BUREAU OF THE CENSUS

STATISTICAL RESEARCH DIVISION REPORT SERIES

SRD Research Report Number: CENSUS/SRD/RR-82/04

A SMOOTHNESS PRIORS APPROACH TO THE MODELING OF TIME SERIES WITH TREND AND SEASONALITY*

by

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

and

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

*This paper was written when both authors were American Statistical Association Fellows in Time Series at the U.S. Bureau of the Census, 1981-1982.

Recommended by: Myron J. Katzoff

Report completed: June 30, 1982

Report issued: November 23, 1982

SUBMITTED TO JASA

A SMOOTHNESS PRIORS APPROACH TO THE MODELING OF TIME SERIES WITH TREND AND SEASONALITY *

GENSHIRO KITAGAWA
The Institute of Statistical Mathematics
4-6-7 Minami Azabu, Minato Ku, Tokyo, Japan

and

WILL GERSCH
Department of Information and Computer Sciences
Univeristy of Hawaii, Honolulu, 96822

ABSTRACT: A smoothness priors approach to the 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. A Kalman predictor yields the likelihood for the unknown variances (hyperparameters) with a computational complexity, O(N). 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 a smoother algorithm. Examples are shown.

Key Words: Box-Jenkins, smoothing, seasonal adjustment, Kalman filter, likelihood.

* The work reported here was done in 1981-1982, when the authors were American Statistical Association Fellows at the Census Bureau.

1. INTRODUCTION

This paper is addressed to the problem of modeling and smoothing of 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 in 1919. (Our earlier work is in Kitagawa (1981) and Brotherton and Gersch (1981).)

Consider the smoothing problem: Let the observations of a discrete time series be:

$$y(n) = f(n) + \varepsilon(n); \; n=1,\ldots,N \qquad (1.1)$$

with $\varepsilon(n)$ i.i.d. from $N(0,\sigma^2)$, σ^2 unknown and $f(\cdot)$ an unknown "smooth" function. The 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 k th 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, parametric in that tradeoff parameter.

A spline smooth - generalized cross validation to determine the smoothness tradeoff parameter approach to the smoothing problem has been developed and extensively exploited in applications by Wahba (1975, 1977) and her 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 (1980), Shiller's (1973) Bayesian smoothness priors idea is fully developed to yield a likelihood computation for determining the smoothness tradeoff parameter. Akaike (1980) is an explicit solution to the problem posed by Whittaker. His contrained least squares computational solution is also $O(N^3)$. Akaike (1980), Akaike and Ishiguro (1981), smooth time series with trends and seasonalities in the BAYSEA seasonal adjustment program. Initially motivated by Akaike (1980), we have achieved an O(N) computational solution to the smoothing problem, have extended some of the ideas of BAYSEA to include provision for the presence of a stationary stochastic component in the trend and seasonal model and have achieved reliable prediction performance of time series with trends and seasonalities, (Gersch and Kitagawa (1982)). The O(N) computations were achieved by casting the computations into a recursive form. Our approach is also a Bayesian-smoothness prior approach that yields the smoothness tradeoff parameters as a likelihood computation.

In our approach stochastically perturbed difference equation constraints on the trend, seasonal and stationary time series 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 a recursive computational Kalman predictor. Akaike's minimum AIC procedure, Akaike (1973, 1974), is used to determine the best of alternative trend and stochastic component difference equation orders and to determine the best model of alternative model classes. Finally, the AIC best modeled data is smoothed by a smoother algorithm.

The subject treated here 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), Schlicht (1981), Hillmer Bell and Tiao (1981), and Hillmer and Tiao (1982). The smoothing problem approach is closely related to work by Wahba (1975 and 1977), and to the maximum penalized likelihood method by I. J. Good and Haskins (1980), (and references therein). Young and Jakeman (1979) is also of interest.

In Section 2 a version of the smoothness prior solution to the smoothing problem is shown. In Section 3 state-space 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 AIC criterion best of alternative candidate difference equation model order of the trend and stationary stochastic autoregressive (AR) components for those state space models and the Kalman predictor and smoother formulas. Examples are shown in Section 4.Our objective there is to illustrate the phenomenology of our smoothing problem approach to the modeling of time series with trends and seasonalities. In section 5, Summary and Discussion, the examples are discussed and we compare our smoothness priors-minimum AIC procedure with the Box-Jenkins-Tiao 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, attribtued to Whittaker (1923) is as follows: Let

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

denote a sequence of observations. $f(n)$ is an unknown "smooth" function, $\varepsilon(n)$, $n=1,\ldots,N$ are independent identically distributed 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.

Whittaker suggested that the solution $f(n)$, $n=1,\ldots,N$ balance a tradeoff between infidelity to the data and infidelity to a k-th 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 equation (2.2) is the infidelity to the data measure, the second is the infidelity to the contraint measure and λ is the smoothness tradeoff parameter. Whittaker left the choice of λ, the smoothness tradeoff parameter, to the investigator.

For given λ and k the solution satisfies the constrained least square problem

$$\min_f \left\| \begin{pmatrix} y \\ 0 \end{pmatrix} - \begin{pmatrix} I \\ \lambda D_k \end{pmatrix} (f) \right\|^2 \tag{2.3a}$$

with

$$f = (I + \lambda^2 D_k' D_k)^{-1} y \tag{2.3b}$$

$$SSE(\lambda,k) = y'y - f'(I + \lambda^2 D_k' D_k)\, f. \tag{2.3c}$$

In (2.3c) SSE $(\lambda,k)$is the sum of squares of the residuals. In (2.3a) D_k is the contraint matrix for the k-th difference equation constraint. For example, for k=2, the constrained least squares set up with $D_k = D_2$ becomes

$$\min_f \left\| \underset{2N\times 1}{\begin{bmatrix} y(1) \\ y(2) \\ \cdot \\ \cdot \\ \cdot \\ y(N) \\ \alpha \\ \beta \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{bmatrix}} = \underset{2N\times N}{\begin{bmatrix} 1 & & & & & \\ & 1 & & & & \\ & & \cdot & & & \\ & & & \cdot & & \\ & & & & \cdot & \\ & & & & & 1 \\ \lambda & & & & & \\ -2\lambda & \lambda & & & & \\ \lambda & -2\lambda & \lambda & & & \\ & \cdot & \cdot & \cdot & & \\ & & \cdot & \cdot & \cdot & \\ & & & \cdot & \cdot & \cdot \\ & & & \lambda & -2\lambda & \lambda \end{bmatrix}} \underset{2X1}{\begin{bmatrix} f(1) \\ f(2) \\ \cdot \\ \cdot \\ \cdot \\ f(N) \end{bmatrix}} \right\|^2 . \tag{2.4a}$$

N x N matrix forms for the k=1 and k=3 difference equation constraints; D_1 and D_3 are

$$D_1 = \begin{bmatrix} -1 & & & & \\ 1 & -1 & & & \\ & & \ddots & \ddots & \\ & & & \ddots & \ddots \\ & & & 1 & -1 \end{bmatrix} ; \quad D_3 = \begin{bmatrix} -1 & & & & & & & \\ 3 & -1 & & & & & & \\ -3 & 3 & -1 & & & & & \\ 1 & -3 & 3 & -1 & & & & \\ & \ddots & \ddots & \ddots & \ddots & & & \\ & & \ddots & \ddots & \ddots & \ddots & & \\ & & & \ddots & \ddots & \ddots & \ddots & \\ & & & & 1 & -3 & 3 & -1 \end{bmatrix} . \tag{2.4b}$$

The top row in D_1, 2 top rows in D_2 and 3 top rows in D_3 are related to initial condition constraints on the D_k, k=1, 2, 3 matrices. In D_2 (2.4a), $\alpha = (2f(0) - f(-1))\lambda$, $\beta = -f(0)\lambda$ and f(0) and f(-1) are estimated by a backcasting method.

Akaike's (1980) smoothness priors solution explicity solves the problem posed by Whittaker in 1923. A version of that solution follows: Consider λ known and exponeniate (2.2). Then,

$$\max_{f,k} \ell(f) = \max_{f,k} \left[\exp\{ \frac{-1}{2\sigma^2} \sum_{n=1}^{N} (y(n)-f(n))^2 \} \cdot \exp\{ \frac{-\lambda^2}{2\sigma^2} \sum_{n=1}^{N} (\nabla^k f(n))^2 \} \right] . \tag{2.5}$$

Under the assumption of normality, equation (2.5) yields a Bayesian posterior distribution interpretation

$$\pi(f|y,\lambda,\sigma^2,k) \propto p(y|\sigma^2,f)\pi(f|\lambda, \sigma^2,k) \tag{2.6}$$

with $\pi(f|\lambda,\sigma^2,k)$ the smoothness prior distribution of f and $p(y|\sigma^2,f)$ the data distribution, conditional on σ^2 and on f. Then, the likelihood for λ and k is given by

$$L(\lambda,\sigma^2,k) = \int p(y|\sigma^2,f) \; \pi(f|\lambda,\sigma^2,k)df . \tag{2.7}$$

In Bayesian terminology, λ is a hyperparameter. This "type II maximum likelimethod" of analysis was suggested by I. J. Good (1965). (See also Good and Gaskins (1980) and references therein.)

Directly integrating equation (2.7) and taking minus two times the logarithm of the likelihood yields an explicit closed form expression for $-2\ell n\, L(\lambda,k)$. Maximization of equation (2.5) is equivalent to the minimization of $-2\ell n\, L(\lambda,k)$. Explicitly, the Bayesian optimal smoothness solution of $-2\ell n\, L(\lambda,k)$ is

$$-2\ell n\, L(\lambda,k) \quad = \quad N\, \ell n\, \frac{1}{N} SSE(\lambda,k) + \ell n| \, I + \lambda^2 D_k{}' D_k| - \ell n|\lambda^2\, D_k{}' D_k| . \qquad (2.8)$$

The solution is achieved by a two parameter search over the paramters λ and k. In (2.8) $|A|$ is the determinant of the matrix A, A' denoted the transpose of A and $SSE(\lambda,k)$ is as defined in (2.3c).

3. A KALMAN SMOOTHER - AIC CRITERION SOLUTION TO THE SMOOTHING PROBLEM

In this section the state-space models for the additive decomposition of the observations into local polynomial and stochastic trend, seasonal and observation error components are shown. The trading day effect model is also shown. Then, Akaike's minimum AIC procedure for the state space model is discussed. The critical role of the computation of the likelihood of the tradeoff or hyperparameters is achieved through the use of the Kalman predictor. That computation, the prediction algorithm and the smoother algorithm are also discussed.

3.1 THE MODELS

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

$$\begin{aligned} x(n) &= Fx(n-1) + Gw(n) \\ y(n) &= Hx(n) + \varepsilon(n) \end{aligned} \qquad (3.1)$$

where the $w(n)$ and $\varepsilon(n)$ are, for convenience, assumed to be i.i.d. zero mean normally white noises. $x(n)$ is the state vector at time n and $y(n)$ is the observation at time n. For any particular model of the observations, the matrices F, G and H are known, and the observations are generated recursively from an initial state that is assumed to be normally distributed with unknown mean $\overline{x}(0)$ and infinite covariance V(0). The difference order constraint k and the variances of w(n) and $\varepsilon(n)$ in equation (3.1) are unknown.

In particular, the general state space model for the observations y(1),..., y(N) that includes the effects of local polynomial trend, stationary AR process, seasonal components, trading day effects and observation errors is written in the following schematic form:

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

$$x(n) = \begin{bmatrix} F_1 & 0 & 0 & 0 \\ 0 & F_2 & 0 & 0 \\ 0 & 0 & F_3 & 0 \\ 0 & 0 & 0 & F_4 \end{bmatrix} x(n-1) + \begin{bmatrix} G_1 & 0 & 0 & 0 \\ 0 & G_2 & 0 & 0 \\ 0 & 0 & G_3 & 0 \\ 0 & 0 & 0 & G_4 \end{bmatrix} w(n) \quad (3.2)$$

$$y(n) = [\; H_1 \quad H_2 \quad H_3 \quad H_4(n)]x(n) + \varepsilon(n) .$$

In (3.2) the overall state space model (F, G, H) is constructed by the component models (F_j, G_j, H_j), (j=1,...,4). In order (j=1,...,4) these respectively represent the polynomial trend, the stationary AR, the seasonal and the trading day effects component models. The number of state components in the particular model (F_j, G_j, H_j) is designated by M_j, (j=1,...,4). (The F_j matrices are square). By the orthogonality of the representation in (3.2), (2^4-1) alternative models of trend and seasonality may be composed of combinations of F_j, G_j, H_j elements (j=1,...,4). The component (F_j, G_j, H_j), (j=1,...,4) are defined by particular difference equation constraints on the components. Those constraints are as follows:

(1) Local Polynomial Trend Model; (F_1, G_1)

The trend constraint satisfies a k-th order stochastically perturbed difference equation

$$\nabla^k t(n) = w_1(n); \quad w_1(n) \sim N(0,\tau_1{}^2) . \tag{3.3}$$

For k=1,2,3 those constraints and the values of M_1 and the corresponding F_1, G_1 matrices are:

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

$$F_1 = [1]; \; G_1 = [1]; \tag{3.4a}$$

$$k = 2 = M_2: t(n) = 2t(n-1) - t(n-2) + w_1(n)$$

$$F_1 = \begin{bmatrix} 2 & -1 \\ 1 & 0 \end{bmatrix} \; ; \quad G_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}; \tag{3.4b}$$

$$k = 3 = M_3: t(n)=3t(n-1) - 3t(n-2) + t(n-3) + w_1(n)$$

$$F_1 = \begin{bmatrix} 3 & -3 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \; ; \quad G_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}. \tag{3.4c}$$

(2) Stochastic Trend Model; (F_2, $\hat{G}_2$)

The stationary stochastic component v(n) is assumed to satisfy an autoregressive (AR) model of order p. That is

$$v(n) = \alpha_1 v(n-1) + \ldots + \alpha_p v(n-p) + w_2(n); \quad w_2 \sim N(0,\tau_2{}^2) . \tag{3.5a}$$

For arbitrary p and M_2 = p the F_2, G_2 matrices are:

$$\begin{bmatrix} \alpha_1 \; \alpha_2 \; \cdots \; \alpha_p \end{bmatrix} \qquad \begin{bmatrix} 1 \end{bmatrix}$$

(3) Local Polynominal Seasonal Constraint Models; (F_3, G_3)

Most often we use the stochastically perturbed seasonal constraint

$$\sum_{i=0}^{L-1} s(n-i) = w_3(n); \quad w_3 \sim \mathcal{N}(0,\tau_3^2) \tag{3.6a}$$

where L is the duration of the seasonality. (L=4, L=12 for quarterly and monthly data respectively.)

Then

$$s(n) = -\sum_{i=1}^{L-1} s(n-i) + w_3(n) \tag{3.6b}$$

or
$$s(n) = -\sum_{i=1}^{L-1} B^i s(n) + w_3(n) \tag{3.6c}$$

where B is the backwards shift operator, defined by $Bs(n) = s(n-1)$. Another seasonal constraint model that we occasionally employ is

$$\left(1 - \sum_{i=1}^{L-1} B^i\right)^2 s(n) = w_3(n)\ . \tag{3.6d}$$

Correspondingly the M_3, F_3, G_3 matrices are

$$M_3 = L-1; \quad F_3 = \begin{bmatrix} -1 & \cdots & & -1 \\ 1 & & & \\ & \ddots & & \\ & & 1 & 0 \end{bmatrix}, \; G_3 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{3.6e}$$

$$M_3 = 2(L-1); \quad F_3 = \begin{bmatrix} 2 \ldots L & L-1 \ldots & 1 \\ 1 & & \\ & \ddots & \\ & 1 & 0 \end{bmatrix}; \; G_3 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \tag{3.6f}$$

The sizes of F_3 with $M_3 = L-1$ and $M_3 = 2(L-1)$ are respectively $(L-1) \times (L-1)$ and $2(L-1) \times 2(L-1)$.

(4) Trading Day Effect Model; (F_4, G_4)

The trading day effect model is an adjustment for the fact that there are a different number of i-th days of the week (i=1,...,7) per month for each successive month [9], [10], [17]. Trading day effeects have been treated by W.S. Cleveland and S.J. Devlin (1979), W.P. Cleveland and U.R. Graupe (1978), and S.C. Hillmer (1982). State space-Kalman filter regression on fixed regressors was suggested by Harvey and Phillips (1979). That effect is modeled by

$$\sum_{i=1}^{7} \beta_i(n) td_i(n) = \sum_{i=1}^{6} \beta_i(n)(td_i(n) - td_7(n)) = \sum_{i=1}^{6} \beta_i(n) td_i^*(n)8 \ . \qquad (3.7)$$

In (3.7), we apply the constraint $\sum_{i=1}^{7} \beta_i = 0$ so that $\beta_7 = -\sum_{i=1}^{6} \beta_i$. The non-perturbed difference constraint on the trading days is:

$$\beta_i(n) = \beta_i(n-1). \qquad (3.8)$$

Then the M_4, F_4, G_4 matrices are

$$M_4 = 6, \ F_4 = \begin{bmatrix} 1 & & & \\ & \cdot & & \\ & & \cdot & \\ & & & \cdot \\ & & & & 1 \end{bmatrix} \quad G_4 = \begin{bmatrix} 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{bmatrix} . \qquad (3.9)$$

The observation vector is a function of time, (to allow for a different number of i-th days/month each month),

$$H_4(n) = [td_1^*(n) \ldots td_6^*(n)]. \qquad (3.10)$$

For the general model including local polynominal and stochastic trends, local polynominal seasonal and trading day components, the state or noise vector w(n) and observation noise $\varepsilon(n)$ are assumed to be normally distributed with zero mean and diagonal covariance matrix

$$\begin{pmatrix} w(n) \\ \varepsilon(n) \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_1^2 & 0 & 0 & 0 \\ 0 & \tau_2^2 & 0 & 0 \\ 0 & 0 & \tau_3^2 & 0 \\ 0 & 0 & 0 & \sigma^2 \end{pmatrix} \right). \qquad (3.11)$$

The variances τ_1^2, τ_2^2, τ_3^2, 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, and $\beta_1 \ldots, \beta_6$ the fixed trading days regression coefficients. Relatively small values of the $\tau_1^2,\tau_2^2,\tau_3^2$ terms imply relatively strict adherence to the corresponding difference equation constraint.

Model class types fitted to data can be designated by a notation which reveals the constraint orders for the components. For example $M = (2, 2, 11)$, $M = (2, 0, 11, 6)$ respectively designate the model with trend constraint order 2, AR model order 2 and (monthly) seasonal order 11 and the model with trend constraint order 2, monthly seasonal order 11 and the trading 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 space structure of a model with $M = (2, 2, 11)$ is

$$x(n) = \left[\begin{array}{cc|cc|cccc} 2 & -1 & 0 & 0 & & 0 & & \\ 1 & 0 & 0 & 0 & & & & \\ \hline 0 & 0 & \alpha_1 & \alpha_2 & & 0 & & \\ 0 & 0 & 1 & 0 & & & & \\ \hline & & & & -1 & \cdots & \cdots & -1 \\ & 0 & & 0 & 1 & & & \vdots \\ & & & & & \ddots & & \vdots \\ & & & & & & 1 & 0 \end{array}\right] x(n-1) + \left[\begin{array}{ccc} 1 & 0 & 0 \\ 0 & 0 & 0 \\ \hline 0 & 1 & 0 \\ 0 & 0 & 0 \\ \hline 0 & 0 & 1 \\ \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots \\ 0 & 0 & 0 \end{array}\right] w(n) \qquad (3.12a)$$

with the state process noise vector $w(n)$ and the observation noise as given by (3.11). The observation equation which explains, $y(n)$, the observed data in terms of the contribution of the local polynomial trend, stationary AR process, seasonal and error components is

$$y(n) = [1\ 0\ 1\ 0\ 1\ 0 \ldots 0]\ x(n) + \varepsilon(n)\ . \qquad (3.12b)$$

If only the trend, $t(n)$, the trend plus AR, $t(n) + v(n)$, or only the seasonal component, $s(n)$, are to be considered, the observation equations $Hx(n)$ become respectively

$$\begin{aligned} Hx(n) &= [1\ 0\ \ldots]\ x(n) \\ Hx(n) &= [1\ 0\ 1\ 0\ \ldots]\ x(n) \qquad (3.12c) \\ Hx(n) &= [0\ 0\ 0\ 0\ 1\ \ldots]\ x(n). \end{aligned}$$

3.2 THE MINIMUM AIC PROCEDURE

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

$$\text{AIC}=-2 \log(\text{maximized likelihood}) + 2(\text{the number of fitted paramters})\ . \qquad (3.13)$$

In fitting state space models of the kind described in Section 3.1.2 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 and $\delta(M) = 1$ 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

$$L\ (\tau|,\overline{x}(o))= \prod_{n=2}^{N} f(y(n)|y(n-1),\ldots,y(1),\tau,\overline{x}(o))\ f(y(1)|(\tau,\overline{x}(o)). \qquad (3.14)$$

Under the Gaussian assumption, by exploiting the innovations representation that is achieved with the Kalman predictor

$$L(\tau,\bar{x}(0)) = \prod_{n=1}^{N}(2\pi v^2(n|n-1))^{-1/2} \exp\left\{\frac{-\nu(n)^2}{2v^2(n|n-1)}\right\} . \qquad (3.15)$$

In (3.15) $\nu(n)=(y(n) - Hx(n|n-1))$, $v^2(n|n-1)$ and $x(n|n-1)$ are respectively, the conditional mean and variance of $\nu(n)$, the innovations at time n, and the conditional mean of the state vector $x(n)$, given $y(n-1), \ldots y(1)$. Also in (3.15) $x(n|n-1)$ and $v^2(n|n-1)$, the one step ahead predictor of the state and the variance of the innovations, are obtained from

$$y(n|n-1) = H(n)\ x(n|n-1) \qquad (3.16)$$
$$v^2(n|n-1) = H(n)'V(n|n-1)H(n) + \sigma^2$$

where $V(n|n-1)$ 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 $\tau_1{}^2,\tau_3{}^2$ $(\tau_4{}^2 = 0)$. When the stationary AR component is included in the model, $\tau_2{}^2$ 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 procedure for each of the points in the $\tau_1^2,\tau_2^2,\tau_3^2$ space. The $\alpha_1,\ldots,\alpha_p$, $\tau_1^2,\tau_2^2,\tau_3^2$ parameters for which the AIC is smallest specifies the AIC criterion best model of the data.

Some comments on computational complexity and the discrete search in hyperparamter space procedure are appropriate here. The basic computation for the minimum AIC procedure, (3.13), is the computation of the maximized likelihood for particular classes of parametric models. With normally distributed correlated data, as is the modeling situation here, the likelihood computation requires the $O(N^3)$ complexity inversion of an NxN covariance matrix. Equation (3.14), 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 $\nu(i)$, $i=1,\ldots,N$. The orthogonalization achieved by the recursive Kalman predictor accounts for the $O(N)$ complexity.

With regard to the discrete hyperparameter search procedure: If the signal model is satisfactory, the influence of the priors and the hyperparameter values become decreasingly significant with increasing data length N. An indication of the insensitivity to the prior is the relative flatness of the likelihood in the vicinity of the location of the maximized likelihood in hyperparameter space.

Additional material relevant to 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.

The state space model is

$$\begin{aligned} x(n) &= F\,x(n-1) + Gw(n) \\ y(n) &= H\,x(n) + \varepsilon(n). \end{aligned} \tag{3.17}$$

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

$$\begin{array}{rl} \text{predicted} & x(n|n-1) \\ & y(n|n-1) \\ \text{filtered} & x(n|n) \\ & y(n|n) \\ \text{smoothed} & x(n|N) \\ & y(n|N) \end{array} \quad . \tag{3.18}$$

In the notation above, $x(n|n-1)$ and $y(n|n-1)$ denote the estimates of the state vector and the observation at time n given the past observations $y(n-1)$, ...,$y(1)$, $x(n|n)$ and $y(n|n)$ are estimates of the state and observations at time n given the current and past data $y(n)$, $y(n-1)$,...,$y(1)$ and $x(n|N)$ and $y(n|N)$ are estimates of the state and observation at time n given all the data $y(1)$, ...,$y(N)$. Meditch (1969) and Anderson and Moore (1979) show very satisfactory derivations for the quantities in (3.18). A first paper in the statistical literature on the Kalman predictor is Duncan and Horn (1972).

Given the initial vector $\overline{x}(0)$ the conditional means required in (3.15), (3.16) are obtained recursively:

$$x(n|n-1) = F\,x(n-1|n-1) \tag{3.19}$$
$$x(n|n) = x(n|n-1) + K(n)[y(n) - H(n)x(n|n-1)],$$

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

$$K(n) = V(n|n-1)\,H'(n)\,v^2(n|n-1)^{-1}\ . \tag{3.20}$$

The update equations for the variance of the state vector are

$$V(n|n-1) = F\,V(n-1|n-1)F' + GQG' \tag{3.21}$$
$$V(n|n) = (I-K(n)H(n))V(n-1|n).$$

The likelihood for each of the particular values of $\tau_1^2, \tau_2^2, \tau_3^2$ 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 is smoothed over the interval $n=N-1,\ldots,1$ by the formulas

$$x(n|N) = x(n|n) + A(n)(x(n+1|N) - x(n+1|n)) \tag{3.22a}$$
$$V(n|N) = V(n|n) + A(n)(V(n+1|N) - V(n+1|n))A(n)' \tag{3.22b}$$

where

$$A(n) = V(n|n)\,F'\,V(n+1|n)^{-1}\ . \tag{3.24c}$$

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

This is Bureau of Labor Statistics, male agricultural workers 20 years and older, data. Computational results are shown in Figure 1 for the models indicated in Table 1.

TABLE 1 - Trend and Seasonal Models Fitted to the BLSAGEMEN data

MODEL	M	T	$\hat{\sigma}^2$	AIC
A	(2 0 11)	(32 0 1)	2014.	1997.
B	(2 0 11)	(1 0 32)	656.	1830.
C	(2 2 11)	(16 1 16)	587.	1789.

Figures $1A_1$, $1B_1$ and C_1, show the original data and the fitted trends of the corresponding models. The seasonal components of the A and B models are in Figures $1A_2$ and $1B_2$ respectively. Figure $1C_2$ shows the local polynominal plus global autoregressive trend. Prediction results are shown in Figures $1A_3$, $1B_3$, $1C_3$, $1A_4$, $1B_4$ and $1C_4$. The model is fitted to the data $y(1),\dots,y(N)$, N=138. Prediction is done to estimate the data $y(N+1),\dots,y(N+M)$, N=138, M=24. Two kinds of predictions are considered. In one-step-ahead prediction, the quantity $y(n+1|n)$, $(n=N,N+1,\dots,N+M-1)$ is computed. In increasing horizon prediction, the quantity $y(N=i|N)$, $(i=1,\dots,M)$ is computed. In these and all subsequent illustrations showing predictions, the true value, the predicted value and the computed plus and minus one sigma confidence intervals are shown. Figures $1A_3$, $1B_3$ and $1C_3$ are the one-step-ahead predictions for the A, B and C models, respectively. Figures $1A_4$, $1A_4$, $1B_4$ and $1C_4$ are the increasing horizon predictions for the A, B and C models, respectively.

Figure $1A_1$ and $1B_1$ reveal that the local polynomial trend is smoother for larger values of τ_1^2. Figures $1A_2$ and $1B_2$ 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 though, the AIC best model, C, does achieve the best increasing horizon prediction performance and does exhibit the narrowest one sigma prediction interval.

EXAMPLE 2. BLSUEM 16-19

This is Bureau of Labor Statistics, unemployed males ages 16-19, data. Computational results are shown in Figure 2 for the models indicated in Table 2.

TABLE 2. Models Fitted to the BLSUEM 16-19 Data.

MODEL	M	N	T	$\hat{\sigma}^2$	AIC
A	(2 0 11)	180	(1 0 4)	628.7	2014.2
B	(2 2 11)	180	(64 1 16)	763.9	1952.5
C	(2 0 11)	48	(16 0 16)	--	--

This data was also analyzed by a different method in Hillmer and Tiao (1981). The trend and seasonal components of model A shown in Figures $2A_1$,A_2 are very similar in appearance to those shown in the Hillmer-Tiao analysis. This is not the AIC best M=(2 0 11) model. The overall AIC best of model types M=(2 0 11), M=(2 2 11) considered in Table 2 is Model B, the M=(2 2 11) model. (Model C was fitted to a different data span than models A and B, so their AIC's can not be compared.) The original data, trend seasonal and autoregressive components are shown respectively in Figures $2B_1$, $2B_2$ and $2B_3$. The one step ahead and increasing horizon prediction performance of Models A and B are shown in Figures $2A_3$,$2A_4$ and $2B_4$,$2B_5$ 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 very much narrower than that of Model A. The models were computed on N=180 data points and predicted for M=24 data points. Some of the computational results for Model C are shown in Figures $2C_1$-$2C_4$. This model was computed on N=48 data points and predicted for M=24 data points.

EXAMPLE 3. CONHSN, N=156, Alternative Seasonal Models.

This is Census Bureau construction series, housing starts, data. Computational results are shown in Figure 3. They correspond to the models for the CONHSN data indicated in Table 3.

TABLE 3 - Trend and Seasonal Models Fitted to the CONHSN Data

MODEL	M	T	$\hat{\sigma}^2$	AIC
A	(2, 0, 11)	(16, 0, 16)	.301	76.85
B	(2, 0, 22)	(16, 0, 8192)	.287	68.25

Figures $2A_1$ and $2B_1$ show the trends of the two models to be very similar. The seasonal component shown in Figures $2A_2$ and $2B_2$ indicate that the M_2 = 2(L-1) L=12, model captures the appearance of the increasing seasonal component that is suggested by the data better than 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

MODEL	M	T	$\hat{\sigma}^2$	AIC
A	(2, 0, 11, 0)	T=(8, 0, 16, 0)	0.245	-429.32
B	(2, 0, 11, 6)	T=(8, 0, 16, 0)	0.241	-439.40

Figure $3A_1$ and $3B_1$ 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 $3A_2$ and $3B_2$ for the two different models are very similar. The trading day effect and trading day plus seasonal components for the trading day model are in Figures $3B_3$ and $3B_4$. The trading day effect appears to be miniscule. The superposition of the trading day effect on the seasonal component does reveal the irregularizing effect of the number of trading days on the seasonality. The trading day effect model is the AIC criterion best model.

5. SUMMARY AND DISCUSSION

A smoothness priors-Kalman filter-Akaike AIC criterion approach to the modeling of time series with trends and seaonalities was shown. In that approach, an observed time series is decomposed into additive local polynomial trend, globally stationary autoregressive, seasonal and observation error components. Those components are each characterized by stochastically perturbed difference equations. The perturbations are uncorrelated with zero means and unknown variances and are independent of each other. The difference equations take 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 uncorrelated sequence forcing term 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 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 of this AIC criterion best model is then smoothed using the 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 seasonality driving input 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 globally stationary autoregressive plus seasonal, model classes are shown. The latter is the overall AIC criterion best model. The one-step ahead prediction performances of the AIC best of both model classes are similar. On the basis of the one-sigma conficence interval width for the increasing horizon prediction and the actual prediction performance, the AIC best model, model C, is strongly preferred to Model B. The evidence is additionally suggestive. 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.

Schlicht (1981), suggested that the value of the smoothness tradeoff parameters could be determined in an ad-hoc manner. That is only true locally. The effect of a sufficiently large amount of data, N, is to wash out local effects of the prior uncertainties. In that case, the particular local value of the hyperparameter is not critical. The prediction performance evidence shows that Schlicht's observation is not true globally. (An additional study of the prediction of time series with trends and seasonalities is in Gersch and Kitagawa (1982).)

The BLSUEM 16-19 data was analyzed by Hillmer and Tiao (1982), using a different signal model analysis. As shown in that example, the trends obtained by that "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, with the accompanying relatively poor increasing horizon prediction performance.

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 is 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.

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 Box-Jenkins-Tiao ARIMA type modeling procedures. See for example Shiskin et al. (1967,1978) and Cleveland and Tiao (1976) for treatments of the X-11 procedure and Box and Jenkins (1970) and Hillmer et al. (1981, 1982) for treatment of the ARIMA procedures. An emphasis in the employment of the Census X-11 method is in achieving an appraisal of the current status or trend of an econometric time series. The X-11 procedures are subject to certain practical public data reporting constraints which 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 not an effective way of evaluating the statistical properties of those procedure.

A critically different technical step between the Box-Jenkins-Tiao (B-J-T) and our methodology is our use of the AIC statistic and the B-J-T 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. The distinguishing practical property of our procedure in comparison with the B-J-T procedure is that ours is essentially a semi-automatic extensive model alternative procedure. The B-J-T 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 telehpone data in Thompson and Tiao (1971), and Hillmer (1982). The Tiao-Thompson 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 BLSUEM16-19, N=48 data point series seems to support the interpretation of our procedure as a semi-automatic procedure even on short duration series. The small sample-large variability properties of the Q statistic does not lend itself 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 (1982) for AIC minimum one step-ahead and twelve-step-ahead modeling and prediction of time series with trends and seasonalities. That prediction performance analysis appears to transcend what has been considered for the Box-Jenkins-Tiao ARIMA model approach.

ACKNOWLEDGEMENT: The enthusiastic encouragement and stimulating conversations with Dr. David Findley are gratefully acknowledged.

6. References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle, in 2nd international symposium on information theory, B.N. Petrov and F. Csaki, eds., (Akademiai Kiado, Budapest),267-281.

Akaike, H. (1974). A new look at the statistical model identification, IEEE Transactions on Automatic Control, AC-19, 716-723.

Akaike, H. (1980). Likelihood and the Bayes procedure in Bayesian Statistics, J. M. Bernado, M. H. DeGroot, D. V. Lindley, and A. F. M. Smith, eds., University Press, Valencia, Spain, 141-166.

Akaike, H. (1980). Seasonal adjustment by a Bayesian modeling, Journal of Time Series Analysis, 1, 1-13.

Akaike, H., and Ishiguro, M. (1981). Comparative study of the X-11 and BAYSEA procedures of seasonal adjustment, Applied Time Series Analysis of Economic Data, ASA-CENSUS-NBER Conference Proceedings.

Anderson, B.D.O., and Moore, J.B. (1979). Optimal Filtering, Prentice Hall, New Jersey.

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

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, 1061-1069.

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. (1981). Modeling time series when calendar effects are present, presented at ASA-Census-NBER Conference on Applied Time Series Analysis.

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, JASA 67, 815-821.

Gersch, W., and Kitagawa, G. (1982). The prediction of time series with trends and seasonalities. Submitted to 21st IEEE, Conf. on Decision and Control.

Good, I.J. (1965). The Estimation of Probabilities. MIT Press, Cambridge, Mass.

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). Modeling and forecasting time series with trading day variation, in preparation.

Hillmer, S.C. and Tiao, G.C. (1982). An ARIMA based approach to seasonal adjustment, Journal of the American Statistical Association, 77, 63-70.

Hillmer, S.C., Bell, W.R., and Tiao, G.C. (1981). Modeling considerations in the seasonal adjustment of economic time series, in Applied Time Series Analysis of Economic Data, ASA-Census-NBER Conference Proceedings.

Kitagawa, G. (1981). A nonstationary time series model and its fitting by a recursive technique, Journal of Time Series Analysis, 2, 103-116.

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

Pierce, D.A. (1978). Seasonal adjustment when both deterministic and stochastic seasonality are present in Seasonal Analysis of Economic Time Series, ed. Arnold Zellner, U. S. Department of Commerce, Bureau of the Census, 242-269.

Shiller, R. (1973). A distributed lag estimator derived from smoothness priors, Econometrica, 41, 775-778.

Schlicht, E.A. (1981). A seasonal adjustment principle and a seasonal adjustment method derived from this principle, Journal of the American Statistical Association, 76, 374-378.

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.

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 Application of Statistics, ed. P.R. Krishnaiah, North Holland, 507-524.

Wahba, G. and Wold, S. (1975). A completely automatic French curve: Fitting spline functions by cross validation, Comm. in Statistics, 4, 1-17.

Whittaker, E.T. (1923). On a new method of graduation. Proceedings Edinborough Math. Society, 41, 63-75.

Young, P.C. and Jakeman, A.J. (1979). The estimation of input variables in stochastic dynamic systems, Centre for Resource and Environmental Studies; Australian National University, Report No.AS/R28.

Legends

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 $M = (2, 0, 11)$, $T = (32, 0, 1)$, $\hat{\sigma}^2 = 2014$, AIC = 1997

A_1 Original data and trend, A_2 Seasonal component, A_3 One step ahead predictions, A_4 Increasing horizon predictions.

B: Model $M = (2, 0, 11)$, $T = (1, 0, 32)$, $\hat{\sigma}^2 = 656$, AIC = 1830

B_1 Original data and trend, B_2 Seasonal component, B_3 One step ahead predictions, B_4 Increasing horizon predictions.

C: Model $M = (2, 2, 11)$, $T = (16, 1, 16)$, $\hat{\sigma}^2 = 587$, AIC = 1789

C_1 Original data and trend, C_2 Original data and trend plus AR component, C_3 One step ahead predictions, C_4 Increasing horizon predictions.

Figure 2: BLSUEM 16-19 Trend and seasonal components, predictions, true values and plus and minus one sigma confidence intervals.

A: Model $M=(1,0,11)$, $T=(1,0,4)$, $\hat{\sigma}^2=628.7$, AIC=2014.2, N=180, M=24, A_1: Original data and trend, A_2 seasonal component, A_3 One step-ahead predictions, A_4 Increasing horizon predictions.

B: Model $M=(2,2,11)$, $T=(64,1,16)$, $\hat{\sigma}^2=763.9$, AIC=1952.5, N=180, M=24, B_1: Original data and trend plus AR component, B_2: Seasonal component, B_3 AR component, B_4 One step ahead prediction, B_5: Increasing horizon prediction.

C: Model $M=(2,0,11)$, $T=(16,0,16)$, N=47, M=24, C_1 Original data and trend, C_2 Seasonal component, C_3 One step-ahead prediction, C_4 Increasing horizon prediction.

Figure 3: Construction Housing Starts North data, trend and seasonal components

A: Model $M = (2, 0, 11)$, $T = (16, 0, 16)$, $\hat{\sigma}^2 = 0.301$, AIC = 76.85

A_1 Original data and trend, A_2 Seasonal component.

B: Model $M = (2, 0, 22)$, $T = (16, 0, 8192)$, $\hat{\sigma}^2 = 287$, AIC = 68.25

B_1 Original data and trend, B_2 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

A_1 Original data and trend, A_2 Seasonal component, A_3 Innovations

B: Model $M = (2,0,11,6)$, $T = (8, 1, 16)$, $\hat{\sigma}^2 = 0.241$, AIC = -439.40

B_1 Original data and trend, B_2 Seasonal component, B_3 Trading Day effect, B_4 Trading day effect plus seasonal, B_5 Innovations.

FIGURE 1

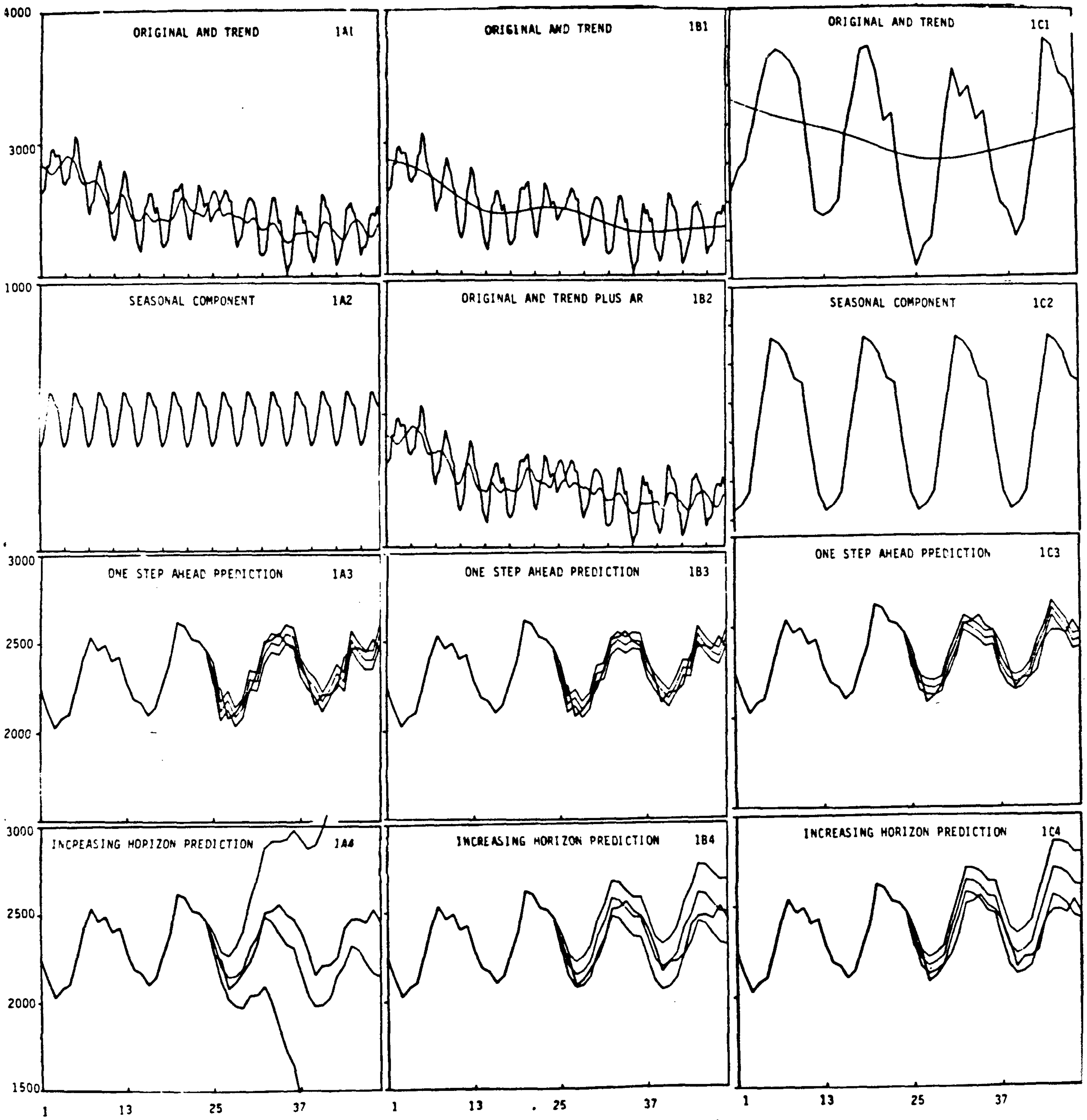
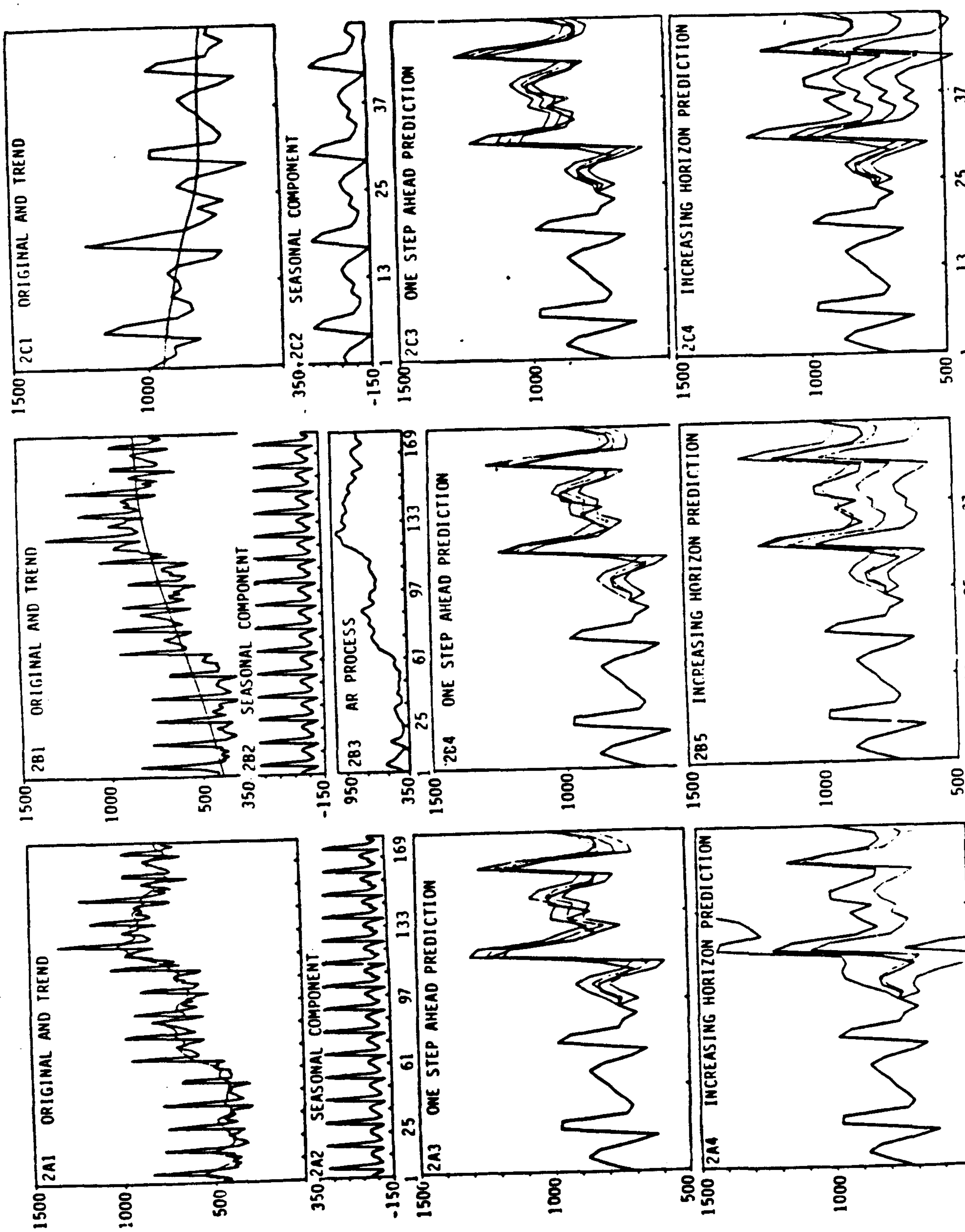
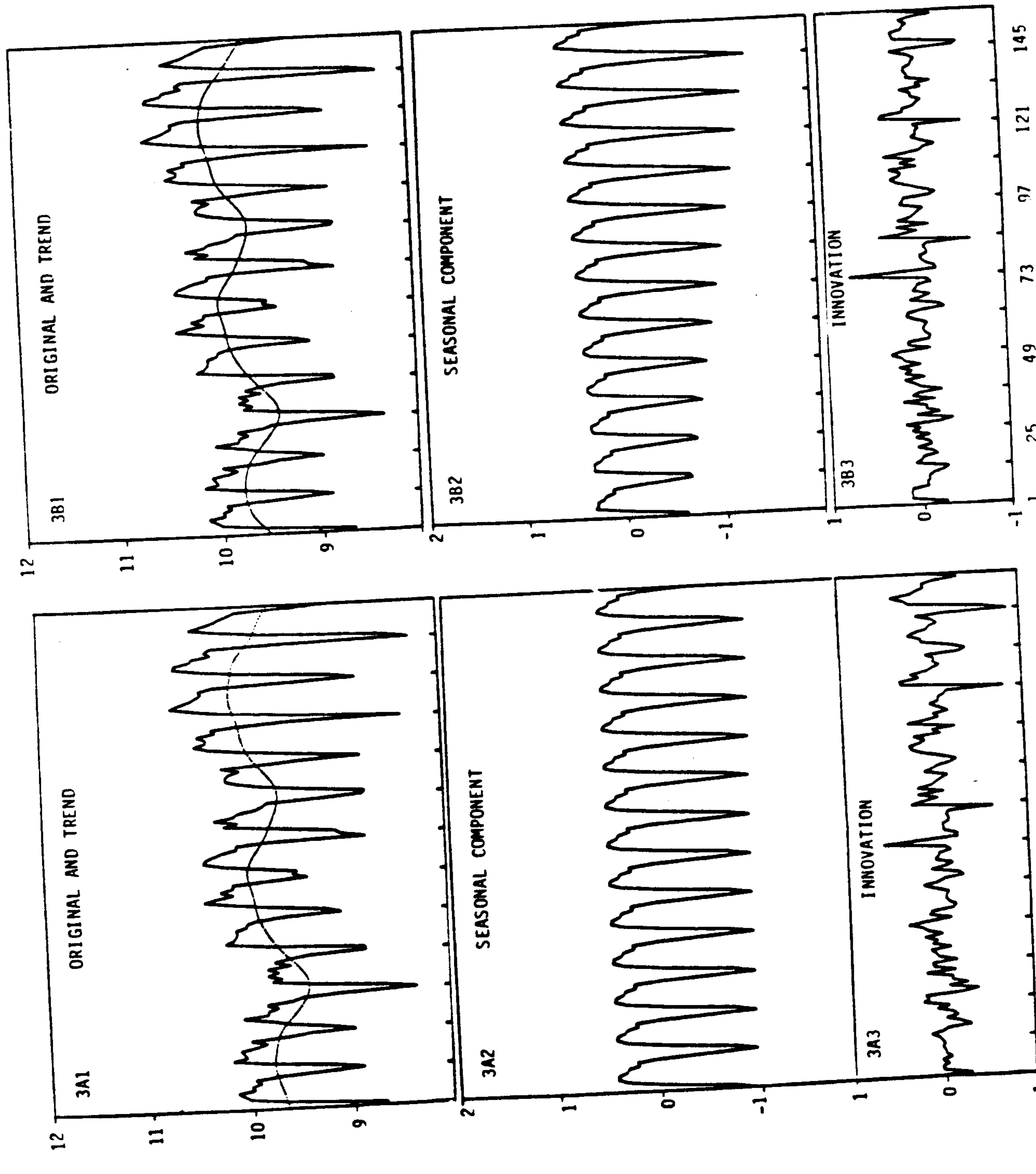

ORIGINAL AND TREND 1A1
ORIGINAL AND TREND 1B1
ORIGINAL AND TREND 1C1
SEASONAL COMPONENT 1A2
ORIGINAL AND TREND PLUS AR 1B2
SEASONAL COMPONENT 1C2
ONE STEP AHEAD PREDICTION 1A3
ONE STEP AHEAD PREDICTION 1B3
ONE STEP AHEAD PREDICTION 1C3
INCREASING HORIZON PREDICTION 1A4
INCREASING HORIZON PREDICTION 1B4
INCREASING HORIZON PREDICTION 1C4
4000
3000
1000
3000
2500
2000
3000
2500
2000
1500
1
13
25
37

FIGURE 2



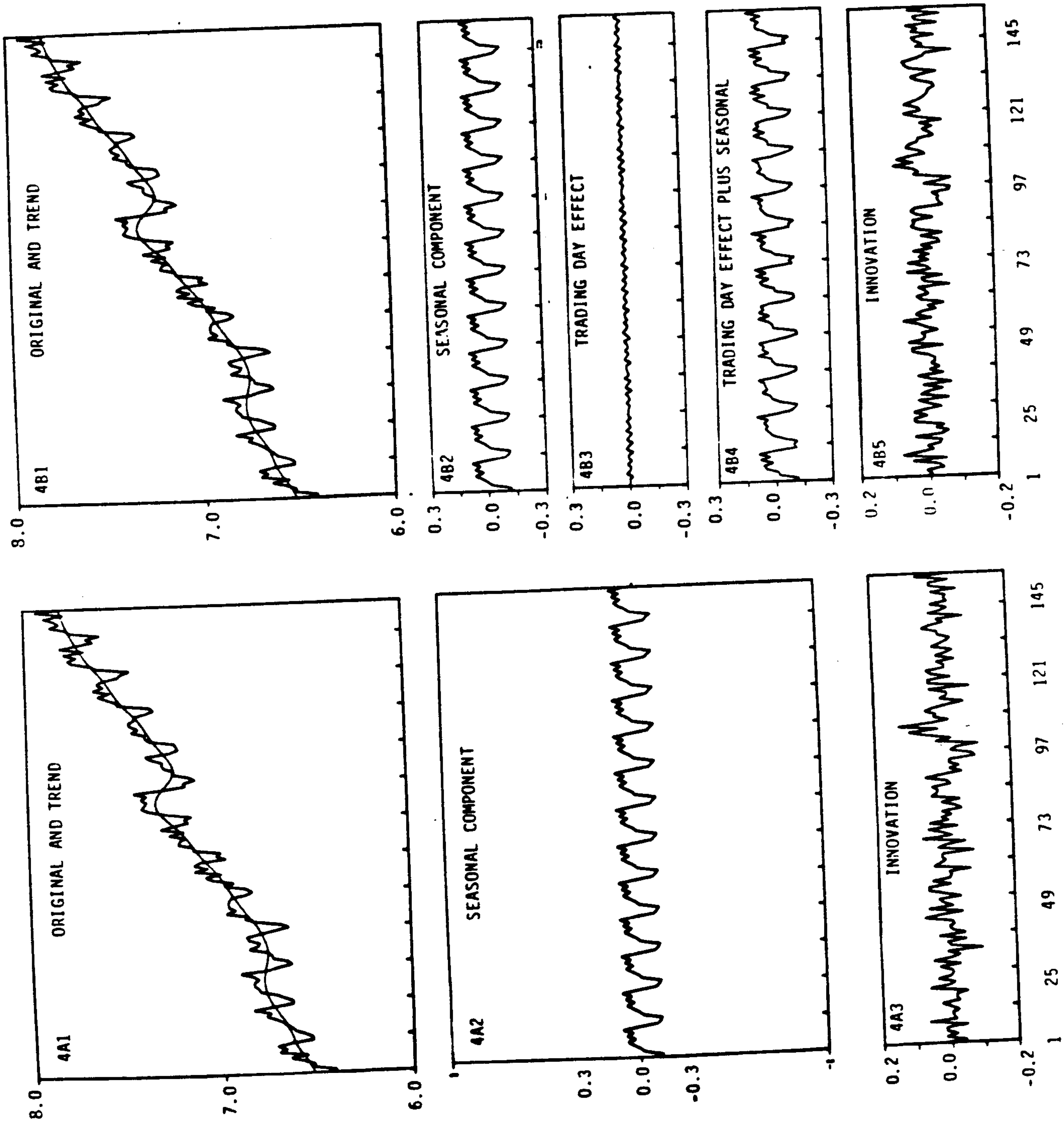
2A1 ORIGINAL AND TREND
2A2 SEASONAL COMPONENT
2A3 ONE STEP AHEAD PREDICTION
2A4 INCREASING HORIZON PREDICTION
2B1 ORIGINAL AND TREND
2B2 SEASONAL COMPONENT
2B3 AR PROCESS
2B4 ONE STEP AHEAD PREDICTION
2B5 INCREASING HORIZON PREDICTION
2C1 ORIGINAL AND TREND
2C2 SEASONAL COMPONENT
2C3 ONE STEP AHEAD PREDICTION
2C4 INCREASING HORIZON PREDICTION

FIGURE 3



ORIGINAL AND TREND
3A1
SEASONAL COMPONENT
3A2
INNOVATION
3A3
ORIGINAL AND TREND
3B1
SEASONAL COMPONENT
3B2
INNOVATION
3B3
12
11
10
9
2
1
0
-1
1
0
-1
1
25
49
73
97
121
145

FIGURE 4


4A1
ORIGINAL AND TREND
4A2
SEASONAL COMPONENT
4A3
INNOVATION
4B1
ORIGINAL AND TREND
4B2
SEASONAL COMPONENT
4B3
TRADING DAY EFFECT
4B4
TRADING DAY EFFECT PLUS SEASONAL
4B5
INNOVATION
8.0
7.0
6.0
0.3
0.0
-0.3
0.2
-0.2
1
25
49
73
97
121
145