A Smoothness Priors–State Space Modeling of Time Series With Trend and Seasonality

GENSHIRO KITAGAWA and WILL GERSCH*

A smoothness priors modeling of time series with trends and seasonalities is shown. An observed time series is decomposed into local polynomial trend, seasonal, globally stationary autoregressive and observation error components. Each component is characterized by an unknown variance–white noise perturbed difference equation constraint. The constraints or Bayesian smoothness priors are expressed in state space model form. Trading day factors are also incorporated in the model. A Kalman predictor yields the likelihood for the unknown variances (hyperparameters). Likelihoods are computed for different constraint order models in different subsets of constraint equation model classes. Akaike's minimum AIC procedure is used to select the best model fitted to the data within and between the alternative model classes. Smoothing is achieved by using a fixed-interval smoother algorithm. Examples are shown.

KEY WORDS: Bayesian modeling; Box-Jenkins; Smoothing; Seasonal adjustment; Kalman filter; Likelihood; Trading day adjustment.

1. INTRODUCTION

This article is addressed to the problem of modeling and smoothing time series with trend and seasonal mean value functions and stationary covariances. A modeling approach is taken. We were motivated by the Shiller-Akaike smoothness priors solution to the smoothing problem originally posed by Whittaker (1923). (Some of our earlier work is in Kitagawa 1981 and Brotherton and Gersch 1981.)

Let the observations of a discrete time series be

$$y(n) = f(n) + \epsilon(n);\ n = 1, \ldots, N \tag{1.1}$$

with $\epsilon(n)$ iid from $\mathcal{N}(0, \sigma^2)$, σ^2 unknown, and $f(\cdot)$ an unknown smooth function. The smoothing problem is to estimate $f(n)$, $n = 1, \ldots, N$ in a statistically satisfactory manner. Whittaker suggested that the solution for $f(n)$, $n = 1, \ldots, N$ balance a tradeoff between infidelity to the data and infidelity to a kth-order difference equation constraint on $f(n)$. The choice of a tradeoff parameter was left to the investigator. For a fixed value of the tradeoff parameter, the solution to Whittaker's problem can be expressed in terms of constrained least squares computations, which are parametric in that tradeoff parameter.

A spline smooth–generalized cross validation to determine the smoothness tradeoff parameter approach to the smoothing problem was developed and extensively exploited in applications by Wahba (1977) and Wahba and Wold (1975) and their colleagues. That solution is of computational complexity $O(N^3)$. Wahba (1977) pointed out that the two critical facets of a solution to the smoothing problem are the determination of the smoothness tradeoff parameter and the realization of a computational procedure. In Akaike (1980a), Shiller's (1973) Bayesian smoothness prior idea is fully developed to yield a likelihood computation for determining the smoothness tradeoff parameter. Akaike (1980a) gives an explicit solution to the problem posed by Whittaker. His constrained least squares computational solution is also $O(N^3)$. Akaike (1980b) and Akaike and Ishiguro (1980,1983) smooth time series with trends and seasonalities in the BAYSEA seasonal adjustment program. Initially motivated by Akaike (1980a), we achieved an $O(N)$ computational solution to the smoothing problem, extended some of the ideas of BAYSEA to include a provision for the presence of a stationary stochastic component in the trend and seasonal model, and achieved reliable prediction performance of time series with trends and seasonalities (Gersch and Kitagawa (1983). Our approach is also a Bayesian–smoothness prior approach that yields the smoothness tradeoff parameters as a likelihood computation.

Stochastically perturbed difference equation constraints on the trend, seasonal, stationary time series, and trading day components of the observed time series are expressed in a state space model. The computation of the likelihood of the hyperparameters that balance the smoothness tradeoffs of the trend, seasonal, stationary stochastic, and observation error components of the data is facilitated by an $O(N)$ computational complexity—recursive computational Kalman predictor. Akaike's (1973,1974) minimum AIC procedure is used to determine the best of alternative trend and stochastic component

* Genshiro Kitagawa is a Researcher at the Institute of Statistical Mathematics, 4-6-7 Minami Azabu, Minato Ku, Tokyo, Japan, and Will Gersch is Professor, Department of Information and Computer Sciences, University of Hawaii, Honolulu, HI 96822. The work reported here was begun in 1981–1982, when the authors were American Statistical Association Fellows at the U.S. Bureau of the Census. Enthusiastic encouragement from, and conversations with, Dr. David Findley are gratefully acknowledged. The authors are also grateful for valuable comments by referees.

© Journal of the American Statistical Association
June 1984, Volume 79, Number 386
Theory and Methods Section

difference equation orders and to determine the best model of alternative model classes. Finally, the AIC best modeled data is smoothed by a fixed-interval smoother algorithm.

Fitting the trend plus seasonal plus stochastic component model to the observed data $y(1), \ldots, y(N)$ implicitly estimates $3N$ parameters (one each for the trend, seasonal, and stochastic components at times $n = 1, \ldots, N$). With Bayesian smoothness priors constraints imposed on those components, they are estimated as the solution of stochastically perturbed difference equations. The hyperparameters, or more precisely the perturbation to observation variance ratios, are the essential parameters of the model. It is the likelihood of the hyperparameters that is computed by the Kalman filter.

The subject treated in this article is very closely related to the subject of seasonal adjustment of time series that is treated, for example, in Shiskin, Young, and Musgrave (1967); Cleveland and Tiao (1976); Pierce (1978); Hillmer, Bell, and Tiao (1983); and Hillmer and Tiao (1982). The smoothing problem approach is closely related to work by Wahba (1977) and Wahba and Wold (1975), and to the maximum penalized likelihood method of Good and Gaskins (1980) and references therein. Young and Jakeman (1979) and Wecker and Ansley (1983) are also of interest.

In Section 2 the Shiller-Akaike smoothness prior solution to the smoothing problem is reviewed. In Section 3 state space smoothness priors models for time series that include trend, seasonality, stationary stochastic, trading day effects, and observation error components are shown. Also included are the minimum AIC method for selecting the best of alternative candidate state space models and the Kalman predictor and smoother formulas. Examples are shown in Section 4; some of the phenomenology of our smoothing problem approach to the modeling of time series with trends and seasonalities is illustrated there as well. In Section 5 we discuss the examples and compare our smoothness priors–minimum AIC procedure with the Box-Jenkins-Tiao (BJT) procedure for the modeling of time series with trends and seasonalities.

2. A BAYESIAN SOLUTION TO THE SMOOTHING PROBLEM

A smoothing problem and an approach to its solution, attributed to Whittaker (1923), is as follows: Let

$$y(n) = f(n) + \epsilon(n) \quad n = 1, \ldots, N \tag{2.1}$$

denote a sequence of observations, where $f(n)$ is an unknown smooth function, and $\epsilon(n)$, $n = 1, \ldots, N$ are iid normal random variables with zero mean and unknown variance σ^2. The problem is to estimate $f(n)$, $n = 1, \ldots, N$ from the observations, $y(1), \ldots, y(N)$, in a statistically sensible way. Here the number of parameters to be estimated is equal to the number of observations. Ordinary least squares or maximum likelihood estimates yield computationally unstable or meaningless results.

Whittaker suggested that the solution $f(n)$, $n = 1, \ldots, N$ balance a tradeoff between infidelity to the data and infidelity to a kth-order difference equation constraint. For fixed values of λ and k, the solution satisfies

$$\min_f \left[\sum_{n=1}^{N} (y(n) - f(n))^2 + \lambda^2 \sum_{n=1}^{N} (\nabla^k f(n))^2 \right]. \tag{2.2}$$

The first term in the brackets in (2.2) is the infidelity-to-the-data measure, the second is the infidelity-to-the-constraint measure, and λ is the smoothness tradeoff parameter. Whittaker left the choice of λ to the investigator.

Akaike's (1980a) smoothness priors solution explicitly solves the problem posed by Whittaker (1923). A version of that solution follows: Multiply (2.2) by $-\frac{1}{2}\sigma^2$; consider σ^2, λ^2, and k known and exponentiate (2.2). Then the solution that maximizes (2.2) achieves the maximization of

$$l(f) = \exp\left\{\frac{-1}{2\sigma^2} \sum_{n=1}^{N} (y(n) - f(n))^2\right\} \cdot \exp\left\{\frac{-\lambda^2}{2\sigma^2} \sum_{n=1}^{N} (\nabla^k f(n))^2\right\}. \tag{2.3}$$

Under the assumption of normality, (2.3) yields a Bayesian posterior distribution interpretation

$$\pi(f \mid y, \lambda, \sigma^2, k) \propto p(y \mid \sigma^2, f)\pi(f \mid \lambda, \sigma^2, k), \tag{2.4}$$

with $\pi(f \mid \lambda, \sigma^2, k)$ the smoothness prior distribution of f and $p(y \mid \sigma^2, f)$ the data distribution, conditional on σ^2 and on f, and $\pi(f \mid y, \lambda, \sigma, k)$ the posterior of f. Akaike (1980a) obtained the marginal likelihood for λ and k by integrating (2.4) with respect to f. He showed the application of this method to several interesting data analysis problems. In Bayesian terminology, λ is a hyperparameter (Lindley and Smith 1972). This "type II maximum likelihood method" of estimation was suggested by Good (1965). (See Good and Gaskins 1980 and references therein.)

3. A KALMAN FILTER–MINIMUM AIC CRITERION SOLUTION TO THE SMOOTHING PROBLEM

Motivated by Akaike (1980a), we developed an equivalent state space smoothness priors approach, which is shown in this section. The time series is decomposed into local polynomial trend, global stochastic trend, seasonal, trading day effect, and observation error components. Difference equation constraints for those components are expressed in state space model form. The state space Kalman filter recursive computation yields the likelihood of the tradeoff parameters. Akaike's (1973,1979) minimum AIC procedure is used to select the best of the alternative state space models. The state space models, Akaike's minimum AIC procedure, and the recursive computational Kalman filtering and smoothing are discussed. (Akaike's least squares computations are of computational complexity $O(N^3)$; ours are of $O(N)$.)

3.1 The Models

The generic state space or signal model for the observations $y(n)$, $n = 1, \ldots, N$ is

$$x(n) = Fx(n-1) + Gw(n),$$
$$y(n) = H(n)x(n) + \epsilon(n), \qquad (3.1)$$

where F, G, and $H(n)$ are $M \times M$, $M \times L$, and $1 \times M$ matrices, respectively, and $w(n)$ and $\epsilon(n)$ are assumed to be zero mean independent and identically distributed normal random variables. $x(n)$ is the state vector at time n, and $y(n)$ is the observation at time n. For any particular model of the time series, the matrices F, G, and $H(n)$ are known, and the observations are generated recursively from an initial state that is assumed to be normally distributed with mean $\bar{x}(0)$ and covariance matrix $V(0)$.

In particular, the general state space model for the time series $y(1), \ldots, y(N)$ that includes the effects of local polynomial trends, stationary AR processes, seasonal components, trading day effects, and observation errors is written in the following schematic form:

$$x(n) = Fx(n-1) + Gw(n)$$

$$x(n) = \begin{bmatrix} F_1 & 0 & 0 & 0 \\ 0 & F_2 & 0 & 0 \\ 0 & 0 & F_3 & 0 \\ 0 & 0 & 0 & F_4 \end{bmatrix} x(n-1) + \begin{bmatrix} G_1 & 0 & 0 & 0 \\ 0 & G_2 & 0 & 0 \\ 0 & 0 & G_3 & 0 \\ 0 & 0 & 0 & G_4 \end{bmatrix} w(n)$$

$$y(n) = [H_1 \quad H_2 \quad H_3 \quad H_4(n)]x(n) + \epsilon(n). \qquad (3.2)$$

In (3.2) the overall state space model $(F, G, H(n))$ is constructed by the component models (F_j, G_j, H_j), $(j = 1, \ldots, 4)$. In order $(j = 1, \ldots, 4)$ these models represent the polynomial trend, stationary AR, seasonal, and trading day effects component models, respectively. The number of state components in the particular model (F_j, G_j, H_j) is designated by M_j, $(j = 1, \ldots, 4)$. (The F_j matrices are square.) By the orthogonality of the representation in (3.2), $(2^4 - 1)$ alternative model classes of trend and seasonality may be constructed from combinations of F_j, G_j, H_j, $(j = 1, \ldots, 4)$. The component models F_j, G_j, H_j, $(j = 1, \ldots, 4)$ satisfy particular difference equation constraints on the components. Some of the particular trend, seasonal, AR, and trading day difference equation constraints that we have employed, and that have representations as the F_j, G_j, H_j, $(j = 1, \ldots, 4)$ matrices in (3.2), are shown immediately following.

1. *Local Polynomial Trend Model:* (F_1, G_1, H_1). The polynomial trend component satisfies a kth-order stochastically perturbed difference equation

$$\nabla^k t(n) = w_1(n); \; w_1(n) \sim \mathcal{N}(0, \tau_1^2), \qquad (3.3a)$$

where $\{w_1(n)\}$ is an iid sequence and ∇ denotes the difference operator defined by $\nabla t(n) = t(n) - t(n-1)$. For $k = 1, 2, 3$, those constraints and the values of M_1, the corresponding F_1, G_1, H_1 matrices, and the state vector components are

$k = 1 = M_1$: $t(n) = t(n-1) + w_1(n)$

$$F_1 = [1],\; G_1 = [1],\; H_1 = [1];\; x(n) = t(n). \qquad (3.3b)$$

$k = 2 = M_1$: $t(n) = 2t(n-1) - t(n-2) + w_1(n)$

$$F_1 = \begin{bmatrix} 2 & -1 \\ 1 & 0 \end{bmatrix},\; G_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix},\; H_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}';$$

$$x(n) = \begin{bmatrix} t(n) \\ t(n-1) \end{bmatrix}. \qquad (3.3c)$$

$k = 3 = M_1$: $t(n) = 3t(n-1) - 3t(n-2) + t(n-3) + w_1(n)$

$$F_1 = \begin{bmatrix} 3 & -3 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix},\; G_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix},\; H_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}';$$

$$x(n) = \begin{bmatrix} t(n) \\ t(n-1) \\ t(n-2) \end{bmatrix}. \qquad (3.3d)$$

The subsequent stochastic, seasonal, and trading day constraints correspond to the state space matrices F_j, G_j, $H_j(n)$, $j = 2, 3, 4$, respectively. For brevity, only the difference equation constraints on those components are shown. The state vector $x(n)$ contains lagged versions of those components. An example of a state space signal model that incorporates each of the constraints is given in (3.9).

2. *Stochastic Trend Model:* (F_2, G_2, H_2). The stationary stochastic component $v(n)$ is assumed to satisfy an autoregressive (AR) model of order p. That is,

$$v(n) = \alpha_1 v(n-1) + \cdots + \alpha_p v(n-p) + w_2(n);$$
$$w_2(n) \sim \mathcal{N}(0, \tau_2^2). \qquad (3.4)$$

In (3.4) $\{w_1(n)\}$ is an iid sequence. The AR model is constrained to be stationary.

3. *Local Polynomial Seasonal Component Models:* (F_3, G_3, H_3). Often we use the seasonal component model

$$\sum_{i=0}^{L-1} s(n-i) = w_3(n);\; w_3(n) \sim \mathcal{N}(0, \tau_3^2), \qquad (3.5a)$$

where L is the number of periods in a season ($L = 4$, $L = 12$ for quarterly and monthly data, respectively), and $\{w_3(n)\}$ is an iid sequence. Then

$$s(n) = -\sum_{i=1}^{L-1} s(n-i) + w_3(n) \qquad (3.5b)$$

or

$$s(n) = -\sum_{i=1}^{L-1} B^i s(n) + w_3(n), \qquad (3.5c)$$

where B^i is the backwards shift operator defined by $B^i s(n) = s(n-i)$. Other seasonal component models that we occasionally employ are

$$\left(1 - \sum_{i=1}^{L-1} B^i\right)^2 s(n) = w_3(n) \tag{3.5d}$$

and

$$s(n) = s(n-L) + w_3(n). \tag{3.5e}$$

The seasonal models in (3.5a) and (3.5e) have been used for comparatively regular and changing seasonal effects, respectively. The seasonal model in (3.5d) is satisfactory for increasing or decreasing seasonal components.

4. *Trading Day Effect Model:* $(F_4, G_4, H_4(n))$. The trading day effect model is an adjustment used because there are a different number of ith days of the week ($i = 1, \ldots, 7$) per month for each successive month (Cleveland and Devlin 1980, Cleveland and Grupe 1983, and Hillmer and Tiao 1982). The adjustment corresponds to the removal of a fixed calendar day effect component from the observed time series. We achieve the trading day adjustment via a state space–Kalman filter regression on fixed regressors. (The use of the Kalman filter for regression on fixed regressors was first suggested by Harvey and Phillips 1979.) The trading day is expressed by

$$\begin{aligned}\sum_{i=1}^{7} \beta_i(n) d_i^*(n) &= \sum_{i=1}^{6} \beta_i(n)(d_i^*(n) - d_7^*(n)) \\ &= \sum_{i=1}^{6} \beta_i(n) d_i(n),\end{aligned} \tag{3.6}$$

where $d_i^*(n)$ denotes the number of ith days of the week in the nth month $d_i = d_i^* - d_7^*$, and $\beta_i(n)$ denotes the trading day factor of ith days of the week at time n. Furthermore, we apply the constraint $\sum_{i=1}^{7} \beta_i(n) = 0$ so that $\beta_7(n) = -\sum_{i=1}^{6} \beta_i(n)$. The nonperturbed difference equation constraint on the trading days is

$$\beta_i(n) = \beta_i(n-1); \; i = 1, \ldots, 6. \tag{3.7}$$

(The trading day effect constraints in (3.6) and (3.7) are shown as functions of n. In fact, the $\beta_i(n)$ converge to steady state values β_i, and the corresponding regression is a fixed-effects regression.)

For a general model including local polynomial and stochastic trends, local polynomial seasonal and trading day components, the state or noise vector $w(n)$, and observation noise $\epsilon(n)$ are assumed to be normal iid with zero mean and diagonal covariance matrix

$$\begin{bmatrix} w(n) \\ \epsilon(n) \end{bmatrix} \sim \mathcal{N}\left[\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_1^2 & 0 & 0 & 0 \\ 0 & \tau_2^2 & 0 & 0 \\ 0 & 0 & \tau_3^2 & 0 \\ 0 & 0 & 0 & \sigma^2 \end{bmatrix}\right]. \tag{3.8}$$

The variances τ_1^2, τ_2^2, τ_3^2, and σ^2 are unknown. The other potentially unknown parameters in the state space model are $\alpha_1, \ldots, \alpha_p$, the AR coefficients of the AR model for the stochastic trend component. Comparatively small values of the τ_1^2, τ_2^2, and τ_3^2 terms imply comparatively strict adherence to the corresponding difference equation constraint.

Model class types fitted to data can be designated by a notation that reveals the constraint orders for the components. For example, $M = (2, 2, 11, 0)$ and $M = (2, 0, 11, 6)$ designate, respectively, the model with trend constraint order 2, AR model order 2 and (monthly) seasonal order 11 without trading day effect, and the model with trend constraint order 2, no AR component monthly seasonal order 11, and the trading day effect component. The vector M plus the values of the hyperparameters for a particular model completely specifies the candidate model to be fitted.

For a specific example, the state and state space structure of a model with $M = (2, 2, 11, 6)$ and observation equation are respectively as in (3.9a) and (3.9b) below.

$$x(n) = \left[\begin{array}{c} t(n) \\ t(n-1) \\ \hline v(n) \\ v(n-1) \\ \hline s(n) \\ \vdots \\ s(n-10) \\ \hline \beta_1(n) \\ \vdots \\ \beta_6(n) \end{array}\right] = \left[\begin{array}{cc|cc|ccc|ccc} 2 & -1 & 0 & 0 & & 0 & & & 0 & \\ 1 & 0 & 0 & 0 & & & & & & \\ \hline 0 & 0 & \alpha_1 & \alpha_2 & & 0 & & & 0 & \\ 0 & 0 & 1 & 0 & & & & & & \\ \hline & & & & -1 & \cdots & -1 & & & \\ & & & & 1 & & 0 & & & \\ 0 & 0 & 0 & 0 & & \ddots & \vdots & & 0 & \\ & & & & & 1 & 0 & & & \\ \hline & & & & & & & 1 & & \\ 0 & 0 & 0 & 0 & & 0 & & & \ddots & \\ & & & & & & & & & 1 \end{array}\right] x(n-1) + \left[\begin{array}{ccc} 1 & 0 & 0 \\ 0 & 0 & 0 \\ \hline 0 & 1 & 0 \\ 0 & 0 & 0 \\ \hline 0 & 0 & 1 \\ \vdots & \vdots & \vdots \\ 0 & 0 & 0 \\ \hline 0 & 0 & 0 \\ \vdots & \vdots & \vdots \\ 0 & 0 & 0 \end{array}\right] w(n) \tag{3.9a}$$

$$y(n) = [1\ 0\ 1\ 0\ 1\ \ldots\ 0\ d_1(n) \ldots d_6(n)]\, x(n) + \epsilon(n). \tag{3.9b}$$

The state process noise vector $w(n)$ and the observation noise $\epsilon(n)$ make up an iid vector with distribution properties given by (3.8). The observation equation (3.9a) explains the observed data $y(n)$ in terms of the contribution of the local polynomial trend, stationary AR process, seasonal, trading day effect, and error components.

If only the trend, $t(n)$, the trend plus AR, $t(n)$, or the seasonal component, $s(n)$, are to be considered, the observation equations revert from the general time form variable $H(n)x(n)$ (as in 3.9b), respectively, to the stationary $Hx(n)$ forms:

$$\begin{aligned} Hx(n) &= [1\ 0\ \quad \ldots]x(n) \\ Hx(n) &= [1\ 0\ 1\ 0\ \quad \ldots]x(n) \\ Hx(n) &= [0\ 0\ 0\ 0\ 1\ \ldots]x(n). \end{aligned} \tag{3.10}$$

In the context of the original Whittaker problem, for the situation in which only the trend and seasonal components are considered, the smoothness priors problem corresponds to the maximization of

$$\begin{aligned} l(f) = \exp&\left\{\frac{-1}{2\sigma^2}\sum_{n=1}^{N}[y(n) - s(n)]^2\right\} \\ &\times \exp\left\{\frac{-\tau_1^2}{2\sigma^2}\sum_{n=1}^{N}[\nabla^k t(n)]^2\right\} \\ &\times \exp\left\{\frac{-\tau_3^2}{2\sigma^2}\sum_{n=1}^{N}\left[\sum_{i=0}^{L-1}s(n-i)\right]^2\right\}. \end{aligned} \tag{3.11}$$

That context emphasizes the role of the hyperparameters τ_1^2 and τ_3^2 as a measure of the uncertainty of belief in the prior. Relatively small τ_1^2 (τ_3^2) imply relatively wiggly trend (seasonal) components. Relatively large τ_1^2 (τ_3^2) imply relatively smooth trend (seasonal) components. Also, the ratio of τ_j^2/σ^2, $j = 1$ or 3, can be interpreted as signal-to-noise ratios. This interpretation suggests that in the vicinity of the maximized likelihood, the likelihood is a rather flat function of the hyperparameters. Indeed, this has been our experience, which suggests an alternative to the usual computationally expensive nonlinear optimization procedures that yield estimates of the hyperparameters via a maximization of the likelihood. We compute the likelihoods, using the efficient Kalman filter algorithm, over a coarse grid of the hyperparameters. Estimation of the hyperparameters is then reduced to a search over the discrete likelihood parameter space. (The value of σ^2 in (3.11) is essentially estimated free of computational charge in the Kalman filter procedures. See the references cited in the section following.)

3.2 The Minimum AIC Procedure

Akaike's (1973,1974) minimum AIC procedure is a statistical estimation procedure for determining the best of alternative parametric models fitted to the data. The AIC of a particular fitted model is

$$\begin{aligned} \text{AIC} = &-2\log(\text{maximized likelihood}) \\ &+ 2(\text{the number of fitted parameters}). \end{aligned} \tag{3.12}$$

In fitting state space models of the kind described in Section 3.1, the total number of parameters fitted is $(M_1 + 2M_2 + M_3 + M_4) + [\delta(M_1) + \delta(M_2) + \delta(M_3)]$, where $(M_1 + M_2 + M_3 + M_4)$ is the dimensionality of the state space, $\delta(M) = 1$ if $M_j \neq 0$, and $\delta(M_j) = 0$ if $M_j = 0$. That is, $M_j = 1$ indicates that the F_j component is included in the signal model. Then the likelihood of the vector of unknown parameters and the initial state given the data is

$$\begin{aligned} L(\theta, \bar{x}(0)) = \prod_{n=2}^{N} f(y(n) \mid y(n-1), \ldots, y(1), \\ \theta, \bar{x}(0)) f(y(1) \mid \theta, \bar{x}(0)), \end{aligned} \tag{3.13}$$

where θ is the parameter vector defined by $\theta = (\tau_1^2, \tau_2^2, \tau_3^2, \sigma^2, \alpha_1, \ldots, \alpha_p)$.

Under the Gaussian assumption, we can exploit the innovations representation achieved with the Kalman predictor as follows:

$$\begin{aligned} &L(\theta, \bar{x}(0)) \\ &\quad = \prod_{n=1}^{N}(2\pi v(n|n-1))^{-1/2}\exp\left\{\frac{-\upsilon(n)^2}{2v(n \mid n-1)}\right\}. \end{aligned} \tag{3.14}$$

In (3.14), $\upsilon(n) = y(n) - Hx(n \mid n-1)$ and $v(n \mid n-1)$ are, respectively, the innovations and the conditional variance of $\upsilon(n)$ at time n. Also, $x(n \mid n-1)$ is the conditional mean of the state vector $x(n)$. The conditioning is on the data $y(n-1), \ldots, y(1)$. The variance of the innovations $v(n \mid n-1)$ is obtained from

$$v(n \mid n-1) = H(n)V(n \mid n-1)H(n)' + \sigma^2, \tag{3.15}$$

where $V(n \mid n-1)$ is the conditional variance of the state vector $x(n)$ given the observations up to time $n-1$.

The likelihood for the hyperparameters is computed for the discrete point set of the values $2^{(j-1)}$ ($j = 1, \ldots, 5$) for each of τ_1^2, τ_3^2 ($\tau_4^2 = 0$). When the stationary AR component is included in the model, τ_2^2 is also searched over $\tau_2^2 = 2^{(j-1)}$, ($j = 1, \ldots, 5$) and the $\alpha_1, \ldots, \alpha_p$ are computed by a quasi-Newton-Raphson type of procedure for each of the points in the τ_1^2, τ_2^2, τ_3^2 space. The parameters $\alpha_1, \ldots, \alpha_p$, τ_1^2, τ_2^2, τ_3^2 for which the AIC is smallest specify the AIC best model of the data.

Some comments on computational complexity are appropriate here. The basic computation for the minimum AIC procedure, (3.12), is the computation of the maximized likelihood for particular classes of parametric models. With correlated data, the likelihood computation usually requires the inversion of an $N \times N$ covariance matrix that needs $O(N^3)$ computations. Equation (3.13), the formula for the likelihood as computed by the Kalman predictor, reveals that the joint density for the observations $y(1), \ldots, y(N)$ has been factored into the product of densities for the innovations $\upsilon(n)$, $n = 1, \ldots, N$. The orthogonalization achieved by the recursive Kalman predictor accounts for the $O(N)$ complexity.

Additional material on the recursive predictor/smoother computations is summarized in the next section.

3.3 Recursive Kalman Filtering and Smoothing

There is a very extensive Kalman methodology literature. Only the barest details and formulas required for our computations are indicated here. Kalman (1960) is the original paper on this subject. Meditch (1969) and Anderson and Moore (1979) give very satisfactory treatments. An early paper in the statistical literature on the Kalman predictor is Duncan and Horn (1972). Chan, Goodwin, and Sin (1982) proved the convergence of the Ricatti equations (equivalently, the convergence of the Kalman gain $K(\cdot)$ in (3.18)) for systems that, like our trend and seasonal models, have zeros on the unit circle. This result provides the theoretical basis for the Kalman filter computations of the likelihoods for our models.

The state space model is

$$x(n) = Fx(n-1) + Gw(n),$$
$$y(n) = H(n)x(n) + \epsilon(n). \tag{3.16}$$

The Kalman methodology yields recursive computations for the predicted, filtered, and smoothed estimates of the state vector $x(n)$ and the signal $H(n)x(n)$ for $n = 1, \ldots, N$. The predicted, filtered, and smoothed state vector and signal are denoted by

$$\begin{array}{ll} \text{predicted} & x(n \mid n-1) \\ & y(n \mid n-1) \\ \text{filtered} & x(n \mid n) \\ & y(n \mid n) \\ \text{smoothed} & x(n \mid N) \\ & y(n \mid N). \end{array} \tag{3.17}$$

In the notation above, $x(n \mid n-1)$ and $y(n \mid n-1)$ denote the estimates of the state vector and the observation at time n given the past observations $y(n-1), \ldots, y(1)$; $x(n \mid n)$ and $y(n \mid n)$ are estimates of the state and observations at time n given the current and past data $y(n), y(n-1), \ldots, y(1)$; and $x(n \mid N)$ and $y(n \mid N)$ are estimates of the state and observation at time n given all the data $y(1), \ldots, y(N)$.

Given the initial vector $x(0 \mid 0)$ and the initial covariance $V(0 \mid 0)$, the quantities required for the computation of the likelihood (3.14) are obtained recursively:

$$x(n \mid n-1) = Fx(n-1 \mid n-1)$$
$$x(n \mid n) = x(n \mid n-1) + K(n)[y(n) - H(n)x(n \mid n-1)], \tag{3.18}$$

where $K(n)$ is the Kalman gain vector

$$K(n) = V(n \mid n-1)H'(n)v(n \mid n-1)^{-1}. \tag{3.19}$$

In (3.19) and subsequently, B' denotes the transpose of B. The updated equations for the variance of the state vector are

$$V(n \mid n-1) = FV(n-1 \mid n-1)F' + GQG',$$
$$V(n \mid n) = (I - K(n)H(n))V(n \mid n-1). \tag{3.20}$$

The likelihood for each of the particular values of τ_1^2, τ_2^2, τ_3^2 is computed, and the parameter set for which the AIC is smallest specifies the AIC criterion best model of the data. For that model, the filtered data are smoothed over the interval $n = N-1, \ldots, 1$ by the fixed interval smoothing formulas

$$x(n \mid N) = x(n \mid n) + A(n)(x(n+1 \mid N) - x(n+1 \mid n)), \tag{3.21a}$$

$$V(n \mid N) = V(n \mid n) + A(n)(V(n+1 \mid N) - V(n+1 \mid n))A(n)', \tag{3.21b}$$

where

$$A(n) = V(n \mid n)F'V(n+1 \mid n)^{-1}. \tag{3.21c}$$

Some comments on the initializing procedure are appropriate here. The Kalman filter algorithm requires the initial values $x(0 \mid 0)$ and $V(0 \mid 0)$. For a stationary system, we can use the theoretical mean value and covariance matrix of the state vector. They are easily computed from the assumed model. For a nonstationary system the theoretical mean and covariance cannot be defined. We use $x(0 \mid 0)$ and $V(0 \mid 0)$, a diagonal matrix with large diagonal values, and then do a first run of the Kalman filter over a time-reversed version of the data to estimate $x(0 \mid 0)$ and $V(0 \mid 0)$. This is equivalent to estimating the initial values from the entire data set.

4. EXAMPLES

In this section some of the phenomenology of the modeling of time series with the additive local polynomial, AR, seasonal, and observation noise components is shown.

Example 1. BLSAGEMEN, $N = 162$. These are Bureau of Labor Statistics data for male agricultural workers 20 years and older. Computational results are shown in Figure 1 for the models indicated in Table 1.

Figures 1A1, 1B1, and 1C1 show the original data and the fitted trends of the corresponding models. The seasonal components of the A and B models are in Figures 1A2 and 1B2, respectively. Figure 1C2 shows the local polynomial plus global autoregressive trend. Prediction results are shown in Figures 1A3, 1B3, 1C3, 1A4, 1B4, and 1C4. The model is fitted to the data $y(1), \ldots, y(N)$, $N = 138$. Prediction is done to estimate the data $y(N+1), \ldots, y(N+K)$, $N = 138$, $K = 24$. Two kinds of predictions are considered. In one-step-ahead prediction, the quantity $y(n+1 \mid n)$, $(n = N, N+1, \ldots, N+K-1)$, is computed. In increasing-horizon prediction, the quantity $y(N+i \mid N)$, $(i = 1, \ldots, K)$, is computed. In

Table 1. Trend and Seasonal Models Fitted to the BLSAGEMEN Data

Model	M	T	$\hat{\sigma}^2$	AIC
A	(2, 0, 11, 0)	(32, 0, 1)	2,014	1,997
B	(2, 0, 11, 0)	(1, 0, 32)	656	1,830
C	(2, 2, 11, 0)	(16, 1, 16)	587	1,789

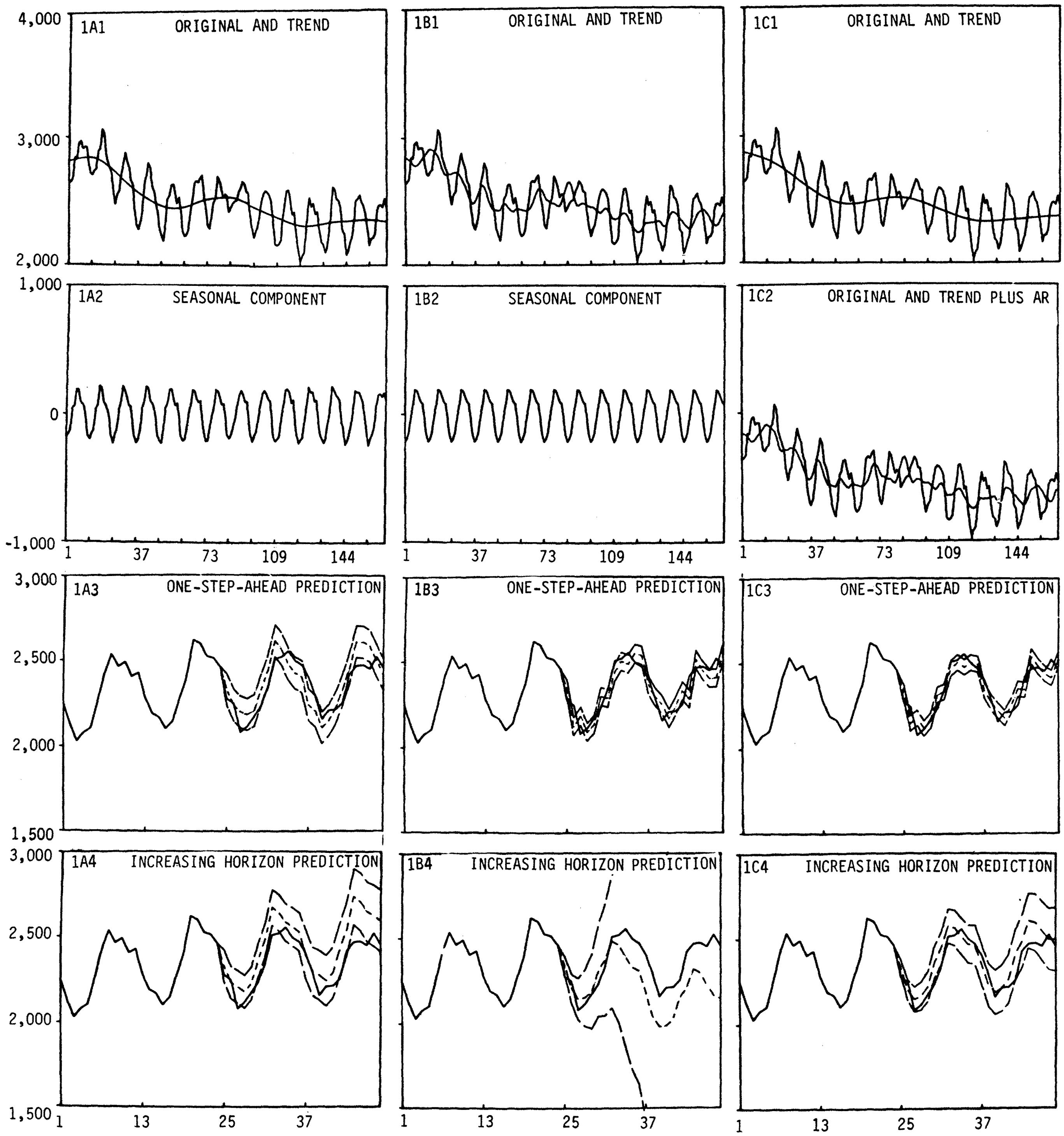

Figure 1. BLSAGEMEN data 1967–October 1980, N = 162. Trend and seasonal components, predictions, true values, and plus and minus one-sigma confidence intervals. (A) Model M = (2, 0, 11, 0), T = (32, 0, 1), $\hat{\sigma}^2$ = 2014, AIC = 1997. A1: Original data and trend; A2: Seasonal component; A3: One-step-ahead predictions; A4: Increasing horizon predictions. (B) Model M = (2, 0, 11, 0), T = (1, 0, 32), $\hat{\sigma}^2$ = 656, AIC = 1830. B1: Original data and trend; B2: Seasonal component; B3: One-step-ahead predictions; B4: Increasing horizon predictions. (C) Model M = (2, 2, 11, 0), T = (16, 1, 16), $\hat{\sigma}^2$ = 587, AIC = 1789. C1: Original data and trend; C2: Original data and trend plus AR component; C3: One-step-ahead predictions; C4: Increasing horizon predictions.

these and all subsequent illustrations showing predictions, the true value, the predicted value (dashed lines), and the computed plus and minus one sigma observation standard deviation (dotted lines) are shown. Figures 1A3, 1B3, and 1C3 are the one-step-ahead predictions for the A, B, and C models, respectively. Figures 1A4, 1B4, and 1C4 are the increasing-horizon predictions for the A, B, and C models, respectively.

Figures 1A1 and 1B1 reveal that the local polynomial trend is smoother for larger values of τ_1^2. Figures 1A2 and 1B2 reveal that the seasonal component is smoother for larger values of τ_2^2. The AIC values of the A, B, and

C models are respectively AIC(A) = 1997, AIC(B) = 1830, and AIC(C) = 1789. The width of the one-step-ahead one-sigma intervals are ranked in order with the AIC, model C having the narrowest one-sigma interval. The AIC ordering of the one-step-ahead prediction performance models does not have any necessary implications on the ordering of increasing horizon prediction performance. In this example, however, the AIC best model, C, does achieve the best increasing horizon prediction performance and does exhibit the narrowest one-sigma prediction interval. The subject of one-step-ahead and k-step-ahead prediction models and their implications for increasing horizon prediction are shown in Gersch and Kitagawa (1983).

Example 2. BLSUEM 16–19. These are Bureau of Labor Statistics data for unemployed males ages 16–19.

Table 2. Models Fitted to the BLSUEM *16–19 Data*

Model	*M*	*N*	*T*	$\hat{\sigma}^2$	AIC
A	(2, 0, 11, 0)	180	(1, 0, 4)	628.7	2,014.2
B	(2, 2, 11, 0)	180	(64, 1, 16)	763.9	1,952.5
C	(2, 0, 11, 0)	48	(16, 0, 16)	—	—

Computational results are shown in Figure 2 for the models indicated in Table 2.

These data were also analyzed by a different method in Hillmer and Tiao (1981). The trend and seasonal components of model A, shown in Figures 2A1 and 2A2, are very similar in appearance to those shown in the Hillmer-Tiao analysis. This is not the AIC best M = (2, 0, 11, 0) model. The overall AIC best of model types M = (2, 0,

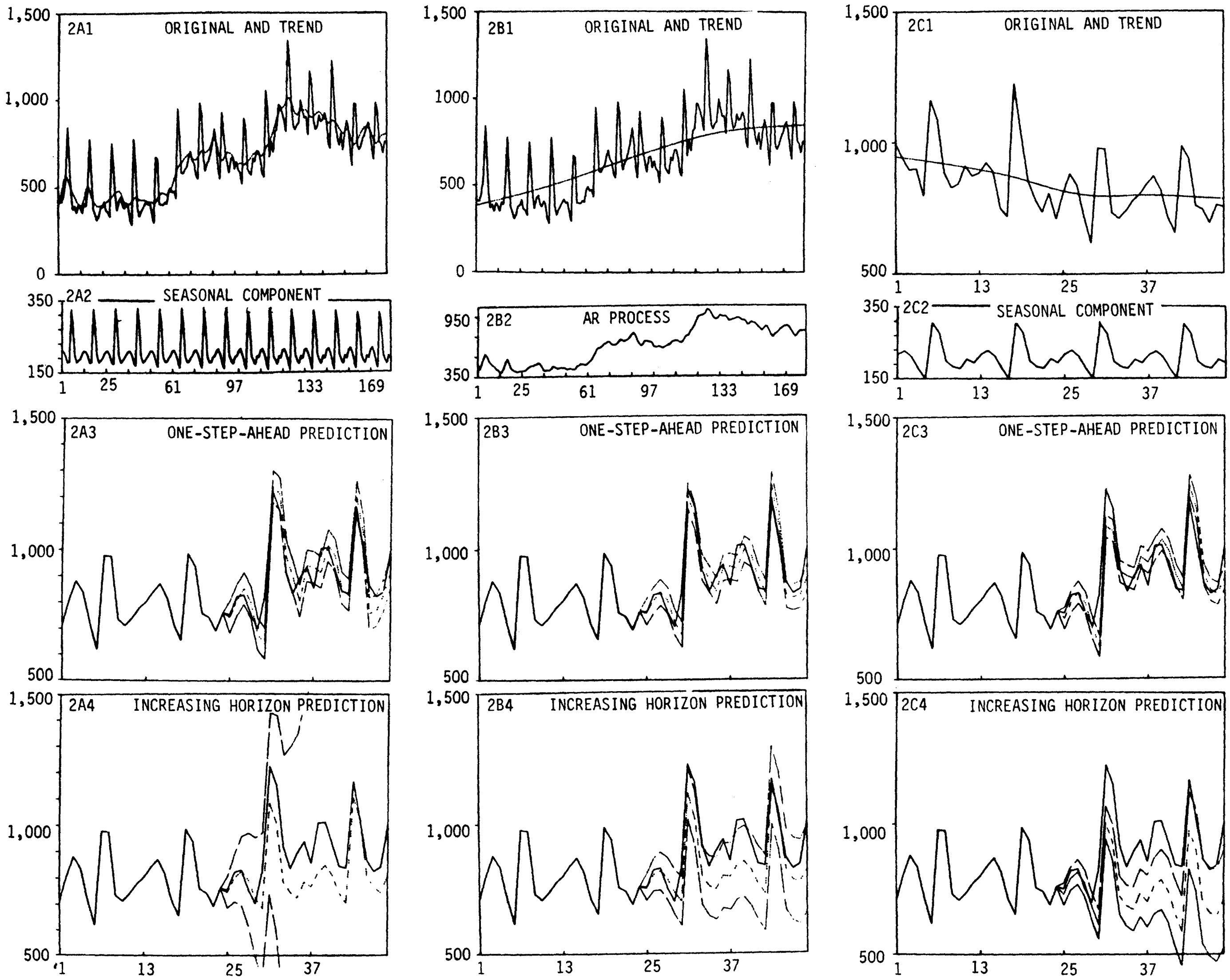

Figure 2. BLSUEM 16–19 Trend and seasonal components, predictions, true values, and plus and minus one-sigma confidence intervals. (A) Model M = (1, 0, 11, 0), T = (1, 0, 4), $\hat{\sigma}^2$ *= 628.7,* AIC *= 2014.2, N = 180, K = 24. A1: Original data and trend; A2: Seasonal component; A3: One-step-ahead predictions; A4: Increasing horizon predictions. (B) Model M = (2, 2, 11, 0), T = (64, 1, 16),* $\hat{\sigma}^2$ *= 763.9,* AIC *= 1952.5, N = 180, K = 24. B1: Original data and trend; B2:* AR *component; B3: One-step-ahead prediction; B4: Increasing horizon prediction. (C) Model M = (2, 0, 11, 0), T = (16, 0, 16), N = 48, K = 24. C1: Original data and trend; C2: Seasonal component; C3: One-step-ahead prediction; C4: Increasing horizon prediction.*

11, 0) and $M = (2, 2, 11, 0)$ considered in Table 2 is Model B, the $M = (2, 2, 11, 0)$ model. (Model C was fitted to a different data span than models A and B, so that its AIC's cannot be compared.) The original data, trend, and autoregressive components estimated by model B, are shown in Figures 2B1, and 2B2, respectively. The seasonal component for Model B is very similar to the seasonal component for Model A, Figure 2A2. The trend plus AR component for Model B is very similar to the trend component of Model A, Figure 2A1. The one-step-ahead and increasing horizon prediction performance of Models A and B are shown in Figures 2A3 and 2A4, and 2B3 and 2B4, respectively. The one-step-ahead one-sigma interval width of Model B is slightly narrower than that of Model A. The increasing horizon prediction one-sigma interval of Model B is much narrower than that of Model A. The models were computed on $N = 156$ data points and predicted for $K = 24$ data points. Some of the computational results for Model C are shown in Figures 2C1–2C4. This model was computed on $N = 24$ data points and predicted for $K = 24$ data points.

Table 3. Trend and Seasonal Models Fitted to the CONHSN *Data*

Model	*M*	*T*	$\hat{\sigma}^2$	AIC
A	(2, 0, 11, 0)	(16, 0, 16)	.301	76.85
B	(2, 0, 22, 0)	(16, 0, 8,192)	.287	68.25

Example 3. CONHSN, $N = 156$, Alternative Seasonal Models. These are Census Bureau construction series data for housing starts. Computational results are shown in Figure 3. They correspond to the models for the CONHSN data shown in Table 3.

Figures 3A1 and 3B1 show the trends of the A and B models to be very similar. The seasonal component shown in Figures 3A2 and 3B2 correspond to the constraint models (3.5a) and (3.5d), respectively, with cor-

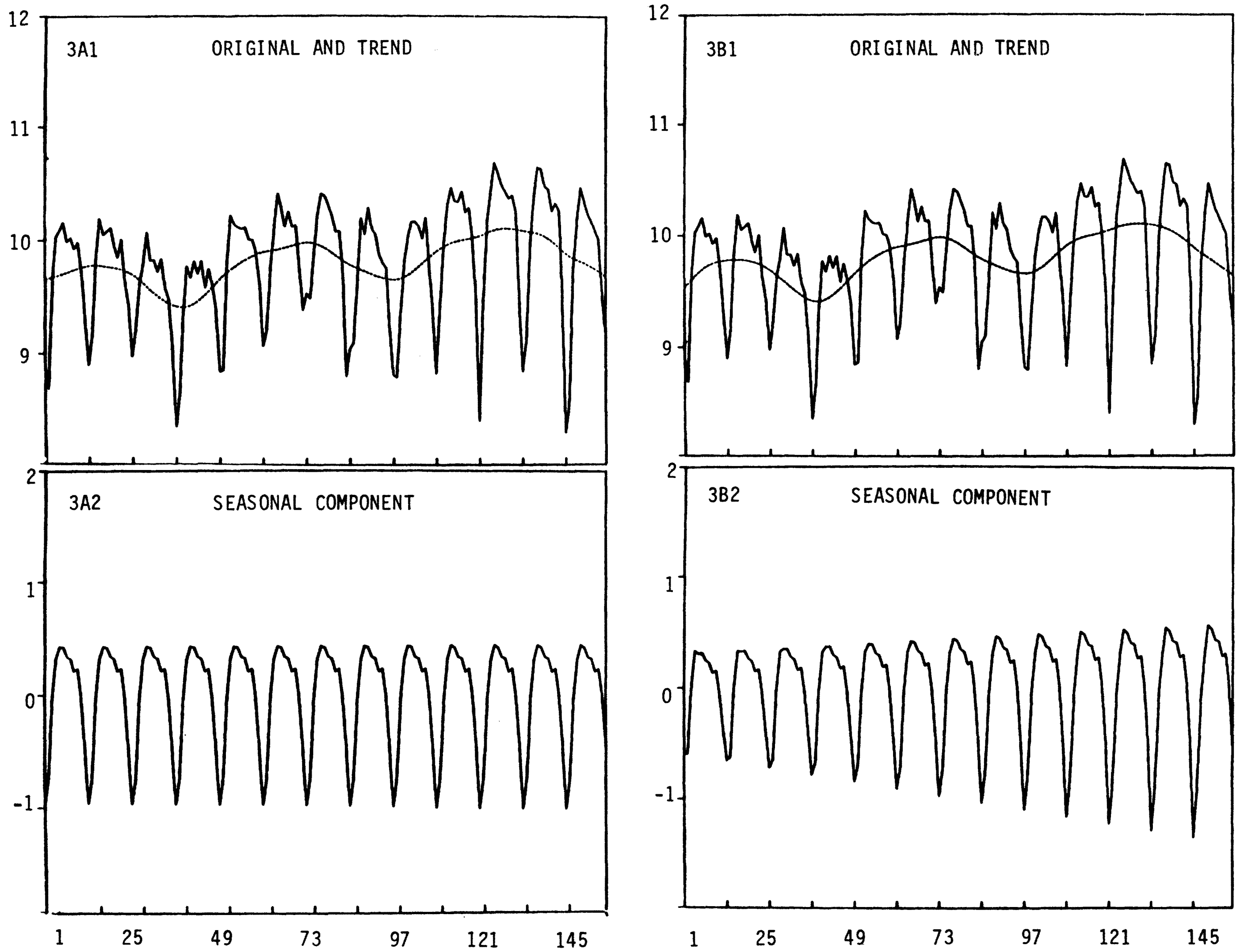

Figure 3. Construction Housing Starts North data, trend, and seasonal components. (A) Model $M = (2, 0, 11, 0)$, $T = (16, 0, 16)$, $\hat{\sigma}^2 = 0.301$, AIC *= 76.85. A1: Original data and trend; A2: Seasonal component. (B) Model $M = (2, 0, 22, 0)$, $T = (16, 0, 8192)$, $\hat{\sigma}^2 = 287$,* AIC *= 68.25. B1: Original data and trend; B2: Seasonal component.*

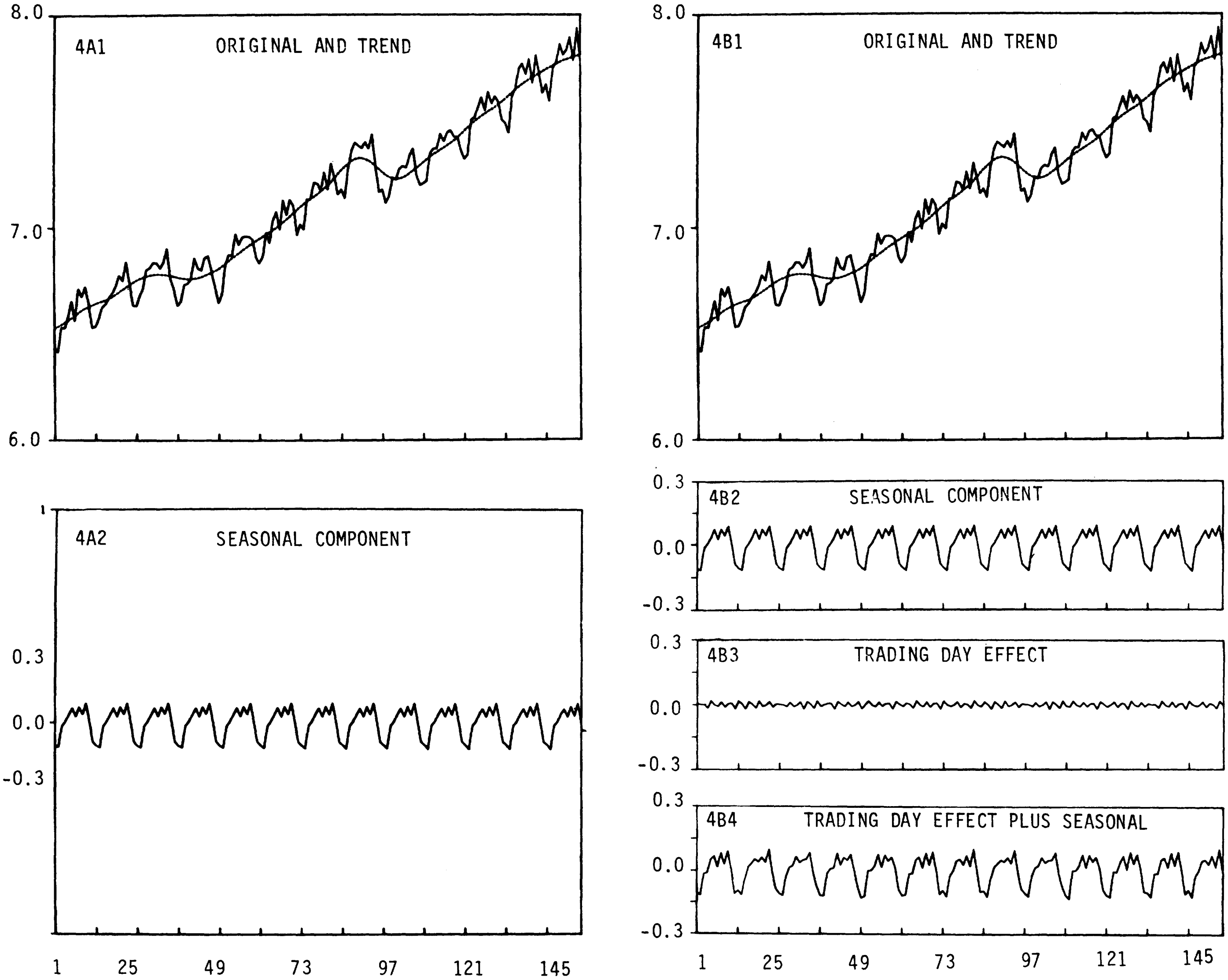

Figure 4. Wholesale Hardware 1967–November 1979 data, N = 156, with and without trading day adjustment. (A) Model M = (2, 0, 11, 0), T = (8, 0, 16), $\hat{\sigma}^2$ = 0.245, AIC = −429.32. A1: Original data and trend; A2: Seasonal component. (B) Model M = (2, 0, 11, 6), T = (8, 1, 16), $\hat{\sigma}^2$ = 0.241, AIC = −439.40. B1: Original data and trend; B2: Seasonal component; B3: Trading day effect; B4: Trading day effect plus seasonal.

responding state vector dimensions 11 and 22. Model B captures the appearance of the increasing seasonal component that is suggested by the data better than does the $M_2 = (L - 1)$ model. Model B is the AIC preferred model.

Example 4. Wholesale Hardware 1/67–11/79, $N = 156$: Trading Day Effect Model. This is Census Bureau data. Computational results are shown in Figure 4. Those results correspond to the models shown in Table 4. Figures 4A1 and 4B1 show the trend of the A and B models, fitted with and without the trading effect, to be very similar. Similarly, the seasonal components shown in Figures 4A2 and 4B2 for the two different models are very similar. The trading day effect and trading day plus seasonal components for the trading day model are shown in Figures 4B3 and 4B4. The trading day effect appears to be minuscule. The superposition of the trading day effect on the seasonal component reveals the irregularizing effect of the number of trading days on the seasonality. The trading day effect model is the AIC criterion best model.

Table 4. Trading Day Effect Model, Wholesale Hardware Data

Model	M	T	$\hat{\sigma}^2$	AIC
A	(2, 0, 11, 0)	T = (8, 0, 16)	.245	−429.32
B	(2, 0, 11, 6)	T = (8, 0, 16)	.241	−439.40

5. SUMMARY AND DISCUSSION

A smoothness priors–Kalman filter–Akaike AIC criterion approach to the modeling of time series with trends and seasonalities was shown. Like the Box-Jenkins procedure, the smoothness priors procedure is a modelbased approach. In our approach, an observed time series is decomposed into additive local polynomial trend, globally stationary autoregressive, seasonal, and observation error components. These components are each characterized by stochastically perturbed difference equations. The perturbations have zero means and unknown variances and are independent of each other. Under the assumption of the normality of the perturbations, the difference equations take on the role of Bayesian priors whose relative uncertainty is characterized by the unknown variances. Alternative time series model classes are characterized by alternative subsets of the constraint equations. Each model class is characterized by models with different order constraint equations and unknown perturbation variances. The constraint equations are expressed in state space model form. The Kalman predictor is employed as an economical computational device to compute the likelihood for the unknown variances and AR coefficients for each of the alternative difference equation model orders in each of the alternative model classes. Akaike's AIC criterion is used to determine the best of the alternative models fitted to the data. The filtered data with this AIC best model is then smoothed using fixed-interval smoother algorithms.

The examples illustrate some of the phenomenology of this smoothness priors approach to the modeling and smoothing of time series with trends and seasonalities. Example 1, BLSAGEMEN data, illustrates the influence of the relative magnitudes of trend and seasonal noise variances on the smoothness of the trend and seasonal components. The modeling performance of two local polynomial trend plus seasonal, and local polynomial plus AR trend plus seasonal, models are shown. The latter is the overall AIC criterion best model. The one-step-ahead prediction performances of the AIC best of the local polynomial trend and local polynomial plus AR trend both model classes are similar. In this example, the AIC best model, model C, has the best increasing horizon prediction performance and the narrowest one-sigma standard deviation interval width. The evidence suggests an interpretation. A relatively smooth trend yields relatively narrow increasing horizon one-sigma prediction intervals. A wiggly trend yields good one-step-ahead prediction performance at the expense of the increasing horizon prediction performance. The local polynomial plus global stationary plus seasonal signal model combines the best predictor properties of the smooth and wiggly trend models.

The BLSUEM 16–19 data were analyzed by Hillmer and Tiao (1982), using a different model analysis. As shown in the example, the trends obtained by the "Wisconsin School" approach are known to be more wiggly than those obtained by the Census X-11 procedure. From the vantage point of our own analysis, the Wisconsin trends appear to be equivalent to some combination of what we refer to as local polynomial and global stochastic components, almost invariably with accompanying relatively wide increasing horizon prediction performance confidence intervals, and quite frequently with relatively poor increasing horizon prediction performance. Insufficient attention has been paid to this tendency.

Examples 3 and 4 exhibit special attributes of our alternative model class characterizations. Example 3, housing starts construction data, illustrates two variations in the modeling of the seasonal component of time series. The data are characterized by an increasing seasonality. The AIC criterion best model clearly captures this pattern. The other seasonality constraint model does not. Example 4, WHARDWARE data, illustrates the modeling of the trading day effect. The AIC criterion best model reveals the impact on the regularity of the seasonal component of the calendar irregularity of the distribution of the number of weekends each month. The trading day effects model achieves regression on fixed regressors within the state space modeling–Kalman filter methodology. These examples reveal that the minimum AIC procedure yields models that agree with human visual judgment when the evidence is clear and also selects models in situations that cannot be handled by the human eye.

The models and examples shown relate to the estimation of trend and seasonal components in the seasonal adjustment of time series. Treatment of that subject has been dominated by the Census X-11 and BJT ARIMA-type modeling procedures. See, for example, Shiskin, Young, and Musgrave (1967), Shiskin and Plewes (1978), and Cleveland and Tiao (1976) for treatments of the X-11 procedure, and see Box and Jenkins (1970), Hillmer, Bell, and Tiao (1983), and Hillmer and Tiao (1982) for treatment of the ARIMA procedures. The X-11 procedures are subject to certain practical public data reporting constraints that influence the trend estimate. There are an extremely large number of variations of smoothing procedures within X-11. Many of the choices of smoothing filters are done subjectively, and there is no effective way of evaluating the statistical properties of those procedures.

There is a critically different attitude toward the diagnostics of modeling between the BJT methodology and ours, which can be seen in our use of the AIC statistic and BJT's use of the Pierce-Box-Ljung Q statistic. The AIC is used to select the best of alternative parametric models within and between model classes. The Q statistic is used to verify the adequacy of a particular candidate model. A distinguishing practical property of our procedure in comparison with the BJT procedure is that ours is essentially a semiautomatic extensive model alternative procedure. The BJT procedure seems to require extensive expert human intervention to achieve satisfactory modeling. Some evidence in support of this appraisal can be seen in the history of the modeling of the Wisconsin telephone data in Thompson and Tiao (1971) and Hillmer (1982). The Thompson-Tiao model is sophisticated, and considerable expertise was required to arrive at that model. Expert experience in the modeling of time series justified Hillmer's use of the trading day effect model. The Q statistic does not.

In addition, the successful AIC criterion modeling of the BLSUEM 16–19, $N = 24$ data point series seems to support the interpretation of our procedure as a semiautomatic procedure even on short-duration series. The small sample–large variability properties of the Q statistic do not lend themselves to reliable diagnostic appraisals of such short duration series. Finally, we suggest that the appropriate testing ground for any time series modeling procedure is in the evaluation of the predictive properties of models fitted by that procedure. A maximization of the expected entropy of the predictive distribution interpretation of the minimum AIC procedure was exhibited in Gersch and Kitagawa (1983) for AIC minimum one-step-ahead and twelve-step-ahead modeling and prediction of time series with trends and seasonalities. Some of that prediction performance analysis appears to transcend the BJT ARIMA model approach.

[*Received July 1982. Revised November 1983.*]

REFERENCES

AKAIKE, H. (1973), "Information Theory and an Extension of the Maximum Likelihood Principle," in *Second 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*, AC-19, 716–723.

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

——— (1980b), "Seasonal Adjustment by a Bayesian Modeling," *Journal of Time Series Analysis*, 1, 1–13.

AKAIKE, H., and ISHIGURO, M. (1980), "BAYSEA, A Bayesian Seasonal Adjustment Program," *Computer Science Monographs*, No. 13, Tokyo: The Institute of Statistical Mathematics.

——— (1983), "Comparative Study of the X-11 and BAYSEA Procedures of Seasonal Adjustment," in *Applied Time Series Analysis of Economic Data*, ed. A. Zellner, Washington, D.C.: U.S. Department of Commerce, Bureau of the Census.

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

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*, 1061–1099.

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

CLEVELAND, W.S., and DEVLIN, S.J. (1980), "Calendar Effects in Monthly Time Series; Detection by Spectrum Analysis and Graphical Methods," *Journal of the American Statistical Association*, 75, 487–496.

CLEVELAND, W.P., and GRUPE, M.R. (1983), "Modeling Time Series When Calendar Effects Are Present," in *Applied Time Series Analysis of Economic Data*, ed. A. Zellner, Washington, D.C.: Department of Commerce, Bureau of the Census.

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., and HORN, S. (1972), "Linear Dynamic Recursive Estimation From the Viewpoint of Regression Analysis," *Journal of the American Statistical Association*, 67, 815–821.

GERSCH, W., and KITAGAWA, G. (1983), "The Prediction of Time Series With Trends and Seasonalities," *Journal of Business & Economic Statistics*, 1, 253–264.

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

GOOD, I.J., and GASKINS, R.A. (1980), "Density Estimation and Bump Hunting by the Penalized Likelihood Method Exemplified by Scattering and Meteorite Data," *Journal of the American Statistical Association*, 75, 42–73.

HARVEY, A.C., and PHILLIPS, G.P.A. (1979), "Maximum Likelihood Estimation of Regression Models With Autoregressive-Moving Average Disturbances," *Biometrika*, 66, 49–58.

HILLMER, S.C. (1982), "Forecasting Time Series With Trading Day Variation," *Journal of Forecasting*, 1, 385–395.

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*, ed. A. Zellner, Washington, D.C.: U.S. Department of Commerce, Bureau of the Census.

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," *Trans. ASME, Journal of Basic Engineering*, Ser. D, 80, 35–45.

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

LINDLEY, D.V., and SMITH, A.F.M. (1972), "Bayes Estimate for the Linear Model," *Journal of the Royal Statistical Society*, Ser. B, 34, 1–41.

MEDITCH, J.S. (1969), *Stochastic Optimal Linear Estimation and Control*, New York: McGraw Hill.

PIERCE, D.A. (1978), "Seasonal Adjustment When Both Deterministic and Stochastic Seasonality Are Present," in *Seasonal Analysis of Economic Time Series*, ed. A. Zellner, Washington, D.C.: U.S. Department of Commerce, Bureau of the Census.

SHILLER, R. (1973), "A Distributed Lag 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, Washington, D.C.: U.S. Bureau of the Census.

THOMPSON, H.E., and TIAO, G.C. (1971), "Analysis of Telephone Data: A Case of Forecasting Seasonal Time Series," *The Bell System Journal of Economics and Management Science*, 2, 515.

WAHBA, G. (1977), "A Survey of Some Smoothing Problems and the Method of Generalized Cross-Validation for Solving Them," in *Applications of Statistics*, ed. P.R. Krishnaiah, Amsterdam: North Holland, 507–524.

WAHBA, G., and WOLD, S. (1975), "A Complete Automatic French Curve: Fitting Spline Functions by Cross Validations," *Communications in Statistics*, 4, 1–17.

WECKER, W.E., and ANSLEY, C.F. (1983), "The Signal Extraction Approach to Nonlinear Regression and Spline Smoothing," *Journal of the American Statistical Association*, 78, 81–89.

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

YOUNG, P.C., and JAKEMAN, A.J. (1979), "The Estimation of Input Variables in Stochastic Dynamic Systems," Report No. AS/R28, Canberra, Australia: Australian National University, Centre for Resource and Environmental Studies.