COMPUTER SCIENCE AND STATISTICS:
The Interface, D.M. Allen (ed.)
Elsevier Science Publishers B.V. (North-Holland), 1986

ON BOOTSTRAP ESTIMATES OF FORECAST MEAN SQUARE ERRORS FOR AUTOREGRESSIVE PROCESSES

David F. Findley

Statistical Research Division
Bureau of the Census
Washington, D.C. 20233

This paper presents several analyses which suggest that the bootstrap procedure used by Freedman and Peters to simulate errors in forecasting future values of an econometrically modelled process is of limited usefulness for estimating mean square forecast errors.

1. INTRODUCTION

Freedman and Peters (1984) recently applied a resampling procedure (the "bootstrap") to obtain estimates of mean square error for the forecasts from an autoregression with exogeneous terms. In this paper, we start with a theoretical analysis of their suggested procedure for the case of (not necessarily stationary) autoregressive models without exogenous terms and later describe two situations in which the same conclusions hold in the presence of exogenous variables.

The theoretical mean square forecast error from an estimated model is the sum of two components, the mean square forecast error of the optimal predictor and the mean square difference between the optimal forecast and the estimated model's forecast. This latter component is of order 1/T, where T is the length of the observed series, and so is negligible with large samples. Our theoretical analysis in Section 2 shows that the bootstrap estimate of mean square forecast error is the sum of the usual (naive) large-sample estimate of the first component, easily obtainable without the bootstrap, and a small-sample estimate of the second. A gaussian Monte Carlo value of the second component is obtained in Section 3 for series of length 25 from the AR(2) models used in the study of Ansley and Newbold, along with the value of the root mean square error (rmse) of the large-sample estimator of the m-step-ahead forecast error, for m = 1, 2 and 5. In these examples, the rmse is always substantially larger than the O(1/T) component, supporting the observation of Stine (1982) that estimates of the second component are of little use in estimating mean square forecast error unless better estimators of the first component are available. In the final section, we discuss conditional forecast mean square errors associated with predictions of the future of the observed sample path, and conclude that in this context as well, the bootstrap's potential contribution seems limited.

2. BOOTSTRAP ESTIMATES OF UNCONDITIONAL MEAN SQUARE FORECAST ERROR

The simple bootstrap procedure of Freedman and Peters we describe below would appear to be appropriate when observations $y_1,\dots,y_T$ are available from a time series obeying a general p-th order autoregression ($p<T$) of the form

$$(2.1) \qquad y_t = \delta + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + e_t \quad (t \geq p+1),$$

where e_t ($t \geq p+1$) are independent, identically distributed random variables with mean 0 and variance σ^2 which are independent of earlier y's; that is, for $k>0$, e_t and y_{t-k} are independent. It is assumed that the order p is known and, only for simplicity of notation, that all of the parameters $\phi_1,\dots,\phi_p$ and δ are unknown. Define $\underline{\theta} = (\delta,\phi_1,\dots,\phi_p)$. For any $m>0$ we can use back substitution in (2.1) to obtain

$$(2.2) \qquad y_{T+m} = \sum_{j=0}^{m-1} \psi_j e_{T+m-j} + f_m[\theta](y_T,\dots,y_{T-p+1}),$$

where the coefficients $\psi_0(=1)$, ψ_1, $\psi_2,\dots$ satisfy

$$(2.3) \qquad \sum_{k=0}^{\min(j,p)} \phi_k \psi_{j-k} = 0 \qquad (\phi_0 = -1),$$

and where $f_m[\theta](y_T,\dots,y_{T-p+1})$ is linear in $y_T,\dots,y_{T-p+1}$ and δ. For example, if $p=1$, then $\psi_j = \phi_1^j$ and $f_m[(\delta,\phi_1)](y_t) = \delta(1 + \phi_1 + \dots + \phi_1^{m-1}) + \phi_1^m y_t$. The two expressions on the right hand side of (2.2) are stochastically independent since e's are independent of earlier y's. It follows from this that $f_m[\theta](y_T,\dots,y_{T+p-1})$ describes the optimal forecast (the conditional mean of y_{T+m} given $y_1,\dots,y_T$) and that $\sum_{j=0}^{m-1} \psi_j e_{T+m-j}$ is the resulting forecast error.

This optimal forecast cannot be precisely determined because $\underline{\theta}$ is unknown. If $\hat{\underline{\theta}} = (\hat{\delta}, \hat{\phi}_1,\dots,\hat{\phi}_p)$ is any estimate of $\underline{\theta}$ obtained using $y_1,\dots,y_T$, then $f_m[\hat{\underline{\theta}}](y_T,\dots,y_{T-p+1})$ is a forecast of y_{T+m} with forecast error

(2.4) $$y_{T+m} - f_m[\hat{\underline{\theta}}](y_T,\dots,y_{T-p+1})$$
$$= \sum_{j=0}^{m-1} \psi_j e_{T+m-j} + \{f_m[\underline{\theta}](y_T,\dots,y_{T-p+1})$$
$$- f_m[\hat{\underline{\theta}}](y_T,\dots,y_{T-p+1})\}.$$

Since the e_{T+m-j}, $j=0,\dots,m-1$ are independent of $\hat{\underline{\theta}}$, the two terms on the right hand side of (2.4) are independent. Consequently, using E to denote expectation, the mean square m-step-ahead forecast error when the forecast is given by $f_m[\hat{\underline{\theta}}](y_T,\dots,y_{T-p+1})$ satisfies

(2.5) $$E\{y_{T+m} - f_m[\hat{\underline{\theta}}](y_T,\dots,y_{T-p+1})\}^2$$
$$= \sigma^2 \sum_{j=0}^{m-1} \psi_j^2 + E\{f_m[\underline{\theta}](y_T,\dots,y_{T-p+1})$$
$$- f_m[\hat{\underline{\theta}}](y_T,\dots,y_{T-p+1})\}^2 .$$

If T is large, and $\hat{\underline{\theta}}$ is a consistent estimator of $\underline{\theta}$ (e.g. from least squares, if $E|e_t|^\alpha < \infty$ for some $\alpha>2$, see Lai and Wei (1983)), then the second term on the right in (2.5) can be ignored and the mean square forecast error can be adequately approximated by

(2.6) $$\hat{\sigma}^2(T-p) \sum_{j=0}^{m-1} \hat{\psi}_j^2$$

where the $\hat{\psi}$'s are obtained by using $\hat{\phi}$'s in (2.3), and $\hat{\sigma}^2(T-p)$ is given by

(2.7) $$\hat{\sigma}^2(T-p) = (T-p)^{-1} \sum_{t=p+1}^{T}$$
$$\{y_t - \hat{\delta} - \hat{\phi}_1 y_{t-1} - \dots - \hat{\phi}_p y_{t-p}\}^2.$$

If T is small, however, then the second term on the right in (2.5) need not be negligible. Also, the quantity (2.6) may be an inadequate approximation to $\sigma^2 \sum_{j=0}^{m-1} \psi_j^2$. For the situation in which T is small, Freedman and Peters (1983) propose the following bootstrap procedure. Define

$$\hat{e}_t = y_t - \hat{\delta} - \hat{\phi}_1 y_{t-1} - \dots - \hat{\phi}_p y_{t-p} ,$$
$$t = p+1,\dots,T$$

Since we are concerned with the situation in which only one realization of the series y_t is observed, we will now regard the $\hat{e}_t$'s and $\hat{\underline{\theta}}$ as fixed. We will assume that the sample mean $\bar{\hat{e}}$ of the $\hat{e}$'s is 0, as happens, for example, when $\hat{\underline{\theta}}$ is chosen to minimize $\hat{\sigma}^2(T-p)$ in (2.7). (Otherwise, use $\hat{e}_t - \bar{\hat{e}}$ in place of $\hat{e}_t$ below.) Then if we define e_t^*, $t>p$, by successive independent draws with replacement from $\{\hat{e}_{p+1},\dots,\hat{e}_T\}$, we obtain a series of identically distributed random variables with mean 0 and variance $\hat{\sigma}^2(T-p)$ whose common distribution is the empirical distribution of $\{\hat{e}_{p+1},\dots,\hat{e}_T\}$. Now we define the so-called psuedo-data series, y_t^*, by means of $y_t^* = y_t$, $1\le t\le p$ and

(2.8) $$y_t^* = \hat{\delta} + \hat{\phi}_1 y_{t-1}^* + \dots + \hat{\phi}_p y_{t-p}^*$$
$$+ e_t^* \quad (t>p).$$

The e^*'s are independent of earlier y^*'s. Let $\underline{\theta}^*$ denote the value corresponding to $\hat{\underline{\theta}}$ when $y_1^*,\dots,y_T^*$ are used in place of the original values $y_1,\dots,y_T$: For example, if $\hat{\underline{\theta}}$ was obtained by least squares, we choose $\underline{\theta}^*$ so that

$$\sum_{t=p+1}^{T} \{y_t^* - \delta^* - \phi_1^* y_{t-1}^* - \dots - \phi_p^* y_{t-p}^*\}^2 .$$

is minimized.

We have now created an analogue of the original situation, but one in which we can use a (psuedo-) random number generator to simulate draws with replacement from $\{\hat{e}_{p+1},\dots,\hat{e}_T\}$ and so obtain as many (psuedo-) independent realizations of $y_1^*, \ldots, y_{T+m}^*$ as we like. With these realizations, finally, we can approximate the distribution of the forecast error process $y_{T+m}^* - f_m[\underline{\theta}^*](y_T^*, \ldots, y_{T-p+1}^*)$ to any desired degree of accuracy. To the extent that this resembles the distribution of $y_{T+m} - f_m[\hat{\underline{\theta}}](y_T, \ldots, y_{T+p-1})$, we thereby gain information about the error process in which we are actually interested.

For example, following Freedman and Peters (1983), given realizations $y_1^{*(n)}, \ldots, y_{T+m}^{*(n)}$, $n=1, \ldots, N$, we can approximate

$$(2.9) \qquad E^*\{y_{T+m}^* - f_m[\underline{\theta}^*](y_T^*, \ldots, y_{T-p+1}^*)\}^2$$

by means of

$$N^{-1} \sum_{n=1}^{N} \{y_{T+m}^{*(n)} - f_m[\underline{\theta}^{*(n)}](y_T^{*(n)}, \ldots, y_{T-p+1}^{*(n)})\}^2 .$$

(In (2.9) and below, we use E^* to denote expectation with respect to the distribution of the series e_t^*.)

The question is, what is the relationship between the quantity (2.9) and $E\{y_{T+m} - f_m[\hat{\underline{\theta}}](y_T, \ldots, y_{T-p+1})\}^2$? To obtain a partial answer, we note that, by analogy with (2.5), the quantity (2.9) is equal to

$$(2.10) \qquad \hat{\sigma}^2(T-p)\sum_{j=1}^{m-1} \hat{\psi}_j^2 + E^*\{f_m[\hat{\underline{\theta}}](y_T^*, \ldots, y_{T-p+1}^*) - f_m[\underline{\theta}^*](y_T^*, \ldots, y_{T-p+1}^*)\}^2 .$$

Thus, this bootstrap procedure inflates the naive estimate of mean square prediction error, (2.6), by an amount

$$(2.11) \qquad E^*\{f_m[\hat{\underline{\theta}}](y_T^*, \ldots, y_{T-p+1}^*) - f_m[\underline{\theta}^*](y_T^*, \ldots, y_{T-p+1}^*)\}^2$$

which is clearly a proxy for the mean square deviation of $f_m[\hat{\underline{\theta}}](y_T, \ldots, y_{T-p+1})$ from $f_m[\underline{\theta}](y_T, \ldots, y_{T-p+1})$,

$$(2.12) \qquad E\{f_m[\hat{\underline{\theta}}](y_T, \ldots, y_{T-p+1}) - f_m[\underline{\theta}](y_T, \ldots, y_{T-p+1})\}^2 ,$$

appearing as the second component on the right hand side of (2.5). Since the quantity (2.6) is known independently of the bootstrap procedure, we conclude that an estimate of (2.11) is, in fact, the only contribution made by this procedure. Further, to estimate (2.11) it is clear that psuedo-future data $y_{T+1}^*, \ldots, y_{T+m}^*$ are not required, but only realizations of $y_1^*, \ldots, y_T^*$. Thus, in place of Freedman and Peters' procedure to estimate the mean square m-step-ahead forecast error, it seems appropriate to only consider quantities

$$(2.13) \qquad N^{-1} \sum_{n=1}^{N} \{f_m[\hat{\underline{\theta}}](y_T^{*(n)}, \ldots, y_{T-p+1}^{*(n)}) - f_m[\underline{\theta}^{*(n)}](y_T^{*(n)}, \ldots, y_{T-p+1}^{*(n)})\}^2 ,$$

using these to estimate (2.12), the component of mean square forecast error due to the use of $\hat{\underline{\theta}}$ instead of $\underline{\theta}$ in the forecast function.

Somewhat analogous observations can be made for the model selection procedure proposed in Freedman and Peters (1983): Suppose two different autoregressive models, of orders p(A) and p(B), are fit to the observed data $y_1, \ldots, y_T$, resulting in estimated parameters $\underline{\theta}_A$ and $\underline{\theta}_B$, residual populations $\{e_{p(A)+1}^A, \ldots, e_T^A\}$ and $\{e_{p(B)+1}^B, \ldots, e_T^B\}$, and psuedo-data series y_t^{A*} and y_t^{B*} as above. Freedman and Peters suggest that each model be fit to, and then used to forecast, the psuedo-data from the other model, and that bootstrap estimates of the mean square forecast error be calculated. The model having the smaller estimated mean square forecast error is to be preferred. Thus, using an obvious notational scheme, the idealized quantities to be compared are

$$E^{A*}\{y_{T+m}^{A*} - f_m^B[\underline{\theta}_B^{A*}](y_T^{A*}, \ldots, y_{T-p(B)}^{A*})\}^2$$

and

$$E^{B*}\{y_{T+m}^{B*} - f_m^A[\underline{\theta}_A^{B*}](y_T^{B*}, \ldots, y_{T-p(A)}^{B*})\}^2 .$$

By the argument used to derive (2.5), these idealized quantities are equal, respectively, to

$$(2.14)\qquad \sigma_A^2(T-p(A))\sum_{j=0}^{m-1}(\psi_j^A)^2 +$$

$$E^{A*}\{f_m^A[\underline{\theta}_A](y_T^{A*},\ldots,y_{T-p(A)}^{A*}) -$$

$$f_m^B[\underline{\theta}_B^{A*}](y_T^{A*},\ldots,y_{T-p(B)}^{A*})\}^2$$

and

$$(2.15)\qquad \sigma_B^2(T-p(B))\sum_{j=0}^{m-1}(\psi_j^B)^2 +$$

$$E^{B*}\{f_m^B[\underline{\theta}_B](y_T^{B*},\ldots,y_{T-p(B)}^{B*}) -$$

$$f_m^A[\underline{\theta}_A^{B*}](y_T^{B*},\ldots,y_{T-p(A)}^{B*})\}^2 .$$

Since the leading expressions in (2.14) and (2.15) can be calculated independently of the bootstrap, we see, as before, that the bootstrap's only contribution is to compare forecasts and that psuedo-data at times later than T are not needed for this.

All of the arguments given above also apply to the case of vector autoregressions, and thus also to the case of autoregressions with exogeneous variables, provided that endogeneous and exogenous variables are simultaneously forecasted from a combined vector autoregression. They also apply if all needed values of the exogenous variables are assumed to be nonrandom and known, as in Freedman and Peters (1984)

3. THE SIZE OF (2.12) IN SOME EXAMPLES

Again using an obvious notation, let us rewrite (2.5) as

$$(3.1)\qquad \sigma_{m,T}^2 = \sigma_m^2 + E\hat{\Delta}_{m,T}^2$$

The analogous formula for the bootstrap estimate (see (2.10)) can be written

$$(3.2)\qquad \sigma_{m,T}^{*2} = \hat{\sigma}_m^2(T-p) + E^*\Delta_{m,T}^{*2}$$

For estimating $\sigma_{m,T}^2$, the practical significance of having an estimate $E^*\Delta_{T,m}^{*2}$ of $E\hat{\Delta}_{m,T}^2$ depends upon the size of $E\hat{\Delta}_{m,T}^2$ relative to σ_m^2 and to the root mean square estimation error of the large-sample estimate $\hat{\sigma}_m^2(T-p)$ of σ_m^2,

$$\mathrm{rmse}(\hat{\sigma}_m^2(T-p)) = \{E(\hat{\sigma}_m^2(T-p) - \sigma_m^2)^2\}^{1/2} .$$

In Table (3.1) below, we present Monte Carlo estimates of the ratios $E\hat{\Delta}_{m,T}^2/\sigma_m^2$ and

$$(3.3)\qquad E\hat{\Delta}_{m,T}^2/\mathrm{rmse}(\hat{\sigma}_m^2(T-p))$$

for the observation length T=25 for some gaussian AR(2) processes

$$(3.4)\qquad y_t = \delta + \phi_1 y_{t-1} + \phi_2 y_{t-1} + e_t$$

utilized in the study of Ansley and Newbold (1981). We note that these quantities are relevant for the estimation of $\sigma_{m,T}$ as well, since, for example,

$$\sigma_{m,T} = \sigma_m\{1 + (E\hat{\Delta}_{m,T}^2/\sigma_m^2)\}^{1/2} ,$$

which is well approximated by

$$\sigma_m\{1 + \tfrac{1}{2}(E\hat{\Delta}_{m,T}^2/\sigma_m^2)\}$$

if $(E\hat{\Delta}_{m,T}^2/\sigma_m^2)^2/8$ is negligible (Taylor polynomial approximation). For each pair of coefficients ϕ_1, ϕ_2 in the Table, we estimated the quantities $E\hat{\Delta}_{m,T}^2$ and $\mathrm{rmse}(\hat{\sigma}_m^2(T-p))$ as the mean of sample estimates obtained from 1000 stationary pseudo-Gaussian series satisfying (3.4) with $\delta = 0$, using least squares to estimate δ, ϕ_1 and ϕ_2. (The IMSL pseudo-Gaussian generator GGNML was utilized.) The tabled results suggest that estimation of $E\hat{\Delta}_{m,T}^2$ is of little consequence when $\hat{\sigma}_m^2(T-p)$ is used to estimate σ_m^2.

Table 3.1 Values of $E\hat{\Delta}^2_{m,T}/\sigma^2_m$ and (3.3) for M=1, 2 and 5, for selected Gaussian AR(2) processes, with T=25.

ϕ_1	ϕ_2	m	$E\hat{\Delta}^2_{m,T}/\sigma^2_m$	(3.3)
.40	-.15	1	.01	.02
		2	.01	.01
		5	.00	.01
.80	-.65	1	.01	.05
		2	.04	.04
		5	.02	.02
.80	-.16	1	.03	.04
		2	.02	.03
		5	.02	.04

We have not included results for those of Ansley and Newbold's AR(2) models whose characteristic polynomials have a root in the annulus $1.0<|z|<1.24$. With T=25, simulations for such models produced large numbers of explosive series (the estimated characteristic polynomials had a root in $|z|<1.0$).

4. CONDITIONAL MEAN SQUARE FORECAST ERROR

In the preceding sections, we investigated unconditional mean square forecast error. However, it is the error associated with predicting a future point on the observed sample path (realization) which usually is most of interest.

4A. Mean Square Error Formulas

Since, by (2.1), the value of y_{T+m} depends on the data $y_1,\ldots,y_T$ only through the last p observations, it is easy to check that we can simply reinterpret the expectation operator E in (2.5) as designating expectation conditional upon $y_T,y_{T-1},\ldots,y_{T-(p+1)}$ and thereby obtain the fundamental decomposition of the mean square forecast error conditional upon the observed sample path. The $y_T,y_{T-1},\ldots,y_{T-(p+1)}$ in the second term on the right in (2.5) are now held constant, with the result that this second term simplifies into a linear expression in the higher order moments of $\hat{\theta} - \theta$. The mean-zero first order case is illustrative: If

$$y_t = \phi y_{t-1} + e_t \quad (\phi\neq 0) \tag{4.1}$$

with e_t, $t>1$, i.i.d. having mean 0 and variance σ^2, and with e_t independent of y_{t-k} whenever $k>0$, then $f_m[\phi](y_T) = \phi^m y_T$. From the the Taylor polynomial expansion of $f_m[\hat{\phi}](y_T)$ about $\hat{\phi} = \phi$, we have

$$f_m[\hat{\phi}](y_T) - f_m[\phi](y_T) =$$

$$y_T \sum_{j=1}^{m} C_{m,j}\phi^{m-j}(\hat{\phi} - \phi)^j , \tag{4.2}$$

where $C_{j,m} = m(m-1)\ldots(m-j+1)/j!$.

Taking the mean square of (4.2) conditional on y_T, we obtain

$$E\{f_m[\hat{\phi}](y_T) - f_m[\phi](y_T)\}^2 =$$

$$y_T^2 \sum_{j,k=1}^{m} C_{m,j}C_{m,k}\phi^{2m-j-k}E\{\hat{\phi} - \phi\}^{j+k} \tag{4.3}$$

To estimate (4.3) via the bootstrap, we replace y_T^* in (2.11) by y_T (ideally generating the pseudo-data in such a way that $y_T^* = y_T$, but see 4B. below). By analogy with (4.3), we then have

$$E^*\{f_m[\phi^*](y_T) - f_m[\hat{\phi}](y_T)\}^2 =$$

$$y_T^2 \sum_{j,k=1}^{m} C_{m,j}C_{m,k}\hat{\phi}^{2m-j-k}E^*\{\phi^* - \hat{\phi}\}^{j+k} . \tag{4.4}$$

The efficacy of the bootstrap procedure is usually related to the extent to which the distribution of $\underline{\theta}^* - \hat{\underline{\theta}}$ resembles that of $\hat{\underline{\theta}} - \underline{\theta}$ and to how insensitive this latter distribution is to the true parameter value θ. However, for our problem, the situation illustrated by (4.3) and (4.4) obviously holds generally: the expected mean square of $f_m[\hat{\underline{\theta}}](y_T,\ldots,y_{T-p+1}) - f_m[\underline{\theta}](y_T,\ldots,y_{T-p+1})$ conditional on $y_T,\ldots,y_{T-p+1}$ depends on the true value of $\underline{\theta}$ as well as on the distribution of $\hat{\underline{\theta}} - \underline{\theta}$, suggesting that the quality of the bootstrap approximation will be influenced by the accuracy of $\hat{\underline{\theta}}$ as an estimate of $\underline{\theta}$.

4B. Bootstrapping Conditional Sample Paths

It would seem like an attractive idea, when, as in this section, statistics associated with the distribution of y_t conditional on $y_T,\ldots,y_{T-p+1}$ are being approximated, to generate pseudodata y_t^* for the bootstrap in such a way that $y_t^* = y_t$ holds for $T-p+1 \le t \le T$.

For example, it would be appealing to estimate ϕ^* in (4.1) from sample paths passing through y_T.

To illustrate a first approach to accomplishing this, suppose we have bootstrapped residuals $e_{p+1}^*,\ldots,e_T^*$ from an estimate $\hat{\phi}$ of ϕ in (4.1). To generate y_t^* satisfying

$$y_t^* = \hat{\phi} y_{t-1}^* + e_t^* \ , \ 2 \le t \le T$$

with $y_T^* = y_T$, we could obviously set $y_T = y_T^*$ and recursively define

$$y_t^* = \hat{\phi}^{-1} y_{t+1}^* - \hat{\phi}^{-1} e_{t+1}^* \ , \quad 1 \le t \le T-1 \ . \tag{4.5}$$

In this case, however, y_t^* is neither independent of nor even uncorrelated with e_{t+1}^* for $1 \le t \le T-1$. Thus the bootstrapped data fail to have a basic property of the original data, and the consequences of this for the estimation of $\hat{\phi}$ from $y_1^*,\ldots,y_T^*$ are an unresolved issue. Furthermore, (4.5) is numerically unstable when $|\hat{\phi}|<1$.

When the series y_t is stationary, a second approach, which avoids the difficulties just encountered, would seem to recommend itself. To illustrate with the first order case again, if y_t satisfying (4.1) is stationary, then it is easy to verify that the random variables a_t defined by

$$a_t = y_t - \phi y_{t+1} \tag{4.6}$$

are uncorrelated with one another, satisfy $Ea_t^2 = Ee_t^2$, and each a_t is uncorrelated with y_{t+j} for all $j \ge 1$. (This equation is sometimes called the time-reversed representation of the process y_t.) We can therefore use, as an estimate of ϕ, the value $\tilde{\phi}$ minimizing $\sum_{t=1}^{T-1} (y_t - \tilde{\phi} y_{t+1})^2$, then define $\tilde{a}_t = y_t - \tilde{\phi} y_{t+1}$, $t=1,\ldots,T-1$, draw randomly with replacement from this set of residuals (after centering about their sample mean) to obtain $a_1^*,\ldots,a_{T-1}^*$ and, finally, define $y_T^* = y_T$ and

$$y_t^* = \tilde{\phi} y_{t+1}^* + a_t^* \tag{4.7}$$

for $t = T-1,\ldots,1$, thus generating a pseudodata sample path containing y_T. This procedure is appropriate only if the a_t defined by (4.6) are i.i.d., since this is a property of the a_t^*.

We will now show, however, that the white noise noise series a_t can be independent only if the cumulants of y_t (or, equivalently, those of e_t) are those of a Gaussian series, i.e., are 0 for orders higher than 2. Indeed, let κ_r denote the r-th order cumulant $\mathrm{cum}(e_t,\ldots,e_t)$ of e_t for some $r>2$ (assumed to exist). Since, from (4.6),

$$y_t = \sum_{j=0}^{\infty} \phi^j a_{t+j}$$

it is easy to see that the a_t's are independent if and only if a_t is independent of y_{t+j} for each $j \ge 1$. In this case, the r-th order cumulants $\mathrm{cum}(a_t, y_{t+j},\ldots,y_{t+j})$ will be 0; see Brillinger (1975, p. 19) for the fundamental properties of cumulants. For $j=1$, in particular, since we can write

$$y_{t+1} = e_{t+1} + \phi \sum_{j=0}^{\infty} \phi^j e_{t-j}$$

and

$$a_t = y_t - \phi y_{t+1} = -\phi e_{t+1} + (1 - \phi^2) \sum_{j=0}^{\infty} \phi^j e_{t-j},$$

we are then led to

$$0 = \mathrm{cum}(a_t, y_{t+1}, \ldots, y_{t+1}) =$$

$$- \phi\ \mathrm{cum}(e_{t+1}, \ldots, e_{t+1})$$

$$+ (1 - \phi^2)\phi^{r-1} \sum_{j=0}^{\infty} \phi^{jr} \mathrm{cum}(e_{t-j}, \ldots, e_{t-j})$$

$$= \kappa_r \{(\phi^{r-1} - \phi)/(1 - \phi^r)\} \quad .$$

Since $0<|\phi|<1$, it follows that $\kappa_r = 0$, as asserted. If the distribution of e_t is determined by its moments and if all moments exist, then e_t, and hence also y_t, is therefore Gaussian. For Gaussian time series, however, pseudo-Gaussian Monte Carlo simulations seem like a more natural device to use to generate sample paths than the bootstrap.

We conclude from the preceding discussion that generally satisfactory methods are lacking for obtaining bootstrap sample paths through the final observations $y_{T-p+1}, \ldots, y_T$.

Remark. The calculation used above, showing that assuming one-step forward and backward prediction are i.i.d. is tantamount to assuming that the observations are Gaussian, can be extended to stationary autoregressive processes of arbitrary order. A much more general assertion is made in Result 2.2 of Donoho (1981), namely, more that a strictly stationary non-Gaussian time series with finite second moments can have (ignoring rescalings) at most one invertible representation as a moving average of an i.i.d. white noise process. Some important details are missing in the proof which is given there, however.

CONCLUSION

Our results suggest that the estimates of mean square forecast error which result from the bootstrap procedure proposed by Freedman and Peters are not significantly more re-reliable than the large sample estimates, which are ill-behaved, in small samples. This does not exclude the possibility that other methods of bootstrapping these statistics could prove useful.

ACKNOWLEDGEMENT

We thank Craig Ansley for suggesting that conditional mean square errors be investigated and William Schucany for stimulating us to consider the time-reversed representation.

REFERENCES

[1] Ansley, C. F. and P. Newbold (1981). "On the Bias in Estimates of Forecast Journal of Multivariate Analysis 13, 1-23. Mean Squared Error." Journal of the American Statistical Association, 76, 569-578.

[2] Donoho, D. L. (1981). "On Minimum Entropy Deconvolution" in Applied Time Series Analysis II (ed. D. F. Findley) 565-608, New York: Academic Press.

[3] Freedman, D. A. and S. F. Peters (1983). "Using the Bootstrap to Evaluate Forecasting Equations." University of California/Berkeley, Department of Statistics Technical Report No. 20.

[4] Freedman, D. A. and S. F. Peters (1984). "Bootstrapping an Economic Model: Some Empirical Results." Journal of Business and Economic Statistics 2, 150-158.

[5] Lai, T. L. and C. Z. Wei (1983). "Asymptotic Properties of General Autoregressive Models and Strong Consistency of Least Squares Estimates of Their Parameters." Journal of Multivariate Analysis 13, 1-23.

[6] Stine, R. A. (1983). "Prediction Intervals for Autoregressions." Paper presented at the 3rd International Symposium on Forecasting.