td>
</tr>
</tbody>
</table>
Figure 1. Bias of \( \hat{\rho} \) as a function of \( \rho \) for \( n=20 \).
Figure 2. Bias of $\hat{\rho}_{ML}$ as a function of $\rho$. 
Figure 3. The Ratio $\frac{\hat{t}_a(\rho)}{\hat{t}(\rho)}$ as a Function of Sample Size.
Figure 4. Annual Wölfher Sunspot Numbers for the Years 1770-1869.
Figure 5. Estimated Innovations vs. Observed Values for the Sunspot Series.
Figure 6. Q-Q Plot of the Estimated Innovations for the Sunspot Series.
Figure 7. Estimated Innovations vs. Observed Values for the Log-Transformed Sunspot Series.
Figure 8. Q-Q Plot of Estimated Innovations for the Log-Transformed Sunspot Series.