Prediction Intervals for ARIMA Models

Author(s): Ralph D. Snyder, J. Keith Ord and Anne B. Koehler

Source: *Journal of Business & Economic Statistics*, Apr., 2001, Vol. 19, No. 2 (Apr., 2001), pp. 217-225

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

Stable URL: https://www.jstor.org/stable/1392165

REFERENCES
Linked references are available on JSTOR for this article:
https://www.jstor.org/stable/1392165?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
152.32.254.10 on Thu, 23 Dec 2021 02:23:34 UTC
All use subject to https://about.jstor.org/terms

Prediction Intervals for ARIMA Models

Ralph D. Snyder
Department of Econometrics and Business Statistics, Monash University, Clayton, Victoria 3168, Australia
(*ralph.snyder@buseco.monash.edu.au*)

J. Keith Ord
McDonough School of Business, Georgetown University, Washington, DC 20057
(*ordk@gunet.georgetown.edu*)

Anne B. Koehler
Department of Decision Sciences and Management Information Systems, Miami University, Oxford, OH 45056 (*koehleab@muohio.edu*)

The problem of constructing prediction intervals for linear time series (ARIMA) models is examined. The aim is to find prediction intervals that incorporate an allowance for sampling error associated with parameter estimates. The effect of constraints on parameters arising from stationarity and invertibility conditions is also incorporated. Two new methods, based on varying degrees of first-order Taylor approximations, are proposed. These are compared in a simulation study to two existing methods, a heuristic approach and the "plug-in" method whereby parameter values are set equal to their maximum likelihood estimates. A comparison of the four methods is also made for quarterly retail sales for 10 Organization for Economic Cooperation and Development countries. The new approaches provide a systematic improvement over existing methods.

KEY WORDS: ARIMA; Bayesian; Forecasting; Holt–Winters; Simulation; State space.

As indicated by Chatfield (1993) in his comprehensive state-of-the-art review, the construction of valid prediction intervals (PI's) for time series continues to present considerable difficulties. In particular, Chatfield noted a number of reasons why PI's may be too narrow; these include the following:

1. Model parameters may have to be estimated.
2. Innovations may not be normally distributed.
3. There may be outliers in the data.
4. The wrong model may be identified.
5. The underlying model may change, either during the period of fit or in the future.

In this article, we focus on the first of these issues. If the uncertainty relating to parameter estimation is not allowed for explicitly, the resulting PI's would be too narrow. Further, the nonlinear nature of the parameter estimates in time series makes the problem intractable as regards an exact analytic solution, so we develop various approximate solutions, which are then explored in a simulation study. Only when we are confident of our ability to produce reliable PI's in the basic case can we address the remaining issues. Thus, in this article, we examine the construction of PI's when the parameters are unknown and the errors are assumed to be normal, leaving the other issues to be addressed in further research.

We identify four approaches to the construction of PI's and report on an extensive simulation study of these alternatives. The particular model used in our simulations is the additive Holt–Winters (HW) scheme; see Example 1.2. Yar and Chatfield (1990) provided PI's for this scheme based on its autoregressive integrated moving average (ARIMA) representation and setting the parameter values equal to their estimates (the "plug-in" approach). These authors found the method to be superior to previous, albeit heuristic, approaches, and the plug-in PI is one of the options considered in our study. However, rather than use an ARIMA framework, we have opted for a state-space scheme; details are given in Section 1.

We are not the first to consider the construction of PI's for autoregressive moving average (ARMA) models by the use of a state-space scheme; see Ansley and Kohn (1986). They proposed a linear approximation method to account for the estimation error of the autoregressive (AR) and moving average (MA) parameters in computing PI's. However, we consider ARIMA models and the equivalent state-space models with a single source of error instead of ARMA models and state-space models with multiple sources of error. The linear approximation that is part of our two proposed methods [linear approximation (LA) method and Bayesian simulation (BS) method] is applied to the time series itself rather than to the variance of the state variables. In the second of our proposed methods, we consider the impact of estimating the error variance as well as the AR and MA parameters. Moreover, we conduct a simulation study to investigate the improvement in PI coverage by using methods that account for the error in the estimation of the parameters.

The second of our two proposed methods in the article is a BS scheme. When Ansley and Kohn (1986) showed how to obtain the conditional mean squared error (MSE) for a time series in the state-space framework, they pointed out that the correction to the MSE has a Bayesian interpretation. Under appropriate conditions we can apply the asymptotic sampling distribution developed by Ansley and Kohn (1986) to generate the predictive distribution, using simulation. De Jong and Whiteman (1994) followed this approach in developing PI's for AR(p) schemes; the resulting simulated distribution

© 2001 American Statistical Association
Journal of Business & Economic Statistics
April 2001, Vol. 19, No. 2

This content downloaded from
152.32.254.10 on Thu, 23 Dec 2021 02:23:34 UTC
All use subject to https://about.jstor.org/terms

is shown to converge to the predictive distribution using a result of Geweke (1989) and the same justification may be employed here.

An alternative approach would be to use a complete Markov-chain Monte Carlo (MCMC) approach; see Barnett, Kohn, and Sheather (1996, 1997) for the development of MCMC estimation procedures for ARMA models. Our method uses an analytic approximation to the posterior distribution of the parameters, which we then feed into the computation of the predictive distribution. Thus, our scheme may be viewed as a "partial" MCMC technique, which should be less demanding computationally, an important consideration when a large number of series is to be analyzed.

Another approach would be to consider the nonparametric bootstrap; see, for example, Thombs and Schucany (1990), Kabaila (1993), and McCullough (1994). However, since our current focus is on getting the correct coverage with a known underlying error process, we have not pursued that line of inquiry at this time.

In the article, we address three main issues—(1) the extension of the Bayesian simulation approach to state-space schemes, (2) the use of approximations to simplify the computational task, and (3) an extensive simulation study to determine whether two suggested approaches to computing PI's (linear approximation method and Bayesian simulation method) provide PI's with improved coverage.

The structure of the remainder of this article is as follows. In Section 1, we compare single-source and multiple-source state-space schemes and justify our use of a single-source model. In Section 2, we describe the various approaches to be considered for the construction of PI's. Section 3 describes the simulation study and summarizes the conclusions from that study. An application to real data is presented in Section 4. The summary and outline of future directions appears in Section 5.

1. STATE–SPACE REPRESENTATION

For the PI's in this article, we use models in their state-space form. In particular, we employ state-space models with a single source of error (SSOE); see Snyder (1985, 1988). The SSOE representation for a time series $\{y_t\}$ is

$$y_t = \mathbf{h}'\mathbf{x}_{t-1} + \varepsilon_t \quad \text{where} \quad \mathbf{x}_t = \mathbf{F}\mathbf{x}_{t-1} + \alpha\varepsilon_t \quad \text{and} \quad \varepsilon_t \sim \text{IID}(0, \sigma^2). \quad (1.1)$$

The k vector $\mathbf{x_{t-1}}$ represents the state of the underlying process at the beginning of period t, ε_t is from an IID$(0, \sigma^2)$ series of disturbances, $\mathbf{h}$ is a fixed k vector, $\mathbf{F}$ is a fixed k by k matrix, and α is a fixed k vector of parameters. It is assumed that ε_t is independent of $\{y_{t-i}, \mathbf{x}_{t-i}, i \geq 1\}$.

Akaike (1974) first showed that any ARMA(p, q) scheme has a Markovian state-space representation with a single source of variation, a small but critical difference. Moreover, any ARIMA$(p, d, q)(P, D, Q)_s$ model can be represented by an SSOE model, and vice versa. However, it should be noted that Akaike's model does not include an error term in the observation equation. In Appendix A, we demonstrate the equivalence of the ARIMA and SSOE schemes.

Example 1.1: ARMA(1,1) Scheme. The state-space scheme $y_t = \mu_{t-1} + \varepsilon_t$, where $\mu_t = \phi\mu_{t-1} + \alpha\varepsilon_t$ corresponds to the ARMA(1,1) scheme $y_t = \phi y_{t-1} + \varepsilon_t - (\phi - \alpha)\varepsilon_{t-1}$ with invertibility condition $|\phi - \alpha| < 1$ and stationarity condition $|\phi| < 1$.

The primary focus in Section 3 is on the following special case of the SSOE, which underlies the HW method of forecasting (Winters 1960).

Example 1.2: Additive Holt–Winters Model. The observation equation $y_t = \ell_{t-1} + b_{t-1} + c_{t-m} + \varepsilon_t$ is accompanied by state equations for the level $\ell_t = \ell_{t-1} + b_{t-1} + \alpha_1\varepsilon_t$, the growth rate $b_t = b_{t-1} + \alpha_2\varepsilon_t$, and the seasonal factors $c_t = c_{t-m} + \alpha_3\varepsilon_t$, where $m =$ number of seasons. By eliminating the state variables, it can be shown that this example reduces to the seasonal ARIMA scheme found by McKenzie (1976) for the HW method. It may then be established that the conditions $0 < \alpha_1 < 2$, $0 < 2\alpha_1 + \alpha_2 < 4$, and $0 < \alpha_3 < 1$ are necessary for invertibility. Necessary and sufficient conditions were given by Archibald (1990).

Multiple source of error (MSOE) models are also available; see Harvey (1990, chap. 2) or West and Harrison (1997, chap. 2). Although ARIMA schemes may be represented by an MSOE scheme, the parameter space is often restricted. For example, the MA(1) scheme with parameter θ has $-1 < \theta < 1$, but the MSOE representation requires $0 < \theta < 1$.

An empirical study by Garcia-Ferrer and Del Hoyo (1992) contrasted the multiple source scheme (Harvey's basic structural model or BSM; Harvey 1990) with ARIMA modeling for a number of series. Garcia-Ferrer and del Hoyo concluded that the ARIMA formulation generally produces better predictions than BSM, a result they attributed to the lack of orthogonality among the components of the state vector, but it may be due to restrictions on the parameter space. Given the equivalence of the ARIMA and SSOE representations, their conclusions imply that the SSOE form of the BSM is superior to its traditional multiple-source counterpart, provided a suitable model-selection procedure is applied.

2. MODEL ESTIMATION AND PREDICTION ERROR

2.1 Exponential Smoothing Versus the Kalman Filter

ML estimates of the parameters α and σ may be obtained using a procedure that incorporates the Kalman filter to expedite the evaluation of the likelihood function (Schweppe 1965). The Kalman filter for the SSOE scheme (Snyder 1985) includes the equations

$$e_t = y_t - \mathbf{h}'\mathbf{x}_{t-1|t-1} \quad (2.1)$$

and

$$\mathbf{x}_{t|t} = \mathbf{F}\mathbf{x}_{t-1|t-1} + \mathbf{a}_t(y_t - \mathbf{h}'\mathbf{x}_{t-1|t-1}), \quad (2.2)$$

where $\mathbf{x}_{t|n}$ denotes the estimator for $\mathbf{x}_t$ based on the sample $\mathrm{Y}_n = (y_1, \ldots, y_n)$, and where $\mathbf{a}_t$ is the Kalman gain.

An inherently simpler strategy is to bypass the Kalman filter and use exponential smoothing from the outset. Conditioning on a trial value for $\mathbf{x}_0$ and assuming that the sample $Y_n = (y_1, \ldots, y_n)$ is known, the SSOE scheme implies that fixed successive values of the state vector $\mathbf{x}_t$ can be computed

This content downloaded from 152.32.254.10 on Thu, 23 Dec 2021 02:23:34 UTC
All use subject to https://about.jstor.org/terms

recursively with the error-correction form of the exponential smoothing equation

$$\mathbf{x}_t = \mathbf{F}\mathbf{x}_{t-1} + \alpha(y_t - \mathbf{h}'\mathbf{x}_{t-1}). \tag{2.3}$$

Strictly speaking, the $\mathbf{x}_t$ should be read as $\mathbf{x}_t|\mathbf{Y}_{t-1}, \theta, \mathbf{x}_0$ in (2.3), where θ denotes the vector of unknown parameters contained in $(\mathbf{h}, \mathbf{F}, \alpha, \sigma)$. The SSOE scheme then implies that $y_t|\mathbf{Y}_{t-1}, \theta, \mathbf{x}_0 \sim \text{IID}(\mathbf{h}'\mathbf{x}_{t-1}, \sigma^2)$, from which it follows that the likelihood function has the form $L(\theta, \mathbf{x}_0|\mathbf{Y}_n) = \prod_{t-1}^{n} p(y_t|\mathbf{Y}_{t-1}, \theta, \mathbf{x}_0)$, where $p()$ is the pdf for ε_t. Since $\mathbf{x}_0$ is treated as a fixed vector of unknown parameters, we have here a conditional, rather than a marginal, likelihood function.

When the likelihood is based on a normal distribution, the ML estimate of $\mathbf{x}_0$, denoted by $\mathbf{x}_{0|n}$ for given θ, is a linear least squares estimate. Estimates of successive state vectors, given a sample of size n, are then obtained with the recurrence relationship based on (2.3)

$$\mathbf{x}_{t|n} = \mathbf{F}\mathbf{x}_{t-1|n} + \alpha(y_t - \mathbf{h}'\mathbf{x}_{t-1|n}). \tag{2.4}$$

This is similar to the updating Equation (2.2) of the Kalman filter, there being two differences, (1) the Kalman gain is replaced by the vector of smoothing parameters and (2) filtered values of the state vectors are replaced by corresponding smoothed values.

The Kalman filter can only be used as part of the ML procedure when the disturbances in the SSOE scheme are normally distributed. The exponential smoothing method outlined in this section, however, can be applied for any disturbance distribution. The seed vector estimates no longer correspond to linear least squares estimates and the links with the Kalman filter disappear. This is of little consequence, however, because the method continues to yield ML estimates.

For the rest of the article the ML estimates of θ and $\mathbf{x}_0$ will be denoted by $\hat{\theta}$ and $\hat{\mathbf{x}}_0$ (replaces $\mathbf{x}_{0|n}$). Likewise $\hat{\mathbf{x}}_t$ will replace $\mathbf{x}_{t|n}$ as the estimate for $\mathbf{x}_t$, obtained by applying (2.3). At some points of the article, σ will not be part of θ and its ML estimate will be denoted by $\hat{\sigma}$.

2.2 Point Predictions and Prediction Intervals

Once the ML estimators have been found, we construct both point predictions and PI's for up to j periods ahead; $j = 1, 2, \ldots, r$. The point predictions for $t = n+1, \ldots, n+r$ are the conditional expectations of Model (1.1): $\hat{\mathbf{y}}_t = \mathbf{h}'\hat{\mathbf{x}}_{t-1}$ and $\hat{\mathbf{x}}_t = \mathbf{F}\hat{\mathbf{x}}_{t-1}$. Four principal approaches for the PI's will be considered—a heuristic (HE) method, the "plug-in" (PL) method, a linear approximation (LA), and a Bayesian simulation (BS) scheme. To simplify the notation, we will use $\mathbf{y}_p = (y_1, \ldots, y_n)$ for the sample of past values and $\mathbf{y}_f = (y_{n+1}, \ldots, y_{n+r})$ for future values.

2.2.1 Heuristic Method (HE). This method was developed by Bowerman and O'Connell (1993, chap. 8). The computation of the PI does not rely on a model that includes the changing nature of the level, trend, and seasonal components. Instead, it relies on the adjustment of the MAD (mean absolute deviation), which is computed from deseasonalized data. The adjustment is based on work by R. G. Brown and takes into account the smoothing constants and length of the forecasting horizon.

2.2.2 Plug-in Method (PL). For the construction of the PI by the PL method, the density function $p(\mathbf{y}_f|\theta, \mathbf{x}_0, \sigma, \mathbf{y}_p)$ of the future time series is approximated by the Gaussian density $\phi(\mathbf{y}_f|\hat{\theta}, \hat{\mathbf{x}}_0, \hat{\sigma}, \mathbf{y}_p)$. The subscripts (p) and (f) refer to past and future values of y_t, respectively. The predictions are determined in the usual way from this distribution. This method is equivalent to the usual procedure in which we assume that the psi weights, ψ_j, which may be determined from the parameters, are known. We anticipate that this method will yield intervals that understate the true width.

2.2.3 Linear Approximation Method (LA). The LA method accounts for the sampling error that is associated with θ, where σ is not included in θ. We expect that this method will produce intervals with better coverage than the PI's from HE and PL. Let e_t denote the tth residual estimated from a sample of size n obtained during the calculations with (2.3) and let the subscripts (p) and (f) refer to past and future values, respectively. Then we may write $\mathbf{e}' = [\mathbf{e}_p', 0]$, where $\mathbf{e}_p' = [e_1, \ldots, e_n]$ and $\varepsilon' = [\varepsilon_p', \varepsilon_f']$ is the corresponding $(n+r) \times 1$ vector of error terms; the vector of zeros corresponds to the predicted values for ε_f. For an invertible process the dependence of $\mathbf{y}_f$ on $\mathbf{x}_0$ will be slight [see the appendix, Eq. (A.8)]. Thus, we assume that $\mathbf{y}$ is approximately a linear function of θ and ε only and write $\mathbf{y}^* = \mathbf{Z}\theta + \mathbf{L}\varepsilon$, where $\mathbf{y}^*$ denotes $\mathbf{y}$ minus the constant term from the Taylor series expansion. The matrices $\mathbf{Z}$ and $\mathbf{L}$ contain the derivatives of $\mathbf{y}$ with respect to θ and ε, evaluated at $[\hat{\theta}, \mathbf{e}]$. Note that $\mathbf{L}$ is a unit lower triangular matrix because the typical y_t cannot depend on future values of the disturbances.

This linear approximation can be expressed as the following equations for the past (p) and future (f) values:

$$\mathbf{y}_p^* = \mathbf{Z}_p\theta + \mathbf{L}_{pp}\varepsilon_p \tag{2.5}$$

and

$$\mathbf{y}_f^* = \mathbf{Z}_f\theta + \mathbf{L}_{pf}\varepsilon_p + \mathbf{L}_{ff}\varepsilon_f. \tag{2.6}$$

Assuming a diffuse prior, we approximate the posterior by a Gaussian distribution with mean $\hat{\theta}$ and variance

$$\text{var}(\theta|\sigma, \mathbf{x}_0, \mathbf{y}_p) = \sigma^2(\overline{\mathbf{Z}}_p'\overline{\mathbf{Z}}_p)^{-1}, \tag{2.7}$$

where $\mathbf{L}_{pp}\overline{\mathbf{Z}}_p = \mathbf{Z}_p$.

To construct PI's we need the variance of the forecast error for future time periods. In the development of this variance, first solve for ε_p in (2.5) to get

$$\varepsilon_p = \mathbf{L}_{pp}^{-1}(\mathbf{y}_p^* - \mathbf{Z}_p\theta). \tag{2.8}$$

Then substitute (2.8) into (2.6) to find $\mathbf{y}_f^* = (\mathbf{Z}_f - \mathbf{L}_{pf}\overline{\mathbf{Z}}_p)\theta + \mathbf{L}_{ff}\varepsilon_f + \mathbf{L}_{pf}\mathbf{L}_{pp}^{-1}\mathbf{y}_p^*$. This equation can be rewritten in the following form:

$$\mathbf{y}_f^* = \overline{\mathbf{Z}}_f\theta + \mathbf{L}_{ff}\varepsilon_f + \text{const}, \tag{2.9}$$

where $\overline{\mathbf{Z}}_f = \mathbf{Z}_f - \mathbf{L}_{pf}\overline{\mathbf{Z}}_p$. As a result, the distribution $p(\mathbf{y}_f|\mathbf{x}_0, \sigma, \mathbf{y}_p)$ is approximated by $p(\mathbf{y}_f|\hat{\mathbf{x}}_0, \hat{\sigma}, \mathbf{y}_p)$, which, in turn, is

This content downloaded from
152.32.254.10 on Thu, 23 Dec 2021 02:23:34 UTC
All use subject to https://about.jstor.org/terms

approximated by multivariate Gaussian distribution with mean $(\mathbf{h}'\hat{\mathbf{x}}_{n+1}, \ldots, \mathbf{h}'\hat{\mathbf{x}}_{n+r})$ and variance matrix $\sigma^2(\overline{\mathbf{Z}}_f(\overline{\mathbf{Z}}_p'\overline{\mathbf{Z}}_p)^{-1}\overline{\mathbf{Z}}_f' + \mathbf{L}_{ff}\mathbf{L}_{ff}')$. The PI's, for a specified coverage probability, are determined from a standard Gaussian distribution.

In our context, some of the elements of θ must be nonnegative. Thus, whenever $\hat{\theta}_i < 0$, we replace it by 0 for the estimation of $\text{var}(\mathbf{y}_f|\theta, \mathbf{x}_0, \sigma, \mathbf{y}_p)$. Since we have yet to include the sampling error for $\hat{\sigma}$ in the variance matrix, it may be possible to improve the coverage of the PI's further. We investigate the addition of this sampling error in the next method.

2.2.4 Bayesian Simulation Method (BS). We may specify the sampling distributions for $\hat{\theta}$, $\hat{\mathbf{x}}_0$, and $\hat{\sigma}$ in terms of the joint pdf: $p(\hat{\theta}, \mathbf{x}_0, \hat{\sigma}|\theta, \mathbf{x}_0, \sigma)$, where $p(.)$ denotes a generic pdf. We may then develop the predictive distribution, in the Bayesian framework of Aitchison and Dunsmore (1975), as

$$p(\mathbf{y}_f|\mathbf{y}_p) = \iiint p(\mathbf{y}_f|\mathbf{y}_p, \theta, \mathbf{x}_0, \sigma) \times p(\theta, x_0, \sigma|\mathbf{y}_p)(d\theta d\mathbf{x}_0 d\sigma), \quad (2.10)$$

where the differential element covers all the items in θ and in $\mathbf{x}_0$ so that the triple integral represents $k+a+1$ dimensions in all, a being the dimensionality of θ and k being the number of states. The posterior density for $(\theta, \mathbf{x}_0, \sigma)$ is determined from the sampling distribution and a suitable prior in the usual way as

$$p(\theta, \mathbf{x}_0, \sigma|\mathbf{y}_p) = p(\theta|\sigma, \mathbf{x}_0, \mathbf{y}_p)p(\sigma|\mathbf{x}_0, \mathbf{y}_p)p(\mathbf{x}_0|\mathbf{y}_p). \quad (2.11)$$

In our numerical work, we found that $\hat{\mathbf{x}}_t$ tended to $\mathbf{x}_t$ quite quickly whenever the estimates of the smoothing parameters were nonzero. This observation has two main consequences: First of all, the results were largely unaffected by the variations in the seed state vector $\hat{\mathbf{x}}_0$; second, this relative insensitivity led to some numerical instabilities, particularly when the seasonal cycle of length $m = 12$. For both reasons, we decided to focus attention on the variations in $\hat{\theta}$ and $\hat{\sigma}$ only. Thus, we now reformulate (2.10) for the current problem as

$$p(\mathbf{y}_f|\mathbf{y}_p) = \iint p(\mathbf{y}_f|\mathbf{y}_p, \theta, \hat{\mathbf{x}}_0, \sigma) \times p(\theta, \sigma|\mathbf{y}_p, \hat{\mathbf{x}}_0)(d\theta d\sigma); \quad (2.12)$$

that is, we perform a simulation to arrive at the predictive distribution, which has the form

$$p(\mathbf{y}_f|\mathbf{y}_p) = \iint p(\mathbf{y}_f|\mathbf{y}_p, \theta, \hat{x}_0, \sigma)p(\theta|y_p, \hat{\mathbf{x}}_0, \sigma) \times p(\sigma|\mathbf{y}_p, \hat{\mathbf{x}}_0)(d\theta d\sigma). \quad (2.13)$$

A Monte Carlo integration method is employed to evaluate $p(\mathbf{y}_f|\mathbf{y}_p)$ as follows; the steps are entirely the same as those described by Ord, Koehler, and Snyder (1997):

1. $p(\sigma|\hat{\mathbf{x}}_0, \mathbf{y}_p)$ is approximated by an inverted gamma distribution. A value of σ^2 is randomly generated from the approximating distribution.
2. $p(\theta|\sigma, \hat{\mathbf{x}}_0, \mathbf{y}_p)$ is approximated by a Gaussian distribution with mean $\hat{\theta}$ and variance matrix (2.7). A value of θ is randomly generated from the approximating distribution. Those elements of θ that violate the invertibility conditions are adjusted. For example, for the additive structural model in Example 1.2, we saw that the smoothing parameters α must be nonnegative. Negative values are truncated to 0. (We ignore the other restrictions on the smoothing parameters for the HW method because we rarely found them to be binding in practice.)
3. The distribution $p(\mathbf{y}_f|\theta, \hat{\mathbf{x}}_0, \sigma, \mathbf{y}_p)$ is approximated by a synthetic sample of M values of the vector $\mathbf{y}_f$ generated from the model in Example 1.2. The future values of e_t required for the calculation of each instance of $\mathbf{y}_f$ are themselves generated from an $N(0, \sigma^2)$ distribution. Thus, for each $\mathbf{y}_f$ we estimate the pdf by

$$p(\mathbf{y}_f|\mathbf{y}_p) = M^{-1}\sum_i p(\mathbf{y}_f|\mathbf{y}_p, \theta_i, \hat{\mathbf{x}}_0, \sigma_i), \quad (2.14)$$

where the M sets of values of θ_i and σ_i are generated in accordance with the preceding steps 1–3.
4. PI's are constructed directly from the sample of the $\mathbf{y}_f$ for a specified confidence interval P by deleting those $M(1-P)$ sample values that are farthest from the associated point prediction for each period t. The smallest and largest values that remain in the culled sample are used as the lower and upper boundaries of the prediction intervals.

The method proposed is similar to that of Thompson and Miller (1986) for stationary AR(p) processes. Thompson (personal communication to Koehler) indicated that they did not extend their method to ARIMA schemes. Our method has two key differences: (1) We impose constraints on the parameter estimates to ensure stationarity after differencing and (2) we use ML estimators. We have not compared the performance of our procedures with those of Thompson and Miller, since our primary interest lay in the construction of PI's for complex models like HW with a strong MA component. Since least squares methods are highly efficient for AR schemes except near the boundaries of the parameter space, it is only for AR processes close to nonstationarity that we would expect to see much difference in performance between their scheme and ours.

3. THE SIMULATION STUDY

A simulation study was undertaken to compare the preceding PI methods. We constrained the scope of the study to the additive HW method of forecasting. Our choice was motivated by the fact that the additive HW method is widely used in practice, that traditionally users of this method have relied on heuristics rather than sound statistical methods to compute associated PI's, and that the structural model underpinning it is a nontrivial example from the ARIMA class (as shown in Example 1.2). Any simulation study is necessarily limited by the range of model options selected, but we believe that the results of our study are reasonably representative of more complex models and likely to provide greater insights than special cases such as AR(1) or MA(1). The case of the multiplicative HW scheme was considered by Ord et al. (1997).

This content downloaded from 152.32.254.10 on Thu, 23 Dec 2021 02:23:34 UTC
All use subject to https://about.jstor.org/terms

3.1 Design

Each original series was generated using the additive HW scheme described in Example 1.2. The initial conditions were $l_0 = 100$, $b_0 = 2$, $c_{j-m} = A\sin(2\pi j/m)$, $j = 1, 2, \ldots, m$, where A denotes the seasonal amplitude. Clearly any distribution may be used in the simulations, although our emphasis has been on the normal distribution. We considered a mixture of normals $(1-q)\mathrm{N}(0,1) + q\mathrm{N}(0,4)$ for ε_t/σ, q representing the proportion of outliers. PI's were generated for future time periods $n+1$ to $n+2m$ at a probability level of .90. We selected the following factors that determine a time series for a base scenario, referred to as case (a): sample size $(n) = 36$, number of seasons per year $(m) = 4$, smoothing parameters $(\alpha) = (.2, .05, .05)'$, seasonal amplitude $(A) = 30$, standard deviation for $\varepsilon_t(\sigma) = 5$, and proportion of outliers $(q) = 0$. Four additional scenarios were selected by modifying the factors as follows: (b) Increase n to 72, (c) set $\sigma = 10$, (d) set $q = .2$, and (e) set $n = 72$ and $m = 12$. Shifts in A and in α did not appear to produce any important changes in the results. PI's were also computed at a probability level of .95 for each scenario, but because the results were qualitatively the same, they are not reported in this article.

Each scenario was replicated 1,000 times. For each replication, we executed the following sequence of steps:

1. Generate a time series $y_1, \ldots, y_n, y_{n+1}, \ldots, y_{n+2m}$ using the specified factors.
2. Generate estimates of $\mathbf{x}_0$, denoted by $\tilde{\mathbf{x}}_0$, using Winters's approach (see Winters 1960; Bowerman and O'Connell 1993, chap. 8), based on the first three years of data.
3. Determine the ML estimators for α conditional on $\tilde{\mathbf{x}}_0$, based on $y_1, \ldots, y_n$.
4. Generate the forecasts for eight leadtime horizons with quarterly data and 24 leadtime horizons with monthly data.
5. Generate the 90% PI's for each of the four approaches described earlier—heuristic, plug-in, Bayesian simulation, and linear approximation.
6. Record the width of each interval constructed in 5 and whether or not the interval covers the value generated in 1 at each leadtime and for each method.

3.2 Results

The results for the five scenarios are reported in Tables 1 and 2. Table 1 gives the percentage (i.e., mean) coverage, across the 1,000 replications, for each method at each leadtime. Ninety-five percent confidence intervals for each proportion are in the range $\pm.02$ to $\pm.03$. Table 2, configured in the same way, gives the mean and standard deviation for the length of the PI's.

3.2.1 Coverage. We found that the patterns that occur in the display for the base scenario [Table 1(a)] are repeated in all the other displays [Table 1(b)–(e)]. The coverage decreases as the length of the horizon increases. The coverage by the BS method is the highest for all horizons and is followed closely by the LA method. Both of these methods provide coverage that is higher that the usual PL method, which does not account for the error in estimating the parameters α and σ. The amount of improvement in coverage that is provided

Table 1. Percentage Coverage by 90% Prediction Intervals (based on 1,000 simulations for each case)

	Method			
Leadtime	HE	PL	LA	BS
(a) Factors for base scenario: N = 36, m = 4, A = 30, q = 0, alpha 1 = .2, sigma = 5				
	.85	.83	.84	.86
	.84	.81	.87	.86
	.81	.82	.83	.86
	.81	.82	.83	.87
	.75	.78	.81	.87
	.73	.77	.81	.85
	.72	.78	.79	.84
	.72	.75	.79	.82
(b) n increased to 72 from base scenario				
	.87	.88	.88	.89
	.85	.85	.86	.89
	.85	.86	.86	.88
	.83	.87	.86	.89
	.82	.84	.87	.87
	.80	.86	.84	.87
	.77	.81	.86	.87
	.75	.80	.83	.86
(c) Sigma increased to 10 from base scenario				
	.83	.82	.86	.87
	.82	.81	.86	.86
	.81	.80	.84	.88
	.78	.81	.84	.85
	.75	.78	.83	.86
	.72	.77	.83	.83
	.70	.76	.82	.84
	.71	.76	.81	.82
(d) q increased to .2 from base scenario				
1	.82	.84	.87	.88
2	.84	.83	.86	.86
3	.83	.81	.85	.86
4	.80	.81	.85	.85
5	.76	.79	.84	.85
6	.73	.77	.81	.82
7	.70	.77	.82	.82
8	.69	.75	.83	.82
(e) n increased to 72 and m increased to 12 from base scenario				
1	.83	.83	.87	.83
2	.81	.82	.85	.88
3	.81	.82	.83	.86
4	.80	.81	.82	.88
5	.82	.80	.83	.87
6	.79	.81	.83	.83
7	.75	.79	.81	.84
8	.74	.78	.80	.82
9	.72	.78	.81	.81
10	.71	.76	.80	.81
11	.69	.76	.80	.83
12	.68	.76	.78	.81
13	.62	.74	.77	.79
14	.62	.71	.76	.79
15	.61	.72	.75	.79
16	.60	.72	.76	.77
17	.57	.73	.75	.78
18	.55	.71	.75	.78
19	.54	.71	.75	.77
20	.53	.72	.75	.77
21	.51	.70	.74	.77
22	.48	.69	.74	.76
23	.50	.68	.73	.76
24	.48	.70	.73	.75

This content downloaded from 152.32.254.10 on Thu, 23 Dec 2021 02:23:34 UTC
All use subject to https://about.jstor.org/terms

Table 2. Length of Interval (mean and standard deviation) for 90% Prediction (based on 1,000 simulations for each case)

	Mean for method				Standard deviation for method			
Leadtime	HE	PL	LA	BS	HE	PL	LA	BS
	(a) Factors for base scenario: $n = 36$, $m = 4$, $A = 30$, $q = 0$, alpha1 $= .2$, sigma $= 5$							
1	15.3	15.3	16.6	17.1	2.2	2.0	2.4	2.6
2	15.7	15.6	17.1	17.6	2.3	2.1	2.5	2.8
3	16.1	16.0	17.6	18.3	2.5	2.2	2.9	3.2
4	16.5	16.5	18.3	19.2	2.8	2.6	3.4	3.8
5	17.0	17.7	20.3	21.4	3.1	3.2	4.6	5.1
6	17.4	18.5	21.4	22.7	3.6	3.8	5.4	5.9
7	17.9	19.5	22.5	24.1	4.1	4.6	6.3	7.0
8	18.4	20.7	23.8	25.7	4.6	5.5	7.4	8.1
	(b) n increased to 72 from base scenario							
1	16.0	15.9	16.4	16.8	1.5	1.5	1.5	1.7
2	16.3	16.3	16.8	17.3	1.6	1.6	1.6	1.8
3	16.6	16.8	17.3	18.0	1.7	1.8	1.8	2.0
4	17.0	17.4	18.0	18.8	1.8	2.0	2.1	2.3
5	17.4	18.8	19.6	20.6	2.0	2.6	2.7	3.0
6	17.8	19.8	20.6	21.8	2.2	3.0	3.2	3.5
7	18.2	20.9	21.8	23.2	2.5	3.6	3.7	4.1
8	18.6	22.2	23.2	24.8	2.8	4.2	4.4	4.8
	(c) Sigma increased to 10 from base scenario							
1	30.4	30.5	33.1	34.1	4.4	4.2	4.9	5.0
2	31.2	31.1	33.9	35.2	4.6	4.3	5.3	5.3
3	32.0	32.0	35.0	36.6	5.0	4.7	6.1	6.0
4	32.9	33.0	36.3	38.2	5.5	5.3	7.0	7.1
5	33.8	35.4	40.3	42.8	6.3	6.8	9.2	9.5
6	34.8	37.0	42.3	45.2	7.1	8.0	10.9	11.0
7	35.8	38.9	44.7	48.0	8.0	9.5	12.7	13.1
8	36.8	41.2	47.2	51.1	9.0	11.4	14.9	15.5
	(d) q increased to .2 from base scenario							
1	18.8	19.1	20.8	21.3	3.1	3.4	3.6	4.0
2	19.3	19.4	21.3	22.0	3.3	3.4	3.9	4.3
3	19.9	20.0	22.1	22.8	3.6	3.7	4.4	4.9
4	20.4	20.6	23.0	23.9	4.1	4.1	5.1	5.6
5	21.0	22.2	25.5	26.7	4.7	5.1	6.6	7.1
6	21.7	23.2	26.9	28.3	5.4	5.8	7.7	8.2
7	22.3	24.5	28.5	30.0	6.1	6.8	8.9	9.4
8	23.0	25.9	30.1	31.9	6.9	7.9	10.5	10.8
	(e) n increased to 72 and m increased to 12 from base scenario							
1	15.6	15.7	16.6	16.8	1.6	1.6	1.8	1.8
2	16.0	16.2	17.2	17.5	1.8	1.8	2.0	2.1
3	16.6	16.9	17.9	18.4	2.1	2.1	2.4	2.5
4	17.1	17.7	18.8	19.4	2.4	2.4	2.7	3.1
5	17.7	18.5	19.8	20.6	2.9	2.9	3.3	3.7
6	18.3	19.5	21.0	22.0	3.3	3.4	3.9	4.4
7	18.9	20.7	22.3	23.5	3.9	4.0	4.6	5.3
8	19.6	22.0	23.7	25.1	4.4	4.8	5.5	6.2
9	20.2	23.4	25.3	26.9	5.0	5.5	6.3	7.2
10	20.9	24.9	26.9	28.7	5.6	6.4	7.3	8.3
11	21.6	26.5	28.7	30.7	6.1	7.4	8.3	9.4
12	22.3	28.2	30.5	32.8	6.7	8.5	9.4	10.6
13	23.1	30.2	33.1	35.6	7.4	9.6	10.7	12.1
14	23.8	32.0	35.1	37.9	8.0	10.7	12.0	13.3
15	24.5	34.0	37.3	40.3	8.6	12.0	13.3	14.8
16	25.3	36.1	39.4	42.6	9.2	13.3	14.6	16.2
17	26.0	38.1	41.7	45.1	9.8	14.7	16.0	17.7
18	26.8	40.3	44.1	47.7	10.5	16.1	17.5	19.3
19	27.6	42.5	46.5	50.4	11.1	17.6	19.0	21.0
20	28.3	44.7	48.9	53.0	11.7	19.0	20.6	22.6
21	29.1	47.1	51.4	55.8	12.4	20.7	22.1	24.3
22	29.9	49.4	53.9	58.6	13.0	22.2	23.8	26.1
23	30.7	51.8	56.5	61.5	13.6	23.8	25.5	27.8
24	31.5	54.3	59.2	64.4	14.3	25.6	27.2	29.7

This content downloaded from
152.32.254.10 on Thu, 23 Dec 2021 02:23:34 UTC
All use subject to https://about.jstor.org/terms

by the BS and LA methods increases as the horizon increases. The HE and the PL methods are comparable in coverage, especially at shorter horizons.

The only differences between the base scenario and the others appear to be functions of sample size and leadtime. An increase in the standard deviation [Table 1 (c)] or in the amplitude caused virtually no change relative to the base scenario. The lack of change in the display for the first eight horizons was also true if the sample size and number of seasons were both increased [Table 1 (e)]. Although the overall coverage in general is lower than the nominal level, it increases with an increase in sample size [Table 1 (b)], as expected. An increase in the proportion of outliers [Table 1 (d)] appears to improve the performance of the LA while leaving other methods unaffected.

3.2.2 Length of Intervals. The results in Table 2 are very much as expected. The improvements in coverage are obtained by lengthening the intervals. Interestingly, the intervals for LA and BS reduce as the sample size increases, yet those for PL and HE *increase*, suggesting that some serious biases exist for those methods when n is small to moderate. Increase in the proportion of outliers, and in σ, lengthens the intervals as expected.

Overall, we regard the performance of BS as the best of the four methods, especially at longer horizons. The LA method is almost as good as the BS method. Although the BS method is computationally quite intensive, it takes less than a minute on a Pentium computer for application to a single series and provides an estimate of the complete predictive distribution, not just the PI; Tsay (1993) and others have argued that such an approach is more desirable. In particular, the entire predictive distribution is required when we consider cost-based loss functions, rather than measures such as squared-error loss. Moreover BS is readily extendable to nonnormal errors, although its performance remains to be explored beyond the normal mixture considered here.

4. AN APPLICATION

We conducted another experiment using quarterly retail sales for 10 countries in the Organization for Economic Cooperation and Development (OECD). The data cover the period from the March quarter 1962 to the December quarter 1993 and are taken from the Main Economic Indicators database maintained by the OECD and published annually by its office in Paris. The 10 countries in the study are listed in Table 3. A log transform was applied to the data to make the seasonality additive and suitable for modeling with the additive HW model of Example 1.2.

The experiment was begun by applying the ML estimation procedures of exponential smoothing (using the additive HW model in Example 1.2) to the first 48 quarters of data, the origin being the first quarter of 1962. For each quarter in the two years 1974 and 1975 following the sample period, PI's were generated with the four methods. The actual time series values over the two-year "pseudo-future" were compared to the resulting PI's. This process was repeated on a rolling basis, with the origin of the sample shifted by two-year increments. The process was repeated for 10 such iterations. Thus 10 sets of PI's were generated for each of the 10 countries, a total of 100 sets of PI's.

Table 3. Percentage Coverage by 90% Prediction Intervals for OECD Quarterly Retail Sales Data (March 1962–December 1996)

		Method			
Leadtime	Country	HE	PL	LA	BS
		(a)			
1	All	.82	.86	.91	.88
2	All	.90	.90	.90	.90
3	All	.82	.91	.91	.92
4	All	.81	.85	.94	.90
5	All	.85	1.00	1.00	1.00
6	All	1.00	1.00	1.00	1.00
7	All	.95	.99	1.00	1.00
8	All	1.00	1.00	1.00	1.00
		(b)			
1–4	Australia	.85	.88	.95	.93
1–4	Austria	.85	.90	.95	.93
1–4	Canada	.83	.90	.93	.90
1–4	Denmark	.83	.88	.88	.90
1–4	Finland	.83	.88	.93	.93
1–4	Japan	.83	.85	.93	.88
1–4	Netherlands	.83	.88	.90	.88
1–4	Norway	.85	.90	.90	.90
1–4	United Kingdom	.83	.88	.90	.90
1–4	United States	.88	.88	.90	.88

The percentages for the coverage by the prediction limits in this experiment are shown in Table 3. Table 3 (a) compares the coverage by each of the four methods across the 10 countries and the 10 iterations for each of the eight leadtimes. Table 3 (b) compares the coverage by each method across the first four leadtimes and 10 iterations for each country. In both parts of the table, we see that LA and BS methods tend to provide more coverage than both the usual PL method and the HE method. The combination of linear trend and multiplicative seasonality make it difficult to model this data adequately with linear models and may account for the unusually high coverage at the longer lead times.

5. SUMMARY

We have proposed two methods (a linear approximation method and a Bayesian simulation method) that account for the estimation error of the parameters in the construction of PI's for linear time series processes. Using the model for the additive Holt–Winters method (Example 1.2), we conducted a simulation study to compare the coverage provided by these two methods with the usual plug-in method and a heuristic method. In addition, we performed an experiment on real data to do the same comparison. We have found significant gains in the accuracy of coverage by the LA and BS methods when the estimation error in the parameters is taken into account, at least in the case of normally distributed errors. The simulations showed that the BS method, which accounts for estimation error in both σ and θ, offers some modest improvements in coverage over the LA method, which only takes into account the estimation error in θ. The application to real data did not

This content downloaded from
152.32.254.10 on Thu, 23 Dec 2021 02:23:34 UTC
All use subject to https://about.jstor.org/terms

support this latter finding, but both methods continue to outperform the PL and HE approaches.

Improved PI's are of natural importance in their own right. In addition, the BS method provides the entire predictive distribution, which is important when considering loss functions. Examples include inventory management, where inaccurate assessment of the distribution could lead to increased inventory or to an excessive risk of stock-outs. Overall, we recommend that one of the two new methods be employed, unless a long series is available for estimation.

Chatfield (1993) noted that several problems remain in the construction of valid PI's, as discussed in the introduction. We believe that the present framework will provide a sound basis for examining a number of these issues.

APPENDIX: RELATIONSHIP BETWEEN ARIMA MODELS AND SSOE MODELS (INCLUDING STATIONARITY AND INVERTIBILITY)

Any ARIMA model may be represented by an SSOE model (Snyder 1985). For example, any ARIMA(p, d, q) scheme may be represented by an SSOE model with

$$\mathbf{h} = (1, 0, \ldots, 0)'$$

$$\mathbf{F} = \begin{bmatrix} \phi & \mathbf{I}_{k-1} \\ \vdots & \\ \phi_k & \mathbf{0} \end{bmatrix} \quad \text{and} \quad \alpha = \begin{bmatrix} \phi_1 + \theta_1 \\ \vdots \\ \phi_k + \theta_k \end{bmatrix},$$

where $k = \max(p + d, q)$. Note that, for this representation and in the rest of this appendix, we assume that the AR and MA polynomials do not possess any common roots.

In the reverse direction, any SSOE model can be expressed as an ARIMA model. We can write the state equation of (1.1) as

$$\mathbf{x}_t = (\mathbf{I} - \mathbf{F}B)^{-1}\alpha\varepsilon_t, \tag{A.1}$$

where we now assume that the polynomial $|\mathbf{I} - \mathbf{F}B| = 0$ has no unit roots, a condition relaxed later. Substituting (A.1) into the observation equation of (1.1),

$$y_t = (\mathbf{h}'(\mathbf{I} - \mathbf{F}B)^{-1}\alpha B + 1)\varepsilon_t \equiv \psi(B)\varepsilon_t. \tag{A.2}$$

Equation (A.2) is the MA form of the state-space model. If we write $\mathbf{F}^j = \mathbf{U}\Lambda^j\mathbf{V}$, where Λ is the diagonal matrix of eigenvalues and $(\mathbf{U}, \mathbf{V})$ are the matrices of eigenvectors, (A.2) becomes $y_t = (\mathbf{h}'\mathbf{U}(\mathbf{I} - \Lambda B)^{-1}\mathbf{V}\alpha B + 1)\varepsilon_t$. If all eigenvalues of $\mathbf{F}$ lie inside the unit circle,

$$y_t = \left(1 + \mathbf{h}'\mathbf{U}\left(\sum_{j=0}^{\infty} \Lambda^j B^{j+1}\right)\mathbf{V}\alpha\right)\varepsilon_t.$$

The convergence of the coefficients in the infinite polynomial $\psi(B)$ corresponds to the roots of $(\psi(B))^{-1} = 0$ lying outside the unit circle. Thus, y_t will be stationary if and only if the eigenvalues of $\mathbf{F}$ are inside the unit circle.

When $\mathbf{F}$ has unit eigenvalues, we can write the state equation of (1.1) as

$$(\mathbf{I} - \mathbf{F}B)\mathbf{x}_t = \alpha\varepsilon_t. \tag{A.3}$$

We can multiply both sides of (A.3) by the adjoint of $\mathbf{I} - \mathbf{F}B$, $\mathbf{W}(B)$, to obtain

$$\det(\mathbf{I} - \mathbf{F}B)\mathbf{x}_t = \mathbf{W}(B)\alpha\varepsilon_t. \tag{A.4}$$

If the eigenvalues do not exceed 1, then $\det(\mathbf{I} - \mathbf{F}B) = G(B)H(B)$, where $G(B)$ is a polynomial whose roots are all the unit eigenvalues of $\mathbf{F}$, and $H(B)$ is a polynomial that has an inverse. Then (A.4) can be written as

$$G(B)\mathbf{x}_t = \frac{\mathbf{W}(B)}{H(B)}\alpha\varepsilon_t. \tag{A.5}$$

The new substitution into the observation equation of (1.1) will produce the following ARIMA model in place of (A.2):

$$G(B)y_t = \left(\mathbf{h}'\frac{\mathbf{W}(B)}{H(B)}\alpha B + G(B)\right)\varepsilon_t = \psi(B)\varepsilon_t.$$

If an eigenvalue of $\mathbf{F}$ exceeds 1, then the roots of $[\Psi(B)]^{-1}$ lie within the unit circle and the process is not stationary and cannot be made stationary by applying unit-root operators.

In a similar manner, we can derive the requirements for invertibility. We may write the transition equation of (1.1) as $\mathbf{x}_t = \mathbf{F}\mathbf{x}_{t-1}\alpha(y_t - \mathbf{h}'\mathbf{x}_{t-1}) = \alpha y_t + \mathbf{D}B\mathbf{x}_t$, where $\mathbf{D} = \mathbf{F} - \alpha\mathbf{h}'$. Thus,

$$\mathbf{x}_t = (\mathbf{I} - \mathbf{D}B)^{-1}\alpha y_t. \tag{A.6}$$

Substituting (A.6) into the observation equation of (1.1), $y_t = \mathbf{h}'(\mathbf{I} - \mathbf{D}B)^{-1}\alpha y_{t-1} + \varepsilon_t$. Hence

$$(1 - \mathbf{h}'(\mathbf{I} - \mathbf{D}B)^{-1}\alpha B)y_t = \pi(B)y_t = \varepsilon_t. \tag{A.7}$$

Equation (A.7) is the AR form of the state-space model. This model will be invertible [i.e., roots of $(\pi(B))^{-1} = 0$ lie outside the unit circle] if and only if the eigenvalues of $\mathbf{D}$ lie inside the unit circle. It can easily be shown that

$$\mathbf{x}_t = \mathbf{D}^t\mathbf{x}_0 + \sum_{i=1}^{t} \mathbf{D}^{t-i}\alpha y_i. \tag{A.8}$$

Hence, an important observation is that $\mathbf{D}^t \to \mathbf{0}$ as $t \to \infty$ if and only if the model is invertible.

[Received June 1999. Revised July 2000.]

REFERENCES

Aitchison, J., and Dunsmore, I. R. (1975), *Statistical Prediction Analysis*, London: Cambridge University Press.

Akaike, H. (1974), "Markovian Representation of Stochastic Processes and its Application to the Analysis of Autoregressive Moving Average Processes," *Annals of the Institute of Statistical Mathematics*, 26, 363–387.

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

Archibald, B. C. (1990), "Parameter Space of the Holt–Winters Model," *International Journal of Forecasting*, 6, 199–209.

Barnett, G., Kohn, R., and Sheather, S. J (1996), "Robust Estimation of an Autoregressive Model Using Markov Chain Monte Carlo," *Journal of Econometrics*, 74, 237–254.

——— (1997), "Robust Bayesian Estimation of Autoregressive-Moving Average Models," *Journal of Time Series Analysis*, 18, 11–28.

This content downloaded from
152.32.254.10 on Thu, 23 Dec 2021 02:23:34 UTC
All use subject to https://about.jstor.org/terms

Bowerman, B. L., and O'Connell, R. T. (1993), *Time Series Forecasting*, Boston: Duxbury.

Chatfield, C. (1993), "Calculating Interval Forecasts," *Journal of Business & Economic Statistics*, 11, 121–135.

De Jong, D. N., and Whiteman, C. H. (1994), "The Forecasting Attributes of Trend- and Difference-Stationary Representations for Macroeconomic Time Series," *Journal of Forecasting*, 13, 279–297.

Garcia-Ferrer, A., and Del Hoyo, J. (1992), "On Trend Extraction Models: Interpretation, Empirical Evidence and Forecasting Performance," *Journal of Forecasting*, 11, 645–665.

Geweke, J. (1989), "Bayesian Inference in Econometric Models Using Monte Carlo Integration," *Econometrica*, 57, 1317–1339.

Harvey, A. C. (1990), *Forecasting, Structural Time Series Models, and the Kalman Filter*, London: Cambridge University Press.

Kabaila, P. (1993), "On Bootstrap Predictive Inference for Autoregressive Processes," *Journal of Time Series Analysis*, 14, 473–484.

McCullough, B. D. (1994), "Bootstrapping Forecast Interval: An Application to AR(p) Models," *Journal of Forecasting*, 13, 51–66.

McKenzie, E. (1984), "Concurrent Seasonal Adjustment With Census X-11," *Journal of Business & Economic Statistics*, 2, 235–249.

Ord, J. K., Koehler, A. B., and Snyder, R. D. (1997), "Estimation and Prediction for a Class of Dynamic Nonlinear Statistical Models," *Journal of the American Statistical Association*, 92, 1621–1629.

Schweppe, F. (1965), "Evaluation of Likelihood Functions for Gaussian Signals," *IEEE Transactions on Information Theory*, 11, 61–70.

Snyder, R. D. (1985), "Recursive Estimation of Dynamic Linear Statistical Models," *Journal of the Royal Statistical Society*, Ser. B, 47, 272–276.

——— (1988), "Statistical Foundations of Exponential Smoothing," Working Paper 5/88, University of Monash, Dept. of Econometrics.

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.

Tsay, R. S. (1993), Discussion of "Calculating Interval Forecasts," by C. Chatfield, *Journal of Business & Economic Statistics*, 11, 140–142.

West, M., and Harrison, J. (1997), *Bayesian Forecasting and Dynamic Models* (2nd ed.), New York: Springer-Verlag.

Winters, P. R. (1960), "Forecasting Sales by Exponentially Weighted Moving Averages," *Management Science*, 6, 324–342.

Yar, M., and Chatfield, C. (1990), "Prediction Intervals for the Holt–Winters Forecasting Procedure," *International Journal of Forecasting*, 6, 127–137.

This content downloaded from
152.32.254.10 on Thu, 23 Dec 2021 02:23:34 UTC
All use subject to https://about.jstor.org/terms