The Annals of Statistics
1977, Vol. 5, No. 4, 700–708

ROBUST INTERVAL ESTIMATION OF THE INNOVATION VARIANCE OF AN ARMA MODEL

BY WILLIAM W. DAVIS

Carnegie-Mellon University

For the autoregressive-moving average time series model, the normal theory procedure for setting confidence intervals for the error variance is not robust against nonnormality. This paper proposes three asymptotically robust techniques: they are a "standard-error" procedure, an analog of Box's simple data splitting technique, and the jackknife procedure. The large sample distribution of each of these techniques is derived.

1. Introduction. A discrete time-series $\{x_t\}$ is called an ARMA (p, q) [autoregressive-moving average] process if

$$\varphi(B)(x_t - \mu) = \theta(B)a_t \tag{1.1}$$

where B is the backshift operator which satisfies $Bx_t = x_{t-1}$, $\varphi(B) = 1 - \sum_1^p \varphi_i B^i$ and $\theta(B) = 1 - \sum_1^q \theta_i B^i$. Models of this form were first used by Yule (1927) and Wold (1938). Throughout this paper it is assumed that $\{a_t\}$ is an i.i.d. sequence with mean 0, variance σ^2, and finite kurtosis γ_2. Also the roots of $\varphi(B) = 0$ and $\theta(B) = 0$ are assumed to be outside the unit circle in the complex plane. Thus if $\boldsymbol{\varphi} = (\varphi_1, \cdots, \varphi_p)$ and $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_q)$, then

$$(\boldsymbol{\varphi}, \boldsymbol{\theta}) \in \Omega = \{(\boldsymbol{\varphi}, \boldsymbol{\theta}) : \text{all roots of } \varphi(B) = 0 \text{ and } \theta(B) = 0 \text{ have norm} > 1\}.$$

The series x_t could be obtained by some transformation (e.g., a difference) of the actual measurements. It is assumed throughout that although x_t satisfies (1.1) for all t it is observed only for $1 \leqq t \leqq T$.

The point and interval estimates for $(\boldsymbol{\varphi}, \boldsymbol{\theta})$ derived under the assumption of normality of the errors can be shown to be asymptotically robust with respect to deviations from this assumption. That this robustness of normal theory procedures does not extend to interval estimation of σ^2 might be expected from knowledge of the i.i.d. case (i.e., $p = q = 0$). As pointed out by Box (1953, page 330), for inferences about σ^2 the fourth moment must be studentized just as the second moment is studentized when making inferences about means.

Three well-known methods of studentizing the kurtosis in the i.i.d. case are shown to be valid in the more general ARMA time-series model. These procedures are a standard error procedure which uses a moment estimator of the kurtosis, an analogue of Box's (1953) simple data split, and the jackknife technique, which was shown to be asymptotically robust in the i.i.d. case by Miller

Received May 1975; revised November 1976.
AMS 1970 *subject classifications.* Primary 62M10; Secondary 62G35.
Key words and phrases. Autoregressive-moving average, discrete time series, variance, data splitting, jackknife, robust inference.

(1968). Analogs of these three techniques were studied by Layard (1972) in the i.i.d. case for testing equality of covariance matrices.

Although the jackknife was originally introduced by Quenouille (1949) in a time-series problem to reduce bias, no application of Tukey's (1958) confidence interval use of the jackknife has been given previously in time-series. The only other application in stochastic processes of interval estimation using the jackknife was by Gray, Watkins and Adams (1972), who assumed a stationary stochastic process with independent increments.

The residuals $\hat{a}_t$ are defined to satisfy (1.1) with estimated values $\hat{\boldsymbol{\beta}} = (\hat{\boldsymbol{\varphi}}, \hat{\boldsymbol{\theta}}, \hat{\mu})$ replacing parameters $\boldsymbol{\beta} = (\boldsymbol{\varphi}, \boldsymbol{\theta}, \mu)$. The method of proof is to show that the residuals are approximately i.i.d. random variables. It is assumed throughout that

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = O_p(T^{-\frac{1}{2}}) \tag{1.2}$$

where Mann and Wald's (1943) notation is used. Most reasonable methods of estimation satisfy (1.2). For one particular method see Hannan (1969).

The correspondence between residuals and errors is derived in Section 2 with some messy algebra deferred to Section 4. Section 3 verifies that the three techniques are asymptotically robust.

2. Residual approximation. Various methods of calculating the residuals for the general ARMA model are discussed by Box and Jenkins (1970, Chapter 7). Here we assume that the residuals are calculated recursively, using estimated values for parameters in (1.1), by setting starting values equal to 0. Thus for arbitrary $\hat{\boldsymbol{\beta}}$ one can calculate the residuals $\hat{a}_t = \hat{a}_t(\hat{\boldsymbol{\beta}})$ recursively from the equations

$$\begin{aligned} \hat{a}_t &= 0 && 1 - q \leqq t \leqq 0 \\ &= \textstyle\sum_1^q \hat{\theta}_i \hat{a}_{t-i} + \hat{x}_t - \sum_1^p \hat{\varphi}_i \hat{x}_{t-i} && 1 \leqq t \leqq T \end{aligned}$$

where

$$\begin{aligned} \hat{x}_t &= 0 && 1 - p \leqq t \leqq 0 \\ &= x_t - \hat{\mu} && 1 \leqq t \leqq T\,. \end{aligned}$$

The residuals are most easily calculated as above, but it is mathematically more convenient to express them as

$$\hat{a}_t = \textstyle\sum_{i=0}^{t-1} \hat{\pi}_i^{(0)} \hat{x}_{t-i} \qquad 1 \leqq t \leqq T$$

where $\pi^{(0)}(B) = \varphi(B)\theta^{-1}(B) = \sum_{i=0}^{\infty} \pi_i^{(0)} B^i$ and $\hat{\pi}^{(0)}(B)$ is the analogous quantity replacing $\boldsymbol{\beta}$ with $\hat{\boldsymbol{\beta}}$. Throughout this work we restrict attention to $\hat{\boldsymbol{\beta}}$ such that its first $p + q$ elements $(\hat{\boldsymbol{\varphi}}, \hat{\boldsymbol{\theta}}) \in \Omega$.

Using a Taylor expansion of $\hat{a}_t(\hat{\boldsymbol{\beta}})$ about the true values $\boldsymbol{\beta}$ (see Box and Jenkins (1970), page 237), one has

$$\hat{a}_t = \textstyle\sum_{i=0}^{t-1} \pi_i^{(0)} (x_{t-i} - \mu) + \hat{s}_t$$

where

$$\hat{s}_t = \sum_1^{p+q+1} (\hat{\beta}_i - \beta_i) \frac{\partial \hat{a}_t}{\partial \hat{\beta}_i}\bigg|_{\tilde{\boldsymbol{\beta}}} \tag{2.1}$$

and $\tilde{\beta}$ is on the line between $\hat{\beta}$ and β. From (1.1) $a_t = \sum_0^\infty \pi_i^{(0)}(x_{t-i} - \mu)$, thus one has $\hat{a}_t = a_t + \hat{r}_t$ where $\hat{r}_t = \hat{s}_t - w_t$ and $w_t = \sum_t^\infty \pi_i^{(0)}(x_{t-i} - \mu)$. So $\hat{\mathbf{a}} = \mathbf{a} + \hat{\mathbf{r}}$ where $\mathbf{a} = (a_1, \cdots, a_T)$, $\hat{\mathbf{a}} = (\hat{a}_1, \cdots, \hat{a}_T)$, and $\hat{\mathbf{r}} = (\hat{r}_1, \cdots, \hat{r}_T)$.

The following calculations show that the residuals $\{\hat{a}_t\}$ are approximately equal to the error $\{a_t\}$. The proofs of

$$\sum_1^T a_t \hat{r}_t = O_p(1) \quad \text{and} \tag{2.2}$$

$$\sum_1^T \hat{r}_t^n = O_p(1) \quad \text{for} \quad n = 2 \quad \text{and} \quad 4 \tag{2.3}$$

are deferred to Section 4. From (2.3) one has

$$\max_{1 \leq t \leq T} |\hat{r}_t| = O_p(1)\,. \tag{2.4}$$

Since

$$\sum_{t=1}^T (\hat{a}_t^2 - a_t^2) = \sum_{t=1}^T ((a_t + \hat{r}_t)^2 - a_t^2) = 2\sum_{t=1}^T a_t \hat{r}_t + \sum_{t=1}^T \hat{r}_t^2\,,$$

by (2.2) and (2.3) one has

$$\sum_1^T (\hat{a}_t^2 - a_t^2) = O_p(1)\,. \tag{2.5}$$

To show

$$\sum_1^T (\hat{a}_t^4 - a_t^4) = O_p(T^{\frac{1}{2}}) \tag{2.6}$$

consider

$$|\textstyle\sum_t (\hat{a}_t^4 - a_t^4)| \leq \max_t (a_t^2 + \hat{a}_t^2) |\sum_t (a_t^2 - \hat{a}_t^2)|\,.$$

So by (2.5) it suffices to show that $\max_{1 \leq t \leq T} (a_t^2 + \hat{a}_t^2) = O_p(T^{\frac{1}{2}})$. By the Cauchy-Schwarz inequality

$$\max_t (a_t^2 + \hat{a}_t^2) \leq \max_t (3a_t^2 + 2\hat{r}_t^2)\,,$$

so by (2.4) it suffices to show

$$\max_{1 \leq t \leq T} a_t^2 = O_p(T^{\frac{1}{2}})\,.$$

Using Chebyshev's inequality, one has

$$P(\max_{1 \leq t \leq T} a_t^2 > NT^{\frac{1}{2}}) \leq \sum_1^T P(a_t^2 > NT^{\frac{1}{2}}) \leq \sum_1^T Ea_t^4/N^2T = \sigma^4(\gamma_2 + 3)/N^2 < \varepsilon$$

for $N = N(\sigma, \gamma_2, \varepsilon)$ chosen sufficiently large, which completes the proof of (2.6).

Using these results Section 3 justifies the claims of Section 1 concerning robust interval estimation of σ^2.

3. Robust interval estimation for σ^2.

3.1. *Standard-error procedure.* This section shows that if confidence intervals for σ^2 are based on the asymptotic distribution of $\hat{\sigma}^2 = \sum_1^T \hat{a}_t^2/T$ the intervals have the same susceptibility to nonnormality as in the i.i.d. case. Furthermore, by using simple moment estimates, one can studentize the kurtosis creating robust interval estimates.

Using (2.5) one has

$$\begin{aligned} T^{\frac{1}{2}}(\hat{\sigma}^2 - \sigma^2) &= \sum_1^T (a_t^2 - \sigma^2)/T^{\frac{1}{2}} + \sum_1^T (\hat{a}_t^2 - a_t^2)/T^{\frac{1}{2}} \\ &\to_{\mathscr{L}} \mathscr{N}(0, \sigma^4(2 + \gamma_2)) \end{aligned} \tag{3.1}$$

by the central limit theorem and Slutsky's theorem (Rao (1973), 2c.4 (x) a, b). The standard dependence of the population kurtosis is illustrated by (3.1). The kurtosis can be studentized as follows:

$$\sum_1^T \hat{a}_t^4/T = \sum_1^T a_t^4/T + \sum_1^T (\hat{a}_t^4 - a_t^4)/T \to_p Ea^4$$

by (2.6), so that

$$\sum_1^T (\hat{a}_t^2 - \hat{\sigma}^2)^2/T = \sum_1^T \hat{a}_t^4/T - (\hat{\sigma}^2)^2 \to_p \sigma^4(2 + \gamma_2)\,. \tag{3.2}$$

Thus by Slutsky's theorem (Rao (1973), 2c.4 (x) b), using (3.1) and (3.2) one has

$$T^{\frac{1}{2}}(\hat{\sigma}^2 - \sigma^2)/(\sum_1^T (\hat{a}_t^2 - \hat{\sigma}^2)^2/(T-1))^{\frac{1}{2}} \to_{\mathscr{L}} \mathscr{N}(0, 1)\,. \tag{3.3}$$

In the i.i.d. case the variance stabilizing log transformation of the sample variance has been shown to have a more nearly normal distribution for moderate samples. Using a well-known theorem (see, for example, Rao (1973) 6a.2 (i)) one has

$$T^{\frac{1}{2}}(\ln \hat{\sigma}^2 - \ln \sigma^2) \to_{\mathscr{L}} \mathscr{N}(0, 2 + \gamma_2)\,.$$

A moment estimator of the asymptotic variance of $\ln \hat{\sigma}^2$ is $2 + \hat{\gamma}_2 = (\sum_1^T \hat{a}_t^4/T(\hat{\sigma}^2)^2) - 1$. By (2.5) and (2.6) one obtains $2 + \hat{\gamma}_2 \to_p 2 + \gamma_2$, so that by Slutsky's theorem

$$T^{\frac{1}{2}}(\ln \hat{\sigma}^2 - \ln \sigma^2)/(2 + \hat{\gamma}_2)^{\frac{1}{2}} \to_{\mathscr{L}} \mathscr{N}(0, 1)\,.$$

From Monte-Carlo work in the i.i.d. case, for moderate sample sizes the following two data splitting procedures seem superior to this simple procedure.

3.2. *Simple data split.* Split the T residuals $\hat{a}_t$ into k groups each of size m as follows. By randomization, advocated by Scheffé (1959, page 87) in the i.i.d. case, or some other procedure partition the first T integers into k mutually exclusive groups of size m ($T = km$). Define the integers in the jth group by $P_j = \{i_{j_1}, \cdots, i_{j_m}\}$ for $j = 1, \cdots, k$. Define k estimates of σ^2 by

$$z_i = \sum_{t \in P_i} \hat{a}_t^2/m \qquad i = 1, \cdots, k\,.$$

It is shown in Section 4 that

$$\sum_{t \in P_i} a_t \hat{r}_t = O_p(1) \qquad i = 1, \cdots, k\,. \tag{3.4}$$

Thus using (2.3) and (3.4) one has

$$\begin{aligned} m^{\frac{1}{2}}(z_i - \sigma^2) &= \sum_{t \in P_i} (a_t^2 - \sigma^2)/m^{\frac{1}{2}} + \sum_{t \in P_i} (\hat{a}_t^2 - a_t^2)/m^{\frac{1}{2}} \\ &= \sum_{t \in P_i} (a_t^2 - \sigma^2)/m^{\frac{1}{2}} + O_p(m^{-\frac{1}{2}})\,. \end{aligned}$$

Thus by the central limit theorem for i.i.d. random vectors and the multivariate form of Slutsky's theorem, one has

$$m^{\frac{1}{2}}(\mathbf{z} - \sigma^2\mathbf{1}) \to_{\mathscr{L}} \mathscr{N}(\mathbf{0}, \sigma^4(2 + \gamma_2)I) \tag{3.5}$$

where $\mathbf{z} = (z_1, \cdots, z_k)$ and $\mathbf{1} = (1, 1, \cdots, 1)$ are k vectors. It follows easily from (3.5) that

$$k^{\frac{1}{2}}(\bar{z} - \sigma^2)/(\sum_1^k (z_i - \bar{z})^2/(k-1))^{\frac{1}{2}} \rightarrow_{\mathscr{L}} t_{k-1} \quad \text{as} \quad m \rightarrow \infty$$

where $\bar{z} = \sum_{i=1}^k z_i/k$.

Box (1953) originally used this data splitting technique, in the 2 sample case, after applying the log transformation to each estimate. Letting $\mathbf{u} = (u_1, \cdots, u_k)$ where $u_i = \log z_i$, then by Rao 6a.2 (iii)

$$m^{\frac{1}{2}}(\mathbf{u} - \ln \sigma^2 \mathbf{1}) \rightarrow_{\mathscr{L}} \mathscr{N}(\mathbf{0}, (2 + \gamma_2)I) .$$

So, exactly as above, one has

$$k^{\frac{1}{2}}(\bar{u} - \ln \sigma^2)/(\sum_{i=1}^k (u_i - \bar{u})^2/(k-1))^{\frac{1}{2}} \rightarrow_{\mathscr{L}} t_{k-1} \quad \text{as} \quad m \rightarrow \infty$$

where $\bar{u} = \sum_i^k u_i/k$.

Small sample Monte-Carlo work in the i.i.d. case indicates that this procedure performs better for small k.

3.3. *Jackknife*. This section shows that the jackknife is trustworthy for interval estimation of σ^2 for the ARMA model.

For notational simplicity let $v = \hat{\sigma}^2$ and $v_i = \sum_{t \notin P_i} \hat{a}_t^2/m(k-1)$ for $i = 1, \cdots, k$ where $\{P_1, \cdots, P_k\}$ is the partition defined in Section 3.2. Let g be a function with bounded second derivative in a neighborhood of σ^2 and define pseudovalues of the jackknife by

$$\hat{\theta}_i = kg(v) - (k-1)g(v_i) \quad \text{for} \quad i = 1, \cdots, k .$$

The jackknife estimate of $g(\sigma^2)$ is

$$\hat{\theta} = \sum_{i=1}^k \hat{\theta}_i/k = g(v) - (k-1) \sum_{i=1}^k (g(v_i) - g(v))/k .$$

THEOREM: *If* $T_{m,k} = k^{\frac{1}{2}}(\hat{\theta} - g(\sigma^2))/(\sum_{i=1}^k (\hat{\theta}_i - \hat{\theta})^2/(k-1))^{\frac{1}{2}}$, *then under the assumptions made above,*

(i) $T_{m,k} \rightarrow_{\mathscr{L}} t_{k-1}$ *as* $m \rightarrow \infty$
(ii) $T_{1,k} \rightarrow_{\mathscr{L}} \mathscr{N}(0, 1)$ *as* $k \rightarrow \infty$.

PROOF. Since the proof of (i) and (ii) are similar, only the proof to (ii) will be given, so it is assumed that $m = 1$ and $k = T$. As pointed out by Miller (1974b, page 2), this is probably the best partition to use if the data base is not too large. The theorem follows easily from Lemmas 1 and 2 which are proven now.

LEMMA 1. $T^{\frac{1}{2}}(\hat{\theta} - g(\sigma^2)) \rightarrow_{\mathscr{L}} \mathscr{N}(0, g'(\sigma^2)^2\sigma^4(2 + \gamma_2))$.

PROOF. Combining (3.1) with Rao 6a.2 (i) one has

$$T^{\frac{1}{2}}(g(\hat{\sigma}^2) - g(\sigma^2) - g'(\sigma^2)(\hat{\sigma}^2 - \sigma^2)) \rightarrow_P 0 .$$

We have

$$\begin{aligned} &T^{\frac{1}{2}}(\hat{\theta} - g(\sigma^2) - g'(\sigma^2)(\hat{\sigma}^2 - \sigma^2)) \\ &\quad = T^{\frac{1}{2}}(g(\hat{\sigma}^2) - g(\sigma^2) - g'(\sigma^2)(\hat{\sigma}^2 - \sigma^2)) - (T-1) \sum_{i=1}^T (g(v_i) - g(v))/T^{\frac{1}{2}} . \end{aligned}$$

If

$$(3.6) \qquad T^{\frac{1}{2}} \sum_{i=1}^{T} (g(v_i) - g(v)) \to_P 0 ,$$

then Lemma 1 is true by (3.1).

To show (3.6) consider

$$v - v_i = \sum_{t=1}^{T} \hat{a}_t^2/T - \sum_{t \neq i} \hat{a}_t^2/(T-1) = (\hat{a}_i^2 - \hat{\sigma}^2)/(T-1) .$$

Thus, applying the triangle inequality, one has

$$(3.7) \qquad \max_{1 \le i \le T} |v - v_i| \leqq \frac{1}{T-1} [|\hat{\sigma}^2 - \sigma^2| + \max_{1 \le i \le T} |a_i^2 - \sigma^2| + \max_{1 \le i \le T} |a_i^2 - \hat{a}_i^2|] .$$

By (2.4), (2.7) and (3.1) one has $\max_{1 \le i \le T} |v - v_i| = O_p(T^{-\frac{1}{2}})$. By an application of the triangle inequality using (3.1), one obtains

$$(3.8) \qquad \Delta_T = \max \{|v - \sigma^2|, |v_1 - \sigma^2|, \cdots, |v_T - \sigma^2|\} \to_P 0 .$$

Let $S(\sigma^2, \delta) = (\sigma^2 - \delta, \sigma^2 + \delta)$ be a neighborhood of σ^2, such that for all $t \in S$ $|g''(t)| \leqq M$ for some constant M. For fixed δ let $C_T = (\Delta_T < \delta)$ then by (3.8) $P(C_T) \to 1$ as $T \to \infty$. Following Miller (1974a, page 886), for an arbitrary sequence E_T one has that $\lim_T P(C_T E_T) = \lim_T P(E_T)$ so that the imposition or removal of the condition C_T has no effect on limiting probabilities.

Letting $\Delta_T < \delta$ and using Taylor's theorem for $i = 1, \cdots, T$, one has

$$(3.9) \qquad g(v_i) - g(v) = (v_i - v)g'(v) + \frac{(v_i - v)^2}{2} g''(\xi_i)$$

where $\xi_i \in S(\sigma^2, \delta)$, since it is on the line between v and v_i. Summing (3.9) for $i = 1, \cdots, T$ and using the identity $\sum_1^T (v_i - v) = 0$, one obtains

$$(3.10) \qquad \begin{aligned} |\textstyle\sum_{i=1}^{T} (g(v_i) - g(v))| &= \tfrac{1}{2} |\textstyle\sum_1^T (v_i - v)^2 g''(\xi_i)| \\ &\leqq M \textstyle\sum_1^T (v_i - v)^2/2 \\ &= M \textstyle\sum_1^T (\hat{a}_i^2 - \hat{\sigma}^2)^2/2(T-1)^2 . \end{aligned}$$

By (3.2) the bounding term in (3.10) is $O_p(T^{-1})$, which establishes Lemma 1.

To show that the jackknife is trustworthy, it suffices to show

LEMMA 2. $\sum_1^T (\hat{\theta}_i - \hat{\theta})^2/(T-1) \to_P g'(\sigma^2)^2 \sigma^4 (2 + \gamma_2)$.

PROOF. Given $\varepsilon > 0$ there exists $\delta_1 = \delta_1(\varepsilon)$ such that if $y \in S(\sigma^2, \delta_1)$ then $|g'(y) - g'(\sigma^2)| < \varepsilon$. Then if $\Delta_T < \delta_1$ applying a first order Taylor expansion, one has

$$(3.11) \qquad \begin{aligned} g(v_i) - g(v) &= (v_i - v)g'(\eta_i) \\ &= -(\hat{a}_i^2 - \hat{\sigma}^2)g'(\eta_i)/(T-1) , \end{aligned}$$

where $\eta_i \in S(\sigma^2, \delta_1)$ for $i = 1, \cdots, T$. Thus, one has

$$(3.12) \qquad \begin{aligned} \hat{\theta}_i - \hat{\theta} &= (T-1)(g(v_i) - g(v)) - (T-1) \textstyle\sum_{i=1}^{T} (g(v_i) - g(v))/T \\ &= b - (\hat{a}_i - \hat{\sigma}^2)g'(\eta_i) \end{aligned}$$

where $b = \sum_{i=1}^{T} (\hat{a}_i^2 - \hat{\sigma}^2)(g'(\eta_i) - g'(\sigma^2))/T$. Thus, one obtains from (3.12)

$$\begin{aligned}\sum_{i=1}^{T} (\hat{\theta}_i - \hat{\theta})^2 &= Tb^2 - 2b \sum_{i=1}^{T} (\hat{a}_i^2 - \hat{\sigma}^2)g'(\eta_i) + \sum_{i=1}^{T} (\hat{a}_i^2 - \hat{\sigma}^2)^2 g'(\eta_i)^2 \\ &= \sum_{i=1}^{T} (\hat{a}_i^2 - \hat{\sigma}^2)^2 g'(\eta_i)^2 - Tb^2 \\ &= g'(\sigma^2)^2 \sum_{i=1}^{T} (\hat{a}_i^2 - \hat{\sigma}^2)^2 \\ &\quad + \sum_{i=1}^{T} (\hat{a}_i^2 - \hat{\sigma}^2)^2 (g'(\eta_i)^2 - g'(\sigma^2)^2) - Tb^2 .\end{aligned}$$

Now using the Cauchy–Schwarz inequality, we have

$$\begin{aligned}&|\textstyle\sum_{i=1}^{T} ((\hat{\theta}_i - \hat{\theta})^2 - g'(\sigma^2)^2(\hat{a}_i^2 - \hat{\sigma}^2)^2)| \\ &\qquad = |\textstyle\sum_1^T (\hat{a}_i^2 - \hat{\sigma}^2)^2 (g'(\eta_i)^2 - g'(\sigma^2)^2) - Tb^2| \\ &\qquad \leqq \textstyle\sum_1^T (\hat{a}_i^2 - \hat{\sigma}^2)^2 |g'(\eta_i)^2 - g'(\sigma^2)^2| \\ &\qquad\quad + \textstyle\sum_1^T (\hat{a}_i^2 - \hat{\sigma}^2)^2 \sum_1^T (g'(\eta_i) - g'(\sigma^2))^2/T .\end{aligned} \tag{3.13}$$

For $\Delta_T < \delta_1$ the bounding term in (3.13) is bounded by

$$2\varepsilon(|g'(\sigma)| + \varepsilon) \textstyle\sum_1^T (\hat{a}_i^2 - \hat{\sigma}^2)^2 .$$

Since $\varepsilon > 0$ is arbitrary, the bounding term in (3.13) is $o_p(T)$ by (3.2). Thus, one has

$$\textstyle\sum_1^T (\hat{\theta}_i - \hat{\theta})^2/(T-1) - g'(\sigma^2)^2 \sum_1^T (\hat{a}_i^2 - \hat{\sigma}^2)^2/(T-1) \to_P 0 ,$$

which establishes Lemma 2 by (3.2).

3.4. *Discussion and summary.* Based on the results of Sections 3.1, 3.2 and 3.3, the methods given in Table 1 are asymptotically correct for setting confidence limits for σ^2 and $\ln \sigma^2$. The symbols $\Phi^{-1}(1-\alpha/2)$ and $t_{k-1}(1-\alpha/2)$ refer to upper percentage points of the normal and t_{k-1} distributions respectively. For calculation of the pseudovalues using the parameter $\ln \sigma^2$, the function $g(u) = \ln u$ is appropriate.

Although all these techniques are asymptotically valid, based on the Monte-Carlo work by Miller (1968) in the i.i.d. case, I would conjecture that the jackknife estimate ($m = 1$) of $\ln \sigma^2$ would be the best of these procedures for moderate sample sizes.

TABLE 1
Robust methods for setting confidence limits for the variance

Procedure	Parameter σ^2	Parameter $\ln \sigma^2$
Standard error	$\hat{\sigma}^2 \pm (\sum_t (\hat{a}_t^2 - \hat{\sigma}^2)^2/(T-1))^{\frac{1}{2}} \times \Phi^{-1}(1-\alpha/2)/T^{\frac{1}{2}}$	$\ln \hat{\sigma}^2 \pm (\sum_t \hat{a}_t^4/T(\hat{\sigma}^2)^2 - 1)^{\frac{1}{2}} \times \Phi^{-1}(1-\alpha/2)/T^{\frac{1}{2}}$
Jackknife $m = 1$	same as above	$\hat{\theta} \pm (\sum_1^T (\hat{\theta}_i - \hat{\theta})^2/(T-1))^{\frac{1}{2}} \times \Phi^{-1}(1-\alpha/2)/T^{\frac{1}{2}}$
Simple Data Split	$\bar{z} \pm (\sum_1^k (z_i - \bar{z})^2/(k-1))^{\frac{1}{2}} \times t_{k-1}(1-\alpha/2)/k^{\frac{1}{2}}$	$\bar{u} \pm (\sum_1^k (u_i - \bar{u})^2/(k-1))^{\frac{1}{2}} \times t_{k-1}(1-\alpha/2)/k^{\frac{1}{2}}$
Jackknife $m > 1$	same as above	$\hat{\theta} \pm (\sum_1^k (\hat{\theta}_i - \hat{\theta})^2/(k-1))^{\frac{1}{2}} \times t_{k-1}(1-\alpha/2)/k^{\frac{1}{2}}$

4. Negligibility of remainder terms. This section proves (2.2), (2.3) and (3.4). From (2.1) one has

$$\hat{s}_t = \textstyle\sum_1^{p+q+1} (\hat{\beta}_j - \beta_j)\tilde{d}_{tj} \tag{4.1}$$

where

$$\tilde{d}_{tj} = \frac{\partial \hat{a}_t}{\partial \hat{\beta}_j}\Big|_{\tilde{\beta}}$$

and $\tilde{\boldsymbol{\beta}}$ is on the line between $\boldsymbol{\beta}$ and $\hat{\boldsymbol{\beta}}$. We have

$$\begin{aligned} \tilde{d}_{tj} &= \textstyle\sum_{i=0}^{t-j-1} \tilde{\pi}_i^{(1)} \tilde{x}_{t-j-i} && 1 \leqq j \leqq p \\ &= \textstyle\sum_{i=0}^{t-j-1} \tilde{\pi}_i^{(2)} \tilde{x}_{t-j-i} && p+1 \leqq j \leqq p+q \\ &= -\textstyle\sum_0^{t-j-1} \tilde{\pi}_i^{(0)} && j = p+q+1 \end{aligned}$$

where $\tilde{\pi}^{(1)}(B) = \tilde{\theta}^{-1}(B) = \sum_0^\infty \tilde{\pi}_i^{(1)} B^i$ and $\tilde{\pi}^{(2)}(B) = -\tilde{\varphi}(B)\tilde{\theta}^{-2}(B) = \sum_0^\infty \tilde{\pi}_i^{(2)} B^i$.

The following easily proven lemma states that with high probability the π weights die out exponentially.

LEMMA 3. *Given $\varepsilon > 0$ there exists a T_0 such that with probability $\geqq 1 - \varepsilon$ for $T \geqq T_0$ $|\tilde{\pi}_i^{(k)}| \leqq MR^{-i}$ $k = 0, 1, 2$ and $i = 0, 1, 2, \cdots$, for some M and $1 < R < \min(R_\varphi, R_\theta)$ where R_φ and R_θ are the radii of convergence of the power series $\varphi(B)$ and $\theta(B)$.*

The following two lemmas are useful in proving the main results of this section. The proof of Lemma 4 (see Davis (1975)) is messy and is omitted here.

LEMMA 4.

(i) $\sum_t a_t \hat{s}_t = O_p(1)$,
(ii) $\sum_t \hat{s}_t^4 = O_p(T^{-1})$.

The sums can be over $1 \leqq t \leqq T$ or $t \in P_i$.

LEMMA 5. *The following are bounded in probability when summed over $1 \leqq t \leqq T$ or $t \in P_i$.*

(i) $\sum a_t w_t$,
(ii) $\sum w_t^n$ *for $n = 2$ and 4.*

PROOF. Since $(\beta_1, \cdots, \beta_{p+q}) \in \Omega$, there exists M and $\mathscr{R} > 1$ such that $|\psi_i^{(0)}| \leqq M\mathscr{R}^{-i}$ for $i = 0, 1, \cdots$. Thus, one obtains

$$E(\textstyle\sum_1^T a_t w_t)^2 = \sigma^2 E(x_0 - \mu)^2 \sum_1^T \sum_{i=t}^\infty \pi_i^{(0)2} \leqq M^2\sigma^2 E(x_0 - \mu)^2 (1 - \mathscr{R}^{-2})^{-2} = O(1)$$

since $E(x_0 - \mu)^2 < \infty$ follows from Lemma 3.

This establishes (i) when $1 \leqq t \leqq T$; the proof for $t \in P_i$ is analogous. To show (ii), one has

$$E \textstyle\sum_1^T w_t^2 = E(x_0 - \mu)^2 \sum_1^T \sum_t^\infty \pi_i^{(0)2} = O(1)$$

as in (i). When $n = 4$, one still has $E \sum_1^T w_t^4 = O(1)$, establishing (ii) in the case $1 \leqq t \leqq T$. Since $\sum_{t\in P_i} w_t^n \leqq \sum_1^T w_t^n$, (ii) is true for summing over $t \in P_i$. This completes the proof of Lemma 5.

Continuing now with the proof of the main results of this section, one has $\sum_{t=1}^T a_t \hat{r}_t = \sum_1^T a_t(\hat{s}_t - w_t) = O_p(1)$ by (i) of Lemmas 4 and 5, which establishes (2.2). By (ii) or Lemmas 4 and 5, one has

$$\sum_{t=1}^T \hat{r}_t^4 \leqq 16 \sum_1^T (\hat{s}_t^4 + w_t^4) = O_p(1) .$$

Applying the Cauchy–Schwarz inequality, one obtains

$$\sum_{t=1}^T \hat{r}_t^2 \leqq 2(\sum_1^T \hat{s}_t^2 + \sum_1^T w_t^2) \leqq 2((T \sum_1^T \hat{s}_t^4)^{\frac{1}{2}} + \sum_1^T w_t^2) = O_p(1)$$

by (ii) of Lemmas 4 and 5. These last two convergences establish (2.3). Equation (3.4) follows from (2.3) and (i) of Lemmas 4 and 5.

Acknowledgment. This is a revised version of the author's doctoral dissertation written under the direction of Stephen M. Stigler at the University of Wisconsin.

REFERENCES

[1] Box, G. E. P. (1953). Non-normality and tests on variances. *Biometrika* **40** 318–335.
[2] Box, G. E. P. and Jenkins, G. M. (1970). *Time-Series Analysis: Forecasting and Control.* Holden-Day, San Francisco.
[3] Davis, W. W. (1975). Robust interval estimation of the error variance of an ARMA model. Technical Report 95, Dept. of Statistics, Carnegie-Mellon Univ.
[4] Gray, H. L., Watkins, T. A. and Adams, J. E. (1972). On the jackknife statistic, its extensions, and its relation to the e_n-transformation. *Ann. Math. Statist.* **43** 1–30.
[5] Hannan, E. J. (1969). The estimation of mixed moving average autoregressive systems. *Biometrika* **56** 579–593.
[6] Layard, M. W. J. (1972). Large sample tests for equality of two covariance matrices. *Ann. Math. Statist.* **43** 123–141.
[7] Mann, H. B. and Wald, A. (1943). On stochastic limit and order relationships. *Ann. Math. Statist.* **14** 217–226.
[8] Miller, R. G. Jr. (1968). Jackknifing variances. *Ann. Math. Statist.* **39** 568–582.
[9] Miller, R. G. Jr. (1974a). An unbalanced jackknife. *Ann. Statist.* **2** 880–891.
[10] Miller, R. G. Jr. (1974b). The jackknife—a review. *Biometrika* **61** 1–15.
[11] Quenouille, M. (1949). Approximate tests of correlation in time series. *J. Roy. Statist. Soc. Ser. B* **11** 68–84.
[12] Rao, C. R. (1973). *Linear Statistical Inference and Its Application.* Wiley, New York.
[13] Scheffé, H. (1959). *The Analysis of Variance.* Wiley, New York.
[14] Tukey, J. W. (1958). Bias and confidence in non-quite large samples (abstract). *Ann. Math. Statist.* **29** 614.
[15] Wold, H. (1938). *A Study in the Analysis of Stationary Time Series.* Almquist and Wiksell, Uppsala.
[16] Yule, G. U. (1927). On a method of investigating periodicities in disturbed series with special reference to Wolfer's sunspot numbers. *Philos. Trans. Roy. Statist. Soc. London Ser. A* **226** 267–298.

Department of Statistics
Carnegie-Mellon University
Schenley Park
Pittsburgh, Pennsylvania 15213