The Annals of Statistics
1987, Vol. 15, No. 3, 1064–1078

MODEL-FREE ONE-STEP-AHEAD PREDICTION INTERVALS: ASYMPTOTIC THEORY AND SMALL SAMPLE SIMULATIONS

BY SINSUP GHO[1] AND ROBERT B. MILLER

University of Wisconsin–Madison

We show that the empirical quantile process from an ARMA(1, q) process which is strongly mixing Δ_s, and is either Gaussian or double exponential, converges to a Gaussian process. This result is used to derive model-free one-step-ahead prediction intervals for such processes. Simulations demonstrate where the asymptotic theory can and cannot be applied to small samples.

1. Introduction. Consider a strictly stationary sequence of random variables (r.v.'s) $\{X_n, n \geq 1\}$ from a stochastic process with marginal c.d.f. $F(\cdot)$. We consider the prediction of a future observation X_{n+1} using prediction intervals. Model-based approaches to forecasting have been suggested by many authors, but, because of the sophistication and computational complexity of the model building approach, simpler methods are increasingly demanded by practitioners [Makridakis and Hibon (1979), Carbone, Anderson, Corriveau and Corson (1983) and Brandon, Jarrett and Khumawala (1983)]. In this paper the simple forecasting technique proposed in Butler (1982) is evaluated for prediction interval construction in ARMA(1, q) processes. The need for prediction intervals instead of just point prediction was emphasized eloquently by Keyfitz (1972).

Let F_n denote the empirical distribution function (e.d.f.) based on a sequence $X_1, X_2, \ldots, X_n$, that is,

$$F_n(t) = n^{-1} \sum_{i=1}^{n} I_{(-\infty, t]}(X_i). \tag{1.1}$$

Gastwirth and Rubin (1975a) showed that the empirical process $n^{1/2}[F_n(t) - F(t)]$ converges weakly to a Gaussian process with a.s. continuous paths for a strongly mixing Δ_s process.

Define the tth sample quantile $F_n^{-1}(t)$ as

$$F_n^{-1}(t) = \begin{cases} \inf\{x: F_n(x) \geq t\} & \text{for } t > 0, \\ F_n^{-1}(0^+) & \text{for } t = 0. \end{cases} \tag{1.2}$$

We will show in Section 3 that the quantile process $n^{1/2}[F_n^{-1}(t) - F^{-1}(t)]$ from either the first order autoregressive Gaussian process (FOAGP) or the first order autoregressive double exponential process (FOADP), which are strongly mixing Δ_s processes, converges to a Gaussian process.

Received July 1984; revised February 1986.

[1] Now at Seoul National University.

AMS 1980 *subject classifications*. 62G30, 62M20.

Key words and phrases. Strong mixing Δ_s, empirical quantile process, prediction interval.

This convergence has an immediate application in model-free prediction for dependent processes. In the context of an independent process Butler (1982) constructed a $100\gamma\%$ prediction interval by locating the $100\gamma\%$ span of the data which supports the smallest trimmed variance. Our result permits extension of this approach to serially correlated sequences. The approach does not require any parametric models and can be applied to Gaussian or double exponential sequences.

The contents of the paper are as follows. Mixing numbers are defined and the weak convergence of empirical processes is reviewed under various conditions in Section 2, and the convergence of the quantile process to a Gaussian process is shown in Section 3. The final section shows how to obtain the model-free prediction intervals and presents simulated comparisons of these prediction intervals with others constructed from the true parametric model.

2. Mixing conditions and the weak convergence of the empirical process. Let $\{X_n;\ -\infty < n < \infty\}$ be a strictly stationary sequence of r.v.'s. Denote by $\mu_{-\infty}^{k}$ and μ_k^{∞} the σ-fields generated by r.v.'s $\{X_n;\ n \le k\}$ and $\{X_n;\ n \ge k\}$, respectively. Let $\phi_0 = 1$, $\alpha_0 = 1$ and for $n \ge 1$ define

$$\phi_n = \sup\{|P(E_2|E_1) - P(E_2)|:\ E_1 \in \mu_{-\infty}^{0},\ E_2 \in \mu_n^{\infty}\} \tag{2.1}$$

and

$$\alpha_n = \sup\{|P(E_1E_2) - P(E_1)P(E_2)|:\ E_1 \in \mu_{-\infty}^{0},\ E_2 \in \mu_n^{\infty}\}. \tag{2.2}$$

In the definition of ϕ_n, we are adopting the convention that $P(E_2|E_1) = 0$ if $P(E_1) = 0$. A stochastic process $\{X_n\}$ is called ϕ-mixing if there exists a sequence $\{\phi_n\}$ such that

$$1 \ge \phi_1 \ge \phi_2 \cdots, \qquad \lim_{n\to\infty} \phi_n = 0$$

and strong mixing if there exists a sequence $\{\alpha_n\}$ such that

$$1 \ge \alpha_1 \ge \alpha_2 \ge \cdots, \qquad \lim_{n\to\infty} \alpha_n = 0.$$

The process $\{X_n\}$ is m-dependent if the random vectors $(X_i, \ldots, X_k)$ and $(X_{k+n}, \ldots, X_j)$ are independent whenever $n > m$. Such a process is ϕ-mixing with $\phi_n = 0$ for $n > m$.

It is known [see, e.g., Theorem 17.3.2 of Ibragimov and Linnik (1971)] that a stationary Gaussian sequence is ϕ-mixing iff it is m-dependent for some m. Gastwirth and Rubin (1975a) showed that the first order autoregressive process is not ϕ-mixing, introduced a new mixing process (strongly mixing Δ_s process) which is easier to calculate and showed that FOAGP and FOADP satisfy the conditions of a strongly mixing Δ_s process. For practical purposes, it is helpful to think of these processes as strong mixing with $\alpha_n = O(e^{-\theta n})$ for some $\theta \in (0, 1)$, although the definition is much more general.

Doob (1949) suggested a novel approach to the study of the empirical process $\{n^{1/2}[F_n(x) - F(x)];\ -\infty < x < \infty\}$, where $X_1, X_2, \ldots$ are i.i.d. r.v.'s with distribution $F(\cdot)$ and Donsker (1952) justified Doob's heuristic approach. Let $\gamma_n^*(y)$

be the continuous version of the empirical process of $Y_i = F(X_i)$. Donsker (1952) showed that

$$h(\gamma_n^*(y)) \to_D h(\beta(y)), \quad \text{as } n \to \infty,$$

for every continuous functional h: $C(0,1) \to R^1$, where $\{\beta(y);\ 0 \le y \le 1\}$ is a Brownian bridge. In the following $E_n(\cdot)$ is defined as the e.d.f. of $Y_i = F(X_i)$. We can see that the sample functions of $\gamma_n(y) = n^{1/2}[E_n(y) - y]$ do not belong to $C(0,1)$. This difficulty was solved by Prohorov (1956) and Skorohod (1956), while working on the so-called $D(0,1)$ function space.

For dependent processes Billingsley (1968, Theorem 22.1), showed that the empirical process from a ϕ-mixing process converges to a Gaussian process.

THEOREM (Billingsley). *Suppose* $\{X_n\}$ *is* ϕ*-mixing with* $\sum n^2\phi_n^{1/2} < \infty$, *and* X_0 *has a continuous distribution function* F *on* $[0,1]$. *Then*

$$Y_n \to_D Y,$$

where Y *is the Gaussian random function specified by* (22.11) *and* (22.12) *of Billingsley and* $P\{Y \in C\} = 1$.

It was shown by Sen (1971) that under a weaker condition on the ϕ-mixing numbers, $\sum n\phi_n^{1/2} < \infty$, the same result can be obtained. Deo (1973) showed that Billingsley's result remains true for strong mixing sequences with $\sum n^2\alpha_n^{1/2-\tau} < \infty$ for some $0 < \tau < \frac{1}{2}$. For a strongly mixing Δ_s process Gastwirth and Rubin (1975a), Theorem 3.1, obtained a similar result.

3. Convergence of the quantile process. For a sequence of independent r.v.'s Bickel (1967) showed that the quantile process converges to a Gaussian process. Let $Z_n(t)$ be a piecewise linear sample quantile function defined as in Bickel (1967). Bickel showed that $n^{1/2}Z_n(t)$ converges to $Z(t)$ in the sense of Prohorov on $[\alpha, \beta]$, $0 < \alpha < \beta < 1$, where $Z(t)$ is a centered Gaussian process on $[\alpha, \beta]$ with continuous sample functions and covariance

$$s(1-t)/\{f[F^{-1}(s)]\, f[F^{-1}(t)]\}, \qquad s \le t.$$

Mehra and Rao (1975) showed that if $\sum n^2\phi_n^{1/2} < \infty$, then the quantile process converges in probability in sup norm to a Gaussian process when the sequence of r.v.'s $\{X_n\}$ are from a ϕ-mixing process. For a strong mixing process, Mehra and Rao also showed that if $\sum n^2\alpha_n^{\delta} < \infty$ for some $0 < \delta < 1$, then the quantile process converges in probability in sup norm to a Gaussian process. However, their results were confined to the sequence of uniformly distributed r.v.'s. In this section we will extend their results to the case when the underlying distribution is not uniform.

In Babu and Singh (1978) the deviation between the empirical and the quantile processes for ϕ-mixing or strong mixing r.v.'s was given for the uniform distribution. As an extension to general distributions, they gave the following theorem. We say that a d.f. F with density f satisfies the condition (*) if for

some interval I,

$$f(x) = 0 \quad \text{for } x \notin I,$$
$$\inf\{f(x), x \in I\} > 0$$

and

$$\sup\{f(x), x \in I\} < \infty.$$

THEOREM (Babu and Singh). *Let $\{X_n\}$ be a strictly stationary ϕ-mixing sequence of r.v.'s such that $\sum \phi_n^{1/2} < \infty$ and the underlying d.f. F satisfies the condition (*). Then*

$$\limsup_{n\to\infty} b_n^{-1} R_n^* \le d_1 \quad a.s., \tag{3.1}$$

where

$$b_n = n^{-3/4}(\log n)^{1/2}(\log\log n)^{1/4}, \tag{3.2}$$

$$R_n^* = \sup_{0\le t\le 1} |R_n^*(t)|, \tag{3.3}$$

$$R_n^*(t) = F_n^{-1}(t) - F^{-1}(t) + \left[F_n(F^{-1}(t)) - t\right]/f\left[F^{-1}(t)\right] \tag{3.4}$$

and d_1 is some constant.

Throughout the paper c, k and d_i's stand for constants.

Using the weak convergence of the empirical process, we can show that the quantile process converges weakly to a Gaussian process following Theorem 4.1 of Billingsley (1968). Furthermore, we have

THEOREM 1. *Let $\{X_n\}$ be a sequence of r.v.'s from a strictly stationary ϕ-mixing process such that $\sum n^2 \phi_n^{1/2} < \infty$ and the underlying d.f. satisfies the condition (*). Then*

$$\sup_{\varepsilon\le t\le 1-\varepsilon} |Z_n(t) - Z(t)| \to_P 0, \quad \text{as } n \to \infty \text{ for some } \varepsilon > 0, \tag{3.5}$$

where

$$Z_n(t) = n^{1/2}\left[F_n^{-1}(t) - F^{-1}(t)\right], \qquad Z(t) = Y_0(t)/f\left[F^{-1}(t)\right],$$
$$Y_0(t) = -Y(t),$$

and $Y(t)$ is defined as in Billingsley (1968).

PROOF. Since

$$E_n^{-1}(t) - t = F\left[F_n^{-1}(t)\right] - F\left[F^{-1}(t)\right]$$
$$= f\left[F^{-1}(t)\right]\left[F_n^{-1}(t) - F^{-1}(t)\right] + f'(\eta)\left[F_n^{-1}(t) - F^{-1}(t)\right]^2,$$

for a random point η in I,

$$F_n^{-1}(t) - F^{-1}(t) = \left[E_n^{-1}(t) - t\right]/f\left[F^{-1}(t)\right]$$
$$-f'(\eta)\left[F_n^{-1}(t) - F^{-1}(t)\right]^2/f\left[F^{-1}(t)\right].$$

Multiplying by $n^{1/2}$ and subtracting $Z(t)$, we obtain

$$\sup_{\varepsilon\le t\le 1-\varepsilon}|Z_n(t)-Z(t)| < d_2 \sup_{\varepsilon\le t\le 1-\varepsilon}\left|n^{1/2}\left[E_n^{-1}(t)-t\right]-Y_0(t)\right| + d_3 \sup_{\varepsilon\le t\le 1-\varepsilon} n^{1/2}\left[F_n^{-1}(t)-F^{-1}(t)\right]^2.$$

By condition (*) and Lemma 4.2 of Babu and Singh (1978), the second term on the right is $O(n^{-1/2}(\log\log n))$. Since the first term goes to 0 in probability following Mehra and Rao (1975), the result follows. □

For a strong mixing process with $\alpha_n = O(e^{-\theta n})$, for some $\theta > 0$, Babu and Singh (1978) also showed that

$$\sup_{0\le t\le 1}\left|F_n^{-1}(t)-t+F_n(t)-t\right| = O\left(n^{-3/4}(\log n)(\log\log n)^{1/4}\right) \quad \text{a.s.}, \tag{3.6}$$

where X_0 is uniformly distributed on the unit interval. We can give the following theorem when the underlying distribution is not uniform.

THEOREM 2. *Let $\{X_n\}$ be a strictly stationary strong mixing sequence of r.v.'s such that $\alpha_n = O(e^{-\theta n})$ for some $\theta > 0$. Assume the underlying distribution F satisfies condition (*) and $f'(x)$ exists and is bounded on I. Then*

$$\limsup_{n\to\infty} C_n^{-1}R_n^* \le d_4 \quad a.s., \tag{3.7}$$

where

$$C_n = n^{-3/4}(\log n)(\log\log n)^{1/4} \tag{3.8}$$

and R_n^* and $R_n^*(t)$ are defined in (3.3) *and* (3.4), *respectively.*

PROOF. Let $R_n(t) = E_n^{-1}(t) + E_n(t) - 2t$. Since $E_n^{-1}(t) = F[F_n^{-1}(t)]$, $E_n(t) = F_n[F^{-1}(t)]$ and $t = F[F^{-1}(t)]$,

$$\begin{aligned} R_n(t) &= F\left[F_n^{-1}(t)\right] + F_n\left[F^{-1}(t)\right] - 2F\left[F^{-1}(t)\right] \\ &= f\left[F^{-1}(t)\right]R_n^*(t) + f'(\eta)\left[F_n^{-1}(t)-F^{-1}(t)\right]^2 \end{aligned}$$

for every $t \in [0,1]$, where η is a random point in I. Using Theorem 4 and the analog of Lemma 4.2 of Babu and Singh (1978) for a strong mixing process, it can be shown that

$$\begin{aligned} d_5 \sup_{0\le t\le 1}|R_n^*(t)| &< \sup_{0\le t\le 1}|R_n(t)| + d_6 \sup_{0\le t\le 1}\left[F_n^{-1}(t)-F^{-1}(t)\right]^2 \\ &= O\left(n^{-3/4}(\log n)(\log\log n)^{1/4}\right) \quad \text{a.s.} \end{aligned}$$

□

Again, using the weak convergence of the empirical process, we can show that the quantile process converges weakly to a Gaussian process for the strong mixing process. Also we can give the following theorem.

THEOREM 3. *Let $\{X_n\}$ be a sequence of r.v.'s from a strictly stationary strong mixing process such that $\alpha_n = O(e^{-\theta n})$ for some $0 < \theta < 1$ and the underlying d.f. satisfies condition (*). If $|\mathrm{Corr}(X, Y)| < C\alpha_n^{\delta}$ for some $0 < \delta < 1$, where X and Y are measurable w.r.t. the σ-fields $\mu_{-\infty}^{0}$ and μ_n^{∞}, respectively, then*

$$\sup_{\varepsilon \le t \le 1-\varepsilon} |Z_n(t) - Z(t)| \to_P 0, \quad \text{as } n \to \infty \text{ for some } \varepsilon > 0. \tag{3.9}$$

PROOF. Following Theorem 3.1 of Mehra and Rao (1975)

$$\sup_{\varepsilon \le t \le 1-\varepsilon} \left|n^{1/2}\left[E_n^{-1}(t) - t\right] - Y_0(t)\right| \to_P 0, \quad \text{as } n \to \infty.$$

The remaining part of the proof is exactly the same as the proof of Theorem 1. □

Following Gastwirth and Rubin (1975a), $\|\Delta(k)\|_1 = O(|\rho|^k)$, for FOAGP and FOADP. Since $\|\Delta(k)\|_1 = O[\exp(k \log|\rho|)] = O[\exp(-\theta k)]$, where $\theta = -\log|\rho| > 0$ for $0 < |\rho| < 1$, we can show that $\alpha_k = O[\exp(-\theta k)]$ for FOAGP and FOADP, using $\alpha_k \le \|\Delta(k)\|_1/4$, and obtain the following corollary.

COROLLARY 3.1. *For FOAGP and FOADP*

$$\sup_{\varepsilon \le t \le 1-\varepsilon} |Z_n(t) - Z(t)| \to_P 0, \quad \text{as } n \to \infty \text{ for some } \varepsilon > 0.$$

4. Use of quantiles in prediction. Butler (1982) proposed a randomly located interval for a future observation from a location model. If F is strictly increasing over the interval of support, then the class of the $100\gamma\%$ prediction interval is

$$\left\{I(\delta) = \left[F^{-1}(\delta), F^{-1}(\delta + \gamma)\right]; 0 \le \delta \le 1 - \gamma\right\}. \tag{4.1}$$

Given a random sample $X_1, X_2, \ldots, X_n$, if F is known, we consider predicting X_{n+1} with the interval (4.1), which supports the smallest trimmed variance, i.e., we use $I(\delta^*)$ where $\delta^* = \delta^*(\gamma)$ is the value of δ, which minimizes

$$\sigma^2(\delta) = \gamma^{-1} \int_{F^{-1}(\delta)}^{F^{-1}(\delta+\gamma)} x^2 \, dF(x) - \left\{\gamma^{-1} \int_{F^{-1}(\delta)}^{F^{-1}(\delta+\gamma)} x \, dF(x)\right\}^2. \tag{4.2}$$

If F is unknown, an estimator of δ^*, $\hat{\delta}^*$ say, is obtained when F^{-1} in (4.2) is replaced by F_n^{-1}, the sample quantile function. Then the $100\gamma\%$ span of untrimmed data,

$$\hat{I}(\hat{\delta}^*) = \left[F_n^{-1}(\hat{\delta}^*), F_n^{-1}(\hat{\delta}^* + \gamma)\right], \tag{4.3}$$

provides a tolerance interval predictor of the next observation from the location sample.

It is natural to use the center of $\hat{I}(\hat{\delta}^*)$ as a point predictor of X_{n+1}. Butler showed that the center of $\hat{I}(\hat{\delta}^*)$ is asymptotically the mean of the data spanned by $\hat{I}(\hat{\delta}^*)$ and represents the mean of the remaining data after the elimination of the $100(1 - \gamma)\%$ "most outlying" subset of the data. In this section we will

extend Butler's result to the dependent case and will present a simulation result for FOAGP and FOADP.

Let $\{X_n\}$ be a sequence of r.v.'s from a ϕ-mixing or a strong mixing process. Butler specified sufficient conditions for the uniqueness of δ^*, Lemma 2.1, which are still applicable to our problem. By the following lemma we can show the consistency of $\hat{\delta}^*$.

LEMMA 4.1. *Let F be continuous, strictly increasing on interval support with a finite mean and a differentiable density f that is unimodel. Then, for a ϕ-mixing process under the same conditions as in Theorem* 1 *and for a strong mixing process under the same conditions as in Theorem* 3, $\hat{\delta}^* \to_p \delta^*$ *as* $n \to \infty$.

PROOF. If we use Theorems 1 and 3, instead of Lemma 2.2 of Butler, then the proof is exactly along the same lines as in Butler (1982). □

Let $P_n(\gamma) = F[F_n^{-1}(\hat{\delta}^* + \gamma) - F[F_n^{-1}(\hat{\delta}^*)]$ be the coverage of $\hat{I}(\hat{\delta}^*)$. Then, following the proof of Theorem 3.1 of Butler (1982), we can obtain the following theorem.

THEOREM 4. *Under the same conditions as in Lemma* 4.1

$$n^{1/2}[P_n(\gamma) - \gamma] \to_D N[0, v(\delta^*, \gamma)], \quad \text{as } n \to \infty, \tag{4.4}$$

where

$$v(\delta^*, \gamma) = E\left[\{Y(\delta^* + \gamma) - Y(\delta^*)\}^2\right].$$

For FOAGP, using Lemma 2.1 of Gastwirth and Rubin (1975b), we have

$$\begin{aligned} v(\delta^*, \gamma) = \gamma(1-\gamma) + \Bigg\{ & \exp\left[F^{-1}(\delta^* + \gamma)^2\right] \sum_{n=1}^{\infty} \sum_{k=1}^{\infty} H_{k-1}^2\left[F^{-1}(\delta^* + \gamma)\right] \rho_n^k / k! \\ & + \exp\left[-F^{-1}(\delta^*)^2\right] \sum_{n=1}^{\infty} \sum_{k=1}^{\infty} H_{k-1}^2\left[F^{-1}(\delta^*)\right] \rho_n^k / k! \\ & - 2\exp\left[-\left\{F^{-1}(\delta^* + \gamma)^2 + F^{-1}(\delta^*)^2\right\}/2\right] \\ & \times \sum_{n=1}^{\infty} \sum_{k=1}^{\infty} H_{k-1}\left[F^{-1}(\delta^* + \gamma)\right] H_{k-1}\left[F^{-1}(\delta^*)\right] \rho_n^k \Bigg\} / \pi, \end{aligned} \tag{4.5}$$

where ρ_n^k is a correlation coefficient between X_n and X_{n+k} in FOAGP and $H_k(a)$ is the kth Hermite polynomial [for the definition, see Hochstadt (1961)].

It is noted that, because of the condition in Lemma 4.1 and the structure of $v(\delta^*, \gamma)$, we need a "nice" form for F to apply Theorems 2 and 3. F is not always nice when we have a nonnormal error distribution in the MA process. In FOADP we have to have the following error distribution to get the double exponential marginal distribution, $f(x) = e^{-|x|}/2$, $-\infty < x < \infty$, Gastwirth,

Rubin and Wolff (1967):

$$f(\varepsilon) = \begin{cases} 0 & \text{with probability } \rho^2, \\ e^{-|\varepsilon|}/2 & \text{with probability } (1-\rho^2). \end{cases} \tag{4.6}$$

We have seen, for any pure MA process, the quantile process converges to a Gaussian process following Billingsley (1968), since any MA process is m-dependent. By Corollary 3.1 the quantile process converges to a Gaussian process for FOAGP and FOADP. It will be seen later that the quantile process converges to a Gaussian process for the mixed autoregressive moving average process of order 1 and q (ARMA(1, q)), $q = 1, 2, \ldots$, with normal or double exponential errors. Hence for these processes we can use the class of $100\gamma\%$ tolerance intervals to predict the next observation. We have not been able to extend this to the general ARMA(p, q) process with $p \geq 2$.

Consider an ARMA(1, q) process, $q \geq 1$,

$$X_n = \rho X_{n-1} + \varepsilon_n - \theta_1 \varepsilon_{n-1} - \cdots - \theta_q \varepsilon_{n-q}. \tag{4.7}$$

It is not too hard to show that (4.7) can be rewritten

$$X_n = \rho^n X_0 + U_n, \tag{4.8}$$

where

$$\begin{aligned} U_n = {} & \varepsilon_n + (\rho - \theta_1)\varepsilon_{n-1} + \left(\rho^2 - \rho\theta_1 - \theta_2\right)\varepsilon_{n-2} \\ & + \cdots + \left(\rho^{q-1} - \rho^{q-2}\theta_1 - \cdots - \theta_{q-1}\right)\varepsilon_{n-q+1} \\ & + \left[\rho^q - \rho^{q-1}\theta_1 - \cdots - \theta_q\right]\left[\varepsilon_{n-q} + \rho\varepsilon_{n-q-1} + \cdots + \rho^{(n-2q)}\varepsilon_q\right] \\ & - \rho^{(n-2q+1)}\left[\rho^{q-1}\theta_1 + \cdots + \theta_q\right]\varepsilon_{q-1} \\ & - \rho^{(n-2q+2)}\left[\rho^{q-2}\theta_2 + \cdots + \theta_q\right]\varepsilon_{q-2} \\ & - \cdots - \rho^{n-q}\theta_q\varepsilon_0, \end{aligned} \tag{4.9}$$

with $\varepsilon_{q-1} = \cdots = \varepsilon_0 = 0$, $E[U_n] = 0$ and

$$\begin{aligned} \operatorname{var}[U_n] = {} & \Big\{1 + (\rho - \theta_1)^2 + \cdots + \left[\rho^{q-1} - \rho^{q-2}\theta_1 - \cdots - \theta_{q-1}\right]^2 \\ & + \left[\rho^q - \rho^{q-1}\theta_1 - \cdots - \theta_q\right]\left[1 - \rho^{2(n-2q+1)}\right]/ \\ & \left(1 - \rho^2\right)\Big\}\operatorname{Var}(\varepsilon_n). \end{aligned} \tag{4.10}$$

With $\operatorname{Var}[\varepsilon_n]$, which satisfies (4.10), we obtain $\operatorname{Var}[U_n] = 1/(1 - \rho^{2n})$ and $\operatorname{Var}[X_n] = 1$ and we can show that the mixing number of (4.7) satisfies $\|\Delta(k)\|_1 = O(|\rho|^k)$.

We present a simulation result comparing the performances of the B-J method [e.g., (5.2.6) of Box and Jenkins (1976)] and the simple method in which the quantile function is used to choose a pair of order statistics to form a predictive interval (P.I.). Coverage frequencies of one-step-ahead P.I.'s are compared for FOAGP and FOADP for three different sample sizes, $n = 10$, 20 and

100, and for $\rho = -0.9$–0.9 in increments of 0.1. Although the B-J method is usually not recommended for small samples of sizes $n = 10$ and 20, P.I.'s are obtained by both methods for comparison purposes. The study favors the B-J method because the correct model is always used. There is no identification error.

To obtain the P.I.'s for $n = 100$ (20 and 10) we generate 101 (21 and 11) observations from FOAGP and FOADP for $\rho = -0.9$–0.9 in increments of 0.1. Using the first 100 (20 and 10) observations 90% (90% and 80%) P.I. is obtained by the B-J method. For the simple method we use the fifth and the 95th observations for $n = 100$ and the smallest and the largest observations for $n = 20$ and 10. For the B-J method, the number of times we have significant (5%) residual ACF's or PACF's after AR(1) fitting are obtained. ACF and PACF stand for autocorrelation function and partial autocorrelation function.

Although the sample quantile functions suggested by Butler (1982) and Parzen (1979) have been recommended because of their accuracy for small samples, we use the sample quantile function (1.2) in using the simple method. The reason we choose (1.2) is that this one is the most conservative one in the sense that if we use the same order statistics it gives the largest coverage probability. For example, if $[X_{(1)}, X_{(10)}]$ is used with $n = 10$, (1.2) ensures 80% while Butler's and Parzen's methods ensure 90%. In the independent case, Hall, Prairie and Motlagh (1975) showed that its coverage probability is 81.82% instead of 90% using the hypergeometric function. We are deliberately avoiding the piecewise linear quantile function for the sake of simplicity. Also, we use the symmetric P.I. because the simulation is performed on normal and double exponential distributions. As Butler (1982) indicated, δ^* picked out the "most symmetric" $100\gamma\%$ of F by centering the interval mean. Since both normal and double exponential distributions are symmetric the interval will also be symmetric about the mean of the distribution, which is not the case with the asymmetric distribution.

To judge the significance of the ACF of the residuals, we use the result of Box and Pierce (1970) for the distribution of the residual ACF. We have obtained a similar result for the residual PACF in the AR(1) model.

Numerical results. We summarize the coverage percentage of both methods in Tables 1 and 2.

For $n = 100$, it is observed that the percentages of the 101th observation covered by the 90% P.I.'s obtained by both methods range from 87–92% in both processes. No significant difference is found in the performance of the methods for FOAGP and FOADP (Figure 1). The coverage percentages of the simple method are observed to be somewhat lower than those of the B-J method only in the neighborhood of $\rho = -0.9$ and $\rho = 0.9$ in both processes. The differences are less than 2.5% in both processes. The average widths of P.I. by the B-J method are found to be narrower than those of the simple method. For small ρ values around 0, the difference is slight, while the difference gets larger as the absolute value of ρ gets larger. This is expected since the criterion in the B-J method is to minimize the mean-square errors, which in turn results in the minimization of

TABLE 1
Coverage percentages of one-step-ahead P.I. Normal distribution (FOAGP)

n / ρ	10 (80%)		20 (90%)		100 (90%)	
	B-J	Simple	B-J	Simple	B-J	Simple
− 0.9	0.7275	0.8000	0.8675	0.8825	0.9100	0.8900
−0.8	0.6900	0.8075	0.8850	0.9075	0.8975	0.9050
−0.7	0.7125	0.7950	0.8425	0.9050	0.8875	0.9050
−0.6	0.7050	0.7925	0.8775	0.9300	0.8800	0.9125
−0.5	0.7225	0.8100	0.8625	0.8950	0.8875	0.8850
−0.4	0.7250	0.8200	0.8575	0.9175	0.8900	0.9025
−0.3	0.7400	0.8500	0.8700	0.9275	0.8950	0.8950
−0.2	0.7400	0.8625	0.8375	0.9050	0.8825	0.8875
−0.1	0.7450	0.8475	0.8525	0.9025	0.8950	0.8950
0.0	0.7300	0.8100	0.8600	0.8850	0.9000	0.8875
0.1	0.7375	0.8075	0.8275	0.8975	0.9175	0.9150
0.2	0.6900	0.8075	0.8650	0.8975	0.8700	0.8825
0.3	0.7350	0.8250	0.8850	0.9225	0.8975	0.8950
0.4	0.7050	0.8050	0.8475	0.8750	0.9050	0.9050
0.5	0.7075	0.7675	0.8775	0.8925	0.8850	0.8725
0.6	0.6975	0.7600	0.8400	0.8625	0.9000	0.8925
0.7	0.7075	0.7300	0.8650	0.8900	0.9100	0.9050
0.8	0.7100	0.7125	0.8575	0.8625	0.9175	0.8975
0.9	0.7100	0.7325	0.8550	0.7975	0.9050	0.8800

TABLE 2
Coverage percentages of one-step-ahead P.I. Double exponential distribution (FOADP)

n / ρ	10 (80%)		20 (90%)		100 (90%)	
	B-J	Simple	B-J	Simple	B-J	Simple
− 0.9	0.8550	0.8575	0.9075	0.9325	0.9250	0.9025
−0.8	0.8150	0.8425	0.8800	0.8975	0.9325	0.9125
−0.7	0.7700	0.8375	0.8750	0.9050	0.8800	0.8875
−0.6	0.7525	0.8125	0.8525	0.8900	0.8975	0.8900
−0.5	0.7475	0.8075	0.8950	0.9175	0.8975	0.8975
−0.4	0.7975	0.8350	0.8675	0.8900	0.8825	0.8700
−0.3	0.7475	0.8100	0.8675	0.8950	0.8800	0.8725
−0.2	0.7750	0.8625	0.8575	0.8875	0.8575	0.8700
−0.1	0.7650	0.8575	0.8425	0.8950	0.8975	0.8725
0.0	0.7775	0.8375	0.8875	0.9400	0.8925	0.8925
0.1	0.7725	0.8600	0.8400	0.8975	0.9075	0.8825
0.2	0.7450	0.8375	0.8975	0.9275	0.9250	0.9275
0.3	0.7600	0.8275	0.8575	0.8825	0.9000	0.8825
0.4	0.7775	0.8125	0.8375	0.8925	0.8950	0.8950
0.5	0.7650	0.8025	0.8650	0.8775	0.9000	0.9175
0.6	0.7225	0.7950	0.8325	0.8850	0.9050	0.8850
0.7	0.7625	0.7725	0.8675	0.9100	0.8900	0.8825
0.8	0.8050	0.7600	0.8675	0.8700	0.8850	0.8850
0.9	0.8250	0.6250	0.8850	0.8050	0.9075	0.8850

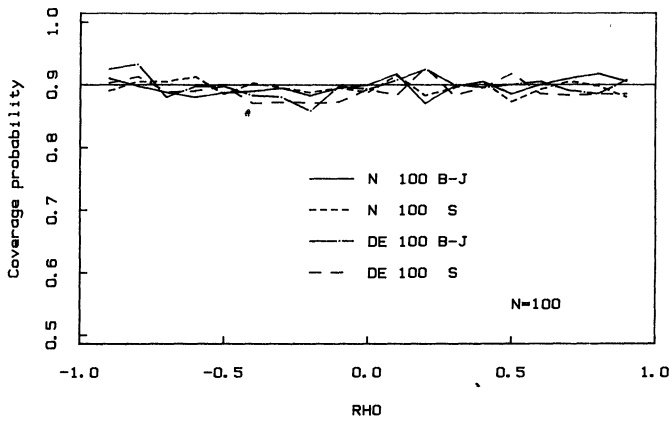

FIG. 1. *B-J method versus simple method one-step-ahead P.I.*, $n = 100$.

the width of P.I. Variance of width is also found to be larger in the simple method than in the B-J method. The same pattern is observed, i.e., for small ρ's there is no significant difference but as $|\rho|$ gets larger the variance gets larger.

To judge the significance of the residual ACF's or PACF's we use intervals of the form est. $\pm$ 2 s.e. Standard errors (s.e.'s) are obtained from

$$V[\hat{r}_1] = V[\hat{v}_1] = \rho/n,$$
$$V[\hat{r}_2] = V[\hat{v}_2] = (1 - \rho^2 + \rho^4)/n, \tag{4.11}$$

and

$$V[\hat{r}_k] = V[\hat{v}_k] \doteq 1/n, \quad \text{for } k = 3,4,\ldots,$$

where $\hat{r}_k$ and $\hat{v}_k$ stand for the residual ACF and PACF, respectively, and ρ must be estimated. It is noted that about 30–40% of the residual ACF's or PACF's we obtained after AR(1) fitting are found to be significant (Table 3). The percentage of significance is about the same over all lags. No significant difference is observed in the percentages of significance in both processes when the B-J method is used. The consequence of a high percentage of significant residual correlations is that one would too often be tempted to fit higher-order models to the data if one were not in a situation where the model is given.

For $n = 20$, it is observed that the performance of the simple method is better than that of the B-J method over all the values of ρ except $\rho = 0.9$ (Figure 2) in both processes. In contrast to the $n = 100$ case, the simple method outperforms the B-J method when $n = 20$. About 90% of the 21st observations are captured by the 90% P.I. (minimum and maximum of 20 observations) using the simple method except for $\rho = 0.9$, while the 90% P.I.'s capture only about 83–88% of the 21st observation by the B-J method. The coverage percentage of the simple method is worse than that of the B-J method only for the case of $\rho = 0.9$. As in the case of $n = 100$, the average width of P.I.'s by the simple method is larger than that of the B-J method.

TABLE 3
Percentage of significant ACF or PACF by the B-J method

Distribution	Normal distribution (FOAGP)			Double exponential distribution (FOADP)		
$\rho \backslash n$	10	20	100	10	20	100
−0.9	0.0950	0.1725	0.2950	0.0625	0.1525	0.3625
−0.8	0.0875	0.1850	0.3350	0.0675	0.1850	0.3575
−0.7	0.0975	0.1850	0.3425	0.0625	0.1425	0.3675
−0.6	0.1000	0.2075	0.3725	0.0650	0.1550	0.3400
−0.5	0.0800	0.2000	0.3725	0.1075	0.1925	0.3900
−0.4	0.1075	0.1975	0.3600	0.1075	0.2000	0.3750
−0.3	0.1125	0.1900	0.3325	0.1000	0.1550	0.3725
−0.2	0.0975	0.1875	0.3775	0.1225	0.1675	0.3625
−0.1	0.1000	0.1350	0.3650	0.1150	0.1850	0.3800
0.0	0.0925	0.1875	0.3775	0.0950	0.1725	0.3550
0.1	0.0925	0.1700	0.3700	0.1250	0.1925	0.3575
0.2	0.1000	0.1775	0.3825	0.1275	0.1750	0.3550
0.3	0.1000	0.2025	0.3825	0.0850	0.1675	0.3675
0.4	0.1200	0.1900	0.3400	0.1050	0.1900	0.3500
0.5	0.1200	0.2025	0.3625	0.1275	0.1575	0.3950
0.6	0.0850	0.1600	0.3725	0.1150	0.1700	0.3325
0.7	0.1100	0.2200	0.3650	0.0950	0.1525	0.3475
0.8	0.1625	0.1600	0.3975	0.0975	0.1775	0.3325
0.9	0.1550	0.1775	0.3600	0.0750	0.1775	0.3975

About 14–22% of the residual ACF's or PACF's are found to be significant (Table 3). Since we used (4.11) to judge the significance of residual ACF's or PACF's, for small n, it is hard to obtain the significance of the residual ACF's or PACF's, especially, at larger lags. To check the effect of (4.11), we use

$$V[\hat{r}_k] = V[\hat{v}_k] \doteq 1/n, \quad \text{for } k = 1, 2, \ldots . \tag{4.12}$$

This time about 8–15% of the residual ACF's or PACF's are found to be significant. Though we achieved some reduction in the percentage of the significant ACF's or PACF's using (4.12) this still implies that, if we followed the B-J script, we would reject the model too often. No significant difference is observed between the FOAGP and FOADP processes.

The conclusions for $n = 10$ are about the same as for $n = 20$. The performance of the simple method is better overall than that of the B-J method (Figure 3). In FOAGP, the coverage percentage by the simple method is always better than that of the B-J method for all ρ's. The 80% P.I. by the simple method covers more than 80% of the 11th observation except for the case of $\rho = 0.5$–0.9, while the 80% P.I. by the B-J method covers only about 70–74% of the 11th observation. In FOADP, the coverage percentage by the simple method is better than that of the B-J method except for $\rho = 0.8$ and 0.9. The 80% P.I. by the simple method covers more than 80% except for the case of $\rho = 0.7$–0.9, while the coverage percentage of the 80% P.I. by the B-J method is less than 80% except for the case of $\rho = 0.8$ or 0.9. The performance of the B-J method is better

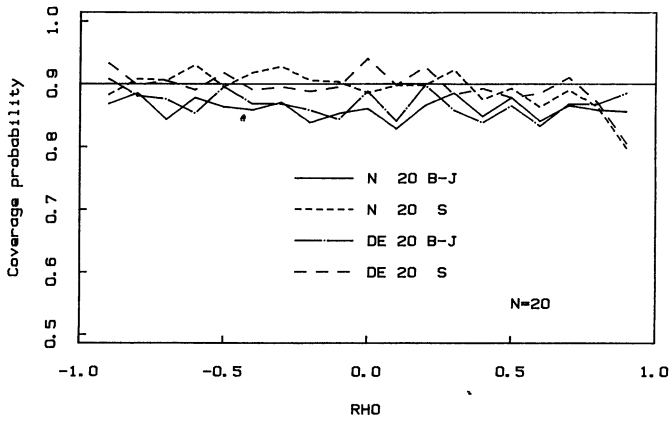

FIG. 2. *B-J method versus simple method one-step-ahead P.I.*, $n = 20$.

than that of the simple method only for $\rho = 0.8$ and 0.9 in FOADP. About 6–16% of the residual ACF's or PACF's are found to be significant using (4.11). When (4.12) is used only 2–4% of the residual ACF's or PACF's are observed to be significant.

Since both methods are based on asymptotic results, the coverage percentage of P.I. is not expected to be consistent with the asymptotic coverage probability for small samples of size like $n = 10$ or 20. We noted that the performance of the simple method is relatively good for either $\rho < 0$ or small positive ρ's for small samples. For large positive ρ's like 0.7–0.9, the coverage is lower than the asymptotic coverage probability. Also, the width of the P.I. varies quite a lot for

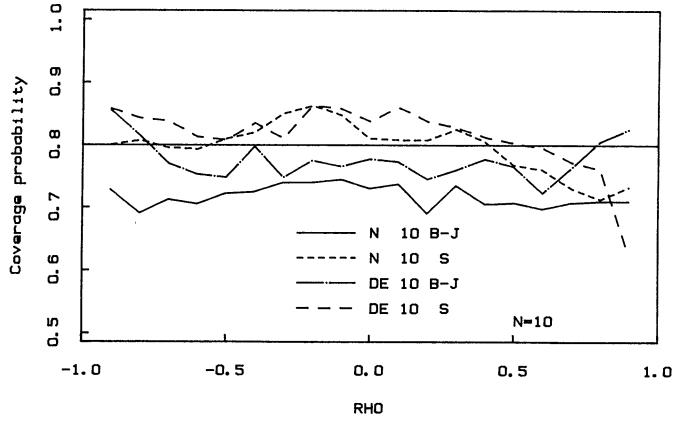

FIG. 3. *B-J method versus simple method one-step-ahead P.I.*, $n = 10$.

small samples. The width is stable for intermediate ρ's like -0.7–0.7 and increases quickly for a large $|\rho|$. This is not the case for a large sample like $n = 100$, where the width of P.I. is much more stable in both processes and the coverage probability is also conformable to the asymptotic results.

From the above simulation study, we conclude that we can use the simple method for both large and small samples. As is expected, the B-J method does not perform well for small samples of size $n = 10$ or 20. The simple method does not perform well for large positive ρ's like 0.7–0.9 when the sample size is small. But this problem is not confined to the simple method. We observed that the B-J method also has some trouble for these ρ's. While neither method can be applied without risk, the simple method appears to be a better bet in small sample situations.

REFERENCES

BABU, G. J. and SINGH, K. (1978). On deviations between empirical and quantile processes for mixing random variables. *J. Multivariate Anal.* **8** 532–549.

BICKEL, P. J. (1967). Some contributions to the theory of order statistics. *Proc. Fifth Berkeley Symp. Math. Statist. Probab.* **1** 575–591. Univ. California Press.

BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

BOX, G. E. P. and JENKINS, G. M. (1976). *Time Series Analysis: forecasting and control*, 2nd ed. Holden-Day, San Francisco.

BOX, G. E. P. and PIERCE, D. A. (1970). Distribution of residual autocorrelations in autoregressive integrated moving average time series models. *J. Amer. Statist. Assoc.* **65** 1509–1526.

BRANDON, C. H., JARRETT, J. E. and KHUMAWALA, S. B. (1983). Revising forecasts of accounting earnings: A comparison with the Box–Jenkins method. *Management Sci.* **29** 256–263.

BUTLER, R. W. (1982). Nonparametric interval and point prediction using data trimmed by a Grubbs-type outlier rule. *Ann. Statist.* **10** 197–204.

CARBONE, R., ANDERSON, A., CORRIVEAU, Y. and CORSON, P. P. (1983). Comparing for different time series methods: The value of technical expertise, individualized analysis, and judgmental adjustment. *Management Sci.* **29** 559–566.

DEO, C. M. (1973). A note on empirical processes of strong-mixing sequences. *Ann. Probab.* **1** 870–875.

DONSKER, M. (1952). Justification and extension of Doob's heuristic approach to the Kolmogorov–Smirnov theorems. *Ann. Math. Statist.* **23** 277–283.

DOOB, J. L. (1949). Heuristic approach to the Kolmogorov–Smirnov theorems. *Ann. Math. Statist.* **20** 393–403.

GASTWIRTH, J. L. and RUBIN, H. (1975a). Asymptotic distribution theory on the empirical c.d.f. for mixing stochastic processes. *Ann. Statist.* **3** 809–824.

GASTWIRTH, J. L. and RUBIN, H. (1975b). The behavior of robust estimators on dependent data. *Ann. Statist.* **3** 1070–1100.

GASTWIRTH, J. L., RUBIN, H. and WOLFF, S. S. (1967). The effect of autoregressive dependence on a non-parametric test. *IEEE Trans. Inform. Theory* **IT-13** 311–313.

HALL, I. J., PRAIRIE, R. R. and MOTLAGH, C. K. (1975). Non-parametric prediction intervals. *J. Quality Tech.* **7** 109–114.

HOCHSTADT, H. (1961). *Special Functions of Mathematical Physics*. Holt, Rinehart and Winston, New York.

IBRAGIMOV, I. A. and LINNIK, Y. V. (1971). *Independent and Stationary Sequences of Random Variables*. Wolters-Noordhoff, Groningen, The Netherlands.

KEYFITZ, N. (1972). On future population. *J. Amer. Statist. Assoc.* **67** 347–363.

MAKRIDAKIS, S. and HIBON, M. (1979). Accuracy of forecasting: An empirical investigation. *J. Roy. Statist. Soc. Ser. A* **142** 97–145.

MEHRA, K. L. and RAO, M. S. (1975). On functions of order statistics for mixing processes. *Ann. Statist.* **3** 874–883.

PARZEN, E. (1979). Nonparametric statistical data modelling. *J. Amer. Statist. Assoc.* **74** 105–131.

PROHOROV, Y. V. (1956). Convergence of random processes and limit theorems in probability theory. *Theory Probab. Appl.* **1** 157–214.

SEN, P. K. (1971). A note on weak convergence of empirical processes for sequences of ϕ-mixing random variables. *Ann. Math. Statist.* **42** 2131–2133.

SKOROHOD, A. V. (1956). Limit theorems for stochastic processes. *Theory Probab. Appl.* **1** 262–290.

DEPARTMENT OF COMPUTER SCIENCE AND STATISTICS
SEOUL NATIONAL UNIVERSITY
SEOUL 151
KOREA

GRADUATE SCHOOL OF BUSINESS
UNIVERSITY OF WISCONSIN, MADISON
1155 OBSERVATORY DRIVE
MADISON, WISCONSIN 53706