© Journal of Business & Economic Statistics, Vol. 1, No. 3, July 1983

The Prediction of Time Series With Trends and Seasonalities

Will Gersch
Department of Information and Computer Sciences, University of Hawaii, Honolulu, HI 96822

Genshiro Kitagawa
The Institute of Statistical Mathematics, 4-6-7 Minami-Azabu, Minato Ku, Tokyo, Japan-106

A maximization of the expected entropy of the predictive distribution interpretation of Akaike's minimum AIC procedure is exploited for the modeling and prediction of time series with trend and seasonal mean value functions and stationary covariances. The AIC criterion best one-step-ahead and best twelve-step-ahead prediction models can be different. The different models exhibit the relative optimality properties for which they were designed. The results are related to open questions on optimal trend estimation and optimal seasonal adjustment of time series.

KEY WORDS: AIC; Bayesian analysis; Entropy; Prediction; Seasonal time series; Smoothness priors.

1. INTRODUCTION

In this article we consider the optimal smoothing and forecasting of nonstationary time series with trend and seasonal mean value components with stationary covariance. Two classes of smoothness priors trend models are considered. The smoothness priors approach to time series modeling was developed earlier in papers by Akaike and by us (Akaike 1980, Kitagawa and Gersch 1982).

In one model the trend is modeled as a stochastically perturbed local polynomial function of time. In the other model, the trend is assumed to consist of both the stochastically perturbed local polynomial plus a "global" stationary time series component. A predictive likelihood interpretation of Akaike's AIC is exploited to determine the best of the models from the alternative trend model classes, for best one-step-ahead and best twelve-step-ahead prediction criteria. The modeling and smoothing of time series is done using a Kalman predictor/smoother-Akaike AIC criterion methodology. The modeling is applied to econometric time series data that are typical of those seasonally adjusted by Census X-11 procedures and by ARIMA-type models. The AIC criterion best one-step-ahead and best twelve-step-ahead prediction models are different. The best models exhibit the relative optimality properties for which they were designed.

The treatment in the article is largely phenomenological.

In detail, we consider two alternative decompositions of the observed time series data into model classes M_1 and M_2, where

$$M_1: y(n) = t(n) + s(n) + \epsilon(n); \quad n = 1, \cdots, N \quad (1.1)$$

$$M_2: y(n) = t(n) + s(n) + v(n) + \epsilon(n); \quad n = 1, \cdots, N. \quad (1.2)$$

In (1.1) and (1.2), $t(n)$ is a local polynomial component, $s(n)$ is a seasonal component, $v(n)$ is a globally stationary autoregressive time series component and $\epsilon(n)$ is an iid $N(0, \sigma^2)$ observation noise component of the observed time series $y(n)$, $n = 1, \ldots, N$. The term $t(n) + s(n)$, $n = 1, \ldots, N$ can be considered to be the unknown mean value function of a nonstationary in the mean time series with the stationary covariance sequence $\{\epsilon(n)\}$ in model M_1 and $\{v(n) + \epsilon(n)\}$ in model M_2.

The smoothness priors approach originated in a paper by E.T. Whittaker (1923). Let $y(n) = f(n) + \epsilon(n)$; $n = 1, \ldots, N$, where $f(\cdot)$ is an unknown smooth function and $\{\epsilon(n)\}$ is an iid sequence from a zero mean finite variance distribution. Whittaker suggested that the solution for the unknown smooth function balance a trade-off between a sum of squares measure of infidelity to the data and infidelity to a kth-order difference equation constraint with the smoothness trade-off parameter to be determined by the investigator. Shiller (1973) suggested an ad hoc method for determining the trade-off parameter. Akaike (1980), in a likelihood of the Bayesian model interpretation of Shiller's work, is an explicit solution to Whittaker's problem. It appears that neither Shiller nor Akaike was aware of Whittaker's work. The key idea of the likelihood of a Bayesian model may have first appeared in Good (1965). In our own earlier work on the modeling of time series with trend and seasonality (Brotherton and Gersch 1981, Kitagawa 1981, and Kitagawa and Gersch 1982), the computation of the likelihood was implemented by the computationally efficient Kalman filter (Kalman 1960). In that work and in this article the trend component $t(n)$, the seasonal component $s(n)$, and the stationary time series $v(n)$ are each expressed in stochastically perturbed contending-model-order dynamic state-space constraint models with unknown process noise variances. It is the likelihood for the unknown variances that is computed with the Kalman filter. (In a Bayesian framework, the process noise variances are hyperparameters (Lindley and Smith 1972). Also, a theorem of Kimeldorf and Wahba (1970, 1971) justifies our use of the stochastic realization of the mean value function of the time series.) Akaike's minimum AIC procedure is used to determine the best of the alternative trend models fitted to the observed data. A smoothing algorithm is subsequently applied to the AIC criterion best model. The final results thus obtained are a "smoothness prior" or Bayesian smooth decomposition of the $t(n)$, $s(n)$, and possibly $v(n)$ components of the observed time series $y(n)$, $n = 1, \cdots, N$. Particular time series that are of interest in the Census Bureau and Bureau of Labor Statistics for seasonal adjustment are analyzed. The one-step-ahead, increasing horizon, and twelve-step-ahead forecasts are computed and shown for both the best-fitted M_1 and M_2 models separately under optimal one-step-ahead and optimal twelve-step-ahead prediction error performance criteria.

A more comprehensive treatment of the smoothing problem approach to the modeling of time series with trend and seasonalities is in Kitagawa and Gersch (1982). Here we emphasize the prediction of such time series. The analysis is in Section 2. State-space representations of the M_1 and M_2 models are described in Section 2.1. The minimum AIC procedure, including the maximization of the entropy of the predictive distribution interpretation of that procedure is in Section 2.2. The Kalman predictor and smoother are described in Sections 2.3 and 2.4, respectively. Examples of M_1 and M_2 modeled time series according to one-step-ahead and twelve-step-ahead prediction criteria are shown in Section 3. In the Summary and Discussion (Sec. 4), we interpret our results and explain why the best short-horizon and long-horizon prediction models can be different.

2. ANALYSIS

In this section the state-space representation of the local polynomial trend or M_1 model and the local polynomial plus globally stationary time series component or M_2 model are shown. Seasonal and observation noise components are also included in both the M_1 and the M_2 models. The minimum AIC procedure for determining the best alternative models is discussed next. The maximization of the expected entropy of a predictive distribution interpretation of the minimum AIC procedure is exploited to determine the AIC criterion best one-step-ahead and twelve-step-ahead predictor models. Following that, the Kalman predictor and smoother are discussed. The discussion includes formulas for the appropriate likelihoods and predictors and smoothers.

2.1 The Models M_1 and M_2

Consider the two alternative model classes M_1 and M_2 for the observed data $y(n)$, $n = 1, \ldots, N$,

$$M_1: y(n) = t(n) + s(n) + \epsilon(n), \qquad n = 1, \ldots, N \quad (2.1)$$

$$M_2: y(n) = t(n) + s(n) + v(n) + \epsilon(n), \qquad n = 1, \ldots, N. \quad (2.2)$$

In (2.1) and (2.2) $t(n)$ is the local polynomial trend, $s(n)$ is the seasonal, $v(n)$ is the globally stationary stochastic component of the observed time series $y(n)$, $n = 1, \ldots, N$ and $\epsilon(n)$, $n = 1, \ldots, N$ is an iid observation-error sequence. For convenience, $\epsilon(n)$ is assumed to be $N(0, \sigma^2)$,σ^2 unknown.

The trend and seasonal components are assumed to be represented by stochastically perturbed difference equation constraints

$$\nabla^k t(n) = w_1(n) \quad \text{for} \quad k = 1, 2, 3, \quad (2.3)$$

$$\sum_{i=0}^{L-1} s(n-i) = w_2(n). \quad (2.4)$$

The stationary process $v(n)$ is assumed to be in the autoregressive (AR) model form

$$v(n) = \alpha_1 v(n-1) + \ldots + \alpha_p v(n-p) + w_3(n). \quad (2.5)$$

In (2.3) $\nabla t(n) = t(n) - t(n-1)$, $\nabla^2 t(n) = t(n) - 2t(n-1) + t(n-2)$ and so on. Also, in (2.3)–(2.5) the process noise components $w_j(n), j = 1, 2, 3$ and observation noise components $\epsilon(n)$ are assumed to be zero-mean independent Gaussian distributed with

$$\begin{pmatrix} w_1(n) \\ w_2(n) \\ w_3(n) \\ \epsilon(n) \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ t0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_1^2 & 0 & 0 & 0 \\ 0 & \tau_2^2 & 0 & 0 \\ 0 & 0 & \tau_3^2 & 0 \\ 0 & 0 & 0 & \sigma^2 \end{pmatrix}\right). \quad (2.6)$$

In a compatible vector-matrix notation (2.6) is

$$\begin{pmatrix} w(n) \\ \epsilon(n) \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} Q & 0 \\ 0 & \sigma^2 \end{pmatrix}\right). \quad (2.7)$$

The constraints in Equations (2.3)–(2.5) and the observation equations in (2.1) and (2.2) are imbedded into the dynamical state space model for the observations,

$$x(n+1) = Fx(n) + Gw(n),$$
$$y(n) = Hx(n) + \epsilon(n), \quad (2.8)$$

The matrices F, G, and H for the M_2 model and an interpretation of the state vector (using an obvious subset notation) are shown in Figure 1. In Figure 1 the matrices F_2, G_2, H_2 are respectively $(k + L' + p) \times (k + L' + p)$, $(k + L' + p) \times 3$, $1 \times (k + L' + p)$, where k is the order of the difference equation constraint in (2.3); $L' = L - 1$ is the seasonal period duration minus one, with $L = 4$, $L = 12$ for quarterly and monthly data respectively; and p is the order of the AR model. The F, G, H matrices for the simpler M_1 model are submatrices of those for the M_2 model. Those M_1 model matrices do not include a provision for the AR component. The vector $(c_1, \ldots, c_k)$ in Figure 1 reflects the trend constraint in (2.3). It is respectively: (1) for $k = 1$; $(2, -1)$ for $k = 2$; and $(3, -3, 1)$ for $k = 3$.

In model M_1 only the process noise parameters τ_1^2, τ_2^2 and the observation noise $\sigma^2$are unknown. In model M_2 the process noise parameters τ_1^2, τ_2^2, τ_3^2, the AR parameters $\alpha_1, \ldots, \alpha_p$, and the observation noise variance σ^2 are unknown. From the smoothing problem point of view, the parameters τ_1^2, τ_2^2, τ_3^2 are trade-off parameters or hyperparameters. The ratio of parameters τ_j^2/σ^2 is a signal-to-noise measure. It expresses the relative uncertainty of the constraints (2.3), (2.4) and (2.5) assumed for the model. Larger values of τ_j^2 imply stricter adherence to the jth class of difference equation constraint.

2.2 The Minimum AIC Procedure

Akaike's minimum AIC procedure (Akaike 1973, 1974), is used to select the best of the alternative parametric models within and between the alternative M_1 and M_2 classes. (Shibata 1980,1981 showed that when the model that is fitted is an approximation to the truth, the minimum AIC procedure is asymptotically efficient under one-step-ahead and k-step-ahead criteria and that the consistent order procedures do not have that property.) That procedure is interpreted here from a maximization of the expected entropy of the predictive distribution approach (Akaike 1981).

Let the true distribution be g and the fitted distribution be f; then the entropy of g with respect to f is

$$\begin{aligned} B(g, f) &= \int g(y) \log\left\{\frac{f(y)}{g(y)}\right\} dy \\ &= E_Y \log\left\{\frac{f(Y)}{g(Y)}\right\} \\ &= E_Y \log f(Y) - E_Y \log g(Y). \end{aligned} \tag{2.9}$$

The true distribution g is unknown. It is known that $B(g, f) \le 0$ and $B(g, f) = 0$ if and only if $f = g$ almost everywhere. The term $E_Y \log g(Y)$ is independent of the distribution f. So it follows that the larger the value of $E_Y \log f(Y)$, the closer f is to g. Correspondingly, the closeness of alternative f's to the unknown true g can be ordered if the quantity $E_Y \log f(Y)$ can be estimated for each alternative f.

Let $y_1, \ldots, y_N$ be the observed data that occur under the true distribution g and let $f(y \mid \theta)$ be an assumed distributional model of the data with parameter θ. Let

$$l(\theta) = \sum_{i=1}^{N} \log f(y_i \mid \theta) \tag{2.10}$$

be the log-likelihood of the parameter θ given the data $y_1, \ldots, y_N$. Then, from the law of large numbers, $(1/N) l(\theta)$ forms a natural estimate of $E_Y \log f(Y \mid \theta)$. Akaike's AIC criterion is a bias corrected estimate of $-2E_Y \log f(Y \mid \hat{\theta})$ for the practical situation in which the value of θ must be estimated from the data, Akaike (1973,1974). The AIC statistic, a bias corrected estimate of $E_Y \log f(Y \mid \hat{\theta})$, is

$$\mathrm{AIC}(\hat{\theta}, y_1, \ldots, y_N) = -2l(\hat{\theta}) + 2k, \tag{2.11}$$

where $\hat{\theta}$ is the MLE of θ, $l(\theta)$ is the maximized log-likelihood, and k is the number of parameters fitted to the model. The minimum AIC procedure preferred model is the one for which the value of the AIC statistic is the smallest.

In the case of a one-step-ahead prediction criterion and a Gaussian model, the likelihood that is computed

$$x(n) = \begin{bmatrix} t(n) \\ \vdots \\ t(n-k+1) \\ v(n) \\ \vdots \\ v(n-p+1) \\ s(n) \\ \vdots \\ s(n-L+2) \end{bmatrix}; \quad F_2 = \begin{bmatrix} \begin{matrix} c_1 & \cdots & c_k \\ 1 & & \\ & \ddots & \\ & 1 & 0 \end{matrix} & 0 & 0 \\ 0 & \begin{matrix} \alpha_1 & \cdots & \alpha_p \\ 1 & & \\ & \ddots & \\ & 1 & 0 \end{matrix} & 0 \\ 0 & 0 & \begin{matrix} -1 & \cdots & -1 \\ 1 & & \\ & \ddots & \\ & 1 & 0 \end{matrix} \end{bmatrix}; \quad G_2 = \begin{bmatrix} 1 & & \\ \vdots & 0 & 0 \\ 0 & & \\ & 1 & \\ 0 & \vdots & 0 \\ & 0 & \\ & & 1 \\ 0 & 0 & \vdots \\ & & 0 \end{bmatrix}; \quad H_2' = \begin{bmatrix} 1 \\ \vdots \\ 0 \\ 1 \\ \vdots \\ 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}.$$

Figure 1. Matrices F, G, and H for the M_2 Model and an Interpretation of the State Vector.

is that of the innovation process. That is, we consider the estimation of $E_Y \log f(y(n) \mid y(n-1), \ldots, y(1), \theta)$ by

$$\frac{1}{N}\sum_{n=1}^{N-1} \log f(y(n+1) \mid y(n), \ldots, y(1), \theta, x(0))$$

$$= - \frac{2}{2(N-1)}\left\{\sum_{n=1}^{N-1} \log 2\pi u^2(n+1 \mid n) - \sum_{n=1}^{N-1} \frac{(y(n+1) - Hx(n+1 \mid n))^2}{u^2(n+1 \mid n)}\right\} \quad (2.12)$$

with

$$u^2(n+1 \mid n) = HV(n+1 \mid n)H' + \sigma^2. \quad (2.13)$$

In (2.12) and (2.13) the notation $x(n \mid n-1)$ and $V(n \mid n-1)$ are respectively the conditional mean and conditional covariance of the state vector $x(n)$ given the past data $y(n-1), y(n-2), \ldots, y(1)$, the parameter vector θ, and the initial state vector $x(0)$. In the model M_1, $\theta = (\tau_1^2, \tau_2^2)$ and in M_2, $\theta = (\tau_1^2, \tau_2^2, \tau_3^2, \alpha_1, \ldots, \alpha_p)$. Also in (2.13) $\nu(n+1) = (y(n+1) - Hx(n+1 \mid n))$, $n = 1, \ldots, N$ are the innovations. The innovations represent the difference between the observed data and the conditional mean of the data given the past. The innovations are a normally distributed zero mean independent process and are uniquely determinable from the observed process, $y(n)$, $n = 1, \ldots, N$ (Anderson and Moore 1979).

In the case of twelve-step-ahead prediction, consider the estimation of $E_Y \log f(y(n+12) \mid y(n), \ldots, y(1), \theta)$ by

$$\frac{1}{N-12}\sum_{n=1}^{N-12} \log f(y(n+12) \mid y(n), \theta, x(0))$$

$$= -\frac{1}{2(N-12)}\left\{\sum_{n=1}^{N-12} \log 2\pi u^2(n+12 \mid n) - \sum_{n=1}^{N-12} \frac{(y(n+12) - Hx(n+12 \mid n))^2}{2u^2(n+12 \mid n)}\right\} \quad (2.14)$$

with

$$u^2(n+12 \mid n) = HV(n+12 \mid n)H' + \sigma^2. \quad (2.15)$$

The last lines in (2.12) and (2.14), the approximations for the one-step-ahead and twelve-step-ahead maximized predictive likelihoods, are computed for discrete sets of τ_1^2, τ_2^2 in the M_1 model and of $\tau_1^2, \tau_2^2, \tau_3^2$ in the M_2 model. In detail, the likelihoods are computed with each τ_j searched over the values $\tau_j^2 = 2^k$, $k = 0, 1, \ldots, 7$. (In the M_2 model the AR parameter estimates $\alpha_1, \ldots, \alpha_p$ are estimated by a quasi Newton-Raphson type procedure for the set of values of τ_1^2, τ_2^2, and τ_3^2.) Formulas for the computation of the relevant state-space conditional mean value and variance terms employed in the predictive likelihoods are in the section immediately following.

In computing the value of the AIC, (2.11), for a particular model, an exact maximum likelihood computation is assumed, and the number of unknown parameters must be specified. In the state-space model, that number is the dimension of the state, for the implicit estimation of $x(0)$, plus the number of hyperparameters and the number of AR parameters fitted. Thus in the M_1 model, the number of parameters fitted is $(k + (L-1) + 2)$, the order of the difference equation constraint, the period of the seasonal duration minus one plus two for the hyperparameters τ_1^2, τ_2^2. Similarly, under the M_2 model the number of parameters fitted is $(k + (L-1) + 2p + 3)$, where the dimensionality of the state is $(k + (L-1) + p)$, p is the number of AR parameters fitted, and there are 3 noise process terms or hyperparameters τ_1^2, τ_2^2, and τ_3^2.

2.3 The Kalman Predictor Formulas

Let the mean and covariance of the Gaussian density function of the state $x(n+1)$ given the observations $y(n)$, $y(n-1), \ldots, y(1)$ be denoted by $x(n+1 \mid n)$ and $V(n+1 \mid n$. Starting with the initial conditions $x(0 \mid 0) = x(0)$, $V(0 \mid 0) = V_0$, one-step-ahead prediction equations are computed recursively from

$$x(n+1 \mid n) = Fx(n \mid n),$$
$$V(n+1 \mid n) = FV(n \mid n)F' + GQG'. \quad (2.16)$$

Time update equations are computed from

$$x(n+1 \mid n+1) = x(n+1 \mid n) + K(n+1)\nu(n+1),$$
$$K(n+1) = V(n+1 \mid n)H'[HV(n+1 \mid n)H' + \sigma^2]^{-1},$$
$$\nu(n+1) = y(n+1) - Hx(n+1 \mid n),$$
$$u^2(n+1 \mid n) = HV(n+1 \mid n)H' + \sigma^2,$$
$$V(n+1 \mid n+1) = (I - K(n+1)H)V(n+1 \mid n). \quad (2.17)$$

In (2.17), $K(n+1)$ is the Kalman filter gain at time $n+1$, $\nu(n+1)$ is the innovations at time $n+1$, and $u^2(n+1 \mid n)$ is the conditional variance of $y(n+1 \mid n)$, the observation process, at time $n+1$ given the past data $y(n), \ldots, y(1)$.

The k-step-ahead predictions formulas for $k = 1, 2, \ldots$ are

$$x(n+k \mid n) = F^k x(n \mid n),$$
$$V(n+k \mid n) = F^k V(n \mid n)F^{k\prime} + \sum_{j=0}^{k-1} F^j GQG'F'^j,$$
$$y(n+k \mid n) = Hx(n+k \mid n),$$
$$u^2(k+n \mid n) = HV(n+k \mid n)H' + \sigma^2. \quad (2.18)$$

These quantities are used for the calculation of the likelihood.

For the examples worked in the next section, models are fitted to the $y(1), \ldots, y(N)$ data and several types of predictors are computed for the future data $y(N+1), \ldots, y(N+M)$ from that model. For convenience the formulas for those predictors are

One-step-ahead prediction,

$$y(n+1\,|\,n) \qquad n = N, N+1, \ldots, N+M-1 \qquad (2.19)$$

Increasing horizon prediction,

$$y(N+i\,|\,N) \qquad i = 1, \ldots, M \qquad (2.20)$$

Twelve-step-ahead prediction,

First:

$$y(N+i\,|\,n) \qquad i = 1, \ldots, 12$$

Then:

$$y(N+12+j\,|\,N+j), \qquad j = 1, \ldots, (M-12). \qquad (2.21)$$

A sensible and insightful development of these Kalman filter/predictor formulas is in Anderson and Moore (1979). The original paper on the subject is Kalman (1960). Duncan and Horn (1972) was an early treatment of the Kalman filter in the statistics literature. Chan, Goodwin, and Sin (1982) demonstrated that a critical quantity, the Kalman filter gain, $K(n+1)$ in (2.7), is convergent for transition matrices with zeros on the unit circle. That situation is present in our trend and seasonal models (as in Figure 1).

2.4 The Backward Smoothing Algorithm

The smooth of the state and of the observation at time n given all of the data $y(1), \ldots, y(N)$ are denoted, respectively, by $x(n\,|\,N)$ and $y(n\,|\,N)$. The smoothed estimates are derived from the forward state estimates by the backward smoothing algorithm for $n = N-1, \ldots, 1$ (Anderson and Moore 1979, also Ansley and Kohn 1982.) It is the smoothed estimates of the trend, the seasonal, and when appropriate, the AR component of the series that are used in the final estimates.

$$\begin{aligned} x(n\,|\,N) &= x(n\,|\,n) + A(n)\,(x(n+1\,|\,N) \\ &\qquad - Fx(n\,|\,n), \\ V(n\,|\,N) &= V(n\,|\,n) + A(n)(V(n+1\,|\,N) \\ &\qquad - V(n+1\,|\,n))A(n)', \\ y(n\,|\,N) &= Hx(n\,|\,N), \end{aligned} \qquad (2.22)$$

where

$$A(n) = V(n\,|\,n)F'V(n+1\,|\,n)^{-1}.$$

3. BACKGROUND AND EXAMPLES

3.1 Background

More than 1,000 time series are processed each month at the Bureau of the Census, U.S. Department of Commerce. A table of seasonal components, that is to be used for the next 12 months, is published annually in December for each time series. Both the current unadjusted and the seasonally adjusted time series values are published monthly. The Census X-11 procedure is used for this processing. It and variations of it are used in other U.S. government agencies, in business and industry, and in many countries.

Implicit in the X-11 procedure is a twelve-month-ahead forecast of the seasonal component and hence a conditional forecast of the trend component. This observation motivated us to examine and contrast the one-month and twelve-month forecast model properties of our smoothness priors modeling approach (Kitagawa and Gersch 1982).

We chose to exhibit some of the results of an analysis of one particular variety of time series, the variety with a slightly irregular trend cycle and a strong seasonal pattern. This variety is typical of several hundred such series analyzed monthly at the Census Bureau. The example chosen is the Bureau of Labor Statistics, all employees in food industries, BLSALLFOOD data, January 1966–December 1979, N = 156. This particular series was one of the set provided for analysis by Sandra McKenzie of the Census Bureau to participants in the October 1981 ASA-CENSUS-NBER Conference on Applied Time Series Analysis of Economic Data.

M_1 and M_2 type models, models with local polynomial trend and local polynomial plus globally stochastic stationary trend components were fitted to the BLSALLFOOD data. The original data, and data decompositions including trend, seasonal, and AR components, and the innovations are exhibited for each of the AIC best M_1 and M_2 models under both the best one-step-ahead prediction and the best twelve-step-ahead prediction criteria. The models are fitted to the observed $y(1), \ldots, y(N)$ data. One-step-ahead ($y(n+1\,|\,n)$, $n = N, N+1, \ldots, N+M-1$), increasing horizon ($y(N+j\,|\,N)$, $j = 1, \ldots, M$), and twelve-step-ahead predictions ($y(N+j\,|\,N)$, $j = 1, \ldots, 12$, $y(N+j+12\,|\,N+j)$, $j = 1, \ldots, M-12$), are shown as are the true data $y(N+1), \ldots, y(N+M)$, and plus and minus one sigma confidence intervals. From a likelihood interpretation of the AIC (Akaike 1978), the AIC best of the best M_1 and M_2 model classes is that for which the AIC is minimum.

3.2 Example 1. BLSALLFOOD Data: One-Step-Ahead Models

BLSALLFOOD data Figures 2A, 2B1, and 2C1 show the trend components on the BLSALLFOOD data computed by the Census X-11, default option, and models M_1 and M_2 respectively. Identifying parameters and computational results for the fitting of the M_1 and M_2 models are shown in the tabulation that follows.

Model	M	T	AIC
M_1	(2, 0, 11)	(1, 0, 16)	1348.7
M_2	(2, 2, 11)	(16, 1, 16)	1309.8

The data points are connected together in Figure 2B1 for easier interpretation. A comparison of Figures 2A

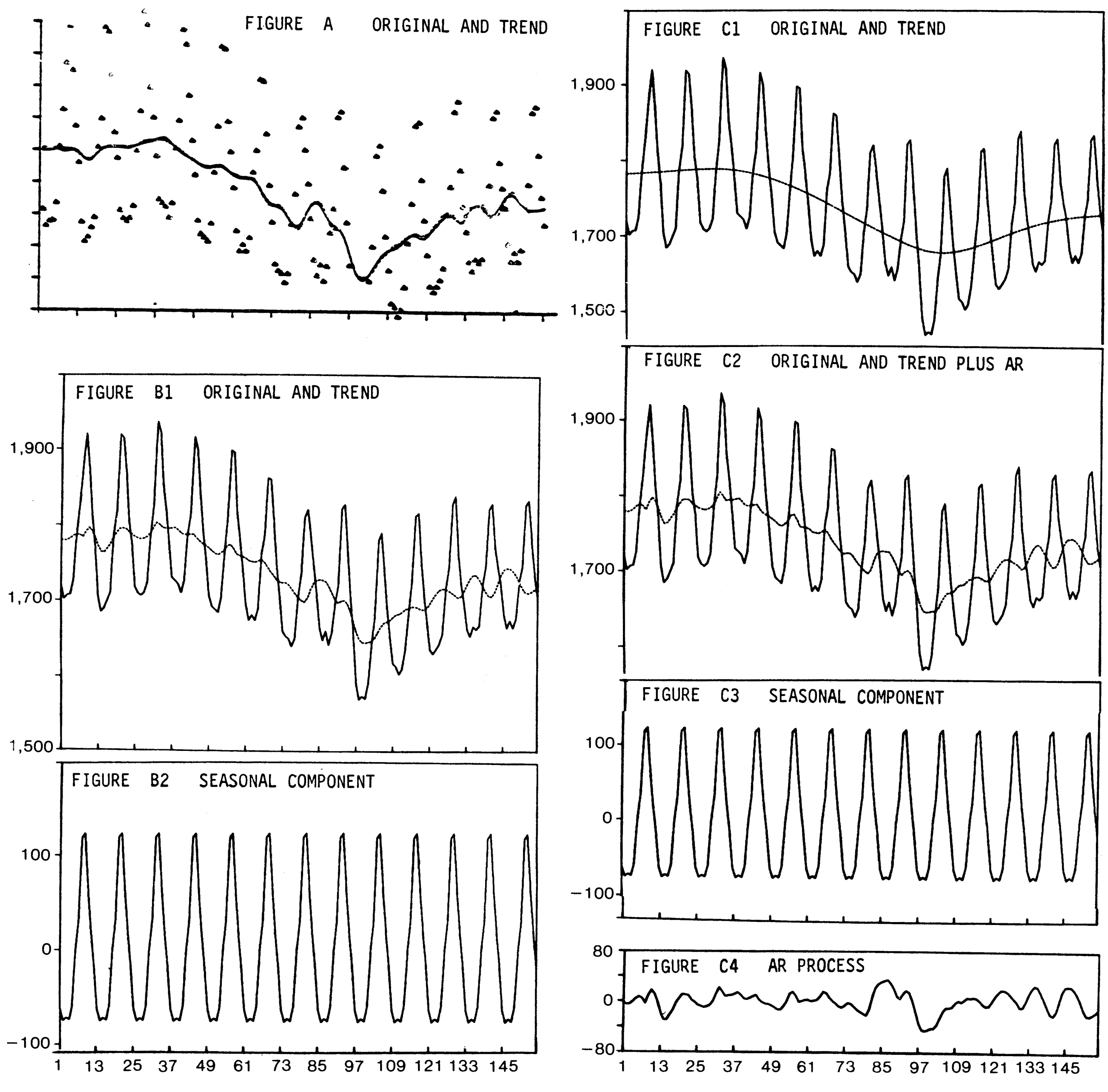

Figure 2. BLSALLFOOD Data and Trends, January 1966–December 1979, N = 156. (A) Census X-11, Default Option. (B) Model M_1: (B1) Original Data Plus Trend, AIC = 1342.49; (B2) Seasonal Component. (C) Model M_2: (C1) Original Data Plus Trend; AIC = 1309.82; (C2) Original Data and Trend Plus AR Component; (C3) Seasonal Component; (C4) AR Component.

and 2B1 illustrates that the M_1 modeled trend is very similar to that obtained by the Census X-11 program. The X-11 trend is computed by an ad hoc two-sided filtering method that was developed to achieve acceptable or pleasing results for the knowledgeable consumer. The local polynomial trend plus AR component computed by the M_2 model is in Figure 2C2. The trend computed by the M_2 model is much smoother than that computed by the M_1 model. The trend plus the AR(2) component of the M_2 model are very similar to the trend of the M_1 model. The seasonal components computed in the M_1 and M_2 models, Figures 2B2 and 2C3, are very similar. The notation, $M = (2, 0, 11)$ $M = (2, 2, 11)$ for the M_1 and M_2 models, respectively, in the tabulation, signifies that the trend polynomial model of order 2, and seasonal model $L = 12$ was fitted in both the M_1 and M_2 model classes and that an AR order-2 model was included in the M_2 model class. The notation $T = (1, 0, 16)$; $T = (16, 1, 16)$ associated with the M_1 and M_2 models signifies the corresponding values of τ_j^2/σ^2 j = 1, 2, 3 parameters. Relatively large values of τ_j^2/σ^2 indicate respectively, relatively strong adherence to the second-

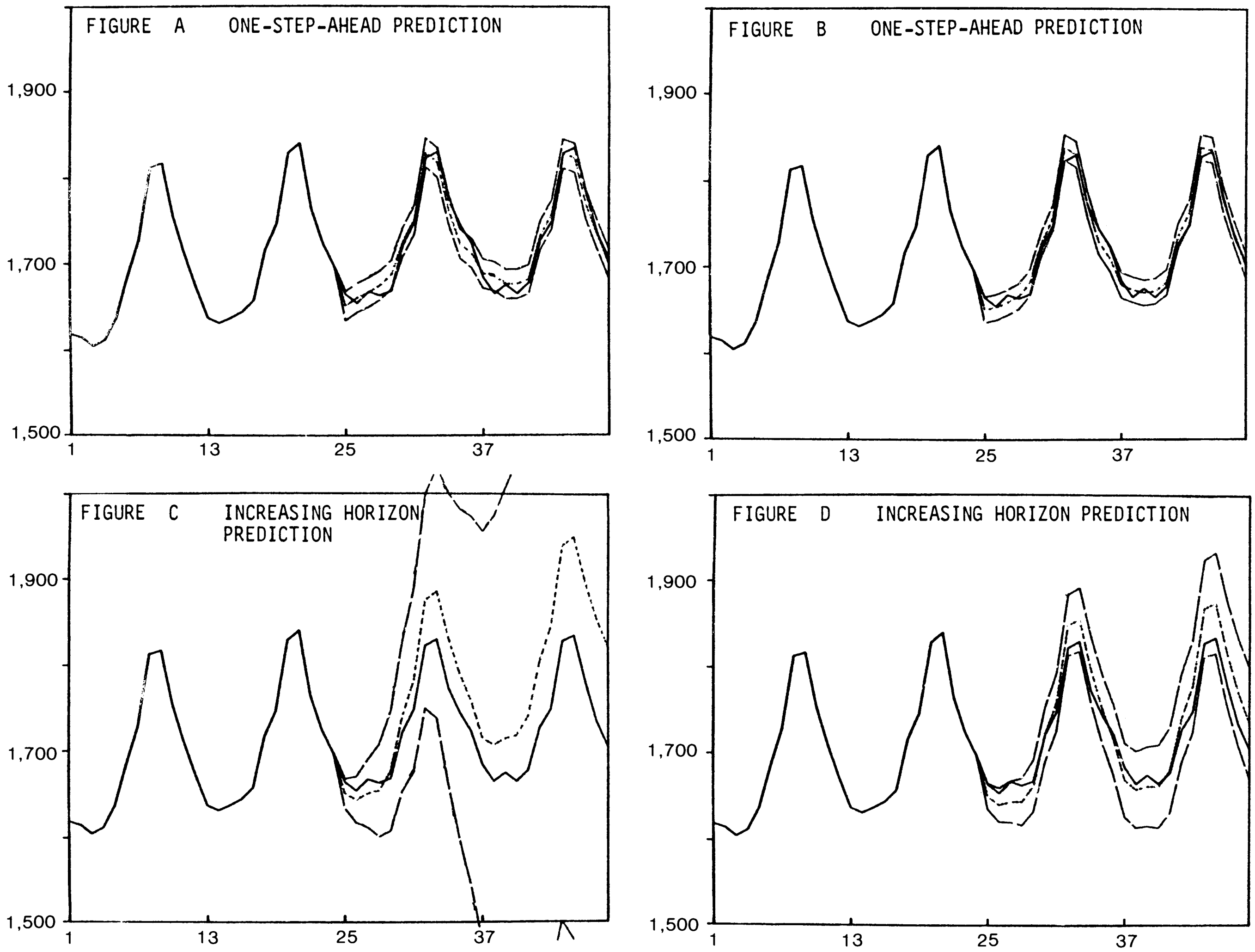

Figure 3. BLSALLFOOD Data Predictions, the Actual Data, and Plus and Minus One Sigma Confidence Intervals. (A) M_1: One-Step-Ahead Predictions. (B) M_2: One-Step-Ahead Predictions. (C) M_1: Increasing Horizon Predictions. (D) M_2: Increasing Horizon Predictions.

order difference equation trend constraint and the corresponding smooth trends. Relatively small values of τ_j^2/σ^2 indicate relatively weak adherence to the difference equation constraint and corresponding "wiggly" trend components.

Prediction performance of the M_1 and M_2 models are shown in Figures 3A–3D. In these and successive prediction performance illustrations, data for 48 months are shown. The model under consideration is fitted from $n = 1, \ldots$ up to the 24th point in the illustration. In each illustration, the actual data, the predictions, and the predictions plus and minus the square root of the predicted variance, $(u^2(k + n \mid n)$ in (2.19), are shown. The true data are indicated by a solid line, the predictions by a dotted line and the prediction standard deviations by dashed lines. In Figures 3A, 3B the statistical performance of the one-step-ahead predictions of the M_1 and M_2 models have similar appearances. An examination of Figures 3A, 3B indicates that the M_2 model prediction performance is superior to that achieved by the M_1 model. The path of the actual BLSAGEMEN data is in closer proximity to the M_2 model predictions than it is to the M_1 model predictions and the plus and minus predictions standard deviation is narrower for the M_2 model than the M_1 model. This superior M_2 model one-step-ahead prediction performance is consistent with a likelihood interpretation of the AIC (Akaike 1978) and the results by Shibata (1980,1981). Shibata showed that the minimum AIC criterion model has an optimal mean squared one-step-ahead prediction property. The increasing horizon performance predictions achieved by the M_1 model, Figure 3C, show the increasing divergence between true and predicted data and the increasing-with-horizon uncomfortably large plus and minus one standard deviation intervals. The increasing-horizon predictions achieved by the M_2 model, Figure 3D, appear to be quite satisfactory. It is important to note that the one-step-ahead best prediction performance does not have any necessary implications about increasing-horizon prediction performance.

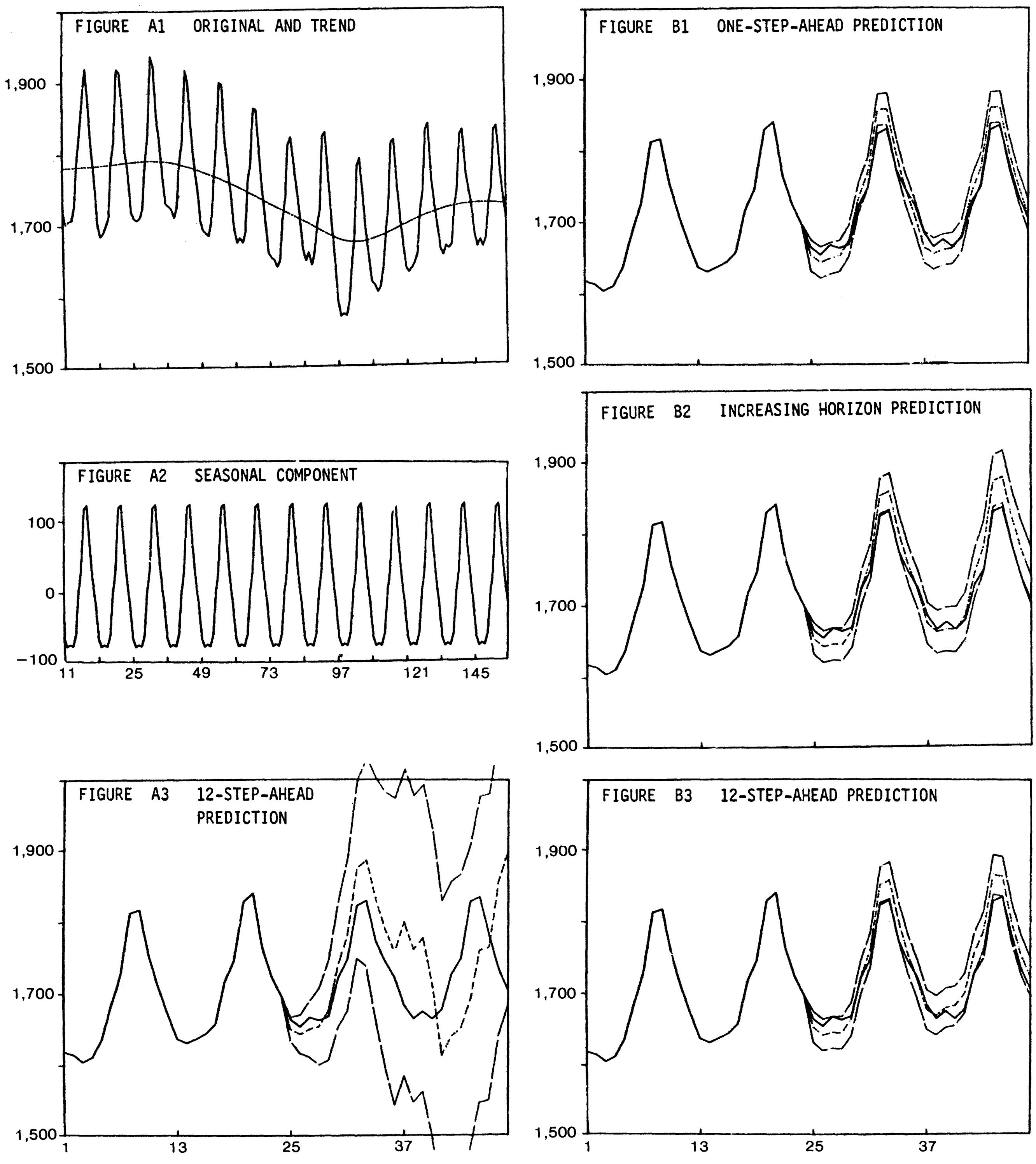

Figure 4. BLSALLFOOD Data, Model M_1: Twelve-Month-Ahead Predictions Criterion, AIC = −725.46. Original Data, Component Decomposition and Predictions, True Values and Plus and Minus One Sigma Confidence Intervals. (A1): Model M_1: Original Plus Trend; (A2) Seasonal Component. (B1) One-Step-Ahead Prediction; (B2) Increasing Horizon Prediction; (B3) Twelve-Step-Ahead Prediction. (C) Model M_1: One-Step-Ahead Criterion Model M_1 Twelve-Month-Ahead Predictions.

3.3 Example 2. BLSALLFOOD Data Reexamined: Twelve-Step-Ahead Models

The following tabulation shows the computational results obtained in fitting the M_1 and M_2 models under the optimum twelve-step-ahead prediction criteria.

Model	M	T	AIC
M_1	(2, 0, 11)	(64, 0, 1)	1538.8
M_2	(2, 2, 11)	(32, 1, 16)	1572.1

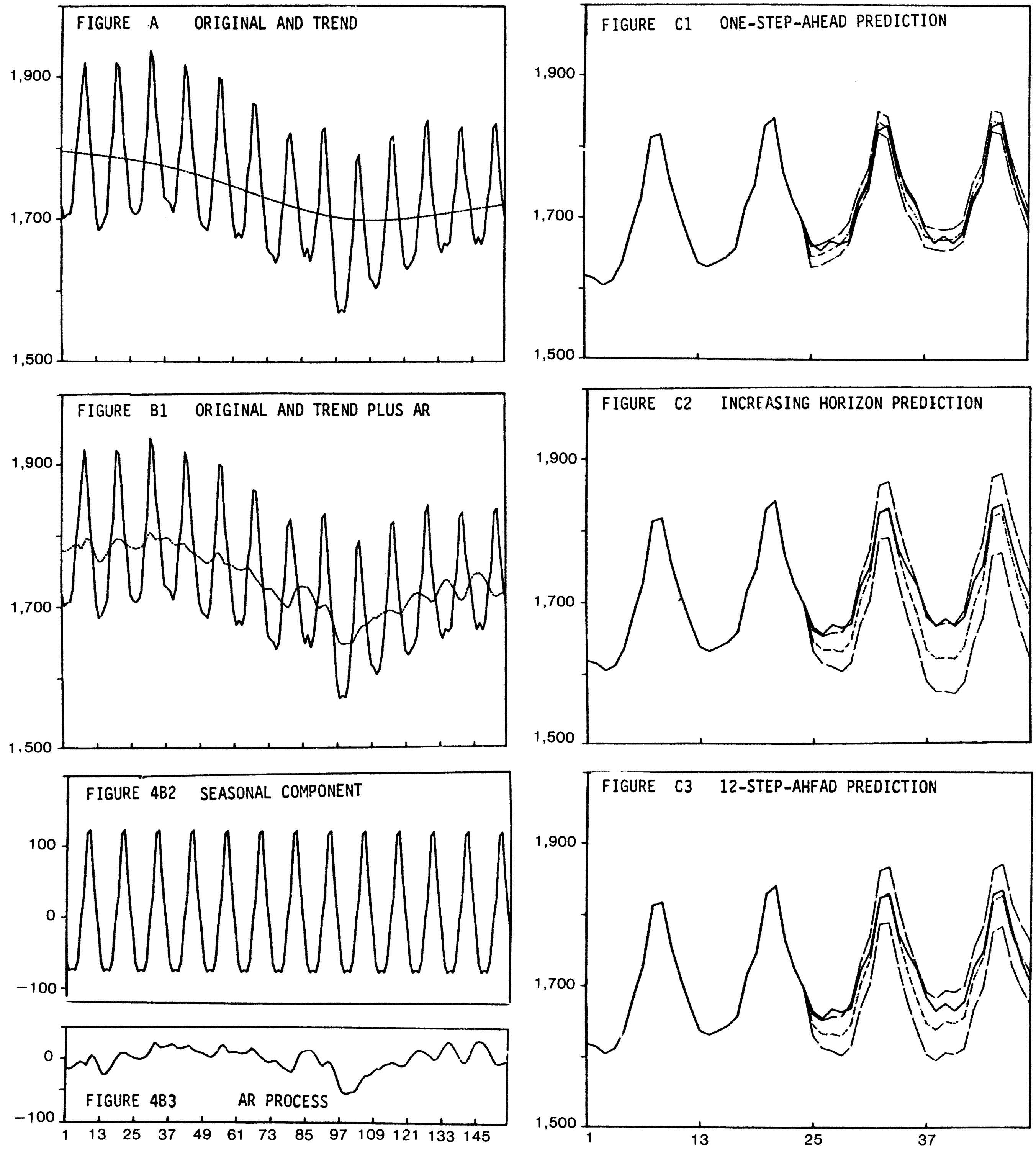

Figure 5. BLSALLFOOD Data, Model M_2: Twelve-Month-Ahead Prediction Criterion, AIC = −715.45. Original Data and Component Decompositions, Predictions, True Values, and Plus and Minus One Sigma Confidence Intervals. (A1) M_1: Original Data and Trend; M_2: (B1) Original Data and Trend Plus AR Components; (B2) AR Component; (B3) Seasonal Component. (C1) M_1: One-Step-Ahead Predictions; (C2) M_s: Increasing Horizon Predictions; (C3) M_2: Twelve-Month-Ahead Predictions.

Figures 4 and 5, respectively, illustrate the original data, the additive component decomposition of the BLSALLFOOD data for the twelve-month-ahead prediction criterion best M_1 and M_2 models and the one-step-ahead, increasing-horizon and twelve-month-ahead prediction performance. For this performance criterion, the 12-month-ahead AIC preferred model is the M_1 model. The twelve-month forecast criterion M_1 trend, Figure 4A, is somewhat similar to the one-month forecast criterion M_2 trend, Figure 2C. The seasonal components of the twelve-month forecast M_1 and M_2 models are very similar to each other. The one-step-ahead prediction performance of the M_2 model is superior to that of the M_1 model. Figures 4B1 and 4C1 reveal that the one-step-ahead predictions of the M_2 model are closer to the truth than are those of the M_1 model, and the standard deviation prediction interval of the M_1 model is narrower than that for the M_2 model.

The M_1 model is superior to the M_2 model in twelve-month and increasing-horizon prediction performance. Note Figures 4B3, 5B3 and Figures 4B2, 5B2. The M_1 model predictions are closer to the actual data than are the M_2 prediction, and the standard deviation interval of the M_1 model is narrower than that for the M_2 model. Figure 4A3 shows the twelve-month-ahead prediction performance achieved by the one-step-ahead best M_1 model. At prediction horizons longer than six months, the prediction performance of this model is quite poor and the plus and minus one standard deviation intervals are excessive. The twelve-month-ahead prediction performance of the M_1 best twelve-step-ahead model is in Figure 4B3. That prediction performance is quite satisfactory. The dramatic superiority in twelve-month prediction performance of the M_1 best twelve-step-ahead model over the M_1 best one-step-ahead model is strong evidence for choosing a long-horizon forecast model rather than a one-step-ahead forecast model for long-horizon forecasting.

4. SUMMARY AND DISCUSSION

A maximization of the expected entropy of the predictive distribution interpretation of Akaike's minimum AIC procedure was exhibited and exploited here in the modeling and prediction of time series with trends and seasonalities. The AIC criterion best one-step-ahead and best twelve-step-ahead prediction models are different, and individually they exhibit the relative optimality properties for which they were designed. These results relate to the trend estimation and seasonal adjustment procedures in the Census X-11 (Shiskin and Plewes 1978, Shiskin, Young, and Musgrave 1967), and the ARIMA-based seasonal adjustment methods (Cleveland and Tiao 1976, Hillmer, Bell, and Tiao 1981, and Hillmer and Tiao 1982).

Employment of the Census X-11 emphasizes achieving an appraisal of the current status or current trend of an econometric time series. The X-11 procedures are subject to certain practical public data reporting constraints that influence the determination of that trend (Shiskin and Plewes 1967). The X-11 seasonal adjustment procedures are implicitly prediction motivated procedures in that seasonalities one year in advance are computed to facilitate deseasonalization of current data.

The invertible ARIMA model is an innovations-type model. Thus, it has optimal one-step-ahead prediction properties under the class of signal and noise model constraints with which it is designed. It is conventional to examine the increasing-horizon forecast performance of such models (Box and Jenkins 1970). With ARIMA models, successive predictions are achieved by concatenations of one-step predictions.

On the other hand, it seems quite reasonable that the model that is best for one-step-ahead prediction and the model that is best for twelve-step-ahead prediction be different. The best one-step-ahead prediction model takes the relatively fast, recent wiggles into account in predicting the near future. Since the cost of the prediction variance of predicting with that wiggling component increases with increasing prediction horizon, the optimal long-range-ahead prediction model should ignore those fast wiggles and instead predict ahead with only a smooth trend. Exactly that situation occurs in our analysis. In both the M_1 and the M_2 model classes, the best twelve-step-ahead model has a smoother trend than the best one-step-ahead model. It is also true in this example that the best one-step-ahead model is in the M_2 model class and the best twelve-step-ahead model is in the M_1 class.

For 5 out of the original 14 time series provided us by Sandra McKenzie for the 1981 ASA-CENSUS-NBER conference, the twelve-month-ahead prediction model was different than the one-month-ahead prediction model. The distinguishing characteristic of those five time series was that their best one-step-ahead trend hyperparameters value was small and correspondingly, their trend estimate was quite wiggly. The hyperparameter of the best twelve-month-ahead prediction model trend was large and the corresponding trend was smooth.

To reconcile the difference between our results and the conventional practice of computing increasing-horizon prediction by concatenation of one-step-ahead predictions, we point out that the implicit assumption in conventional practice is that the fitted models are "true." Certainly this is not the case here. Indeed, the concepts of trend and seasonality are arbitrary (Durbin 1983). Consequently, models fitted to data that exhibits trend and seasonal behavior can only be approximate models. Under that circumstance it is quite reasonable that the choice of the best approximate model would depend on the particular purpose of the model. D. F. Findley (1983) showed an example in which the model that achieves the best one-step-ahead prediction model is not the same as the model that achieves the best k-step-ahead prediction. Let $y(t)$ be a stationary time series width autocorrelation sequence $\rho(j), j = 1, 2, \ldots$ such that $\rho(k) \neq [\rho(1)]^k$. The parameter, a, of the AR_1 model fitted to these data satisfies $a = \rho(1)$. The parameter $\tilde{a}$ of the AR_1 best k-step-ahead prediction model is the minimizer of $E[y(t + k) - \tilde{a}^k y(t)]^2$. Since we assumed that $\rho(k) \neq \rho(1)$, then $\tilde{a} \neq \rho(1)$ achieves the indicated minimization. Here then, the AR_1 best k-step-ahead predictor is not the same as the AR_1 best one-step predictor. As before, the situation is that the fitted model is not in the same model class that generated the data, that is, the model is not the truth.

We compared the one-step-ahead prediction performance of AR, Box-Jenkins-type ARIMA, and our own smoothness priors models on trend plus seasonal type time series data. The AR model class alternatives included AR modeling of the original data, the differenced data, the seasonally differenced data, and the differenced plus seasonally differenced data. We concluded that the overall one-step-ahead statistical prediction performance of those different model classes was very similar. It appears that for data that are dominated by trend and seasonality, almost any reasonable way of modeling those features will give reasonable one-step-ahead prediction performance. The increasing-horizon prediction performance of those alternative models is another much more complex subject.

The following seems to be reasonable: (a) The one-step-ahead prediction performance of almost any modeling method yields quite satisfactory one-step-ahead prediction performance; (b) Superior twelve-step-ahead (or more generally k-step-ahead) prediction performance can be obtained from, say, smoothness-priors models designed to be optimal for twelve-step-ahead (or k-step-ahead) prediction. Therefore for the problem of obtaining the increasing horizon prediction up to M steps ahead, we advocate fitting the set of individual different models that yields the collectively best $k = 1, 2, \ldots, M$-step-ahead predictions. That is quite simple to realize computationally and is logically similar to conventional econometric practice in which different models are employed for short- and long-range prediction horizons.

To conclude, the exhibited statistical performance of the AIC maximized predictive distribution performance procedure suggests new inquiries as to what is really the problem in the seasonal adjustment of time series. Our evidence suggests that rather reliable one-step-ahead and twelve-step-ahead predictions can be obtained by our methodology. The models, and hence the estimate of trend, differ according to whether the desired optimal prediction performance is one step ahead or twelve steps ahead. The implication of the existence of different best models for different prediction horizons is that the best estimate of current trend then becomes dependent on the objectives of the analysis. In light of this discussion, trend estimation and consequently seasonal adjustments are then, once again, open subjects.

ACKNOWLEDGMENTS

The authors would like to express their appreciation to David Findley for his concerned interest and support of their work. Work on this article was initiated when the authors were American Statistical Association Fellows in Time Series at the Census Bureau, U.S. Department of Commerce during 1981–1982.

[Received July 1982. Revised March 1983.]

REFERENCES

AKAIKE, H. (1973), "Information Theory and an Extension of the Maximum Likelihood Principle," in *2nd International Symposium on Information Theory*, eds. B. N. Petrov and F. Csaki, Budapest: Akademiai Kiado, 267–281.

——— (1974), "A New Look at the Statistical Model Identification," *IEEE Transactions on Automatic Control*, ,19, 716–723.

——— (1978), "On the Likelihood of a Time Series Model," *The Statistician*, 27, 217–235.

——— (1980), "Likelihood and the Bayes Procedure," *Bayesian Statistics*, eds. J. M. Bernado, M. H. DeGroot, D. V. Lindley and A. F. M. Smith, Valencia, Spain: University Press, 141–166.

——— (1981), "Likelihood of a Model and Information Criteria," *Journal of Econometrics*, 16.

ANDERSON, B. D. O., and MOORE, J. B. (1979), *Optimal Filtering*, Englewood Cliffs, N.J.: Prentice Hall.

ANSLEY, C. F., and KOHN, R. (1982), "A Geometrical Derivation of the Fixed Interval Smoothing Algorithm," *Biometrika*, 69, 486–487.

BOX, G. E. P., and Jenkins, G. M. (1970), *Time Series Analysis. Forecasting and Control*, San Francisco: Holden-Day.

BROTHERTON, T., and GERSCH, W. (1981). "A Data Analytic Approach to the Smoothing Problem and Some of Its Variations," in *Proceedings of the 20th IEEE Conference on Decision and Control*, 3, 1061–1069.

CHAN, S. W., GOODWIN, G. C., and SIN, K. W. (1982), "Convergence and Properties of the Solutions of the Ricatti Difference Equations," *Proceedings 21st IEEE Conference on Decision and Control*, IEEE Control System Society, 1, 300–305.

CLEVELAND, W. P., and TIAO, G. C. (1976), "Decomposition of Seasonal Time Series: A Model for the Census X-11 Program," *Journal of the American Statistical Association*, 71, 581–587.

DUNCAN, D. B., and HORN, S. D. (1972), "Dynamic Recursive Estimation From the Viewpoint of Regression Analysis, *Journal of the American Statistical Association*, 67, 815–835.

DURBIN, J. (1983), "Theory and Practice in Time Series Analysis," keynote address at ASA/NBER/Census Bureau Conference on Time Series Analysis in Washington, D.C., October 1981.

FINDLEY, D. F. (1983), "On the Indistinguishability of Autoregressive Process From Moving Average Processes," *Journal of Time Series Analysis*, forthcoming.

HILLMER, S. C., BELL, W. R., and TIAO, G. C., (1983), "Modeling Considerations in the Seasonal Adjustment of Economic Time Series, in *Applied Time Series Analysis of Economic Data*, ASA/NBER/Census Bureau Conference Proceedings, forthcoming.

HILLMER, S. C., and TIAO, G. C., (1982), "An ARIMA Based Approach to Seasonal Adjustment," *Journal of the American Statistical Association*, 77, 63–70.

KALMAN, R. E. (1960), "A New Approach to Linear Filtering and Prediction Problems," *Journal of Basic Engineering Transactions*, Ser. D, 82, 35–45.

KIMMELDORF, G. S., and WAHBA, G. (1970), "A Correspondence Between Bayesian Estimation on Stochastic Processes and Smoothing by Splines," *Annals of Mathematical Statistics*, 41, 495–502.

——— (1971), "Some Results on Tchebycheffian Spline Functions," *Journal of Mathematical Analysis and Applications*, 38, 82–05.

KITAGAWA, G. (1981), "A Nonstationary Time Series Model and Its Fitting by Recursive Technique," *Journal of Time Series Analysis*, 2, 103–116.

KITAGAWA, G., and GERSCH, W. (1982), "A Smoothing Problem Approach to the Modeling of Time Series With Trends and Seasonalities," *American Statistical Association Proceedings of the Business and Economic Statistics Section*, 403–408.

LINDLEY, D., and SMITH, A. F. M. (1972), "Bayes Estimates for the

Linear Model," *Journal of the Royal Statistical Society*, Ser. B, 34, 1–41.

GOOD, I. J. (1965), "The Estimation of Probabilities," Cambridge, Mass.: MIT Press.

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

——— (1981), "An Optimal Selection for Regression Variables," *Biometrika*, 68, 45–54.

SHILLER, R. (1973), "A Distributed Log Estimator Derived From Smoothness Priors," *Econometrica*, 41, 775–778.

SHISKIN, J., and PLEWES, T. J. (1978), "Seasonal Adjustment of the U.S. Employment Rate," *The Statistician*, 27, 181–202.

SHISKIN, J., YOUNG, A. H., and MUSGRAVE, J. C. (1967), "The *X*-11 Variant of Census Method II Seasonal Adjustment Program," Technical Paper 15, Bureau of the Census, U.S. Department of Commerce.

WHITTAKER, E. T. (1923), "On a New Method of Graduation," *Proceedings of the Edinborough Mathematical Society*, Ser. II, 41, 63–75.