Estimating Properties of Autoregressive Forecasts

ROBERT A. STINE*

Forecasting requires estimates of the error of prediction; however, such estimates for autoregressive forecasts depend nonlinearly on unknown parameters and distributions. Substitution estimators of mean squared error (MSE) possess bias that varies with the underlying model, and Gaussian-based prediction intervals fail if the data are not normally distributed. This article proposes methods that avoid these problems. A second-order Taylor expansion produces an estimator of MSE that is unbiased and leads to accurate prediction intervals for Gaussian data. Bootstrapping also suggests an estimator of MSE, but it is approximately the problematic substitution estimator. Bootstrapping also yields prediction intervals, however, whose coverages are invariant of the sampling distribution and asymptotically approach the nominal content. Parameter estimation increases the error in autoregressive forecasts. This additional error inflates one-step prediction mean squared error (PMSE) by a factor of $1 + p/T$, where p is the number of parameters and T is the series length (Bloomfield 1972; Box and Jenkins 1976; Davisson 1965). The increase at greater extrapolation involves parameters of the process (Yamamoto 1976). Simple substitution estimators of squared error possess bias that can dominate attempts to estimate the inflation. The proposed bias-corrected estimator alleviates this problem. For example, in simulations of short series ($T = 24$) from a first-order model with a coefficient of .8 (.4), the substitution estimator at forecast leads 2 and 3 underestimated the true PMSE by 4.7% and 6.4% (−.4% and −2.2%). The corrected estimator erred by less than .5%. Prediction intervals based on the bootstrap are preferable unless the sampling distribution is known. The bias-corrected estimator of PMSE leads to a very accurate prediction interval for Gaussian data, but its coverage depends on the normality assumption. A bootstrap interval asymptotically approaches the desired coverage but is less efficient. For non-Gaussian data, only the bootstrap intervals necessarily tend toward the correct coverage. Although numerical results show bootstrap intervals tend to lack the nominal coverage by several percentages, this deficiency is consistent across sampling distributions and rapidly decays with increasing series length.

KEY WORDS: Bootstrap; Least squares; Mean squared error; Prediction interval; Simulation.

1. INTRODUCTION

This article proposes and compares several strategies for obtaining estimates of prediction mean squared error (PMSE) and prediction intervals. These forecast measures are intimately related, since one commonly uses an estimate of PMSE to set the width of a prediction interval. A more traditional Gaussian strategy uses an improved estimator of PMSE to get an interval that obtains the nominal coverage even with short Gaussian time series. A bootstrap approach eschews the Gaussian focus on squared error and uses a distribution estimate to find an interval that performs well over a range of distributions.

My focus is on prediction intervals and associated estimators of squared error. A prediction interval is a random interval, say $I(\mathbf{Y})$, based on observations $\mathbf{Y}$ and designed to capture a future value Y_f with chosen probability, the nominal coverage. For example, if $I(\mathbf{Y}) = [L(\mathbf{Y}), U(\mathbf{Y})]$, then we might want $\Pr\{L(\mathbf{Y}) \le Y_f \le U(\mathbf{Y})\} = .8$. This criterion is equivalent to requiring that the mean of an *unobservable* random variable be .8:

$$E_{\mathbf{Y}}\left\{\int_{L(\mathbf{Y})}^{U(\mathbf{Y})} dF\,(Y_f \mid \mathbf{Y})\right\} = E\{C(\mathbf{Y})\} = .8. \quad (1.1)$$

The random variable $C(\mathbf{Y})$ is the coverage of the interval. We usually attempt to obtain (1.1) by forming an estimator of the conditional distribution of Y_f given $\mathbf{Y}$, often by assuming normality and centering the interval about some prediction. An alternative approach that seeks a specific conditional expectation for $C(\mathbf{Y})$ appears in Section 6.

Section 2 reviews autoregressive models and describes the PMSE of their forecasts. Parametric, Gaussian-oriented procedures appear in Section 3. The bootstrap follows in Section 4. A numerical comparison of these methods with short series appears in Section 5. Finally, a conditional approach appears in Section 6.

2. AUTOREGRESSIVE MODELS AND FORECASTS

Assume that the observed time series $\mathbf{Y} = (y_1, \ldots, y_T)'$ is a segment from an autoregressive (AR) process of known order p (AR(p)),

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \cdots + \alpha_p y_{t-p} + \varepsilon_t, \quad (2.1)$$

where stationarity constrains the coefficients $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_p)'$ so that the roots of the polynomial $1 - \sum \alpha_j z^j$ lie outside the unit circle. The errors $\{\varepsilon_t\}$ are independent random variables with distribution F, mean 0, and variance σ^2. The observations also have mean 0, as in many studies of PMSE. If $\mathbf{Y}_t = (y_t, y_{t-1}, \ldots, y_{t-p+1})'$, then (2.1) becomes $\mathbf{Y}_t = \mathbf{A}\mathbf{Y}_{t-1} + \mathbf{J}\varepsilon_t$, where the vector $\mathbf{J}$ has a leading 1 followed by $(p - 1)$ zeros and the $(p \times p)$ matrix $\mathbf{A}$ is

$$\mathbf{A} = \begin{bmatrix} & \boldsymbol{\alpha}' & \\ \mathbf{I}_{p-1} & & \mathbf{0} \end{bmatrix},$$

where $\mathbf{I}_k$ denotes a $(k \times k)$ identity matrix. Given $\boldsymbol{\alpha}$ and $\mathbf{Y}$ (with $T \ge p$), the minimum MSE forecast of y_{T+f} is $\tilde{y}_{T+f} = \boldsymbol{\theta}_f'\mathbf{Y}_T$, where $\boldsymbol{\theta}_f$ is the first row of $\mathbf{A}^f$. The PMSE of $\tilde{y}_{T+f}$ is

$$\mathrm{PMSE}(\tilde{y}_{T+f}) = \sigma^2\left(\sum_0^{f-1} \omega_j^2\right) := \sigma^2 W_f^2,$$

where ω_j is the jth coefficient in the infinite moving average representation of the processes, $y_t = \sum \omega_j \varepsilon_{t-j}$. A summary of these basic results appears in Fuller (1976, chap. 2).

The least squares method provides an asymptotically efficient estimator of $\boldsymbol{\alpha}$. This estimator is

$$\hat{\boldsymbol{\alpha}} = (\textstyle\sum \mathbf{Y}_{t-1}\mathbf{Y}_{t-1}')^{-1}(\sum \mathbf{Y}_{t-1}y_t),$$

* Robert A. Stine is Assistant Professor, Department of Statistics, Wharton School, University of Pennsylvania, Philadelphia, PA 19104. This research was partially supported by National Heart, Lung, and Blood Institute Grants HL19869 and HL26898. The author appreciates the helpful comments of referees who improved the clarity and accuracy of the presentation.

Journal of the American Statistical Association
December 1987, Vol. 82, No. 400, Theory and Methods

where the sums run from $p + 1$ to T. Associated with $\hat{\boldsymbol{\alpha}}$ are $(T - p)$ residuals, $\hat{\varepsilon}_t = y_t - \hat{\boldsymbol{\alpha}}'\mathbf{Y}_{t-1}$ $(t = p + 1, \ldots, T)$, that lead to an estimator of σ^2,

$$\hat{\sigma}^2 = \sum_{p+1}^{T} \hat{\varepsilon}_t^2/(T - 2p). \tag{2.2}$$

This estimator is virtually unbiased, $E(\hat{\sigma}^2) = \sigma^2 + o(1/T)$ (Shaman 1983), and nearly uncorrelated with $\hat{\boldsymbol{\alpha}}$, since $\operatorname{cov}(\hat{\alpha}_j, \hat{\sigma}^2) = o(1/T)$, $j = 1, \ldots, p$ (Stine 1982).

The PMSE grows when $\hat{\boldsymbol{\alpha}}$ replaces $\boldsymbol{\alpha}$ in forecasts. The estimated predictor is computed by replacing $\boldsymbol{\alpha}$ in the first row of $\mathbf{A}$ by $\hat{\boldsymbol{\alpha}}$ so that $\hat{y}_{T+f} = \hat{\mathbf{A}}^f\mathbf{Y}_T$, where I use a "^" to denote substitution estimators. The PMSE of $\hat{y}_{T+f}$ exceeds that of $\tilde{y}_{T+f}$ by a term of order $O(1/T)$,

$$\begin{aligned}\operatorname{PMSE}(\hat{y}_{T+f}) &= E\left\{\left(\sum_{0}^{f-1} \omega_j \varepsilon_{T+f-j}\right) + (\boldsymbol{\theta}_f - \hat{\boldsymbol{\theta}}_f)'\mathbf{Y}_T\right\}^2 \\ &= \sigma^2\{W_f^2 + \eta_f/T + o(1/T)\}. \end{aligned} \tag{2.3}$$

Under some general conditions, the inflation term η_f is

$$\eta_f = \sigma^2 \sum_{j,k=0}^{f-1} \sum \omega_j\omega_k \operatorname{tr}(\boldsymbol{\Gamma}\mathbf{A}^{f-k-1\prime}\boldsymbol{\Gamma}^{-1}\mathbf{A}^{f-j-1}), \tag{2.4}$$

where tr denotes the trace operator and $\boldsymbol{\Gamma} = E(\mathbf{Y}_t\mathbf{Y}_t')$ (Fuller and Hasza 1981; Kunitomo and Yamamoto 1985). Thus PMSE can rise and then fall as forecasts are extrapolated into the future.

3. THE GAUSSIAN MODEL

The methods of this section assume that the errors in (2.1) follow a normal distribution:

Assumption 1. $\varepsilon_t \sim N(0, \sigma^2)$ for all t.

If a prediction interval is centered on $\hat{y}_{T+f}$, then a plausible estimator of the scale of the conditional distribution of Y_{T+f} is $\hat{\sigma}^2(\hat{W}_f^2 + \hat{\eta}_f/T)$ in which we replace $\boldsymbol{\alpha}$ in (2.3) and (2.4) by $\hat{\boldsymbol{\alpha}}$ (Fuller and Hasza 1981). This estimator is biased, however, and leads to intervals that lack the nominal average coverage.

A second-order Taylor series expansion suggests a simple, improved estimator of PMSE that corrects for the sources of bias described by Ansley and Newbold (1981). Let $\hat{\boldsymbol{\omega}}_f = (1, \hat{\omega}_1, \ldots, \hat{\omega}_{f-1})'$ denote the estimators of $\omega_1, \ldots, \omega_{f-1}$ based on $\hat{\boldsymbol{\alpha}}$. Viewing $\boldsymbol{\omega}_f$ as a function of $\boldsymbol{\alpha}$ gives

$$\begin{aligned}\hat{\boldsymbol{\omega}}_f &= \omega_f(\boldsymbol{\alpha}) + \mathbf{M}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) \\ &\quad + \{\mathbf{I}_f \otimes (\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})'\}\mathbf{H}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) + \mathbf{R}, \end{aligned} \tag{3.1}$$

where $\mathbf{M}$ is the $(f \times p)$ array of first derivatives,

$$m_{i,j} = \partial\omega_i(\boldsymbol{\alpha})/\partial\alpha_j|_{\alpha}, \qquad i = 0, \ldots, f - 1, \quad j = 1, \ldots, p,$$

and $\mathbf{H}$ is $\frac{1}{2}$ times the $(fp \times p)$ array of second derivatives partitioned as

$$\mathbf{H}' = \tfrac{1}{2}\,(\mathbf{H}_0|\mathbf{H}_1 \mid \cdots \mid \mathbf{H}_{f-1}).$$

The (i, j) element of $\mathbf{H}_k$ is then

$$h_k(i, j) = \partial^2\omega_k(\boldsymbol{\alpha})/\partial\alpha_i\partial\alpha_j|_{\alpha}, \qquad i, j = 1, \ldots, p.$$

In (3.1) $\otimes$ denotes the Kronecker product and $E(R_i) = o(1/T)$, $i = 0, \ldots, f - 1$. From this expansion, we find (Stine 1982) that

$$E(\hat{\sigma}^2\hat{W}_f^2) = \sigma^2(W_f^2 + D_f/T) + o(1/T) \tag{3.2}$$

and

$$\begin{aligned}\operatorname{var}(\hat{\sigma}^2\hat{W}_f^2) &= (2\sigma^4/T)\{(W_f^2)^2 \\ &\quad + 2\sigma^2\boldsymbol{\omega}_f'\mathbf{M}\boldsymbol{\Gamma}^{-1}\mathbf{M}'\boldsymbol{\omega}_f\} + o(1/T), \end{aligned} \tag{3.3}$$

where the bias term is

$$D_f = \sigma^2\operatorname{tr}(\mathbf{M}'\mathbf{M}\boldsymbol{\Gamma}^{-1}) + 2\boldsymbol{\delta}'\mathbf{M}'\boldsymbol{\omega}_f + \sigma^2 \sum_{0}^{f-1} \omega_j\operatorname{tr}(\mathbf{H}_j\boldsymbol{\Gamma}^{-1}). \tag{3.4}$$

The vector $\boldsymbol{\delta}$ in (3.4) is the $O(1/T)$ bias of $\hat{\boldsymbol{\alpha}}$, $E(\hat{\boldsymbol{\alpha}}) = \boldsymbol{\alpha} + \boldsymbol{\delta}/T + o(1/T)$, and simple expressions for $\boldsymbol{\delta}$ exist for low-order models [Tanaka 1984; Yamamoto and Kunitomo 1984, whose eq. (17) ought to read $\operatorname{ABIAS}[\hat{\beta}_2] = -(1 + 3\beta_2)/n$]. Patterns in $\mathbf{M}$ and $\mathbf{H}$ also simplify calculation of (3.2)–(3.4). The matrix $\mathbf{M}$ is banded; m_{ij} only depends on $(i - j)$ and $m_{ij} = 0$ for $j \geq i$ (Godolphin 1977). Similarly, the $\mathbf{H}_i$ are related. For example, $h_k(i, i) = h_{k+1}(i, i + 1) = h_{k-2}(i - 1, i - 1)$. These adjustments lead to the interval

$$I_G(\mathbf{Y}; \beta, f) = [\hat{y}_{T+f} \pm t_{\beta,\mathrm{df}}\hat{\sigma}\{\hat{W}_f^2 + (\hat{\eta}_f - \hat{D}_f)/T\}^{1/2}],$$

where $t_{\beta,\mathrm{df}}$ is the two-sided β percentile of a t distribution with df degrees of freedom.

An empirical estimator of the degrees of freedom improves this interval. Analogy to regression suggests using the divisor $(T - 2p)$ of $\hat{\sigma}^2$ for df. For leads $f > 1$, however, the width of I_G is a stochastic multiple of $\hat{\sigma}$. Borrowing from spectral estimation (Koopmans 1974, chap. 8), consider using the equivalent degrees of freedom and let df be the ratio of twice the square of the mean of $\hat{\sigma}^2\hat{W}_f^2$ to its variance (ignoring the smaller-order term involving $\hat{\eta}_f$). A straightforward estimator of this ratio is

$$\mathrm{edf} := \left[\frac{(T - 2p)(\hat{W}_f^2 + \hat{D}_f/T)^2}{\{(\hat{W}_f^2)^2 + 2\hat{\sigma}^2\hat{\boldsymbol{\omega}}_f'\hat{\mathbf{M}}\hat{\boldsymbol{\Gamma}}^{-1}\hat{\mathbf{M}}'\hat{\boldsymbol{\omega}}_f\}} + .5\right],$$

where $[x]$ is the greatest integer less than or equal to x and edf represents equivalent degrees of freedom. Note that $\mathrm{edf} \equiv (T - 2p)$ when $f = 1$. Without edf, the average coverage of I_G drops below β as f increases. In a simulation of 1,000 Gaussian series ($\alpha = .6$, $\beta = .95$, $T = 24$), the average coverage of I_G with $\mathrm{df} = T - 2p$ was .949, .945, and .943 (standard error < .002) for the first three leads. By comparison, the average coverage increases to .950 and .951 (leads 2, 3) with edf.

4. BOOTSTRAP METHODS

We cannot expect the Gaussian intervals of Section 3 to perform very well when the errors are not normally distributed. Since I_G becomes $[\hat{y}_{T+f} \pm z_\beta\sigma W_f]$ for large T, its average coverage can be quite far from β, especially when β differs from .95 (Sharpe 1970). Bootstrap resampling allows a more general model in which we assume a symmetric error distribution with finite moments:

Assumption 2. $\Pr(\varepsilon_t \le x) = F(x) = 1 - F(-x)$, F is continuous and strictly increasing, and $E(|\varepsilon_t|^k) < \infty$, for $k = 1, 2, \ldots$.

The assumption of finite moments implies that the PMSE results of Section 2 hold, and the smoothness insures that the coverage of bootstrap intervals converges to the nominal amount. The basic idea of bootstrapping is to generate replicates $\mathbf{Y}^*$ of the observed data $\mathbf{Y}$ and extend these replicates into the "future." We can then compare forecasts from the first T values of $\mathbf{Y}^*$ to y^*_{T+1}, y^*_{T+2}, and so forth. The resulting, observable prediction errors provide the basis for estimating PMSE and prediction intervals.

Given $\hat{\boldsymbol{\alpha}}$, generating $\mathbf{Y}^*$ proceeds by recursively mimicking nature. Begin by initially choosing a block of p consecutive observations from $\mathbf{Y}$, labeled $\mathbf{Y}^*_0 = (y^*_{-p+1}, \ldots, y^*_0)'$. The recursion parallels (2.1), with $\hat{\boldsymbol{\alpha}}$ in place of $\boldsymbol{\alpha}$ and ε_t replaced by ε^*_t drawn from some estimate of F,

$$y^*_t = \hat{\alpha}_1 y^*_{t-1} + \cdots + \hat{\alpha}_p y^*_{t-p} + \varepsilon^*_t, \qquad t = 1, \ldots, T. \quad (4.1)$$

Repeating this recursion B times leads to a set of bootstrap series that I denote by $\mathbf{Y}^{*(1)}, \ldots, \mathbf{Y}^{*(B)}$. For small T and moderate p, the restriction to $T - p$ possible initial vectors might be considered limiting. One can obtain a slightly richer collection of initial values with $\mathbf{Y}^*_0 = \hat{\mathbf{C}}^{1/2}\boldsymbol{\varepsilon}^*$, where $\hat{\mathbf{C}}^{1/2}$ is the Cholesky square root of the estimate of $\boldsymbol{\Gamma}/\sigma^2$ defined by $\hat{\boldsymbol{\alpha}}$ and $\boldsymbol{\epsilon}^* = (\varepsilon^*_{1-p}, \ldots, \varepsilon^*_0)'$ is a sample from the estimate of F. The infinite moving-average form, as in $y^*_0 = \sum_j \hat{\omega}_j \varepsilon^*_{-j}$, leads to even more starting values. In Section 5 I use the block initialization, since the alternatives require more computation (and truncation of an infinite sum) and appear to have no real effect on these applications, even with T near 20. Further discussion appears in Efron and Tibshirani (1986) and Findley (1985).

Given a reasonable estimator of $\boldsymbol{\alpha}$, the key to bootstrapping is the distribution from which $\{\varepsilon^*_t\}$ are sampled. Rather than use the empirical distribution, Assumption 2 allows for a more efficient, symmetric estimator that possesses certain minimax properties (Millar 1979; Schuster 1975). In addition, a scaling adjustment gives the ε^*_t the desired variance:

$$\begin{aligned} F_T(x) &= .5 + \#(k|\hat{\varepsilon}_t| \le x)/2(T-p), \\ &\qquad\qquad x \ge 0, \quad t = p+1, \ldots, T, \\ &= 1 - F_T(-x), \qquad x < 0, \end{aligned} \quad (4.2)$$

where $k = \{(T-p)/(T-2p)\}^{1/2}$. Thus if E_* denotes expectation under F_T given $\{\hat{\varepsilon}_{p+1}, \ldots, \hat{\varepsilon}_T\}$ and $\varepsilon^*_t \sim F_T$, then $E_*(\varepsilon^*_t) = 0$ and $E_*(\varepsilon^*_t)^2 = \hat{\sigma}^2$, the nearly unbiased estimator (2.2) of σ^2.

Since the bootstrap errors have mean 0, no drift is introduced into $\mathbf{Y}^*$ as would occur if $\{\varepsilon^*_t\}$ were drawn from an uncentered distribution. If, for example, one were to resample the residuals $\{\hat{\varepsilon}_t\}$, the bootstrap recursions in an AR(1) model would be $y^*_t = \hat{\alpha} y^*_{t-1} + \varepsilon^*_t = \hat{\alpha} y^*_{t-1} + \bar{\varepsilon} + e^*_t$, where $\bar{\varepsilon} = \sum \hat{\varepsilon}_t/(T-p)$. Fitting an AR(1) model to $\mathbf{Y}^*$ without a constant term introduces specification error that can distort bootstrap estimates. An expansion for the asymptotically equivalent Yule–Walker estimator $\tilde{\alpha}$, which is bounded, suggests the problem in bias estimation. Using the approach of Kendall (1954), one can show that $E(\tilde{\alpha} - \alpha)$ is approximately $-3\alpha/T + (1 + \alpha)(\tau/\sigma)^2$ for Gaussian series $y_t = \alpha y_{t-1} + \tau + \varepsilon_t$ and $\tilde{\alpha} = r_1/r_0$, $r_i = \sum_{t=i+1}^T y_t y_{t-i}$. Thus we expect $E_*[\tilde{\alpha}^* - \tilde{\alpha}]$ to be near $-3\tilde{\alpha}/T + (1 + \tilde{\alpha})(1/T) = (1 - 2\tilde{\alpha})/T$, or 0 for $\alpha = .5$. In a simulation of 250 Gaussian series ($\alpha = .5$, $T = 24$), resampling uncentered residuals led to a bias estimate of $-.007$ (with a standard error of .004) as opposed to the value $-.051$ (with a standard error of .002) obtained by sampling F_T. The "correct" answer is $-2\alpha/T = -.042$ to order $o(1/T)$. A similar concern arises in regression analysis (Freedman 1981, p. 1220).

The direct bootstrap estimate of PMSE reduces to the substitution estimator. Let $PE^*_f = y^*_{T+f} - \hat{y}^*_{T+f}$, where $\hat{y}^*_{T+f}$ is the forecast of y^*_{T+f} using the estimator $\hat{\boldsymbol{\alpha}}^*$ calculated from $\mathbf{Y}^*$. These observable prediction errors suggest the bootstrap estimator $\text{PMSE}^*(\hat{y}_{T+f}) = \sum_{b=1}^B (y^{*(b)}_{T+f} - \hat{y}^{*(b)}_{T+f})^2$. Since the parameters of the bootstrap process generating $\mathbf{Y}^*$ are $\hat{\boldsymbol{\alpha}}$ and $\hat{\sigma}^2$, PMSE* approaches (2.3) with these estimators in place of $\boldsymbol{\alpha}$ and σ^2. Thus PMSE* is essentially the naive substitution estimator of Section 3.

Bootstrapping easily extends to prediction intervals. If we knew the distribution $H(x; f)$ of the prediction error $y_{T+f} - \hat{y}_{T+f}$, then one prediction interval would be $[\hat{y}_{T+f} + H^{-1}\{(1-\beta)/2; f\}, \hat{y}_{T+f} + H^{-1}\{(1+\beta)/2; f\}]$, where $H^{-1}(p; f) = \inf\{x\colon H(x; f) \ge p\}$. The obvious bootstrap interval replaces H in this interval by an estimator of the distribution of PE^*_f, such as $\#\{PE^{*(b)}_f \le x\}/B$. Such an estimator ignores the structure of PE^*_f, however, which is a sum of two independent terms,

$$PE^*_f = \sum_{j=0}^{f-1} \hat{\omega}_j e^*_{T+f-j} + (\hat{\boldsymbol{\theta}}_f - \hat{\boldsymbol{\theta}}^*_f)'\mathbf{Y}^*_T. \quad (4.3)$$

Thus we can approximate $H(x; f)^*$ by the convolution of $C^*_T(x; f) = F_T(x) * F_T(x/\hat{\omega}_1) * \cdots * F_T(x/\hat{\omega}_{f-1})$ with an estimate of the distribution of the second term of PE^*_f, such as $W^*_B(x; f) = \#\{(\hat{\boldsymbol{\theta}}_f - \hat{\boldsymbol{\theta}}^{*(b)}_f)'\mathbf{Y}^{*(b)}_T \le x\}/B$. The resulting estimator is

$$H^*_B(x; f) = C^*_T(x; f) * W^*_B(x; f), \quad (4.4)$$

which leads to the bootstrap prediction interval (Stine 1982)

$$\begin{aligned} I_B(\mathbf{Y}; \beta, f) = [&\hat{y}_{T+f} + H^{*-1}_B\{(1-\beta)/2; f\}, \\ &\hat{y}_{T+f} + H^{*-1}_B\{(1+\beta)/2; f\}]. \end{aligned}$$

In practice, it may not be practical to obtain exact order statistics from (4.4); for the simulations of Section 5, sampling (4.4) 200 times gives sufficiently accurate estimates of the desired quantiles. Thus bootstrapping prediction intervals does not require extrapolating $\mathbf{Y}^*$ into the future, though this characterization makes the ideas more clear.

Unlike Gaussian intervals, I_B obtains the nominal coverage without assuming a specific sampling distribution F. This invariance arises because the effects of estimation and bootstrapping are asymptotically negligible. For large T, I_B is essentially a distribution-free tolerance interval based on residuals. For example, let B be arbitrarily large and consider the first lead for an AR(1) model. If we write $I_B = [\hat{y}_{T+1} + L^*(\mathbf{Y}), \hat{y}_{T+1} + U^*(\mathbf{Y})]$, then the average coverage is

$$\begin{aligned} E_Y C(\mathbf{Y}) &= E_Y[P_\varepsilon\{L^*(\mathbf{Y}) + (\hat{\alpha} - \alpha)y_T \leq \varepsilon_{T+1} \\ &\qquad \leq U^*(\mathbf{Y}) + (\hat{\alpha} - \alpha)y_T\}] \\ &= E_Y[F\{U^*(\mathbf{Y}) + (\hat{\alpha} - \alpha)y_T\} \\ &\qquad - F\{L^*(\mathbf{Y}) + (\hat{\alpha} - \alpha)y_T\}]. \end{aligned}$$

Since the variance of $\hat{\alpha}^*$ is $O(1/T)$ (Freedman 1984, theorem 4.1), the effect of bootstrapping (and estimation) in (4.4) is $o_p(1)$ and $H^{*-1} \rightarrow_p F^{-1}$. Thus $U^*(\mathbf{Y}) + (\hat{\alpha} - \alpha)Y_T \rightarrow_p F^{-1}\{(1 + \beta)/2\}$, and the average coverage of I_B

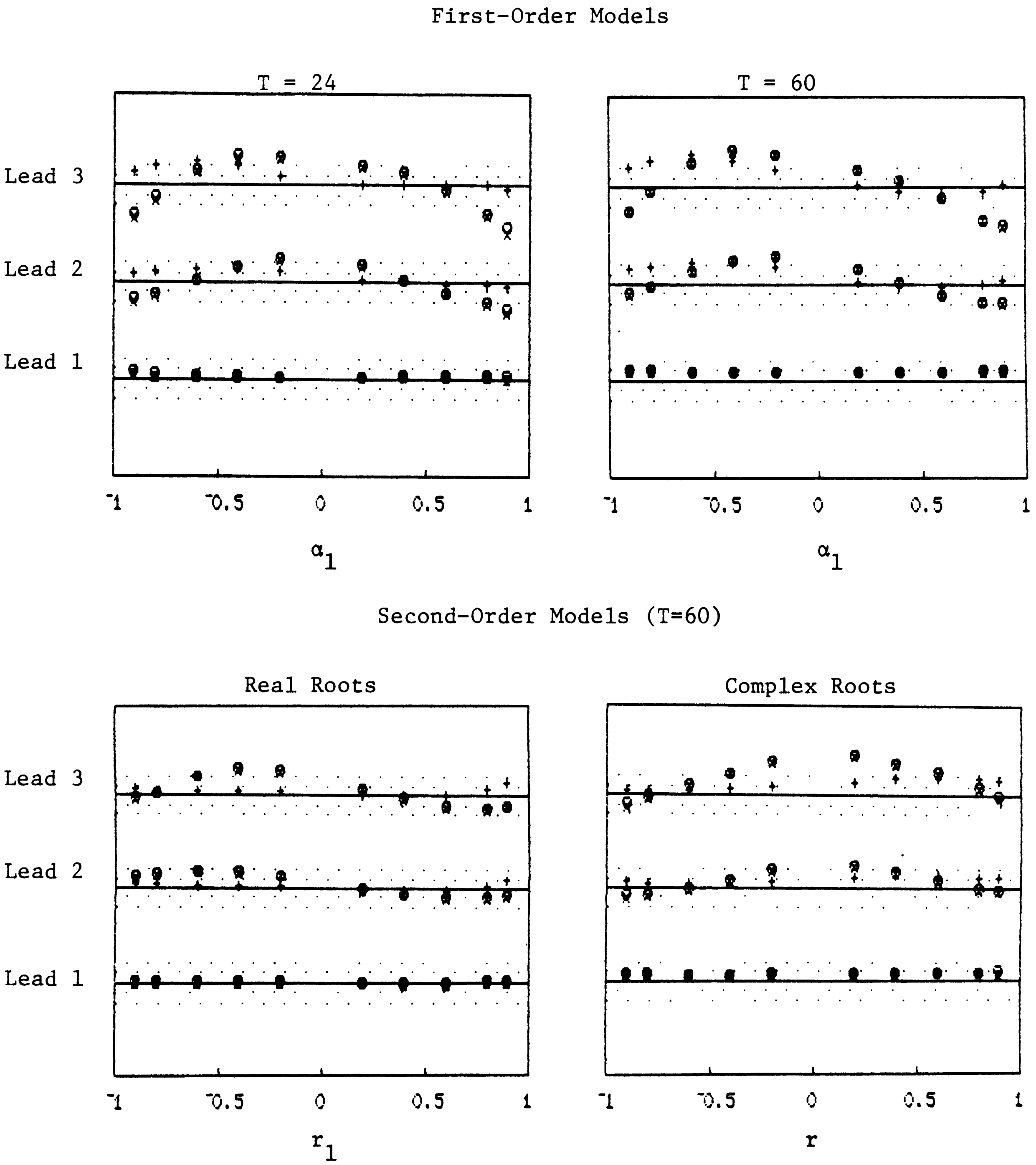

Figure 1. Percentage Error of Substitution (x), Bias-Corrected (+), and Bootstrap (o) Estimates of Prediction Mean Squared Error. Results are from a simulation of 1,000 Gaussian series. Dotted lines are 2% apart for T = 24 and 1% elsewhere.

converges to β. Similar invariance properties obtain in regression models (Stine 1985).

5. COMPARISONS

Comparing these estimators of PMSE illustrates the value of the bias correction and the similarity of the bootstrap to the simple substitution estimator. Figure 1 shows the percentage error 100 × (estimate − true)/true from a simulation of 1,000 realizations of 10 AR(1) and 20 AR(2) Gaussian processes. If r_1 and r_2 are the roots of $z^2 - \alpha_1 z - \alpha_2$, then 10 of the AR(2) models have real roots with $r_2 \equiv .5$ and 10 have complex roots with $r_1 = re^{i\theta}$ and $r_2 = re^{-i\theta}$ for $\theta = \pi/4$. To avoid start-up transients, $\varepsilon_1, \ldots, \varepsilon_p$ were transformed so that $y_1, \ldots, y_p$ have covariance matrix Γ. The shorter series comprised the last 24 observations of the longer so that the results are correlated across sample size. The same random errors were used for all processes in each cell of the figure for more efficient comparisons. The standard error of each estimate is about 1.25% when $T = 24$ and .8% when $T = 60$.

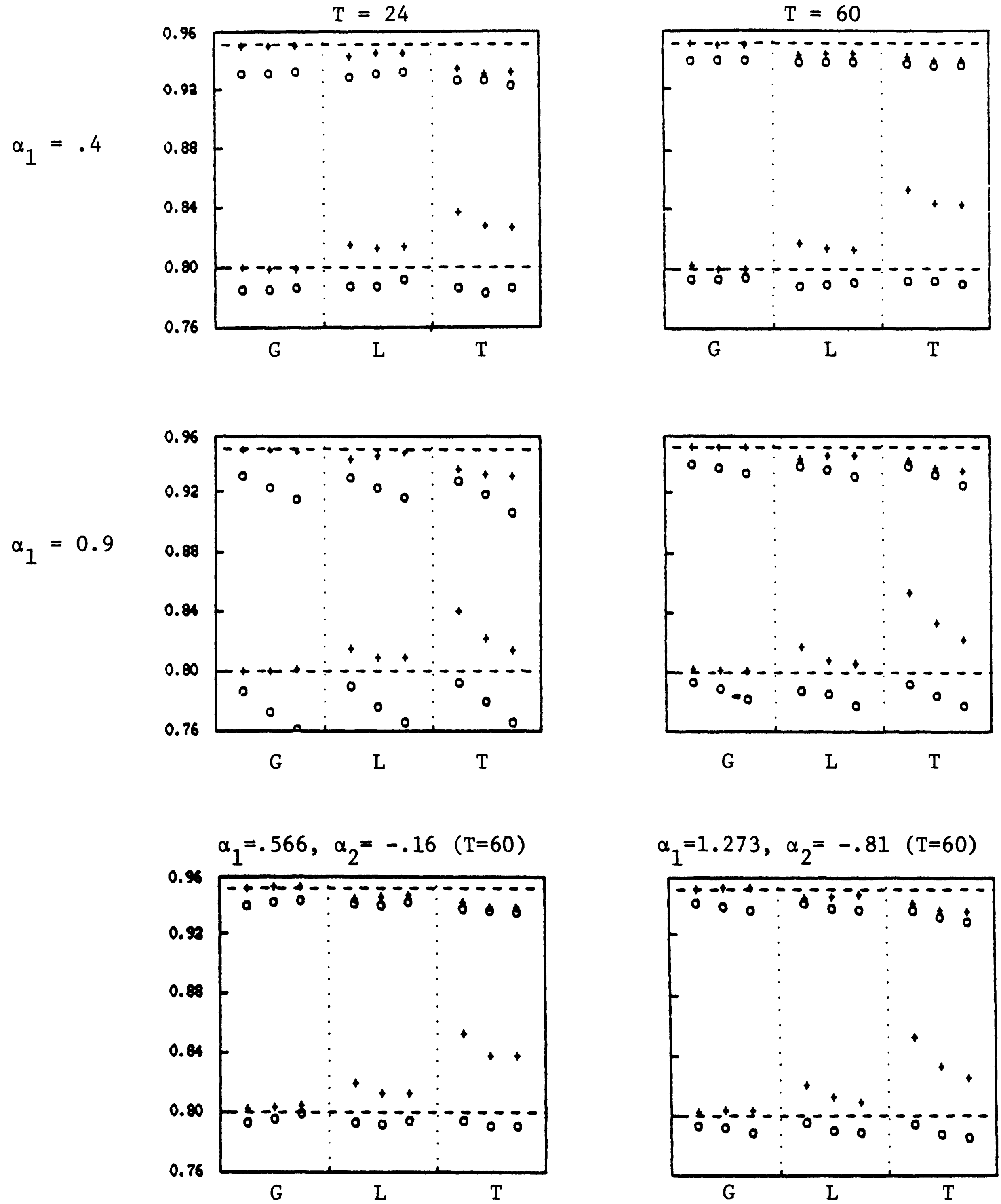

Figure 2. Average Coverage of Gaussian-Based (+) and Bootstrap (o) Prediction Intervals for Gaussian, Logistic, and Student-t Distributed Processes.

The bootstrap estimators of PMSE only rely on the bootstrap for estimating the inflation due to using $\hat{\boldsymbol{\alpha}}$. These estimators are the sum of the resampled average of $(\hat{\boldsymbol{\theta}}_f - \hat{\boldsymbol{\theta}}_f^*)'\mathbf{Y}_T^*$ and $\hat{\sigma}^2\hat{W}_f^2$ with $B = 100$. Increasing the amount of resampling would be of little benefit, since the residual distribution F_T is not affected by B and it dominates the shape of the bootstrap intervals. If we replace L_B^* in (4.4) by the limiting Gaussian distribution, exploratory simulations show negligible effects on the coverage. The bootstrap series are initialized with randomly selected blocks from the observed series.

Figure 1 shows that the bias-corrected estimator is most consistent and accurate. It is generally within two standard errors of (2.3), which is itself an approximation. The bootstrap estimators are indeed very similar to the simple substitution estimators. Both are too large when the roots are near 0 and too small when the roots are large. The bias (3.4) and inflation (2.4) cause this behavior, since they can either reinforce or weaken each other. For example, if $p = 1$ and $\alpha = .3$, $D_2 = .55$ and $\eta_2 = .36$ so that the substitution estimator is larger on average than necessary. In contrast, if $\alpha = .7$, the two factors almost cancel ($D_2 = -1.45$ and $\eta_2 = 1.96$) so that the substitution estimator is too small and estimates PMSE$(\tilde{y}_{T+f})$ rather than PMSE$(\hat{y}_{T+f})$.

Figure 2 compares the coverages of I_G and I_B in a simulation of four processes that appear in Figure 1. The two second-order processes have complex roots with norms $.4(\alpha_1 = .566, \alpha_2 = -.16)$ and $.9(\alpha_1 = 1.273, \alpha_2 = -.81)$. Three sampling distributions for the errors $\{\varepsilon_t\}$ appear in the figure: Gaussian, logistic, and Student's t with 3 df. Note that the latter violates Assumption 2, of finite moments of arbitrary order. These simulated results use the same series used to produce Figure 1; the standard errors of the average coverages are less than .003 ($T = 24$) and .002 ($T = 60$). Since the intervals are computed on the same series, the standard error of the difference in average coverage is much smaller.

The Gaussian interval outperforms the bootstrap interval for Gaussian series, but I_B performs consistently across distributions. When the errors are normally distributed, I_G obtains the nominal coverage, even for $T = 24$. For these long-tailed, non-Gaussian distributions, I_G is conservative with $\beta = .80$ and liberal with $\beta = .95$. In contrast, coverages of the bootstrap intervals are similar across distributions. The average coverage is consistently several percentages below β and decreases with increasing extrapolation. Not only is the average coverage of I_B invariant of the sampling distribution, but the simulations suggest a more general invariance. Tests for differences among the empirical distributions of the coverage of I_B for the three error distributions (for some lead, length, and process) were not significant using the methods of Conover (1980, sec. 6.4).

6. CONDITIONAL INTERVALS

The bootstrap also allows us to consider conditional prediction intervals. Both I_G and I_B estimate the conditional distribution of the forecast error and are centered on the forecast $\hat{y}_{T+f}$. Both, however, ignore specific values in $\mathbf{Y}$ and the dependence of $\hat{\boldsymbol{\alpha}}$ on $\mathbf{Y}$. It is evident, though, that if the last p observations of $\mathbf{Y}$ are near 0, then error in estimating the coefficients does not affect the prediction error. In this case, one might expect the aforementioned intervals to be conservative, since they include a correction for estimation error. This perspective makes it appealing to obtain an interval conditional on the last p observations. Phillips (1979) used Edgeworth approximations to estimate the conditional distributions of future values in first-order Gaussian autoregressions. Ansley and Kohn (1986) offered a conditional estimator of PMSE in Gaussian state-space models. One could apply the latter approach in this context by replacing $\hat{\eta}_f$ in the Gaussian interval I_G with a conditional estimate of the increase in squared error due to parameter estimation.

Bootstrapping provides another means to achieve this sort of conditioning. Rather than start the bootstrap with a randomly chosen block, require $\mathbf{Y}_T^*$ to be the same as the last p observations in $\mathbf{Y}$. Since the second-order properties of stationary autoregressions are invariant of time order, generate $\mathbf{Y}^*$ backwards by sampling the reverse-time residuals, $v_t = y_t - \sum \hat{\alpha}_j y_{t+j}$, so that $y_t^* = \sum \hat{\alpha}_j y_{t+j}^* + v_t^*$ $(t = T - p, \ldots, 1)$, as described by Thoombs and Schucany (1986). The bootstrap prediction error is then $\sum_0^{f-1} \hat{\omega}_j e_{T+f-j}^* + (\hat{\boldsymbol{\theta}}_f - \hat{\boldsymbol{\theta}}_f^*)'\mathbf{Y}_T$, where, as before, e_t^* is a random draw from the forward-time residuals. Although covariances are invariant of time reversal, independence is not. The reverse-time errors that the $\{v_t\}$ estimate need only be uncorrelated, not independent. This bootstrap procedure treats them as independent and so favors normality.

As in the case of I_B, these conditional intervals lack the nominal probability content, have comparable length and variability, and are consistent across sampling distributions. To illustrate, simulations of 1,000 Gaussian AR(1) series with $\alpha = .9$ and $T = 60$ found differences in average coverage between the two bootstrap intervals ($B = 100$) to be 0, .007, and .016 for the first three leads, with the conditional intervals having less and less coverage. These results are similar to those reported by Thoombs and Schucany (1986).

Since this approach conditions on $\mathbf{Y}_T$, the conditional coverage of the intervals is of interest. Some results for AR(1) series with $\alpha = .5$ appear in Table 1. When $y_T = 0$, we find that I_G is not conservative as suggested before and that the unconditional bootstrap interval has slightly

Table 1. Conditional Coverages of Prediction Intervals for a Gaussian AR(1) Process

Last observation	Interval	Forecast lead 1	2	3
$y_T = 0$	I_G	.797	.793	.791
	I_B	.781	.775	.773
	Conditional BS	.771	.768	.770
$y_T = 3$	I_G	.792	.820	.847
	I_B	.772	.793	.817
	Conditional BS	.809	.832	.853

NOTE: From a simulation of 1,000 series with $\alpha = .5$, $\sigma = 1$, $B = 50$, $\beta = .80$, and $T = 24$.

greater coverage than the conditional. With $y_T = 3$, we might expect an unconditional approach to underestimate the effect of estimation and lack the nominal coverage. Except at the first lead, the reverse occurs for I_G. As we would expect, the coverage of I_B is less than that of the conditional approach, which is very conservative.

The reasons behind this perhaps surprising behavior lie in the conditional properties of $\hat{\alpha}$ and $\hat{\sigma}$. The conditional expectation of $\hat{\sigma}$ is less than σ when $y_T = 0$ and greater than σ when $y_T = 3$ (.96 and 1.10, respectively). Similarly, $\hat{\alpha}$ is small on average when y_T is small, and it is large when y_T is large (.45 and .60). The bias in $\hat{\sigma}$ mitigates the effect of adding the unneeded term for estimation error when $y_T = 0$. With $y_T = 3$, positive bias in $\hat{\alpha}$ leads us to overestimate the needed width of the interval at leads greater than 1.

A naive conditioning approach like that suggested here is not particularly effective and offers little beyond I_B. The conditional approach has great appeal, however, and I am currently exploring other methods that take into account how the estimators depend on $\mathbf{Y}_T$.

[*Received April 1986. Revised February 1987.*]

REFERENCES

Ansley, C. F., and Kohn, R. (1986), "Prediction Mean Squared Error for State Space Models With Estimated Parameters," *Biometrika*, 73, 467–473.

Ansley, C. F., and Newbold, P. (1981), "On the Bias in Estimates of Forecast Mean Squared Error," *Journal of the American Statistical Association*, 76, 569–578.

Bloomfield, P. (1972), "On the Error of Prediction of a Time Series," *Biometrika*, 59, 501–507.

Box, G. E. P., and Jenkins, G. M. (1976), *Time Series Analysis: Forecasting and Control*, San Francisco: Holden-Day.

Conover, W. J. (1980), *Practical Nonparametric Statistics*, New York: John Wiley.

Davisson, I. D. (1965), "The Prediction Error of a Stationary Gaussian Time Series of Unknown Coefficients," *IEEE Transactions on Information Theory*, 11, 527–532.

Efron, B., and Tibshirani, R. (1986), "Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy," *Statistical Science*, 1, 54–74.

Findley, D. F. (1985), "On Bootstrap Estimates of Forecast Mean Squared Errors for Autoregressive Processes," in *17th Symposium on the Interface*, Amsterdam: North-Holland, pp. 154–173.

Freedman, D. A. (1981), "Bootstrapping Regression Models," *The Annals of Statistics*, 9, 1218–1228.

——— (1984), "On Bootstrapping Two-Stage Least Squares Estimates in Stationary Linear Models," *The Annals of Statistics*, 12, 827–842.

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

Fuller, W. A., and Hasza, D. P. (1981), "Properties of Predictors for Autoregressive Time Series," *Journal of the American Statistical Association*, 76, 155–161.

Godolphin, E. J. (1977), "A Direct Representation for the Maximum Likelihood Estimator of a Gaussian Moving Average Process," *Biometrika*, 64, 365–384.

Kendall, M. G. (1954), "Note on Bias in the Estimation of Autocorrelations," *Biometrika*, 41, 403–405.

Koopmans, L. (1974), *The Spectral Analysis of Time Series*, New York: Academic Press.

Kunitomo, N., and Yamamoto, T. (1985), "Properties of Predictors in Misspecified Autoregressive Time Series Models," *Journal of the American Statistical Association*, 80, 941–950.

Millar, P. W. (1979), "Asymptotic Minimax Theorems for the Sample Distribution Function," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 48, 233–252.

Phillips, P. C. B. (1979), "The Sampling Distribution of Forecasts From First-Order Autoregressions," *Journal of Econometrics*, 9, 241–261.

Schuster, E. F. (1975), "Estimating the Distribution Function of a Symmetric Distribution," *Biometrika*, 62, 631–635.

Shaman, P. (1983), "Properties of Estimates of the Mean Squared Error of Prediction in Autoregressive Models," in *Studies in Econometrics, Time Series, and Multivariate Statistics*, eds. S. Karlin, T. Amemiya, and L. A. Goodman, New York: Academic Press, pp. 331–342.

Sharpe, K. (1970), "Robustness of Normal Tolerance Intervals," *Biometrika*, 57, 71–78.

Stine, R. A. (1982), "Prediction Intervals for Time Series," unpublished Ph.D. dissertation, Princeton University, Dept. of Statistics.

——— (1985), "Bootstrap Prediction Intervals for Regression," *Journal of the American Statistical Association*, 80, 1026–1031.

Tanaka, K. (1984), "An Asymptotic Expansion Associated With the Maximum Likelihood Estimators in ARMA Models," *Journal of the Royal Statistical Society*, Ser. B, 46, 58–67.

Thoombs, L., and Schucany, W. (1986), "Bootstrap Prediction Intervals for Autoregression," unpublished manuscript, University of South Carolina, Dept. of Statistics.

Yamamoto, T. (1976), "Asymptotic Mean Squared Prediction Error for an Autoregression Model With Estimated Coefficients," *Applied Statistics*, 25, 123–127.

Yamamoto, T., and Kunitomo, N. (1984), "Asymptotic Bias of the Least Squares Estimator for Multivariate Autoregressive Models," *Annals of the Institute for Statistical Mathematics*, 36, 419–430.