Calculating Interval Forecasts

Author(s): Chris Chatfield

Source: *Journal of Business & Economic Statistics*, Apr., 1993, Vol. 11, No. 2 (Apr., 1993), pp. 121-135

Published by: Taylor & Francis, Ltd. on behalf of American Statistical Association

Stable URL: https://www.jstor.org/stable/1391361

REFERENCES
Linked references are available on JSTOR for this article:
https://www.jstor.org/stable/1391361?seq=1&cid=pdf-reference#references_tab_contents
You may need to log in to JSTOR to access the linked references.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at https://about.jstor.org/terms

Taylor & Francis, Ltd. and *American Statistical Association* are collaborating with JSTOR to digitize, preserve and extend access to *Journal of Business & Economic Statistics*

This content downloaded from
192.12.247.244 on Mon, 25 Mar 2024 08:44:27 +00:00
All use subject to https://about.jstor.org/terms

© 1993 American Statistical Association | Journal of Business & Economic Statistics, April 1993, Vol. 11, No. 2

Editor's Note: The following article was the *JBES* 1992 Invited Lecture at the Joint Statistical Meetings, Boston, Massachusetts, August 8–12, 1992.

Calculating Interval Forecasts

Chris Chatfield
School of Mathematical Sciences, University of Bath, Bath, Avon BA2 7AY, United Kingdom

The importance of interval forecasts is reviewed. Several general approaches to calculating such forecasts are described and compared. They include the use of theoretical formulas based on a fitted probability model (with or without a correction for parameter uncertainty), various "approximate" formulas (which should be avoided), and empirically based, simulation, and resampling procedures. The latter are useful when theoretical formulas are not available or there are doubts about some model assumptions. The distinction between a forecasting *method* and a forecasting *model* is expounded. For large groups of series, a forecasting method may be chosen in a fairly *ad hoc* way. With appropriate checks, it may be possible to base interval forecasts on the model for which the method is optimal. It is certainly unsound to use a model for which the method is *not* optimal, but, strangely, this is sometimes done. Some general comments are made as to why prediction intervals tend to be too narrow in practice to encompass the required proportion of future observations. An example demonstrates the overriding importance of careful model specification. In particular, when data are "nearly nonstationary," the difference between fitting a stationary and a nonstationary model is critical.

KEY WORDS: Bayesian forecasting; Bootstrapping; Box–Jenkins method; Forecasting; Holt–Winters method; Prediction intervals; Resampling; Simulation.

Predictions are often given as point forecasts with no guidance as to their likely accuracy (and perhaps even with an unreasonably high number of significant digits implying spurious accuracy!). Of course, point forecasts may sometimes be adequate; for example, a sales manager may be happier with a single point forecast of demand than with an interval forecast. But should he be? It is often important, however, to give interval forecasts as well as (or instead of) point forecasts so as to

1. Assess future uncertainty.
2. Enable different strategies to be planned for the range of possible outcomes indicated by the interval forecast.
3. Compare forecasts from different methods more thoroughly (is a narrower interval forecast necessarily better?) and explore different scenarios based on different assumptions.

An interval forecast usually consists of upper and lower limits associated with a prescribed probability. The limits are sometimes called *forecast limits* (Wei 1990) or *prediction bounds* (Brockwell and Davis 1987, p. 175), whereas the interval is sometimes called a *confidence interval* (e.g., Granger and Newbold 1986). This article prefers the more widely used term *prediction interval* (e.g., Abraham and Ledolter 1983; Bowerman and O'Connell 1987; Chatfield 1989; Harvey 1989; Montgomery and Johnson 1976), both because it is more descriptive and because the term "confidence interval" is usually applied to estimates of (fixed but unknown) parameters. In contrast, a prediction interval (PI) is an estimate of an (unknown) future value that can be regarded as a random variable at the time the forecast is made. This involves a different sort of probability statement to a confidence interval as discussed, for example, by Hahn and Meeker (1991, sec. 2.3) in a non-time-series context.

Given their importance, it is perhaps surprising and rather regrettable that many companies do not regularly produce PI's (e.g., Dalrymple 1987) and that most economic predictions are given as a single value. Several reasons can be suggested for this; namely,

1. The topic has been rather neglected in the literature. Textbooks on time series analysis and forecasting often give little guidance, except perhaps for regression and autoregressive integrated moving average (ARIMA) modeling, and much of the relevant journal literature is demanding, sometimes unhelpful, and occasionally misleading or even wrong.
2. There is no generally accepted method of calculating PI's except for procedures based on fitting a probability model for which the variance of forecast errors can be readily evaluated.
3. Theoretical PI's are difficult or impossible to evaluate for many econometric models, especially multivariate models containing many equations or depending on nonlinear relationships. In any case, judgmental "adjustment" is often used in the forecasting process (e.g., to forecast exogenous variables or to compensate for known changes in external conditions), and it is not clear how this will affect PI's.

This content downloaded from
192.12.247.244 on Mon, 25 Mar 2024 08:44:27 +00:00
All use subject to https://about.jstor.org/terms

4. A forecasting method is sometimes chosen for a group of series (e.g., in inventory control) in a fairly *ad hoc* way with no attempt to fit a probability model. Then it is not clear if theoretical PI's should be used even when the method is known to be optimal for a particular model. In other cases, a method may not be based explicitly, or even implicitly, on a probability model, and it is unclear how to proceed.

5. Empirically based methods for calculating PI's are not widely understood, and their properties have been little studied.

6. Various "approximate" procedures for calculating PI's have been suggested, but many users are unsure as to their validity.

7. Many software packages produce invalid PI's or do not produce them at all, partly because of 1–6.

Note that we restrict ourselves to PI's for a single observation and do not consider the more difficult calculation of a simultaneous prediction region for a set of future observations (e.g., Lütkepohl 1991, sec. 2.2.3; Ravishankar, Hochberg, and Melnick 1987; Ravishankar, Wu, and Glaz 1991).

1. NOTATION

The clarity of the time series literature is not helped by a diverse and sometimes confusing notation. For example, Z_t is often used to denote an innovation process but is occasionally used to denote a measured variable. The number of observations is variously denoted by n, N, or T. The innovations process is variously denoted by u_t, ε_t, e_t, η_t, and a_t, among others, as well as by Z_t. More confusingly, the same symbol is often used to denote both the random variable at time t and its observed value. Moreover, this one symbol is sometimes written in upper case and sometimes in lower case.

This article adopts the following notation. An observed time series, containing n observations, is denoted by $x_1, x_2, \ldots, x_n$. This is regarded as a finite realization of a stochastic process $\{X_t, t \in T\}$, where the index set T could, for example, be the positive integers. Suppose that we wish to forecast X_{n+k}, where the integer k is called the lead time (or the forecasting horizon). The point forecast of X_{n+k} made conditional on data up to time n for k steps ahead will be denoted by $\hat{X}_n(k)$ when regarded as a random variable and by $\hat{x}_n(k)$ when it is a particular value determined by the observed data. Note that it is essential to specify both the time at which a forecast is made *and* the lead time. Some of the literature does not do this and uses an ambiguous notation such as $\hat{X}_{n+k}$ for forecasts of X_{n+k} regardless of when the forecast was made. I denote the (conditional) forecast error corresponding to $\hat{x}_n(k)$ by

$$e_n(k) = X_{n+k} - \hat{x}_n(k) \qquad (1.1)$$

which is, of course, a random variable even though $\hat{x}_n(k)$ is not. The observed value of $e_n(k)$—namely, $(x_{n+k} - \hat{x}_n(k))$—may later become available. In Section 3, I also refer to the unconditional forecast error—namely, $(X_{n+k} - \hat{X}_n(k))$, where both terms are random variables.

It is important to understand the distinction between the forecast errors, $e_n(k)$, the model innovations, and the fitted residuals (or within-sample "forecast" errors). I use $\{\varepsilon_t\}$ to denote the innovations process so that a model with additive innovations can generally be represented by $X_t = \mu_t + \varepsilon_t$, where μ_t denotes the "signal" at time t. The innovations are usually assumed to be a sequence of independent normally distributed random variables with zero mean and constant variance σ_ε^2, which I write as NID $(0, \sigma_\varepsilon^2)$.

The within-sample one-step-ahead observed "forecasting" errors—namely, $[x_t - \hat{x}_{t-1}(1)]$ for $t = 2, 3, \ldots, n$—can also be called the *residuals* because they are the differences between the observed and the fitted values. They will not be the same as the true-model innovations because the residuals depend on estimates of the model parameters and perhaps also on estimated starting values. They are also not true forecasting errors because the model and parameters are determined from all the data up to time n.

The standard notation ARIMA (p, d, q) is used to denote a process with p autoregressive items and q moving average terms, which requires dth differencing to make it stationary.

2. FORECASTING METHODS

Forecasting methods come in a wide variety of forms, (e.g., see Chatfield [1988] for a review and a discussion of the criteria for choosing between them). It is helpful to categorize methods as *univariate*, where $\hat{x}_n(k)$ depends only on $x_n, x_{n-1}, \ldots$; *multivariate*, where $\hat{x}_n(k)$ may also depend on other explanatory variables; and *judgmental* methods. Another useful distinction is between *automatic* methods, requiring no human intervention, and nonautomatic methods.

A further useful distinction, particularly for computing PI's, is between methods that involve fitting an "optimal" probability model and those that do not. The former situation is perhaps more familiar to the statistician and the latter to the operational researcher. For example, the Box–Jenkins forecasting procedure involves formulating, fitting, and then checking an appropriate ARIMA model. The corresponding optimal forecasts from that model can then be computed. The strategy of fitting an optimal model is recommended with only a few series to forecast and with adequate statistical expertise available. At the other extreme, the practitioner with many series to forecast may decide to use the same all-purpose procedure whatever the individual series look like. For example, the Holt–Winters forecasting procedure generalizes exponential smoothing to cope with trend and seasonality but does not depend explicitly on any probability model. The method may be selected for a group of series showing trend and seasonal variation. No model identification is involved, however, and it is quite wrong to talk about the Holt–

This content downloaded from
192.12.247.244 on Mon, 25 Mar 2024 08:44:27 +00:00
All use subject to https://about.jstor.org/terms

Winters forecasting *model*, as some authors have done. Forecasts need not be optimal for an individual series, and the approach can rightly be criticized as somewhat ad hoc. This sort of approach is widely used, however, particularly in inventory control, and cannot be ignored. In summary, a forecasting *method* may, or may not, depend on a forecasting *model*, and it is helpful to keep this distinction in mind.

The choice of method depends on a variety of factors such as the objectives and the type of data. Regrettably, the results of previous forecasting "competitions" apply mainly to the use of automatic methods on large groups of disparate series and do not generally apply to a single-series situation or to a homogeneous group of series such as sales of similar items in the same company (Fildes 1992). Thus the analyst still has the difficult task of choosing a method using background knowledge, a preliminary examination of the data, and perhaps a comparative evaluation of a short list of methods.

Given such a wide range of methods and strategies, it follows that a variety of approaches may be needed to compute PI's.

3. EXPECTED MEAN SQUARED PREDICTION ERROR

The usual expression for assessing the uncertainty in forecasts is the expected mean squared prediction error (PMSE)—namely, $E[e_n(k)^2]$. This quantity is used in computing PI's; see Equation (4.1). If the forecast is *unbiased*, meaning that $\hat{x}_n(k)$ is the mean of the predictive distribution (i.e., the conditional expectation of X_{n+k} given data up to time n), then $E[e_n(k)] = 0$ and $E[e_n(k)^2] = \text{var}[e_n(k)]$. Forecasters often assume unbiasedness (explicitly or implicitly) and work with the latter quantity.

A potential pitfall here is to think that the quantity required to assess forecast uncertainty is the variance of the forecast rather than the variance of the forecast error. In fact, given data up to time n and a particular method or model, the forecast $\hat{x}_n(k)$ will be determined exactly and hence have a conditional variance of 0, whereas X_{n+k} and $e_n(k)$ are random variables, albeit conditioned by the observed data.

At first sight, the evaluation of the expression $E[e_n(k)^2]$ seems to pose no particular problems. In fact, it is not always clear how the expectation should be evaluated and what assumptions should be made. Textbooks rarely consider the problem thoroughly, if at all, though that of Kendall and Ord (1990, chap. 8) is a partial exception. There have been several technically difficult papers on different aspects of the problem, but they give few numerical illustrations and little qualitative comment and say little or nothing about the construction of PI's, which should surely be one of the main objectives.

Consider for simplicity the zero-mean (autoregressive) AR(1) process given by

$$X_t = \alpha X_{t-1} + \varepsilon_t, \tag{3.1}$$

where $\{\varepsilon_t\}$ are NID $(0, \sigma_\varepsilon^2)$. If one assumes complete knowledge of the model, including the values of α and σ_ε^2, then it can be shown [Box and Jenkins 1970, eq. (5.4.16)] that

$$E[e_n(k)^2] = \sigma_\varepsilon^2(1 - \alpha^{2k})/(1 - \alpha^2). \tag{3.2}$$

I shall call this the *true-model* PMSE. Formulas for true-model PMSE's can readily be derived for many types of time series model (see Sec. 4.2).

In practice the model parameters will not be known exactly, and it will be necessary to replace them with sample estimates when computing forecasts. Thus $\hat{x}_n(k)$ will be $\hat{\alpha}^k x_n$ rather than $\alpha^k x_n$. Restricting attention to the case $k = 1$ for simplicity and conditioning on $X_n = x_n$, we find that

$$\begin{aligned} e_n(1) = X_{n+1} - \hat{x}_n(1) &= \alpha x_n + \varepsilon_{n+1} - \hat{\alpha} x_n \\ &= (\alpha - \hat{\alpha})x_n + \varepsilon_{n+1}. \end{aligned} \tag{3.3}$$

Finding the expectation of the square of expressions like this is not easy, and the rest of this section considers the effect of parameter uncertainty on the true-model PMSE. I assume throughout that parameter estimates are obtained by a procedure that is asymptotically equivalent to maximum likelihood.

First, look at the expected value of $e_n(1)$ rather than its square. Looking back at Equation (3.3), for example, it is clear that if x_n is fixed and $\hat{\alpha}$ is a biased estimator for α (did you realize the least squares estimator can have a sizable bias for short series?), then the expected value of $e_n(1)$ need not be 0 (Phillips 1979). If, however, we average over all possible values of x_n, as well as over ε_n, then it can be shown that the expectation will indeed be 0 giving an unbiased forecast. The former operation is *conditional* on x_n, but the latter involves the *unconditional* forecast error—namely, $(X_{n+1} - \hat{X}_n(1))$. It is important to be clear about what one is or is not conditioning on, and it could also be useful to have additional notation to distinguish between forecasts involving true and estimated parameters. There has been much confusion because of a failure to distinguish the different types of situation.

Box and Jenkins (1970, appendix A7.3) made an early contribution to assessing the effect of parameter uncertainty on the PMSE. They concluded that correction terms would generally be of order $1/n$. Their approach, and that of some later authors [e.g., Yamamoto 1976, for AR processes; Baillie 1979 and Reinsel 1980, for vector AR processes; Yamamoto 1981, for vector autoregressive moving average (ARMA) processes] was to find the PMSE by averaging not only over the distribution of future innovations [e.g., ε_{n+1} in Eq. (3.3)] but also over the distribution of the current observed values, [e.g., x_n in Eq. (3.3)], to give what is generally called the *unconditional* PMSE. This quantity can be useful to assess the "success" of a forecasting method *on average*. If used to compute PI's, however, it effectively assumes that the observations used to estimate

This content downloaded from
192.12.247.244 on Mon, 25 Mar 2024 08:44:27 +00:00
All use subject to https://about.jstor.org/terms

the model parameters are independent of those used to construct the forecasts. Although this assumption can be justified asymptotically, Phillips (1979) pointed out that it "is quite unrealistic in practical situations" (p. 241) and went on to look at the distribution of the forecast errors for the AR(1) case *conditional* on the final observed value x_n. The resulting mean squared error (MSE) is called the *conditional* PMSE.

Phillips's results for the conditional PMSE of an AR(1) process were extended, for example, by Fuller and Hasza (1981) to AR(p) processes and by Ansley and Kohn (1986) to state-space models. Because the general ARMA model can be formulated as a state-space model, the latter results also cover ARMA models (and hence AR processes) as a special case. Note that PMSE formulas for regression models are typically of conditional form and *do* allow for parameter uncertainty; see Section 4.2.

From a practical point of view, it is important to know if the effect of incorporating parameter uncertainty into PMSE's has a nontrivial effect. Unfortunately, the literature appears to have made little attempt to quantify the effect.

Consider, for example, a K-variable vector AR(p) process with known mean value. It can be shown that the true-model PMSE at lead time 1 has to be multiplied by the correction factor $[1 + Kp/n] + o(1/n)$ to give the corresponding unconditional PMSE allowing for parameter uncertainty [e.g., Lütkepohl 1991, eq. (3.5.13)]. Thus the more parameters there are, and the shorter the series, the greater will be the correction term, as would intuitively be expected. When $n = 50$, $K = 1$, and $p = 2$, for example, the correction to the square root of PMSE is only 2%. Findley's (1986, table 3.1) bootstrapping results also suggest that the correction term is often small, though results from more complex models suggest it can be somewhat larger. When $n = 30$, $K = 3$, and $p = 2$, for example, the correction to the square root of PMSE rises to 6%. The effect on *probabilities*, however, is much smaller (see Lütkepohl 1991, table 3.1). This may be readily demonstrated with the standard normal distribution in which 95% of values lie in the range ± 1.96, but 93.5% lie in the range $\pm(1.96 \times .94)$ or ± 1.84. Thus a mistaken reduction of 6% in the standard deviation would only reduce the resulting probability from 95% to 93.5%. This nonlinear relation between corrections to square roots of PMSE's, and corrections to the resulting probabilities is worth noting.

In the AR(p) case, the formula for the *conditional* one-step-ahead PMSE will also involve the last p observed values. In particular, when $p = 1$, the correction term involves an expression proportional to x_n^2. Thus, if the last observed value is "large," then the conditional PMSE will be inflated, which is also intuitively reasonable. [There is a natural analogy here with standard linear regression in which the PI for a future observation on the response variable at a "new" value of the explanatory variable, say x_0, relies on an expression for the conditional PMSE that involves a term proportional to $(x_0 - \bar{x})^2$ (e.g., Weisberg [1985, eq. (1.36)], where $\bar{x}$ is the average x value in the sample used to fit the model).

More work needs to be done to assess the size of correction terms in the conditional case for time series models, but few authors actually use the conditional PMSE in computing PI's (except when using regression models), presumably because of the difficulty involved in evaluating the conditional expression and in having to recompute it every time a new observation becomes available.

Overall, the effect of parameter uncertainty seems likely to be of a smaller order of magnitude in general than that due to model uncertainty and the effect of errors and outliers (see Sec. 6). Thus I agree with Granger and Newbold (1986, p. 158) that "for most general purposes, it should prove adequate to substitute the parameter estimates" into the true-model PMSE. For models with many parameters in relation to the length of the observed series, however (e.g., vector ARMA models and econometric simultaneous-equation models), this strategy could lead to a serious underestimate of the length of PI's.

4. PROCEDURES FOR CALCULATING PI's

This section reviews various approaches for calculating PI's.

4.1 Introduction

Most PI's used in practice are essentially of the following general form. A $100(1 - \alpha)\%$ PI for X_{n+k} is given by

$$\hat{x}_n(k) \pm z_{\alpha/2} \sqrt{\operatorname{var}[e_n(k)]}, \qquad (4.1)$$

where $z_{\alpha/2}$ denotes the appropriate percentage point of a standard normal distribution.

Equation (4.1) effectively assumes that the forecast is unbiased with PMSE given by $E[e_n(k)^2] = \operatorname{var}[e_n(k)]$ and that the forecast errors are normally distributed.

When $\operatorname{var}[e_n(k)]$ has to be estimated (as it usually must), some authors (e.g., Harvey 1989, p. 32) suggest replacing $z_{\alpha/2}$ in Equation (4.1) by the percentage point of a t distribution with an appropriate number of degrees of freedom, but this makes little difference except for very short series (e.g., $n < 20$) in which other effects (e.g., model and parameter uncertainty) are likely to be more serious.

The normality assumption may be true asymptotically, but it can be shown that the one-step-ahead conditional forecast-error distribution will not in general be normal, even for a linear model with normally distributed innovations, when model parameters have to be estimated from the same data used to compute forecasts. This also applies to k-steps-ahead errors (Phillips 1979), although the normal approximation does seem to improve as k increases, at least for an AR(1) process. Looking back at Equation (3.3), for example, we have

This content downloaded from
192.12.247.244 on Mon, 25 Mar 2024 08:44:27 +00:00
All use subject to https://about.jstor.org/terms

already noted that the expected value of the conditional distribution of $e_n(1)$ need not be 0, and Phillips (1979) showed that the conditional distribution will generally be skewed. This is another point to bear in mind when computing PI's.

Whether the departure from normality caused by parameter uncertainty is of practical importance seems doubtful. The only guidance offered by Phillips (1979) is that when n is "small" (how small?) the correction to the normal approximation can be "substantial" and that the normal approximation becomes less satisfactory for the AR(1) process in Equation (3.1) as the parameter α increases in size. The correction term is of order $1/n$, however, and seems likely to be of a smaller order of magnitude in general than that due to model uncertainty, and to the effect of errors and outliers and other departures from normality in the distribution of the innovations. Thus, rightly or wrongly, Equation (4.1) is the formula that is generally used to compute PI's, though preferably after checking that the underlying assumptions are at least reasonably satisfied. For any given forecasting method, the main problem will then lie with evaluating var$[e_n(k)]$.

4.2 PI's Derived From a Fitted Probability Model

If the true model for a given time series is known, then it will usually be possible to derive minimum MSE forecasts and the corresponding PMSE and hence evaluate PI's, probably using Equation (4.1). In practice, the true model will not be known and must be *formulated from the data*. This is usually done by prespecifying a broad class of models, such as ARIMA or state-space models, and choosing the most appropriate model from within that class. The practitioner then typically acts as though the selected model is the true model. In addition, the practitioner typically ignores the effect of parameter uncertainty (see Sec. 3) and acts as though the estimated model parameters are the true values. Thus it is the true-model PMSE that is usually substituted into Equation (4.1). Note that models are customarily fitted by minimizing one-step-ahead "errors" in some way even when k-step-ahead forecasts are required. This is valid provided that one has specified the correct model (Stoica and Nehorai 1989).

True-model PMSE's are available for many classes of model. Perhaps the best-known equation is for Box–Jenkins ARIMA forecasting. By writing an ARIMA model in infinite-moving average form as $X_t = \varepsilon_t + \Psi_1\varepsilon_{t-1} + \Psi_2\varepsilon_{t-2} + \ldots$, it can be shown that

$$\text{var}[e_n(k)] = [1 + \Psi_1^2 + \cdots + \Psi_{k-1}^2]\sigma_\varepsilon^2. \quad (4.2)$$

Although this is a true-model PMSE, we would of course have to insert estimates of $\{\Psi_i\}$ and σ_ε^2 into this equation to use it in practice. The combination of Equations (4.1) and (4.2) is sometimes the one and only equation for PI's given in textbooks.

True-model PMSE's can also be derived for vector ARMA models (e.g., Lütkepohl 1991, Secs. 2.2.3 and 6.5) and for structural state-space models [Harvey 1989, eq. (4.6.3)].

Note also that PMSE formulas are available for various regression models (e.g., Kendall and Ord 1990, eq. 12.32; Miller 1990, sec. 6.2; Weisberg 1985) and that these formulas typically allow for parameter uncertainty and are conditional in the sense that they depend on the particular values of the explanatory variables from where a prediction is being made. The regression results can be applied when the explanatory variable is "time" as in the global-constant-mean and linear-trend models. I do not believe, however, in models that include deterministic functions of time. Current thinking [e.g., Chatfield 1989, chap. 10; Newbold 1988] generally prefers *local* to *global* models so that trend, for example, is allowed to evolve through time in a stochastic rather than deterministic way.

Finally, I mention two classes of model in which the preceding approach may not be available. For some complicated simultaneous-equation econometric models, it is not possible to derive var$[e_n(k)]$, particularly when some of the equations incorporate nonlinear relationships and when judgment is used in producing forecasts (e.g., in specifying future values of exogenous variables). Then an empirically based approach must be used; see Sections 4.5–4.6. Equation (4.1) is also inappropriate for nonlinear models (e.g., Granger and Newbold 1986, chap. 10; Tong 1990, especially chap. 6) such as threshold and bilinear models. Rather little work has been done on forecasting aspects of such models. It can be difficult to evaluate conditional expectations more than one step ahead. In addition, the width of PI's need not necessarily increase with lead time. Perhaps the most serious difficulty from the point of view of calculating PI's is that the predictive distribution will not in general be normal and may, for example, be bimodal. In the latter case, a single point forecast could be particularly misleading (see Tong 1990, fig. 6.1) and a sensible PI could even comprise two disjoint intervals. Here there seems little alternative to evaluating the complete predictive distribution, even though this may be computationally demanding. Note that economists also work with what they call nonlinear models, though they are different in kind to the mainly univariate models considered by Tong (1990). Note also that nonlinear transformations of variables introduce nonlinearities (see Sec. 4.8).

4.3 PI's Derived by Assuming That a Method is Optimal

As noted in Section 2, a forecasting method is sometimes selected without applying any formal model-identification procedure (although one should certainly choose a method that does or does not cope with trend and/or seasonality as appropriate). The question then arises as to whether PI's should be calculated by some empirical procedure (see Secs. 4.5–4.6) or by assuming that the method is "optimal" in some sense.

This content downloaded from
192.12.247.244 on Mon, 25 Mar 2024 08:44:27 +00:00
All use subject to https://about.jstor.org/terms

Consider exponential smoothing (ES), for example. This well-known forecasting procedure can be used for series showing no obvious trend or seasonality without trying to identify the underlying model. Now ES is well known to be optimal for an ARIMA (0, 1, 1) model or for the structural (state-space) steady model, and both of these models lead to the true-model PMSE formula (Box and Jenkins 1970, p. 145; Harrison 1967)

$$\operatorname{var}[e_n(k)] = [1 + (k - 1)\alpha^2]\sigma_e^2, \qquad (4.3)$$

where α denotes the smoothing parameter and $\sigma_e^2 = \operatorname{var}[e_n(1)]$ denotes the variance of the one-step-ahead forecast errors. Should this formula then be used in conjunction with Equation (4.1) for ES even though a model has not been formally identified? My answer would be that it is reasonable (or at least not unreasonable) to use Equation (4.3) provided that the observed one-step-ahead forecast errors show no obvious autocorrelation and provided that there are no other obvious features of the data (e.g., trend) that need to be modeled.

There are, however, some PI formulas for ES that I would argue should be disregarded. Montgomery and Johnson (1976) and Abraham and Ledolter (1983) follow Brown (1963) in deriving formulas for various smoothing methods based on a general regression model. In particular, it can be shown that ES arises from the model

$$X_t = \beta + \varepsilon_t \qquad (4.4)$$

when β is estimated by discounted least squares and expressed as an updating formula. Then PI's can be found for Model (4.4), assuming that β is constant. Although these formulas can take account of the sampling variability in $\hat{\beta}$, they have the unlikely feature that they are of constant width as the lead time increases [e.g., Abraham and Ledolter 1983, Eq. (3.60); Bowerman and O'Connell 1987, p. 266]. Intuitively this is not sensible. It arises because β is assumed constant in Equation (4.4). But if this were true, then ordinary least squares should be used to estimate it. The use of discounted least squares suggests that β is thought to be changing. If indeed β follows a random walk, then we are back to the structural steady model for which Equation (4.3) is indeed appropriate. More generally, the many formulas given in the literature based on general ES derived by applying discounted least squares to global models such as (4.4) should be disregarded because ordinary least squares is optimal for a global model (Abraham and Ledolter 1983, p. 126). As a related example, McKenzie (1986) derived the variance of the forecast error for the Holt–Winters method with additive seasonality by employing a deterministic trend-and-seasonal model for which Holt–Winters is not optimal. These results should also be disregarded.

Some forecasting methods are not based, even implicitly, on a probability model. What can be done then? Suppose that we assume that the method is optimal in the sense that the one-step-ahead errors are uncorrelated (this can easily be checked by looking at the correlogram of the one-step-ahead errors; if there is correlation, then there is more structure in the data that it should be possible to "capture" so as to improve the forecasts). From the updating equations, it may be possible to express $e_n(k)$ in terms of the intervening one-step-ahead errors—namely, $e_n(1)$, $e_{n+1}(1), \ldots, e_{n+k-1}(1)$. If we assume that the one-step-ahead errors are not only uncorrelated but also have equal variance, then it should be possible to evaluate $\operatorname{var}[e_n(k)]$ in terms of $\operatorname{var}[e_n(1)]$. It may also be possible to examine the effects of alternative assumptions about $\operatorname{var}[e_n(1)]$.

Yar and Chatfield (1990) and Chatfield and Yar (1991) have applied this approach to the Holt–Winters method with additive and multiplicative seasonality, respectively. In the additive case, it is encouraging to find that the results turn out to be equivalent to those resulting from the seasonal ARIMA model for which additive Holt–Winters is optimal (although this model is so complicated that it would never be identified in practice). The results in the multiplicative case are of particular interest because $\operatorname{var}[e_n(\mathrm{k})]$ does not necessarily increase monotonically with k. Rather PI's tend to be wider near a seasonal peak, as would intuitively be expected, and more self-consistent results are obtained if the one-step-ahead error variance is assumed to be proportional to the seasonal effect rather than constant. The phenomenon of getting wider PI's near a seasonal peak is not captured by most alternative approaches (except perhaps by using a variance-stabilizing transformation). The lack of monotonicity of $\operatorname{var}[e_n(k)]$ with k is typical of behavior resulting from nonlinear models (see Tong 1990, chap. 6, and comments in sec. 4.2) and arises because multiplicative Holt–Winters is a nonlinear method in that forecasts are not a linear combination of past observations.

4.4 PI's Based on "Approximate" Formulas

For some forecasting methods, theoretical PI formulas are not available. As one alternative, a variety of approximate formulas have been suggested, either for forecasting methods in general or for specific methods. Because of their simplicity, they have sometimes been used even when better alternatives *are* available. This is most unfortunate given that the approximations will be shown to be poor. Some readers may think that the proposed approximations are too silly to warrant serious discussion, but they *are* being used and do need to be explicitly repudiated.

1. One general "approximate" formula for the PMSE is that

$$\operatorname{var}[e_n(k)] = k\,\sigma_e^2, \qquad (4.5)$$

where $\sigma_e^2 = \operatorname{var}[e_n(1)]$ denotes the variance of the one-step-ahead forecast errors. This formula is then substituted into Equation (4.1) to give PI's. Equation (4.5)

This content downloaded from 192.12.247.244 on Mon, 25 Mar 2024 08:44:27 +00:00
All use subject to https://about.jstor.org/terms

was given by Makridakis, Hibon, Lusk, and Belhadjali (1987 Eq. 1), Lefrancois (1989, Eq. 1), and verbally by Makridakis and Winkler (1989, p. 336). It is stated to depend on an unchanging model having independent normal errors with zero mean and constant variance. In fact these assumptions are not enough, and Equation (4.5) is only true for a random-walk model (Koehler 1990). For other methods and models, it can be seriously in error (e.g., see Table 1 in Sec. 5 and Yar and Chatfield 1990) and should not be used.

2. Equation (4.5) may have arisen from confusion with a similar-looking approximation for the error in a cumulative forecast (see Lefrancois 1990 and the reply by Chatfield and Koehler 1991). Let

$$E_n(k) = e_n(1) + \cdots + e_n(k)$$
$$= \text{error in cumulative demand over next } k \text{ periods.}$$

Brown (1963, p. 239) suggested verbally that

$$\operatorname{var}[E_n(k)] = \textstyle\sum_{i=1}^{k} \operatorname{var}[e_n(i)], \qquad (4.6)$$

and by assuming that $\operatorname{var}[e_n(i)]$ is a constant (!??) it is but a short step to the approximation (e.g., Johnston and Harrison 1986, p. 304) that

$$\operatorname{var}[E_n(k)] = k\,\sigma_e^2. \qquad (4.7)$$

Equation (4.6), however, ignores the correlations between errors in forecast made from the same time origin, and Brown's statement has been the source of much confusion. There is no theoretical justification for Equation (4.7), which, like Equation (4.5), can give very inadequate results.

3. Brown (1967, p. 144) also proposed an alternative approximation for the variance of the cumulative error—namely, that

$$\operatorname{var}[E_n(k)] = (.659 + .341k)^2\sigma_e^2. \qquad (4.8a)$$

This approximation, like Equations (4.5) and (4.7), cannot possibly be accurate for all methods and models and may also be seriously inadequate. Recently Makridakis et al. (1987, eq. 2) cited Brown's formula but applied it not to the cumulative error but to a single forecast error so that they effectively take

$$\operatorname{var}[e_n(k)] = (.659 + .341k)^2\sigma_e^2. \qquad (4.8b)$$

This appears to be a simple error from misreading Brown's book and is another example of the confusion between single-period and cumulative forecasts (Chatfield and Koehler 1991). Equation (4.8b) should therefore not be used.

4. Only one example of an approximation aimed at a specific method will be given here. Bowerman and O'Connell (1987, sec. 6.4) gave approximate formulas for PI's for the Holt–Winters method. The formulas are rather complicated and depend on the maximum of the three smoothing parameters. As such, they appear to be producing conservative limits in some way, but the exact reasoning behind these formulas is unclear. They are not compatible with the exact results given by Yar and Chatfield (1990) and Chatfield and Yar (1991). (Note that Bowerman and O'Connell's [1987] formulas look unfamiliar because they effectively estimate σ_e as 1.25 times the mean absolute one-step-ahead forecasting error over the fit period, which is a standard alternative to the use of root mean squared error.)

4.5 Empirically Based PI's

When theoretical formulas are not available, a more promising approach than the use of "approximate" formulas involves using the properties of the *observed* distribution of "errors."

1. The simplest type of procedure (e.g., Gilchrist 1976, p. 242) involves applying the forecasting method to all of the past data, finding the within-sample "forecast" errors at 1, 2, 3, . . . steps ahead from all available time origins, and then finding the variance of these errors at each lead time over the period of fit. Let $s_{e,k}$ denote the standard deviation of the k-steps-ahead errors. Then an approximate empirical 100 $(1 - \alpha)$% PI for X_{n+k} is given by $\hat{x}_n(k) \pm z_{\alpha/2}s_{e,k}$. If n is small, $z_{\alpha/2}$ could be replaced by an appropriate value from a t distribution. This approach often seems to work reasonably well and gives results comparable to theoretical formulas when the latter are available (Bowerman and Koehler 1989; Yar and Chatfield 1990). The value of $s_{e,k}$ is unreliable for small n and large k, however, and is based on *model-fitting errors* rather than true post-sample forecast errors. There is evidence that the characteristics of these two types of error may not be the same (e.g., Makridakis and Winkler 1989) in that true forecast errors tend to be somewhat larger on average; see Section 6.

2. An earlier method (Williams and Goodman 1971) involves splitting the past data into two parts. Fit the method or model to the first part and make predictions of the second part. The resulting "errors" are much more like true forecast errors than those in property 1, especially for long series in which model parameters can be estimated with high precision. The model is then refitted with one additional observation in the first part and one less in the second part and so on. Williams and Goodman found that the distribution of forecast errors approximated a gamma distribution rather than a normal distribution. PI's were constructed using the percentage points of the empirical distribution, thereby avoiding any distributional assumptions. Promising results were obtained. Although the approach is attractive in principle, however, it seems to have been little used in practice, presumably because of the heavy computational demands. The latter problem, however, has not prevented developments as in Section 4.6, and it may be that the Williams–Goodman method was ahead of its time and should now be reexamined.

4.6 Simulation and Resampling Methods

This type of approach is even more computationally intensive than that described in Section 4.5 but is in-

This content downloaded from
192.12.247.244 on Mon, 25 Mar 2024 08:44:27 +00:00
All use subject to https://about.jstor.org/terms

creasingly used for the construction of PI's (and many other problems). The approach can be used when theoretical formulas are not available, for short series when only asymptotic results are available and when there are doubts about model assumptions.

Given a probability time series model, it is possible to *simulate* both past and future behavior by generating an appropriate series of random innovations and hence constructing a sequence of possible past and future values. This process can be repeated many times, leading to a large set of possible sequences, sometimes called pseudodata. From such a set, it is possible to evaluate PI's at different horizons. The use of simulation is sometimes called a *Monte Carlo* approach. It generally assumes that the model has been identified correctly.

Instead of sampling the innovations from some assumed parametric distribution (usually normal), an alternative is to sample from the empirical distribution of past fitted "errors." This is called *resampling* or *bootstrapping*. The procedure effectively approximates the theoretical distribution of innovations by the empirical distribution of the observed residuals. Thus it is a distribution-free approach. It may also be possible to extend the use of resampling to forecasting methods that are not based on a proper probability method but rely instead on a set of recursive equations involving observed and forecast values.

The literature in this area, particularly in bootstrapping, is growing rapidly. Veall (1989) reviewed the use of computationally intensive methods for complex econometric models, where they are particularly useful, and gave many references. He suggested that "most applied econometric exercises should include bootstrapping or some other form of simulation as a check" (p. 77). Bootstrapping can of course be used for other aspects of time series analysis such as evaluating the standard errors of estimates of model parameters (e.g., Freedman and Peters 1984a,b), where the normal assumption may be less critical than in the evaluation of PI's (Veall 1989, sec. 3.2). It can be a mistake, however, to concentrate on departures from the secondary model assumptions (e.g., normal errors) when departures from the primary assumptions (or specification error) can be much more serious.

One classic simulation study reviewed by Veall (1989) is that of Fair (1980), who showed how to assess four sources of uncertainty in forecasts from econometric models—namely, (1) the model innovations, (2) having estimates of model parameters rather than true values, (3) having forecasts of exogenous variables rather than true values, and (4) misspecification of the model. Fair sampled from a multivariate normal distribution for two example models, one a large (97 equations!) model and the other a much simpler autoregressive model.

I now concentrate on references not covered by Veall (1989). Early work on the use of resampling for calculating PI's (e.g., Butler and Rothman 1980) was for regression models and not really concerned with time series forecasting. Freedman and Peters (1984b, sec. 6) gave one example of forecasting no less than 24 years ahead and showed how to compute "standard errors of forecasts." Peters and Freedman (1985) showed how to use bootstrapping to get standard errors for multistep-ahead forecasts for the complex 10-equation model of Freedman and Peters (1984a). They showed that the results are more reliable for short series than those given by the so-called delta method (Schmidt 1977). Note that exogenous variables are forecast by some process external to the equation. Bianchi, Calzolari, and Brillet (1987) discussed various simulation and resampling methods as applied to a large macro model of the French economy involving over 20 equations. Findley (1986) showed how to compute bootstrap estimates of the unconditional PMSE and the conditional PMSE for an AR(p) process. Since forecasts for an AR(p) model depend on the last p observed values, Findley said that it is the error associated with conditional predictions from sample paths through these last p observations that is usually of most interest, but he went on to say that satisfactory methods for obtaining such sample paths were not then available. Latterly, Stine (1987) and Thombs and Schucany (1990) have both shown how to overcome this problem for AR processes by using the *backward* representation of the series conditional on the last p observations. For example, in the zero-mean AR(1) case, fix y_n equal to the latest observation, x_n and generate backward sample paths from [c.f. Eq. (3.1)]

$$y_t = \hat{\alpha} y_{t+1} + \hat{\varepsilon}_t \tag{4.9}$$

for $t = (n - 1), (n - 2), \ldots$, where $\hat{\alpha}$ is the least squares estimate of α from the (original) observed series and $\hat{\varepsilon}_t$ are iid samples from the empirical distribution of the observed backward residuals. Each sample path can then be used to reestimate the parameter α in Equation (3.1), after which conditional bootstrap replicates of the future can be constructed using the bootstrap estimate of α and further random drawings from the empirical distribution of forward residuals. Full details are in an article by Thombs and Schucany (1990). Stine (1987) looked at unconditional PI's for AR(p) processes, as well as conditional PI's, and carried out various simulations. He showed that bootstrap PI's compare favorably with normal-based PI's, particularly when the innovations are not normal as would intuitively be expected. Thombs and Schucany (1990) also simulated various AR(1) and AR(2) models with innovations that are normal, exponential, or a mixture of two normals. They also concluded that bootstrap PI's are a useful nonparametric alternative to the usual Box–Jenkins intervals. Masarotto (1990) also looked at bootstrap PI's for AR models but appeared to only consider the unconditional (and less interesting?) case. Most authors have assumed that the order of the AR process is known, but Masarotto did explicitly take into account that the order, as well as the model parameters, is generally

This content downloaded from
192.12.247.244 on Mon, 25 Mar 2024 08:44:27 +00:00
All use subject to https://about.jstor.org/terms

unknown. Masarotto presented simulation results for AR(1), AR(3), and AR(5) processes with innovations that are normal or from the centered extreme-value distribution. One feature of interest is that innovations are sampled not only from a parametric distribution and from the empirical distribution of residuals but also from a smoothed version of the latter using a kernel-density estimate. There seems to be little difference in the results from the last two types of distribution, so the extra computation needed to smooth the empirical residual distribution was not worthwhile in this case.

Monte Carlo simulation was also used by Pflaumer (1988) to calculate PI's for population projections by letting fertility and net immigration rates vary as random variables with specified distributions. This is arguably superior to the "alternative scenarios" approach where "high," "medium," and "low" assumptions are made about different components, leading to a range of possible population trajectories. No probabilities can be attached to the latter, however, and a single PI may well be easier to interpret. An alternative approach for a large (nonlinear) econometric model, which involves stochastic perturbation of input variables, was described by Corker, Holly, and Ellis (1986).

Finally, I mention that Thompson and Miller (1986) also used simulation, but this will be discussed in Section 4.7.

4.7 The Bayesian Approach

I am not a Bayesian, so the following brief summary should be read with this in mind. In principle, the Bayesian approach will give the complete probability distribution for a future value. From this distribution, it should be possible to derive interval forecasts, either by a decision-theoretic approach along the lines of Winkler (1972) or (more usually) by calculating symmetric intervals, using the Bayesian version of Equation (4.1), when the predictive distribution is normal. If the predictive distribution has some other symmetric distribution (e.g., Student-t), then Equation (4.1) can readily be adapted by inserting appropriate percentiles. Unfortunately the multiperiod-ahead predictive density does not have a convenient closed form for many forecasting models, so Bayesian statisticians may need to use some sort of approximation when interval forecasts are required (Thompson and Miller 1986, sec. 3). Alternatively, it is possible to simulate the predictive distribution rather than try to obtain or approximate its analytic form. Thompson and Miller (1986) compared the resulting percentiles of the simulated predictive distribution for an AR(2) process with the Box–Jenkins PI's based on Equations (4.1) and (4.2). As the latter do not allow for parameter uncertainty, the Bayesian intervals are naturally somewhat wider (although they still do not include all the ensuing observed values; see Sec. 7).

The most comprehensive description of Bayesian forecasting was given by West and Harrison (1989), based on a general class of models called *dynamic linear models*, which are much more amenable to the Bayesian approach than ARIMA models, for example. Unfortunately, West and Harrison said little explicitly about interval forecasts although they are computed in several examples. It seems to be assumed implicitly that the mean and variance of the forecast distribution are substituted into Equation (4.1), together with normal or t percentiles as appropriate. The results are called "probability limits" (p. 262), "prediction intervals" (p. 329), "symmetric intervals" (p. 342), "intervals for the one-step-ahead forecast" (p. 343), and "forecast intervals" (p. 396). If the error terms in the observation and system equations are assumed to be normal with *known* variances and (conjugate) normal priors are assumed, then forecast distributions will also be normal. The error variances, however, will generally be unknown (as will the parameters in the corresponding ARIMA model), and if they are allowed to evolve as a normal process, then a Student-t distribution will result. All in all, there are a lot of assumptions, but the identification process seems (to me) to lack the cohesive strength of the Box–Jenkins approach. There will, of course, be certain situations in which a "dynamic" or local model is indicated, but then the non-Bayesian may prefer the conceptual approach of Harvey's (1989) structural modeling.

4.8 PI's for Transformed Variables

Whichever approach is used to calculate PI's, the possibility of working with a transformed variable needs to be considered (e.g., Granger and Newbold 1986, sec. 10.5; West and Harrison 1989, sec. 10.6). It may be sensible for a variety of reasons to work not with the observed variable X_t but with some nonlinear transformation of it, say $Y_t = g(X_t)$, where g may, for example, be the logarithmic transformation or the more general Box–Cox transformation. This may be done to stabilize the variance, to make the seasonal effect additive, to make the data more normally distributed, or because the transformed variable "makes more sense."

PI's may be calculated for Y_t in an appropriate way, but the literature says very little about transforming the PI's back to get PI's for the original observed variable if this is desired. (Collins [1991] was an exception, but he considered regression models). It is well known that the "naive" point forecast of X_{n+k}—namely, $g^{-1}[\hat{y}_n(k)]$—is generally not unbiased, essentially because the expected value $E[g^{-1}(Y)]$ is not generally equal to $g^{-1}[E(Y)]$. If the predictive distribution of Y_{n+k} is symmetric with mean $\hat{y}_n(k)$, then $g^{-1}[\hat{y}_n(k)]$ will be the *median*, rather than the mean, of the predictive distribution of X_{n+k}. A naive PI can also be constructed by retransforming the upper and lower values of the PI for Y_{n+k}. If the PI for Y_{n+k} has a prescribed probability, say $(1 - \alpha)$, then the retransformed PI for X_{n+k} should have the same prescribed probability (e.g., Harvey 1989, eq. 4.6.7) apart from any additional uncertainty introduced when the identification of the transformation, g,

This content downloaded from 192.12.247.244 on Mon, 25 Mar 2024 08:44:27 +00:00
All use subject to https://about.jstor.org/terms

involves estimating one or more parameters. Fortunately, the results of Collins (1991) suggest that uncertainty about the Box–Cox transformation parameter may be relatively unimportant. Note also that Collins's results indicate that model parameter estimates are likely to be correlated with the transformation parameter when the latter has to be estimated.

Note that Collins (1991) used the description "plug-in density" to describe estimates arising from replacing unknown parameters with their estimated values and the term "deterministic prediction" to describe forecasts made by applying the inverse Box–Cox transformation to forecasts of Y_t, which *do* allow for parameter uncertainty (but I do not understand why this is termed "deterministic").

If the PI for Y_{n+k} is based on a normal assumption and hence symmetric, then the transformed PI for X_{n+k} will be asymmetric (which can be intuitively sensible). Note also that the width of the transformed PI will depend on the level as well as the lead time and the variability. It may be possible to derive symmetric PI's (Collins [1991] referred to this as a "mean-squared error analysis") but this is not sensible unless X_t is thought to be normally distributed—in which case a transformation will probably not be indicated anyway. [But note that some forecasting programs do retransform the point forecast and its standard error to compute symmetric PI's using Equation (4.1). This is not recommended!]

Although point forecasts may be affected relatively little by whether or not a transformation is used, the PI's will typically be affected much more, particularly in being asymmetric. My own preference, stemming from early problems (Chatfield and Prothero 1973), is to avoid transformations wherever possible except where the transformed variable is of interest in its own right (e.g., taking logarithms to analyze percentage increases) or is clearly indicated by theoretical considerations (e.g., Poisson data have nonconstant variance). Thus I agree with West and Harrison (1989, p. 360) that "uncritical use of transformations for reasons of convenience should be guarded against" and that generally "it is preferable to model the series on the original data scale."

4.9 Judgmental PI's

Judgment may be used in time-series forecasting not only to produce point forecasts but also to produce PI's. Empirical evidence (e.g., Armstrong 1985, pp. 138–145; O'Connor and Lawrence 1989, 1992) suggested that the PI's will generally be too narrow, indicating overconfidence in the forecasts. This topic is outside the main time series emphasis of this article and will not be pursued here.

5. A COMPARATIVE ASSESSMENT

The choice of procedure for calculating PI's in a particular situation depends on various factors. The most important of these is the choice of forecasting method, which depends in turn on such factors as the objectives and types of data (see Sec. 2). Yar and Chatfield (1990) compared several approaches using additive Holt–Winters as the reference method. This section makes a more general comparison.

Theoretical PI formulas based on a fitted probability model are easy to implement, but they do assume the fitted model to be "true," not only in regard to the primary assumptions [e.g., X_t follows an AR(p) process], but also in regard to the secondary "error" assumptions [e.g., ε_t are NID(0, σ^2)]. Allowance can be made for parameter uncertainty, but the formulas are complex and generally give corrections of order $1/n$ and so are rarely used. A more serious problem is that the model may be misspecified or may change in the forecast period. Nevertheless, the formulas are widely used, although they are not available for some complex and/or nonlinear models.

Formulas that simply assume that a given forecasting method is optimal (see Sec. 4.3) are also widely used because of their simplicity, but it is then important to check that the method really is a sensible one (see Sec. 6). Formulas based on a model for which the method is *not* optimal should be disregarded.

Empirically based and resampling methods are always available and require fewer assumptions but can be much more computationally demanding (especially for resampling). Nevertheless, they have much potential promise, particularly when theoretical formulas are not available or there are doubts about the error assumptions. It is important to remember, however, that they do still depend on the primary assumptions of the model. For example, Thombs and Schucany (1990) computed bootstrap PI's under the assumption that an AR(p) model really does fit the data even though no "error" assumptions are made. This is an area in which further research is particularly required.

It is hard to envisage any situation in which the approximate formulas should be used. They have no theoretical basis and cannot possibly capture the varied behavior of PI's. Table 1 shows the ratio of var[$e_n(k)$] to var[$e_n(1)$] for the two approximate formulas in Equations (4.5) and (4.8b), as well as the theoretical results for various methods and models. The disparate relationship with k is evident. For stationary models, such as an AR(1) or MA(1) model, the width of PI's will increase rather slowly with k to a finite upper bound [or even be constant for the MA(1) model), whereas for nonstationary models the width of PI's will increase without bound. This applies to the random-walk and ARIMA(1,1,0) models, to exponential smoothing [optimal for an ARIMA (0,1,1) model] and to Holt's linear trend method [optimal for an ARIMA (0,2,2) model]. The inadequacy of Equations (4.5) and (4.8b) is clear.

6. WHY ARE PI's TOO NARROW?

Wallis (1974) noted that it is a "common experience for models to have worse error variances than they

This content downloaded from
192.12.247.244 on Mon, 25 Mar 2024 08:44:27 +00:00
All use subject to https://about.jstor.org/terms

Table 1. Values of $var[e_n(k)]/var[e_n(1)]$ for Various Equations, Models, and Methods

					Nonstationary models				
	Approximate		Stationary		(ARIMA(0,1,1)) ES		(ARIMA(0,2,2)) Holt linear $\gamma = .1$		ARIMA(1,1,0)
k	(4.5) (or RW)	(4.8b)	AR(1) $\varphi = .5$	MA(1) $\theta = .5$	$\alpha = .3$	$\alpha = .7$	$\alpha = .3$	$\alpha = .9$	$\varphi = .5$
2	2	1.8	1.25	1.25	1.09	1.5	1.1	2.0	3.2
4	4	4.1	1.328	1.25	1.27	2.5	1.4	4.5	9.8
8	8	11.5	1.333	1.25	1.63	4.4	2.2	12.3	25.3
12	12	22.6	1.333	1.25	1.99	6.4	3.6	24.7	41.3

NOTE: The approximate formulas in Equations (4.5) and (4.8b) are k and $(.659 + .341k)^2$, respectively; Formula (4.5) gives the same ratios as the random walk (RW) or ARIMA(0,1,0) model: the AR(1), MA(1), and ARIMA(1,1,0) models use the notation of Box and Jenkins (1970); the smoothing parameters for exponential smoothing (ES) and Holt's linear smoothing use the notation of Chatfield and Yar (1988); ES is optimal for an ARIMA(0,1,1) model, and Holt's linear method is optimal for an ARIMA(0,2,2).

'should' when used in forecasting outside the period of fit" (p. 156). Subsequent empirical studies have generally borne this out by showing that PI's tend to be too narrow on average. In other words, more than 5% of future observations will typically fall outside the 95% PI's on average. Newbold and Granger (1974, p. 161) reported this with respect to Box–Jenkins models, Williams and Goodman (1971) with respect to regression models, Gardner (1988) with respect to empirical formulas, and Makridakis et al. (1987) with respect to both approximate formulas and Box–Jenkins models. Of course, the latter study is flawed in its use of "approximate" formulas (see Sec. 5), but the results for one-step-ahead and for Box–Jenkins should still be relevant.

There are various possible reasons why PI's are too narrow, not all of which need apply in any particular situation. They include the following:

1. Model parameters have to be estimated.
2. For multivariate forecasts, exogenous variables may have to be forecasted.
3. Innovations may not be normally distributed.
4. There may be outliers or errors in the data.
5. Unconditional, rather than conditional, PI's are typically calculated.
6. The wrong model may be identified.
7. The underlying model may change, either during the period of fit or in the future.

Problem 1 can often be dealt with by using PMSE formulas incorporating correction terms for parameter uncertainty, though the corrections are typically of order $1/n$ and smaller than those due to other sources of uncertainty.

Problem 2 goes some way toward explaining why multivariate forecasts need not be as accurate as univariate forecasts, contrary to many people's intuition (e.g., Ashley 1988). PI's for multivariate models can be found that take account of the need to forecast other endogenous variables and also exogenous variables when the latter are forecast separately using a probability model (Lütkepohl 1991, sec. 10.5.1). When future values of the exogenous variables are assumed known or "guessed," however (e.g., assumed to grow at a constant inflation rate), then the PI's will not take account of this additional uncertainty.

Problem 3 may be circumvented by adopting an alternative parametric distribution, though this may be analytically intractable, or by simulating the empirical distribution of innovations. Problems 3 and 4 are related since, if outliers occur, then the innovation distribution may not be normal, whereas errors may also appear to affect the innovation distribution. It is well known that outliers can have a disastrous effect, both on point forecasts and on PI's (e.g., see Ledolter 1989), when the outlier is near the forecast origin. Fortunately, "the impact of outliers that occur well before the forecast origin is usually discounted rapidly" (Ledolter 1989, p. 233). The presence of outliers and errors will also complicate model identification; see problem 6.

Problem 5 reminds us that even when the innovations are normal the conditional prediction errors need not be normal but will typically be asymmetric and have a larger variance than the usually calculated unconditional value.

Problem 6 may arise for various reasons. In particular, it is always tempting to overfit the data with more and more complicated models to improve the fit, but empirical evidence suggests that more complicated models, which give a better fit, do not necessarily give better forecasts. Indeed, it is strange that we admit model uncertainty by searching for the best fitting model and then ignore this uncertainty by making forecasts as if the fitted model is known to be true. It is clearly essential that different forecasting models and methods be compared on the basis of out-of-sample forecasts rather than on measures of fit.

Problem 6 should, of course, be circumvented whenever possible by carrying out appropriate diagnostic checks. For example, when fitting ARIMA models, model checking is an integral part of the identification process (Box and Jenkins 1970, chap. 8). Even when using a forecasting method that does not depend explicitly on a probability model, checks should still be made on the (possibly implicit) assumptions. In particular, checks should be made on the one-step-ahead fit-

This content downloaded from
192.12.247.244 on Mon, 25 Mar 2024 08:44:27 +00:00
All use subject to https://about.jstor.org/terms

ted errors to see if they (a) are uncorrelated and (b) have constant variance. It may be sufficient to calculate the first-order autocorrelation coefficient and the autocorrelation at the seasonal lag (if there is a seasonal effect). If the values are significantly different from 0 (i.e., exceed about $2/\sqrt{n}$ in modulus), then this suggests that the optimal method or model is not being used and there is more structure to find. To check constant variance, it is a good idea to compare the residual variances in the first and second halves of the data and also to compare periods near a seasonal peak with periods near a seasonal trough. It often seems to be the case that the residual variance increases with the mean level and is higher near a seasonal peak. These features need to be explicitly dealt with (e.g., see Chatfield and Yar 1991) or alternatively the data could be transformed so as to stabilize the variance (see Sec. 4.8).

Problem 7 may arise because of a slowly changing structure or because of a sudden shift or turning point such as that caused by the 1973 oil crisis or the 1990 Gulf war. The prediction of change points is a topic of much current interest but is notoriously difficult to do (see Makridakis 1988).

For all of the preceding reasons, post-sample forecast errors tend to be larger than model-fitting errors, as already noted in Section 4.5. Because of this, Gardner (1988) suggested modifying Equation (4.1) to

$$\hat{x}_n(k) \pm \sqrt{\text{var}[e_n(k)]}/\sqrt{\alpha}, \qquad (6.1)$$

where the constant $1/\sqrt{\alpha}$ (which replaces $z_{\alpha/2}$) is selected using an argument based on Chebychev's inequality. Bowerman and Koehler (1989), however, pointed out that this may give very wide PI's in some cases, which are of little practical use. In any case, they may be unnecessarily wide for reasonably stable series in which the usual normal values will be adequate. On the other hand, when there is a substantial change in the forecast period (e.g., a change in trend), then the Chebychev PI's may still not be wide enough. My own preference is generally to use Equation (4.1) rather than (6.1) but to recognize explicitly that this assumes the future to be like the past with all the dangers that entails. Thus there is no general method for taking problem 7 into account. As regards problem 6, it is worth noting that bootstrapping may be able to take account of the possibility of identifying the wrong model from within a given class (e.g., Masarotto 1990).

But whatever checks are made and whatever precautions are taken, it is still impossible to be certain that one has fitted the correct model or to rule out the possibility of structural change in the present or future, and problems 6 and 7 are, in my view, the most important reasons why PI's are too narrow. Section 7 gives an instructive example in which two plausible models give substantially different PI's for the same data.

7. AN EXAMPLE

This example is designed not to compare different approaches to computing PI's (see Sec. 5) but to illustrate the overriding importance of "good" model identification. The data shown in Figure 1 were analyzed by Thompson and Miller (1986) to compare Box–Jenkins PI's with a Bayesian approach that simulates the predictive distribution. The fourth quarter in 1979 was taken as the base month and forecasts were computed up to 12 steps (3 years) ahead. The point forecasts were generally poor because of the large (unforeseeable?) increase in unemployment. The Bayesian PI's were somewhat wider than the Box–Jenkins PI's because the Bayesian method allows for parameter uncertainty. The Bayesian PI's still fail to include the actual values for 10, 11, and 12 steps ahead, however. Thompson and Miller (1986) went on to hypothesize a shift in level that does produce better forecasts. But could this hypothesis reasonably have been made in 1979? If it could, then perhaps a multivariate model, including any known leading variable(s), should have been constructed. (Indeed some readers may think that such a model is intrinsically more sensible anyway than a univariate model, but that is not the main point of this example.

Thompson and Miller (1986) fitted an AR(2) model with a nonzero mean. Looking at Figure 1, it seemed to me that the series is nonstationary, rather than stationary. This view is reinforced by the slowly decreasing autocorrelation function (ACF) that is positive up to lag 9. The ACF of the first differences could be explained by an AR(1) model, given that autocorrelations less than about $\pm 2/\sqrt{48} \simeq .3$ are not significantly different from 0. An ARIMA(1,1,0) model was therefore fitted. The resulting sum of squared residuals is a little higher (4.08 rather than 3.64), but diagnostic checks on the residuals suggest that the model is adequate [the modified Box–Pierce χ^2 statistic of 12.3 on 11 df is actually better than the 12.2 on 10 df for the AR(2) model].

The point forecasts from the ARIMA(1,1,0) model are perhaps a little better than those from the AR(2) model (e.g., 6.00 versus 5.69 for 12 steps ahead), but it is the PI's that concern us here. At 12 steps ahead, for example, the 95% PI for the ARIMA(1,1,0) model

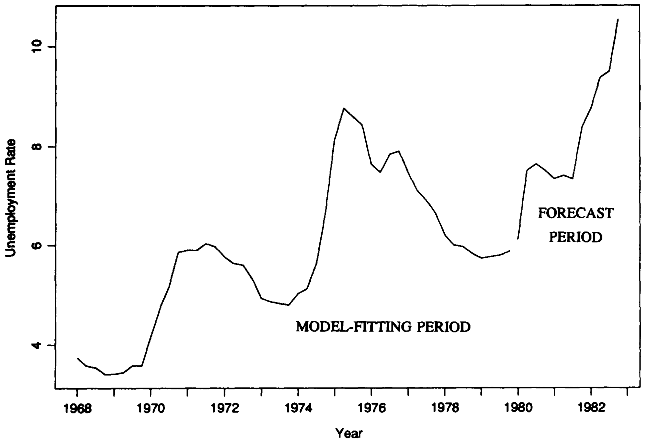

Figure 1. U.S. Quarterly Unemployment Rate 1968–1982 (seasonally adjusted).

This content downloaded from 192.12.247.244 on Mon, 25 Mar 2024 08:44:27 +00:00
All use subject to https://about.jstor.org/terms

is (.86, 11.14), but that for the AR(2) model is (2.66, 8.73). Thus the PI's from the ARIMA(1,1,0) model are much wider and *do* include the actual values. Wide PI's are sometimes seen as indicating "failure," either to fit the right model or to get a usable interval, but here the wider PI's are realistic in allowing for high uncertainty.

The crucial difference between an AR(2) and an ARIMA(1,1,0) model is that the first is stationary, but the second is nonstationary. For stationary processes, var$[e_n(k)]$ tends to the variance of the process as $k \rightarrow \infty$. In other words, PI's will tend to a constant finite width as the lead time increases. But for nonstationary processes, there is no upper bound to the width of PI's. This was noted in passing by Box and Jenkins (1970, p. 151), but the result does not appear to have received the attention it deserves. The greater the degree of differencing needed to make a series stationary, the greater in general will be the divergence of the width of the resulting PI's. Thus, for an ARIMA(0,1,1) process, var$[e_n(k)]$ is of order k, but for an ARIMA(0,2,2) process, var$[e_n(k)]$ is of order k^3 [Box and Jenkins 1970, eqs. (5.4.7) and (5.4.14)].

There is a similar dichotomy for multivariate time series models. For stationary series, the PMSE for each series tends to its variance, but that for nonstationary series increases without bound (e.g., Lütkepohl 1991, p. 377), although it is relevant to note that for nonstationary series that are "tied together" or cointegrated, a multistep forecast that satisfies the cointegrating relationship will have a *finite* limiting PMSE (Engle and Yoo 1987).

Returning to the example, the wider PI's for the nonstationary ARIMA(1,1,0) model seem to capture the observed uncertainty better than the narrower stationary alternative, given that a univariate model is to be fitted. The difference in the PI widths is much larger than that resulting from parameter uncertainty for example, and confirms the overriding importance of identifying the "correct" model, particularly in regard to deciding whether the data are stationary or not.

8. SUMMARY AND RECOMMENDATIONS

The computation of interval forecasts can be of vital importance in planning and deserves more attention. A variety of approaches for computing them has been described and compared. The main findings and recommendations can be summarized as follows:

1. Various "approximate" formulas can be very inaccurate and should not be used.
2. The distinction between a forecasting *method* and a forecasting *model* should be borne in mind. The former may, or may not, depend (explicitly or implicitly) on the latter.
3. A theoretically satisfying way of computing PI's is to formulate a model that provides a reasonable approximation for the process generating a given series, derive the resulting PMSE, and then use Equation (4.1). A correction term to allow for parameter uncertainty can be incorporated in the PMSE, but this is usually of order $1/n$. This approach assumes not only that the "correct" model has been fitted but also that "errors" are normally distributed. The normality assumption for forecast errors implicitly assumes that unconditional results are used. The arguably more correct application of conditional results is rarely used.
4. For large groups of series, an ad hoc forecasting method is sometimes selected. Then PI formulas are sometimes based on the model for which the method is optimal, but this should only be done after appropriate checks on the one-step-ahead forecasting errors. Conversely, it seems silly to derive PI's from a model for which a given method is *not* optimal as happens, for example, in general ES (see Sec. 4.3).

   Formulas that are based not on a model but on assuming that the method is "optimal" in the sense of giving uncorrelated one-step-ahead errors are also now available for the Holt–Winters method. In the multiplicative case, these results have the interesting property that the PMSE does not necessarily increase with the lead time, such behavior being typical of nonlinear systems.

5. When theoretical formulas are not available or there are doubts about model assumptions, the use of empirically based or resampling methods should be considered as a general-purpose alternative. More research on this sort of approach is to be encouraged.
6. PI's tend to be too narrow in practice for a variety of reasons, not all of which can be foreseen. Rather than systematically widening all intervals as in Equation (6.1), it is recommended that Equation (4.1) be generally used but that the implied assumptions be more clearly recognized. In particular, it assumes that a model has been identified correctly and that the future will be like the past.
7. The example in Section 7 demonstrates the overriding importance of model specification. In particular, for a series that is nearly nonstationary, the difference between the finite limiting PMSE, which results from fitting a stationary model, and the unbounded PMSE, which results from a nonstationary model, is critical.

[*Received June 1990. Revised November 1991.*]

REFERENCES

Abraham, B., and Ledolter, J. (1983), *Statistical Methods for Forecasting*, New York: John Wiley.

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

Armstrong, J. S. (1985), *Long-Range Forecasting* (2nd ed.), New York: John Wiley.

Ashley, R. (1988), "On the Relative Worth of Recent Macroeconomic Forecasts," *International Journal of Forecasting*, 4, 363–376.

Baillie, R. T. (1979), "Asymptotic Prediction Mean Squared Error for Vector Autoregressive Models," *Biometrika*, 66, 675–678.

This content downloaded from
192.12.247.244 on Mon, 25 Mar 2024 08:44:27 +00:00
All use subject to https://about.jstor.org/terms

Bianchi, C., Calzolari, G., and Brillet, J.-L. (1987), "Measuring Forecast Uncertainty," *International Journal of Forecasting*, 3, 211–227.

Bowerman, B. L., and Koehler, A. B. (1989), "The Appropriateness of Gardner's Simple Approach and Chebychev Prediction Intervals," unpublished paper presented at the 9th International Symposium on Forecasting in Vancouver, British Columbia, June 18–20.

Bowerman, B. L., and O'Connell, R. T. (1987), *Time Series Forecasting* (2nd ed.), Boston: Duxbury Press.

Box, G. E. P., and Jenkins, G. M. (1970), *Time-Series Analysis, Forecasting and Control*, San Francisco: Holden-Day (revised ed. published 1976).

Brockwell, P. J., and Davis, R. A. (1987), *Time Series: Theory and Methods*, New York: Springer-Verlag.

Brown, R. G. (1963), *Smoothing, Forecasting and Prediction*, Englewood Cliffs, NJ: Prentice-Hall.

——— (1967), *Decision Rules for Inventory Management*, New York: Holt, Rinehart & Winston.

Butler, R., and Rothman, E. D. (1980), "Predictive Intervals Based on Reuse of the Sample," *Journal of the American Statistical Association*, 75, 881–889.

Chatfield, C. (1988), "What Is the Best Method of Forecasting?" *Journal of Applied Statistics*, 15, 19–38.

——— (1989), *The Analysis of Time Series* (4th ed.), London: Chapman and Hall.

Chatfield, C., and Koehler, A. B. (1991), "On Confusing Lead Time Demand With *h*-period-ahead Forecasts," *International Journal of Forecasting*, 7, 239–240.

Chatfield, C., and Prothero, D. L. (1973), "Box–Jenkins Seasonal Forecasting: Problems in a Case Study" (with discussion), *Journal of the Royal Statistical Society*, Ser. A, 136, 295–352.

Chatfield, C., and Yar, M. (1988), "Holt–Winters Forecasting—Some Practical Issues," *The Statistician*, 37, 129–140.

——— (1991), "Prediction Intervals for Multiplicative Holt–Winters," *International Journal of Forecasting*, 7, 31–37.

Collins, S. (1991), "Prediction Techniques for Box–Cox Regression Models," *Journal of Business & Economic Statistics*, 9, 267–277.

Corker, R. J., Holly, S., and Ellis, R. G. (1986), "Uncertainty and Forecast Precision," *International Journal of Forecasting*, 2, 53–69.

Dalrymple, D. J. (1987), "Sales Forecasting Practices: Results From a United States Survey," *International Journal of Forecasting*, 3, 379–391.

Engle, R. F., and Yoo, B. S. (1987), "Forecasting and Testing in Co-integrated Systems," *Journal of Econometrics*, 35, 143–159.

Fair, R. C. (1980), "Estimating the Expected Predictive Accuracy of Econometric Models," *International Economic Review*, 21, 355–378.

Fildes, R. (1992), "The Evaluation of Extrapolative Forecasting Methods," *International Journal of Forecasting*, 8, 81–98.

Findley, D. F. (1986), "On Bootstrap Estimates of Forecast Mean Square Errors for Autoregressive Processes," in *Computer Science and Statistics: 17th Symposium on the Interface*, ed. D. M. Allen, Amsterdam: North-Holland, pp. 11–17.

Freedman, D. A., and Peters, S. C. (1984a), "Bootstrapping a Regression Equation: Some Empirical Results," *Journal of the American Statistical Association*, 79, 97–106.

——— (1984b), "Bootstrapping an Econometric Model: Some Empirical Results," *Journal of Business & Economic Statistics*, 2, 150–158.

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

Gardner, E. S., Jr. (1988), "A Simple Method of Computing Prediction Intervals for Time-Series Forecasts," *Management Science*, 34, 541–546.

Gilchrist, W. (1976), *Statistical Forecasting*, London: John Wiley.

Granger, C. W. J., and Newbold, P. (1986), *Forecasting Economic Time Series* (2nd ed.), New York: Academic Press.

Hahn, G. J., and Meeker, W. Q. (1991), *Statistical Intervals: A Guide for Practitioners*, New York: John Wiley.

Harrison, P. J. (1967), "Exponential Smoothing and Short-Term Sales Forecasting," *Management Science*, 13, 821–842.

Harvey, A. C. (1989), *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge, U.K.: Cambridge University Press.

Johnston, F. R., and Harrison, P. J. (1986), "The Variance of Lead-Time Demand," *Journal of the Operational Research Society*, 37, 303–308.

Kendall, M., and Ord, J. K. (1990), *Time Series* (3rd ed.), Sevenoaks, U.K.: Edward Arnold.

Koehler, A. B. (1990), "An Inappropriate Prediction Interval," *International Journal of Forecasting*, 6, 557–558.

Ledolter, J. (1989), "The Effect of Additive Outliers on the Forecasts From ARIMA Models," *International Journal of Forecasting*, 5, 231–240.

Lefrancois, P. (1989), "Confidence Intervals for Non-stationary Forecast Errors: Some Empirical Results for the Series in the *M*-Competition," *International Journal of Forecasting*, 5, 553–557.

——— (1990), "Reply to: Comments by C. Chatfield," *International Journal of Forecasting*, 6, 561.

Lütkepohl, H. (1991), *Introduction to Multiple Time Series Analysis*, Berlin: Springer-Verlag.

Makridakis, S. (1988), "Metaforecasting," *International Journal of Forecasting*, 4, 467–491.

Makridakis, S., Hibon, M., Lusk, E., and Belhadjali, M. (1987), "Confidence Intervals: An Empirical Investigation of the Series in the *M*-competition," *International Journal of Forecasting*, 3, 489–508.

Makridakis, S., and Winkler, R. L. (1989), "Sampling Distributions of Post-sample Forecasting Errors," *Applied Statistics*, 38, 331–342.

Masarotto, G. (1990), "Bootstrap Prediction Intervals for Autoregressions," *International Journal of Forecasting*, 6, 229–239.

McKenzie, E. (1986), "Error Analysis for Winters' Additive Seasonal Forecasting System," *International Journal of Forecasting*, 2, 373–382.

Miller, A. J. (1990), *Subset Selection in Regression*, London: Chapman & Hall.

Montgomery, D. C., and Johnson, L. A. (1976), *Forecasting and Time Series Analysis*, New York: McGraw-Hill.

Newbold, P. (1988), "Predictors Projecting Linear Trend Plus Seasonal Dummies," *The Statistician*, 37, 111–127.

Newbold, P., and Granger, C. W. J. (1974), "Experience With Forecasting Univariate Time-Series and the Combination of Forecasts" (with discussion), *Journal of the Royal Statistical Society*, Ser. A, 137, 131–165.

O'Connor, M., and Lawrence, M. (1989), "An Examination of the Accuracy of Judgmental Confidence Intervals in Time Series Forecasting," *Journal of Forecasting*, 8, 141–155.

——— (1992), "Time Series Characteristics and the Widths of Judgmental Confidence Intervals," *International Journal of Forecasting*, 7, 413–420.

Peters, S. C., and Freedman, D. A. (1985), "Using the Bootstrap to Evaluate Forecasting Equations," *Journal of Forecasting*, 4, 251–262.

Pflaumer, P. (1988), "Confidence Intervals for Population Projections Based on Monte Carlo Methods," *International Journal of Forecasting*, 4, 135–142.

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

Ravishankar, N., Hochberg, Y., and Melnick, E. L. (1987), "Approximate Simultaneous Prediction Intervals for Multiple Forecasts," *Technometrics*, 29, 371–376.

Ravishankar, N., Wu, S.-Y., and Glaz, J. (1991), "Multiple Prediction Intervals for Time Series: Comparison of Simultaneous and Marginal Intervals," *Journal of Forecasting*, 10, 445–463.

This content downloaded from
192.12.247.244 on Mon, 25 Mar 2024 08:44:27 +00:00
All use subject to https://about.jstor.org/terms

Reinsel, G. (1980), "Asymptotic Properties of Prediction Errors for the Multivariate Autoregressive Model Using Estimated Parameters," *Journal of the Royal Statistical Society*, Ser. B, 42, 328–333.

Schmidt, P. (1977), "Some Small Sample Evidence on the Distribution of Dynamic Simulation Forecasts," *Econometrica*, 45, 997–1005.

Stine, R. A. (1987), "Estimating Properties of Autoregressive Forecasts," *Journal of the American Statistical Association*, 82, 1072–1078.

Stoica, P., and Nehorai, A. (1989), "On Multistep Prediction Error Methods for Time Series Models," *Journal of Forecasting*, 8, 357–368.

Thombs, L. A., and Schucany, W. R. (1990), "Bootstrap Prediction Intervals for Autoregression," *Journal of the American Statistical Association*, 85, 486–492.

Thompson, P. A., and Miller, R. B. (1986), "Sampling the Future: A Bayesian Approach to Forecasting From Univariate Time Series Models," *Journal of Business & Economic Statistics*, 4, 427–436.

Tong, H. (1990), *Non-Linear Time Series*, Oxford, U.K.: Clarendon Press.

Veall, M. R. (1989), "Applications of Computationally-Intensive Methods to Econometrics," in *Bulletin of the I.S.I., 47th Session*, Amsterdam, I.S.I., pp. 75–88.

Wallis, K. F. (1974), Discussion of "Experience With Forecasting Univariate Time-Series and the Combination of Forecasts," by P. Newbold and C. W. J. Granger, *Journal of the Royal Statistical Society*, Ser. A, 137, 155–156.

Wei, W. W. S. (1990), *Time Series Analysis*, Redwood City, CA: Addison-Wesley.

Weisberg, S. (1985), *Applied Linear Regression*, New York: John Wiley.

West, M., and Harrison, J. (1989), *Bayesian Forecasting and Dynamic Models*, Berlin: Springer-Verlag.

Williams, W. H., and Goodman, M. L. (1971), "A Simple Method for the Construction of Empirical Confidence Limits for Economic Forecasts," *Journal of the American Statistical Association*, 66, 752–754.

Winkler, R. L. (1972), "A Decision-Theoretic Approach to Interval Estimation," *Journal of the American Statistical Association*, 67, 187–191.

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

——— (1981), "Predictions of Multivariate Autoregressive Moving Average Models," *Biometrika*, 68, 485–492.

Yar, M., and Chatfield, C. (1990), "Prediction Intervals for the Holt–Winters Forecasting Procedure," *International Journal of Forecasting*, 6, 1–11.

This content downloaded from
192.12.247.244 on Mon, 25 Mar 2024 08:44:27 +00:00
All use subject to https://about.jstor.org/terms