BUREAU OF THE CENSUS

STATISTICAL RESEARCH DIVISION REPORT SERIES

SRD Research Report Number: CENSUS/SRD/RR-83/06

ON USING A DIFFERENT TIME SERIES FORECASTING MODEL FOR EACH FORECAST LEAD

by

David F. Findley
U.S. Bureau of the Census
Washington, D.C. 20233

Recommended by: Myron J. Katzoff

Report completed: July 18, 1983

Report issued: August 12, 1983

On Using a Different Time Series Forecasting Model for Each Forecast Lead

David F. Findley
U.S. Bureau of the Census, Washington, D.C. 20233

It is a common practice to fit a single time series model for the purpose of forecasting an observed time series at several different leads. By examining autoregressive models, we demonstrate both theoretically and empirically that there are situations in which better forecasts can be obtained by using a different model for each lead.

KEYWORDS: Forecasting; Misspecified time series models. Autoregressions.

1. INTRODUCTION

This paper advances simple theoretical arguments in favor of selecting and estimating a different linear forecasting "model" for each prediction period (lead) m for which a forecast is desired. For these arguments, it is assumed that the series being forecast is covariance stationary and is not perfectly modeled by the one-period-ahead forecasting model which is fit to it. Two examples are given using well-known series which illustrate one possible implementation of a multi-model forecasting procedure for autoregressive forecasting.

2. CHOOSING AN OPTIMAL MODEL FOR M-PERIOD FORECASTING

For ease of exposition our theoretical discussion will be confined to the situation in which a first order autoregression is to be used for prediction, although the results are valid more generally. To achieve further simplicity, we shall start with the (very large sample) situation in which the autocorrelation sequence ρ_m of the observed covariance stationary series y_t is known, and we assume that y_t has mean zero. From the easily verified fact that $y_{t+m} - \rho_m y_t$ is uncorrelated with y_t , it follows that, for any constant ψ, the expected mean square of $y_{t+m} - \psi y_t$ satisfies

$$E\{y_{t+m} - \psi y_t\}^2 = E\{y_{t+m} - \rho_m y_t\}^2 + (\psi - \rho_m)^2 E y_t^2 \qquad (1).$$

Thus, <u>if</u> ψ <u>is to be chosen without constraint to minimize</u> (1), <u>we must set</u> $\psi=\rho_m$.

If we assume that y_t is an AR(1) process, i.e.

$$y_t = \phi y_{t-1} + e_t$$

for all t, where $Ee_t^2 = \sigma_e^2$ and e_t is uncorrelated with $y_{t-1}, y_{t-2}, \ldots$, then m applications of this recursion formula yield

$$y_{t+m} - \phi^m y_t = e_{t+m} + \phi e_{t+m-1} + \ldots + \phi^{m-1} e_{t+1} \qquad (2),$$

from which the familiar facts follow that $\rho_m = \phi^m$ and that $\psi = \phi^m$ minimizes (1). Without this assumption, if we wish to "fit" an AR(1) process to y_t for the purpose of m-step-ahead prediction, i.e. if we wish to choose ϕ to minimize $E\{y_{t+m} - \phi^m y_t\}^2$, then it follows from (1) that ϕ must minimize $(\phi^m - \rho_m)^2$, so that

$$\phi = \begin{cases} \{\rho_m\}^{1/m} & \text{, if m is odd or } \rho_m > 0 \\ 0 & \text{, if m is even and } \rho_m < 0 \quad . \end{cases} \qquad (3).$$

Hence, if y_t is not an AR(1) process, then different choices of ϕ, i.e. different AR(1) models, depending on the forecast period m, will be required to obtain optimal forecasts. Also, if m is even and $\rho_m < 0$, it follows from (3) that the optimal AR(1) model's forecast of y_{t+m} from origin t will be inferior in the mean square sense to that obtained from minimizing the unconstrained expression $E\{y_{t+m} - \psi y_t\}^2$. Thus it would seem to be advantageous not to constrain the forecast function to coincide with that of an AR(1) model.

Finally, note that if an AR(1)-model is fit in the usual way, i.e., with $\phi = \rho_1$, then by setting $\psi = \rho_1^m$ in (1), the loss in mean square accuracy which results from using this model to do m-period forecasting can be seen to be $(\rho_1^m - \rho_m)^2 E y_t^2$.

3. CHOOSING AN OPTIMAL SINGLE MODEL FOR INCREASING HORIZON FORECASTS.

It can be imagined that finding an optimal model is more difficult if the model's purpose is to provide forecasts not for a single period m but for an interval $1 \leq m \leq M$. This is easy to illustrate. To fit an AR(1) for this task, we could, for example, seek a value of ϕ minimizing

$$E\{y_{t+1} - \phi y_t\}^2 + \delta_2 E\{y_{t+2} - \phi^2 y_t\}^2 + \ldots + \delta_M E\{y_{t+m} - \phi^M y_t\}^2 \qquad (4),$$

where $\delta_2, \ldots, \delta_M$ are non-negative weights chosen according to the relative importance of the different forecasts compared to the one-period forecast. It follows from (1) that (4) can be written as the sum of a term not depending on ϕ, plus Ey_t^2 times a non-negative polynomial $F(\phi)$ of degree 2m in ϕ,

$$F(\phi) = (\rho_1 - \phi)^2 + \delta_2(\rho_2 - \phi^2)^2 + \ldots + \delta_M(\rho_M - \phi^M)^2 .$$

If $\rho_m = \phi^m$, $1 \leq m \leq M$ as happens when y_t is an AR(1) process, the minimum of $F(\phi)$, is obtained at $\phi = \rho_1$. Otherwise, to minimize (4), we shall need to examine the real zeros of $f(\phi) = F'(\phi)$, a polynomial of degree 2M-1. When M=2, a minimizing ϕ will satisfy

$$\phi^3 + \{\delta_2^{-1} - \rho_2\}\phi - \delta_2^{-1}\rho_1 = 0$$

and so will have an explicit formula, perhaps a not very informative one, except for special values of δ_2, such as $\delta_2 = (\rho_2)^{-1}$, when $\phi = \{\rho_2\rho_1\}^{1/3}$. For M>2, the roots of $f(\phi)$ will have to be sought numerically, and the amount of effort required to minimize ϕ is clearly greater than that involved in fitting models for each forecast period m=1,...,M using (3).

4. OTHER LOSS FUNCTIONS FOR FORECASTING

We now wish to examine two more approaches for obtaining an optimal linear predictor for a given forecast period m. First we consider the method utilized by Gersch and Kitagawa (1983), who suggest choosing ψ to minimize an estimator of -2 times the expected gaussian log likelihood of y_{t+m} conditioned on y_t, $y_{t-1}, \ldots$. Their estimator can be approximated by

$$\log 2\pi\sigma_m^2 + (N-m)^{-1} \sum_{t=1}^{N-m} \frac{\{y_{t+m} - \psi y_t\}^2}{\sigma_m^2} \qquad (5),$$

with $\sigma_m^2 = \sigma^2 W_m(\psi)$, where $W_m(\psi)$ is scale-invariant. (For example, if ψ is constrained to have the form $\psi = \phi^m$, i.e. if an AR(1) model is being fit, then, by (2), a natural choice is $W_m(\phi) = 1 + \phi^2 + \ldots + \phi^{2m-2}$.)

Under ergodicity, the large sample limit of (5) is

$$B_m(\sigma^2, \psi) = \log 2\pi\sigma^2 W(\psi) + \{\sigma^2 W(\psi)\}^{-1} E\{y_{t+m} - \psi y_t\}^2.$$

From the critical point equation $\partial B_m(\sigma^2, \psi)/\partial\sigma^2 = 0$, we obtain

$$\sigma^2(\psi) = \{W_m(\psi)\}^{-1} E\{y_{t+m} - \psi y_t\}^2 .$$

Substituting this expression into the formula for $B(\sigma^2, \psi)$, one obtains that the optimal choice of ψ minimizes $\log 2\pi E\{y_{t+m} - \psi y_t\}^2 + 1$, and so is equal to ρ_m, as before.

The other theoretical approach we consider is one in which the coefficient ϕ is obtained by maximizing a log likelihood function $L(\phi,\sigma^2,N)$ for the m-period prediction errors which has the correct quadratic form for data from a gaussian AR(1) process:

$$-2(N-m)^{-1}L(\phi,\sigma^2,N) =$$

$$\log 2\pi\sigma^2 + \sigma^{-2}(N-m)^{-1}\sum_{s,t=1}^{N-m}(y_{t+m} - \phi^m y_t)M_{s,t}(N-m)(y_{s+m} - \phi^m y_s),$$

where $M_{s,t}(N-m)$ is the (s,t)-entry of the inverse of the covariance matrix of order N-m of the MA(m-1) process defined by the right hand side of (2) when $\sigma_e^2=1$. It can be shown, using results of Ljung (1979), Galbraith and Galbraith (1974) and a calculation similar to that given for the formula (2.15) of Findley (1983), that the estimates of ϕ obtained by maximizing $L(\phi,\sigma^2,N)$ tend almost surely, as $N\to\infty$, to ρ_1, the large-sample limit for the case m=1. Said more elaborately, the almost sure limit is the value of ϕ minimizing

$\lim_{N \to \infty} -2(N-m)^{-1} L(\phi, \sigma^2, N)$, which, with $f(\lambda)$ denoting the spectral density of y_t, is given by the function

$$\log 2\pi\sigma^2 + \sigma^{-2} \int_{-\pi}^{\pi} f(\lambda) |1 - \phi^m e^{im\lambda}|^2 / |1 + \phi e^{i\lambda} + \ldots + \phi^{m-1} e^{i(m-1)\lambda}|^2 d\lambda$$

$$= \log 2\pi\sigma^2 + \sigma^{-2} \int_{-\pi}^{\pi} f(\lambda) |1 - \phi e^{i\lambda}|^2 d\lambda$$

$$= \log 2\pi\sigma^2 + \sigma^{-2} E\{y_{t+1} - \phi y_t\}^2 \quad .$$

Thus $-2(N-m)^{-1} L(\phi, \sigma^2, N)$ would not always be an appropriate loss function for m-period prediction.

5. A CRITERION FOR SELECTING THE ORDER OF AN AUTOREGRESSION FOR m-PERIOD FORECASTING.

To fit a p-term autoregression for m-period forecasting, one could choose the coefficients $\tilde{\phi}_1, \ldots, \tilde{\phi}_p$ which minimize the expression

$$SSQ_m(\phi_1, \ldots, \phi_p) = \sum_{t=p_{max}}^{N-m} (y_{t+m} - \phi_1 y_t - \ldots - \phi_p y_{t-p+1})^2 ,$$

where p_{max} is some preassigned largest order. (Alternatively, one could minimize the estimator of Gersch and Kitagawa (1982).)

If we (i) set $N_0 = N - p_{max} - m + 1$, (ii) define

$$m\text{-AIC}(p) = N_0 \log 2\pi\{SSQ_m(\tilde{\phi}_1,\ldots,\tilde{\phi}_p)/N_0\} + N_0 + 2(p+1),$$

(iii) choose $p_{max} = p_{max}(N)$ in such a way that $p_{max}(N) \to \infty$ and that $N^{-1/2}p_{max}(N) \to 0$ as $N \to \infty$, and (iv) select $\tilde{p}$ so that

$$m\text{-AIC}(\tilde{p}) = \min_{1 \leq p \leq p_{max}(N)} m\text{-AIC}(p),$$

then, according to Shibata (1980), an asymptotically optimally efficient autoregressive m-period predictor will result, provided, among other assumptions, that x_t is stationary and is not an autoregressive process. However, Findley (1983) shows that the penalty term 2(p+1) in m-AIC(p) cannot be considered a full bias correction when $m \geq 2$. In fact, the analysis given in this reference shows that the asymptotic mean of m-AIC(p) depends in a complicated way on m and on the underlying processes, and suggests that the problem of selecting the length of an autoregression for m-period forecasting becomes more delicate with increasing m.

It seems very likely that Shibata's results can be extended to the case of nonstationary series admitting moving average transformations (e.g., differencing) to stationarity, see Tsay and Tiao (1982). We apply the three-step procedure of this section to such a nonstationary series in the following section.

6. EXAMPLES: FORECASTS OF SERIES C AND E.

The minimum m-AIC procedure decribed above was used to select models for forecasting Series C (226 chemical process temperature readings) and Series E (annual Wolfer sunspot numbers, 1770-1869) from Box and Jenkins (1976) at prediction periods 1, 2, 5 and 10 for an interval of consecutive forecast origins N_{origin}, $N_{min} \le N_{origin} \le N_{max}$. Using $p_{max}(N_{origin}) = [N_{origin}^{.48}]$, where [ ] indicates integer part, different autoregressions were fit for each value of N_{origin} using the data segment $1 \le y_t \le N_{origin}$, for each forecasting period m. The m-period forecasts from the m-period-forecasting autoregressions were compared with the m-period forecasts obtained in the usual way by assuming that the one-period forecasting autoregression correctly models the series, see Box and Jenkins (1976). The forecast origins were chosen in such a way that actual observations y_t were available to compare with the forecasts. A forecast error statistic was calculated for each forecast period m=2,5,10,

$$\text{RMSQ} = \{ \sum_{N_{origin}=N_{min}}^{N_{max}} (y_{N_{origin}+m} - \sum_{j=1}^{\tilde{p}} \tilde{\phi}_j(N_{origin}) y_{N_{origin}-j+1})^2 \}^{1/2} .$$

for each m-period-forecasting autoregression, along with an analogous statistics for the forecasts obtained from the one-period-forecasting models. For each forecast period m = 2, 5, 10 the ratios of the values of RMSQ for the two different forecasting procedures are tabulated below, using the value obtained for the usual forecasts from the one-period-forecasting autoregressive model as the numerator and the value from the m-period forecasting autoregression as the denominator.

The results indicate that a modest average improvement is often obtained by using the m-period-forecasting autoregressions. Also, depend- on the series, individual forecasts can differ markedly from those obtained from the one-period forecast autoregressive model, see Tables 2 and 4. This suggests that it can be worthwhile to calculate forecasts from both procedures in order to obtain two somewhat independently conceived forecasts. The fitting of these m-period forecasting autoregressions can also yield additional information about forecast error variances: It is shown in Findley (1983) that $SSQ(\phi_1,\ldots,\phi_p)/(N_{origin} - p_{max})$ is a downwardly biased estimate of the forecast error variance. (The bias is of order $(N-p_{max})^{-1}$.) When this quantity is observed to be larger than σ_m^2, as happens 10 of 11 times with Series E for m = 5,10, see Table 5, it can be taken as an indication that σ_m^2 might be underestimating the actual forecast error variance, which would be a sign of model inadequacy.

Table 1. RMSQ Ratios for Series C, $150 \le N_{origin} \le 200$.

m =	2	5	10
RMSQ	1.001	1.036	1.043

Table 2. Average Absolute Difference (AAD) between Forecasts from the Two Methods, and Average Absolute Percent Difference (AAPD) as a Percentage of the Observed Value (the Correct Forecast), for Series C, $150 < N_{origin} < 200$.

m =	2	5	10
AAD	.009	.036	.090
AAPD	0.0%	0.0%	0.0%

Table 3. RMSQ Ratios for Series E, $80 < N_{origin} < 90$.

m =	2	5	10
RMSQ	.994	1.026	1.106

Table 4. Average Absolute Difference (AAD) between Forecasts from the Two Methods, and Average Absolute Percent Difference (AAPD) as a Percentage of the Observed Value (the Correct Forecast), for Series E, $80 < N_{origin} < 90$.

m =	2	5	10
AAD	1.19	2.42	3.89
AAPD	6.8%	12.4%	10.0%

Table 5. Average Values of σ_m and $(SSQ_m)^{1/2}$ for Series E over the Interval $80 < N_{origin} < 90$.

m =	2	5	10
Avg. σ_m	23.3	31.6	32.0
Avg. $(SSQ)^{1/2}$	22.9	32.5	33.1

The Series C and E represent two different extremes in the sense that Series C seems well-modeled as an AR(2) ($\tilde{p}$ was always 2), whereas the sunspot number series is known to be difficult to model ($\tilde{p}$ ranged from 2 to 8 depending mainly on the value of m). Over the course of years, more elaborate models have been proposed for the sunspot series, see Woodward and Gray (1978). However, our example serves to illustrate the situation when a series is modeled for the first time.

7. CONCLUDING REMARKS

We have demonstrated theoretically that the widespread practice of using a single model for the purpose of forecasting several different future values of a series is not an optimal procedure when the data do not perfectly conform to the type of model being fit. A simple procedure for selecting m-period forecasting autoregressions was introduced and applied to two series which have traditionally been modeled as autoregressive processes. Compared with a naive model for Series E, this produced a modest forecast improvement at the longer leads. Another reason, described above, for augmenting the usual forecasting procedure with this new procedure is the gain in information obtained about the forecast variability and the fit of the model.

For series which seem well-modeled by ARMA(p,q) models with $q \neq 0$ it would be appropriate to seek rational predicting functions $\phi(B)/\theta(B)$ rather than simple autoregressions. Another issue of potential importance is deciding when certain coefficients in the predicting formulas should be set equal to zero. Ploberger (1982) describes an estimator of the variance-covariance matrix of $\tilde{\phi}_1,\ldots,\tilde{\phi}_p$ which is consistent even in cases in which the series is not an autoregressive or ARMA process, but this estimator does not appear to have been implemented yet.

REFERENCES

Box, G.E.P. and Jenkins, G.M. (1976). Time Series Analysis 2nd Ed. San Francisco: Holden-Day.

Findley, D. (1983). "On the Unbiasedness Property of AIC for Exact or Approximating ARMA Models," Statistical Research Division Report 83/01, U.S. Census Bureau.

Galbraith, R. F. and Galbraith, J. I. (1974)). "On the Inverses of Some Patterned Matrices Arising in the Theory of Stationary Time Series," Journal of Applied Probability, II, 63-71.

Gersch, W. and Kitagawa, G. (1983). "The Prediction of Time Series with Trends and Seasonalities," to appear in Journal of Business and Economic Statistics.

Ljung, L. (1978). "Convergence Anlaysis of Parametric Identification Models," IEEE Transactions on Automatic Control, 23, 770-782.

Ploberger, W. (1982). "The Estimation of the Asymptotic Distribution of the Specification Error for Misspecified Models," Institute for Econometrics and Operations Research, Technical University of Vienna, Research Report Number 18.

Shibata, R. (1980). "Asymptotically Efficient Selection of the Order of the Model for Estimating Parameters of a Linear Process," Annals of Statistics 8, 147-164.

Tsay, R. S. and Tiao, G. (1982). "Consistent Least Squares Estimates of Autoregressive Parameters for Stationary and Nonstationary ARMA Models," Department of Statistics, University of Wisconsin-Madison.

Woodward, W. A. and Gray, H. L. (1978). "New ARMA Models for Wolfer's Sunspot Data," Communications in Statistics B, 7, 97-115.