The Annals of Statistics
1983, Vol. 11, No. 1, 326–329

CONSISTENCY OF THE LEAST SQUARES ESTIMATOR OF THE FIRST ORDER MOVING AVERAGE PARAMETER

By Brian D. Macpherson and Wayne A. Fuller

University of Manitoba and Iowa State University

The consistency of the least squares estimator of the parameter of the first order moving average time series is proven for the parameter in the interval $[-1, 1]$.

1. Introduction. The first order moving average process is given by

$$Y_t = e_t + \beta e_{t-1}, \tag{1.1}$$

where the e_t are independent, identically distributed random variables with mean zero, variance σ^2, and $E(e_t^4) = \nu\sigma^4$, $\nu < \infty$.

The model (1.1) is said to be invertible if $|\beta| < 1$. When $|\beta| \geq 1$, model (1.1) is noninvertible. The autocorrelation function at lag one for the first order moving average model (1.1) is given by

$$\rho(1) = (1 + \beta^2)^{-1}\beta, \tag{1.2}$$

and it can be shown that $\rho(1) \in [-0.5, 0.5]$. If β_0 is a solution to (1.2) and $\beta_0 \neq 0$, then β_0^{-1} is also a solution to (1.2). Therefore, for every noninvertible model with $|\beta| > 1$, there is an equivalent invertible model. Most work with model (1.1) has restricted the parameter space to the open interval $(-1, 1)$ so that the model is strictly invertible.

The estimation of the parameter β has been the subject of a substantial amount of work. Whittle (1951, 1953), Durbin (1959), Walker (1961), Box and Jenkins (1970), and Fuller (1976) provide estimation techniques and properties of some estimators. All authors cited have restricted the parameter space to be the open interval $(-1, 1)$ and all of the estimators and their properties have been determined under this constraint.

The noninvertible situation with $\beta = \pm 1$ was considered by Plosser and Schwert (1977). They note the difficulties inherent in this situation, particularly with respect to the estimation of β using approximate likelihood procedures. Plosser and Schwert provide several situations in which it is reasonable to consider the model (1.1) with $\beta = \pm 1$, and report on an extensive Monte Carlo experiment using such a model.

2. Consistency of the least squares estimator. We consider the first order moving average model with the process initiated at time one with $e_0 = 0$. Given a sample of n observations, $Y_1, Y_2, \cdots, Y_n$, we write, for model (1.1),

$$e_0(Y; \beta) = 0, \quad e_t(Y; \beta) = Y_t - \beta e_{t-1}(Y; \beta), \quad t = 1, 2, \cdots, n.$$

where the notation $e_t(Y; \beta)$ is used to emphasize the fact that the e_t depend on the observations, Y_t, and on β.

The least squares estimator of β is that value of $\theta \in [-1, 1]$ that minimizes

$$Q_n(\theta) = \textstyle\sum_{t=1}^n \{e_t(Y; \theta)\}^2. \tag{2.1}$$

Let $W_t(Y; \theta)$ denote the negative of the partial derivative of $e_t(Y; \beta)$ with respect to β evaluated at $\beta = \theta$. Then

Received November 1981; revised September 1982.

AMS 1970 *subject classification.* Primary 62M10; secondary 62F10.

Key words and phrases. Moving average process, consistent estimation, least squares estimator, noninvertible case.

$$\frac{\partial}{\partial \theta} Q_n(\theta) = -2ng_n(\theta),$$

where

(2.2) $$g_n(\theta) = n^{-1} \sum_{t=1}^{n} e_t(Y; \theta) W_t(Y; \theta).$$

If $Q_n(\theta)$ has a minimum in the interval $(-1, 1)$, then $g_n(\theta) = 0$ at that minimum.

The function $e_t(Y; \theta)$ satisfies the difference equation

(2.3) $$e_t(Y; \theta) + \theta e_{t-1}(Y; \theta) = Y_t$$

and the function $W_t(Y; \theta)$ satisfies the difference equation

(2.4) $$W_t(Y; \theta) + 2\theta W_{t-1}(Y; \theta) + \theta^2 W_{t-2}(Y; \theta) = Y_{t-1}.$$

It follows that

(2.5) $$e_t(Y; \theta) = \sum_{i=0}^{t-1} a_i e_{t-i},$$

(2.6) $$W_t(Y; \theta) = \sum_{i=0}^{t-2} b_i e_{t-1-i},$$

where $a_0 = b_0 = 1$, $a_i = (\beta - \theta)(-\theta)^{i-1}$, $i = 1, 2, \cdots$, $b_i = (-\theta)^{i-1}\{(\beta - \theta)i - \theta\}$, $i = 1, 2, \cdots$.

We now prove two lemmas.

Lemma 1. *Let model* (1.1) *hold with* $\beta = 1$ *and let* $g_n(\theta)$ *be defined by* (2.2). *Then* $E\{g_n(\theta)\} > 0$ *for all* $\theta \in [-1, 1)$.

Proof. From (2.5) and (2.6) we have

$$E\{ng_n(\theta)\} = \sum_{t=2}^{n} E(\sum_{i=0}^{t-1} \sum_{j=0}^{t-2} a_i b_j e_{t-i} e_{t-1-j}).$$

Because the e_t are independent, identically distributed random variables with mean zero and variance σ^2,

$$E\{W_t(Y; \theta) e_t(Y; \theta)\} = \sigma^2\{(1 - \theta) + (1 - \theta)^2 \sum_{j=1}^{t-2} j(-\theta)^{2j-1} - \theta(1 - \theta) \sum_{j=1}^{t-2} (-\theta)^{2j-1}\}$$

for $t = 2, 3, \cdots$. We note that this expression is equal to zero when $\theta = 1$, and is positive for $\theta \in [-1, 0]$. For $\theta \in (0, 1)$, we obtain

(2.7) $$\begin{aligned} E\{W_t(Y; \theta) e_t(Y; \theta)\} = \sigma^2[&(1 - \theta)^2(1 - \theta^2)^{-2}\{1 + \theta(1 - \theta - \theta^2)(-\theta)^{2t-4}\} \\ &- (1 - \theta^2)^{-1}\{(t - 2)(1 - \theta)^2(-\theta)^{2t-3}\}]. \end{aligned}$$

The last term in (2.7) is positive for all $\theta \in (0, 1)$ because the exponent $2t - 3$ is always odd. The function $\theta(1 - \theta - \theta^2)$ defined on $[-1, 1]$ has a minimum value of -1 at $\theta = 1$. Therefore, $E\{g_n(\theta)\} > 0$ for all $\theta \in [-1, 1)$.

Lemma 2. *Let* $g_n(\theta)$ *be defined by* (2.2), *let* $g_n'(\theta)$ *denote the derivative of* $g_n(\beta)$ *with respect to* β *evaluated at* $\beta = \theta$. *Then*

$$|g_n'(\theta)| \le 3(1 - |\theta|)^{-4}\hat{\gamma}_Y$$

for $\theta \in (-1, 1)$ *and all* n.

Proof. We have

$$-2ng_n'(\theta) = 2 \sum_{t=1}^{n} \{e_t'(Y; \theta)\}^2 + 2 \sum_{t=1}^{n} e_t(Y; \theta) e_t''(Y; \theta),$$

where

$$e_t(Y; \theta) = \sum_{j=0}^{t-1} (-\theta)^j Y_{t-j} = \sum_{j=0}^{t-1} a_j^* Y_{t-j}.$$

$$e_t'(Y; \theta) = \sum_{j=1}^{t-1} (-j)(-\theta)^{j-1} Y_{t-j} = \sum_{j=1}^{t-1} b_j^* Y_{t-j},$$

and

$$e_t''(Y;\theta) = \sum_{j=2}^{t-1} j(j-1)(-\theta)^{j-2}Y_{t-j} = \sum_{j=2}^{t-1} c_j^* Y_{t-j}.$$

Using the convolution inequality,

$$\sum_{t=1}^{n} \left(\sum_{i=1}^{t-1} a_{t-i}b_i\right)^2 \le \left(\sum_{t=1}^{n} |b_t|\right)^2 \sum_{t=1}^{n} a_t^2,$$

we obtain

$$\sum_{t=1}^{n} \{e_t'(Y;\theta)\}^2 = \sum_{t=1}^{n} \left(\sum_{j=1}^{t-1} b_j^* Y_{t-j}\right)^2 \le \left(\sum_{t=1}^{n} Y_t^2\right)\left(\sum_{j=1}^{n} |b_j^*|\right)^2.$$

Using the Cauchy-Schwarz Inequality, and the convolution inequality, we have

$$\left\{\sum_{t=1}^{n} e_t(Y;\theta)e_t''(Y;\theta)\right\}^2 \le \left(\sum_{t=1}^{n} Y_t^2\right)^2\left(\sum_{j=0}^{n} |a_j^*|\right)^2\left(\sum_{j=2}^{n} |c_j^*|\right)^2.$$

Therefore,

$$|g_n'(\theta)| \le \left\{\left(\sum_{j=1}^{n} |b_j^*|\right)^2 + \sum_{j=0}^{n} |a_j^*| \sum_{j=2}^{n} |c_j^*|\right\}\hat{\gamma}_y \le 3(1-|\theta|)^{-4}\hat{\gamma}_Y. \square \tag{2.8}$$

We now prove the following theorem.

THEOREM. *Let Y_t satisfy the model*

$$Y_t = e_t + \beta e_{t-1},\ t = 1, 2, \cdots,$$

where $e_0 = 0$ and the e_t, $t = 1, 2, \cdots$, are i.i.d. *random variables with mean zero, variance σ^2, and $E(e_t^4) = \nu\sigma^4$. Let $\beta \in [-1, 1]$. Then the least squares estimator $\hat{\beta}$ converges to β in probability as $n \to \infty$.*

PROOF. Given $\beta \in (-1, 1)$ and $\delta_1 > 0$, let δ be the minimum of δ_1, ½ $(1-\beta)$ and ½ $(1+\beta)$. Then $e_t(Y;\theta)$ defined in (2.3) is converging to a stationary autoregressive moving average (1, 1) with autoregressive parameter equal to $-\theta$ and moving average parameter equal to β as $t \to \infty$. Therefore, for $\theta \in [-1+\delta, 1-\delta]$,

$$n^{-1}Q_n(\theta) \to (1-\theta^2)^{-1}(1+\beta^2-2\beta\theta)\sigma^2 \tag{2.9}$$

in mean square. For example, see Fuller (1976, page 68 and page 237). The function on the right of (2.9) is continuous and has continuous first and second derivatives on $[-1+\delta, 1-\delta]$. Furthermore, the variance of $n^{-1}Q_n(\theta)$ is bounded on $[-1+\delta, 1-\delta]$. It follows that $\hat{\beta}$ is consistent for $\beta \in [-1+\delta, 1-\delta]$.

We now consider $\beta = 1$ and first demonstrate that for $\beta = 1$

$$\lim_{n\to\infty} P\{Q_n(-1) > Q_n(1)\} = 1. \tag{2.10}$$

We have

$$e_t(Y;-1) = \begin{cases} e_t + 2\sum_{j=1}^{t-1} e_{t-j}, & t = 2, 3, \cdots, \\ e_1 & t = 1, \end{cases}$$

and letting $Z_{t-1} = 2\sum_{j=1}^{t-1} e_{t-j}$,

$$\sum_{t=1}^{n} e_t^2(Y;-1) - \sum_{t=1}^{n} e_t^2(Y;1) = \sum_{t=1}^{n} Z_{t-1}^2 + 2\sum_{t=1}^{n} Z_{t-1}e_t.$$

By the results of Dickey and Fuller (1979)

$$\left(\sum_{t=1}^{n} Z_{t-1}^2\right)^{-1} \sum_{t=1}^{n} Z_{t-1}e_t = O_p(n^{-1})$$

and $n^{-2}\sum_{t=1}^{n} Z_{t-1}^2$ converges in distribution to a linear combination of Chi squared random variables. Conclusion (2.10) then follows.

We now consider the function $g_n(\theta)$ given in (2.2) and demonstrate that, for given $\delta >$

0, the probability that $g_n(\theta)$ is positive on the interval $[-1 + \delta, 1 - \delta]$ approaches one as $n \to \infty$. The function $g_n(\theta)$ is a continuous differentiable function of θ. By Lemma 1, there is a K such that $E\{g_n(\theta)\} > K$ for $\theta \in [-1 + \delta, 1 - \delta]$. The variance of $g_n(\theta)$ is $O(n^{-1})$, (for example, see Fuller, 1976, page 237) and, by Chebyshev's inequality, $P\{g_n(\theta) > K\} \to 1$ as $n \to \infty$ for any $\theta \in [-1 + \delta, 1 - \delta]$. By Lemma 2 and because $\text{Var}\{\hat{\gamma}_Y(0)\} = O(n^{-1})$, it follows that given $\varepsilon > 0$, there exists an N_1, and $M < \infty$ such that

$$P\{|g_n'(\theta)| < M, \quad \text{for all} \quad \theta \in [-1 + \delta, 1 - \delta]\} \geq 1 - \tfrac{1}{2}\varepsilon$$

for all $n > N_1$. Hence, there exists an η, $0 < \eta < (2M)^{-1}K$ such that for all $n > N_1$ and $(\theta_1, \theta_2) \in [-1 + \delta, 1 - \delta]$,

$$P\{\sup_{|\theta_1 - \theta_2| \leq \eta}[|g_n(\theta_1) - g_n(\theta_2)| < \tfrac{1}{2} K]\} \geq 1 - \tfrac{1}{2}\varepsilon.$$

We now subdivide the interval $[-1 + \delta, 1 - \delta]$ using the points $d_i = -1 + \delta + i\eta$, $i = 0, 1, \cdots, A$, and $d_{A+1} = 1 - \delta$, where A is the largest integer such that $2(1 - \delta) - \eta A > 0$. Given $\varepsilon > 0$, there exists an N_2 such that for all $n > N_2$,

$$P\{g_n(d_i) > K\} \geq 1 - [2(A + 2)]^{-1}\varepsilon, \quad i = 0, 1, \cdots, A + 1$$

and hence,

$$P\{g_n(\theta) > K;\, i = 0, 1, \cdots, A + 1\} \geq 1 - \tfrac{1}{2}\varepsilon.$$

Let $N = \max(N_1, N_2)$, then for all $n > N$, it follows that

$$P\{g_n(\theta) > 0 \quad \text{for all} \quad \theta \in [-1 + \delta, 1 - \delta]\} \geq 1 - \varepsilon.$$

Thus, with probability greater than or equal to $1 - \varepsilon$, the derivative of $Q_n(\theta)$ is negative and $Q_n(\theta)$ is a decreasing function on the interval $[-1 + \delta, 1 - \delta]$. Therefore, with probability greater than or equal to $1 - \varepsilon$, $Q_n(\theta)$ achieves its minimum value for $|\theta| \geq 1 - \delta$. Combining this result with (2.10), we obtain the conclusion that $\hat{\beta}$ is a consistent estimator of β for $\beta = 1$. The argument for $\beta = -1$ is completely analogous to that for $\beta = 1$ and is omitted.

Acknowledgments. We wish to express our thanks to Dr. David Hasza, Dr. Sastry Pantula and a referee for helpful suggestions.

REFERENCES

BOX, G. E. P., and JENKINS, G. M. (1970). *Time Series Analysis, Forecasting and Control.* Holden-Day, San Francisco.

DICKEY, D. A. and FULLER, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *J. Amer. Statist. Assoc.* **74** 427–431.

DURBIN, J. (1959). Efficient estimation of parameters in moving average models. *Biometrika* **46** 306–316.

FULLER, W. A. (1976). *Introduction to Statistical Time Series.* Wiley, New York.

PLOSSER, CHARLES I. and SCHWERT, G. WILLIAM. (1977). Estimation of a noninvertible moving average process: the case of overdifferencing. *J. Econometrics* **6** 199–224.

WALKER, A. M. (1961). Large sample estimation of parameters for moving average models. *Biometrika* **48** 343–357.

WHITTLE, P. (1951). *Hypothesis Testing in Time Series Analysis.* Hafner, New York.

WHITTLE, P. (1953). Estimation and information in stationary time series. *Ark. Mat.* **2** 423–434.

DEPARTMENT OF STATISTICS
UNIVERSITY OF MANITOBA, WINNIPEG
MANITOBA, CANADA R3T 2N2

DEPARTMENT OF STATISTICS
IOWA STATE UNIVERSITY
AMES, IOWA 50011