ESTIMATING AND FORECASTING UNIVARIATE AND MULTIVARIATE TIME SERIES USING DYNAMIC LINEAR MODELS

a thesis
submitted in fulfillment
of the requirements for the degree
of
Doctor of Philosophy
at the
University of Waikato

William M. Bolstad

April 1990

Abstract

This thesis investigates the place of dynamic linear models in time series analysis. These models, which are based on the state space representation of the time series where the dynamic parameters are the state variables, are particularly useful because they are applicable to non-stationary (as well as stationary) time series. In this thesis some forecasting algorithms using dynamic linear models are developed for univariate and multivariate time series, and the effectiveness of these algorithms is investigated. The algorithms are applied to some examples in biometrics and economics. The thesis is based on original research papers published by the author, which are included as appendices 1-5.

Chapter one surveys the field of time series analysis in order to get a broad overview of the subject. There are two main goals in time series analysis: the search for the underlying pattern in the time series, and the prediction of future values of the series. The first leads to the the study of the time series in the frequency domain where a harmonic analysis of the time series decomposes it into a sum of sine waves at various frequencies. The second leads to the study of the time series in the time domain. Here the relationship between values a given distance apart is studied for each possible distance, and these relationships are used to predict future values. Chapter two develops the class of second order stationary stochastic processes in the frequency and time domains. The main tool of the frequency domain analysis, the spectral distribution of the process, and the main tool of the time domain analysis, the autocovariance function of the process, are seen to be Fourier transforms of each other. The effect of a time invariant linear filter on a second order stationary stochastic process is determined. It establishes the important results that if a stochastic process is Markovian, it must be the output of a first order linear system excited by Gaussian white noise. The Wiener-Hopf integral equation for the optimal (in terms of minimizing expected squared error) linear filter is developed using the principle of orthogonality. Chapter three introduces the state space representation of a linear dynamic system, develops the Kalman filter by using orthogonal projections, and establishes the recursive equations. Section 3 investigates the efficiency of the dynamic linear model applied to a linearly aggregated time series, and determines the efficiency of estimation using the aggregated time series relative to using the weighted sum of the component model estimators. While it is not surprising that the latter method is more efficient, the exact relative efficiency is developed here for the first time. Chapter four surveys the field of filtering in some non-Gaussian models. While the Kalman filter is still the optimal *linear* filter in the non-Gaussian case, it may be far from the optimal filter. Chapter five introduces multiprocess dynamic (regression) models to give a filter that is effective for short term forecasting in time series that are subject to abrupt, as well as evolutionary, changes in pattern, and may contain transients (outliers). In Section 3, the multivariate extension of the multiprocess dynamic linear model developed by the author is presented. The forecasting algorithm is used to forecast the quarterly inflation rate in the consumer price index, by using the inflation rates for the six main consumer price subindices, food, housing, household operations, apparel, transportation, and miscellaneous. This method was found to give an improvement over the univariate

dynamic linear model forecasts, which in turn are improvements over the naive model forecasts. In Section 4, the biased perturbation multiprocess dynamic linear model developed by the author is presented and used to analyze growth hormone level data in young steers that are subject to dietary stress. The growth hormone level data exhibited the characteristics of pulsatile data, that is exponential decay towards a base level with occasional jumps in level when a pulse of growth hormone is released. The model was found to give an effective method of analyzing pulsatile data that can be implemented in real time. In Section 5, the multiprocess dynamic generalized linear model developed by the author is presented. It is shown to be an effective method for estimating a time varying population proportion.

This thesis investigates dynamic linear time series models and has shown their place in the time series analysis framework. The relative efficiency of a dynamic linear model applied to a linearly aggregated time series has been determined. Estimation and forecasting algorithms based on dynamic linear time series models for univariate and multivariate time series have been developed and analyzed. They are effective for time series that contain outliers and exhibit occasional abrupt as well as evolutionary changes in pattern. These algorithms have been applied successfully to non-stationary time series in biology and economics.

Preface

The form of this thesis follows a model suggested to me by Prof. Roger Hosking. The original research is incorporated in separate (previously published) papers which are included as appendices. The main body of the thesis shows the position of dynamic linear models in the time series analysis framework, and relates my work to previous work in the field. This thesis does not incorporate, without acknowledgement, any material previously submitted for a degree or diploma in any University, and does not contain any material previously published or written by another person except with due reference, to the best of my knowledge and belief.

Acknowledgement

I wish to thank Prof. Les Foulds, Assoc. Prof. John Turner, and Prof. Roger Hosking for the valuable guidance and advice they each gave me when they were acting as my supervisor. The careful reading and comments from Prof. Foulds and Assoc. Prof. Turner have helped me create a more clearly focused and better presented thesis. Any errors that remain, are solely due to me. Also, I would like to thank my collegues for their interest and encouragement.

I wish to thank my Mother and my late Father for raising me in an supportive environment that greatly valued acquiring knowledge. Their belief in me allowed me to develop confidence in my own ability.

Last but not least, I want to dedicate this thesis to my wife Sylvie and my children Ben and Rachel. They have given me the reason to create.

Table of contents

Appendices:

[1] Bolstad, W. M., (1990), "The Efficiency of Dynamic Linear Model Estimators Applied to a Linearly Combined Time Series", Communications in Statistics: Theory and Methods, Vol. A19 No. 1, pp.

[2] Bolstad, W. M. (1986 A). "Harrison-Stevens Forecasting and the Multiprocess Dynamic Linear Model", The American Statistician, Vol. 40, pp. 129-135.

[3] Bolstad, W. M. (1986 B). "An Efficient Algorithm for Harrison-Stevens Forecasting Using the Multiprocess Multivariate Dynamic Linear Model", Communications in Statistics: Simulation and Computation, Vol. B15, pp. 819-828.

[4] Bolstad, W. M., (1988 A), "The Multiprocess Dynamic Linear Model with Biased Perturbations: A Real Time Model for Growth Hormone Level", Biometrika, Vol. 75 No. 4, pp. 685-692.

[5] Bolstad, W. M. (1988 B), "Estimation in the Multiprocess Dynamic Generalized Linear Model", Communications in Statistics: Theory and Methods, Vol. A17 No. 12, pp. 4179-4204.

List of Tables

1. Introduction To Time Series Models

1.1 Introduction

A time series is a sequence of observations of a variable which occur over time. Time series arise in the sciences of biology, meteorology, astronomy, geophysics, and others; in economics with series such as gross national product measurements and consumer price indices; in social science with series such as demographic records; and in engineering and technology with time series used in control systems. Some time series are deterministic, but following an unknown law; others exhibit stochastic properties; and some such as the signal-plus-noise model exhibit both deterministic and stochastic behaviour. The search for underlying patterns in time series and the prediction of future values goes back to antiquity. In the time of the ancient Greeks, observations on the positions of the planets were analyzed and the Ptolemaic system based on cycles and epicycles was created to explain the positions. This system was used to predict future astronomical positions. The need to explain the errors in the predictions that eventually emerged gave a major impetus to the development of science and mathematics. The search for an underlying pattern, and the prediction of future values are still the main goals of time series analysis. They lead to the study of time series in the frequency domain and time domain respectively. Predicting future values of the time series normally requires the assumption that the process generating the time series must not change over time. One important class of processes having this property is the class of second-order stationary processes. For this class, the mean for each member of the process is constant, and the covariance between any two members of the process depends only on the (time) lag between them. We now briefly discuss the historical development of time series analysis in both the frequency and time domains.

1.2 Time series in the frequency domain

The search for an underlying pattern gives rise to the study of time series in the frequency domain, where the time series is considered to be a sum of sinusoidal waves, and the goal is to determine the amplitude and phase at each frequency. Schuster (1898) developed the periodogram to search for "hidden periodicities" in the series and Fisher (1929) developed periodogram analysis, a method for testing the statistical significance of the suspected periods. Wiener (1930) developed generalized harmonic analysis, a general theory for spectral analysis of time series that exhibit stationarity in the sense that they have finite power. A wide variety of time series can be analyzed this way, including deterministic time series.

Khintchine (1934) developed the theory of second-order stationary stochastic processes in the frequency domain, where the stochastic process is represented by the Lebesgue-Stieltjes integral of a weighting function times a complex exponential function integrated over the frequency axis. The weight function, similar to complex random variables indexed by frequency, is called random spectral measure and it must be uncorrelated over disjoint intervals of frequencies for the process to be second-order stationary. The expansion of a second-order

stationary process in frequency space is into orthogonal components. Cramer (1942) discovered the spectral decomposition of second-order stationary stochastic processes. The time series is considered to be a partial realization of the stochastic process. The underlying pattern to be estimated from the time series is the spectral distribution (spectrum) of the process which is the expected value of the square of the random spectral measure as a function of the frequency. The spectral distribution can have both discrete and continuous components. Wold's decomposition theorem states that every zero mean second-order stationary process can be decomposed uniquely into the sum of a (linearly) purely deterministic component and a (linearly) purely non-deterministic component. The two components are uncorrelated. The purely deterministic component is contributed by the discrete part of the spectrum and its future development can be completely predicted from its past in a linear manner. The purely non-deterministic component is contributed by the continuous part of the spectrum and the linear relationship of its past history and future development and tends to zero over time. The sample analogue of the spectral distribution of the process is the periodogram of the time series, the squared norm of the finite Fourier transform of the time series. It is not a consistent estimate for the spectral distribution since its asymptotic variance is not zero. However, the periodogram estimates computed at different frequencies are asymptotically independent. Daniell (1946) introduced smoothed periodogram estimators which averaged the periodogram over a window of adjacent frequencies in order to get consistent estimates of the spectrum.

1.3 Time series in the time domain

The need for methods to predict future values of the time series led to the study of time series in the time domain. The assumption underlying most methods of forecasting is that the future behavior of the process generating the time series will be similar to its past behavior. Members of the class of second-order stationary processes satisfy this assumption. Although not second-order stationary themselves, processes whose derivatives (differences) are second-order stationary processes also satisfy this assumption. The traditional tools for studying how these processes evolve in time are the autocorrelation function and the partial autocorrelation function of the process. A class of processes which can be made second-order stationary by putting suitable restrictions on the parameters of the process is the general linear process, where the process is the output of a realizable time-invariant linear filter with white noise input. That is, the process is a fixed linear sum (integral) of past values of the input process which consists of independent identically distributed random variables. All purely non-deterministic second-order stationary stochastic processes are of this form.

When a discrete-time linear process is represented as the output of a time-invariant linear filter, it is called the *moving average* representation of the process. There is another representation for discrete-time processes called the *autoregressive* representation, where the process is regressed on past values of the process plus a random shock. The class of discrete-time linear processes is very large, as there are an infinite number of coefficients in the impulse response function which describes the moving average representation. Discrete-time general linear processes can be either stationary or non-stationary. The principle of parsimony suggests that

this class of models has too many parameters and would overfit a time series. That is, it would fit the observed series very well, but give poor forecasts compared to a model containing fewer parameters. Autoregressive-moving average (ARMA) processes are a subset of second-order stationary general linear processes which have a finite number of parameters. Every purely non-deterministic second-order stationary stochastic process can be approximated to any desired accuracy by an ARMA process. This corresponds to approximating the impulse response function of the process by a rational function. Autoregressive-integrated-moving average (ARIMA) processes are a subset of non-stationary general linear processes that have a finite number of parameters.

The Box-Jenkins method for fitting and forecasting time series using ARIMA models is a comprehensive development in the traditional time domain mode. This method requires the time series to be the realization of a stochastic process which is either stationary or can be made stationary by appropriate differencing.

A more general problem than forecasting is the problem of estimating the signal in the presence of noise. It is more general because it is desired to estimate the true underlying signal in the past, present, as well as the future, based on observations of the signal plus noise up to the present. This problem has been considered by control system engineers. Their goal was to develop a filter through which they could pass the noisy signal and extract the signal as efficiently as possible. The Weiner-Hopf integral equation gives the filter having minimum mean-squared prediction error, however it is difficult to solve except in special situations. These difficulties were not overcome until later when Kalman developed a filter using the state-space representation of a linear dynamic system.

1.4 State-space models for time series

Markovian (state-space) models are a relatively recent formulation of time series models in the time domain. With these models, the process is described in terms of a vector of unobserved auxilliary variables called *state variables*. Most of the early work on these models was done by engineers designing control systems for missiles. For these applications, the state-vector represents some real, but not directly measureable quantities. In these models the signal itself is not measured directly, rather a linear function of it is measured subject to a white noise measurement error.

A linear dynamic system is a system of linear difference or differential equations excited by an input process for discrete or continuous time processes respectively. If the input process is deterministic, the output process is deterministic, and if the input process is a stochastic process, so is the output process. The state-vector for a linear dynamic system contains the least amount of information which one must know about the past history of the output process which, taken together with the future input values, is sufficient to determine the output process at any future time. Loosely speaking, the state-vector can be thought of as boundary values of the process which are passed from the past through the present into the future in order that the future development of the process, given the present state, depends only on future input values. Zadeh and Doseur (1963) described the state as a parameterization of the input-output

space of a linear dynamic system in such a way that the output process from time t_0 depends only on the state at time t_0 and the input process after time t_0. This definition applies to both deterministic and stochastic processes.

In this thesis, we consider a stochastic process which is the output of a linear dynamic system with white noise input process. The output of a higher order linear difference or differential equation excited by white noise can be reformulated as the output of a system of first-order linear difference or differential equations excited by white noise in a higher dimensional space. The state-vector consists of the boundary values of the process which are passed from the past through the present in order that given the present state, the future development of the process is Markovian, i.e., it does not depend on the past values of the process. The state-vector could be considered to be an implicit parameterization of the input-output space of the process with no physical meaning. The traditional autoregressive (AR), moving-average (MA), and autoregressive-moving-average (ARMA) models can be put into state-space form and here the state-vectors are "minimal dimensional" Markovian models, see Aoki (1987). Other non-stationary processes as well as integrated-autoregressive-moving-average (ARIMA) models can be formulated in this way.

1.5 The Kalman filter

In Chapter 3, we discuss a more recent approach to linear filtering introduced by Kalman (1960), which is based on the state-space representation of a linear dynamic system. The problem addressed is to find the best (in terms of minimizing the expected squared error) estimate of the state at time $t+s$ based on observations up to time t. When $s<0$, $s=0$, or $s>0$ the estimation problem is called *smoothing*, *filtering*, or *predicting* respectively. Kalman's approach is general enough to include all these estimation problems. Also of great importance is that it is not restricted to stationary processes. In Kalman's method, the search for the optimal linear filter is approached from the point of view of conditional distributions and expectations found by orthogonal projections. The coefficients of the optimal linear filter are obtained recursively and can be done in real time. Kalman also shows that the dual of the optimal estimation problem is the noise-free optimal regulator problem. He showed that the optimal linear filter is actually the optimal filter for the case where the perturbation and observation distributions are both normal. Ho and Lee (1964) developed the Kalman filter under those normal assumptions by using a posterior Bayesian approach.

Duncan and Horn (1972) introduced the Kalman filter into the statistical literature by relating the dynamic linear model to random-β regression theory using the time-varying parameters as the state variables. They showed by an extension of the Gauss-Markov theorem that the Kalman filter estimates are minimum mean-squared linear estimators as well as being the minimum variance unbiased linear estimators. Meinhold and Singpurwalla (1983) gave a clear exposition of the Kalman filter as a sequential Bayesian estimation procedure in a time-varying linear regression model. Bolstad (1990) investigated the efficiency of dynamic linear model estimators applied to linearly combined time series.

In Chapter 4 we discuss some non-Gaussian state-space models. Masrieliez and Martin (1977) developed robust Bayesian estimates for a state-space model where either the state noise is Gaussian and the observation noise distribution has heavier tails than the normal distribution, or vice-versa. Smith and Miller (1986) developed a model for an exponential observation distribution and censored data. Kitagawa (1987) developed state-space models for non-stationary time series where the non-normal observation and perturbation noise distributions are approximated by piecewise linear functions and the recursive updating is done numerically. Martin and Raftery (1987) noted that the computational effort required precludes real time applications of this method for multi-dimensional state-vectors. West, Harrison, and Migon (1985) developed the dynamic generalized linear model where the parameters change over time according to a linear pattern, and the observations are from a one-dimensional exponential family distribution (not necessarily normal) on a non-linear scale. The recursions they developed for updating the estimates have the same form as the Kalman filter. A "guide" relationship is used to choose the conjugate prior, hence the estimates are only approximations.

1.6 Multiprocess dynamic models and applications

In Chapter 5, we discuss multiprocess dynamic regression models. These are appropriate for modelling time series which contain outliers and are subject to abrupt, as well as evolutionary changes in pattern. The multiprocess dynamic linear model was introduced by Harrison and Stevens (1971). They assume the perturbation distribution to be a mixture of distributions, thus allowing for the various types of changes. Harrison-Stevens forecasting, sometimes called Bayesian forecasting or the multiprocess Kalman filter, was shown to be an effective forecasting tool for non-stationary time series. It responds quickly to changes in pattern, yet is robust against outliers. Bolstad (1986 A) analyzed the Harrison-Stevens forecasting algorithm and provided a bibliography of previous applications. Bolstad (1986 B) has developed a computationally efficient algorithm for this method for multivariate time series having contemporaneously correlated errors. Inflation data based on consumer price index for the consumer price index groups is analyzed using this algorithm. It is shown to give improved forecasts in this very noisy time series.

Bolstad (1988 A) introduced the use of biased perturbations in the dynamic linear model in order to develop a real-time model for growth-hormone level data which exhibits the pattern of pulsatile release and exponential decay. The biased perturbation distribution allows the detection of a pulse on the subsequent observation, and incorporates it into future forecasts. Data showing the growth-hormone level of young cattle subject to varying degrees of dietary stress was analyzed using this algorithm. It was shown to be an effective model for this data, and the mean smoothed posterior pulse probability and mean smoothed posterior growth-hormone level estimate are useful tools for showing the treatment effect on the daily cycle of pulsatile activity rate and growth-hormone level.

Bolstad (1988 B) developed the multiprocess version of the dynamic generalized linear model. This model permits quick response to underlying changes in the time series while being robust against outliers. It was shown to be an effective forecasting tool for estimating a

dynamic proportion.

1.7 Other applications

Filters developed for linear dynamic systems have also been used as a fault detection procedure in control systems where abrupt shifts in the series indicate a degradation or breakdown in some instrument. Willsky (1976) examined the statistical performance of failure detection methods with respect to both speed of response to failure, and robustness to normal operations. Smith and West (1983) and Smith et al. (1983) used the multiprocess dynamic linear model when monitoring renal transplant data in order to detect an abrupt shift (organ rejection) as quickly as possible.

Dynamic linear estimation algorithms can also be used in situations where the parameters are fixed, but there is a time dependence in the error structure. In this case their main advantage is their computational efficiency. Harvey and Phillips (1979) found the maximum likelihood estimates for a regression model with autoregressive-moving-average disturbances. Harville (1980) used a similar algorithm to find the best linear unbiased predictors of the outcomes of National Football League gridiron games. Sallas and Harville (1981) developed a method for finding the optimal estimates in a mixed linear model by first using the Kalman filter to estimate the fixed and realized values of the random effects, and then taking limits to obtain the mixed model estimates. Bolstad (1987) developed a method of estimating the parameters of the seemingly unrelated regression model which is based on the use of an efficient recursive algorithm for estimating the state-vector in the presence of known contemporaneous covariances.

1.8 Summary and conclusions

In this chapter we have briefly surveyed the field of time series analysis, in both the frequency and time domains. We have noted that ARMA and ARIMA time series models can be put into the more recently formulated state-space form. The Kalman filter is the general linear filtering technique based on this form, and it is the optimal filter in the mean-squared error sense under the assumption of normality. Some alternative state-space models are noted which are applicable in non-Gaussian situations. Multiprocess models are introduced to model non-Gaussian estimator and perturbations distributions. These models will be shown to be very effective for short-term forecasting in time series that contain outliers, and are subject to abrupt as well as evolutionary changes.

2. Second-Order Stationary Stochastic Processes

2.1 Introduction

In this chapter we review the historical development of time series analysis of processes whose evolution contains a random or stochastic element. These are called stochastic processes and are modelled by a set of jointly distributed random variables indexed by time. A time series is considered to be a partial realization of the process, that is, a time indexed set of observed values of the process over a period of time. The stochastic processes that satisfy the intuitive stationarity requirement that "the mechanism generating the time series does not change with time" are defined and are used in time series analysis in both the time and frequency domains.

Time-invariant linear filters which transform one stochastic process into another are introduced and studied in the frequency and time domains. The relationship between the input process and the output process of a realizable linear filter (where the output process only depends on past and present values of the input process) is often defined by a linear differential (difference) equation. If the filter is time-invariant, the coefficients of the linear differential equation are constant. Stochastic processes where the future development of the process, given the present and past history of the process, depends only on the present are called Markov processes. They may or may not be stationary. Markov processes can be characterized as the output of a first-order linear differential (difference) equation when the input process is white noise. If the coefficient is a positive constant, the process is second-order stationary.

The general estimation problem is introduced and the Wiener-Hopf equation for the optimal linear filter is developed.

2.2 Mathematical model

The mathematical model for a stochastic process is a set of random variables $\{ Y(t) \}$ indexed by time $t \in T$ where the index set T is usually the set of real numbers (continuous time) or the set of integers (discrete-time). The set of all possible realizations of the stochastic process is called the *ensemble*. Kolmogorov (1933) showed that the probabilistic mechanism that determines the distribution of the ensemble is contained in

$$f_{t_0,\ldots,t_n}(y(t_0), \ldots, y(t_n)) \text{ , where } t_i \in T \text{ for } i = 0,\ldots,n \text{ and } 0 < n < \infty,$$

the set of joint probability (density) functions for every possible finite subset of the random variables.

A time series is a single (partial) realization of the process; a set of observed values of the process, $y(t)$ such that $t \in S$, for some set $S \subseteq T$. Time series analysis includes two types of inferences; inferences about the ensemble, and inferences about the realization, such as forecasts. Inferences about the ensemble, made from a single realization of the process require that the single realization be representative of the ensemble in some sense. Certainly, in order

for the past values of a time series to be useful for making inferences about the ensemble, the behavior of the underlying stochastic process must be the same in the future as it has been in the past. This means that the underlying mechanism of the stochastic process does not change as time progresses. A stochastic process is said to be strictly stationary if and only if the joint random variables: $Y(t_0),\ldots,Y(t_n)$ and the joint random variables: $Y(t_0+h),\ldots,Y(t_n+h)$ are identically distributed for all possible sets $t_0, \cdots, t_n$ and all h. That is, the joint distribution of any subset of the random variables is invariant under a time shift. Stationary stochastic processes are models for physical phenomena that are in steady-state operation. In practice, it is difficult to establish that a stochastic process is strictly stationary. Fortunately, a weaker property than strict stationarity is sufficient for our purposes.

The first two moments contain the most important information about the distribution of a single random variable, namely its location and scale; although, they do not generally specify the distribution completely. Analagously for a stochastic process the first moment function and the second joint moment function contain the most important information about the joint distribution. Define the mean function and the covariance kernel to be

$$\mu(t) = E(Y(t)) \quad \text{and} \quad K(s,t) = E[(Y(s)-\mu(s))\times(Y(t)-\mu(t))]$$

respectively. A stochastic process is said to be *covariance stationary* if and only if the mean function is a constant and the covariance kernel is a function only of the absolute difference between the times, not on the actual times themselves. That is,

$$\mu(t) = \mu \quad \text{and} \quad K(s,t) = \Gamma(|t-s|).$$

Covariance stationary processes are also called *weakly stationary* or *second-order stationary* processes. A strictly stationary process is always covariance stationary. However, the converse is not generally true. For normal (or Gaussian) processes the first two moment functions are sufficient to determine the joint normal distribution. Hence covariance stationary Gaussian processes are strictly stationary. If the covariance between any two of the random variables from a normal process is zero, the random variables are independent. Also, any set of linear combinations of multivariate normal random variables are jointly normally distributed.

In order to make inferences about the first two moment functions of a stochastic process from the time series, the ensemble averages must be estimated by the corresponding time averages over a single realization. The first *ergodic* theorem, which showed these to be consistent estimators when the stochastic process is strictly stationary, was proved by Birkoff (1931). Parzen (1958) proved that necessary and sufficient conditions for the sample mean of a stochastic process to be ergodic are that the correlation between the sample mean and the last observation approach zero as the sample size tends to infinity. This implies that the mean of a covariance stationary stochastic process is ergodic if $\lim_{\tau\to\infty} \Gamma(\tau) = 0$.

Every second-order stationary stochastic process with zero mean has the spectral representation

$$Y(t) = \int_{-\infty}^{\infty} e^{i\lambda t}\, Z(d\lambda)$$

where the random spectral measure, $Z(d\lambda)$, is a complex-valued random set function, similar to complex random variables. In order that the process be real-valued,

$$Z(-A) = \overline{Z(A)},$$

the complex conjugate of $Z(A)$, for any measureable set A of real numbers on the frequency axis. The covariance

$$\begin{aligned} E(Y(t)\times Y(t-\tau)) &= E[\int_{-\infty}^{\infty} e^{i\lambda_1 t} Z(d\lambda_1) \int_{-\infty}^{\infty} e^{-i\lambda_2(t-\tau)} \overline{Z(d\lambda_2)}] \\ &= \int_{-\infty}^{\infty} e^{i\lambda_2\tau} \int_{-\infty}^{\infty} e^{i(\lambda_1-\lambda_2)t} E(Z(d\lambda_1)\times \overline{Z(d\lambda_2)}. \end{aligned}$$

Since the process is weakly stationary, the covariance cannot depend on t. Hence, the random spectral measure over disjoint sets must be uncorrelated. Furthermore, the autocovariance function of the process can be written as

$$\Gamma(\tau) = \int_{-\infty}^{\infty} e^{i\lambda\tau} F(d\lambda),$$

where

$$F(d\lambda) = E|Z(d\lambda)|^2 = E[Z(d\lambda)\times \overline{Z(d\lambda)}]$$

is called the spectral distribution of the process. The reader is referred to Koopmans (1974 p. 61) for a proof of this. The spectral distribution has all the properties of a probability distribution except that its total power $\int_{-\infty}^{\infty} F(d\lambda)$ is finite, but not necessarily equal to one.

Wold (1938) showed that any stationary process can be expressed as

$$Y(t) = Y_d(t) + Y_c(t)$$

where the two uncorrelated processes $Y_d(t)$ and $Y_c(t)$ are called purely deterministic and purely non-deterministic respectively. The process $Y_d(t)$ can be (linearly) predicted from its past with no error and the process $Y_c(t)$ is ergodic. Thus the random spectral measure can be decomposed into

$$Z(d\lambda) = Z_d(d\lambda) + Z_c(d\lambda)$$

where the discrete and continuous components are uncorrelated. Hence

$$F(d\lambda) = E|Z(d\lambda)|^2 = E|Z_d(d\lambda)|^2 + E|Z_c(d\lambda)|^2 = f_d(\lambda) + f_c(\lambda)d\lambda,$$

where f_d is the discrete spectral function and f_c is the spectral density function. When the spectrum has only a discrete component, the stochastic process is called *almost periodic* and has spectral representation

$$Y(t) = \sum_{k=-\infty}^{\infty} e^{i\lambda_k t} Z_d(\lambda_k).$$

If the spectrum has only a continuous component, the stochastic process is called a *noise* process and has spectral representation

$$Y(t) = \int_{-\infty}^{\infty} e^{i\lambda t} Z_c(d\lambda).$$

If the spectrum has both components, then

$$Y(t) = Y_d(t) + Y_c(t),$$

where $Y_d(t)$ is the purely deterministic component and $Y_c(t)$ is the purely non-deterministic component.

A process is called white noise if the random variables at different times are uncorrelated with each other. First let us consider a noise process with autocovariance function

$$\Gamma(\tau) = \Gamma(0)\,\delta(\tau),$$

where $\delta(\tau)$ is the Dirac delta function. This means that the white noise process consists of instantaneously uncorrelated random variables having a finite variance. In this case the spectral density is given by

$$f_c(\lambda) = (2\pi)^{-1}\int_{-\infty}^{\infty} e^{-i\lambda t}\Gamma(t)dt = 0.$$

Thus, its power at every frequency is zero. Next we consider a noise process having constant spectral density $f_c(\lambda) = c$. For such a process

$$\Gamma(\tau) = \int_{-\infty}^{\infty} e^{i\lambda\tau} c\, d\lambda \quad = \infty \qquad \text{if} \quad \tau = 0, \text{ and}$$

$$= 0 \qquad \text{if} \quad \tau \neq 0.$$

Hence it has infinite variance. Thus continuous time white noise has either a spectrum equal to 0 (zero power), or an infinite variance (infinite instantaneous power), so it cannot exist as a second-order stationary stochastic process except as a limit. Nevertheless, it can be used quite satisfactorily as the input to a linear filter.

2.3 Time-invariant linear filters

A time-invariant linear filter is a transformation satisfying the properties of (*i*) scale preservation, (*ii*) superposition, and (*iii*) time invariance, which transforms a stochastic process $W(t)$ into another stochastic process $V(t)$. It is denoted by

$$L(W(t)) = V(t).$$

A consequence of these three properties is that the linear filter transforms a complex exponential function having frequency λ into a multiple of itself. Thus

$$L(e^{i\lambda t}) = B(\lambda)e^{i\lambda t},$$

where $B(\lambda)$ is called the transfer function of the filter. Hence for any weakly stationary stochastic process

$$V(t) = L(W(t)) = L\left(\int_{-\infty}^{\infty} e^{i\lambda t} Z_W(d\lambda)\right) = \int_{-\infty}^{\infty} L(e^{i\lambda t})\, Z_W(d\lambda) = \int_{-\infty}^{\infty} B(\lambda)e^{i\lambda t} Z_W(d\lambda).$$

Clearly the spectral representation of the output process is

$$V(t) = \int_{-\infty}^{\infty} e^{i\lambda t} Z_V(d\lambda),$$

where the random spectral measure

$$Z_V(d\lambda) = B(\lambda)Z_W(d\lambda).$$

Hence the spectral distribution of $V(t)$ is given by

$$F_V(d\lambda) = |B(\lambda)|^2 F_W(d\lambda).$$

The output process has finite power (variance) provided the matching condition

$$\int_{-\infty}^{\infty} |B(\lambda)|^2 F_W(d\lambda) < \infty$$

is satisfied.

Let the impulse response function of the linear filter be the Fourier transform of $B(\lambda)$,

$$h(u) = (2\pi)^{-1}\int_{-\infty}^{\infty} B(\lambda)e^{i\lambda u}d\lambda.$$

Then

$$B(\lambda) = \int_{-\infty}^{\infty} h(u)e^{-i\lambda u}du,$$

so

$$L(W(t)) = \int_{-\infty}^{\infty} B(\lambda)e^{i\lambda t}Z_W(d\lambda) = \int_{-\infty}^{\infty} h(u)[\int_{-\infty}^{\infty} e^{i\lambda(t-u)}Z_W(d\lambda)]du = \int_{-\infty}^{\infty} h(u)W(t-u)du.$$

The output process is the convolution of the impulse response function and the process. In general, the output process is a function of the entire history of the process, past present and future. If the output process is to be a function of the past and present only, then

$$h(u) = 0 \quad \text{for } u < 0.$$

In that case, the linear filter is called *realizable*. For realizable linear filters

$$L(W(t)) = \int_0^{\infty} h(u)W(t-u)du,$$

the present value of the output process depends on the input process only up to and including the present time.

If the input process $W(t)$ is white noise, the output process $V(t)$ will be second-order stationary provided

$$\int_{-\infty}^{\infty} |B(\lambda)|^2 < \infty.$$

Thus, even though second-order stationary continous time white noise exists only as a limit, it can be used perfectly satisfactorily as the input process into a linear filter.

2.4 Time-invariant linear systems

A time-invariant linear system is an input-output relationship that is a realizable linear filter. Many linear systems are defined by linear differential equations. If the coefficients are constants, the linear system is time-invariant. Let $W(t)$ and $V(t)$ be the input and output

processes respectively. Let a first-order time-invariant linear system be governed by the equation

$$(D+a)V(t) = W(t)$$

where $a > 0$. The solution is given by

$$V(t) = \int_{-\infty}^{t} e^{-a(t-s)}W(s)ds = \int_{0}^{\infty} e^{-au}W(t-u)du.$$

This is the output of a realizable linear filter with impulse response function

$$h(u) = 0 \quad \text{for} \quad u < 0$$

$$= e^{-au} \quad \text{for} \quad u \geq 0.$$

The differential equation governing the general $p'th$ order time-invariant linear system is

$$\Phi(D)V(t) = W(t)$$

where $\Phi(D)$ is a $p'th$ degree polynomial in the differential operator D. If the polynomial $\Phi(D) = (D+a_1)\times \cdots \times (D+a_p)$ where $a_i > 0$ for all i then the solution is

$$V(t) = \int_{-\infty}^{t} e^{-a_1(t-s_1)} \int_{-\infty}^{s_2} e^{-a_2(s_1-s_2)} \int_{-\infty}^{s_3} \cdots \int_{-\infty}^{s_{p-1}} e^{-a_p(s_{p-1}-s_p)}W(s_p)ds_p \ \ldots \ ds_1.$$

Clearly $V(t)$ is the output when $W(t)$ is passed through of a sequence of realizable linear filters with respective impulse response functions $h_i(u) = e^{-a_i u}$ for $u > 0$ and $i = 1,\ldots,p$. If the input process $W(t)$ is white noise, the output process $V(t)$ is second-order stationary, since the filter clearly satisfies $\int_{-\infty}^{\infty} |B(\lambda)|^2 < \infty$. Hence second-order stationary processes can be modelled as the output of time-invariant linear system ($p'th$ order linear differential equation) with white noise input. Clearly if the input process is Gaussian white noise, the output process is also Gaussian.

2.5 Markov processes

If the conditional distribution of the process at some specific future time, given the past history of the process up to the present time, depends only on the value of the process at the present time, the process is said to have the *Markov* property. If a process is the output of a time-invariant first-order linear system excited by Gaussian white noise, then

$$Y(t) = \int_{-\infty}^{t} e^{-a(t-v)}W(v)dv = \int_{-\infty}^{s} e^{-a(t-v)}W(v)dv + \int_{s}^{t} e^{-a(t-v)}W(v)dv$$

$$= e^{-a(t-s)} \int_{-\infty}^{s} e^{-a(s-v)}W(v)dv + \int_{s}^{t} e^{-a(t-v)}W(v)dv.$$

Let $u = s - v$ in the first term. Then

$$Y(t) = e^{-a(t-s)} \int_{0}^{\infty} e^{-au}W(s-u)du + \int_{s}^{t} e^{-a(t-v)}W(v)dv$$

$$= e^{-a(t-s)}Y(s) + \int_{s}^{t} e^{-a(t-v)}W(v)dv.$$

The output process $Y(s)$ depends only on the white noise process $W(v)$ for times $v \leq s$. But Gaussian white noise is independent over disjoint intervals. Hence the white noise process at

times $v > s$ is independent of the process at all times before s, so the the output process $Y(t)$ satisfies the Markov property.

The converse is also true. Let

$$U(t) = Y(t) - (\Gamma(t-s)/\Gamma(0)) \times Y(s),$$

where $\Gamma(t)$ is the autocovariance function of the process. If a second-order stationary Gaussian process $Y(t)$ satisfies the Markov property, then the conditional distribution of $U(t)$, given the history of the process up to time $s < t$ depends only on $Y(s)$. Because the process is Gaussian, $U(t)$ and $Y(s)$ are independent since $E[U(t) \times Y(s)] = 0$. In a small interval $0 < t - s < \varepsilon$

$$\Gamma(t-s) = \Gamma(0) + \Gamma'(\delta) \times (t-s)$$

for some $0 < \delta < t - s$. Hence

$$U(t) = Y(t) - Y(s) - (\Gamma'(\delta)/\Gamma(0)) \times (t-s) \times Y(s),$$

so

$$W(t) = \lim_{s \to t} U(t)/(t-s) = (D+a)Y(t),$$

where $a = -\Gamma'(0)/\Gamma(0)$. $W(t)$ is a multiple of $U(t)$ which depends only on $Y(s)$, not previous values. Taking the limit as $s \to t$, the $W(t)$ process will consist of independent random variables. Thus $Y(t)$ is the output of a first-order linear system with Gaussian white noise input.

The autocovariance function of a second-order stationary Markov process is

$$\Gamma(\tau) = \int_{-\infty}^{t+\tau} \int_{-\infty}^{t} e^{-a(t+\tau-s_1)} e^{a(t-s_2)} E(W(s_1) \times W(s_2)) ds_1 ds_2.$$

Since $W(t)$ is Gaussian white noise with variance σ^2, $E[W(s_1) \times W(s_2)] = \sigma^2$ when $s_1 = s_2$ and 0 otherwise. Hence

$$\Gamma(\tau) = \sigma^2 e^{-a|\tau|}.$$

Conversely, if $Y(t)$ is a process with autocovariance function of the form

$$\Gamma(\tau) = \sigma^2 e^{-a|\tau|},$$

then let $W(t) = (D+a)Y(t)$. The spectral density of $Y(t)$ is

$$f_Y(\lambda) = a\pi^{-1} \sigma^2 (\lambda^2 + a^2)^{-1},$$

and the spectral density of $W(t)$ is

$$f_W(\lambda) = |i\lambda + a|^2 f_Y(\lambda) = a\pi^{-1}\sigma^2,$$

which makes $W(t)$ a white noise process. Also if $Y(t)$ is Gaussian, so is $W(t)$. A process with autocovariance function $\Gamma(\tau) = \sigma^2 e^{-a|\tau|}$ is defined to be a second-order Markov process by Brieman (1969 p. 289). It is important to realize that a second-order Markov process does not satisfy the Markov property. It merely satisfies the weaker property that the first two moments of the process at time t, given the history of the process up to time $s < t$, depends

only on the value of process at time s. This is sufficient to determine the distribution of the process at time t in the Gaussian case, but not in the general case.

It is important to realize that the output of a time-varying first-order linear system with Gaussian white noise input also satisfies the Markov property. If

$$(D+a(t))Y(t) = W(t),$$

then

$$Y(t) = e^{-A(t)}\int_{-\infty}^{t} e^{A(v)}W(v)dv = e^{-A(t)}\int_{-\infty}^{s} e^{A(v)}W(v)dv + e^{-A(t)}\int_{s}^{t} e^{A(v)}W(v)dv$$
$$= e^{-A(t)+A(s)}Y(s) + e^{-A(t)}\int_{s}^{t} e^{A(v)}W(v)dv,$$

where $A(t)$ is an antiderivative of $a(t)$. Since $W(t)$ is a Gaussian white noise process, the output process $Y(t)$ satisfies the Markov property.

Conversely, let $Y(t)$ be a time-varying Gaussian process satisfying the Markov property. Let

$$U(t) = Y(t) - (K(s,t)/K(t,t))\times Y(s)$$

where $K(s,t)$ is the covariance kernel of the process. In a small interval $0 < t-s < \varepsilon$,

$$K(s,t) = K(t,t) + K_s(\delta,t)\times(t-s) \quad \text{for some } 0 < \delta < t-s$$

where $K_s(s,t)$ is the partial derivative of the covariance kernel with respect to the first component. Hence

$$U(t) = Y(t) - Y(s) - (K_s(\delta,t)/K(t,t))\times(t-s)\times Y(s),$$

so

$$W(t) = \lim_{s\to t} U(t)/(t-s) = (D+a(t))Y(t),$$

where $a(t) = -K_s(t,t)/K(t,t)$. In the limit as $s\to t$, $W(t)$ will consist of Gaussian white noise.

2.6 The general estimation problem and the Wiener-Hopf filter

In this section we look at the general estimation problem. Here, inferences are to be made about the actual realization of the process, not the ensemble of all possible realizations. In general, the stochastic process need not be stationary. Thus, the mechanism generating the stochastic process may be changing with time, but it is necessary to know how it is changing.

There is an unobserved stochastic process $V(t)$. A related stochastic process $Y(t)$ has been observed up until time t. We wish to find the best estimate of the unobserved process $V(t+\tau)$ at some fixed time lag τ. When $\tau < 0$, $\tau = 0$, and $\tau > 0$ the problem is called smoothing, filtering, or prediction respectively. Often the relationship between the two processes is either that $Y(t)$ is the output process when $V(t)$ is passed through a linear filter, or that $Y(t)$ is an observation of $V(t)+N(t)$, where $N(t)$ is a noise process, or a combination of both, that is $Y(t)$ is the output of a linear filter plus noise. The criterion we shall use for the "best

estimate" is minimizing expected squared error. That is, $\hat{V}(t+\tau)$ is the function of $Y(t')$ for $t' \leq t$ that minimizes the expected squared error $E(V(t+\tau)-\hat{V}(t+\tau))^2$. The function that minimizes this is the conditional mean

$$\hat{V}(t+\tau) = E(V(t+\tau) | Y(t') \text{ for } t' \leq t).$$

In practice, it is not feasible to evaluate this conditional mean.

An alternative to this is to restrict ourselves to second-order stationary stochastic processes and linear estimators which are the output of realizable linear filters operating on the observed series. The best linear estimator is

$$\hat{V}(t+\tau) = \int_0^\infty h(u')Y(t-u')du',$$

where $h(u)$ is the impulse response function of the optimal realizable linear filter. $\hat{V}(t+\tau)$ is the point in the linear space spanned by the values of the process up to time t that is closest to $V(t+\tau)$. By the orthogonality principle $\hat{V}(t+\tau)-V(t+\tau)$ must be perpendicular to $Y(t-u)$ for all $u \geq 0$. Hence

$$E[(V(t+\tau)-\int_0^\infty h(u')Y(t-u')du') \times Y(t-u)] = 0.$$

The impulse response function of the optimal linear filter is the solution of the integral equation

$$\phi_\tau(u) = \int_0^\infty h(u')\Gamma(u-u')du',$$

where $\phi_\tau(u) = E[V(t+\tau) \times Y(t-u)]$. In general, this integral equation is too difficult to solve for the function $h(u')$. Similarly, for discrete-time processes the impulse response function of the optimal linear filter is the solution of the analagous equation

$$\phi_\tau(u) = \sum_0^\infty h(u')\Gamma(u-u').$$

An important special case is where we are trying to predict $Y(t+\tau)$ at some future instant $t+\tau$. Then $V(t+\tau) = Y(t+\tau)$, and the integral equation simplifies to the Wiener-Hopf equation,

$$\Gamma(\tau+u) = \int_0^\infty h(u')\Gamma(u-u')du'.$$

This can be solved for the impulse response function using the methods developed in Wiener (1949). For discrete-time processes, the Wiener-Hopf equation is

$$\Gamma(\tau+u) = \sum_0^\infty h(u')\Gamma(u-u').$$

2.7 Summary and conclusions

In this chapter we have modelled discrete and continuous time second-order stationary stochastic processes in the time and frequency domains. We have noted that every second-order stationary stochastic process can be decomposed into the sum of two uncorrelated

second-order stochastic processes; the purely deterministic part contributed by the discrete part of the spectrum, and the purely non-deterministic part contributed by the continuous part of the spectrum. We have defined a white noise processes and noted that a second-order stationary continuous time white noise process exists only in the limiting sense as it either has zero power, or infinite variance. We have observed that a second-order stationary stochastic process can be considered the output of a time-invariant linear filter with white noise input, and that the output of a first-order linear system with white noise input also satisfies the Markov property. We have introduced the general estimation problem and shown how the optimal linear filter satisfies the Wiener-Hopf equation. The Wiener-Hopf equation is difficult to implement. In the next chapter, we shall see how higher order linear systems satisfy the Markov property when they are put in the state-space, and how the Kalman filter constructs the Wiener-Hopf filter for higher order discrete-time linear dynamic systems, both stationary and non-stationary.

3. The Kalman Filter

3.1 Introduction

In this chapter we review the development of the Kalman filter. This technique, which uses the state-space representation, finds the optimal linear filter for the output of a discrete-time linear dynamic system with white noise input. A higher order linear dynamic system with white noise input is shown to be Markovian in the state-space. The linear filter that minimizes expected mean-squared error is the conditional mean which is found by orthogonal projections. This leads to very simple recursive relations for updating the filter each time an additional observation becomes available. The filter actually gives the optimal linear solution to the general estimation problem described in the previous chapter. Hence it applies to smoothing, filtering, and prediction. In Section 3.3, new results by the author on the relative efficiency of the dynamic linear model, when applied to a linearly aggregated time series, are presented. It is shown that it is always more efficient to estimate the parameters of each component dynamic linear model and then estimate the parameters of the linearly aggregated dynamic linear model by the weighted sum of component parameter estimates rather than estimate them by using the dynamic linear model directly on the aggregated series.

3.2 The Kalman filter

Kalman (1960) developed a new method for approaching linear filtering and prediction problems which introduced a new way of determining the Wiener filter. It overcame many of the difficulties associated with previous methods such as: (*i*) the optimal filter is specified by its impulse response function, and it is not a simple task to synthesize the filter from the data; (*ii*) numerical determination of the optimal impulse response function is often quite involved, and gets rapidly worse with increasing complexity of the problem; (*iii*) important generalizations such as prediction in nonstationary processes require new derivations, frequently very difficult; and (*iv*) the assumptions and the mathematical derivation is not transparent, so consequences of fundamental assumptions are obscured.

Kalman approached the problem of estimating the conditional mean by using orthogonal projections. Stochastic processes are represented as the outputs of a linear dynamic systems excited by white noise. The linear dynamic system is reduced to a system of coupled first-order linear difference (differential) equations excited by white noise in a higher dimension space. In this state-space representation, the process is Markovian, and a single derivation finds the solution of the Wiener filter for both stationary and nonstationary processes. The solution depends only on the first two moment properties of the series. He showed that the dual of the Wiener filter problem is found to be the noise-free optimal regulator problem.

Kalman and Bucy (1961) expanded upon these results. They showed that the optimal linear filter is determined by five equations: (*i*) The differential equation govering the optimal filter which is excited by the observed signals and generates the best linear estimate of the message. (*ii*) The differential equations governing the error of the best linear estimate. (*iii*)

The time-varying gains of the optimal filter expressed in terms of the error variances. (*iv*) The nonlinear differential equation called the *variance equation* which governs the covariance matrix of the errors of the best linear estimate. The variance equation is of the Riccati type. The solution of the variance equation is the key to the filter. (*v*) The prediction formula. They showed that a sufficient condition for the existence of a steady-state solution to the variance equation is that the information matrix be nonsingular, and a sufficient condition for the optimal filter to be stable is the dual of the preceeding condition.

3.2.1 The state-space representation of a linear dynamic system

Let $v(t)$ be the output of the linear dynamic system $\Phi_t(D)v(t) = r(t)$ where

$$\Phi(D) = D^p + b_1(t)\, D^{p-1} + \ldots + b_{p-1}(t)D^1 + b_p(t)$$

is an $p'th$ degree polynomial in the differential operator D and $r(t)$ is a white noise process. If $v(t)$ is a covariance stationary stochastic process, the coefficients do not change with time. However, in general the coefficients are allowed to vary with time. Define the variables $u_1(t), \cdots, u_p(t)$ as follows:

$$u_1(t) = v(t)$$

$$u_2(t) = D(u_1(t)) = D(v(t))$$

$$u_p(t) = D(u_{p-1}(t)) = D^{p-1}(v(t)).$$

Then

$$\beta(t) = \begin{bmatrix} u_1(t) \\ \cdot \\ \cdot \\ u_p(t) \end{bmatrix}$$

satisfies the matrix first-order differential equation

$$[D + \mathbf{A}(t)]\beta(t) = \mathbf{r}(t),$$

where

$$\mathbf{A}(t) = \begin{bmatrix} 0 & -1 & 0 & \ldots & 0 \\ 0 & 0 & -1 & \ldots & \cdot \\ \cdot & \cdot & \cdot & \ldots & \cdot \\ \cdot & \cdot & \cdot & \ldots & \cdot \\ \cdot & \cdot & \cdot & \ldots & \cdot \\ \cdot & 0 & 0 & \ldots & -1 \\ b_p(t) & b_{p-1}(t) & b_{p-2}(t) & \ldots & b_1(t) \end{bmatrix} \quad \text{and} \quad \mathbf{r}(t) = \begin{bmatrix} 0 \\ 0 \\ \cdot \\ \cdot \\ 0 \\ r(t) \end{bmatrix}.$$

Hence $\beta(t)$, the unobservable state-vector, is Markovian in the p dimensional space. Actually,

Kalman included another term $\mathbf{G(t)}\alpha(t)$ where $\mathbf{G(t)}$ is a known function and $\alpha(t)$ is the control vector. We shall omit this term, because it is our goal to estimate and forecast the process, not to regulate it. Let the observed process be the linear function of the state-vector plus a noise process

$$y(t) = \mathbf{X}(t)'\beta(t) + e(t),$$

where $\mathbf{X}(t)'$ is a row vector of known coefficients and e_t is a white noise process.

In the discrete-time case, the linear dynamic system satisfies the matrix first-order difference equation

$$(\nabla + \mathbf{A}_t)\beta_t = \mathbf{r}_t,$$

where ∇ is the difference operator and the perturbation process $\mathbf{r}_t$ is a white noise process in the p dimension space. Hence the state-vector satisfies the matrix equation

$$\beta_{t+1} = \mathbf{A}_t\beta_t + \mathbf{r}_t,$$

where $\mathbf{A}_t$ is the state transition matrix of known coefficients at time t. Note that $\mathbf{A}_t$ now stands for $\mathbf{I} - \mathbf{A}_t$, the identity matrix minus the former $\mathbf{A}_t$. The observed process is the known linear function of the state-vector plus a discrete white noise process

$$y_t = \mathbf{X}_t'\beta_t + e_t,$$

where $\mathbf{X}_t'$ is a vector of known coefficients and the white noise process e_t is independent of the perturbation process.

3.2.2 The Kalman filter in terms of orthogonal projections

In the discrete-time general estimation problem, we are trying to estimate β_{t+k} given all the observations $y_1, \cdots, y_t$. If $k < 0$, $k = 0$, or $k > 0$ the problem is called smoothing, filtering, or prediction respectively. The mean-squared error $E(\tilde{\beta}_{t+k} - \beta_{t+k})^2$ is minimized by the conditional mean $\tilde{\beta}_{t+k} = E[\beta_{t+k} \mid y_1, \cdots, y_t]$, hence it is the best estimate. However in general, it is a difficult function to evaluate.

The Kalman filter finds the linear estimate

$$\hat{\beta}_{t+k} = a_1 y_1 + \cdots + a_t y_t$$

that minimizes $E(\hat{\beta}_{t+k} - \beta_{t+k})^2$. The set of all possible linear combinations of the observations forms an t dimensional vector space V_t. By using the Gram-Schmidt orthogonalization procedure, an orthonormal basis for the space is found. The state-vector estimate $\hat{\beta}_{t+k}$ is uniquely expressed in terms of that basis as the projection of the vector β_{t+k} onto the space V_t. The coefficients $a_1, \cdots, a_t$ are determined by the covariance matrix of $y_1, \cdots, y_t, \beta_{t+k}$. The orthogonal projection is the vector in the space V_t that is closest to β_{t+k}, hence it is the optimal linear estimator in terms of minimizing the expected squared-error loss function. Thus it is the solution to the Wiener-Hopf difference equation. When $y_1, \cdots, y_t, \beta_{t+k}$ are jointly normally distributed then the conditional mean $E[\beta_{t+k} \mid y_1, \cdots, y_t]$ is a linear function. Thus, in the Gaussian case, the optimal linear

estimator $\hat{\beta}_{t+k}$ is the optimal estimator in terms of minimizing expected squared-error.

One of the advantages of the Kalman filter is that the state-space approach gives a computationally efficient recursion for updating the filter without explicitly recalculating all the coefficients of the filter each time a new observation becomes available. The vector space spanned by the observations up to time t is a subspace of the vector space spanned by the observations up to time $t+1$. Hence y_{t+1} is the sum of $\hat{y}_{t+1|t}$, its projection onto V_t, and $y_{t+1}-\hat{y}_{t+1|t}$ which lies in the orthogonal complement of V_t. Similarly, $\hat{\beta}_{t+1}$ is the sum of $\hat{\beta}_{t+1|t}$ which is its projection onto V_t, and $\hat{\beta}_{t+1}-\hat{\beta}_{t+1|t}$ which lies in the orthogonal complement of V_t. Also $\hat{\beta}_t$ lies in V_t, hence $\hat{\beta}_{t+1|t}$ is a linear transformation of $\hat{\beta}_t$ since both lie in V_t, and the correction $\hat{\beta}_{t+1}-\hat{\beta}_{t+1|t}$ is a linear function of the prediction error $y_{t+1}-\hat{y}_{t+1|t}$ since both lie in the orthogonal complement of V_t. This means the estimate at time t is extrapolated to time $t+1$ by a linear transformation, and corrected to time $t+1$ by a linear function of the prediction error.

3.2.3 The recursive Kalman filter equations

Let the state-vector satisfy the discrete-time Markov process

$$\beta_{t+1} = \mathbf{A}_t\beta_t + \mathbf{r}_t,$$

where $\mathbf{A}_t$ is a matrix of known coefficients and $\mathbf{r}_t$, the perturbation vector at time t has mean vector 0 and covariance matrix $\mathbf{R}_t$. The sequence of perturbation vectors are independent of each other. Let the observed process be the known linear function of the state-vector plus random error

$$y_{t+1} = \mathbf{X}'_{t+1}\beta_{t+1} + e_{t+1},$$

where $\mathbf{X}'_{t+1}$ is a known row vector of coefficients, and e_{t+1} is a random error that has mean 0 and variance σ^2_{t+1}. The sequence of observation errors are independent of each other and of the perturbation errors. The estimator $\hat{\beta}_{t+1|t}$ is the projection of $\mathbf{A}_t\beta_t + \mathbf{r}_t$ onto V_t. Thus

$$\hat{\beta}_{t+1|t} = \mathbf{A}_t\hat{\beta}_t,$$

since the independence of the perturbation sequence implies $\mathbf{r}_t$ is orthogonal to the space V_t. Let $\hat{\beta}_t \sim W.S.(\beta_t\ ;\ \mathbf{V}_t)$, the notation from Duncan and Horn (1972) for the wide-sense distribution, denote that the distribution of $\hat{\beta}_t$ has mean vector β_t and covariance matrix $\mathbf{V}_t$. Then

$$\hat{\beta}_{t+1|t} \sim W.S.(\beta_{t+1};\mathbf{C}_{t+1}),$$

where the covariance matrix $\mathbf{C}_{t+1} = \mathbf{A}_t\mathbf{V}_t\mathbf{A}'_t + \mathbf{R}_t$.

Similarly the estimator $\hat{y}_{t+1|t}$ is the projection of $\mathbf{X}'_{t+1}\beta_{t+1} + e_{t+1}$ onto V_t. Since the independence of the observation error sequence implies e_{t+1} is orthogonal to the space V_t,

$$\hat{y}_{t+1|t} = \mathbf{X}'_{t+1}\hat{\beta}_{t+1|t}.$$

Hence

$$\begin{bmatrix} \hat{\beta}_{t+1|t}-\beta_{t+1} \\ y_{t+1}-\hat{y}_{t+1|t} \end{bmatrix} \sim W.S.\left[\begin{pmatrix}0\\0\end{pmatrix};\begin{bmatrix} \mathbf{C}_{t+1} & -\mathbf{C}_{t+1}\mathbf{X}_{t+1} \\ -\mathbf{X}'_{t+1}\mathbf{C}_{t+1} & \mathbf{X}'_{t+1}\mathbf{C}_{t+1}\mathbf{X}_{t+1}+\sigma^2_{t+1} \end{bmatrix}\right].$$

Thus the mean and variance of $(\hat{\beta}_{t+1|t} - \beta_{t+1} | \, y_{t+1} - \hat{y}_{t+1|t})$ are

$$[0 - C_{t+1}X_{t+1}(X'_{t+1}C_{t+1}X_{t+1} + \sigma^2_{t+1})^{-1}(y_{t+1} - \hat{y}_{t+1|t})]$$

and

$$[C_{t+1} - C_{t+1}X_{t+1}(X'_{t+1}C_{t+1}X_{t+1} + \sigma^2_{t+1})^{-1}X'_{t+1}C_{t+1}]$$

respectively. Hence the Kalman filter recursive equations for the updated parameter estimate and covariance matrix are

$$\hat{\beta}_{t+1} = \hat{\beta}_{t+1|t} + C_{t+1}X_{t+1}(X'_{t+1}C_{t+1}X_{t+1} + \sigma^2_{t+1})^{-1}(y_{t+1} - \hat{y}_{t+1|t})$$

and

$$V_{t+1} = C_{t+1} - C_{t+1|t}X_{t+1}(X'_{t+1}C_{t+1}X_{t+1} + \sigma^2_{t+1})^{-1}X'_{t+1}C_{t+1}$$

respectively. The above derivation is for the general (nonstationary) case and univariate observations. The stationary case occurs when σ_t^2 and the matrices A_t and R_t are all constant. The derivation for a multivariate observation vector y_t follows the same lines with X'_t being a matrix and the observation variance σ_t^2 being replaced by the observation covariance matrix Σ_t.

Ho and Lee (1969) used a posterior Bayesian approach to derive the Kalman filter equations under the Gaussian assumption. They note that the posterior density actually factors into a part which is dependent on the parameter, the updated covariance matrix, and the posterior mean; and a part which is dependent on the observations alone. Hence the updated covariance matrix and posterior mean are jointly sufficient statistics for the parameter.

Duncan and Horn (1972) introduced the Kalman filter into the statistics literature by relating it to a random-β regression. They note that

$$\begin{bmatrix} y_{t+1} \\ A_t\hat{\beta}_t \end{bmatrix} - \begin{bmatrix} X'_{t+1} \\ I \end{bmatrix}\beta_{t+1} \sim W.S. \left[\begin{bmatrix} 0 \\ 0 \end{bmatrix} ; \begin{bmatrix} \Sigma_{t+1} & 0 \\ 0 & C_{t+1} \end{bmatrix} \right],$$

where $C_{t+1} = A_t V_t A'_t + R_t$ is the prediction covariance matrix. The generalized least squares estimate of β_{t+1} given $\hat{\beta}_t$ and y_{t+1} minimizes the quadratic form

$$\left[\begin{bmatrix} y_{t+1} \\ (A_t\hat{\beta}_t \end{bmatrix} - \begin{bmatrix} X'_{t+1} \\ I \end{bmatrix}\beta_{t+1} \right]' \begin{bmatrix} \Sigma_{t+1} & 0 \\ 0 & C_{t+1} \end{bmatrix}^{-1} \left[\begin{bmatrix} y_{t+1} \\ (A_t\hat{\beta}_t \end{bmatrix} - \begin{bmatrix} X'_{t+1} \\ I \end{bmatrix}\beta_{t+1} \right].$$

The quadratic form to be minimized has the same form as that in fixed-β regression theory, so the minimum is given by the solution of the normal equations

$$[X_{t+1} \;\; I] \begin{bmatrix} \Sigma^{-1}_{t+1} & 0 \\ 0 & C^{-1}_{t+1} \end{bmatrix} \begin{bmatrix} X'_{t+1} \\ I \end{bmatrix} \beta_{t+1} = [X_{t+1} \;\; I] \begin{bmatrix} \Sigma^{-1}_{t+1} & 0 \\ 0 & C^{-1}_{t+1} \end{bmatrix} \begin{bmatrix} y_{t+1} \\ A_t\hat{\beta}_t \end{bmatrix}.$$

Hence it has the traditional regression interpretation as the projection of the augmented vector $\tilde{y}$ onto the vector space spanned by the columns of the augmented matrix $\tilde{X}$ with the inner product matrix Σ where

$$\tilde{\mathbf{y}} = \begin{bmatrix} \mathbf{y}_{t+1} \\ \mathbf{A}_t\hat{\boldsymbol{\beta}}_t \end{bmatrix} \quad \text{and } \tilde{\mathbf{X}} = \begin{bmatrix} \mathbf{X}'_{t+1} \\ \mathbf{I} \end{bmatrix} \quad \text{and } \boldsymbol{\Sigma} = \begin{bmatrix} \Sigma_{t+1} & 0 \\ 0 & \mathbf{C}_{t+1} \end{bmatrix}.$$

The solution to the normal equations is

$$\hat{\boldsymbol{\beta}}_{t+1} = (\mathbf{X}_{t+1}\Sigma_{t+1}^{-1}\mathbf{X}'_{t+1} + \mathbf{C}_{t+1}^{-1})^{-1}(\mathbf{X}_{t+1}\Sigma_{t+1}^{-1}\mathbf{y}_{t+1} + \mathbf{C}_{t+1}^{-1}\mathbf{A}_t\hat{\boldsymbol{\beta}}_t),$$

which in the more familiar Kalman filter form is

$$= \mathbf{A}_t\hat{\boldsymbol{\beta}}_t - (\mathbf{X}_{t+1}\Sigma_{t+1}^{-1}\mathbf{X}'_{t+1} + \mathbf{C}_{t+1}^{-1})^{-1}\mathbf{X}_{t+1}\Sigma_{t+1}^{-1}(\mathbf{y}_{t+1} - \hat{\mathbf{y}}_{t+1}),$$

where $\hat{\mathbf{y}}_{t+1} = \mathbf{X}'_{t+1}\mathbf{A}_t\hat{\boldsymbol{\beta}}_t$ is the predicted value.

3.3 The dynamic linear model applied to a linearly aggregated time series.

Often a time series is a weighted sum of component time series. In Appendix 1, Bolstad (1990) showed that if the same dynamic linear model generates all the component series, it also generates the linearly aggregated time series. The parameters of the aggregated model are weighted sums of the parameters of the component models. There are two possible estimators for the parameters of the linearly aggregated series; the dynamic linear model estimators from the linearly aggregated series, and the weighted sum of the dynamic linear model estimators for the component series. It established the relative efficiency of the two estimators. The second was found to be more efficient in all realistic cases.

Consider the case where the prior distribution of the dynamic parameter estimators is the same for the dynamic linear model applied to the linearly aggregated time series as for the weighted sum of the component dynamic linear model estimators. In that case the posterior covariance matrix for the linearly aggregated model estimators equals the posterior covariance matrix for the weighted sum of the component model estimators plus the weighted sum of the covariance matrices of the posterior mean vectors averaged over all possible component observations such that the weighted sum is the linearly aggregated observation. If the component models are being generated independently, the covariance of the linearly aggregated model parameters is

$$COV(\hat{\boldsymbol{\beta}}_t^A - \boldsymbol{\beta}_t^A \mid \mathbf{y}_t^A) = COV(\hat{\boldsymbol{\beta}}_t^W - \boldsymbol{\beta}_t^A \mid \mathbf{y}_t^{(1)}, \cdots, \mathbf{y}_t^{(k)})$$

$$+ \sum_{i=1}^{k} (w^{(i)})^2[\mathbf{C}_t^{(i)}\mathbf{X}_t(\mathbf{X}'_t\mathbf{C}_t^{(i)}\mathbf{X}_t + \sigma_i^2)^{-1}\mathbf{X}'_t\mathbf{C}_t^{(i)} - (w^{(i)})^2\mathbf{C}_t^{(i)}\mathbf{X}_t[\sum_{j=1}^{k}(w^{(j)})^2(\mathbf{X}'_t\mathbf{C}_t^{(i)}\mathbf{X}_t + \sigma_j^2)]^{-1}\mathbf{X}'_t\mathbf{C}_t^{(i)}].$$

Hence the relative efficiency of the linearly aggregated model relative to the weighted sum of the component models is given by

$$eff(\hat{\boldsymbol{\beta}}_t^A : \hat{\boldsymbol{\beta}}_t^W) = COV(\hat{\boldsymbol{\beta}}^W \mid \mathbf{y}_t^{(1)}, \cdots, \mathbf{y}_t^{(k)}) \times [COV(\hat{\boldsymbol{\beta}}^W \mid \mathbf{y}_t^{(1)}, \cdots, \mathbf{y}_t^{(k)})$$

$$+ \sum_{i=1}^{k} (w^{(i)})^2[\mathbf{C}_t^{(i)}\mathbf{X}_t(\mathbf{X}'_t\mathbf{C}_t^{(i)}\mathbf{X}_t + \sigma_i^2)^{-1}\mathbf{X}'_t\mathbf{C}_t^{(i)} - (w^{(i)})^2\mathbf{C}_t^{(i)}\mathbf{X}_t[\sum_{j=1}^{k}(w^{(j)})^2(\mathbf{X}'_t\mathbf{C}_t^{(i)}\mathbf{X}_t + \sigma_j^2)]^{-1}\mathbf{X}'_t\mathbf{C}_t^{(i)}]]^{-1}.$$

The covariance of $\mathbf{C}_t^{(i)}\mathbf{X}_t(\mathbf{X}'_t\mathbf{C}_t^{(i)}\mathbf{X}_t + \sigma_i^2)^{-1}(y_t^{(i)} - \hat{y}_t^{(i)} \mid y_t^A)$, the i'th posterior mean vector, equals 0 if and only if $y_t^{(i)} - \hat{y}_t^{(i)}$ has covariance matrix identically equal the zero matrix, or if it and $y_t^A - \hat{y}_t^A$ are perfectly correlated. Hence, for all realistic cases, the inequality is a strict

one. Starting from the same prior distribution, the weighted sum estimator always has a smaller Wilk's generalized variance. Thus the parameters of the linearly aggregated time series are more efficiently estimated by the weighted sum of the dynamic linear model component series estimators than by the dynamic linear model estimator for the linearly aggregated series.

This can also be seen to be a consequence of the projection theorem. The set of observations on the linearly aggregated time series is a subspace of the set of observations on each of the component time series. The length of the perpendicular to the subspace must be at least as long as the length of the perpendicular to the total observation space.

Kohn (1982) investigated conditions in which the estimate using the linearly aggregated series is efficient for the special cases of ARIMA and stationary Gaussian models. Bolstad (1990) showed that for all realistic cases, the weighted sum of component estimators is more efficient.

The conclusion that the estimator from the linearly aggregated time series is not as good an estimator for the weighted sum of parameters as the weighted sum of component estimators holds even if the dynamic linear models are not independent. The posterior covariance matrix for the linearly aggregated model estimators will still equal the posterior covariance matrix for the weighted sum of the component model estimators plus the weighted sum of the covariance matrices of the posterior mean vectors averaged over all possible component observations such that the weighted sum is the aggregated observation. In that case, however, the formula for the relative efficiency would be somewhat more complicated.

3.4 Summary and conclusions

In this chapter we have used the state-space representation of a linear dynamic system in order to develop the Kalman filter. This allows the construction of the solution to the discrete-time general linear estimation problem in a recursive manner. The Kalman filter is the optimal linear filter, and in the case where the perturbations and observation errors are normally distributed, it is the optimal filter. In the next two chapters, we shall investigate filtering models for non-Gaussian observation and perturbation errors.

The relative efficiency of the dynamic linear model estimators applied to a linearly aggregated time series is developed. We note that if the same dynamic linear model generates all the component time series it also generates the linearly aggregated time series. We have shown that it is always more efficient to estimate the parameters of each component dynamic linear model separately and linearly combine the estimates than to estimate the parameters of the aggregated model using the aggregated time series

4. Filtering In Non-Gaussian Models

4.1 Introduction

The Kalman filter is known to be the optimal linear filter in the sense of minimizing expected squared error. In that sense, the optimal filter is the conditional mean, given the observations. Thus the Kalman filter is the optimal filter whenever the conditional mean is a linear function of the observations. This is the case when the perturbations and observation errors are normal. However, when the perturbation or observation distributions are non-normal, it may be possible to find non-linear filters which are considerably better than the Kalman filter. Many authors have developed methods for non-linear filtering that preserve the recursive nature of the Kalman filter. We now discuss some of them.

4.2 Non-Gaussian filtering models

Masriclez (1975) developed two filters for dynamic linear models, the first is applicable when the distribution for the predicted state estimator is approximately Gaussian, and the observation has a known non-Gaussian distribution. Let $f(y_t \mid \mathbf{y}_{t-1})$, the prediction density of the observation at time t given the observations up to time $t-1$, be a function that is twice differentiable. Under the assumption that the prior state estimator density is $N(\hat{\beta}_{t\mid t-1}; \mathbf{C}_{t\mid t-1})$, the mean vector and covariance matrix of the posterior density satisfy

$$\hat{\beta}_t = \hat{\beta}_{t\mid t-1} + \mathbf{C}_{t\mid t-1}\mathbf{X}_t'\mathbf{g}_t(y_t) \quad \text{and} \quad \mathbf{V}_t = \mathbf{C}_{t\mid t-1} - \mathbf{C}_{t\mid t-1}\mathbf{X}_t'\mathbf{G}_t(y_t)\mathbf{X}_t\mathbf{C}_{t\mid t-1}$$

respectively, where $\mathbf{g}_t(y_t)$ is the score vector and $\mathbf{G}_t(y_t)$ is Fisher's information matrix of the observation density $f(y_t \mid \mathbf{y}_{t-1})$. Note that if the observation density is univariate Gaussian, the correction term is the linear function through the prior mean given by the Kalman filter. However, for univariate non-Gaussian observation distributions, the score function and hence the correction term is not a linear function of the distance the observation from the prior mean. For symmetric heavy-tailed observation distributions the score function de-emphasizes the extreme observations, hence for large deviations the correction function should lie closer to the horizontal axis than the line of the Kalman filter.

The second filter is appropriate when there is Gaussian observation noise, but non-Gaussian perturbations. Provided $\mathbf{T}_t^{-1} = \mathbf{X}_t\Sigma^{-1}\mathbf{X}_t'$ is nonsingular, the minimum variance estimator can be calculated from the following relations:

$$\hat{\beta}_t = \mathbf{T}_t\mathbf{X}_t(\Sigma^{-1}y_t - \mathbf{g}(y_t)) \quad \text{and} \quad \mathbf{V}_t; = \mathbf{T}_t - \mathbf{T}_t\mathbf{X}_t\mathbf{G}_t(y_t)\mathbf{X}_t'\mathbf{T}_t,$$

where $\mathbf{g}_t(y_t)$ and $\mathbf{G}_t(y_t)$ are the score and information respectively as defined above.

These two filters retain the recursive nature of the Kalman filter by using the form of a prediction plus a correction. However, they require either the state prediction or the observation noise to be Gaussian. The use of the score function and Fisher's information for the prediction density establishes links with classical statistical estimation theory, and indicates how filters could be modified by changing the shape of the score function.

Masreliez and Martin (1977) developed modifications of the Kalman filter that are robust for heavy-tailed symmetric distributions for the same two cases as Masreliez (1975); Gaussian state estimation and heavy-tailed observation distribution, and heavy-tailed non-Gaussian state estimation and Gaussian observation distribution. They used influence functions developed for the least favorable alternative distribution and found the bounds on the asymptotic variance. They performed Monte Carlo studies using the filters on a variety of distributions and they were found to be almost as efficient as the Kalman filter in the Gaussian case, and much more efficient in the heavy-tailed cases.

Kitagawa (1987) developed a non-Gaussian state-space modelling methodology for nonstationary time series which follow a dynamic linear model. The recursive formulae for one-step prediction, filtering, and smoothing are found by using Bayes' theorem. The formulae are implemented by approximating the initial state-vector density, the perturbation density, and the observation density by piecewise linear functions and doing all integrations numerically. The unknown parameters of the model are found by maximizing the log-likelihood. If there are competing models, choice is made by maximizing the Akaike information criterion. This essentially chooses the model with the largest log-likelihood adjusted for the number of parameters. Kitagawa (1987) demonstrated this procedure on simulated data and also on a non-stationary binary process (Tokyo rainfall data). Martin and Raftery (1987) comment that the computational difficulties associated with the method preclude its use for high dimensional state-vectors.

Nelder and Wedderburn (1972) introduced the (static) generalized linear model in order to model situations where an independent sequence of random variables are taken, and a function of their means is linked to a linear predictor on known covariates. This allows regression techniques to be used for non-normal (but exponential family) distributions on a non-linear scale. West, Harrison, and Migon (1985) extended this into the *dynamic* generalized linear model, where the mean values of the parameters (covariates) change over time according to some linear pattern, and the covariance matrix of the estimates is discounted as in Ameen and Harrison (1985). They developed simple recursions for updating the parameter estimates and covariance estimates (posterior means and covariances of the parameter distribution) that have the same form as the recursions used in the Kalman filter.

The dynamic generalized linear model allows the parameter vector β to change over time according to the parameter dynamic equation

$$\beta_t = \mathbf{A}_t \beta_{t-1} + \mathbf{r}_t,$$

where $\mathbf{A}_t$ is a matrix of known coefficients and $\mathbf{r}_t$, the perturbation vector at time t, has mean vector $\vec{0}$ and covariance matrix $\mathbf{R}_t$. The perturbation vectors at different times are independent of each other. This allows the parameter vector to behave in a Markovian manner.

i) The systematic component of the model is η_t, and the linear predictor is given by

$$\eta_t = \mathbf{X}_t' \beta_t,$$

where $\mathbf{X}_t'$ is the known row vector of the covariates for the observation at time t and β_t

is the vector of unknown parameters.

ii) The link between the systematic part of the model and the random part of the model is given by

$$g(\mu_t) = \eta_t,$$

where μ_t is the expected value of the observation at time t and η_t is the linear predictor. The link function g is assumed known.

iii) Instead of using the observation equation which relates the observation to the linear parameters and measurement error, the observation distribution of y_t is explicitly assumed to be from the one-dimensional exponential family which has probability density function

$$f(y_t \mid \theta_t, \phi) = \exp[c(y_t, \phi) + \phi(\theta_t y_t - b(\theta_t))],$$

where θ_t is the unknown parameter of observation y_t and ϕ is a known scale parameter. (When y is binomial or Poisson, $\phi = 1$.) The mean and variance of observation y_t are given by $\mu_t = E(y_t) = b'(\theta_t)$ and $var(y_t) = \phi^{-1} b''(\theta_t)$ respectively. The sampling parameter is related to the linear parameters β_t by the guide relationship $g(b'(\theta_t)) = \eta_t$.

Guttman and Pena (1985) developed a filter that is robust against outliers by allowing the observation distribution to be a mixture of two distributions, the usual distribution and an outlier distribution. The posterior probability of each is computed, and the posterior estimator distribution is replaced by a single normal distribution having the same first two moments as the mixture. This gives very little weight to any outlying observations, thus providing an outlier filter for the dynamic linear model. Meinhold and Singpurwalla (1989) developed a robustified Kalman filter model by using a Bayesian approximation to the posterior distribution where the prior distribution is a mixture of Student-t distributions and the observation distribution is Student-t.

4.3 Summary and conclusions

In this chapter we have investigated some non-Gaussian filtering models. Two filtering models that have simple recursions like the Kalman filter were developed by Masreliez (1975), to be applicable when either the observation is normal and the prior state estimator is heavy-tailed, or vice versa. The use of the score vector and Fisher's information for the predictive density indicates how the correction function shape should be modified for non-Gaussian distributions. We noted that the Bayesian state-space model developed by Kitagawa (1987) which is implemented by numerical integration using piecewise linear approminations to the prior state-vector density, perturbation density, and observation density, requires too much computation to be useful when there are high dimensional state-vectors. The dynamic generalized linear model developed by West (1985) is introduced for non-normal exponential family observation distribution and where a function of the mean is linked to the linear parameters.

In the next chapter we shall introduce models where the observation and perturbation distributions are mixtures of distributions (often normal). This will enable non-linear filters to be

developed which are effective forecasting tools where the nonstationary time series contains unpredictable abrupt as well as evolutionary changes in pattern as well as outliers.

5. Multiprocess Dynamic Regression Models

5.1 Introduction

In this chapter we investigate multiprocess dynamic regression models. In these models the perturbation and observation distributions are considered to be mixtures of distributions, often normal distributions. This allows simple recursions to be developed for heavy-tailed distributions, making effective forecasting tools for non-stationary time series that contain outliers and abrupt, as well as evolutionary, changes in pattern. The multivariate multiprocess dynamic linear model is developed for the case where the observation errors are correlated, and it is used for forecasting the inflation rate in the consumer price index. The multiprocess dynamic linear model with biased perturbations is developed for modelling pulsatile data and used to model growth-hormone level in young steers subject to dietary stress. The multiprocess dynamic generalized linear model is developed. In this model the observation distribution is a mixture of known one-dimensional exponential family distributions. It is demonstrated to be an effective forecasting model for a time-varying proportion.

5.2 The multiprocess dynamic linear model

Harrison and Stevens (1971) developed a filtering and forecasting procedure based on the multiprocess dynamic linear model in order to model non-stationary time series which contain outliers and are subject to abrupt, as well as evolutionary, changes in pattern. In the dynamic linear model the state-vector contains all the information from the past and present that is relevant to the future development of the process. That is, the state-vector has the property that given the past history of the process up to and including time t_0 and the state-vector at time t_0, the conditional distribution of the stochastic process at time $t_1 > t_0$ depends only on the state-vector at time t_0. The multiprocess dynamic linear model deals with non-Gaussian perturbation distributions by using a mixture of k Gaussian distributions. Each type of pattern change is caused by a different type of perturbation and corresponds to a different element of the mixture. The element of the mixture which occurs at time t depends on the current value of the perturbation-index variable which comes from an independent sequence of multinomial random variables with known prior probabilities. (Harrison and Stevens (1971) called this the state variable, but this overloads the term *state* because the dynamic parameter vector corresponds to the state-vector in the Kalman filter terminology.) The prior probabilities do not have to remain constant over time. This flexibility allows prior knowlege by the forecaster to be introduced into the model. This model can deal with outliers by augmenting the state-vector with a transient variable which is included in the current observation, but does not carry forward. Thus the model is robust for both non-normal perturbation noise and non-normal observation noise.

Harrison and Stevens (1976) summarized the foundations of Bayesian forecasting. Essential to the methods are: (*i*) the parametric or state-space model, (*ii*) the probabilistic information on the model parameters, (*iii*) the sequential model definition which describes the

dynamic behavior of the model parameters, and (*iv*) some uncertainty in choosing the underlying model from a number of possibilities.

The parameter dynamics are governed by the equation

$$\boldsymbol{\beta}_t = \mathbf{A}_t \boldsymbol{\beta}_{t-1} + \mathbf{r}_t,$$

where $\boldsymbol{\beta}_t$ is the parameter vector at time t, $\mathbf{A}_t$ is a known matrix of coefficients at time t, and $\mathbf{r}_t$ is the perturbation vector. When the current perturbation-index $I_t = j$, the perturbation $\mathbf{r}_t$ is normally distributed with mean vector 0 and known covariance matrix $\mathbf{R}_t^{(j)}$ which depends on the perturbation-index j and can change over time. All of the perturbation-index random variables are independent of each other and each can be considered the outcome of a single multinomial trial with the prior probabilities given by

$$P(I_t = j) = \pi_t^{(j)} \qquad \text{for } j = 1,\ldots,k,$$

which can change with time.

The observation y_t is governed by the observation equation

$$y_t = \mathbf{X}_t' \boldsymbol{\beta}_t + e_t,$$

where the observation errors e_t are iid $N(0,\sigma_t^2)$ random variables which are independent of the perturbation random vectors and the perturbation-index random variables, and $\mathbf{X}_t'$ is the known row vector of coefficients at time t.

In the Harrison-Stevens estimation algorithm, sometimes called the *multiprocess Kalman filter*, the prior distribution is a mixture of k normal distributions, dependent on the previous value of the perturbation-index; and the parameter perturbation distribution is a mixture of k normal distributions dependent on the current value of the perturbation-index variable. Hence, the posterior distribution of the parameter vector is calculated conditionally for each of the k^2 possible pairs of perturbation indices. The posterior probability of each pair of perturbation indices is calculated (using the assumption of the normal distribution). The posterior distribution (which is a mixture of k^2 normal distributions depending on the two most recent values of the perturbation-index) is approximated by a mixture of k normal distributions (depending only on the most recent perturbation-index) having the same mean and variance. Pena and Guttman (1989) show that this is the optimal approximation in the sense of minimizing the Kullbeck-Leibler distance.

In the Harrison-Stevens algorithm, the correction applied to the parameter vector after the observation is a weighted average of the corrections which are calculated (conditionally on each pair of current and previous perturbation-index values) by the Kalman filter, where the weights are the posterior probabilities of the perturbation-index values given the observation. Thus Bayes' theorem is used twice; first in calculating the conditional corrections using the Kalman filter, and second in calculating the posterior probabilities of the perturbation-index. The posterior probabilities of the current perturbation-index values are found by summing the joint posterior probabilities over the previous perturbation-index values, and the algorithm then is ready to analyze another observation. Makov (1983) concluded that condensing the mixture

distributions into a single distribution, referred to as "probabilistic editing", is in the long run more accurate than the "quasi Bayes" procedure where the most likely posterior distribution is used.

In Appendix 2, Bolstad (1986 A) analyzed the Harrison-Stevens algorithm and showed how it is an effective forecasting tool for non-stationary time series since it responds very quickly when there is an underlying change in the pattern of the time series, yet is robust against outliers. After the current observation, the posterior probabilities of both the current and previous values of the perturbation indices are calculated using Bayes theorem. Thus the posterior perturbation-index probabilities at time $t-1$ are calculated after observation y_{t-1} and again after observation y_t. The use of two observations enables the algorithm to make the distinction between the types of perturbations, and allows the rapid adjustment of the parameters on the second observation subsequent to a change in the pattern of the series.

Smith and West (1983) and Smith et al. (1983) used the model as a fault detection procedure when monitoring renal transplant data, in order to detect a change of state (organ rejection) as quickly as possible, thus allowing early medical intervention.

5.3 The multivariate multiprocess dynamic linear model

In Appendix 3, Bolstad (1986 B) developed a computationally efficient algorithm for Harrison-Stevens forecasting involving multivariate observations with contemporaneously correlated errors. This algorithm uses an augmented state-vector which includes the parameter vector and the observation mean vector. The algorithm analyzes the multivariate observation one component at a time using the Harrison-Stevens algorithm on the augmented state-vector. As each component is analyzed, the dimension of the state-vector is decreased by one. This algorithm avoids matrix inversions necessary in a straightforward application of the Harrison-Stevens algorithm to multivariate observations, thus increasing the computational efficiency.

5.3.1 The forecasting algorithm

The algorithm starts with the following conditions. At time $t-1$ after all M components of observation $\mathbf{y}_{t-1}$ have been analyzed, the posterior parameter estimators are $\hat{\beta}_{t-1}^{(i)}$ and the posterior covariance matrices are $\mathbf{V}_{t-1}^{(i)}$, conditional on the perturbation-index $I_{t-1}=i$. The posterior perturbation-index probabilities are given by $q_{t-1}^{(i)} = P(I_{t-1}=i \mid \mathbf{Y}_{t-1})$ where $\mathbf{Y}_{t-1}$ denotes all the observations up to time $t-1$ and the prior probabilities of the perturbation-index are given by $\pi_t^{(j)} = P(I_t=j)$. The steps of the algorithm are:

1. Extrapolate the estimators and covariance matrices forward to time t conditionally on the perturbation indices $I_{t-1}=i$, $I_t=j$ by

$$\hat{\beta}_t^{(i,j)} = \mathbf{A}_t \hat{\beta}_{t-1}^{(i)} \quad \text{and} \quad \mathbf{C}_t^{(i,j)} = \mathbf{A}_t \mathbf{V}_{t-1}^{(i)} \mathbf{A}_t' + \mathbf{R}_t^{(j)},$$

respectively, for $i = 1,\ldots,k$ and $j = 1,\ldots,k$.

2. Let the prior mean vectors $\Theta_{M\,t}^{(i,j)}$ be a $(p+M)\times 1$ vector of zeros, and the prior covariance matrices be

$$\mathbf{W}_{M\,t}^{(i,j)} = \begin{bmatrix} \mathbf{C}_t^{(i,j)} & -\mathbf{C}_t^{(i,j)}\mathbf{X}_t \\ -\mathbf{X}_t'\mathbf{C}_t^{(i,j)} & \mathbf{X}_t'\mathbf{C}_t^{(i,j)}\mathbf{X}_t + \Sigma_e \end{bmatrix}$$

for $i = 1,\ldots,k$ and $j = 1,\ldots,k$. Σ_e is the known covariance matrix of the observations. The first p components of $\Theta_{M\,t}^{(i,j)}$ are the mean vector of the parameter vector estimates, and the remaining M components are for the mean vector of the observation.

3. Set $m = M$.

4. Let $p_{m\,t}^{(i,j)} = q_{t-1}^{(i)}\pi_t^{(j)}$.

5. Partition off the last component of the prior mean vectors and covariance matrices. Thus

$$\Theta_{m\,t}^{(i,j)} = \begin{bmatrix} \Theta_{m-1\,t}^{(i,j)} \\ \theta_{m\,t}^{(i,j)} \end{bmatrix} \quad \text{and} \quad \mathbf{W}_{m\,t}^{(i,j)} = \begin{bmatrix} \mathbf{U}_{11}^{(i,j)} & \mathbf{U}_{12}^{(i,j)} \\ \mathbf{U}_{21}^{(i,j)} & \mathbf{U}_{22}^{(i,j)} \end{bmatrix}$$

for $i = 1,\ldots,k$ and $j = 1,\ldots,k$.

6. Update the mean vectors and covariance matrices posterior to $\hat{e}_{m\,t}^{(i,j)}$ by

$$\Theta_{m-1\,t}^{(i,j)} = \Theta_{m-1\,t}^{(i,j)} + \mathbf{U}_{12}^{(i,j)}(\mathbf{U}_{22}^{(i,j)})^{-1}(\hat{e}_{m\,t}^{(i,j)} - \theta_{m\,t}^{(i,j)})$$

and

$$\mathbf{W}_{m-1\,t}^{(i,j)} = \mathbf{U}_{11}^{(i,j)} - \mathbf{U}_{12}^{(i,j)}(\mathbf{U}_{22}^{(i,j)})^{-1}\mathbf{U}_{21}^{(i,j)}$$

respectively, where $\hat{e}_{m\,t}^{(i,j)}$ is the $m'th$ component of $\mathbf{y}_t - \mathbf{X}_t'\hat{\beta}_t^{(i,j)}$, the prediction error of the observation vector at time t, given all the observations up to time $t-1$. Note that the matrices to be inverted are 1 by 1.

7. Recalculate the posterior perturbation-index probabilities by

$$p_{m-1\,t}^{(i,j)} = p_{m\,t}^{(i,j)}(f(\hat{e}_{m\,t}))^{-1}(2\pi)^{-1/2}|\mathbf{U}_{22}^{(i,j)}|^{-1/2}\exp[-1/2(\mathbf{U}_{22}^{(i,j)})^{-1}(\hat{e}_{m\,t}^{(i,j)} - \theta_{m\,t}^{(i,j)})^2.$$

The unconditional density, $f(\hat{e}_{m\,t}^{(i,j)})$ is a constant in each of the terms and they must sum (over i and j) to one. Note the assumption of normality is required.

8. If $m > 1$ then let $m = m-1$ and go to step 5.

9. Calculate the posterior probabilities of the most recent perturbation-index to be $q_t^{(j)} = \sum_{i=1}^{k} p_{0\,t}^{(i,j)}$ and condense the mixture of k^2 distributions into a mixture of k distributions conditional on $I_t = j$. The condensed mean is

$$\mu_t^{(j)} = (q_t^{(j)})^{-1}\sum_{i=1}^{k} p_{0\,t}^{(i,j)}\Theta_{0\,t}^{(i,j)},$$

and the condensed covariance matrix is

$$\mathbf{V}_t^{(j)} = (q_t^{(j)})^{-1}\sum_{i=1}^{k} p_{0\,t}^{(i,j)}[(\mathbf{W}_{0\,t}^{(i,j)} + [(\Theta_{0\,t}^{(i,j)} - \mu_t^{(j)})(\Theta_{0\,t}^{(i,j)} - \mu_t^{(j)})']]$$

for $j = 1,\ldots,k$.

10. The posterior estimators conditional on $I_t = j$ are

$$\hat{\beta}_t^{(j)} = (q_t^{(j)})^{-1} \sum_{i=1}^{k} p_{0t}^{(i,j)} \hat{\beta}_t^{(i,j)} - \mu_t^{(j)}$$

for $j = 1,\ldots,k$.

11. Let $t = t+1$ and go to step 1.

5.3.2 The multivariate multiprocess dynamic linear model applied to inflation data

The consumer price index is a measure of price inflation over time. It measures the prices on a "basket of goods and services" that a "typical" household uses. The goods and services are classified into six main categories and the consumer price indices are also calculated for each of these groups. The categories are: Food, Housing, Household Operation, Apparel, Transportation, and Miscellaneous. This data was obtained from the Department of Statistics INFOS on-line computer database. The aggregate consumer price index is a linear aggregate of these six consumer price subindices, where the weights are updated from time to time to bring them in line with the expenses of a "typical" household. The weights for the period since 1960 are given in Table 1.

Table 1.

Consumer price index group weights						
Years	Food	Housing	H. Op.	Apparel	Trans.	Misc.
60-64	32.26	15.40	10.60	15.43	8.76	17.55
65-74	30.09	17.99	11.21	13.12	9.44	18.15
75-77	17.82	27.80	14.31	8.55	14.70	16.82
78-80	19.12	23.53	17.04	7.95	13.84	18.52
81-83	19.62	18.38	16.08	7.13	18.26	20.53
84-	18.35	21.00	16.00	6.37	18.22	20.06

The aggregate quarterly inflation rate in the consumer price index is defined to be the ratio of the aggregate CPI values at successive times.

The (univariate) Harrison-Stevens forecasting algorithm was applied to the aggregate quarterly inflation rate in order to determine a suitable dynamic linear model. The autocorrelation function of the differences in the quarterly inflation factor was computed and the first autocorrelation was found to be equal to $-.365$. The negative autocorrelation indicated a model where the inflation factor overshoots the required correction. The model with four possible values of the perturbation-index was selected with three parameters, mean, slope, and transient. To model the overshooting, the transient was allowed to be observed at time t and to partially feed back into the mean at time $t-1$. The perturbation-index, perturbation type, perturbation variance, and prior probabilities chosen are in Table 2.

Table 2.

Perturbation variances and prior probabilities			
Index	Perturbation	Variance	Prior Probability
1	no change	0	.9
2	transient	$36\times\sigma^2$	.05
3	mean change	$16\times\sigma^2$	.04
4	slope change	$.25\times\sigma^2$	.01

The parameter vector is

$$\beta_t = \begin{bmatrix} m_t \\ b_t \\ \delta_t \end{bmatrix}$$

where m_t, b_t, and δ_t are the local mean, slope, and transient respectively. The dynamic matrix is

$$\mathbf{A}_t = \begin{bmatrix} 1 & 1 & .8 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

The model was found to give forecasts that (in terms of prediction errors) were somewhat of an improvement over the naive forecasts $\hat{y}_{t+1} = y_t$.

The aggregate quarterly inflation rate Q_t equals the weighted average of the group quarterly inflation rates $Q_t^{(i)}$ since

$$Q_t = \frac{CPI_t}{CPI_{t-1}} = \frac{\sum_i w^{(i)} CPI_t^{(i)}}{\sum_j w^{(j)} CPI_{t-1}^{(j)}} = \frac{\sum_i (CPI_t^{(i)}/CPI_{t-1}^{(i)})(w^{(i)} CPI_{t-1}^{(i)})}{\sum_j w^{(j)} CPI_{t-1}^{(j)}} = \sum_{i=1} c^{(i)} Q_t^{(i)},$$

where the weight $c^{(i)}$ is the proportion of the previous quarter price index times $w^{(i)}$, the base period weight for group i.

If the same multiprocess dynamic linear model generates all the quarterly group inflation rates, it also generates the aggregate inflation rate. However, the multiprocess dynamic linear models are connected through the perturbation-index variables, so the results of Section 3.3 cannot be applied directly. We cannot prove that forecasting the component series and then aggregating the forecasts is always more efficient than forecasting from the aggregated series. It is likely, however, that in practice the former method will give better forecasts than the latter, since it does use more complete information. We will compare the methods on the quarterly inflation rates since 1960.

The dynamic behavior of each component follows the pattern found above. However, we allowed the observations to be contemporaneously correlated. To estimate the contemporaneous covariance matrix, quarterly inflation rate differences were taken for each of the groups and the

covariance matrix of the differences calculated and multiplied by $2^{-1/2}$. This gave

$$\Sigma = \begin{bmatrix} .0005597 & .0000749 & .0001319 & .0001968 & .0000826 & .0001414 \\ .0000749 & .0001300 & .0000755 & .0002038 & -.0000191 & .0000130 \\ .0001319 & .0000755 & .0003700 & .0002602 & .0001020 & .0001453 \\ .0001968 & .0002038 & .0002602 & .006739 & .0000269 & .0000261 \\ .0000826 & -.0000191 & .0001020 & .0000269 & .0003232 & .0001398 \\ .0001414 & .0000130 & .0001453 & .0000261 & .0001398 & .0003712 \end{bmatrix}.$$

To model this as a multivariate multiprocess dynamic linear model, we let

$$\beta_t = \begin{bmatrix} \beta_{1t} \\ \beta_{2t} \\ \vdots \\ \beta_{6t} \end{bmatrix} \quad \text{and} \quad y_t = \begin{bmatrix} y_{1t} \\ y_{2t} \\ \vdots \\ y_{6t} \end{bmatrix}$$

be the parameter vector and the observation vector respectively where $\beta_{i\,t}$ and $y_{i\,t}$ are the parameter vector and the observation for group i at time t. The dynamic matrix and the observation matrix are given by

$$\mathbf{A}_t = \begin{bmatrix} \mathbf{A}_t & 0 & 0 & 0 & 0 & 0 \\ 0 & \mathbf{A}_t & 0 & 0 & 0 & 0 \\ 0 & 0 & \mathbf{A}_t & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathbf{A}_t & 0 & 0 \\ 0 & 0 & 0 & 0 & \mathbf{A}_t & 0 \\ 0 & 0 & 0 & 0 & 0 & \mathbf{A}_t \end{bmatrix} \quad \text{and} \quad \mathbf{X}_t' = \begin{bmatrix} \mathbf{X}_t' & 0 & 0 & 0 & 0 & 0 \\ 0 & \mathbf{X}_t' & 0 & 0 & 0 & 0 \\ 0 & 0 & \mathbf{X}_t' & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathbf{X}_t' & 0 & 0 \\ 0 & 0 & 0 & 0 & \mathbf{X}_t' & 0 \\ 0 & 0 & 0 & 0 & 0 & \mathbf{X}_t' \end{bmatrix}$$

respectively, where $\mathbf{A}_t$ and $\mathbf{X}_t'$ are the univariate dynamic matrix and observation matrix used above for the aggregate inflation rate. The entries in the perturbation covariance matrices are all zero except

$$R^{(2)}_{3k\,3k} = 36\Sigma_{k\,k}\ , \quad R^{(3)}_{3k-2\,3k-2} = 16\Sigma_{k\,k}\ , \quad \text{and} \quad R^{(4)}_{3k-1\,3k-1} = .25\Sigma_{k\,k}, \text{ for } k = 1,\ldots,6.$$

This model is the dynamic analogue of the seemingly unrelated regression model.

The forecast errors for the naive model, univariate model, and the multivariate model were compared and summary statistics computed and put in Table 3.

Table 3.

Forecast error summary statistics					
Model	Mean	Median	Q_1	Q_2	σ
naive	.017	-.014	-.558	.475	1.234
univariate	-.003	-.135	-.512	.461	1.170
multivariate	.018	-.048	-.443	.448	1.130

The forecast and observed values for the two models are shown in figure 1 and figure 2 respectively. They were examined and it was found that the two largest errors occurred at

Figure 1
forecast and observed quarterly CPI inflation factors
univariate model

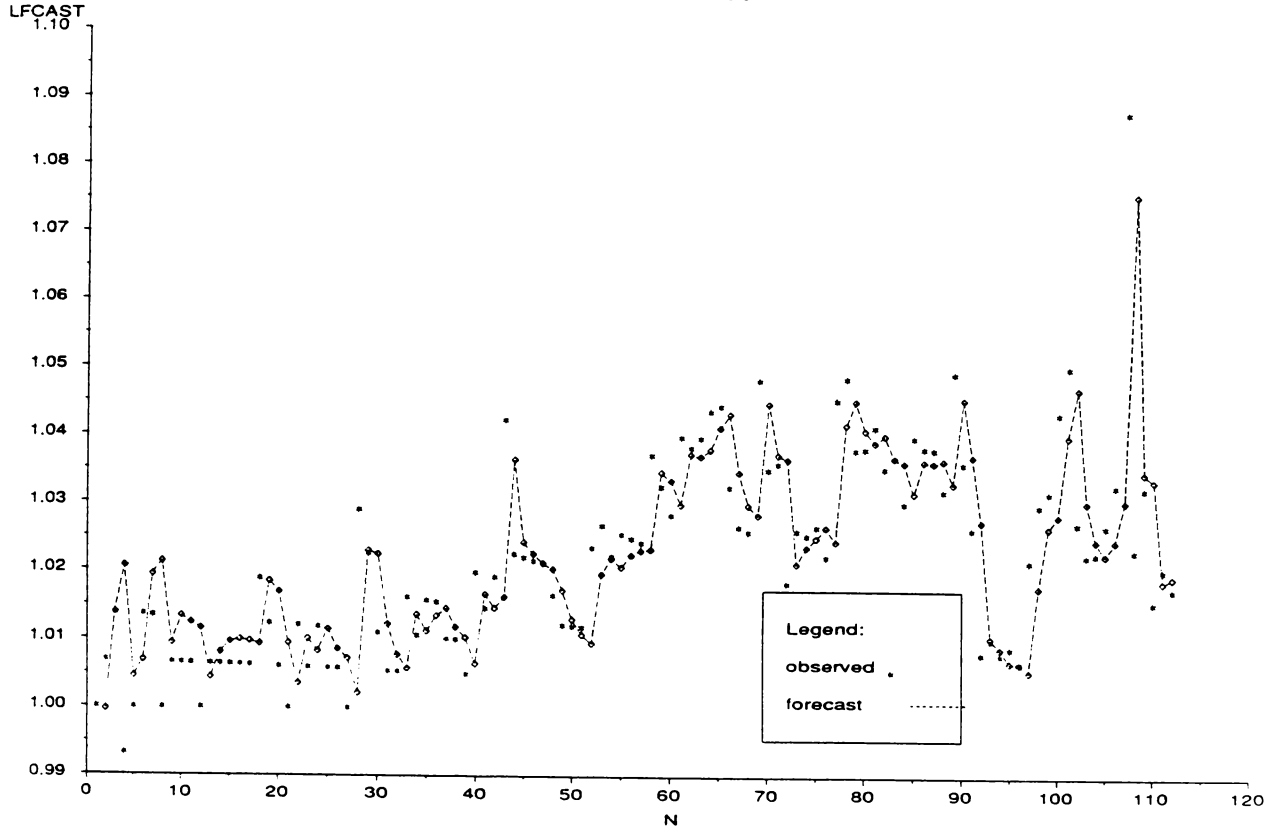
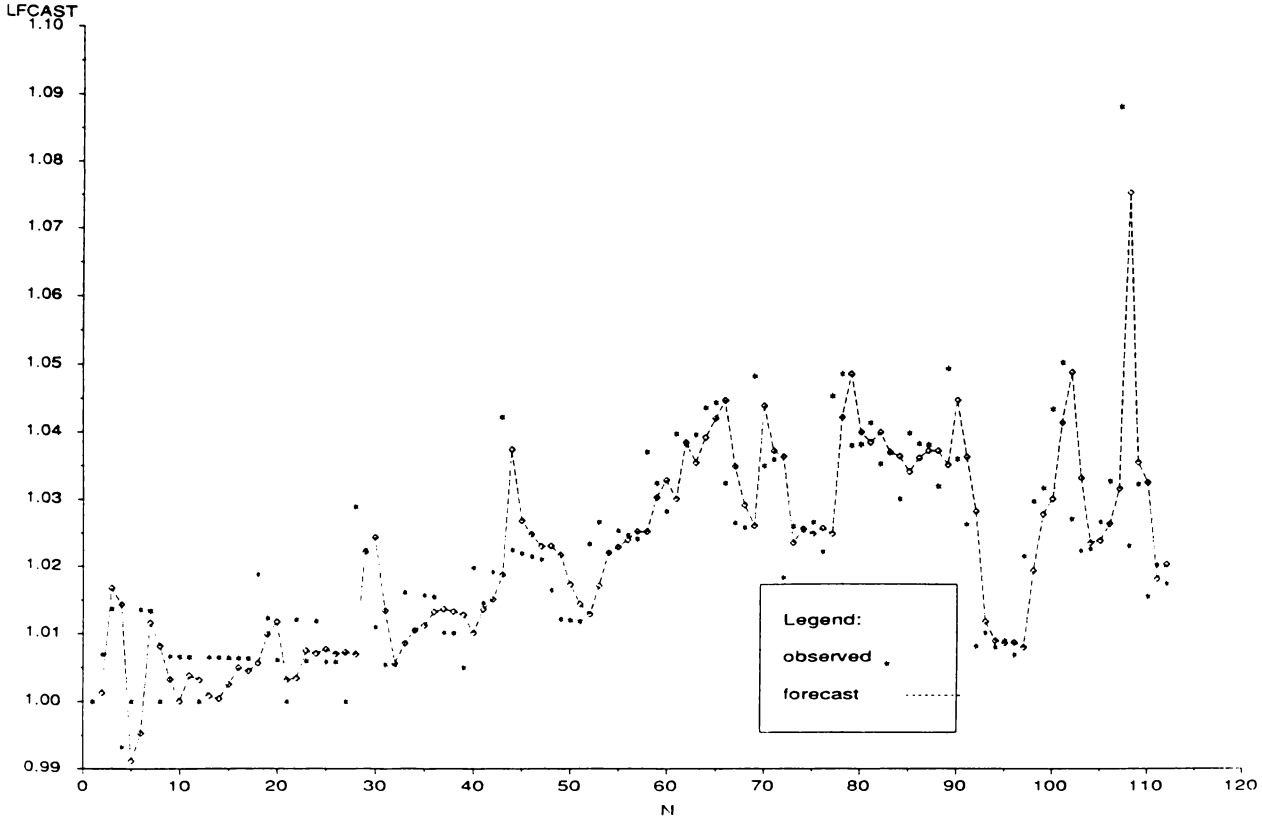


Figure 2
forecast and observed quarterly CPI inflation factors
multivariate model



quarter 107 and 108. This was caused by the large jump in the consumer price index quarter 107 due to the introduction of a 10% Goods and Services Tax. The forecasting model had been on automatic mode and hadn't been set to expect a large one-off jump that quarter that did not persist the next quarter. It was felt a more fair assessment of the model could be made by omitting these two observations. The first 10 observations were also omitted in order that any effect from the initial conditions had time to die out. The results were computed and they are shown in Table 4.

Table 4.

Revised forecast error summary statistics					
model	mean	median	Quartile 1	Quartile 3	σ
naive	.021	-.015	-.552	.448	.916
univariate	-.016	-.102	-.504	.453	.896
multivariate	.019	-.055	-.444	.413	.842

Comparing the forecast errors summary statistics for the three methods we see the multivariate multiprocess dynamic linear model gives an improvement over both the univariate dynamic linear model and the naive model. This is shown as both a smaller interquartile range and standard deviation.

5.4 The multiprocess dynamic linear model with biased perturbations applied to growth-hormone data

Hormone level concentration profiles often exhibit the characteristics of pulsatile data. Most of the time the level is decaying exponentially towards a base level; however, occasionally a pulse occurs and the concentration rises by a large amount. Both the time between pulses, and the pulse magnitudes, are unpredictable; and the sample values are subject to measurement error. In Appendix 4, Bolstad (1988 A) introduced biased perturbations into the multiprocess dynamic linear model to account for the sudden pulse in the dynamic behavior of the series. The bias allows for large jumps to be detected only in the direction of increased concentration, e.g. a pulse. The model assumes that any outlying observations are due to a pulse; the measurement errors are a sequence of independent normally distributed random errors with equal variances. Because outliers are not allowed in the model, the use of biased perturbations allows a pulse to be detected on the first subsequent observation rather than the second as in the Harrison-Stevens algorithm.

The multiprocess dynamic linear model with biased perturbations for pulsatile data used the parameter vector

$$\beta_t = \begin{bmatrix} \lambda_t \\ b \end{bmatrix}$$

where λ_t is the level at time t and b is the base level. The dynamic matrix is

$$\mathbf{A}_t = \begin{bmatrix} a & 1-a \\ 0 & 1 \end{bmatrix},$$

where a is the decay factor. The perturbation distributions for "no pulse" and "pulse" are

$$(\mathbf{r}_t \mid I_t = 1) \sim N\left[\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}\right]$$

and

$$(\mathbf{r}_t \mid I_t = 2) \sim N\left[\begin{bmatrix} \eta_t \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_\lambda^2 & 0 \\ 0 & 0 \end{bmatrix}\right].$$

The row vector of observation coefficients is $\mathbf{X}_t' = (1,0)$, the observation errors e_t are independent $N(0,\sigma_e^2)$ random variables. The variances and prior probabilities satisfy $\sigma_e^2 < \cdots < \sigma_\lambda^2$, and $\pi_t^{(1)} > \cdots > \pi_t^{(2)}$ respectively .

5.4.1 The algorithm

1. Extrapolate the parameter estimators from time $t-1$ to time t by

$$(\hat{\beta}_t \mid I_t = i, \mathbf{y}_{t-1}) = \mathbf{A}_t(\hat{\beta}_{t-1} \mid \mathbf{y}_{t-1}) + \eta_t^{(i)}$$

and the covariance matrices by

$$\mathbf{C}_t^{(i)} = \mathbf{A}_t \mathbf{V}_{t-1} \mathbf{A}_t' + \mathbf{R}_t^{(i)}$$

for $i = 1,2$.

2. Compute the predicted values of the observation at time t by

$$\hat{y}_t^{(i)} = \mathbf{X}_t'(\hat{\beta}_t \mid I_t = i, \mathbf{y}_{t-1})$$

for $i = 1,2$.

3. Compute the posterior means by

$$\mu_t^{(i)} = 0 - \mathbf{C}_t^{(i)}\mathbf{X}_t(\mathbf{X}_t'\mathbf{C}_t^{(i)}\mathbf{X}_t + \sigma_e^2)^{-1}(y_t - \hat{y}_t^{(i)})$$

and the posterior covariance matrices by

$$\mathbf{V}_t^{(i)} = \mathbf{C}_t^{(i)} - \mathbf{C}_t^{(i)}\mathbf{X}_t(\mathbf{X}_t'\mathbf{C}_t^{(i)}\mathbf{X}_t + \sigma_e^2)^{-1}\mathbf{X}_t'\mathbf{C}_t^{(i)}$$

for $i = 1,2$.

4. Compute the posterior probability of the perturbation-index variable by

$$q_t^{(i)} = \pi_t^{(i)}(f(y_t))^{-1}(2\pi)^{-1/2} \mid \mathbf{X}_t'\mathbf{C}_t^{(i)}\mathbf{X}_t + \sigma_e^2 \mid^{-1/2}$$
$$\times \exp[-1/2(\mathbf{X}_t'\mathbf{C}_t^{(i)}\mathbf{X}_t + \sigma_e^2)^{-1}(y_t - \hat{y}_t^{(i)})^2]$$

for $i = 1,2$. Note this requires the assumption of normality.

5. Condense the posterior mixture distribution into a single distribution having mean vector

$$\mathbf{m}_t = \sum_{i=1}^{2} q_t^{(i)}\mu_t^{(i)}$$

and covariance matrix

$$\mathbf{V}_t = \sum_{i=1}^{2} q_t^{(i)}[\mathbf{V}_t^{(i)} + (\mu_t^{(i)} - \mathbf{m}_t) \times (\mu_t^{(i)} - \mathbf{m}_t)'].$$

6. Correct the estimator by

$$(\hat{\beta}_t \mid \mathbf{y}_t) = (\hat{\beta}_t \mid \mathbf{y}_t) - \mathbf{m}_t.$$

7. Let $t = t + 1$ and go to step 1.

In Appendix 4, Bolstad (1988 B) used this model to analyze growth-hormone levels in experimental groups of steers subject to various levels of dietary stress. The model was found to give good agreement with the experimental data. The mean smoothed estimated growth-hormone levels and mean smoothed posterior pulse probabilities reflected the effect of the treatments on the daily growth-hormone level and pulsatile activity rates, and the synchronicity between steers. This method of analyzing pulsatile data promises to become even more fruitful when measuring techniques become automated so the effects of intervention can be monitored in real time.

5.5 The multiprocess dynamic generalized linear model

In Appendix 5, Bolstad (1988 B) developed the multiprocess extension of the dynamic generalized linear model from West et al (1985). This allows quick reacting robust multiprocess algorithms to be used in the generalized linear model setting. Here it is not the mean, but a link function of the mean that equals the linear predictor, which is itself a known linear function of the linear parameters. The linear parameter vector at time t is a known linear function of the linear parameter vector at time $t-1$ plus a zero-mean perturbation vector. The perturbation covariance matrix takes one of k known alternatives dependent on the value of the perturbation-index at that time. The perturbation indices are the outcomes of a sequence of independent multinomial trials with known prior probabilities, and the perturbations at different times are independent of each other. Thus the linear parameter vector behaves in a Markovian manner. The observation vector comes from a known one-dimensional exponential family, not necessarily normal. The sampling parameter is related to the linear parameters by the knowledge of the particular exponential family and the known link function.

The multiprocess dynamic generalized discount Bayesian model is similar to the above model, except that the covariance matrix of the linear parameters is inflated by pre and post multiplication by a discount matrix instead of being inflated by the addition of the perturbation covariance matrix. The discount matrix at time t is determined by the discount-index variable, which comes from an independent sequence of multinomial trials with known prior probabilities. This uses a technique introduced by Ameen and Harrison (1985) to overcome some practical disadvantages of using perturbations. They felt that many modellers have a more intuitive feeling for inflating the covariance matrix this way than by adding a perturbation covariance matrix.

The recursive algorithms for estimation and forecasting with these models are contained in Appendix 5. The prior distribution of the sampling parameter is chosen to be the member of the conjugate family that has the mean and variance indicated by the "guide relationship" to the linear parameters. The correction mechanism of the algorithm combines two uses of Bayes' theorem. The first is in the calculation of the posterior mean vector and covariance matrix, conditionally on each pair of previous and current perturbation-index (or discount-index) variables. The second is in the calculation of the posterior probabilities of the previous and current index variables. Hence the posterior probabilities of the index variable at time $t-1$ are calculated twice; first after the observation at time $t-1$, and again after the observation at time t. This enables the distinctions to be drawn between the different types of perturbations (discounts). The two aspects are brought together in the condensation of the mixture of k^2 distributions into a mixture of k distributions, where the weights are the posterior probabilities of the previous index variable, given the current index variable and all the observations up to and including the most recent.

When it appears that a change in the underlying process generating the series has occurred due to a particular type of perturbation, each of the k^2 conditional posterior distributions is somewhat changed due to the prediction error, the greatest change occurring in those that are conditional on that type of perturbation. The posterior probabilities of the previous perturbation-index change very little, so the condensation phase only changes the k conditional posterior distributions a modest amount. However, the posterior probabilities of the current perturbation-index values change considerably, and in particular the posterior probability of that change becomes much greater. If the next observation also indicates that the same type of change has occurred, again each of the k^2 conditional posterior distributions changes. However, in this case the posterior probabilities of the previous perturbation-index are changed considerably. Thus in the condensation phase the most weight is given to the distribution showing the greatest change, so there is a large change due to the condensation phase emphasizing that particular type of perturbation. This allows a rapid adjustment of the parameters two observations after a change in the time series occurs. Thus if the model were being used as a fault detection method, the fault would become apparent on the second subsequent observation.

5.5.1 Estimating a Dynamic Population Proportion

In this section we consider how to estimate and forecast a population proportion that evolves in the following manner. The proportion is stable most of the time, but it is subject to occasional relatively large shifts that may or may not be transitory. The population is sampled at time t, and we wish to estimate the proportion at that time, and forecast the proportion at time $t+1$ using the data up to time t. The traditional approach is to base the estimate on the data at time t only, and disregard the previous data. This does not require any knowlege of how the process evolves, since the estimate is based only on the current data, which is used to get a "snapshot" of the current proportion. If the true proportion is not changing, data from the recent past is also relevant to the current proportion. However if the true proportion has changed, the data from the recent past is no longer very relevant. The traditional approach

takes the cautious attitude to past data and ignores it, thus losing efficiency when the proportion has not changed. This cautious approach has other drawbacks as well. One is that it is particularly bad at picking up a trend because the standard error is multiplied by $2^{1/2}$. A second shortcoming of this approach occurs when the overall series consists of a linear sum of component subseries, for instance if the proportions for the population are broken down by region, age group, or sex. Estimates for these subseries are often just as important as for the series as a whole, and the sample size which is sufficient to give sufficient accuracy for the series as a whole is completely inadequate to give sufficient accuracy for the component subseries.

The multiprocess dynamic generalized linear model avoids some of these difficulties. Thus it (*i*) gives good estimates when the series is in a stable pattern, by using past data as well as present data; (*ii*) quickly reacts when a transient or a change occurs and quickly distinguishes between them; (*iii*) provides standard errors of the estimates at every step; and (*iv*) can give good estimates and forecasts for component subseries, since it doesn't require as large a sample size. Furthermore the model for the combined series is the linear sum of the component series models. Although the results of Section 3.3 cannot be directly applied to multiprocess models, it is likely that an analysis like that performed in Section 5.3.2 would show the same conclusion still holds, namely that better estimates for the aggregated series could be found by linearly combining estimates from the component subseries than by estimating from the aggregate series directly.

In Appendix 5, this model is applied to data from 50 (simulated) independent random samples of size 100, from a population where the population proportion is .35 for the first 10 samples, increases by .01 for each of the next 5 samples, continues at .4 for the next 10 samples, abruptly changes to .3 for the next 10 samples, decreases by .005 for each of the next 10 samples, and remains at .25 for the last 5 samples. Thus the sampling distribution of y_t is binomial $(100,\theta_t)$. The multiprocess dynamic generalized linear model used for this has parameter dynamic equation $\beta_t = \mathbf{A}_t\beta_{t-1}$ where

$$\beta = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} \quad \text{and} \quad \mathbf{A}_t = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

and b_0, b_1, and b_2 are the local mean, slope and transient respectively. The linear predictor is $\eta_t = \mathbf{X}_t'\beta_t$ where $\mathbf{X}_t' = (1,0,1)$. Thus the transient is included in the current linear predictor and hence the current observation, but it does not have any effect on future observations. The sampling parameter is related to the parameter vector through the linear predictor by the logit link function

$$\eta_t = g(\mu_t) = \log[\mu_t/(n-\mu_t)]$$

where $\mu_t = b(\theta_t) = n\theta_t$ is the mean of the sampling distribution of y_t, hence the guide relationship is

$$h(g(\theta_t)) = \log[\theta_t/(1-\theta_t)].$$

The four values of the perturbation-index variable and their prior probabilities are given in Table 5.

Table 5.

Perturbation type and prior probability		
perturbation-index	perturbation	prior probability
1	no change	.9
2	transient	.08
3	mean change	.015
4	slope change	.005

The perturbation covariance matrices are given by

$$\mathbf{R}^{(1)} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \mathbf{R}^{(2)} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \mathbf{R}^{(3)} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \mathbf{R}^{(1)} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & .01 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

The summary statistics for the true value, estimate, prediction, observation error, estimation error, naive prediction error $(y_t/n - \theta_{t+1})$, and the prediction error $(\hat{\theta}_{t+1} | \mathbf{y}_t - \theta_{t+1})$ are calculated and shown in Table 6.

Table 6.

Multiprocess dynamic generalized linear model summary statistics								
variable	formula	mean	med	min	max	q1	q3	σ
value	y_t/n	.331	.325	.200	.430	.280	.383	.0588
estimate	$\hat{\theta}_t \mid \mathbf{y}_t$	.335	.328	.257	.401	.299	.384	.0459
prediction	$\hat{\theta}_{t+1} \mid \mathbf{y}_t$	.332	.326	.254	.409	.280	.385	.0496
observation error	$y_t/n - \theta_t$	.004	.000	-.060	.120	-.023	.030	.0400
estimation error	$\hat{\theta}_t \mid \mathbf{y}_t - \theta_t$	-.004	.001	-.080	.074	-.031	.017	.0327
naive prediction error	$y_t/n - \theta_{t+1}$	-.006	-.000	-.120	.060	-.038	.028	.0442
prediction error	$\hat{\theta}_{t+1} \mid \mathbf{y}_t - \theta_{t+1}$	-.007	-.012	-.093	.084	-.024	.012	.0363

The multiprocess dynamic generalized linear model algorithms extend the attributes of multiprocess models, into non-stationary time series where the sampling distribution is not normal, rather a known member of the one-dimensional exponential family of distributions; and it is some function of the mean, rather than the mean itself that is a linear function of the dynamic parameters. The use of the conjugate family of prior distributions enables computationally simple recursions similar to those of the Kalman filter to be developed, conditionally on the perturbation-index variables. The condensing of the mixture of k^2 posterior distributions

into a mixture of k posterior distributions enables the algorithm to react quickly to abrupt real changes, yet be insensitive to transients.

5.6 Summary, conclusions, and future research

In this chapter we have developed some forecasting algorithms which adapt multiprocess models for time series of various types, including correlated multivariate normal observations, pulsatile observations, and non-normal observations from a one-dimensional exponential family. These algorithms have the advantages that characterize Harrison-Stevens forecasting, namely quick reaction to real change, and robustness against outliers.

The multivariate multiprocess dynamic linear model was used to forecast the quarterly inflation rate in the overall consumer price index using the inflation rates for the main consumer price groups. In this example, it was shown to be effective in forecasting the quarterly inflation rate, which is a very noisy time series.

The multiprocess dynamic linear model with biased perturbations was developed. In Appendix 4, it was used to model growth-hormone level in young cows subject to various levels of dietary stress. This model was found to be a very effective tool for analyzing this data which is representative of noisy *pulsatile* data. The posterior pulse probabilities and smoothed estimated growth-hormone levels when averaged over cows under the same treatment, gave evidence of the treatment effect on the diurnal cycle of pulsatile activity and hormone levels.

The multiprocess dynamic generalized linear model was developed to forecast where the observation distribution is from a known one-dimensional exponential family distribution (not necessarily normal), and a function of the mean (not necessarily the mean itself) is a linear function of the dynamic parameters. In Appendix 5, the algorithm was demonstrated on a simulated time series of a dynamic population proportion. It was shown to be very effective in estimating the proportion, adapting quickly to a real shift, but being robust against outliers (transients). This algorithm will be particularly useful when the population is broken down into subgroups, since the subgroup sample sizes are often inadequate to get sufficiently accurate estimates the subgroup proportions in a non-dynamic way.

Future plans include adapting the observation score functions from Masreliez (1975) to multiprocess models. This will be done conditionally on the perturbation indices. These will be condensed to get the influence functions. The shape of these influence functions will depend on the values of the perturbation prior probabilities and variances chosen. This will help in finding sensible choices for these values.

6. Bibliography

Ameen, J. R. M., and Harrison, P. J. (1985), "Normal Discount Bayesian Models", in Bayesian Statistics 2, eds. J. M. Bernardo, M. H. Degroot, D. V. Lindley, and A. F. M. Smith, Elsevier Science Publishers, Amsterdam; North Holland .

Aoki, M. (1987), "State Space Modeling of Time Series", Springer-Verlag, Berlin, Heidelberg, New York.

Birkhoff, G. D. (1931), "Proof of the Ergodic Theorem", Proc. Natl. Acad. Sci. USA 17, pp. 656-660.

Bolstad, W. M. (1986 A), "Harrison-Stevens Forecasting and the Multiprocess Dynamic Linear Model", The American Statistician, Vol. 40, pp. 129-135.

Bolstad, W. M. (1986 B), "An Efficient Algorithm for Harrison-Stevens Forecasting Using the Multiprocess Multivariate Dynamic Linear Model", Communications in Statistics: Simulation and Computation, Vol. B15, No. 3, pp. 819-828.

Bolstad, W. M. (1987), "An Estimation Method for the Seemingly Unrelated Regression Model with Contemporaneous Covariances Based on an Efficient Recursive Algorithm", Communications in Statistics: Simulation and Computation, Vol. B16, No. 3, pp. 689-698.

Bolstad, W. M., (1988 A), "The Multiprocess Dynamic Linear Model with Biased Perturbations: A Real Time Model for Growth Hormone Level", Biometrika, Vol. 75, No. 4, pp. 685-692.

Bolstad, W. M. (1988 B), "Estimation in the Multiprocess Dynamic Generalized Linear Model", Communications in Statistics: Theory and Methods Vol. 17, No. 12, pp. 4179-4204.

Bolstad, W. M., (1990), "The Efficiency of Dynamic Linear Model Estimators Applied to a Linearly Combined Time Series", Communications in Statistics: Theory and Methods, Vol. 19, No. 1,

Breiman, L. (1969), "Probability and Stochastic Processes: With a View Toward Applications", Houghton Mifflin, Boston.

Chatfield, C. (1980), "The Analysis of Time Series: An Introduction", Chapman and Hall, London

Cramer, H. (1942), "On Harmonic Analysis in Certain Functional Spaces", Ark. Mat. Astron. Fys. 283 No. 1.

Daniell, P. J. (1946), Discussion following "On the Theoretical Specification and Sampling Properties of Autocorrelated Time Series" by M. S. Bartlett, J. Roy. Statist. Soc. Suppl. 8, pp. 27-41.

Duncan, B., and Horn, S. D. (1972), "Linear Dynamic Recursive Estimation From the Viewpoint of Regression Analysis", JASA 67, pp. 815-821.

Fisher, R. A. (1929), "Tests of Significance in Harmonic Analysis", Proc. Roy. Soc. Series A., Vol. 125, pp. 54-59.

Guttman, I., and Pena, D. (1985), "Robust Filtering", JASA 80, pp. 91-92.

Harrison, P. J., and Stevens, C. F. (1976). "Bayesian Forecasting" (With Discussion)", J. Roy. Statist. Soc. Series B, Vol. 38, pp. 205-247.

Harrison, P. J., and Stevens, C. F. (1971), "A Bayesian Approach to Short-term Forecasting", Operational Research Quarterly Vol. 22, pp. 341-362.

Harvey, A. C., and Phillips, G. D. A. (1979), "Maximum Likelihood Estimation in Regression Models with Autoregressive Moving Average Disturbances", Biometrika 66, pp.49-58.

Harville, D. A. (1980), "Predictors for National Football League Games via Linear Model Methodology", JASA 75, pp. 516-524.

Ho, Y. C., and Lee, R. C. K. (1964), "A Bayesian Approach to Problems in Stochastic Estimation and Control", IEEE Transactions in Automatic Control, pp. 33-39.

Kalman, R. E. (1960), "A New Approach to Linear Filtering and Prediction Problems", Journal of Basic Engineering, pp. 35-45.

Kalman, R. E., and Bucy, R. S. (1961), "New Results in Linear Filtering and Prediction Theory", Journal of Basic Engineering, pp. 95-108.

Khintchine, A. (1934), "Korrelationstheorie der Stationare Stochastichen Processe", Math. Ann. 109, pp. 604-615.

Kitagawa, G. (1987), "Non-Gaussian State-Space modelling of Non-stationary Time Series (with Comments)", JASA Vol. 82, pp. 1032-1063.

Kohn, R. (1982), "When is an Aggregate of a Time Series Efficiently Forecast by its Past", Journal of Econometrics 18, pp. 337-350.

Kolmogorov, A. N. (1933) "Grundbegriffe der Wahrscheinlichkeitrechnung", Ergebnisse der Mathematik. (Published in English as "Foundations of the Theory of Probability", Chelsea, Bronx, New York.)

Koopmans, L. H. (1974), "The Spectral Analysis of Time Series", Academic Press, New York.

Makov, U. E. (1983), "Approximate Bayesian Procedures for Dynamic Linear Models in the Presence of Jumps", The Statistician 32, pp. 207-213.

Martin, R. D., and Raftery, A. (1987), "Comments on Non-Gaussian State-Space Modeling of Nonstationary Time Series", JASA 82, pp. 1044-1050.

Masreliez, C. J. (1975), "Approximate Non-Gaussian Filtering with Linear State and Observation Relations", IEEE Transactions on Automatic Control, pp. 107-110.

Masricliez, C. J., and Martin, R. D. (1977), "Robust Bayesian Estimation for the Linear Model and Robustifying the Kalman Filter", IEEE Transactions on Automatic Control 22, pp. 361-371.

Meinhold, L. R., and Singpurawalla, N. (1983), "Understanding the Kalman Filter", The American Statistician vol. 37, pp. 207-213.

Meinhold, L. R., and Singpurawalla, N. (1989), "Robistification of Kalman Filter Models", JASA 84, pp. 479-486. XP Nelder, J. A., and Wedderburn, R. W. M. (1972), "Generalized Linear Models", J. Roy. Statist. Soc. Series A, Vol. 35, pp. 370-384.

Parzen, E. (1958), "Conditions that a Stochastic Process be Ergodic", Ann. Math. Stat. Vol. 29, pp. 299-301.

Parzen, E. (1962), "Stochastic Processes", Holden-Day, San Francisco.

Pena,D., and Guttman, I. (1989), "Optimal Collapsing of Mixture Distributions in Robust Recursive Estimation", Communications in Statistics: Theory and Methods, Vol. 18, No. 3, pp. 817-834.

Sallas, W. M., and Harville, D. A. (1981), "Best Linear Recursive Estimation for Mixed Linear Models", JASA 76, pp. 860-869.

Schuster, A. (1898), "On the Investigation of Hidden Periodicities with Application to a Supposed 26-day Period of Meteorological Phenomena", Terrestial Magnetism, Vol. 3, pp. 13-41.

Smith, A. F. M., West, M. (1983), "An Application of Multiprocess Kalman Filter to Renal Transplant Data", Biometrics 39, pp. 867-878.

Smith, A. F. M., West, M., Gordon, K., Knapp, M., and Trimble, I. C. (1983), "Monitoring Kidney Transplant Patients", Statistician 32, pp. 46-54.

Smith, R. L. and Miller, J. E. (1986), "A Non-Gaussian State Space Model and Application to Prediction of Records", J. Roy. Statist. Soc. Series B, Vol. 48, No. 1, pp. 79-88.

West, M. (1985), "Generalized Linear Models; Scale Parameters, Outlier Accomodation and Prior Distributions", in Bayesian Statistics 2, eds J. M. Bernardo, M. H. DeGroot, D. V. Lindley, and A. F. M. Smith, Amsterdam: North Holland.

West, M., Harrison, P. J., and Migon, H. S. (1985), "Dynamic Generalized Linear Models and Bayesian Forecasting (With Discussion)", JASA 80, pp. 73-97.

Wiener, N. (1930), "Generalized Harmonic Analysis", Acta. Math. 99, pp. 117-258.

Wiener, N. (1949), "Extrapolation, Interpolation, and Smoothing of Stationary Time Series with Engineering Applications", John Wiley & Sons, New York.

Willsky, A. S., (1976), "A Survey of Design Methods for Failure Detection in Dynamic Systems", Automatica 12, pp. 601-611.

Wold, H., (1938), "A Study of Stationary Time Series", Almquist and Wiksell, Stockholm.

Zadeh, L. and Desoer, C. (1963) "Linear System Theory", McGraw Hill, New York.

APPENDIX 1
Communications in Statistics: Theory and Methods
Vol. 19, No, 1

(To appear)

THE EFFICIENCY OF DYNAMIC LINEAR MODEL ESTIMATORS APPLIED TO A LINEARLY AGGREGATED TIME SERIES

William M. Bolstad

Department of Mathematics and Statistics
University of Waikato
Hamilton, New Zealand

ABSTRACT

Many time series are weighted sums of component time series. If the same dynamic linear model generates all of the component time series, it also generates the linearly aggregated time series. This paper establishes the result that if the prior distribution of the dynamic parameter estimators is the same for the dynamic linear model applied to the linearly aggregated time series as for the weighted sum of the component dynamic linear model estimators, the posterior covariance matrix for the linearly aggregated model estimators equals the posterior covariance matrix for the weighted sum of the component model estimators plus the weighted sum of the covariance matrices of the posterior mean vectors averaged over all possible component observations such that the weighted sum is the linearly aggregated observation. This paper determines the relative efficiency of the two estimators and shows that the generalized variance of the posterior estimators for the linearly aggregated model is always greater than or equal to the generalized variance of the weighted sum of component posterior estimators.

Key Words and Phrases: Kalman filter, state vector, Markov property.

1. INTRODUCTION

In a linear dynamic system, the stochastic behavior of the observations is described in terms of an unobserved state vector whose dynamic behavior is Markovian, and consists of a linear deterministic mean plus a random perturbation. Kalman (1960) developed an algorithm for the recursive estimation of the state variables of a linear dynamic system, and he showed it to be the optimal linear filter for both symmetric convex and squared error loss functions. The dynamic linear model introduced by Duncan and Horn (1972) is a random β regression model where the regression coefficients change with time in a manner that has a linearly deterministic part and a Markovian part. They related the time varying parameters to the state variables, and showed the Kalman filter estimates are the minimum variance linear unbiased estimates and also the minimum mean square linear estimates of the parameters at the current time, given all observations up to and including the present. The dynamic linear model and the Kalman filter can operate under either of two alternative sets of assumptions: either the normal assumptions where all the variables are assumed normal with known first two moments, or the less restrictive assumptions where only the first two moments are known for all the variables. The optimal properties mentioned above hold under both sets of assumptions. Additionally, the posterior distribution of the estimators are normal, and the Kalman filter estimates are the minimum variance unbiased and minimum mean squared estimates under the normal assumptions.

Many time series are weighted sums of component time series, for example the consumer price index is a weighted sum of consumer price subindices. If each of these components is considered to be an observation sequence from the same dynamic linear model (same dynamic coefficient matrix and same observation coefficient vector) then the linearly aggregated time series is also an observation sequence from the same model. This would be the case when the local behavior of the components are all being modelled by the same dynamic linear model. There are two possible approaches for estimating the parameters of the linearly aggregated series: using the Kalman filter estimates from the linearly aggregated series, or taking the weighted sum of the Kalman filter estimates from each component series. This paper determines the relative efficiencies of the two methods, and establishes that the posterior covariance matrix for the linearly aggregated model is at least as large as the posterior

covariance matrix for the weighted sum of the component models, and it is a strict inequality unless all the component series are completely correlated. Thus in all realistic situations it is preferable to estimate the parameters of each component model separately and then linearly combine the estimates. Kohn (1982) investigates conditions in which the estimate using the linearly aggregated series is efficient for the special cases of ARIMA and stationary Gaussian models.

2. THE DYNAMIC LINEAR MODEL AND THE KALMAN FILTER

In the dynamic linear model the parameter vector changes through time according to the parameter dynamic equation

$$\boldsymbol{\beta}_t = \mathbf{A}_t \boldsymbol{\beta}_{t-1} + \mathbf{r}_t \qquad [1]$$

where $\boldsymbol{\beta}_t$ is the parameter vector at time t, $\mathbf{A}_t$ is a known matrix of coefficients at time t, and $\mathbf{r}_t$, the perturbation vector at time t, is normally distributed with mean vector 0 and known covariance matrix $\mathbf{R}_t$. The perturbation vectors at different times are independent of each other.

The observation y_t is governed by the observation equation

$$y_t = \mathbf{X}_t' \boldsymbol{\beta}_t + e_t \qquad [2]$$

where $\mathbf{X}_t'$ is the known row vector of observation coefficients at time t, and the observation errors e_t are independent identically distributed $N(0,\sigma^2)$ random variables which are independent of the perturbation random vectors.

At time $t-1$ the initial condition for the Kalman filtering algorithm is that there is an unbiased parameter estimator vector, conditional on $\mathbf{y}_{t-1} = y_{t-1}, y_{t-2}, \cdots, y_1$, the set of observations up to time $t-1$. Its distribution is assumed to be normal with mean vector equal to the parameter vector and to have a known covariance matrix. Then

$$(\hat{\boldsymbol{\beta}}_{t-1} - \boldsymbol{\beta}_{t-1} \mid \mathbf{y}_{t-1}) \quad \text{is} \quad N(0, \mathbf{V}_{t-1}). \qquad [3]$$

At time t the perturbation vector $\mathbf{r}_t$ is $N(0,\mathbf{R}_t)$. The prior estimator of the parameter vector at time t is

$$(\hat{\boldsymbol{\beta}}_t \mid \mathbf{y}_{t-1}) = \mathbf{A}_t (\hat{\boldsymbol{\beta}}_{t-1} \mid \mathbf{y}_{t-1}), \qquad [4]$$

the conditional parameter estimator vector at time $t-1$ brought forward to time

t, but prior to observation y_t. The distribution of the conditional random variable $(\hat{\boldsymbol{\beta}}_t - \boldsymbol{\beta}_t \mid \mathbf{y}_{t-1})$ is $N(0, \mathbf{C}_t)$ where the covariance matrix $\mathbf{C}_t = \mathbf{A}_t \mathbf{V}_{t-1} \mathbf{A}_t' + \mathbf{R}_t$. The predicted value of the observation at time t is

$$\hat{y}_t = \mathbf{X}_t'(\hat{\boldsymbol{\beta}}_t \mid \mathbf{y}_{t-1}). \qquad [5]$$

The joint random variables

$$\begin{bmatrix} \hat{\boldsymbol{\beta}}_t - \boldsymbol{\beta}_t \mid \mathbf{y}_{t-1} \\ y_t - \hat{y}_t \mid \mathbf{y}_{t-1} \end{bmatrix} \quad \text{are} \quad \mathbf{N}\left[\begin{bmatrix} 0 \\ 0 \end{bmatrix}; \begin{bmatrix} \mathbf{C}_t & -\mathbf{C}_t \mathbf{X}_t \\ -\mathbf{X}_t' \mathbf{C}_t & \mathbf{X}_t' \mathbf{C}_t X_t + \sigma^2 \end{bmatrix}\right]. \qquad [6]$$

Hence the conditional random variable

$$(\hat{\boldsymbol{\beta}}_t - \boldsymbol{\beta}_t \mid \mathbf{y}_t) \text{ is } N(\boldsymbol{\mu}_t; \mathbf{V}_t) \qquad [7]$$

where

$$\boldsymbol{\mu}_t = 0 - \mathbf{C}_t \mathbf{X}_t (\mathbf{X}_t' \mathbf{C}_t \mathbf{X}_t + \sigma^2)^{-1} (y_t - \hat{y}_t) \qquad [7A]$$

and

$$\mathbf{V}_t = \mathbf{C}_t - \mathbf{C}_t \mathbf{X}_t (\mathbf{X}_t' \mathbf{C}_t \mathbf{X}_t + \sigma^2)^{-1} \mathbf{X}_t' \mathbf{C}_t. \qquad [7B]$$

The estimator vector $(\hat{\boldsymbol{\beta}}_t \mid \mathbf{y}_t)$ is no longer an unbiased estimator of $\boldsymbol{\beta}_t$, so we correct the estimator by subtracting the mean vector to give the posterior estimator

$$(\hat{\boldsymbol{\beta}}_t \mid \mathbf{y}_t) = (\hat{\boldsymbol{\beta}}_t \mid \mathbf{y}_t) - \boldsymbol{\mu}_t. \qquad [8]$$

At this point we have found the unbiased parameter estimator vector conditional on all the observations up to and including the present. We are now in the same position as we were when the algorithm started so we are ready to repeat the algorithm when the next observation becomes available.

3. THE DYNAMIC LINEAR MODEL FOR A LINEARLY AGGREGATED TIME SERIES

Suppose that there are k independent time series $y_t^{(1)}, \cdots, y_t^{(k)}$ each of which is the observation sequence from the same dynamic linear model. That is

$$y_t^{(i)} = \mathbf{X}_t' \boldsymbol{\beta}_t^{(i)} + e_t^{(i)} \quad \text{and} \quad \boldsymbol{\beta}_t^{(i)} = \mathbf{A}_t \boldsymbol{\beta}_{t-1}^{(i)} + \mathbf{r}_t^{(i)} \qquad [9]$$

where all the dynamic matrices and the observation coefficient vectors are respectively equal. The observation errors $e_t^{(i)}$ and the perturbations errors $\mathbf{r}_t^{(i)}$ are independent identically distributed $\mathbf{N}[0; \sigma_t^2]$ and $\mathbf{N}[0; \mathbf{R}_t^{(i)}]$ random variables

respectively. Assume that the models generating the the series are all independent of each other, that is the observation errors and perturbations are all independent for $i \neq j$. Let $w^{(i)}$ for $i = 1, \cdots, k$ be a set of positive weights that sum to 1. Then

$$y_t^A = \sum_{i=1}^{k} w^{(i)} y_t^{(i)} \quad \text{and} \quad \beta_t^A = \sum_{i=1}^{k} w^{(i)} \beta_t^{(i)} \qquad [10]$$

are the linearly aggregated time series and the linearly aggregated parameter series respectively. Clearly

$$y_t^A = \mathbf{X}_t' \beta_t^A + e_t^A \quad \text{and} \quad \beta_t^A = \mathbf{A}_t \beta_{t-1}^A + \mathbf{r}_t^A \qquad [11]$$

where $\mathbf{r}_t^A = \sum_{i=1}^{k} w^{(i)} \mathbf{r}_t^{(i)}$ and $e_t^A = \sum_{i=1}^{k} w^{(i)} e_t^{(i)}$. Thus the linearly aggregated time series also comes from the same dynamic linear model where the perturbation covariance matrix is $\mathbf{R}_t^A = \sum_{i=1}^{k} (w^{(i)})^2 \mathbf{R}_t^{(i)}$ and the observation variance is $\sigma_L^2 = \sum_{i=1}^{k} (w^{(i)})^2 \sigma_i^2$.

Theorem: Let $\hat{\beta}_t^{(i)}$ be the unbiased estimators of $\beta_t^{(i)}$ prior to taking the observations $y_t^{(i)}$ for $i = 1,\ldots,k$ respectively. Assume the prior prediction errors, $\hat{\beta}_t^{(i)} - \beta_t^{(i)}$ for $i = 1,\ldots,k$ are all independent of each other. Let $\hat{\beta}_t^A$ be the prior estimator for the linearly aggregated time series, and $\hat{\beta}_t^W = \sum_{i=1}^{k} w^{(i)} \hat{\beta}_t^{(i)}$ be the weighted sum of the prior estimators of component series. Assume that the prior distributions of the estimation errors $\hat{\beta}_t^A - \beta_t^A | \, \mathbf{y}_{t-1}^A$ and $\hat{\beta}_t^W - \beta_t^A | \, \mathbf{y}_{t-1}^{(1)}, \cdots, \mathbf{y}_{t-1}^{(k)}$ are identical. Then

$$COV(\hat{\beta}_t^A - \beta_t^A | \, \mathbf{y}_t^A) = COV(\hat{\beta}_t^W - \beta_t^A | \, \mathbf{y}_t^{(1)}, \cdots, \mathbf{y}_t^{(k)}) \qquad [12]$$

$$+ \sum_{i=1}^{k} (w^{(i)})^2 [\mathbf{C}_t^{(i)} \mathbf{X}_t (\mathbf{X}_t' \mathbf{C}_t^{(i)} \mathbf{X}_t + \sigma_i^2)^{-1} \mathbf{X}_t' \mathbf{C}_t^{(i)} - (w^{(i)})^2 \mathbf{C}_t^{(i)} \mathbf{X}_t [\sum_{j=1}^{k} (w^{(j)})^2 (\mathbf{X}_t' \mathbf{C}_t^{(i)} \mathbf{X}_t + \sigma_j^2)]^{-1} \mathbf{X}_t' \mathbf{C}_t^{(i)}]$$

Proof:

The covariance matrices posterior to observing y_t^A are identical. Hence

$$COV(\hat{\beta}_t^A - \beta_t^A | \, y_t^A) = E[(\sum_{i=1}^{k} w^{(i)} (\hat{\beta}_t^{(i)} - \beta_t^{(i)} | \, y_t^A))(\sum_{j=1}^{k} w^{(j)} (\hat{\beta}_t^{(j)} - \beta_t^{(j)} | \, y_t^A))'] \qquad [13]$$

$$- E(\sum_{i=1}^{k} w^{(i)} (\hat{\beta}_t^{(i)} - \beta_t^{(i)} | \, y_t^A)) \times E(\sum_{j=1}^{k} w^{(j)} (\hat{\beta}_t^{(j)} - \beta_t^{(j)} | \, y_t^A))'$$

$$= \sum_{i=1}^{k}\sum_{j=1}^{k} w^{(i)}w^{(j)}E\,[E\,((\hat{\beta}_t^{(i)}-\beta_t^{(i)}|\,y_t^{(i)})\times(\hat{\beta}_t^{(j)}-\beta_t^{(j)}|\,y_t^{(j)})')|\,y_t^{A}]$$

$$- \sum_{i=1}^{k}\sum_{j=1}^{k} w^{(i)}w^{(j)}E\,(\hat{\beta}_t^{(i)} - \beta_t^{(i)}|\,y_t^{A})\times E\,(\hat{\beta}_t^{(j)} - \beta_t^{(j)}|\,y_t^{A})'.$$

Since all the prior distributions and all the observation errors for the component series are assumed to be independent, the terms for $i \neq j$ cancel. Thus

$$COV\,(\hat{\beta}_t^{A}-\beta_t^{A}|\,y_t^{A}) = \sum_{i=1}^{k}(w^{(i)})^2E\,[E\,((\hat{\beta}_t^{(i)}-\beta_t^{(i)}|\,y_t^{(i)})\times(\hat{\beta}_t^{(i)}-\beta_t^{(i)}|\,y_t^{(i)})')|\,y_t^{A}] \qquad [14]$$

$$- \sum_{i=1}^{k}(w^{(i)})^2E\,[E\,(\hat{\beta}_t^{(i)}-\beta_t^{(i)}|\,y_t^{(i)})|\,y_t^{A}]\times E\,[E\,(\hat{\beta}_t^{(i)}-\beta_t^{(i)}|\,y_t^{(i)})|\,y_t^{A}]'.$$

Subtracting and adding

$$\sum_{i=1}^{k}(w^{(i)})^2E\,[(E\,(\hat{\beta}_t^{(i)}-\beta_t^{(i)}|\,y_t^{(i)})\times(E\,(\hat{\beta}_t^{(i)}-\beta_t^{(i)}|\,y_t^{(i)}))'|\,y_t^{A}] \qquad [14A]$$

gives the result

$$COV\,(\hat{\beta}_t^{A}-\beta_t^{A}|\,y_t^{A}) = \sum_{i=1}^{k}(w^{(i)})^2E\,[COV\,(\hat{\beta}_t^{(i)}-\beta_t^{(i)}|\,y_t^{(i)})|\,y_t^{A}] \qquad [15]$$

$$+ \sum_{i=1}^{k}(w^{(i)})^2COV\,[E\,(\hat{\beta}_t^{(i)}-\beta_t^{(i)}|\,y_t^{(i)})|\,y_t^{A}].$$

The posterior covariance matrices do not depend on the y values, so their expectations equal themselves. Thus

$$COV\,(\hat{\beta}_t^{A}-\beta_t^{A}|\,\mathbf{y}_t^{A}) = \sum_{i=1}^{k}(w^{(i)})^2COV\,(\hat{\beta}_t^{(i)}-\beta_t^{(i)}|\,y_t^{(i)}) \qquad [16]$$

$$+ \sum_{i=1}^{k}(w^{(i)})^2COV\,[\mathbf{C}_t^{(i)}\mathbf{X}_t(\mathbf{X}_t'\mathbf{C}_t^{(i)}\mathbf{X}_t+\sigma_i^2)^{-1}(y_t^{(i)}-\hat{y}_t^{(i)}|\,y_t^{A})].$$

The first term is the covariance matrix of the weighted sum of the component estimators. When we use a conditional covariance inversion formula similar to equation [7B] on the second term we get the result

$$COV\,(\hat{\beta}_t^{A}-\beta_t^{A}|\,\mathbf{y}_t^{A}) = COV\,(\hat{\beta}_t^{W}-\beta_t^{A}|\,\mathbf{y}_t^{(1)},\cdots,\mathbf{y}_t^{(k)}) \qquad [17]$$

$$+ \sum_{i=1}^{k}(w^{(i)})^2[\mathbf{C}_t^{(i)}\mathbf{X}_t(\mathbf{X}_t'\mathbf{C}_t^{(i)}\mathbf{X}_t+\sigma_i^2)^{-1}\mathbf{X}_t'\mathbf{C}_t^{(i)} - (w^{(i)})^2\mathbf{C}_t^{(i)}\mathbf{X}_t[\sum_{j=1}^{k}(w^{(j)})^2(\mathbf{X}_t'\mathbf{C}_t^{(j)}\mathbf{X}_t+\sigma_j^2)]^{-1}\mathbf{X}_t'\mathbf{C}_t^{(i)}]$$

Q.E.D.

The second term is the weighted sum of the covariance matrices of the posterior component mean vectors averaged over all possible component observations $y_t^{(1)}, \cdots, y_t^{(k)}$ such that $\sum_{i=1}^{k} w^{(i)} y_t^{(i)} = y_t^A$. It is always non-negative definite, hence the efficiency of the weighted sum of the component estimators is always greater than or equal to that of the estimator from the linearly aggregated time series because it has a smaller Wilk's generalized variance. This can also be seen to be a consequence of the projection theorem. The set of observations on the linearly aggregated time series is a subspace of the set of observations on each of the component time series. The length of the perpendicular to the subspace must be at least as long as the length of the perpendicular to the total observation space.

The covariance of $\mathbf{C}_t^{(i)}\mathbf{X}_t(\mathbf{X}_t'\mathbf{C}_t^{(i)}\mathbf{X}_t + \sigma_i^2)^{-1}(y_t^{(i)} - \hat{y}_t^{(i)} \mid y_t^A)$, the i'th posterior mean vector, equals 0 if and only if $y_t^{(i)} - \hat{y}_t^{(i)}$ has covariance matrix identically equal the zero matrix, or if it and $y_t^A - \hat{y}_t^A$ are perfectly correlated. Hence, for all realistic cases, the inequality is a strict one and it is always more efficient to estimate the parameters of each of the component models separately, and then linearly combine the estimates to get the estimates for the linearly aggregated model parameters. The relative efficiency of the linearly aggregated model relative to the weighted sum of the component models is given by

$$eff(\hat{\beta}_t^A : \hat{\beta}_t^W) = COV(\hat{\beta}^W \mid \mathbf{y}_t^{(1)}, \cdots, \mathbf{y}_t^{(k)}) \times [COV(\hat{\beta}^W \mid \mathbf{y}_t^{(1)}, \cdots, \mathbf{y}_t^{(k)}) + \qquad [18]$$

$$\sum_{i=1}^{k} (w^{(i)})^2 [\mathbf{C}_t^{(i)}\mathbf{X}_t(\mathbf{X}_t'\mathbf{C}_t^{(i)}\mathbf{X}_t + \sigma_i^2)^{-1}\mathbf{X}_t'\mathbf{C}_t^{(i)} - (w^{(i)})^2 \mathbf{C}_t^{(i)}\mathbf{X}_t [\sum_{j=1}^{k} (w^{(j)})^2 (\mathbf{X}_t'\mathbf{C}_t^{(j)}\mathbf{X}_t + \sigma_j^2)]^{-1}\mathbf{X}_t'\mathbf{C}_t^{(i)}]]^{-1}$$

4. SUMMARY

Kalman (1960) showed that the problems of forecasting, filtering, and smoothing are all equivalent to estimating the unobservable value of a signal at some time point given the signal plus noise over some time period. Since the Kalman filter applies to all these problems, the result shown in this paper for the filtering problem applies to the forecasting and smoothing problems as well.

The conclusion that the estimator from the linearly aggregated time series is not as good an estimator for the weighted sum of parameters as the weighted sum of component estimators holds even if the dynamic linear models aren't independent. In that case, the cross terms where $i \neq j$ in equation [13] would not cancel, so the equivalent to equation [13] would be a double sum over i

and j. So a double sum cross term similar to equation [14A] would be subtracted and added, yielding the equations equivalent to [15] and [16] having two double sums on the right hand side. Thus the posterior covariance matrix for the linearly aggregated model estimators would still equal the posterior covariance matrix for the weighted sum of the component model estimators plus the weighted sum of the covariance matrices of the posterior mean vectors averaged over all possible component observations such that the weighted sum is the aggregated observation. Thus the formula for the relative efficiency would be somewhat more complicated.

ACKNOWLEGEMENT

The Author wishes to thank the referee for his useful comments.

BIBLIOGRAPHY

Duncan, B., and Horn, S. D. (1972), "Linear Dynamic Recursive Estimation From the Viewpoint of Regression Analysis", JASA 67, pp. 815-821.

Kalman, R. E. (1960), "A New Approach to Linear Filtering and Prediction Problems", Journal of Basic Engineering, pp. 95-108.

Kohn, R. (1982), "When is an Aggregate of a Time Series Efficiently Forecast by its Past", Journal of Econometrics 18, pp. 337-350.

Harrison–Stevens Forecasting and the Multiprocess Dynamic Linear Model

WILLIAM M. BOLSTAD*

This is an expository article. The Harrison–Stevens forecasting algorithm using the multiprocess dynamic linear model is a robust method for forecasting in a nonstationary time series. The purpose of this article is to help statisticians become familiar with the method.

KEY WORDS: Bayesian forecasting; Kalman filter; Mixture of distributions; Multiprocess Kalman filter; State variable.

1. INTRODUCTION

Harrison–Stevens forecasting is a technique devised by Harrison and Stevens (1971) as a tool for short-term forecasting in a time series in which there may be unknown underlying changes in the process generating the series. In their paper they referred to it as Bayesian forecasting, but I will not use that term, since it could lead to confusion because there are other Bayesian techniques used in forecasting. For example, Meinhold and Singpurwalla (1983) showed that the Kalman filter can be considered a Bayesian technique, using Bayes's theorem in the parameter-estimating algorithm. Harrison–Stevens forecasting uses Bayes's theorem in two ways: parameter estimation and model selection. Fildes (1983) surveyed the history of Harrison–Stevens and other Bayesian methods used in forecasting. Makov (1983) investigated several approximate Bayesian solutions to the dynamic linear model subject to jumps.

The Harrison–Stevens forecasting method uses the multiprocess dynamic linear model. It is an approximate method based on the notion of an independent sequence of state random variables coming from multinomial trials. At each time point, the state random variable determines which type of perturbation is applied to the parameter vector of the dynamic linear model. Actually this is only one of the set of state random variables of the process.

For a stochastic dynamic system, the set of state random variables has the property that given the past history of the process up to and including time t_0 and the set of state random variables at time t_0, the conditional distribution of the stochastic process at time $t_1 > t_0$ depends only on the state random variables at time t_0. The set of state random variables contains all of the information from the past and present that is relevant to the future development of the process. When the Kalman filter is applied to the dynamic linear model, the parameter vector is the set of state random variables (see Duncan and Horn 1972). The set of state random variables for the Harrison–Stevens forecasting method applied to the multiprocess dynamic linear model is the parameter vector plus the additional random variable that determines the type of perturbation most recently applied to the parameter vector. This additional random variable is called the state variable.

In the Harrison–Stevens algorithm, the correction applied to the parameter vector after the observation is a weighted average of the corrections calculated (conditionally on the value of the state random variable) by the Kalman filter, where the weights are the posterior probabilities of the state values. The method has also been referred to as the multiprocess Kalman filter.

The purpose of this article is to present the method in a terminology familiar to statisticians and to show that it can be only an approximation, for it relies on the assumption of normality, and a mixture of normal distributions is not itself normal. This is unlike the Kalman filter, which needs only the mean vector and covariance matrix, not the normality assumption. Despite this fact, the Harrison–Stevens method is a very effective forecasting tool and should be useful in a wide variety of situations.

Harrison and Stevens (1971) used the method on simulated data, and it proved to be effective in discriminating between transient events and sudden real changes in the process and to adapt quickly when real changes occur. It produces not only a forecast but also the joint parameter distribution that can be used to express uncertainty in the forecast or to construct prediction intervals. It also produces a posterior probability for each state. These can be used to detect when a change of state occurs for situations in which the state is of more interest than the forecast itself.

2. THE MULTIPROCESS DYNAMIC LINEAR MODEL

The multiprocess dynamic linear model is like the dynamic linear model in that the parameter vector is subject to perturbations. In the multiprocess model, however, the distribution of the perturbation depends on the state random variable at that time. The sequences of state random variables are independent of each other, and each can be considered the outcome of a single multinomial trial with known prior probabilities. The prior probabilities do not have to remain constant over time. This allows prior knowledge by the forecaster to be provided for in the model, making the forecasting system very flexible. Let S_t be the state random variable at time t. Then $\Pr(S_t = j) = \pi_t^{(j)}$ for $j = 1, \ldots, k$. When $S_t = j$, the parameter dynamics are governed by the equation $\beta_t = \mathbf{A}_t\beta_{t-1} + \mathbf{r}_t$, where β_t is the parameter vector at time t, $\mathbf{A}_t$ is a known matrix of coefficients at time t, and $\mathbf{r}_t$ is the perturbation vector, which is normally distributed with mean vector 0 and known covariance matrix

*William M. Bolstad is Mathematics Lecturer at the University of Waikato, Private Bag, Hamilton, New Zealand. The author would like to thank the editor and the referees for their helpful comments.

Table 1. States and Their Prior Probabilities

State number	Perturbation	Prior probability
1	No change	.9
2	Transient	.08
3	Mean change	.01
4	Slope change	.01

$\mathbf{R}_t^{(j)}$. The covariance matrix depends on the state j and can change over time.

The observation y_t is governed by the observation equation $y_t = \mathbf{X}_t'\beta_t + e_t$, where the observation errors e_t are iid $N(0, \sigma_e^2)$ random variables that are independent of the perturbation random vectors and the state random variables and $\mathbf{X}_t'$ is the known row vector of coefficients at time t.

3. THE HARRISON–STEVENS RECURSIVE ALGORITHM

The Harrison–Stevens algorithm starts at time $t - 1$ with k unbiased parameter estimator vectors, each conditional on the most recent state variable having been one of the k values and on all of the observations up to and including the most recent, y_{t-1}. Each conditional distribution is given to be normal with mean vector equal to the parameter vector and known covariance matrix. Let $\mathbf{y}_{t-1} = y_{t-1}, y_{t-2}, \ldots, y_1$ be the set of observations up to time $t - 1$. Then $(\hat{\beta}_{t-1} - \beta_{t-1}|S_{t-1} = i, \mathbf{y}_{t-1})$ is $N(0, \mathbf{V}_{t-1}^{(i)})$, where $\mathbf{V}_{t-1}^{(i)}$ is the covariance matrix given the state value $S_{t-1} = i$ and the past observations $\mathbf{y}_{t-1}$. In addition, there is the posterior (to observation $\mathbf{y}_{t-1}$) probability $S_{t-1} = i$:

$$q_{t-1}^{(i)} = \Pr(S_{t-1} = i|\mathbf{y}_{t-1}).$$

The joint probability–probability density function of $\hat{\beta}_{t-1}, S_{t-1}|\mathbf{y}_{t-1}$ is given by

$$f(\hat{\beta}_{t-1}, i|\mathbf{y}_{t-1}) = q_{t-1}^{(i)}(2\pi)^{-v/2}|\mathbf{V}_{t-1}^{(i)}|^{-1/2} \times \exp[-\tfrac{1}{2}(\hat{\beta}_{t-1} - \beta_{t-1})'[\mathbf{V}_{t-1}^{(i)}]^{-1}(\hat{\beta}_{t-1} - \beta_{t-1})]$$

for $i = 1, \ldots, k$.

The k conditional estimators together with the posterior probabilities can be used to give the composite estimator vector

$$\hat{\beta}_{t-1}|\mathbf{y}_{t-1} = \sum_{i=1}^{k} q_{t-1}^{(i)}(\hat{\beta}_{t-1}, S_{t-1} = i|\mathbf{y}_{t-1}).$$

Because it is a mixture of random vectors, the composite estimator vector $\hat{\beta}_{t-1}|\mathbf{y}_{t-1}$ has the mixture multivariate normal probability density function

$$f(\hat{\beta}_{t-1}|\mathbf{y}_{t-1}) = \sum_{i=1}^{k} q_{t-1}^{(i)}(2\pi)^{-v/2}|\mathbf{V}_{t-1}^{(i)}|^{-1/2} \times \exp[-\tfrac{1}{2}(\hat{\beta}_{t-1} - \beta_{t-1})'[\mathbf{V}_{t-1}^{(i)}]^{-1}(\hat{\beta}_{t-1} - \beta_{t-1})].$$

In Section 4 the composite estimator vector and its distribution will be brought forward in time and used to compute the forecast and its distribution.

If at time t the state variable $S_t = j$, then the perturbation vector $\mathbf{r}_t$ is $N(0, \mathbf{R}_t^{(j)})$. Let

$$(\hat{\beta}_t|S_t = j, S_{t-1} = i, \mathbf{y}_{t-1}) = \mathbf{A}_t(\hat{\beta}_{t-1}|S_{t-1} = i, \mathbf{y}_{t-1})$$

be the prior estimator of the parameter vector at time t. It is the conditional parameter estimator vector brought forward to time t, given $S_t = j$, but prior to observation y_t. The distribution of the conditional random variable $(\hat{\beta}_t - \beta_t|S_t = j, S_{t-1} = i, \mathbf{y}_{t-1})$ is $N(0, \mathbf{C}_t^{(i,j)})$, where the covariance matrix $\mathbf{C}_t^{(i,j)} = \mathbf{A}_t\mathbf{V}_{t-1}^{(i)}\mathbf{A}_t' + \mathbf{R}_t^{(j)}$. The predicted value of the observation at time t given $S_t = j$, $S_{t-1} = i$ is

$$\hat{y}_t^{(i,j)} = \mathbf{X}_t'(\hat{\beta}_t|S_t = j, S_{t-1} = i, \mathbf{y}_{t-1}).$$

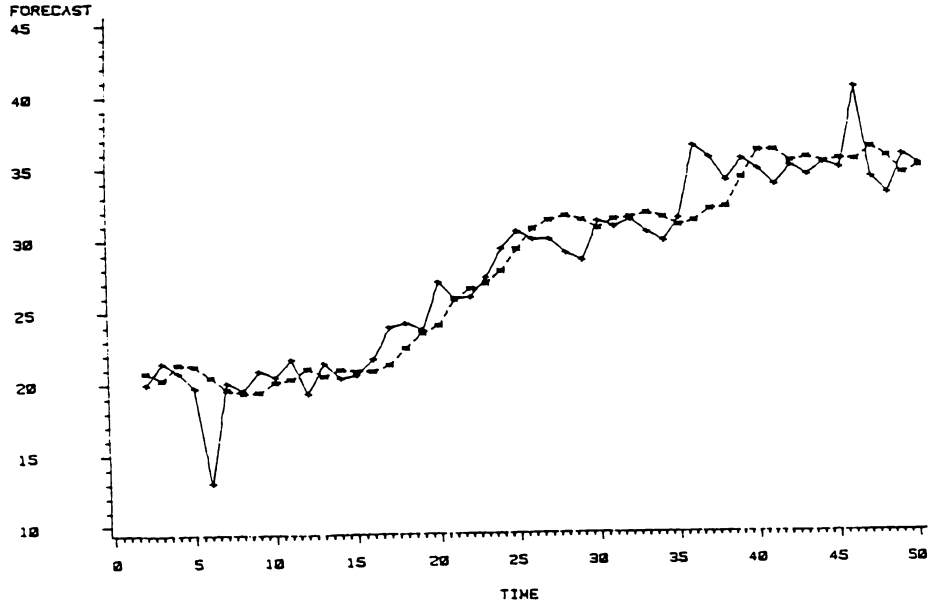


Figure 1. Forecast (•) and Observed Values (+).

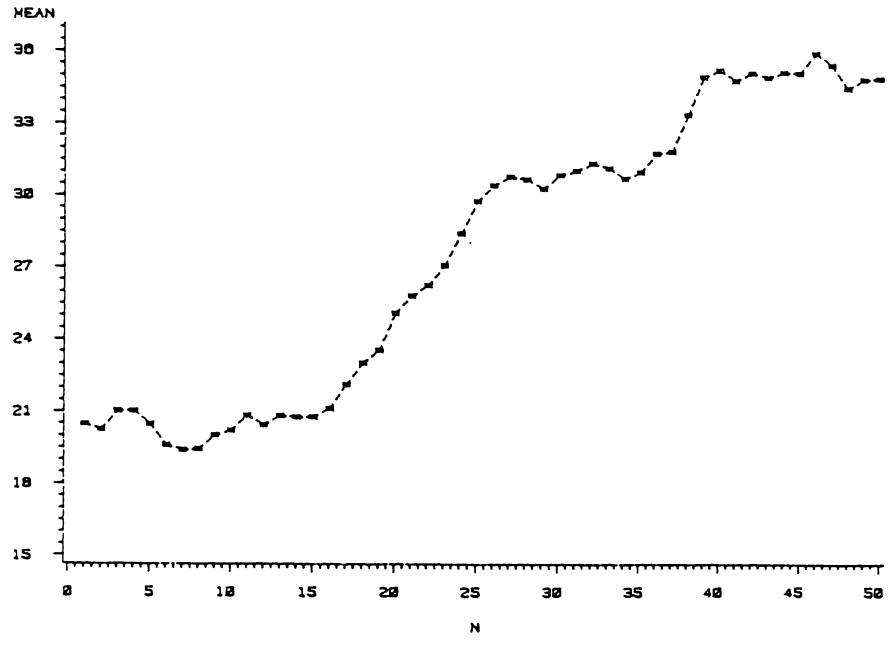


Figure 2. Local Mean Estimate.

The joint random variables

$$\begin{pmatrix} \hat{\beta}_t - \beta_t | S_t = j, S_{t-1} = i, \mathbf{y}_{t-1} \\ y_t - \hat{y}_t^{(i,j)} | S_t = j, S_{t-1} = i, \mathbf{y}_{t-1} \end{pmatrix}$$

are

$$\mathbf{N}\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}; \begin{pmatrix} \mathbf{C}_t^{(i,j)} & -\mathbf{C}_t^{(i,j)}\mathbf{X}_t \\ -\mathbf{X}_t'\mathbf{C}_t^{(i,j)} & \mathbf{X}_t'\mathbf{C}_t^{(i,j)}\mathbf{X}_t + \sigma_\epsilon^2 \end{pmatrix}\right].$$

Hence the conditional random variable $(\hat{\beta}_t - \beta_t | S_t = j, S_{t-1} = i, \mathbf{y}_t)$ is $N(\mu_t^{(i,j)}; \mathbf{V}_t^{(i,j)})$, where

$$\mu_t^{(i,j)} = 0 - \mathbf{C}_t^{(i,j)}\mathbf{X}_t(\mathbf{X}_t'\mathbf{C}_t^{(i,j)}\mathbf{X}_t + \sigma_\epsilon^2)^{-1}(y_t - \hat{y}_t^{(i,j)})$$

and

$$\mathbf{V}_t^{(i,j)} = \mathbf{C}_t^{(i,j)} - \mathbf{C}_t^{(i,j)}\mathbf{X}_t(\mathbf{X}_t'\mathbf{C}_t^{(i,j)}\mathbf{X}_t + \sigma_\epsilon^2)^{-1}\mathbf{X}_t'\mathbf{C}_t^{(i,j)}.$$

The joint probability–probability density function of $(\hat{\beta}_t, S_t, S_{t-1} | \mathbf{y}_t)$ is

$$f(\hat{\beta}_t, j, i | \mathbf{y}_t) = p_t^{(i,j)}(2\pi)^{-w/2}|\mathbf{V}_t^{(i,j)}|^{-1/2} \times \exp[-\tfrac{1}{2}(\hat{\beta}_t - \beta_t - \mu_t^{(i,j)})[\mathbf{V}_t^{(i,j)}]^{-1}(\hat{\beta}_t - \beta_t - \mu_t^{(i,j)})]$$

for $i = 1, \ldots, k$, and $j = 1, \ldots, k$, where $p_t^{(i,j)}$ is the posterior probability of the two states given the observation at time t:

$$p_t^{(i,j)} = \Pr(S_t = j, S_{t-1} = i | \mathbf{y}_t).$$

Because the sequence of state random variables is independent, the prior probability $\Pr(S_t = j, S_{t-1} = i) = \pi_t^{(j)} q_{t-1}^{(i)}$, and since $(y_t | S_t = j, S_{t-1} = i)$ is a normally distributed random variable with mean $\hat{y}_t^{(i,j)}$ and variance $\mathbf{X}_t'\mathbf{C}_t^{(i,j)}\mathbf{X}_t + \sigma_\epsilon^2$,

$$p_t^{(i,j)} = \pi_t^{(j)} q_{t-1}^{(i)} (f(y_t))^{-1}(2\pi)^{-w/2}|\mathbf{X}_t'\mathbf{C}_t^{(i,j)}\mathbf{X}_t + \sigma_\epsilon^2|^{-1/2} \times \exp[-\tfrac{1}{2}(\mathbf{X}_t'\mathbf{C}_t^{(i,j)}\mathbf{X}_t + \sigma_\epsilon^2)^{-1}(y_t - \hat{y}_t^{(i,j)})^2]$$

for $i = 1, \ldots, k$ and $j = 1, \ldots, k$. This is completely determined because $f(y_t)$ is a proportionality constant in all of the probabilities, and they sum to 1. Note that the calculation of these posterior probabilities requires the assumption of normality. This completes the determination of the joint probability–probability density function of $(\hat{\beta}_t, S_t, S_{t-1} | \mathbf{y}_t)$.

To find the marginal probability density function of $(\hat{\beta}_t | \mathbf{y}_t)$, we sum the joint probability–probability density function over i and j. Thus

$$f(\hat{\beta}_t | \mathbf{y}_t) = \sum_{i=1}^{k}\sum_{j=1}^{k} p_t^{(i,j)}(2\pi)^{-w/2}|\mathbf{V}_t^{(i,j)}|^{-1/2} \times \exp[-\tfrac{1}{2}(\hat{\beta}_t - \beta_t - \mu_t^{(i,j)})'[\mathbf{V}_t^{(i,j)}]^{-1}(\hat{\beta}_t - \beta_t - \mu_t^{(i,j)})].$$

This posterior distribution is a mixture of k^2 normal distributions depending on the two most recent state values. The process could be repeated each time another observation becomes available, increasing the number of components of the mixture by a factor of k each time. This would quickly get out of hand computationally. Instead, it is desirable to get the posterior distribution as a mixture of k normal distributions depending on the most recent state value. Let $q_t^{(j)}$ be the posterior probability $S_t = j$:

$$q_t^{(j)} = \Pr(S_t = j | \mathbf{y}_t) = \sum_{i=1}^{k} \Pr(S_t = j, S_{t-1} = i | \mathbf{y}_t) = \sum_{i=1}^{k} p_t^{(i,j)}.$$

Then

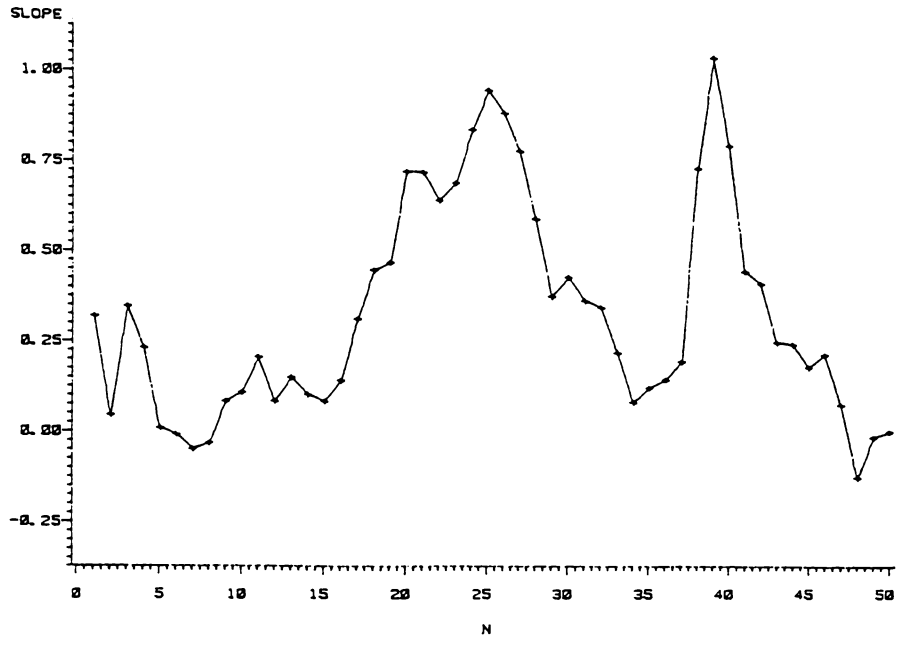


Figure 3. Local Slope Estimate.

$$f(\hat{\beta}_t|\mathbf{y}_t) = \sum_{j=1}^{k} q_t^{(j)} \sum_{i=1}^{k} p_t^{(i,j)} (q_t^{(j)})^{-1} (2\pi)^{-w/2} |\mathbf{V}_t^{(i,j)}|^{-1/2}$$

$$\times \exp[-\tfrac{1}{2}(\hat{\beta}_t - \beta_t - \mu_t^{(i,j)})' [\mathbf{V}_t^{(i,j)}]^{-1} (\hat{\beta}_t - \beta_t - \mu_t^{(i,j)})].$$

Hence the estimator vector is the mixture random vector

$$(\hat{\beta}_t|\mathbf{y}_t) = \sum_{j=1}^{k} q_t^{(j)} (\hat{\beta}_t|S_t = j, \mathbf{y}_t),$$

where

$$(\hat{\beta}_t|S_t = j, \mathbf{y}_t) = (q_t^{(j)})^{-1} \sum_{i=1}^{k} p_t^{(i,j)} (\hat{\beta}_t|S_t = j, S_{t-1} = i, \mathbf{y}_t).$$

The mean vector of a mixture is the mixture of the mean vectors, and the covariance matrix of a mixture is the mixture of the covariance matrices plus the mixture of the cross products of the deviations of the mean vectors from the overall mean vector. Hence the mixture of k normal distributions with $S_t = j$ fixed and $S_{t-1} = i$ for $i = 1, \ldots, k$ has mean vector $\beta_t + \mathbf{m}_t^{(j)}$, where

$$\mathbf{m}_t^{(j)} = (q_t^{(j)})^{-1} \sum_{i=1}^{k} p_t^{(i,j)} \mu_t^{(i,j)}$$

and covariance matrix

$$\mathbf{V}_t^{(j)} = (q_t^{(j)})^{-1} \sum_{i=1}^{k} p_t^{(i,j)} \times (\mathbf{V}_t^{(i,j)} + (\mu_t^{(i,j)} - \mathbf{m}_t^{(j)})(\mu_t^{(i,j)} - \mathbf{m}_t^{(j)})').$$

From this we see that the prior estimator vector $(\hat{\beta}_t|S_t = j, \mathbf{y}_t)$ is no longer an unbiased estimator of β_t.

We correct the estimator by subtracting the mean vector to give the posterior estimator

$$(\hat{\beta}_t|S_t = j, \mathbf{y}_t) = (\hat{\beta}_t|S_t = j, \mathbf{y}_t) - \mathbf{m}_t^{(j)}.$$

The correction applied is a weighted average of the corrections calculated by the Kalman filter conditional on each possible state at time $t - 1$, where the weights are the posterior probabilities of the states at time $t - 1$ given the state at time t and the observation. At this time we approximate the mixture distribution by a normal distribution having the same mean vector and covariance matrix. This can be only an approximation, since a mixture of normal distributions is normal only in the trivial case that all mean vectors and covariance matrices are respectively equal. At this point we have found k unbiased parameter estimator vectors, each conditional on the most recent state value and on all past observations, and posterior probabilities for the most recent state given all the past observations. We are now in the same position as we were when the algorithm started, so we are ready to repeat the algorithm when the next observation becomes available.

4. DISTRIBUTION OF THE FORECAST

The predicted value of the parameter vector at time $t + 1$ is

$$(\hat{\beta}_{t+1}|\mathbf{y}_t) = \mathbf{A}_{t+1}(\hat{\beta}_t|\mathbf{y}_t) = \sum_{i=1}^{k} q_t^{(i)} \mathbf{A}_{t+1} (\hat{\beta}_t|S_t = i, \mathbf{y}_t),$$

the composite estimator at time t brought forward to time $t + 1$. This is an unbiased estimator of the parameter vector at time $t + 1$, since the expected value of the perturbation error equals 0. $(\hat{\beta}_{t+1} - \beta_{t+1}|\mathbf{y}_t)$ is

$$N\left(0; \sum_{i=1}^{k} \sum_{j=1}^{k} q_t^{(i)} \pi_{t+1}^{(j)} \mathbf{C}_{t+1}^{(i,j)}\right),$$

where $\mathbf{C}_{t+1}^{(i,j)} = \mathbf{A}_{t+1} \mathbf{V}_t^{(i)} \mathbf{A}_{t+1}' + \mathbf{R}_{t+1}^{(j)}$. The forecast value of the observation is $\hat{y}_{t+1} = \mathbf{X}_{t+1}'(\hat{\beta}_{t+1}|\mathbf{y}_t)$, since the expected value of the observation error is 0. The variance of the prediction error $y_{t+1} - \hat{y}_{t+1}$ is given by

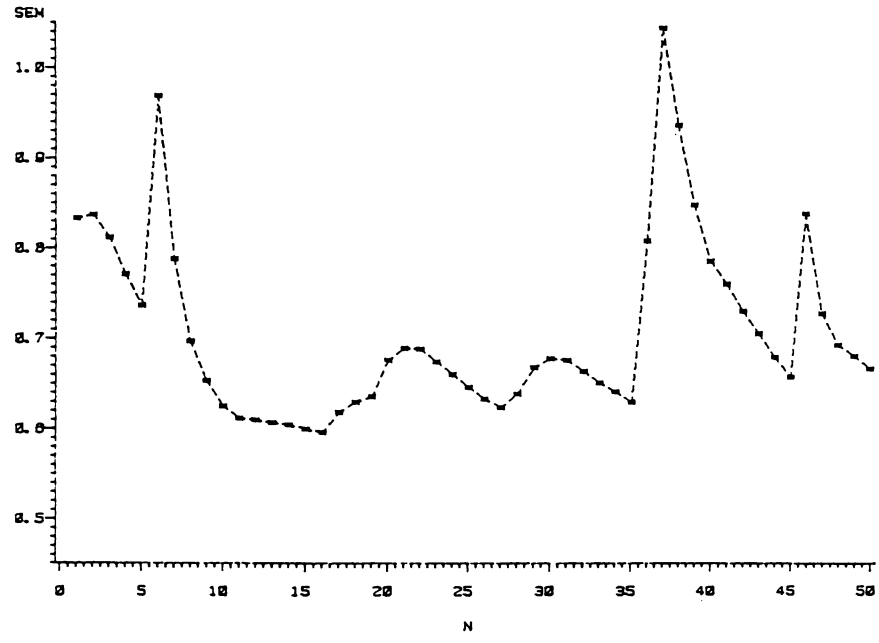


Figure 4. Standard Error of Mean.

$$\sigma^2_{y_{t+1}-\hat{y}_{t+1}} = \sum_{i=1}^{k}\sum_{j=1}^{k}(q_t^{(i)}\pi_{t+1}^{(j)}\mathbf{X}'_{t+1}\mathbf{C}_{t+1}^{(i,j)}\mathbf{X}_{t+1}) + \sigma_e^2.$$

Under the assumptions, the prediction error is approximately normally distributed, so prediction intervals can be constructed.

5. THE METHOD IN PRACTICE

The correction mechanism of the algorithm combines two uses of Bayes's theorem. The first is the calculation of the posterior mean vector and covariance matrix by the Kalman filter. This is done conditionally for each pair of previous and current state values. This gives the corrections necessary conditional on the two known state values.

The second is the use of the current forecast error to calculate the posterior probabilities of both the current state and the previous state. Thus the posterior state probabilities at time $t - 1$ are calculated once after observation y_{t-1} and again after observation y_t. The use of two observations is what enables the method to make distinction between the

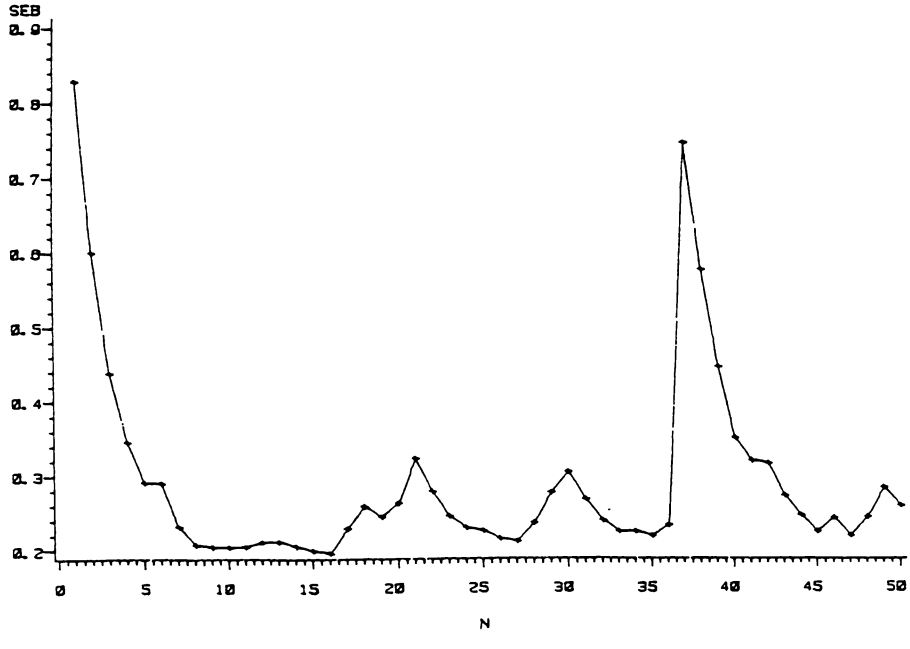


Figure 5. Standard Error of Slope.

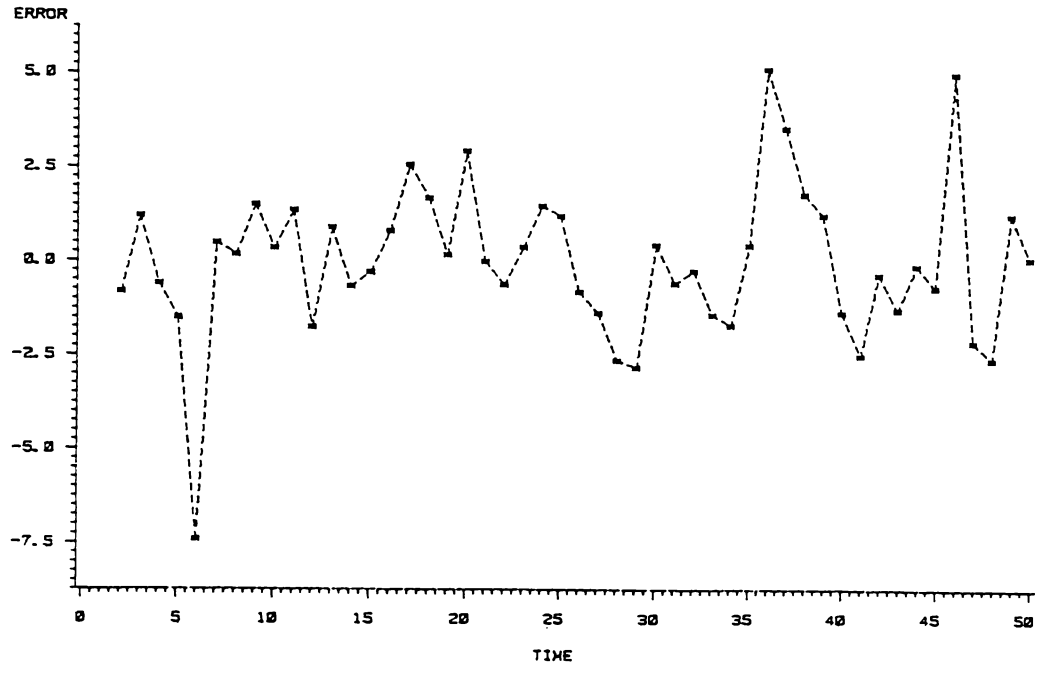


Figure 6. Forecast Error.

types of perturbations and adjust the state probabilities accordingly.

The two aspects are brought together in the condensation of a mixture of k^2 normal random variables to a mixture of k normal random variables. The final correction is the weighted average of the conditional corrections, where the weights are the conditional probability of the previous state given the present state after analyzing the present observation. This allows the rapid adjustment of the parameters two observations after a change in the time series occurs.

These will be illustrated in the following example on an artificially generated time series. The time series consists of 50 normally distributed random variables with variance 1 and the following mean pattern. The mean starts at 20 with slope = 0. There is a transient of -6 at observation 6. At step 16 the slope changes to 1, and it changes back to 0 at step 25, at which time the mean is 30. There is a mean change to 35 at step 36 and a transient of 5 at step 46.

The parameter vector

$$\beta_t = \begin{bmatrix} \mu_t \\ \beta_t \\ \gamma_t \end{bmatrix},$$

where μ_t, β_t, and γ_t are, respectively, the mean, the slope, and the transient at time t. The matrices of parameter dynamic coefficients and observation coefficients are, respectively,

$$\mathbf{A}_t = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad \mathbf{X}_t = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}.$$

The four states and their prior probabilities are given in Table 1.

The perturbation covariance matrices are

$$\mathbf{R}^{(1)} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \mathbf{R}^{(2)} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \sigma_\gamma^2 \end{pmatrix},$$

$$\mathbf{R}^{(3)} = \begin{pmatrix} \sigma_\mu^2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \mathbf{R}^{(4)} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \sigma_\beta^2 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

where $\sigma_\gamma^2 = \sigma_\mu^2 = 100\sigma_\epsilon^2$ and $\sigma_\beta^2 = \sigma_\epsilon^2$.

The forecast and the actual observations are shown in Figure 1. The estimated values of the parameters μ and β are shown in Figures 2 and 3, respectively, and the standard errors of the estimates are shown in Figures 4 and 5. The forecast errors are shown in Figure 6. Note that the transients at step 6 and step 46 hardly affected the parameter estimates; yet when real changes occurred, such as the slope change at step 16 and the mean change at step 36, the parameter estimates quickly tracked the new values. What happens is that when the first observation far from its forecast value occurs, the posterior probabilities of the states at that time drastically change in that the probabilities of "no change" go down and the probabilities of "mean change" and "transient" increase greatly. The posterior probabilities of the states at the previous time, however, do not change much, and these are the probabilities that are used in the condensation phase. Thus the parameter estimates do not change much, but the uncertainty in the parameter estimates increases. If the next observation is also far from its forecast value, the posterior probabilities for both the present and previous state values are drastically changed in that at the previous state "mean change" becomes much more likely and the state "transient" becomes less likely. These are the probabilities used in the condensation phase, so the parameter estimates change a great deal. If the next observation

Table 2. The State Probabilities at Time 6, Where a Transient Occurs

State number	Prior probability	Posterior observation 6	Posterior observation 7
1	.9	.000129	.000004
2	.08	.888773	.998467
3	.01	.111097	.001529
4	.01	.000001	.000000

Table 3. The State Probabilities at Time 36, Where a Mean Shift Occurs

State number	Prior probability	Posterior observation 36	Posterior observation 37
1	.9	.021643	.110507
2	.08	.869437	.155448
3	.01	.108680	.732930
4	.01	.000240	.001114

is close to its forecast value, however, the opposite occurs; the probability of "transient" goes up and the probability of "mean change" goes down. When these probabilities are used in the condensation phase, the parameter estimates do not change much. The changes in the probabilities that occur for a transient and for a mean change are illustrated in Tables 2 and 3.

6. PREVIOUS APPLICATIONS OF THE METHOD

Johnston and Harrison (1980) used the method to forecast demand in the alcoholic drink industry over a period that included record demand followed by a drought and the imposition of a new excise duty. The Harrison–Stevens algorithm is particularly useful in situations like these, for the forecaster can allow for these known shocks in the model by changing the prior probabilities. J. Q. Smith (1983) used the method to forecast accident insurance claims.

A. F. M. Smith, West, Gordon, Knapp, and Trimble (1983) and A. F. M. Smith and West (1983) used the method to monitor renal transplant data to detect a change of state (rejection of organ) if it occurs. The Harrison–Stevens algorithm is used to provide on-line probabilities of the state.

7. COMPARISON WITH OTHER METHODS

Taylor and Thomas (1982), in their comparison with the Box–Jenkins method for forecasting natural gas demand, concluded that

> It is the nature of the time series which determines the most appropriate form of forecasting to be used. In particular, the likelihood of the model remaining valid into the future. If there is good reason to suppose the model will persist, then exhaustive analysis on past data is clearly justified and the Box–Jenkins approach has proved effective in this role. If there is little reason to suppose the underlying model will remain valid, then some form of adaptive system is required and the Harrison–Stevens method seems to have the advantage. (p. 693)

Fildes (1983) empirically compared Harrison–Stevens forecasting with a number of other methods. He concluded that although it did not distinguish itself in terms of mean squared error, it did appear to be robust and avoid large errors. Fildes and Lusk (1984), in their study of how accuracy studies influence model choice, noted that perceived accuracy is related to familiarity with the method, and U.S. forecasters have a low familiarity with Harrison–Stevens forecasting and "accuracy studies cannot tell forecasters which method to adopt, rather they provide information which is relevant to prior belief in the accuracy of the method in the forecasters particular problem" (p. 429).

8. CONCLUSION

The Harrison–Stevens forecasting algorithm is a very flexible method that is particularly suitable for forecasting in nonstationary time series. Bayes's theorem is used two ways, and despite the assumption of normality, it can be only an approximation. It is flexible in that it allows an interaction between the forecaster and the method by allowing the forecaster to vary the prior state probabilities. It has also been used to determine the posterior state probabilities in situations in which the state itself is of more interest than the forecast.

I believe that as more statisticians become familiar with the method, it will become a widely used tool, and it will be adapted into other models, for instance, econometric models, and other multivariate models.

[*Received March 1985. Revised October 1985.*]

REFERENCES

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

Fildes, R. (1983), "An Evaluation of Bayesian Forecasting," *Journal of Forecasting*, 2, 137–150.

Fildes, R., and Lusk, E. J. (1984), "Choice of Forecasting Model," *Omega*, 12, 427–435.

Harrison, P. J., and Stevens, C. F. (1971), "A Bayesian Approach to Short-term Forecasting," *Operational Research Quarterly*, 22, 341–362.

Johnston, F. R., and Harrison, P. J. (1980), "An Application of Forecasting in the Alcoholic Drinks Industry," *Journal of the Operational Research Society*, 31, 699–709.

Makov, U. (1983), "Approximate Bayesian Procedures for Dynamic Linear Models in the Presence of Jumps," *Statistician*, 32, 207–213.

Meinhold, L. R., and Singpurwalla, N. (1983), "Understanding the Kalman Filter," *The American Statistician*, 37, 123–127.

Smith, A. F. M., and West, M. (1983), "An Application of Multi-process Kalman Filter to Renal Transplant Data," *Biometrics*, 39, 867–878.

Smith, A. F. M., West, M., Gordon, K., Knapp, M., and Trimble, I. C. (1983), "Monitoring Kidney Transplant Patients," *Statistician*, 32, 46–54.

Smith, J. Q. (1983), "An Application of Multi-process Dynamic Linear Model to Forecasting Accident Insurance Claims," *Statistician*, 32, 109–116.

Taylor, P. E., and Thomas, M. E. (1982), "Short Term Forecasting; Horses for Courses," *Journal of the Operational Research Society*, 33, 685–694.

```
c  PERFORMS THE HARRISON-STEVENS FORECASTING ON THE DYNAMIC LINEAR MODEL
       implicit double precision (a-h,o-z)
       dimension betai(4,3),xbeta(4,4)
       DIMENSION beta(4,4,3),c(4,4,3,3),UMU(4,4,3),q(4),ve(3),se(3)
       DIMENSION pi(4),po(4),po2(4),umuu(4,3)
       DIMENSION P(4,4),y(99),W1(4,4,3)
       DIMENSION A(3,3),r(4,3,3),X(3),AT(3,3),w(4,4,3,3),xc(4,4,3),cx(4,4,3)
       dimension yhat(4,4),vhat(4,4),vhatinv(4,4)
       dimension v(4,4,3,3),vi(4,3,3)
       dimension betapred(4,3),cpred(4,4,3,3),betahat(3),chat(3,3),ypred(4)
       dimension betaest(3),VEST(3,3)
       INTEGER I,J,K,l,N,ichk,ii,iny,iy,nn
       REAL em,ev,OBS,u,err,sum,
       real sigma2
       real vcast,secast
C beta CONTAINS THE STATE VECTORS -THIRD INDEX
c betai contains the condensed state vector -second index
C c CONTAINS THE STATE VECTOR COVARIANCE MATRICES-THIRD & FOURTH INDICES
C pi CONTAINS THE PRIOR PROBABILITIES
C po CONTAINS THE POSTERIOR PROBABILITIES
C A CONTAINS THE STATE TRANSISTION EQUATION
c sigma2 is the error variance
C   R contains the perturbation covariance matrix second & Third indices
C X CONTAINS THE OBSERVATION EQUATION
C Y IS THE TIME SERIES
C yhat contains the predicted value of y
C vhat contains the variance of the predicted value
c betapred is the conditional prediction
c betahat is the unconditional prediction
c cpred is the conditional cov mat of the pred
c chat is the unconditional cov mat of the pred
       ichk=1
       open(unit=4,file='y.dat',status='old')
       do 50 n=1,50
   50  read(4,60)  y(n)
   60  format(f10.0)
       do 63 i=1,4

       do 63 k=1,3

   63  vi(i,k,k)=99
       sigma2=1
       R(2,3,3)=100*sigma2
       R(3,1,1)=100*sigma2
       R(4,2,2)=4*sigma2
       pi(1)=.9
       pi(2)=.045
       pi(3)=.045
       pi(4)=.01
       po(1)=.9
       po(2)=.045
       po(3)=.045
       po(4)=.01
       phi=1
       n=100
       call assign(8,'dlm.out')
       CALL ASSIGN(6,'dlmwork.OUT')
C FOR pi AND po, 1=NO CHANGE,2=TRANSIENT,3=STEP CHANGE,4=SLOPE CHANGE
       x(1)=1
       X(2)=0
       X(3)=1
       DO 89 K=1,3
       DO 89 L=1,3
       A(K,L)=0
  89   CONTINUE
       A(1,1)=1
```

```
      A(1,2)=1
      A(2,2)=1
      DO 90 K=1,3
      DO 90 L=1,3
 90   AT(K,L)=A(L,K)
      if (ichk .eq. 1)  write(6,91) ((K,L,A(K,L),L=1,3),K=1,3)
 91   FORMAT(3(3(' A(',I1,',',I1,')=',F10.6)/))
C FOR A,1=MEAN,2=SLOPE,3=TRANSIENT
      do 300 nn=1,50
c the section up to statement 110 computes the updated (conditional) mean and
c variance of the linearly dynamic parameters
      write(6,60) y(nn)
      if (ichk .eq. 1) write(6,92) ((i,k,betai(i,k),k=1,3),i=1,4)
      if (ichk .eq. 1) write(6,272) (((j,k,l,vi(j,k,l),l=1,3)
     c  ,k=1,3),j=1,4)
 92   format(4(3(' betai(',i1,',',i1,')=',f10.6)/)/)

      do 100 i=1,4
      do 100 j=1,4
      do 100 k=1,3
      beta(i,j,k)=0
      do 100 l=1,3
100   beta(i,j,k)=beta(i,j,k)+a(k,l)*betai(i,l)
      if (ichk .eq. 1) write(6,102) (((i,j,k,beta(i,j,k),k=1,3),j=1,4),i=1,4)
102   format(4(4(3(' beta(',i1,',',i1,',',i1,')=',f10.6)/)/))
      do 110 i=1,4
      do 110 j=1,4
      do 110 k=1,3
      do 110 l=1,3
      w(i,j,k,l)=0
      do 105 k1=1,3
      do 105 l1=1,3
105   w(i,j,k,l)=w(i,j,k,l)+a(k,k1)*vi(i,k1,l1)*at(l1,l)
110   c(i,j,k,l)=w(i,j,k,l)+R(j,k,l)
      if (ichk .eq. 1) write(6,111) ((((i,j,k,l,w(i,j,k,l),l=1,3)
     c  ,k=1,3),j=1,4),i=1,4)
111   format(4(4(3(3(' w(',i1,',',i1,',',i1,',',i1,')=',f10.6)/)/)/))
      if (ichk .eq. 1) write(6,112) ((((i,j,k,l,c(i,j,k,l),l=1,3)
     c  ,k=1,3),j=1,4),i=1,4)
112   format(4(4(3(3(' c(',i1,',',i1,',',i1,',',i1,')=',f10.6)/)/)/))
      do 193 i=1,4
      do 193 j=1,4
      yhat(i,j)=0
      do 193 k=1,3
193   yhat(i,j)=yhat(i,j)+x(k)*beta(i,j,k)
      do 194 i=1,4
      do 194 j=1,4
      vhat(i,j)=0
      do 194 k=1,3
      xc(i,j,k)=0
      do 194 l=1,3
      xc(i,j,k)=xc(i,j,k) + x(l)*c(i,j,l,k)
194   vhat(i,j)=vhat(i,j)+x(k)*C(i,j,k,l)*x(l)
      if (ichk .eq. 1) write(6,195) ((i,j,yhat(i,j),j=1,4),i=1,4)
195   format(4(4(' yhat(',I1,',',I1,')=',F10.6)/))
      if (ichk .eq. 1) write(6,196) ((i,j,vhat(i,j),j=1,4),i=1,4)
196   format(4(4(' vhat(',I1,',',I1,')=',F10.6)/))
      if (ichk .eq. 1) write(6,197) (((i,j,k,xc(i,j,k),k=1,3),j=1,4),i=1,4)
197   format(4(4(3(' xc(',i1,',',i1,',',i1,')=',f10.6)/)/))
      do 198 i=1,4
      do 198 j=1,4
      vhat(i,j)=vhat(i,j)+sigma2
198   vhatinv(i,j)=1/vhat(i,j)
c the section up until statement 215 calculates the corrected mean of the
c dynamic parameter (conditional)
      do 215 i=1,4
```

```
      do 268 k=1,3
      do 268 l=1,3
268   vi(j,k,l)=0
      do 270 k=1,3
      do 270 l=1,3
      do 270 i=1,4
      if (po(j) .ne. 0.) vi(j,k,l)=vi(j,k,l)+p(i,j)/po(j)
     c  *(v(i,j,k,l)+(umu(i,j,k)-umuu(j,k))*(umu(i,j,l)-umuu(j,l)))
270   continue
      if (ichk .eq. 1) write(6,271) ((j,k,umuu(j,k),k=1,3),j=1,4)
      if (ichk .eq. 1) write(6,92) ((i,k,betai(i,k),k=1,3),i=1,4)
      if (ichk .eq. 1) write(6,272) (((j,k,l,vi(j,k,l),l=1,3)
     c  ,k=1,3),j=1,4)
271   format(4(3(' umuu(',i1,',',i1,')=',f10.6)/)/)
272   format(4(3(3(' vi(',i1,',',i1,',',i1,')=',f10.6)/)/)/)
      do 275 j=1,4
      do 275 k=1,3
      betapred(j,k)=0
      do 275 l=1,3
275   betapred(j,k)=betapred(j,k)+a(k,l)*betai(j,l)
      do 277 j=1,4
      do 277 i=1,4
      do 277 k=1,3
      do 277 l=1,3
      cpred(i,j,k,l)=0
      do 276 k1=1,3
      do 276 l1=1,3
276   cpred(j,i,k,l)=cpred(j,i,k,l)+a(k,k1)*vi(j,k1,l1)*at(l1,l)
277   cpred(j,i,k,l)=cpred(j,i,k,l)+R(i,k,l)
      if (ichk .eq. 1) write(6,278) ((j,k,betapred(j,k),k=1,3),j=1,4)
278   format(4(3(' betapred(',i1,',',i1,')=',f10.6)/)/)
      if (ichk .eq. 1) write(6,279) ((((i,j,k,l,cpred(i,j,k,l),l=1,3)
     c  ,k=1,3),j=1,4),i=1,4)
279   format(4(4(3(3(' cpred(',i1,',',i1,',',i1,',',i1,')=',
     c  f10.6)/)/)/))
      do 281 k=1,3
      betahat(k)=0
      do 281 j=1,4
281   betahat(k)=betahat(k)+po(j)*betapred(j,k)
      do 284 k=1,3
      do 284 l=1,3
      chat(k,l)=0
      do 284 i=1,4
      do 284 j=1,4
284   chat(k,l)=chat(k,l)+po(j)*pi(i)*cpred(j,i,k,l)
      if (ichk .eq. 1) write(6,285) (k,betahat(k),k=1,3)
285   format(3(' betahat(',i1,')=',f10.6))
      if (ichk .eq. 1)  write(6,286) ((K,L,chat(K,L),L=1,3),K=1,3)
286   FORMAT(3(3(' chat(',I1,',',I1,')=',F10.6)/))
      yfcst =  0
      vcast=0
      ypred(1)=0
      ypred(2)=0
      ypred(3)=0
      ypred(4)=0
      do 287 k=1,3
      ypred(1)=ypred(1)+x(k)*betapred(1,k)
      ypred(2)=ypred(2)+x(k)*betapred(2,k)
      ypred(3)=ypred(3)+x(k)*betapred(3,k)
      ypred(4)=ypred(4)+x(k)*betapred(4,k)
287   yfcst=yfcst+x(k)*betahat(k)
      do 288 k=1,3
      do 288 l=1,3
      vcast=vcast+x(k)*chat(k,l)*x(l)
288   secast=sqrt(vcast+sigma2)
      if (ichk .eq. 1) write(6,289) y(nn),yfcst,(ypred(k),k=1,4)
```

```
      do 215 j=1,4
      do 215 k=1,3
      umu(i,j,k)=0
      do 214 l=1,3
  214 cx(i,j,l)=xc(i,j,l)
  215 umu(i,j,k)=umu(i,j,k)+cx(i,j,k)*(yhat(i,j)-y(nn))*vhatinv(i,j)
      if (ichk .eq. 1) write(6,217) (((i,j,k,wl(i,j,k),k=1,3),j=1,4),i=1,4)
  217 format(4(4(3(' umu(',i1,',',i1,',',i1,')=',f10.6)/)/))
      if (ichk .eq. 1) write(6,102) (((i,j,k,beta(i,j,k),k=1,3),j=1,4),i=1,4)

c the section up to statement 230 calculates the corrected covariance
c matrix of the dynamic parameter (conditional)
      do 230 i=1,4
      do 230 j=1,4
      do 230 k=1,3
      do 230 l=1,3
  230 v(i,j,k,l)=c(i,j,k,l)-cx(i,j,k)*cx(i,j,l)*vhatinv(i,j)
      if (ichk .eq. 1) write(6,242) ((((i,j,k,l,v(i,j,k,l),l=1,3)
     c ,k=1,3),j=1,4),i=1,4)
  242 format(4(4(3(3(' v(',i1,',',i1,',',i1,',',i1,')=',f10.6)/)/)/))
c the section up to statement 260 is determining the posterior probabilities
c of the perturbation index variable
      sum=0
      do 258 i=1,4
      do 258 j=1,4
      err=Y(nn)-yhat(i,j)
      err2=err**2
      if (ichk .eq. 1) write(6,254) y(nn),yhat(i,j),err,err2
  254 Format (4f10.5)
  258 p(i,j)=po(i)*pi(j)*.3989422803/(sqrt(vhat(i,j))) *
     c exp(-.5*err2/vhat(i,j))
      if (ichk .eq. 1) write(6,261) ((i,j,p(i,j),j=1,4),i=1,4)
      do 259 i=1,4
      do 259 j=1,4
  259 sum=sum+p(i,j)
      do 260 i=1,4
      do 260 j=1,4
  260 if (sum .ne. 0) p(i,j)=p(i,j)/sum
      if (ichk .eq. 1) write(6,261) ((i,j,p(i,j),j=1,4),i=1,4)
  261 format(4(4(' p(',I1,',',I1,')=',F10.6)/))
c the section up to statement 270 finds the condensed posterior means and
c covariance matrices
c po(j) is the posterior probof most recent pert index,
c po2(j) is the post prob of previous pert index
      do 262 j=1,4
      po(j)=0
      po2(j)=0
      do 262 i=1,4
      po2(j)=po2(j)+p(j,i)
  262 po(j)=po(j)+p(i,j)
      if (ichk .eq. 1) write(6,263) (j,po(j),j=1,4),(i,po2(i),i=1,4)
  263 format(4(' po(',I1,')=',F10.6)/,4(' po2(',i1,')=',f10.6))
      do 266 j=1,4
      do 266 k=1,3
      umuu(j,k)=0
      betai(j,k)=0
      do 264 i=1,4
      if (po(j) .ne. 0.) betai(j,k)=betai(j,k)+p(i,j)/po(j)*beta(i,j,k)

  264 if (po(j) .ne. 0.) umuu(j,k)=umuu(j,k)+p(i,j)/po(j) * umu(i,j,k)
  266 continue
      if (ichk .eq. 1) write(6,92) ((i,k,betai(i,k),k=1,3),i=1,4)
      do 267 j=1,4
      do 267 k=1,3
  267 betai(j,k)=betai(j,k)-umuu(j,k)
      do 270 j=1,4
```

```
  289   format(6f10.6)
  290   format(i3,9(1x,f10.6))
  291   format(4(4(1x,f10.6)/))
c  the section up to 299 calculates the combined posterior estimators
        do 295 k=1,3
        betaest(k)=0
        ve(k)=0
        do 295 j=1,4
        ve(k)=ve(k)+po(j)*vi(j,k,k)
        se(k)=sqrt(ve(k))
  295   betaest(k)=betaest(k)+po(j)*betai(j,k)
        if (ichk .eq. 1) write(6,298) vcast,secast
  298   format(2f10.6)
        write(8,290) nn,y(nn),(betaest(k),k=1,3),(se(k),k=1,3),yfcst,secast
        ichk=0
  300   continue
        close(unit=4)
        END
```

APPENDIX 3

COMMUN. STATIST.-SIMULA., 15(3), 819-828 (1986)

AN EFFICIENT ALGORITHM FOR HARRISON-STEVENS FORECASTING USING THE MULTI-PROCESS MULTIVARIATE DYNAMIC LINEAR MODEL

William M. Bolstad

Department of Mathematics
University of Waikato
Hamilton
New Zealand

Key words and phrases: Bayesian forecasting, dynamic linear model, Kalman filter, mixture of distributions, multi-process Kalman filter, state vector, state vector estimator.

ABSTRACT

This paper develops a computationally efficient algorithm for Harrison-Stevens forecasting in a multivariate time series which has correlated errors. The algorithm uses the observation vector one component at a time on the multi-process multivariate dynamic linear model. This gives a computationally efficient, robust, quick adapting forecasting method for non stationary multivariate time series.

1. INTRODUCTION

Recursive estimation techniques have been used in time series analysis for many years. One of their main attractions is their computational efficiency. Although time series models and statistical models are similar in that they are both based on probability, techniques

developed in one only slowly diffused into the other. Recursive estimation seems natural in time series analysis where the observations come in sequence, but not so in statistical analysis; yet the computational efficiency is there for both models. On the other hand, statistics has developed more effective techniques for dealing with multivariate observations. Clearly, cross fertilization between the two fields will yield benefits to both.

Recursive estimation algorithms first appeared in the statistical literature with the recursive least squares algorithm developed by R. L. Plackett (1950). R. E. Kalman (1960) discovered the optimal linear filter for a linear dynamic time varying system. This was quickly incorporated into control system engineering, however statisticians were slow to adapt the method to statistical problems even though it has a simple interpretation as a Bayesian technique. Duncan and Horn (1972) related the Kalman filter to the dynamic linear model. Salas and Harville (1981) adapted the method for estimating the parameters of the (static) general linear model. Bolstad (1986 A) developed a computationally efficient recursive estimation algorithm for the seemingly unrelated regression model with known contemporaneous covariance matrix. This algorithm enlarged the state vector by adjoining the predicted observation error vector at that time. The algorithm used the observation vector components one at a time, thus eliminating the need for time consuming matrix inversions. As each component is analyzed the dimension of the state variable is reduced by one.

Harrison-Stevens forecasting is a technique devised by Harrison and Stevens (1971) as a forecasting tool for short term forecasting in a time series where there may be underlying changes in the process generating the time series. Baye's theorem is used two ways each time an observation becomes available; parameter estimation and model selection. It is based on the multi-process dynamic linear model. A random variable which Harrison and Stevens call "the state variable" coming from a single multinomial trial determines the distribution of the perturbation applied to the parameter vector of the dynamic linear model at that time. Bolstad (1986 B).describes how Harrison and Stevens use the term "state variable" differently than it is used in the Kalman filter applied to the dynamic linear model . In order to prevent confusion, this paper will use the term "perturbation index variable" instead of "state variable".

Harrison and Stevens (1976) described the foundations of Bayesian forecasting in terms of (*i*) the parametric or statespace model, (*ii*) the probabilistic information on the model parameters, (*iii*) the sequential model definition which describes how the model parameters change with time, and (*iv*) the uncertainty as to the underlying model itself as between a number of discrete alternatives. They show how their method of Bayesian forecasting can be used for a variety of processes.

West, Harrison, and Migon (1985) have extended the dynamic linear model to the dynamic generalized linear model. Their algorithm is a generalization of the Kalman filtering algorithm for a univariate time series from a dynamic exponential family model when the parameter process is linear. Guttman and Pena (1985) developed a similar method called robust filtering. They incorporate two states, "outlying observation" and "acceptable observation" into the observation equation, not the parameter dynamic equation as is done in

Harrison-Stevens forecasting. Thus it will be robust against outliers, but it will not have the quick reacting capability of Harrison-Stevens forecasting because no allowance has been made for an abrupt change in the parameters.

In the Harrison-Stevens algorithm, the correction applied to the parameter vector after analyzing the observation is a weighted average of the corrections calculated (conditionally on the value of the perturbation index variable) by the Kalman filter where the weights are the posterior probabilities of the perturbation index values. It has been referred to as the multi-process Kalman filter. However, unlike the ordinary Kalman filter it is only an approximation, for it relies on the assumption of normality, and a mixture of normal distributions is not actually normal. Despite this fact, the Harrison-Stevens forecasting method is very effective for short term forecasts, and also for detecting changes in the underlying process generating the time series.

The purpose of this paper is to develop a computationally efficient algorithm for Harrison-Stevens forecasting in a multivariate time series where there are correlated observation errors. While a straight forward application of the Harrison-Stevens algorithm is possible for the multivariate case, it would be computationally inefficient due to the matrix inversions required. The algorithm developed here will adapt the one component at a time approach from Bolstad (1986 A) to the multi-process multivariate dynamic linear model. Thus the only matrices to be inverted will be one by one.

2. THE MULTI-PROCESS MULTIVARIATE DYNAMIC LINEAR MODEL

The multi-process multivariate dynamic linear model is like the dynamic linear model in that the parameter vector is subject to perturbations. In the multi-process model however, the distribution of the perturbation depends on the perturbation index random variable at that time. The sequence of perturbation index variables are independent of each other and each can be considered to be the outcome of a single multinomial trial with known prior probabilities. The prior probabilities do not have to remain constant over time. This allows prior knowledge by the forecaster into the model , hence the forecasting system is very flexible. Let I_t be the perturbation index random variable at time t.

$$Pr(I_t = j) = \pi_t^{(j)} \quad \text{for } j = 1,\ldots,k$$

When $I_t = j$, the parameter dynamics are governed by the equation

$$\beta_t = \mathbf{A}_t \beta_{t-1} + \mathbf{r}_t$$

where β_t is the p by 1 parameter vector at time t , $\mathbf{A}_t$ is a known matrix of coefficients at time t, and $\mathbf{r}_t$ is the perturbation vector which is normally distributed with mean vector 0 and known covariance matrix $\mathbf{R}_t^{(j)}$. The covariance matrix depends on the perturbation index j and can change over time. The M by 1 observation vector $\mathbf{y}_t$ is governed by the observation equation

$$\mathbf{y}_t = \mathbf{X}_t' \beta_t + \mathbf{e}_t$$

where X_t is the known matrix of coefficients at time t and the error vectors e_t are independent $N(0, \Sigma_e)$ random vectors.

3. THE ALGORITHM

This section develops an algorithm for estimating the parameter vector β_t. It starts at time $t-1$ with k unbiased parameter estimation vectors, each conditional on the most recent perturbation index variable having been one of the k values and on all the observation vectors up to and including the most recent y_{t-1}. Let $Y_{t-1}=y_{t-1},y_{t-2},\cdots,y_1$ be the set of observation vectors up to and including time $t-1$. Then

$$(\hat{\beta}_{t-1}-\beta_{t-1}|I_{t-1}=i,Y_{t-1}) \text{ is } N(0,V_{t-1}^{(i)})$$

where $V_{t-1}^{(i)}$ is the known covariance matrix given the perturbation index variable $I_{t-1}=i$ and the past observations Y_{t-1}. Also given are the posterior probabilities of the perturbation index given the observations, $q_{t-1}^{(i)}=Pr(I_{t-1}=i \mid Y_{t-1})$. These starting conditions are the same as those for the univariate Harrison-Stevens algorithm.

When the observation vector at time t becomes available, the k parameter vector estimators are projected forward to time t becoming k^2 parameter vector estimators given the perturbation index value at time t as well as the perturbation index value at time $t-1$ and the observations Y_{t-1}. In this algorithm $S_{m\,t-1}$, the state vector at time $t-1$ with m out of the M components of the observation vector yet to be analyzed consists of two parts; the parameter vector β_{t-1} and an m by 1 zero vector which is the expected values of the prediction errors of the m components of the observation vector that haven't yet been analyzed. Each of the k^2 corresponding state vector estimators also consists of two parts; the corresponding parameter estimator, and an estimator for the prediction errors of the components of the error vector not yet analyzed. As each component of the observation vector is used, the dimension of the state vector estimators are reduced by one, and a change in the expected value of the state vector estimator is calculated based on the prediction error of that component. The joint probabilities of the two most recent perturbation index values are also recalculated. At each step the only matrices to be inverted are one by one matrices so the algorithm is computationally efficient.

After all the components of the observation vector have been analyzed, the state vector consists of the parameter vector and the state vector estimators consists of the parameter vector estimators. Then the k^2 parameter vector estimators which are conditional on the two most recent perturbation index values are condensed into k parameter vector estimators which are conditional on the most recent perturbation index value. The expected value of these parameter vector estimators are no longer the parameter vector, so the posterior parameter vector estimators are found by subtraction. The posterior probabilities of the most recent perturbation index are also calculated. Then the algorithm is ready to be repeated when another observation vector becomes available.

Steps of the algorithm

1. At time t-1 after all the components of Y_{t-1} have been analyzed the state vector $S_{0,t-1} = \beta_{t-1}$ and the state vector estimators $(\hat{S}_{0,t-1}|I_{t-1}=i, Y_{t-1})=(\hat{\beta}_{t-1}|I_{t-1}=i, Y_{t-1})$. At this stage the state vector and k state vector estimators are the parameter vector and the k corresponding parameter vector estimators respectively, and let $q_{t-1}^{(i)}$ be the posterior probability of the perturbation index being i at time $t-1$.

2. At time t, before using observation y_t, the prior probabilities of the perturbation indices are given by $\pi_t^{(j)}=Pr(I_t=j)$. Because the sequence of perturbation index variables are assumed to be independent, the prior probabilities $p_{M\,t}^{(i,j)} = Pr(I_t=j, I_{t-1}=i \mid Y_{t-1}) = \pi_t^{(j)} q_{t-1}^{(i)}$. At that time there are M components of the observation vector not yet analyzed and the state vector is

$$S_{M,t} = \begin{bmatrix} \beta_t \\ \bar{0} \end{bmatrix}$$

where $\bar{0}$ is a M by 1 vector of 0's. Let $(\hat{\beta}_t | I_t=j, I_{t-1}=i, Y_{t-1}) = A_t(\hat{\beta}_{t-1}|I_{t-1}=i, Y_{t-1})$ be the prior estimator of the parameter vector at time t conditional on the perturbation index variables $I_t=j, I_{t-1}=i$ and the observation vectors Y_{t-1}. Each of these is the conditional parameter vector estimator at time $t-1$ brought forward to time t, given $I_t=j$, but prior to analyzing observation vector y_t. At this stage there are k^2 conditional prior estimators of the parameter vector. If at time t the perturbation index variable $I_t=j$ then r_t, the perturbation vector is $N(\bar{0}, R_t^{(j)})$. Hence the distribution of $(\hat{\beta}_t | I_t=j, I_{t-1}=i, Y_{t-1})$ is $N(\beta_t, C_t^{(i,j)})$ where $C_t^{(i,j)} = A_t V_{t-1}^{(i)} A_t' + R_t^{(j)}$

3. Adjoin to each of these a M by 1 vector representing the prediction error vector at time t given the perturbation index variables $I_t=j, I_{t-1}=j$ and the observation vectors Y_{t-1}.

$$(\hat{S}_{M,t} | I_t=j, I_{t-1}=i, Y_{t-1}) = \begin{bmatrix} \hat{\beta}_t | I_t=j, I_{t-1}=i, Y_{t-1} \\ \hat{e}_t | I_t=j, I_{t-1}=i, Y_{t-1} \end{bmatrix}.$$

where $(\hat{e}_t | I_t=j, I_{t-1}=i, Y_{t-1}) = (y_t | I_t=j, I_{t-1}=i, Y_{t-1}) - X_t(\hat{\beta}_t | I_{t-1}=i, Y_{t-1})$ is the prediction error vector. Under the assumption that the observation vector is normally distributed the distribution of $(\hat{S}_{M,t} - S_{M,t} | I_t=j, I_{t-1}=i, Y_{t-1})$ is $N(\bar{0}, W_t^{(i,j)})$ where the covariance matrix

$$W_t^{(i,j)} = \begin{bmatrix} C_t^{(i,j)} & -C_t^{(i,j)} X \\ -X_t' C_t^{(i,j)} & X_t' C_t^{(i,j)} X_t + \Sigma_e \end{bmatrix}$$

Let $\Theta_{M\,t}^{(i,j)}=\bar{0}$ be a $p+M$ by 1 vector of 0's. The first p components represent the error in the parameter vector estimator, and the last M components represent the prior mean values of the prediction error vector.

4. Set $m=M$

5. Partition each of the conditional state vector estimators into

$$(\hat{S}_{m,t} | I_t=j, I_{t-1}=i, Y_{t-1}, \hat{e}_{m+1,t}, \cdots, \hat{e}_{M,t}) = \begin{bmatrix} \hat{S}_{m-1,t} | I_t=j, I_{t-1}=i, Y_{t-1}, \hat{e}_{m+1,t}, \cdots, \hat{e}_{M,t} \\ \hat{e}_{m,t} | I_t=j, I_{t-1}=i, Y_{t-1}, \hat{e}_{m+1,t}, \cdots, \hat{e}_{M,t} \end{bmatrix}$$

where $\hat{e}_{m,t}$ is the the $m'th$ component of the prediction error vector at time t. The corresponding partitions of the mean vectors are

$$\Theta_{m,t}^{(i,j)} = \begin{bmatrix} \Theta_{m-1,t}^{(i,j)} \\ \theta_{m,t} \end{bmatrix}$$

And the corresponding partitions of the covariance matrices are

$$W_t^{(i,j)} = \begin{bmatrix} U_{1,1}^{(i,j)} & U_{1,2}^{(i,j)} \\ U_{2,1}^{(i,j)} & U_{2,2}^{(i,j)} \end{bmatrix}$$

where $U_{1,1}^{(i,j)}$ is a $p+m-1$ by $p+m-1$ matrix, $U_{1,2}^{(i,j)}$ is a $p+m-1$ by 1 matrix, $U_{2,1}^{(i,j)}$ is a 1 by $p+m-1$ matrix and $U_{2,2}^{(i,j)}$ is a 1 by 1 matrix. The conditional distribution of the prior state vector estimator $(\hat{S}_{m-1,t} \mid i_{t-1}, I_{t-1}=i, Y_{t-1}, \hat{e}_{m,t}, \ldots, \hat{e}_{M,t})$ is normal with mean vector $S_{m-1,t}+\Theta_{m-1,t}^{(i,j)}$, where the new value of vector $\Theta_{m-1,t}^{(i,j)}=\Theta_{m-1,t}^{(i,j)}+U_{1,2}^{(i,j)}(U_{2,2}^{(i,j)})^{-1}(\hat{e}_{m,t}-\theta_{m,t}^{(i,j)})$ and the covariance matrix $U_{1,1}^{(i,j)}-U_{1,2}^{(i,j)}(U_{2,2}^{(i,j)})^{-1}U_{2,1}^{(i,j)}$. (Graybill, page 63). Note that the only matrices to be inverted are 1 by 1

6. Calculate the posterior probabilities of the parameter perturbation indices by

$$p_{m-1,t}^{(i,j)} = Pr(I_t=j, I_{t-1}=i \mid Y_{t-1}, \hat{e}_{m,t}, \cdots, \hat{e}_{M,t})$$

$$= p_{m,t}^{(i,j)}(f(\hat{e}_{m,t}))^{-1}(2\pi)^{-1/2}\mid U_{2,2}^{(i,j)}\mid^{-1/2}\exp[-1/2(U_{2,2}^{(i,j)})^{-1}(\hat{e}_{m,t}-\theta_{m,t})^2$$

for $i=1,\ldots,k$ and $j=1,\ldots,k$. These are known since $f(\hat{e}_{m,t})$ is a constant in each of the terms and they must sum to one (over i and j) to make a probability distribution. $f(\hat{e}_{m,t})$ is the unconditional density of the random variable $\hat{e}_{m,t}$ which has as its distribution a mixture of normal distributions. Hence the calculation of these posterior probabilities requires the assumption of normality.

7. If $m>1$ then let $m=m-1$ and go to step 5

8. If $m=1$ then each of the k^2 state vector estimators is now conditional on observation y_t as well as all the previous observations Y_{t-1}. Hence it can be written as $(\hat{S}_{0,t} \mid I_t=j, I_{t-1}=i, Y_t)$ and it equals $(\hat{\beta}_t \mid I_t=j, I_{t-1}=i, Y_t)$. The distributions of the k^2 estimators of the parameter vector at time t

$$(\hat{\beta}_t \mid I_t=j, I_{t-1}=i, Y_t) = (\hat{S}_t \mid I_t=j, I_{t-1}=i, Y_t) \text{ are } N(\beta_t+\Theta_{0,t}^{(i,j)}; U_{2,2}^{(i,j)})$$

The posterior probabilities of the perturbation index variables are also conditional on observation y_t as well as on all the previous observations Y_{t-1}. They can be written as

$$p_{0,t}^{(i,j)} = Pr(I_t=j, I_{t-1}=i \mid Y_t)$$

9. The joint probability-probability density function of $(\hat{\beta}_t, I_t, I_{t-1} \mid Y_t)$ is

$$f(\hat{\beta}_t, j, i \mid Y_t) = p_{0,t}^{(i,j)}\mid U_{2,2}^{(i,j)}\mid^{-1/2}\times\exp[-1/2(\hat{\beta}_t-\beta_t-\Theta_{0,t}^{(i,j)})'[U_{2,2}^{(i,j)}]^{-1}(\hat{\beta}_t-\beta_t-\Theta_{0,t}^{(i,j)})$$

for $i=1,\ldots,k$ and $j=1,\ldots,k$. This distribution is a mixture of k^2 normal distributions depending on the two most recent values of the perturbation index variable. The pro-

cess could be repeated each time another observation vector becomes available, however this would quickly get out of hand computationally. Instead we want to get the posterior distribution as a mixture of k normal distributions depending on the most recent perturbation index value. The composite estimator vector is the mixture estimator

$$(\hat{\beta}_t \mid \mathbf{Y}_t) = \sum_{j=1}^{k}\sum_{i=1}^{k} p_{0\,t}^{(i,j)}(\hat{\beta}_t \mid I_t=j, I_{t-1}=i, \mathbf{Y}_t) = \sum_{j=1}^{k} q_t^{(j)}(\hat{\beta}_t \mid I_t=j, \mathbf{Y}_t)$$

where

$$(\hat{\beta}_t \mid I_t=j, \mathbf{Y}_t) = (q_t^{(j)})^{-1}\sum_{i=1}^{k} p_{0\,t}^{(i,j)}(\hat{\beta}_t \mid I_t=j, I_{t-1}=i, \mathbf{Y}_t)$$

and the posterior probability is given by

$$q_t^{(j)} = Pr(i_t=j \mid \mathbf{Y}_t) = \sum_{i=1}^{k} Pr(I_t=j, I_{t-1}=i \mid \mathbf{Y}_t) = \sum_{i=1}^{k} p_{0\,t}^{(i,j)}$$

The mean vector of a mixture is the mixture of the mean vectors and the covariance matrix of a mixture is the mixture of the covariance matrices plus the mixture of the cross products of the deviations of the mean vectors from the overall mean vector. Hence, $(\hat{\beta}_t \mid I_t=j, \mathbf{Y}_t)$ is has mean vector $\beta_t + \mu_t^{(j)}$ and covariance matrix $\mathbf{V}_t^{(j)}$ where

$$\mu_t^{(j)} = (q_t^{(j)})^{-1}\sum_{i=1}^{k} p_{0\,t}^{(i,j)}\Theta_{0\,t}^{(i,j)}$$

and

$$\mathbf{V}_t^{(j)} = (q_t^{(j)})^{-1}\sum_{i=1}^{k} p_{0\,t}^{(i,j)}[\mathbf{U}_{1\,t}^{(i,j)} + (\Theta_{0\,t}^{(i,j)} - \mu_t^{(i,j)})(\Theta_{0\,t}^{(i,j)} - \mu_t^{(i,j)})']$$

From this we see that the prior estimator vector $(\hat{\beta}_t \mid I_t=j, \mathbf{Y}_t)$ is no longer an unbiased estimator of β_t. We correct the estimator by subtracting the mean vector t give the posterior estimator

$$(\hat{\beta}_t \mid I_t=j, \mathbf{Y}_t) = (\hat{\beta}_t \mid I_t=j, \mathbf{Y}_t) - \mu_t^{(i,j)}.$$

The correction applied is a weighted average of the corrections calculated by the Kalman filter conditional on each possible perturbation index value at time $t-1$ where the weights are the posterior probabilities of the perturbation index at time $t-1$ given the perturbation index at time t and the observations $\mathbf{Y}_t$. At this time we approximate the mixture distribution by a normal distribution having the same mean vector and covariance matrix. This can only be an approximation since a mixture of normal distrubutions is normal only in the trivial case that all mean vectors and covariance matrices are respectively equal. At this point there are k normally distributed unbiased parameter estimator vectors each conditional on the most recent state value and on all past observations, and posterior probabilities for the most recent parameter perturbation value given the past observations.

10. Let $t=t+1$ and go to step 1.

4. THE DISTRIBUTION OF THE FORECAST

The predicted value of the parameter vector at time $t+1$ is the composite estimator at time t brought forward to time $t+1$, $(\hat{\beta}_{t+1}|Y_t) = A_{t+1}(\hat{\beta}_t|Y_t) = \sum_{i=1}^{k} q_t^{(i)} A_{t+1}(\hat{\beta}|I_t=i, Y_t)$. Since the expected value of the perturbation vector equals the zero vector this is an unbiased estimator of the parameter vector at time $t+1$. Hence

$$(\hat{\beta}_{t+1}-\beta_{t+1}|Y_t) \text{ is } N[0; \sum_{i=1}^{k}\sum_{j=1}^{k} q_t^{(i)}\pi_{t+1}^{(j)} C_{t+1}^{(i,j)}]$$

where $C_{t+1}^{(i,j)} = A_{t+1}V_t^{(i)}A_{t+1}' + R_{t+1}^{(j)}$. The unbiased forecast of the observation vector is $y_{t+1} = X_{t+1}'(\hat{\beta}_{t+1}|Y_t$ since the expected value of the observation error vector is the zero vector. The covariance matrix of the prediction errors is given by

$$\Sigma_{y-y} = \sum_{i=1}^{k}\sum_{j=1}^{k}(q_t^{(i)}\pi_{t+1}^{(j)} X_{t+1}' C_{t+1}^{(i,j)} X_{t+1}) + \Sigma_\epsilon$$

Under the assumption that a mixture of normal random vectors has the normal distribution, the prediction error vector has the multivariate normal distribution, so prediction regions can be constructed.

5. PRACTICAL CONSIDERATIONS FOR IMPLEMENTATION

The correction mechanism of the algorithm combines two uses of Baye's theorem for each component of the observation vector. The first is the calculation of the posterior mean vector and covariance matrix of the state vector estimator by the Kalman filter, conditionally for each pair of previous and current perturbation index values. The second is the use of the prediction error for that component to recalculate the posterior probabilities of both the current and the previous perturbation indices. Thus the posterior probabilities of the perturbation index at time $t-1$ are calculated once after each component of y_{t-1} and again after each component of y_t. The use of two observation vectors is what enables the method to make distinctions between the types of perturbations and adjust the perturbation index probabilities accordingly.

The two aspects are brought together in the condensation of a mixture of k^2 normal random vectors into a mixture of k normal random vectors. The final correction is the weighted average of the conditional corrections where the weights are the conditional probabilities of the previous perturbation index values given the present perturbation index value and after analyzing the present observation vector. When it appears that a change in the underlying process generating the time series has occured due to particular type perturbation, there is a small correction due to each of the conditional applications of the Kalman filter. However the posterior probabilities of the previous perturbation index values do not change much, so the condensation phase doesn't generate much correction. However, the posterior probabilities of the current perturbation index values change considerably, in particular that type perturbation

becomes much more probable. Hence if the next observation also indicates that the change has occurred in the underlying process, there is a correction due to each of the conditional applications of the Kalman filter and the posterior probabilities of the previous perturbation values also change. Hence there is a large correction due to the condensation phase emphasizing that particular type perturbation. This allows the rapid adjustment of the parameters two observations after a change in the time series occurs.

The method requires knowledge of the observation error covariance matrix. In many cases this will have to be estimated from past data by using an initial smoothing process. A learning system could be incorporated and the covariance estimates improved over time.

The method also requires the perturbation index distribution and the covariance matrices of the perturbation distributions to be specified. Harrison and Stevens (1971) have shown that for univariate time series, the method is not sensitive to the covariances of the perturbation distributions assumed. In fact, changes of the order of 400 % had negligable effect on the root mean square forecast error. It is believed that the multivariate case will be similarly robust.

6. CONCLUSION

This algorithm is an efficient extension of the Harrison-Stevens forecasting method into multivariate time series which have correlated observation errors. It is a very robust method that is particularly suitable for forecasting in non-stationary time series. It uses Baye's theorem two ways, and despite the assumption of normality it can only be an approximation. The forecaster can intervene in the method by varying the prior perturbation index probabilities, hence the method is quite flexible. It also determines the posterior probabilities of the perturbation indices in situations where it is important to know when and what type change in the underlying process has occurred.

7. BIBLIOGRAPHY

Bolstad, W. M. (1986 A), "An Efficient Estimation Algorithm for the Seemingly Unrelated Regression Model With Known Contemporaneous Covariance Matrix", Mathematics Research Report No. 148, University of Waikato Mathematics Department, Hamilton, New Zealand.

Bolstad, W. M. (1986 B), "Harrison-Stevens Forecasting and the Multi-process Dynamic Linear Model", The American Statistician 40 No. 1, PP.

Duncan, B., and Horn, S. D. (1972), "Linear Dynamic Recursive Estimation from the Viewpoint of Regression Analysis", JASA 67, PP. 815-821.

Graybill, F. A., (1961), "an Introduction to Linear Statistical Models : Volume 1 " McGraw-Hill, New York.

Guttman, I., and Pena, D., (1985), "Comment, Robust Filtering", JASA 80, PP. 91-92.

Harrison, P. J., and Stevens, C. F. (1971), "A Bayesian Approach to Short-term Forecasting", Operations Research Quarterly 22, PP. 341-362.

Harrison, P. J., and Stevens, C. F. (1976), "Bayesian Forecasting" (with discussion), Journal of the Royal Statistical Society, Series B, 38, PP. 205-247.

Kalman, R. E. (1960), "A New Approach to Linear Filtering and Prediction Problems", Journal of Basic Engineering, PP. 95-108.

Plackett, R. L., (1950), "Some Theorems in Least Squares", Biometrika 37, PP. 149-157.

Sallas, W. M., and Harville, D. A. (1981), "Best Linear Recursive Estimation for Mixed Linear Models", JASA 76 , PP. 860-869.

West, M., Harrison, P. J., and Migon, H. S. (1985), "Dynamic Generalized Linear Models and Bayesian Forecasting", JASA 80 PP. 73-83.

Received by Editorial office April, 1986; Revised June, 1986.

Recommended Anonymously.

```
THIS PROGRAM USES THE MULTIVARIATE MULTIPROCESS DYNAMIC LINEAR MODEL
      implicit double precision (a-h,o-z)
      DIMENSION Bpost(4,18), bprior(4,4,18),bp(18)
      dimension A(18,18),at(18,18),xt(6,18),x(18,6),xta(6,18)
      dimension v(4,18,18),r(4,18,18),c(4,4,18,18),sigma(6,6)
      DIMENSION PI(4),q(4),post(4,4,6),pj(4,4,24),q2(4,4)
      DIMENSION theta(4,4,24),w(4,4,24,24),wtheta(4,4,24),umu(4,4)
      dimension thmimu(4,4,18)
      dimension yhat(4,4,6),ehat(4,4,6),obs(120,6),err(4,4)
      dimension ratio(120)
      dimension u11(4,4,24,24),u12(4,4,24),u22(4,4),pred(6)
      integer index,ipar,iobs,i,j,il,jl,k,k1,k2,k3
c bpost is the posterior estimator second index
c bprior is the prior estimator third index
c bp is the predicted parameter vector
c A is the dynamic matrix, x is the observation matrix
c xt is transpose of x xta is xt*a
C V is the posterior covariance matrix second and third indices
C r is the perturbation covariance matrix second and third indices
C c is the prior covariance matrix  third and fourth indices
C sigma is the observation covariance matrix
C pi is the prior probability,q is the posterior probability
C post is the posterior joint probability first and second indices
C theta is the mean vector third index
C w is the covariance matrix of the estimate and the mean vectors 3 & 4 index
C yhat is the predicted observation vector third index
C y is the observation vector
C ehat is the prediction error third index
c index is the number of possible index values
c 1=no change,2 is transient, 3 is mean change, 4 is slope change,
c ipar is the dimension of the parameter vector
c iobs is the dimension of the observation vector
      CALL ASSIGN(6,'mvHS.OUT')
      CALL ASSIGN(8,'mvhsinf.OUT')
      index=4
      pi(1)=.9
      pi(2)=.05
      pi(3)=.04
      pi(4)=.01
      q(1)=.9
      q(2)=.05
      q(3)=.04
      q(4)=.01
      ipar=18
      iobs=6
      do 40 k=1,ipar
      do 39 k1=1,iobs
      x(k,k1)=0
39    xt(k1,k)=0
      do 40 k1=1,ipar
      at(k1,k)=0
40    a(k,k1)=0
      do 41 k=1,iobs
      a(3*(k-1)+1,3*(k-1)+1)=1
      a(3*(k-1)+1,3*(k-1)+2)=1
      a(3*(k-1)+1,3*(k-1)+3)=.75
      a(3*(k-1)+2,3*(k-1)+2)=1
      xt(k,3*(k-1)+1)=1
      xt(k,3*(k-1)+3)=1
41    continue
      do 42 k=1,iobs
      do 42 k1=1,ipar
      x(k1,k)=xt(k,k1)
42    continue
      do 43 k=1,ipar
      do 43 k1=1,ipar
```

```
      a(K,k1)=a(k1,K)
      ichk=1
      if (ichk .eq. 1) write(6,45)
45    format( ' A(i,j) ' )
      if (ichk .eq. 1)  write(6,46) ((a(K,k1),k1=1,ipar),K=1,ipar)
46    FORMAT(1x,18f4.1)
      if (ichk .eq. 1) write (6,47)
47    format ( ' X(i,j) ')
      if (ichk .eq. 1)  write(6,48) ((x(K1,k),K=1,iobs),k1=1,ipar)
48    FORMAT(1x,6f4.1)
      if (ichk .eq. 1) write (6,49)
49    format ( ' XT(i,j) ')
      if (ichk .eq. 1)  write(6,50) ((xt(K,k1),k1=1,ipar),K=1,iobs)
50    FORMAT(1x,18f4.1)
      do 51 k=1,iobs
      do 51 k1=1,ipar
      xta(k,k1)=0
      do 51 k2=1,ipar
51    xta(k,k1)=xta(k,k1)+xt(k,k2)*a(k2,k1)
      if (ichk .eq. 1) write (6,52)
52    format ( ' XTA(i,j) ')
      if (ichk .eq. 1)  write(6,53) ((xta(K,k1),k1=1,ipar),K=1,iobs)
53    FORMAT(1x,18f4.1)
      open(unit=4,file='cpigrinf.dat',status='old')
      nobs=112
      do 55 n=1,nobs
     read(4,58) obs(n,1),obs(n,2),obs(n,3),obs(n,4),obs(n,5),obs(n,6)
   c ,ratio(n)
55    if (ichk .eq. 1) write(6,60) obs(n,1),obs(n,2),obs(n,3),obs(n,4)
   c ,obs(n,5),obs(n,6)
58    format(8x,7f9.5)
60    format(1x,6f8.5)
      open(unit=5,file='sigma.dat',status='old')
      do 80 k=1,iobs
      read(5,75) sigma(k,1),sigma(k,2),sigma(k,3),sigma(k,4),sigma(k,5),
   c sigma(k,6)
75    format(1x,6f13.10)
80    continue
      do 81 k=1,6
      do 81 k1=1,6
81    sigma (k,k1)=.1*sigma(k,k1)
      if (ichk .eq. 1) write(6,82)
82    format( ' SIGMA ' )
      if (ichk .eq. 1)  write(6,83) ((sigma(K,k1),k1=1,iobs),K=1,iobs)
83    FORMAT(1x,6f13.10)
      do 87 j=1,index
      do 87 k=1,ipar
      do 87 k1=1,ipar
87    r(j,k,k1)=0
      do 88 k=1,iobs
      do 88 k1=1,3
      r(2,(k-1)*3+3,(k-1)*3+3)=36*sigma(k,k)
      r(3,(k-1)*3+1,(k-1)*3+1)=16*sigma(k,k)
88    r(4,(k-1)*3+2,(k-1)*3+2)=.25*sigma(k,k)
      if (ichk .eq. 1) write(6,89)
89    format( ' R(i,j) ' )
      if (ichk .eq. 1)  write(6,90) (((R(j,K,k1),k1=1,ipar),K=1,ipar)
   c  ,j=1,index)
90    FORMAT(1x,18f5.3)

      do 95 i=1,index
      bpost(i,1)=1
      bpost(i,4)=1
      bpost(i,7)=1
      bpost(i,10)=1
      bpost(i,13)=1
```

```
      bpost(i,16)=1
      do 95 k=1,ipar
 95   v(i,k,k)=1
      if (ichk .eq. 1) write (6,96)
 96   format( ' bpost ' )
      if (ichk .eq. 1) write(6,97) ((bpost(i,k),k=1,ipar),i=1,index)
 97   FORMAT(1x,18(F7.4))
      if (ichk .eq. 1) write(6,98)
 98   format( ' V(i,k,k1) ' )
      if (ichk .eq. 1) write(6,99) (((v(i,K,k1),k1=1,ipar)
     c ,K=1,ipar),i=1,index)
 99   FORMAT(1x,18F7.4)
      do 300 n=1,nobs

      ichk=0
c statements to 110 are extrapolating the estimator from t-1 to t
 100  do 110 i=1,index
      do 110 j=1,index
      do 110 k=1,ipar
      bprior(i,j,k)=0
      do 110 k1=1,ipar
      bprior(i,j,k)=bprior(i,j,k)+a(k,k1)*bpost(i,k1)
 110  continue
      if (ichk .eq. 1) write(6,111)
 111  format ( ' bprior(i,j,k) ' )
      if (ichk .eq. 1) write(6,112) (((bprior(i,j,k),k=1,ipar)
     c ,j=1,index),i=1,index)
 112  FORMAT(1x,18F7.4)

c statements to 120 are extrapolating the covariance mat from t-1 to t
      do 120 i=1,index
      do 120 j=1,index
      do 120 k=1,ipar
      do 120 k1=1,ipar
      c(i,j,k,k1)=R(j,k,k1)
      do 120 k2=1,ipar
      do 120 k3=1,ipar
      c(i,j,k,k1)=c(i,j,k,k1)+a(k,k2)*V(i,k2,k3)*at(k3,k1)
 120  continue
      if (ichk .eq. 1) write(6,121)
 121  format ( ' c(i,j,k,k1)' )
      if (ichk .eq. 1) write(6,122) ((((c(i,j,k,k1),k1=1,ipar)
     c ,k=1,ipar),j=1,index),i=1,index)
 122  FORMAT(18F7.4)

c statemants to 130 are setting prior mean to zero
      m=ipar+iobs
      do 130 i=1,index
      do 130 j=1,index
      do 130 k=1,m
      theta(i,j,k)=0
 130  continue
c statements to 150 are setting prior covariance matrix
c and setting prior probabilities of perturbation index
      do 135 i=1,index
      do 135 j=1,index
      do 135 k=1,ipar
      do 135 k1=1,ipar
      w(i,j,k,k1)=c(i,j,k,k1)
 135  continue
      do 140 i=1,index
      do 140 j=1,index
      do 140 k=1,ipar
      do 140 k1=1,m-ipar
      w(i,j,k,ipar+k1)=0
      do 140 k2=1,ipar
```

```
        w(i,j,k,ipar+k1)=w(i,j,k,ipar+k1)-c(i,j,k,k2)*xt(k1,k2)
        w(i,j,ipar+k1,k)=w(i,j,k,ipar+k1)
 140    continue

        do 145 i=1,index
        do 145 j=1,index
        do 145 k=1,m-ipar
        do 145 k1=1,m-ipar
        w(i,j,ipar+k,ipar+k1)=sigma(k,k1)
        do 145 k2=1,ipar
        do 145 k3=1,ipar
        w(i,j,ipar+k,ipar+k1)=w(i,j,ipar+k,ipar+k1)+xt(k,k2)*c(i,j,k2,k3)
     c *xt(k1,k3)
 145    continue
        do 150 i=1,index
        do 150 j=1,index
        pj(i,j,ipar+iobs)=pi(j)*q(i)
 150    continue

c statements to 160 calculate estimation errors

        do 160 i=1,index
        do 160 j=1,index
        do 160 k=1,iobs
        yhat(i,j,k)=0
        do 155 k1=1,ipar
 155    yhat(i,j,k)=yhat(i,j,k)+xt(k,k1)*bprior(i,j,k1)
        ehat(i,j,k)=obs(n,k)-yhat(i,j,k)
 160    continue
 162    continue
        if (ichk .eq. 1) write (6,221)
        if (ichk .eq. 1) write (6,222) ((pj(i,j,m),j=1,index),i=1,index)
c       type *,n,m
        if (ichk .eq. 1) write(6,163)
  163   format(' w(i,j,k,k1) ')
        if (ichk .eq. 1) write(6,164) ((((w(i,j,k,k1),k1=1,m)
     c ,k=1,m),j=1,index),i=1,index)
  164   FORMAT(1x,24f5.1)


c statements to 200 update the mean vector
        do 180 i=1,index
        do 180 j=1,index
        do 175 k=1,m-1
        u12(i,j,k)=w(i,j,k,m)
        do 175 k1=1,m-1
 175    u11(i,j,k,k1)=w(i,j,k,k1)
        u22(i,j)=w(i,j,m,m)
 180    continue
        if (ichk .eq. 1) write (6,181)
  181   format ( ' u22(i,j) ')
        if (ichk .eq. 1) write (6,182) ((u22(i,j),j=1,index),i=1,index)
  182   format(1x, 4f18.15)
c statements to 200 update the mean vector
        do 200 i=1,index
        do 200 j=1,index
        do 200 k=1,m-1
        theta(i,j,k)=theta(i,j,k)+u12(i,j,k)*(ehat(i,j,m-ipar)-theta(i,j,m))/
     c  u22(i,j)
 200    continue
        if (ichk .eq. 1) write(6,201)
 201    format (' theta(i,j,k)')
        if (ichk .eq. 1) write(6,202) (((theta(i,j,k),k=1,m),j=1,index)
     c  ,i=1,index)
 202    format (1x,24f5.2)
```

```
c statements to 210 update the covariance matrix
        do 210 i=1,index
        do 210 j=1,index
        do 210 k=1,m-1
        do 210 kl=1,m-1
        w(i,j,k,kl)=u11(i,j,k,kl)-u12(i,j,k)*u12(i,j,kl)/u22(i,j)
 210    continue
        if (ichk .eq. 1) write(6,163)
        if (ichk .eq. 1) write(6,164) ((((w(i,j,k,kl),kl=1,m)
     c ,k=1,m),j=1,index),i=1,index)
c statements to 230 recalculate the last two posterior index probs
        sum=0
        do 220 i=1,index
        do 220 j=1,index
        err(i,j)=ehat(i,j,m-ipar)-theta(i,j,m)
      pj(i,j,m-1)=pj(i,j,m)*exp(-.5*err(i,j)**2/u22(i,j))/sqrt
     c (u22(i,j))
        sum=sum+pj(i,j,m-1)
 220    continue
 221    format( ' pj(i,j) ')
 222    format(1x,4f10.8)
        if (ichk .eq. 1) write (6,223)
  223   format ( ' err(i,j) ')
        if (ichk .eq. 1) write (6,224) ((err(i,j),j=1,index),i=1,index)
  224   format(1x, 4f18.15)
c       type *,sum
        sum1=0
        do 230 i=1,index
        do 230 j=1,index
        pj(i,j,m-1)=pj(i,j,m-1)/sum
        sum1=sum1+pj(i,j,m-1)
 230    continue
        m=m-1

        if (m .GT. ipar) go to 162
        ichk=0
c the statements to 250 calculate the condensed mean and estimator
        do 235 j=1,index
        sum=0
        do 233 i=1,index
 233    sum=sum+pj(i,j,ipar)
 235    q(j)=sum

        do 240 i=1,index
        do 240 j=1,index
        if (q(j) .gt. 0) then
        q2(i,j)=pj(i,j,ipar)/q(j)
        else
        q2(i,j)=pi(i)
        end if
 240    continue
        if (ichk .eq. 1) write (6,221)
        if (ichk .eq. 1) write (6,222) ((pj(i,j,ipar),j=1,index),i=1,index)
        do 250 j=1,index
        do 250 k=1,ipar
        umu(j,k)=0
        bpost(j,k)=0
        do 250 i=1,index
        umu(j,k)=umu(j,k)+q2(i,j)*theta(i,j,k)
        bpost(j,k)=bpost(j,k)+q2(i,j)*bprior(i,j,k)
 250    continue
        if (ichk .eq. 1) write (6,251)
 251    format ( ' q2(i,j) ')
        if (ichk .eq. 1) write (6,252) ((q2(i,j),j=1,index),i=1,index)
 252    format ( 1x,4f10.8)
        if (ichk .eq. 1) write (6,253)
```

```
 250    format ( ' q(j) ')
        if (ichk .eq. 1) write (6,252) (q(j),j=1,index)
        if (ichk .eq. 1) write (6,261)
 261    format ( ' umu(j,k)')
        if (ichk .eq. 1) write (6,262) ((umu(j,k),k=1,ipar),j=1,index)
 262    format ( 1x,18f7.4)

c the statements to 270 calculate the condensed cov matrix
        do 265 i=1,index
        do 265 j=1,index
        do 265 k=1,ipar
 265    thmimu(i,j,k)=theta(i,j,k)-umu(j,k)

        do 270 j=1,index
        do 270 k=1,ipar
        do 270 k1=1,ipar
        V(j,k,k1)=0
        do 270 i=1,index
        v(j,k,k1)=v(j,k,k1)+q2(i,j)*(w(i,j,k,k1)+(theta(i,j,k)-umu(j,k))
     c  *(theta(i,j,k1)-umu(j,k1)))
 270    continue
c the statements to 280 calculate the posterior estimator
        do 280 j=1,index
        do 280 k=1,ipar
        bpost(j,k)=bpost(j,k)-umu(j,k)
 280    continue
c       type *,((bpost(j,k),k=1,ipar),j=1,index)
c       type *,(q(j),j=1,index)
        if (ichk .eq. 1) write (6,281)
 281    format ( ' bpost(j,k)')
        if (ichk .eq. 1) write (6,282) ((bpost(j,k),k=1,ipar),j=1,index)
 282    format ( 1x,18f7.4)
        if (ichk .eq. 1) write (6,283)
 283    format ( ' v(j,k,k1)')
        if (ichk .eq. 1) write (6,284) (((v(j,k,k1),k1=1,ipar)
     c  ,k=1,ipar),j=1,index)
 284    format ( 1x,18f7.4)
        do 286 k=1,iobs
 286    pred(k)=0
        do 288 k=1,ipar
        bp(k)=0
        do 288 j=1,index
 288    bp(k)=bp(k)+bpost(j,k)*q(j)
c       type *,(bp(k),k=1,ipar)
        do 290 k=1,iobs
        do 290 k1=1,ipar
 290    pred(k)=pred(k)+xta(k,k1)*bp(k1)
c       type *,(pred(k),k=1,iobs)
        write (8,295)((obs(n,k),k=1,iobs),(pred(k),k=1,iobs),ratio(n) )
 295    format(13f10.7)

 300    continue

        end
```

Biometrika (1988), 75, 4, *pp.* 685-92
Printed in Great Britain

The multiprocess dynamic linear model with biased perturbations: A real time model for growth hormone level

BY WILLIAM M. BOLSTAD
Department of Mathematics, University of Waikato, Hamilton, New Zealand

SUMMARY

This paper introduces biased perturbation distributions into the multiprocess dynamic linear model in order to represent growth hormone levels in animals, which are characteristically noisy pulsatile time series. The use of biased perturbation distributions allows a pulse to be detected on the first observation immediately after it occurs, thus allowing the process to be modelled in real time. This approach is found to be quite effective at detecting the occurrence of a pulse in growth hormone level, and revising the parameter estimates and predictions accordingly.

Some key words: Harrison-Stevens forecasting; Mixture distribution; Multiprocess Kalman filter; Pulsatile time series.

1. INTRODUCTION

The measured concentration levels of certain biochemicals in biological organisms frequently exhibit the characteristics of noisy pulsatile time series, as does the plasma growth hormone concentration described by Brier et al. (1986). Thus, the concentration usually declines exponentially towards a base level, however, sometimes a pulse occurs and the concentration rises a large amount. Both the time between pulses and the pulse magnitudes are unpredictable. The process is sampled at equally spaced time points and the observations are subject to measurement errors. Models for this process must account for both the dynamic behaviour and the measurement error characteristics of the series. Recursive estimation of the parameters of a dynamic model makes use of both the dynamic and error characteristics of the series. Since the parameter estimates are updated each time an observation is taken, it is possible to model the series in real time, thus enabling researchers to monitor the effects of experimental intervention.

The multiprocess dynamic linear model was used by Harrison & Stevens (1971) to make short term forecasts in a time series that is subject to abrupt changes in pattern and transient effects. The changes are modelled by adding a random perturbation vector having zero mean to the linearly updated parameter vector. The covariance matrix of the perturbation takes on one of k possibilities depending on the current value of the perturbation index variable, the outcome of a multinomial trial with known prior probabilities. After each observation, the parameter vector estimator, conditional on the perturbation index variables, is updated by the Kalman filter applied to the dynamic linear model (Duncan & Horn, 1972). Then the posterior probabilities of the perturbation index variable are calculated. The resulting posterior distribution of the parameter vector estimator, a mixture of k mixtures of k normal distributions, is condensed into a mixture of k normal distributions, each normal having the same first two moments as the corresponding mixture component. This returns the algorithm to its initial conditions

ready for the next observation. A clear exposition of the algorithm, and why it is able to distinguish between a real change and a transient effect following the second subsequent observation, is given by Bolstad (1986).

West, Harrison & Migon (1985) developed the dynamic generalized linear model where the observation distribution is from a known one-dimensional exponential family, and a known function of the mean is a linear function of the parameters. A multiprocess extension of this model has been developed by Bolstad (1988).

In the present paper a known bias is introduced into the perturbation distribution, involving a nonzero perturbation mean to allow for the sudden pulse in the time series. The algorithm outlined in §§ 2–4 resembles that of Harrison & Stevens (1971) in that Bayes's theorem is used both for parameter estimation and for model selection, but using the nonzero perturbation mean allows the pulse to be detected following the first subsequent observation rather than the second. Thus the model is appropriate for real time estimation and prediction in noisy pulsatile time series. In § 5, the growth hormone data described by Brier et al. (1986) are analysed.

2. THE MULTIPROCESS DYNAMIC LINEAR MODEL WITH BIASED PERTURBATIONS

In the multiprocess dynamic linear model, the dynamic behaviour of the parameter vector has both a deterministic and a random part. The parameter dynamic equation is

$$\beta_t = A_t\beta_{t-1} + r_t,$$

where β_t is the $\nu \times 1$ parameter vector at time t, A_t is a known matrix of coefficients at time t, and r_t is the perturbation vector at time t, whose distribution depends on the value of the perturbation index variable at time t. In this model we are allowing the perturbation distribution to be biased, so $r_t \sim N(\eta_t^{(i)}, R_t^{(i)})$ when the perturbation index variable $I_t = i$ for $i = 1, \ldots, k$. The sequence of perturbation index variables are independent of each other, and each can be considered the outcome of a single multinomial trial with known prior probabilities, $\pi_t^{(i)} = \text{pr}\,(I_t = i)$ for $i = 1, \ldots, k$. Prior knowledge can be incorporated in the system by allowing the prior probabilities to change with time.

The observation y_t is governed by the observation equation

$$y_t = X_t'\beta_t + e_t,$$

where the observation errors e_t are independent $N(0, \sigma_e^2)$ random variables that are independent of the perturbation random vectors and the perturbation index random variables. Here X_t' is the known row vector of observation coefficients at time t.

3. THE RECURSIVE ALGORITHM

The algorithm starts at time $t-1$ with an unbiased parameter estimator vector conditional on all the observations up to and including the most recent, y_{t-1}. Its conditional distribution is assumed to be normal with mean vector equal to the parameter vector, and a known covariance matrix. Let $y_{t-1} = y_{t-1}, y_{t-2}, \ldots, y_1$ denote the set of observations up to and including time $t-1$. Then $(\hat{\beta}_{t-1} - \beta_{t-1}|y_{t-1}) \sim N(0, V_{t-1})$, where V_{t-1} is the covariance matrix given the past observations y_{t-1}.

If at time t the perturbation index variable $I_t = i$, then the perturbation vector $r_t \sim N(\eta_t^{(i)}, R_t^{(i)})$. Let

$$(\hat{\beta}_t|I_t = i, y_{t-1}) = A_t(\hat{\beta}_{t-1}|y_{t-1}) + \eta_t^{(i)}$$

be the prior estimator of the parameter vector at time t. It is the conditional parameter estimator vector brought forward to time t given $I_t = i$, but prior to observation y_t. The prior distribution of the conditional random variable $(\hat{\beta}_t - \beta_t | I_t = i, y_{t-1})$ is $N(0, C_t^{(i)})$, where the covariance matrix $C_t^{(i)} = A_t V_t A_t' + R_t^{(i)}$. The predicted value of the observation at time t given $I_t = i$ is

$$\hat{y}_t^{(i)} = X_t'(\hat{\beta}_t | I_t = i, y_{t-1}).$$

The joint random variables

$$\begin{pmatrix} \hat{\beta}_t - \beta_t | I_t = i, y_{t-1} \\ y_t - \hat{y}_t^{(i)} | I_t = i, y_{t-1} \end{pmatrix} \sim N\left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}; \begin{pmatrix} C_t^{(i)} & -C_t^{(i)} X_t \\ -X_t' C_t^{(i)} & X_t' C_t^{(i)} X_t + \sigma_e^2 \end{pmatrix} \right\}.$$

Hence the conditional random variable $(\hat{\beta}_t - \beta_t | I_t = i, y_t)$ has the distribution $N(\mu_t^{(i)}, V_t^{(i)})$, where

$$\mu_t^{(i)} = 0 - C_t^{(i)} X_t (X_t' C_t^{(i)} X_t + \sigma_e^2)^{-1} (y_t - \hat{y}_t^{(i)}),$$
$$V_t^{(i)} = C_t^{(i)} - C_t^{(i)} X_t (X_t' C_t^{(i)} X_t + \sigma_e^2)^{-1} X_t' C_t^{(i)}.$$

The prior probability of the perturbation index variable is $\text{pr}(I_t = i) = \pi_t^{(i)}$. Since

$$(y_t | I_t = i, y_{t-1}) \sim N(\hat{y}_t^{(i)}, X_t' C_t^{(i)} X_t + \sigma_e^2),$$

the posterior probability of the perturbation index variable is

$$q_t^{(i)} = \pi_t^{(i)} \{f(y_t)\}^{-1} (2\pi)^{-\frac{1}{2}} |X_t' C_t^{(i)} X_t + \sigma_e^2|^{-\frac{1}{2}} \exp\{-\tfrac{1}{2}(X_t' C_t^{(i)} X_t + \sigma_e^2)^{-1} (y_t - \hat{y}_t^{(i)})^2\}$$

for $i = 1, \ldots, k$. This is completely determined, because the unconditional density function $f(y_t)$ is a proportionality constant in all the probabilities, and they sum to one. Note that calculation of these posterior probabilities implicitly assumes normality. Thus the joint probability-probability density function of $(\hat{\beta}_t, I_t | y_t)$ is

$$f(\hat{\beta}_t, i | y_t) = q_t^{(i)} (2\pi)^{-\frac{1}{2}\nu} |V_t^{(i)}|^{-\frac{1}{2}} \exp\{-\tfrac{1}{2}(\hat{\beta}_t - \beta_t - \mu_t^{(i)})' (V_t^{(i)})^{-1} (\hat{\beta}_t - \beta_t - \mu_t^{(i)})\}$$

for $i = 1, \ldots, k$.

To find the marginal probability density function of $(\hat{\beta}_t | y_t)$, we sum the joint probability-probability density function over $i = 1, \ldots, k$. Thus

$$f(\hat{\beta}_t | y_t) = \sum_{i=1}^{k} q_t^{(i)} (2\pi)^{-\frac{1}{2}\nu} |V_t^{(i)}|^{-\frac{1}{2}} \exp\{-\tfrac{1}{2}(\hat{\beta}_t - \beta_t - \mu_t^{(i)})' (V_t^{(i)})^{-1} (\hat{\beta}_t - \beta_t - \mu_t^{(i)})\}.$$

This posterior distribution is a mixture of k normal distributions. The mean vector of a mixture is the mixture of the mean vectors, and the covariance matrix of a mixture is the mixture of the covariance matrices plus the mixture of the cross products of the deviations of the mean vectors from the overall mean vector. Hence the estimator

$$\hat{\beta}_t | y_t = \sum_{i=1}^{k} q_t^{(i)} (\hat{\beta}_t | I_t = i, y_t)$$

has mean vector $\beta_t + m_t$, where the bias vector and covariance matrix are

$$m_t = \sum_{i=1}^{k} q_t^{(i)} \mu_t^{(i)}, \quad V_t = \sum_{i=1}^{k} q_t^{(i)} \{V_t^{(i)} + (\mu_t^{(i)} - m_t)(\mu_t^{(i)} - m_t)'\}.$$

Thus the prior estimator $(\hat{\beta}_t | y_{t-1})$ is no longer an unbiased estimator of β_t. The posterior estimator is found by subtracting the bias vector to give

$$(\hat{\beta}_t | y_t) = (\hat{\beta}_t | y_{t-1}) - m_t.$$

The correction m_t is a weighted average of the corrections calculated by the Kalman filter conditional on each possible value of the perturbation index at time t, where the weights are the posterior probabilities of the perturbation index at time t given the observation. At this time we approximate the distribution of the posterior estimator by a normal distribution having the same mean vector and covariance matrix. This is only an approximation since a mixture of normals is only normal in the trivial case when all the mean vectors and covariance matrices are respectively equal. We are now in the same position as we were when the algorithm started, so we are ready to repeat the algorithm when the next observation becomes available.

4. MODELLING PULSATILE TIME SERIES

In a pulsatile time series, the level of a variable is geometrically decaying towards a base level, except when a pulse occurs and the level rises by an unpredictable amount. This can be modelled by a multiprocess dynamic linear model with biased perturbations using the parameter vector $\beta_t = (\lambda_t, b)'$, where λ_t is the level at time t and b is the base level. The dynamic matrix is

$$A_t = \begin{bmatrix} a & 1-a \\ 0 & 1 \end{bmatrix},$$

where a is the decay factor between 0 and 1. There are two possible values of the perturbation index variable, $I_t = 1$ when there is no pulse, and $I_t = 2$ when there is a pulse. The corresponding perturbation distributions are that, if $I_t = 1$, r_t is zero with probability one; and, if $I_t = 2$, the first component of r_t is $N(\eta_t, \sigma_t^2)$ and the second component is zero. The row vector of observation coefficients is $X_t' = (1\ \ 0)$, the observation errors e_t are independent $N(0, \sigma_e^2)$ random variables, $\sigma_e^2 \ll \sigma_\lambda^2$, and $\pi_t^{(1)} \gg \pi_t^{(2)}$.

The algorithm is easily able to detect when a pulse occurs because the posterior probability that the perturbation index variable equals one is nearly one whenever an unexpectedly large value of the observation occurs. Thus, when a pulse occurs, there is a large correction of the estimated parameter value on the first subsequent observation. The algorithm was tested on simulated series and was found to be very effective at detecting a pulse and quickly reacting to it.

5. MODELLING GROWTH HORMONE LEVEL

The growth hormone concentration data from the experiments described by Brier et al. (1986) relating growth hormone level with dietary stress were analysed using the model developed in the previous sections. In these experiments, a group of fifteen steers were divided into a control group and two treatment groups of five steers each. Blood samples from these steers were taken every 15 minutes over a 25 hour interval, at the end of three time periods. During time period I, five weeks, each steer was on a normal nutrition plane. During time period II, four weeks, the control group remained on a normal nutrition plane while the two treatment groups of the steers were subjected to moderate and severe dietary stress, a reduced nutrition plane at one of two levels, respectively. During time period III, three weeks, all the steers were back on the normal nutrition plane.

There was a considerable variation in the growth hormone level exhibited between individual steers and between the treatment groups. In particular, it appeared that the

base level was different for each steer for each of the treatments, so a different base level was assigned to each steer-treatment combination. The geometric decay rate $a = 0{\cdot}55$ and the variances $\sigma_e^2 = 20$ and $\sigma_\lambda^2 = 60$ were found to minimize the sum of squares of one-step prediction errors over all steer-treatment combinations. In Fig. 1(a), 2(a) and 3(a) are shown the estimated growth hormone level profiles and the observations for three steers, one from each of the three respective treatments; and in Fig. 1(b), 2(b) and 3(b) are the corresponding posterior probabilities of a pulse. Steer 8 is in the control group, steer 14 is in the moderately stressed group, and steer 9 is in the severely stressed group. These animals exhibit typical patterns for their respective groups. Note the increased pulsatility of growth hormone level for the observations taken during time period II for steer 14 and steer 9 while they are under dietary stress, in both pulse magnitude and pulse frequency. The results for these steers during time period III show a residual effect of increased pulse magnitude, but decreased frequency.

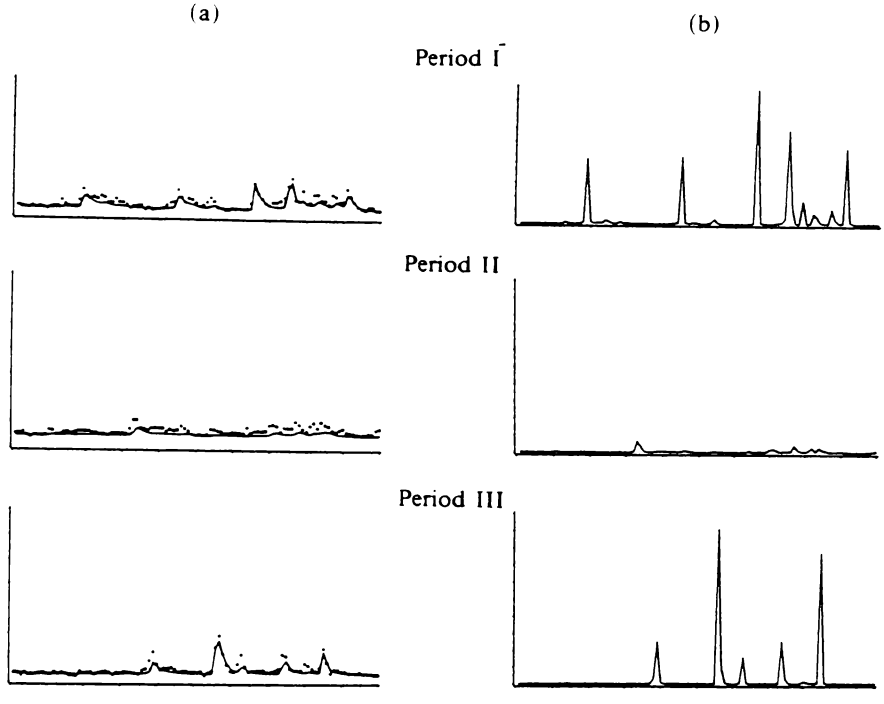


Fig. 1. Steer 8. Time periods I, II, III. Control group: normal nutrition throughout. (a) Growth hormone level versus time, (b) Posterior probability of pulse versus time.

A smoothed growth hormone profile was found for each steer at each time period by taking a five point moving average. The mean of these smoothed growth hormone levels was taken over each group at each time period, and these are shown in Fig. 4. During the period of dietary stress, not only is there a general increase in growth hormone level, but periods of high growth hormone release occurred episodically during the 24 hours with considerable synchronization between steers. This effect is still apparent, although at a reduced level, during time period III. Similarly, mean smoothed posterior probability profiles were found for each group at each time, and are shown in Fig. 5. These show that dietary stress, time period II, groups 2 and 3, are associated with increased pulsatility in growth hormone release; periods of increased pulsatile activity occur episodically

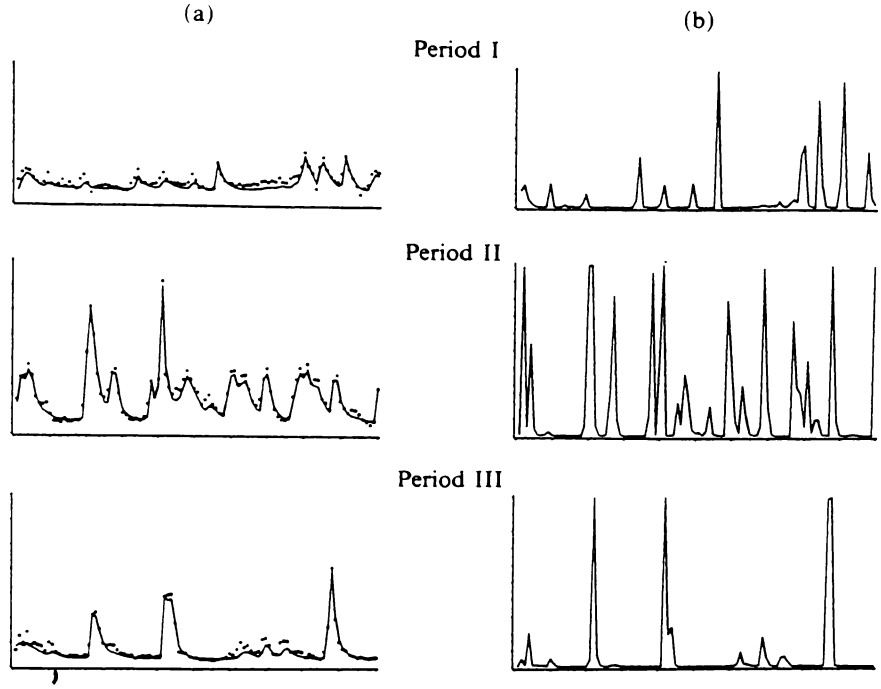


Fig. 2. Steer 14. Time periods I, II, III. Moderate dietary stress in period II. (a) Growth hormone level versus time, (b) Posterior probability of pulse versus time.

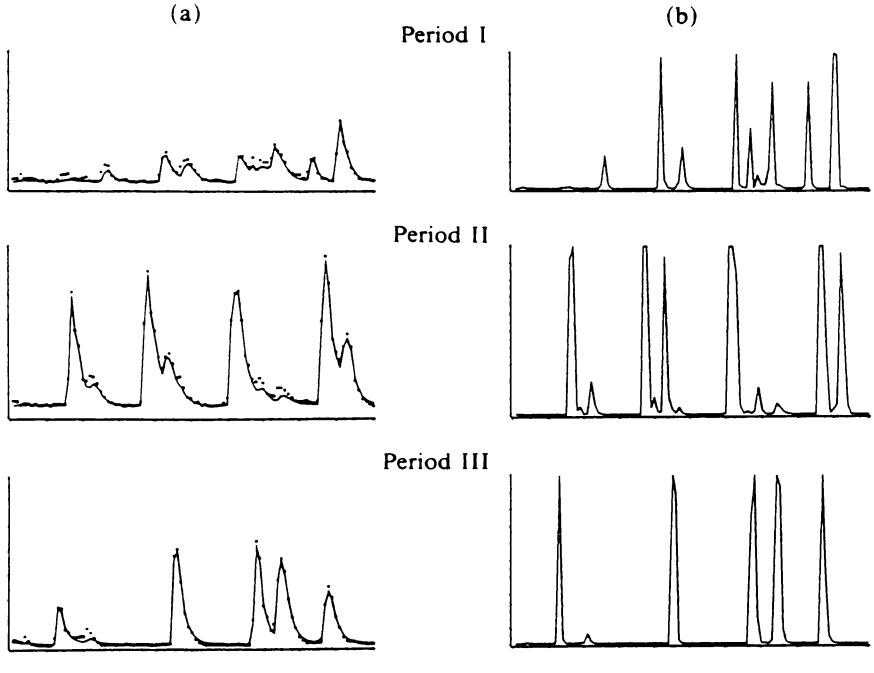


Fig. 3. Steer 9. Time periods I, II, III. Severe dietary stress in period II. (a) Growth hormone level versus time, (b) Posterior probability of pulse versus time.

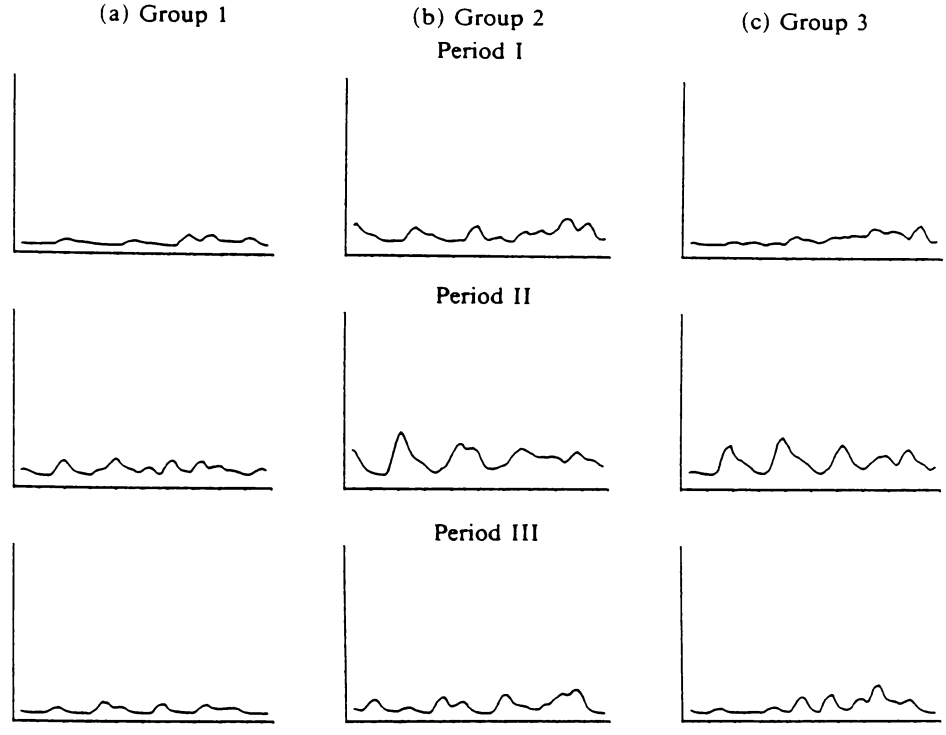


Fig. 4. Mean smoothed group hormone level versus time. Time periods I, II, III. (a) Group 1, control; (b) Group 2, moderate stress; (c) Group 3, severe stress.

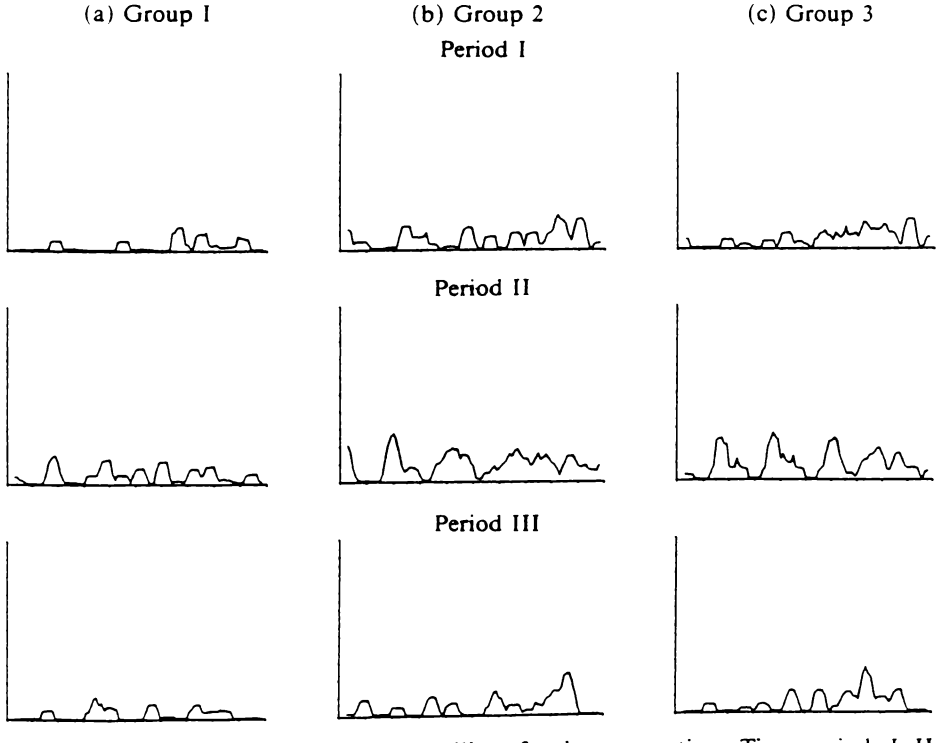


Fig. 5. Mean smoothed posterior probability of pulse versus time. Time periods I, II, III. Group 1, control; Group 2, moderate stress; Group 3, severe stress.

Table 1. *Mean and standard deviation of smoothed posterior probability and smoothed growth hormone level for each treatment group at each time period*

		Smoothed posterior prob.		Smoothed growth hormone level	
Group	Period	Mean	St. dev.	Mean	St. dev.
1	I	0·029	0·034	46·0	12·0
	II	0·053	0·044	73·9	17·6
	III	0·030	0·031	44·4	12·4
2	I	0·064	0·051	88·1	24·8
	II	0·100	0·062	120·0	37·3
	III	0·054	0·053	69·5	25·4
3	I	0·059	0·046	77·6	21·3
	II	0·091	0·072	111·6	37·4
	III	0·055	0·055	60·0	27·5

throughout the 24 hour day; and while the pulses themselves are not synchronized between steers, the pulsatile activity rates exhibit considerable synchronization. Table 1 gives the mean and standard deviations for smoothed posterior probabilities and smoothed growth hormone level for each group at each time period. The increase in mean that occurs in groups 2 and 3 during time period II reflects the increased pulsatile release of growth hormone and the increase in the standard deviation reflects that the pulsatile activity is synchronized.

ACKNOWLEDGEMENTS

The author is grateful to the editor, the referee, and to Professor R. Hosking for their helpful comments, and to Dr J. Bass of Ruakura Animal Research Station and Dr B. Brier of the University of Auckland for the data.

REFERENCES

BOLSTAD, W. M. (1986). Harrison-Stevens forecasting and the multiprocess dynamic linear model. *Am. Statistician* **40**, 129-36.

BOLSTAD, W. M. (1988). Estimation in the multiprocess dynamic generalized linear model. *Comm. Statist.* A **17** (12). To appear.

BRIER, B. H., BASS, J. J., BUTLER, J. H. & GLUCKMAN, P. D. (1986). The somatotrophic axis in young steers: Influence of nutritional status on pulsatile release of growth hormone and circulating concentrations of insulin-like growth factor 1. *J. Endocrinol.* **111**, 209-15.

DUNCAN, B. & HORN, S. D. (1972). Linear dynamic recursive estimation from the viewpoint of regression analysis. *J. Am. Statist. Assoc.* **67**, 815-21.

HARRISON, P. J. & STEVENS, C. F. (1971). A Bayesian approach to short-term forecasting. *Oper. Res. Quart.* **22**, 341-62.

WEST, M., HARRISON, P. J. & MIGON, H. (1985). Dynamic generalized linear models and Bayesian forecasting (with discussion). *J. Am. Statist. Assoc.* **80**, 73-97.

[*Received July* 1987. *Revised June* 1988]

```
C THIS PROGRAM ANALYZES THE GROWTH HORMONE LEVEL USING THE MULTIPROCESS
C DYNAMIC LINEAR MODEL WITH BIASED PERTURBATIONS
        implicit double precision (a-h,o-z)
        DIMENSION B(2,2),V(2,2,2),rmu(2,2),q(2),se(2),beta(2),umu(2)
        DIMENSION PI(2), eta(2,2)
        DIMENSION y(100),BEST(2),BPRED(2),VEST(2,2),VPRED(2,2),W1(2,2)
        dimension w(2,2),w3(2),w2(2),w4(2)
        dimension u(2,2),u2(2,2),c(2,2,2),yhat(2)
        DIMENSION A(2,2),r(2,2,2),X(2),cov(2,2)
        dimension s(15,3,100)
        INTEGER m,I,J,K,l,N,ichk
        REAL  VE,OBS,u,err,vcast,secast
C B CONTAINS THE conditional parameter VECTORS -second INDEX
c beta contains the parameter vector
c bpred contains the conditional prediction for the parameter vector
c vpred contains the conditional predicted covariance matrix
C V CONTAINS THE conditional parameter VECTOR COVARIANCE
C             MATRICES-second & third INDICES
c cov contains the parameter vector covariance matrix
c u contains the matrix A*cov
c u2 contains the matrix A*cov*A'
c C contains the matrix A*cov*A'+R the parameter estimator covariance matrix
C PI CONTAINS THE PRIOR PROBABILITIES q CONTAINS THE POSTERIOR PROBABILITIES
C l=level 2=base level
C A CONTAINS THE parameter TRANSISTION EQUATION coefficients
C R CONTAINS THE PERTURBAseTION COVARIANCE MATRICES-SECOND & third INDICES
C X CONTAINS THE OBSERVATION EQUATION coefficients
C Y IS THE TIME SERIES
C VE IS THE VARIANCE OF THE TIME SERIES
c s(i,j,n) is the n'th obs on steer i at occasion j
        ichk=0
        ve=30**2
        open(unit=4,file='hormone.dat',status='old')
        do 45  n=1,100
   45   read(4,60) s(1,2,n),s(1,3,n)
   60   format(4x,f3.0,x,f3.0)
        beta(1)=100
        eta(1,1)=0
        eta(2,1)=200
        cov(1,1)=9*ve
        cov(1,2)=0
        cov(2,1)=0
        cov(2,2)=0
        R(2,1,1)=9*ve
        R(2,2,2)=0
        PI(1)=.9
        PI(2)=.1
C FOR PI AND PO, 1=NO CHANGE,2=impulse
        x(1)=1
        X(2)=0
        DO 189 K=1,2
        DO 189 L=1,2
        A(K,L)=0
  189   CONTINUE
        ak=.6
        A(1,1)=ak
        A(1,2)=1-ak
        A(2,2)=1
        do 70 m=1,2
        if (m .eq. 1) then
        CALL ASSIGN(6,'ss1.OUT')
        else if (m .eq. 2) then
        call ASSIGN(6,'st1.OUT')
        end if
        if (m .eq. 1) then
        beta(2)=30
```

```
      else if (m .eq. 2) then
      beta(2)=15
      end if
      DO 320 N=1,100
      if (m .eq. 1) then
      y(n)=s(1,2,n)
      else if (m .eq. 2) then
      y(n)=s(1,3,n)
      end if
  320 continue
      do 330 n=1,100
      DO 190 K=1,2
      DO 190 L=1,2
  190 b(k,l)=0
  191 FORMAT(2(2(' A(',I1,',',I1,')=',F10.4)/))
C FOR A the dynamic coefficient matrix first = level, second = base
      DO 193 i=1,2
      DO 193 j=1,2
      do 192 k=1,2
  192 b(i,j)=b(i,j)+a(j,k)*beta(k)
  193 b(i,j)=b(i,j)+eta(i,j)
      bpred(1)=b(1,1)
      bpred(2)=b(2,1)
      if (ichk .eq. 1)  write(6,291) ((k,l,b(k,l),l=1,2),k=1,2)
      if (ichk .eq. 1)  write(6,218) (((i,j,k,v(i,j,k),k=1,2),j=1,2),i=1,2)
  291 format(2(2( ' b(',i1,',',i1,')=',F10.4)/))
      DO 194 i=1,2
      DO 194 j=1,2
  194 u(i,j)=0
      DO 195 j=1,2
      DO 195 k=1,2
      DO 195 l=1,2
  195 u(j,k)=u(j,k)+a(j,l)*cov(l,k)
      if (ichk .eq. 1)  WRITE(6,199) ((K,L,u(K,L),L=1,2),K=1,2)
  199 format(2(2( ' u(',i1,',',i1,')=',F10.4)/))
      OBS=Y(N)
      if (ichk .eq. 1)  write(6,196) obs
  196 format('obs=',F10.4)
      DO 197 I=1,2
      DO 197 J=1,2
  197 u2(i,j)=0
      do 198 i=1,2
      do 198 j=1,2
      do 198 k=1,2
  198 u2(i,j)=u2(i,j)+u(i,k)*a(j,k)
      if (ichk .eq. 1)  write(6, 93) (( i,j,u2(i,j),j=1,2),i=1,2)
   93 FORMAT(2(2( ' u2(',I2,',',I2,')=',F10.4)/))
      do 200 i=1,2
      do 200 j=1,2
      do 200 k=1,2
  200 c(i,j,k)=u2(k,j)+r(i,j,k)
      if (ichk .eq. 1) write(6,299) (((i,j,k,c(i,j,k),k=1,2),j=1,2),
     c i=1,2)
  299 format(2(2(2( ' c(',i1,',',i1,',',i1,')=',F10.4)/)/))
      do 201 i=1,2
  201 yhat(i)=0
      do 202 i=1,2
      do 202 j=1,2
  202 yhat(i)=yhat(i)+x(j)*b(i,j)
      if (ichk .eq. 1) write(6,203) (i,yhat(i),i=1,2)
  203 format(2( ' yhat(',i1,')=',F10.4))
      do 204 i=1,2
      do 204 j=1,2
      w(i,j)=0
  204 continue
      do 206 i=1,2
```

```
      do 206 j=1,2
      do 206 k=1,2
206   w(i,j)=w(i,j)+c(i,j,k)*x(k)
      if (ichk .eq. 1) write(6,294) ((i,j,w(i,j), j=1,2),i=1,2)
294   format(2(2( ' w(',i1,',',i1,')=',F10.4)/))
      do 208 i=1,2
      w2(i)=0
208   continue
      do 210 i=1,2
      do 210 j=1,2
      w2(i)=w2(i)+x(j)*w(i,j)
210   CONTINUE
      do 212 i=1,2
212   w2(i)=w2(i)+ve
      if (ichk .eq. 1) write (6,213) (i,w2(i),i=1,2)
213   format(2( ' w2(',i1,')=',F10.4))
      do 214 i=1,2
      do 214 j=1,2
214   rmu(i,j)=-w(i,j)*(obs-yhat(i))/(w2(i))
      if (ichk .eq. 1) write(6,215) ((i,j,rmu(i,j),j=1,2),i=1,2)
215   format(2(2( ' rmu(',i1,',',i1,')=',F10.4)/))
      do 216 i=1,2
      do 216 j=1,2
      do 216 k=1,2
216   v(i,j,k)=w(i,j)*w(i,k)/w2(i)
      if (ichk .eq. 1) write(6,218) (((i,j,k,v(i,j,k),k=1,2),
     c  j=1,2),i=1,2)
218   format(2(2(2(' v(',I1,',',I1,',',I1,')=',F10.4)/)/))
      cpi=1/sqrt(2*3.14159)
      do 219 i=1,2
      w3(i)=cpi/(sqrt(w2(i)))
219   w3(i)=w3(i)*exp(-.5*(yhat(i)-obs)**2/w2(i))
      sum=0
      DO 220 I=1,2
      SUM=SUM+pi(i)*w3(i)
220   CONTINUE
      do 222 i=1,2
222   q(i)=pi(i)*w3(i)/sum
      if (ichk .eq. 1) write(6,224) (i,q(i), i=1,2)
224   format(2( ' q(',i1,') =',F10.4)/)
      do 230 j=1,2
230   umu(j)=0
      do 232 j=1,2
      do 232 i=1,2
232   umu(j)=umu(j)+q(i)*rmu(i,j)
      if (ichk .eq. 1) write(6,233) (j,umu(j),j=1,2)
233   format(2( ' umu(',i1,')=',F10.4)/)
      do 234 j=1,2
      beta(j)=0
      do 234 k=1,2
234   cov(j,k)=0
      do 235 j=1,2
      do 235 i=1,2
235   beta(j)=beta(j)+q(i)*b(i,j)
      do 238 j=1,2
238   beta(j)=beta(j)-umu(j)
      if (ichk .eq. 1) write(6,240) (j,beta(j),j=1,2)
      do 239 j=1,2
      do 239 k=1,2
      do 239 i=1,2
239   cov(j,k)=cov(j,k)+q(i)*(v(i,j,k)+(rmu(i,j)-umu(j))
     c  *(rmu(i,k)-umu(k)))
240   format( 2( ' beta(',i1,') =',F10.4))
      if (ichk .eq. 1 ) write (6,242) ((j,k,cov(j,k),k=1,2),j=1,2)
242   format( 2(2( ' cov(', i1,',',i1,') =', F10.4)/))
      do 290 k=1,2
```

```
      DO 290 L=1,2
290   b(k,l)=0
      DO 293 i=1,2
      DO 293 j=1,2
      do 292 k=1,2
292   b(i,j)=b(i,j)+a(j,k)*beta(k)
293   b(i,j)=b(i,j)+eta(i,j)
      bpred(1)=b(1,1)
      bpred(2)=b(2,1)
      write(6,300) (obs,beta(1),beta(2),b(1,1),q(1),q(2),v(1,1,1),
   c  v(2,1,1), cov(1,1),bpred(1),bpred(2))
300   format (11F10.4)
330   CONTINUE
      close(unit=6)
 70   continue
      close(unit=4)
      END
```

COMMUN. STATIST.-THEORY METH., 17(12), 4179-4204 (1988)

ESTIMATION IN THE MULTIPROCESS DYNAMIC GENERALIZED LINEAR MODEL

William M. Bolstad

Department of Mathematics
University of Waikato
Hamilton
New Zealand

Key words and phrases: Bayesian forecasting; dynamic discount Bayesian model; fault detection; Harrison-Stevens forecasting; Kalman filter; multiprocess models; state space models.

ABSTRACT

The dynamic generalized linear model and the dynamic discount Bayesian model have been used to describe processes involving time-varying parameters. This paper develops an estimation algorithm for the multiprocess extension of these models. These algorithms have the same characteristics as Harrison-Stevens forecasting, namely insensitivity to outliers and quick reaction to real change in the parameters.

1. INTRODUCTION

Engineers have been analyzing the behavior of control systems in the time domain by using state space models for nearly thirty years. The stochastic

behavior of the observations is described in terms of an unobserved state vector whose dynamic behavior is Markovian, and consists of a linear deterministic component plus a random component called the perturbation. In particular, Kalman (1960) developed an algorithm for the recursive estimation of the state variables of a linear dynamic system, and he showed it to be the optimal linear filter for both symmetric convex and squared error loss functions. Ho and Lee (1964) approached the dynamic linear model under the general unifying framework of Bayesian decision theory. Assuming normality, they found that the mean and variance of the exact posterior distribution of the state vector are given by the Kalman filter equations. Meinhold and Singapuralla (1983) showed how the Bayesian interpretation of the Kalman filter simplifies its derivation and hence aids understanding.

The study of linear dynamic systems was also used as a fault detection procedure, in control systems where abrupt shifts in the estimated series may indicate a degradation or breakdown in some instrument. Willsky (1976) examined the statistical performance of failure detection methods, with respect to both speed of response to failure and robustness to normal operations. He concluded that sophisticated digital failure detection algorithms may allow the reduction in hardware redundancy without loss of system reliability.

Although control systems engineers used the Kalman filter extensively, for several years it was largely ignored in the field of statistics. Duncan and Horn (1972) introduced the Kalman filter into the statistical literature by relating the dynamic linear model to random β regression theory, using the time varying random parameters as the state variables. They showed the Kalman filter estimates are the minimum variance unbiased linear estimators and found by an extension of the Gauss-Markov theorem that they are also minimum mean square linear estimators The Kalman filter and dynamic linear model can operate under either of two alternative sets of assumptions; either the normal assumptions where all the variables assumed normal with known first two moments, or the less restrictive assumptions where only the first two moments are known for all the variables. The optimal properties mentioned above hold under both sets of assumptions. In addition, under the normal assumptions the posterior distribution of the estimator is normal, and the Kalman filter estimates are the minimum variance unbiased estimates and minimum mean squared estimates.

Dynamic linear models with other non-normal perturbation distributions or observation errors have also been analyzed. Masrielez and Martin (1977) developed robust Bayesian estimates for a state space model where either the state noise is Gaussian and the observation noise is heavy tailed, or vice-versa. Smith and Miller (1986) developed a model for an exponential observation distribution and censored data which they applied to the prediction of athletic records. Kitagawa (1987) developed state space models for non-stationary time series, where the observation and perturbation noise distributions are non-normal. The Bayesian recursion formulae for prediction, smoothing, and filtering are developed and the updating is done numerically, where the densities are approximated by piecewise linear functions. However, Martin and Raftery (1987) note that when the state vector is multi-dimensional, heavy computational effort required precludes real time applications of this method.

Ameen and Harrison (1985) developed normal discount Bayesian models, in order to overcome some practical disadvantages of dynamic linear models, and they also demonstrated a correspondence between the two models. In normal discount Bayesian models, the parameter covariance matrix is updated by pre and post multiplication by a discount matrix, instead of by the addition of the covariance matrix of the perturbation error. They felt that many practicioners had better intuition for this method of allowing for the increasing uncertainty due to the passing of time.

Nelder and Wedderburn (1972) introduced the (static) generalized linear model in order to model situations where an independent sequence of random variables are taken, and a function of their means is linked to a linear predictor on known covariates. This allows regression techniques to be used for non-normal (but exponential family) distributions on a non-linear scale. West, Harrison, and Migon (1985) then developed the dynamic generalized linear model, where the mean values of the parameters change over time according to some linear pattern, and the covariance matrix is discounted as in Ameen and Harrison (1985). They developed simple recursions for updating the parameter estimates and covariance estimates (posterior means and covariances of the parameter distribution) that have the same form as the recursions used in the Kalman filter. Their algorithm uses a "guide relationship" to choose the conjugate prior, and thus the estimates are only approximations. Guttman and

Pena (1985) allow the observation distribution to be a mixture of two distributions, the usual distribution and an outlier distribution. The posterior probability of each is computed, and the posterior estimator distribution is replaced by a single normal distribution having the same first two moments as the mixture. This gives very little weight to any outlying observations, thus providing an outlier filter for the dynamic linear model.

The multiprocess dynamic linear model was introduced by Harrison and Stevens (1971) to model time series that contain outliers and are subject to abrupt, as well as evolutionary, changes in pattern. The different types of pattern change correspond to values of a variable called the perturbation index variable. In the estimating algorithm, the parameter perturbation distribution is a mixture of k normal distributions, each depending on the current value of the perturbation index variable which comes from an independent sequence of multinomial random variables; and the prior distribution is a mixture of k normal distributions, each depending on the previous value of the perturbation index. The posterior distribution of the state vector, a mixture of k^2 normal distributions, conditional on the last two values of the perturbation index, is calculated and then condensed into a mixture of k normal distributions conditional on only the most recent value of the index. Makov (1983) concluded that condensing the mixture distributions into a single distribution, referred to as "probabilistic editing", is in the long run more accurate than the "quasi Bayes" procedure where the most likely posterior distribution is used. Harrison-Stevens forecasting, sometimes called Bayesian forecasting or the multiprocess Kalman filter, was shown to be an effective forecasting tool for non-stationary time series. It responds very quickly when there is an underlying change in the pattern of the time series, yet is not sensitive to outliers, so it can be used as a fault detection procedure. Smith and West (1983) and Smith et al. (1983) did so when monitoring renal transplant data, in order to detect a change of state (organ rejection) as quickly as possible

Harrison and Stevens (1976) have summarized the foundations of Bayesian forecasting. Essential to the methods are i) the parametric or statespace model, ii) the probabilistic information on the model parameters, iii) the sequential model definition which describes the dynamic behavior of the model

parameters, and iv) some uncertainty in choosing the underlying model from a number of possibilities. The present author has analyzed the Harrison-Stevens forecasting algorithm and provided a bibliography of previous applications of Harrison-Stevens methods (Bolstad 1986 A), and subsequently developed a computationally efficient algorithm for Harrison-Stevens forecasting involving multivariate observations with contemporaneously correlated errors (Bolstad 1986 B).

The purpose of this paper is to develop the multiprocess dynamic generalized linear model, by incorporating the perturbation index variable which determines the perturbation distribution. This extension of the Harrison-Stevens estimation algorithm to the multiprocess dynamic generalized linear model permits a quick response to underlying changes in the model, yet without heavy influence from outliers. The estimation algorithm for the multiprocess generalized discount Bayesian model is also developed. Here the index variable determines which of several possible discount matrices is used, not the perturbation distribution.

2. THE MULTIPROCESS DYNAMIC GENERALIZED LINEAR MODEL

The *generalized linear model* is an extension of the well known general linear model in two ways. First, the sampling distribution of the random variable is assumed to be a known distribution from the one-dimensional exponential family, but not necessarily normal. Second, it is not the mean itself, but rather a function of the mean that is an unknown linear function of known covariates. This function of the mean is called the link function. Thus, the generalized linear model allows the extension of regression techniques to non-normal sampling models with effects on a nonlinear scale.

The underlying assumptions for the generalized linear model are:

i) The random variables $y_1, \cdots, y_n$ are mutually independent.

ii) y_i is a random variable from the one dimensional exponential family which has probability density function

$$f(y_i \mid \theta_i, \phi) = \exp[c(y_i, \phi) + \phi(\theta_i y_i - b(\theta_i))] \tag{2.1}$$

where θ_i is the unknown parameter of observation y_i and ϕ is a known scale parameter. (When y is binomial or Poisson, $\phi = 1$.) The mean and

variance of observation y_i are given by $\mu_i = E(y_i) = b'(\theta_i)$ and $var(y_i) = \phi^{-1}b''(\theta_i)$ respectively.

iii) The systematic component of the model is η, the linear predictor given by

$$\eta_i = X_i'\beta \tag{2.2}$$

where $X_i'=(X_{i1}, \ldots, X_{ik})$ are the known values of the covariates for the $i'th$ observation and β is the vector of unknown parameters.

iv) The link between the systematic part of the model and the random part of the model is given by

$$g(\mu_i)=\eta_i \tag{2.3}$$

where μ_i is the expected value of the $i'th$ observation and η_i is the linear predictor. The link function g is assumed known.

The *dynamic generalized linear model* allows the parameter vector β to change over time according to the parameter dynamic equation

$$\beta_t = A_t\beta_{t-1} + r_t \tag{2.4}$$

where A_t is a matrix of known coefficients and r_t, the perturbation vector at time t, has mean vector $\vec{0}$ and the covariance matrix R_t. The perturbation vectors at different times are independent of each other. This allows the parameter vector to behave in a Markovian manner. Instead of using the observation equation which relates the observation to the linear parameters and measurement error, the observation distribution is explicitly assumed to be the known one-dimensional exponential family member with sampling parameter θ_t, where the sampling parameter is related to the linear parameters β_t by the guide relationship defined in West, Harrison, and Migon (1985).

In the multiprocess extension of this model, the perturbation distribution at time t is allowed to take on one of k possibilities conditional on the current value of the perturbation index variable. The different perturbation distributions in the model allow for possible abrupt changes in the time series, by allowing the parameters to change very quickly when an abrupt change occurs. The term "perturbation index variable" is used in preference to the term "state variable" as used by Harrison and Stevens (1971) as it avoids dual meanings.

for the word state - see Bolstad (1986 B). The perturbation index variables $\{I_t\}$ are a sequence of independent multinomial trials with known prior probabilities $P(I_t=j)=\pi_t^{(j)}$ for $j=1,\cdots,k$. The modeller can introduce prior knowledge into the system by varying the prior probabilities over time. At time t, the conditional mean vector and covariance matrix of the perturbation for the perturbation index variable $I_t=j$ are 0 and $\mathbf{R}_t^{(j)}$ respectively.

3. THE RECURSIVE ESTIMATION ALGORITHM

Initial conditions

The algorithm for recursively estimating the parameters of the multiprocess dynamic generalized linear model is developed in this section. It generalizes the algorithm given by West, Harrison, and Migon (1985), by incorporating the multiprocess approach of Harrison and Stevens (1971).

The initial conditions for the algorithm at time $t-1$ include the requirement that the first two moments for each of k conditional posterior distributions are known. Each distribution is conditional on the perturbation index variable at time $t-1$ having been one of the k possible values, and posterior to analyzing all the observations up to and including the most recent, y_{t-1}. The notation $(\beta_{t-1}\,|\,I_{t-1}=i,\mathbf{y}_{t-1}) \sim (\hat{\beta}_{t-1}^{(i)},\mathbf{V}_{t-1}^{(i)})$ indicates the mean vector and covariance matrix of $(\beta_{t-1}\,|\,I_{t-1}=i,\mathbf{y}_{t-1})$ are equal to $\hat{\beta}_{t-1}^{(i)}$ and $\mathbf{V}_{t-1}^{(i)}$ respectively. The notation $\mathbf{y}_{t-1}=y_{t-1},y_{t-2},\cdots$ denotes all the observations up to and including y_{t-1}. Also required is that $q_{t-1}^{(i)}=P(i_{t-1}=i\,|\,\mathbf{y}_{t-1})$, the posterior probabilities of the perturbation index variable at time $t-1$ posterior to all the observations up to and including y_{t-1}, are known.

Each of these k distributions is updated to time t conditional on $I_t=j$, the perturbation index variable at time t being equal to j for $j=1,\ldots,k$.

$$(\beta_t\,|\,I_{t-1}=i,I_t=j,\mathbf{y}_{t-1}) \sim (\hat{\beta}_t^{(i,j)},\mathbf{C}_t^{(i,j)}) \tag{3.1}$$

where $\hat{\beta}_t^{(i,j)}=\mathbf{A}_t\hat{\beta}_{t-1}^{(i)}$ and $\mathbf{C}_t^{(i,j)}=\mathbf{A}_t\mathbf{V}_{t-1}^{(i)}\mathbf{A}_t'+\mathbf{R}_t^{(j)}$. The first two moments of the joint conditional distribution of the parameters and the linear predictor are

$$\begin{bmatrix}\beta_t\,|\,I_{t-1}=i,I_t=j,\mathbf{y}_{t-1}\\ \eta_t\,|\,I_{t-1}=i,I_t=j,\mathbf{y}_{t-1}\end{bmatrix} \sim \left[\begin{pmatrix}\hat{\beta}_t^{(i,j)}\\ \mathbf{X}_t'\hat{\beta}_t^{(i,j)}\end{pmatrix},\ \begin{pmatrix}\mathbf{C}_t^{(i,j)} & \mathbf{C}_t^{(i,j)}\mathbf{X}_t\\ \mathbf{X}_t'\mathbf{C}_t^{(i,j)} & \mathbf{X}_t'\mathbf{C}_t^{(i,j)}\mathbf{X}_t\end{pmatrix}\right]. \tag{3.2}$$

Hence the conditional distribution of the parameter vector given the linear predictor is

$$(\beta_t \mid I_{t-1}=i, I_t=j, \eta_t, y_{t-1}) \sim [(\hat{\beta}_t^{(i,j)}+C_t^{(i,j)}X_t(X_t'C_t^{(i,j)}X_t)^{-1}(\eta_t - X_t'\hat{\beta}_t^{(i,j)})), \quad (3.3)$$

$$(C_t^{(i,j)}-C_t^{(i,j)}X_t(X_t'C_t^{(i,j)}X_t)^{-1}X_t'C_t^{(i,j)})].$$

These equations are the same form as the Kalman filter equations applied conditionally on the two previous perturbation indices, where η_t is used in place of the observation. At this point η_t is not yet known, but in the next phase its conditional estimates will be found posterior to observation y_t.

Correction phase

In this phase we find the posterior distributions of the parameters, given the observation. The corrections required are determined by the prediction errors. Let the prior distribution of the sampling distribution parameter $(\theta_t \mid I_{t-1}=i, I_t=j, y_{t-1})$ be approximated by the conjugate prior distribution $CP(\gamma_t^{(i,j)}, \delta_t^{(i,j)})$ where $\gamma_t^{(i,j)}$ and $\delta_t^{(i,j)}$ are chosen to give the same first two moments as $h^{-1}(\eta_t \mid I_{t-1}=i, I_t=j, y_{t-1})$ where the function $h(\,)=g(b(\,))$. The relationship between the first two moments of the sampling parameter prior distribution and the moments of the linear predictor is called the "guide relationship" by West, Harrison, and Migon (1985), and is meant to be a device for the choice of a reasonable prior distribution, rather than an exact reality. Clearly the joint density is.

$$f(y_t, \theta_t \mid I_{t-1}=i, I_t=j, y_{t-1}) = \exp[c(y_t,\phi)+c(\gamma_t^{(i,j)}, \delta_t^{(i,j)})+\theta_t(\gamma_t^{(i,j)}+y_t\phi) \quad (3.4)$$

$$-(\delta_t^{(i,j)}+\phi)b(\theta_t)]$$

and

$$f(y_t \mid I_{t-1}=i, I_t=j, y_{t-1}) = \int f(y_t, \theta_t \mid I_{t-1}=i, I_t=j, y_{t-1}) d\theta_t \quad (3.5)$$

$$= \exp[c(y_t,\phi)+c(\gamma_t^{(i,j)}, \delta_t^{(i,j)})-c(\gamma_t^{(i,j)}+\phi y_t, \delta_t^{(i,j)}+\phi)].$$

Hence the posterior distribution of $(\theta_t \mid I_{t-1}=i, I_t=j, y_t)$ is the conjugate posterior $CP(\gamma_t^{(i,j)}+\phi y_t, \delta_t^{(i,j)}+\phi)$ having probability density function

$$g(\theta_t \mid I_{t-1}=i, I_t=j, y_t) = \exp[c(\gamma_t^{(i,j)}+\phi y_t, \delta_t^{(i,j)}+\phi)+\theta_t(\gamma_t^{(i,j)}+\phi y_t)-(\delta_t^{(i,j)}+\phi)b(\theta_t)]. \quad (3.6)$$

The guide relationship is used to relate the posterior distribution of the sampling parameter back to the distribution of the linear predictor. A first order

Taylor series approximation to the linear predictor is $\eta_t = h(\theta_t) = h(\hat{\theta}_t) + h'(\hat{\theta}_t)(\theta_t - \hat{\theta}_t)$, where $\hat{\theta}_t$ is the mean of the posterior distribution. Hence

$$(\eta_t \mid I_{t-1}=i, I_t=j, y_t) \sim (\hat{\eta}_t^{(i,j)}, U_t^{(i,j)}) \tag{3.7}$$

where $\hat{\eta}_t^{(i,j)} = h(\hat{\theta}_t^{(i,j)})$ and $U_t^{(i,j)} = (h'(\hat{\theta}_t^{(i,j)}))^2 var(\theta_t^{(i,j)})$. This completes the determination of the first two moments of the conditional distributions of the linear predictor at time t posterior to the observations y_t.

Now we need to relate this back to find the moments of the conditional posterior distributions of the parameters. Noting

$$f(\beta_t, \eta_t \mid I_{t-1}=i, I_t=j, y_t) = f(\beta_t \mid I_{t-1}=i, I_t=j, \eta_t, y_{t-1}) \times f(\eta_t \mid I_{t-1}=i, I_t=j, y_t) \tag{3.8}$$

and taking expectations and variances, we have

$$E(\beta_t \mid I_{t-1}=i, I_t=j, y_t) = E(E(\beta_t \mid I_{t-1}=i, I_t=j, \eta_t, y_{t-1}) \mid y_t) \tag{3.9}$$

and

$$\begin{aligned} Var(\beta_t \mid I_{t-1}=i, I_t=j, y_t) &= E(var(\beta_t \mid I_{t-1}=i, I_t=j, \eta_t, y_{t-1}) \mid y_t) \\ &\quad + var(E(\beta_t \mid I_{t-1}=i, I_t=j, \eta_t, y_{t-1}) \mid y_t). \end{aligned} \tag{3.10}$$

Hence from (3.3) we have

$$(\beta_t \mid I_{t-1}=i, I_t=j, y_t) \sim (\hat{\beta}_t^{(i,j)}, V_t^{(i,j)}) \tag{3.11}$$

where

$$\hat{\beta}_t^{(i,j)} = \hat{\beta}_t^{(i,j)} + C_t^{(i,j)} X_t (X_t' C_t^{(i,j)} X_t)^{-1} (\hat{\eta}_t^{(i,j)} - X_t' \hat{\beta}_t^{(i,j)}) \tag{3.12}$$

and

$$\begin{aligned} V_t^{(i,j)} &= var(\beta_t \mid I_{t-1}=i, I_t=j, y_t) \\ &= C_t^{(i,j)} - C_t^{(i,j)} X_t (X_t' C_t^{(i,j)} X_t)^{-1} X_t' C_t^{(i,j)} \\ &\quad + C_t^{(i,j)} X_t (X_t' C_t^{(i,j)} X_t)^{-1} var(\eta_t \mid I_{t-1}=i, I_t=j, y_t) (X_t' C_t^{(i,j)} X_t)^{-1} X_t' C_t^{(i,j)}. \end{aligned} \tag{3.13}$$

This completes the determination of the posterior distributions of the parameters.

Determining the posterior index probabilities

Now we need to determine the posterior probabilities of the perturbation indices given the present observation. Using Bayes theorem, we have

$$\begin{aligned} p_t^{(i,j)} &= P(I_{t-1}=i, I_t=j \mid y_t) \qquad (3.14)\\ &= P(I_{t-1}=i, I_t=j, y_t \mid y_{t-1})/f(y_t \mid y_{t-1}) \\ &= P(y_t \mid I_{t-1}=i, I_t=j, y_{t-1}) \times P(I_{t-1}=i, I_t=j \mid y_{t-1})/f(y_t \mid y_{t-1}) \\ &= q_{t-1}^{(i)} \pi_t^{(j)} \exp[c(y_t,\phi)+c(\gamma_t^{(i,j)}+y_t\phi, \delta_t^{(i,j)}+\phi)]/f(y_t \mid y_{t-1}) , \end{aligned}$$

for $i = 1,\ldots,k$ and $j = 1,\ldots,k$. The quantity $f(y_t \mid y_{t-1})$ is a constant in each term and they sum to one, hence the $p_t^{(i,j)}$ are all completely determined.

Condensation phase

At this stage there are k^2 posterior mean vectors and covariance matrices, one for each of the distributions conditional on the last two values of the perturbation index variables and on all the observations including the most recent; and there are the posterior probabilities of the last two values of the perturbation index variable, given all the observations. What we want are the k posterior mean vectors and covariance matrices for the distributions conditional only on the most recent value of the perturbation index variable and all the observations. Noting

$$f(\beta_t \mid I_t=j, y_t) = \sum_{i=1}^{k} (q_t^{(j)})^{-1} p_t^{(i,j)} f(\beta_t \mid I_{t-1}=i, I_t=j, y_t) , \qquad (3.15)$$

where $q_t^{(j)} = \sum_{i=1}^{k} p_t^{(i,j)}$, we have

$$\hat{\beta}_t^{(j)} = E(\beta_t \mid I_t=j, y_t) = \sum_{i=1}^{k} (q_t^{(j)})^{-1} p_t^{(i,j)} \hat{\beta}_t^{(i,j)} \qquad (3.16)$$

because the mean vector of a mixture is the mixture of the mean vectors; and

$$V_t^{(j)} = var(\beta_t \mid I_t=j, y_t) = \sum_{i=1}^{k} (q_t^{(j)})^{-1} p_t^{(i,j)} [V_t^{(i,j)} + (\hat{\beta}_t^{(i,j)} - \hat{\beta}_t^{(j)})(\hat{\beta}_t^{(i,j)} - \hat{\beta}_t^{(j)})'] \qquad (3.17)$$

because the covariance matrix of a mixture is the mixture of the covariance matrices plus the mixture of the cross products of the deviations of the mean vectors from the overall mean vector. At this point, we are in the same

position as we were when we started the algorithm, so we are ready to repeat the process when the next observation becomes available.

Prediction using the model

The first two moments of the random vector $(\beta_{t+1} | I_t = i, I_{t+1} = j, y_t)$ are $\hat{\beta}_{t+1}^{(i,j)}$ and $C_{t+1}^{(i,j)}$ where $\hat{\beta}_{t+1}^{(i,j)} = A_{t+1}\hat{\beta}_t^{(i)}$ and $C_{t+1}^{(i,j)} = A_{t+1}V_t^{(i)}A_{t+1}' + R_{t+1}^{(j)}$. The prior distribution of $(\theta_{t+1} | I_t = i, I_{t+1} = j, y_t)$ is the conjugate prior $CP(\gamma_{t+1}^{(i,j)}, \delta_{t+1}^{(i,j)})$, where $\gamma_{t+1}^{(i,j)}$ and $\delta_{t+1}^{(i,j)}$ are chosen to give the first two moments as indicated by the guide relationship. Hence the conditional distribution of the next observation is

$$f(y_{t+1} | y_t) = \sum_{i=1}^{k}\sum_{j=1}^{k} q_t^{(i)}\pi_{t+1}^{(j)} \exp[c(y_{t+1}, \phi) + c(\gamma_{t+1}^{(i,j)}, \delta_t^{(i,j)}) - c(\gamma_{t+1}^{(i,j)}) + \phi y_{t+1}, \delta_{t+1}^{(i,j)} + \phi)] \quad (3.18)$$

This can be used to construct prediction intervals.

Summary of the effects of the algorithm

The correction mechanism of the algorithm combines two uses of Bayes' theorem. First is the calculation of the posterior mean vector and covariance matrix, conditionally for each pair of previous and current perturbation index variables. The second is the calculation of the posterior probabilities for the previous and current perturbation index variables. Hence the posterior probabilities of the perturbation index at time $t-1$ are calculated twice, first after observation y_{t-1}, and then again after observation y_t. The use of two observations enables the distinction to be drawn between the different types of perturbations, and the perturbation index probabilities are adjusted accordingly. The two aspects are brought together in the condensation of a mixture of k^2 posterior distributions into a mixture of k posterior distributions. The final posterior mean vectors are the weighted average of the conditional posterior mean vectors, where the weights are the conditional probabilities of the previous perturbation index variable, given the present perturbation index variable and all the observations including the present.

When it appears that a change in the underlying process generating the series has occurred due to a particular type of perturbation, each of the k^2 conditional posterior distributions is somewhat changed due to the prediction error, the greatest change occurring in those conditional on that type of perturbation. The posterior probabilities of the previous perturbation index change very little, so the condensation phase only changes the k conditional posterior distributions

a modest amount. However, the posterior probabilities of the current perturbation index values change considerably, and in particular the posterior probability of that change becomes much greater. If the next observation also indicates that the same type of change has occurred, again each of the k^2 conditional posterior distributions changes. However, in this case the posterior probabilities of the previous perturbation index are changed considerably. Thus in the condensation phase the most weight is given to the distribution showing the greatest change, so there is a large change due to condensation phase emphasizing that particular type perturbation. This allows a rapid adjustment of the parameters two observations after a change in the time series occurs. Thus if the model was being used as a fault detection method, the fault would become apparent on the second subsequent observation.

4. MULTIPROCESS GENERALIZED DISCOUNT BAYESIAN MODEL

Ameen and Harrison (1985) introduced normal discount Bayesian models to overcome some practical disadvantages of the dynamic linear model. Modeling the parameter change by introducing a perturbation may not be appropriate. Instead of updating the parameter covariance matrix by adding the perturbation covariance matrix, the normal discount Bayesian model updates it by pre and post multiplication by a discount matrix. This also has the same effect of increasing the variances, and in many cases modellers and forecasters have a more intuitive feel for the appropriate discount matrix than for a perturbation covariance matrix. In this section we introduce the multiprocess generalized discount Bayesian model, and develop the estimation algorithm for it.

The assumptions of the dynamic discount Bayesian model are the same as for the dynamic generalized linear model, except that if the previous posterior conditional distribution is

$$(\beta_{t-1} \mid y_{t-1}) \sim (\hat{\beta}_{t-1}, V_{t-1}) , \tag{4.4}$$

then the next prior conditional distribution is given by

$$(\beta_t \mid y_{t-1}) \sim (A_t \hat{\beta}_{t-1} , B_t A_t V_t A_t' B_t)$$

where A_t is the known matrix of dynamic coefficients at time t and B_t is the discount matrix, a diagonal matrix of discount factors.
The effect is similar to

that of adding a perturbation in the dynamic generalized linear model. The mean of the subsequent prior distribution is unchanged, and the covariance matrix has been inflated to allow for increased uncertainty. However, the covariance matrix inflation is multiplicative instead of additive, and this produces some slight differences. If the posterior is very precise, the subsequent prior would be more precise for the generalized discount Bayesian model than for the dynamic generalized linear model. If the posterior is very imprecise, the opposite would hold. The generalized discount Bayesian model could be put in the dynamic generalized linear model form by allowing the perturbation covariance matrix to depend on the posterior covariance matrix.

The multiprocess extension of this model allows the discount matrix to have one of k possible values, $\mathbf{B}_t^{(1)}, \cdots, \mathbf{B}_t^{(k)}$, depending on the value of the discount index variable, I_t. The discount index variables $\{I_t\}$ are an independent sequence of multinomial random trials with known prior probabilities $P(I_t=j) = \pi_t^{(j)}$, which may change over time.

The recursive estimation algorithm

The algorithm starts at time $t-1$ with the first two moments of k posterior conditional distributions $(\beta_{t-1} \mid I_{t-1}=i, \mathbf{y}_{t-1}) \sim (\hat{\beta}_{t-1}, \mathbf{V}_{t-1}^{(i)})$, and with the posterior probabilities $q_{t-1} = P(I_{t-1}=i \mid \mathbf{y}_{t-1})$. Each of these k distributions is updated to time t conditional on $I_t=j$ to give $(\beta_t \mid I_{t-1}=i, I_t=j, \mathbf{y}_{t-1}) \sim (\hat{\beta}_t^{(i,j)}, \mathbf{C}_t^{(i,j)})$, where $\hat{\beta}_t^{(i,j)}=\mathbf{A}_t \hat{\beta}_{t-1}^{(i)}$ and $\mathbf{C}_t^{(i,j)}=\mathbf{B}_t^{(j)}\mathbf{A}_t \mathbf{V}_t^{(i)}\mathbf{A}_t'\mathbf{B}_t^{(j)}$

The rest of the algorithm follows along the same lines as the algorithm for the multiprocess dynamic generalized linear model developed in section 3. The (moments of the) conditional distribution of $(\beta_t^{(i,j)} \mid I_{t-1}=i, I_t=j, \eta_t, \mathbf{y}_{t-1})$ are found using (3.3) and the moments of the conditional predictors $(\eta_t \mid I_{t-1}=i, I_t=j, \mathbf{y}_t)$ are found using (3.7). The posterior moments of $(\beta_t \mid I_{t-1}=i, I_t=j, \mathbf{y}_t)$ are found using (3.12) and (3.13), and the posterior probabilities of the index variables are determined by (3.14). The moments of $(\beta_t \mid I_t=j, \mathbf{y}_t)$ are found by condensing the mixture of k^2 random variables into a mixture of k random variables by using (3.16) and (3.17). At this point, we have returned to the starting conditions of the algorithm, so we are ready to take another observation.

Prediction using the model

The random vector $(\beta_{t+1} \mid I_t=i, I_{t+1}=j, \mathbf{y}_t) \sim (\hat{\beta}_{t+1}^{(i,j)}, \mathbf{C}_{t+1}^{(i,j)})$ where $\hat{\beta}^{(i,j)} = \mathbf{A}_{t+1}\hat{\beta}_t^{(i)}$ and $\mathbf{C}_t^{(i,j)} = \mathbf{B}_{t+1}^{(j)}\mathbf{A}_{t+1}\mathbf{V}_t^{(i)}\mathbf{A}_{t+1}'\mathbf{B}_{t+1}^{(j)}$. The prior distribution of

$(\theta_{t+1} | I_t=i, I_{t+1}=j, y_t)$ is the conjugate prior $CP(\gamma_{t+1}^{(i,j)}, \delta_{t+1}^{(i,j)})$. Hence the conditional observation density is the same as that given in (3.18), and it can be used to construct prediction intervals.

Summary of the effects of the algorithm

The algorithm for the multiprocess dynamic discount Bayesian model has similar characteristics to the algorithm for the multiprocess dynamic generalized linear model. The posterior distributions, the posterior probabilities of the discount index variable, and also the condensation of k^2 posterior distributions (conditional on the two most recent values of the discount index variable) into k posterior distributions dependent on the most recent value of the discount index variable all proceed in the analagous manner. Thus in the condensation phase, the most weight is again given to the distributions showing the greatest change. Hence there is a large change due to the condensation phase emphasizing heavy discounting of the past values. This allows the rapid adjustment of the parameters to occur two observations after a change in the process occurs.

5. ESTIMATING THE SCALE PARAMETER

The algorithms developed in sections 3 and 4 assumed the scale parameter ϕ was a known constant. This is correct for the discrete exponential family models, since in both the binomial and Poisson cases $\phi=1$, however in other models (such as the normal or gamma) ϕ is not necessarily known. In this section we look at a method of recursively estimating ϕ as well as θ_t.

The algorithm developed in this section extends the work of West (1985) into the multiprocess dynamic generalized linear model. At each step, the sampling model will be approximated by one that has the same first two moments as, and is a scaled version of, the original sampling model at that step. It is also possible to embed the discrete exponential family models, such as the binomial and Poisson models, in a scaled sampling model to allow for an unknown scale parameter. When this is done, estimates of ϕ can be used to assess the fit of the exponential family model.

To begin this analysis, we look at a standard one-dimensional exponential family model having fixed scale parameter $\phi=\phi_0$. In these models, a measure of goodness of fit is provided by the deviance, the logarithm of a likelihood

ratio. In the normal distribution, the deviance simplifies to a sum of squares. Thus, the deviance is a generalization of sum of squares to non-normal exponential family models, and it can be partitioned in a similar manner. In order to analyze the effect of the scale parameter, we need the following deviance functions:

1. Conditional observation deviance

$$D(y_t \mid \theta_t, \phi_0, I_{t-1}=i, I_t=j) = -2\log[f(y_t \mid \theta_t, \phi_0, I_{t-1}=i, I_t=j)] + 2\log[f(y_t \mid \hat{\theta}_t, \phi_0, I_{t-1}=i, I_t=j)] \tag{5.1}$$

where $\hat{\theta}_t$ is the maximum likelihood estimator of θ_t.

2. Conditional prior deviance

$$D(\theta_t \mid \phi_0, I_{t-1}=i, I_t=j, y_{t-1}) = -2\log[g(\theta_t \mid \phi_0, I_{t-1}=i, I_t=j, y_{t-1})] + 2\log[g(\tilde{\theta}_t \mid \phi_0, I_{t-1}=i, I_t=j, y_{t-1})] \tag{5.2}$$

where $\tilde{\theta}_t$ is the mode of the prior distribution.

3. Conditional posterior deviance

$$D(\theta_t \mid \phi_0, I_{t-1}=i, I_t=j, y_t) = -2\log[g(\theta_t \mid \phi_0, I_{t-1}=i, I_t=j, y_t)] + 2\log[g(\ddot{\theta}_t \mid \phi_0, I_{t-1}=i, I_t=j, y_t)] \tag{5.3}$$

where $\ddot{\theta}_t$ is the mode of the posterior distribution.

4. Conditional residual deviance

$$D(y_t \mid \phi_0, I_{t-1}=i, I_t=j, y_{t-1}) = [D(\theta_t \mid \phi_0, I_{t-1}=i, I_t=j, y_{t-1}) + D(y_t \mid \theta_t, \phi_0, I_{t-1}=i, I_t=j)]_{\theta=\ddot{\theta}}. \tag{5.4}$$

The deviance functions decompose in a manner similar to sums of squares. Note the important relationship

$$D(\theta_t \mid \phi_0, I_{t-1}=i, I_t=j, y_{t-1}) + D(y_t \mid \theta_t, \phi_0, I_{t-1}=i, I_t=j) = D(\theta_t \mid \phi_0, I_{t-1}=i, I_t=j, y_t) + D(y_t \mid \phi_0, I_{t-1}=i, I_t=j, y_{t-1}). \tag{5.5}$$

The prior deviance plus the observation deviance equals the posterior deviance plus the residual deviance.

The approximate distributions

Note that the standard exponential family sampling distribution satisfies the identity

$$f(y_t \mid \theta_t, \phi_0, I_{t-1}=i, I_t=j) = f(y_t \mid \hat{\theta}_t, \phi_0, I_{t-1}=i, I_t=j) \times \exp[-D(y_t \mid \theta_t, \phi_0, I_{t-1}=i, I_t=j)/2], \tag{5.6}$$

and we may approximate the sampling distribution for an arbitrary scale parameter ϕ by

$$f^*(y_t \mid \theta_t, \phi, I_{t-1}=i, I_t=j) = [\phi/\phi_0]^{1/2} f(y_t \mid \hat{\theta}_t, \phi_0, I_{t-1}=i, I_t=j) \times \exp[-(\phi/\phi_0) D(y_t \mid \theta_t, \phi_0, I_{t-1}=i, I_t=j)/2]. \tag{5.7}$$

Clearly $f^*(y_t \mid \theta_t, \phi, I_{t-1}=i, I_t=j)$ is proportional to a power of $f(y_t \mid \theta_t, \phi_0, I_{t-1}=i, I_t=j)$ as a function of θ_t, and reduces to it when $\phi=\phi_0$.

The conjugate prior distribution for θ_t satisfies a similar identity

$$g(\theta_t \mid \phi_0, I_{t-1}=i, I_t=j, y_{t-1}) = g(\tilde{\theta}_t \mid \phi_0, I_{t-1}=i, I_t=j, y_{t-1}) \times \exp[-(\phi/\phi_0) D(\theta_t \mid \phi_0, I_{t-1}=i, I_t=j, y_{t-1})/2], \tag{5.8}$$

and we may approximate the conjugate prior distribution having an arbitrary ϕ by

$$g^*(\theta_t \mid \phi, I_{t-1}=i, I_t=j, y_{t-1}) = (\phi/\phi_0)^{1/2} g(\tilde{\theta}_t \mid \phi_0, I_{t-1}=i, I_t=j, y_{t-1}) \times \exp[-(\phi/\phi_0) D(\theta_t \mid \phi_0, I_{t-1}=i, I_t=j, y_{t-1})/2]. \tag{5.9}$$

Clearly $g^*(\theta_t \mid \phi, I_{t-1}=i, I_t=j, y_{t-1})$ is proportional to a power of $g(\theta_t \mid \phi, I_{t-1}=i, I_t=j, y_{t-1})$ as a function of θ_t and equal to it when $\phi=\phi_0$.

The algorithm

The initial conditions for the algorithm at time $t-1$ are that there are k prior distributions for $(\phi \mid I_{t-1}=i, y_{t-1})$,and each density $f(\phi \mid I_{t-1}=i, y_{t-1}) \propto (\phi/\phi_0)^{\nu_0^{(i)}/2-1} \exp[-(\phi/\phi_0)\psi_0^{(i)}/2]$. Hence $\psi_0^{(i)}/\phi_0 \times (\phi \mid I_{t-1}=i, y_{t-1})$ has the chi squared distribution with ($\nu_0^{(i)}$) degrees of freedom. (The k posterior distributions for $\beta_{t-1} \mid I_{t-1}=i, y_{t-1}$ and the posterior and prior probabilities of the perturbation indices $q_{t-1}^{(i)}$ and $\pi_t^{(j)}$ for $i=1,\ldots,k$ and $j=1,\ldots,k$ are also known, hence the estimation of β_t proceeds as outlined in section 3).

The joint density is

$$f(\theta_t,\phi \mid I_{t-1}=i, I_t=j, y_{t-1}) = f(\phi \mid I_{t-1}=i, y_{t-1}) \times g^*(\theta_t \mid \phi, I_{t-1}=i, I_t=j, y_{t-1}) \quad (5.10)$$

$$\propto (\phi/\phi_0)^{(\nu_0^{(i)}+1)/2-1} \exp[-(\phi/\phi_0)(\psi^{(i)}$$

$$+D(\theta_t \mid \phi_0, I_{t-1}=i, I_t=j, y_{t-1}))/2],$$

and the marginal density is

$$f(\theta_t \mid I_{t-1}=i, I_t=j, y_{t-1}) \propto [\psi^{(i)}+D(\theta_t \mid \phi, I_{t-1}=i, I_t=j, y_{t-1})]^{(\nu_0-1)/2}, \quad (5.11)$$

so the conditional density is

$$f(\phi \mid \theta_t, I_{t-1}=i, I_t=j, y_{t-1}) \quad (5.12)$$

$$\propto (\phi/\phi_0)^{(\nu_0^{(i)}+1)/2-1} \exp[-(\phi/\phi_0)(\psi_0^{(i)}+D(\theta_t \mid \phi_0, I_{t-1}=i, I_t=j, y_{t-1}))].$$

In other words, $[\psi_0^{(i)}+D(\theta_t \mid I_{t-1}=i, I_t=j, y_{t-1})] \times [(\phi \mid \theta_t, I_{t-1}=i, I_t=j)/\phi_0], y_{t-1})$ has the chi square distribution with $\nu_0^{(i)}+1$ degrees of freedom.

The results after analyzing observation y_t are found in the same manner. The joint conditional density is

$$f(\phi, \theta_t, y_t \mid I_{t-1}=i, I_t=j, y_{t-1}) \propto (\phi/\phi_0)^{\nu_0^{(i)}/2} \quad (5.13)$$

$$\times \exp[-(\phi/\phi_0)(\psi_0^{(i)}+D(\theta_t \mid \phi_0, I_{t-1}=i, I_t=j, y_{t-1})+D(y_t \mid \phi_0, I_{t-1}=i, I_t=j))/2],$$

and the marginal conditional density is

$$f(y_t, \theta_t \mid I_{t-1}=i, I_t=j, y_{t-1}) \quad (5.14)$$

$$\propto [\psi_0^{(i)}+D(\theta_t \mid \phi_0, I_{t-1}=i, I_t=j, y_{t-1})$$

$$+D(y_t \mid \phi_0, I_{t-1}=i, I_t=j)]^{\nu_0^{(i)}/2+1}.$$

The posterior conditional density is found by using (5.5) to yield

$$f(\phi \mid \theta_t, I_{t-1}=i, I_t=j, y_t) \propto (\phi/\phi_0)^{(\nu_0^{(i)}+1)/2-1} \times \quad (5.15)$$

$$\exp[-(\phi/\phi_0)(\psi_0^{(i)}+D(\theta_t \mid \phi_0, I_{t-1}=i, I_t=j, y_{t-1}) \ +D(y_t \mid \theta_t, \phi_0, I_{t-1}=i, I_t=j))/2]$$

$$\propto (\phi/\phi_0)^{\nu_1^{(i,j)}/2-1} \exp[-(\phi/\phi_0)(\psi_1^{(i,j)}+D(\theta_t \mid \phi_0, I_{t-1}=i, I_t=j, y_t))]$$

where $\nu_1^{(i,j)}=\nu_0^{(i)}+1$ and $\psi_1^{(i,j)}=\psi_0^{(i)}+D(y_t \mid \phi_0, I_{t-1}=i, I_t=j, y_{t-1})$. Thus $(\psi_1^{(i,j)} + D(\theta_t \mid \phi_0, I_{t-1}=i, I_t=j, y_t))/\phi_0 \times (\phi \mid \theta_t, I_{t-1}=i, I_t=j, y_t)$ has the chi squared distribution with $\nu_1^{(i,j)}$ degrees of freedom. The mean value and variance of $(\phi \mid \theta_t, I_t=j, y_t)$ are

$$\hat{\phi}_1^{(j)}=(q_t^{(j)})^{-1}\sum_{i=1}^{k} p_t^{(i,j)}\hat{\phi}_1^{(i,j)} \tag{5.16}$$

and

$$\hat{\omega}_1^{(j)}=(q_t^{(j)})^{-1}\sum_{i=1}^{k} p_t^{(i,j)}[\hat{\omega}_1^{(i,j)}+\hat{\phi}_1^{(i,j)}-\hat{\phi}_1^{(j)})^2] , \tag{5.17}$$

where $\hat{\phi}_1^{(i,j)}=\nu_1^{(i,j)}/(\psi_1^{(i,j)}+D(\theta_t \mid \phi_0, I_{t-1}=i, I_t=j, y_t))$ and $\hat{\omega}_1^{(i,j)}=2\nu_1^{(i,j)}/[\psi_1^{(i,j)}+D(\theta_t \mid \phi_0, I_{t-1}=i, I_t=j, y_t)]^2$ are respectively the mean value and variance of $(\phi \mid \theta_t, I_{t-1}=i, I_t=j, y_t)$.

To get back to the same conditions that, we started with we need to assume that $(\psi_\delta^{(j)}/\phi_0)(\phi \mid \theta_t, I_t=j, y_t)$ has the chi square ($\nu_\delta^{(j)}$) distribution with the mean and variance given above. The density is

$$f(\phi \mid I_t=j, y_t) \propto (\phi/\phi_0)^{\nu_\delta^{(j)}/2-1}\exp[-(\phi/\phi_0)\psi_\delta^{(j)}/2] \tag{5.18}$$

where $\nu_\delta^{(j)}=2(\phi_1^{(j)})^2/\omega_1^{(j)}$ and $\psi_\delta^{(j)}=2(\phi_1^{(j)})/\omega_1^{(j)}$. In practice, the distribution of $(\theta_t \mid I_{t-1}=i, I_t=j, y_t)$ is found using (3.6) with $\phi=\phi_0^{(i)}$. The moments of $(\beta_t \mid I_{t-1}=i, I_t=j, y_t)$ are found by using (3.12) and (3.13), and condensed using (3.15), (3.16), and (3.17) to give the final distributions of $(\beta_t \mid I_t=j, y_t)$. The values of $\nu_1^{(i,j)}$ and $\psi_1^{(i,j)}$ are calculated using (5.15), and $\nu_\delta^{(j)}$ and $\psi_\delta^{(j)}$ are calculated using (5.18). The final estimate of $(\phi \mid y_t)$ is $\sum_{j=1}^{k} q_t^{(j)}\nu_1^{(j)}/\psi_1^{(j)}$

6. EXAMPLE

In this section we consider how to estimate and forecast a population proportion that evolves in the following manner. The proportion is stable most of the time, but it is subject to occasional relatively large shifts that may or may not be transitory. The population is sampled at time t, and we wish to estimate the proportion at that time, and forecast the proportion at time $t+1$ using the data up to time t . The traditional approach is to base the estimate on the data at time t only, and disregard the previous data. This does not require any knowledge of how the process evolves, since the estimate is only based on the

the current data, which is used to get a "snapshot" of the current proportion. One drawback of this approach is that it is particularly bad at picking up a trend because the standard error is multiplied by $\sqrt{2}$. A second shortcoming of this approach occurs when the overall series consists of a linear sum of component subseries, for instance the proportions for the population broken down by region, age group, or sex. Estimates for these subseries are often just as important as the series as a whole, and the sample size sufficient to give sufficient accuracy for the series as a whole is completely inadequate to give sufficient accuracy for the component subseries.

The multiprocess dynamic generalized linear model avoids some of these difficulties. Thus it (i) gives good estimates when the series is in a stable pattern, by using past data as well as present data; (ii) quickly reacts when a transient or a change occurs and quickly distinguishes between them; (iii) provides standard errors of the estimates at every step; and (iv) can be used for component subseries, since the model for the combined series is the linear sum of the component series models.

The data we consider here consists of 50 simulated random samples of size 100, from a population where the population proportion is .35 for the first 10 samples, increases by .01 for each of the next 5 samples, continues at .4 for the next 10 samples, abruptly changes to .3 for the next 10 samples, decreases by .005 for each of the next 10 samples, and remains at .25 for the last 5 samples. Thus the sampling distribution of y_t is binomial $(100,\theta_t)$. The multiprocess dynamic generalized linear model we use for this has parameter dynamic equation $\beta_t = A_t\beta_{t-1}$ where

$$\beta = \begin{pmatrix} b_0 \\ b_1 \\ b_2 \end{pmatrix} \text{ and } A_t = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

and b_0, b_1, and b_2 are the local mean, slope and transient respectively. The linear predictor is $\eta_t = X_t'\beta_t$ where $X_t' = (1,0,1)$. Thus the transient is included in the current linear predictor and hence the current observation, but it does not have any effect on future observations. The sampling parameter is related to the parameter vector through the linear predictor by the logit link function

$$\eta_t = g(\mu_t) = \log[\mu_t/(n-\mu_t)]$$

Table 1. States and their prior probabilities.

perturbation index	perturbation	prior probability
1	no change	.9
2	transient	.08
3	mean change	.015
4	slope change	.005

where $\mu_t = b(\theta_t) = n\theta_t$ is the mean of the sampling distribution of y_t, hence the guide relationship is

$$h(g(\theta_t)) = \log[\theta_t/(1-\theta_t)].$$

The four values of the perturbation index variable and their prior probabilities are given in table 1.

The perturbation covariance matrices are given by

$$\mathbf{R}^{(1)} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \mathbf{R}^{(2)} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \mathbf{R}^{(3)} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \mathbf{R}^{(1)} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & .01 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

The prediction $(\hat{\theta}_{t+1}|y_t)$, observation (y_{t+1}), and true proportion (θ_{t+1}) are given in figure 1. Figure 2 shows the true prediction errors $(\theta_{t+1}-\hat{\theta}_{t+1}|y_t)$ and the naive prediction errors $(y_t/n-\theta_{t+1})$, and their boxplots. It is clear that the distribution of true prediction errors is much more concentrated than that of the observation errors. Figure 3 shows the observations, the posterior estimators,and the true proportions, and figure 4 shows the posterior estimator errors and observation errors, and their boxplots. Again, it is clear that the posterior estimator error distribution is much more concentrated than the observation error distribution. Figure 5 shows the equivalent sample size required to have the same variance as the posterior estimator of the mean. Summary statistics are shown in table 2.

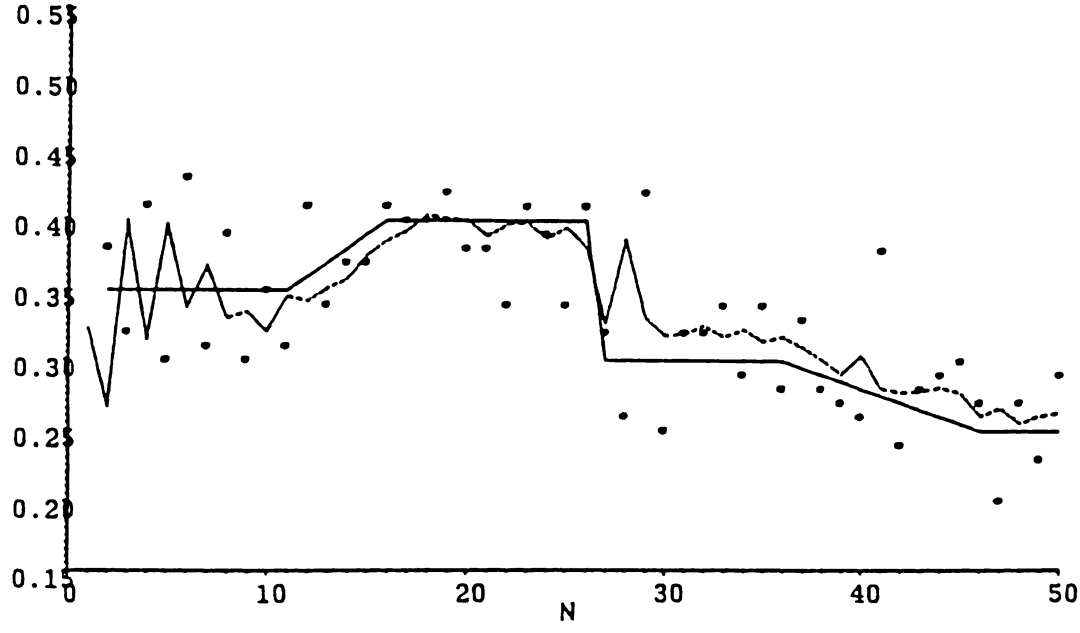


Figure 1. Prediction, observation, and true proportion

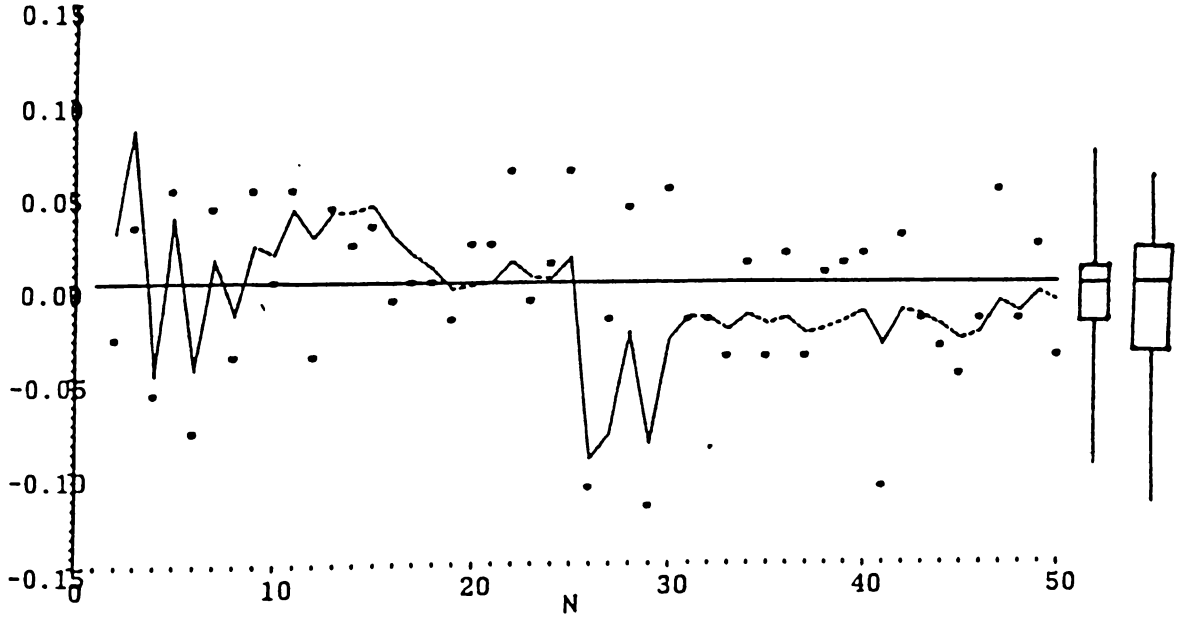


Figure 2. True prediction error and naive prediction error

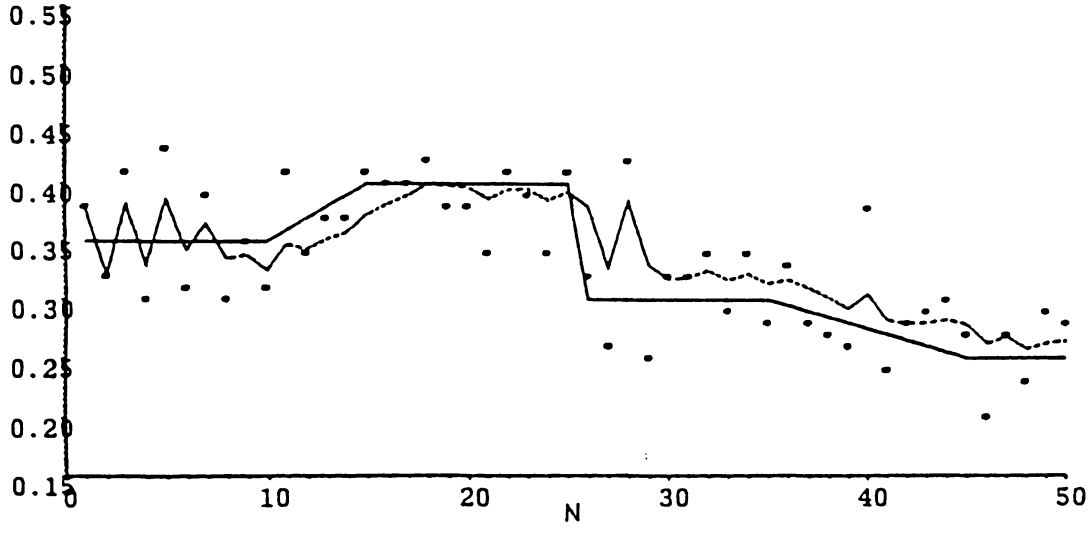


Figure 3. Observation, posterior estimator, and true proportion

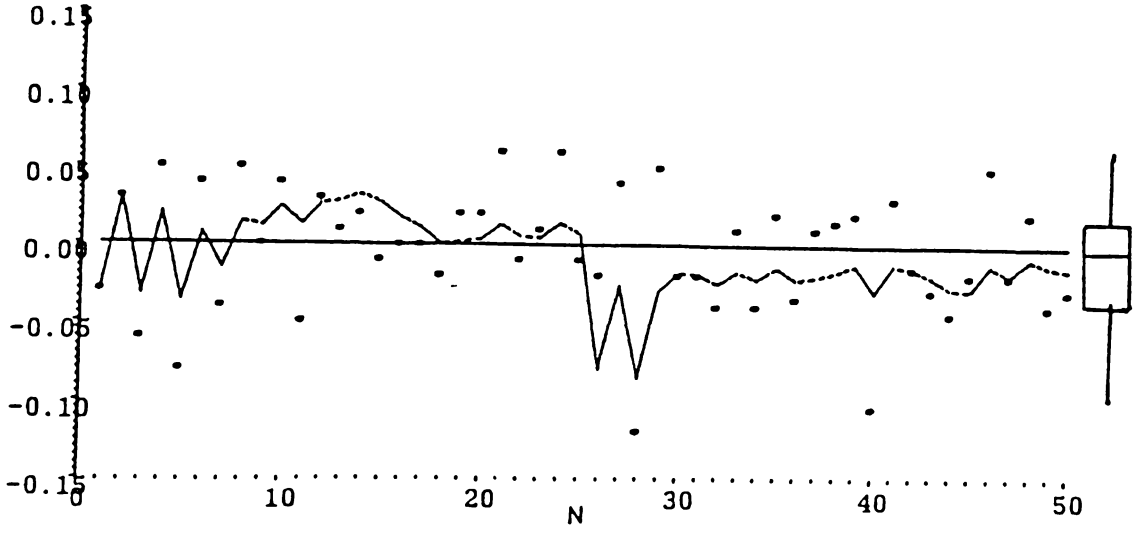


Figure 4. Posterior estimator and observation errors

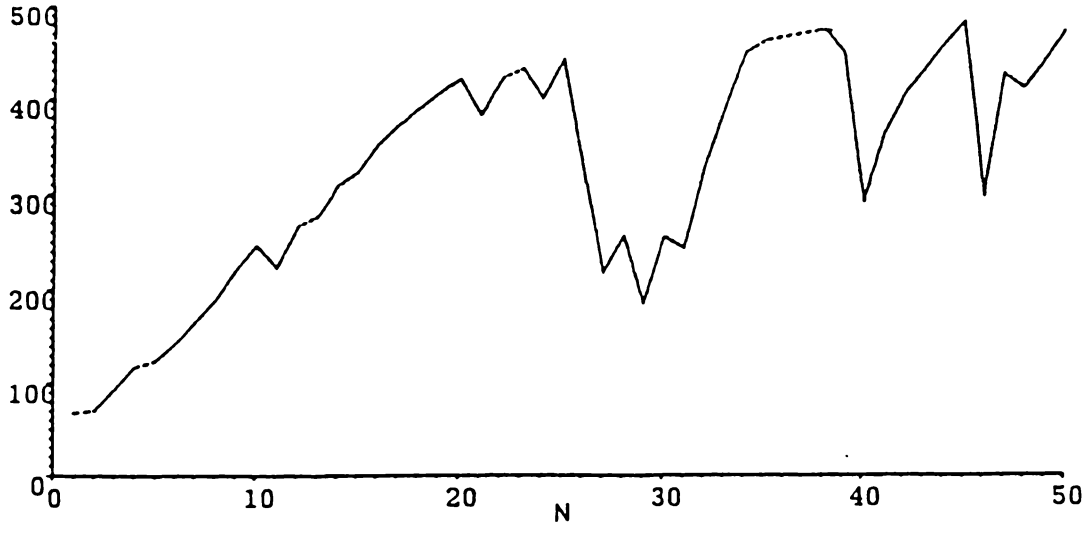


Figure 5. Equivalent sample size

Table 2. Summary statistics

var	mean	med	min	max	q1	q3	σ
y_t/n	.331	.325	.200	.430	.280	.383	.0588
$\hat{\theta}_t \mid y_t$	.335	.328	.257	.401	.299	.384	.0459
$\hat{\theta}_{t+1} \mid y_t$	.332	.326	.254	.409	.280	.385	.0496
$y_t/n - \theta_t$	.004	.000	-.060	.120	-.023	.030	.0400
$\hat{\theta}_t \mid y_t - \theta_t$	-.004	.001	-.080	.074	-.031	.017	.0327
$y_t/n - \theta_{t+1}$	-.006	-.000	-.120	.060	-.038	.028	.0442
$\hat{\theta}_{t+1} \mid y_t - \theta_{t+1}$	-.007	-.012	-.093	.084	-.024	.012	.0363

7. CONCLUSION

The algorithms developed in this paper extend the attributes of multiprocess models, into non-stationary time series where the sampling distribution is not normal, rather a known member of the one dimensional family of distributions; and it is some function of the mean, rather than the mean itself that is a linear function of the dynamic parameters. The use of the conjugate family of prior distributions enables computationally simple recursions similar to those of the Kalman filter to be developed, conditionally on the perturbation index variables. The condensing of the mixture of k^2 posterior distributions into a mixture of k posterior distributions enables the algorithm to quickly react to abrupt

ACKNOWLEGEMENTS

The author is grateful to the referee and the editor for their very constructive criticism, and to Professor R. Hosking for his helpful comments.

BIBLIOGRAPHY

Ameen, J. R. M., and Harrison, P. J. (1985). "Normal Discount Bayesian Models" , in Bayesian Statistics 2, eds. J. M. Bernardo, M. H. Degroot, D. V. Lindley, and A. F. M. Smith, Amsterdam; North Holland .

Bolstad, W. M. (1986 A). "Harrison-Stevens Forecasting and the Multiprocess Dynamic Linear Model" , The American Statistician, Vol. 40, pp. 129-135.

Bolstad, W. M. (1986 B). "An Efficient Algorithm for Harrison-Stevens Forecasting Using the Multiprocess Multivariate Dynamic Linear Model", Communications in Statistics: Simulation and Computation, Vol. B15, pp. 819-828.

Duncan, B., and Horn, S. D. (1972) "Linear Dynamic Recursive Estimation From the Viewpoint of Regression Analysis", JASA 67, pp. 815-821.

Guttman, I., and Pena, D. (1985) "Robust Filtering", JASA 80, pp. 91-92.

Harrison, P. J., and Stevens, C. F. (1971). "A Bayesian Approach to Short-term Forecasting", Operations Research Quarterly 22, pp. 341-362.

Harrison, P. J., and Stevens, C. F. (1976). "Bayesian Forecasting" (With Discussion)", Journal of the Royal Statistical Society, Series B 38, pp. 205-247.

Ho, Y. C., and Lee, R. C. K. (1964), "A Bayesian Approach to Problems in Stochastic Estimation and Control", IEEE Transactions in Automatic Control, pp. 33-39.

Kalman, R. E. (1960). "A New Approach to Linear Filtering and Prediction Problems", Journal of Basic Engineering, pp. 95-108.

Kitagawa, G., (1987), "Non-Gaussian State-Space Modeling of Nonstationary Time Series" (with discussion), JASA 82, pp. 1041-1061.

Makov, U. E. (1983), "Approximate Bayesian Procedures for Dynamic Linear Models in the Presence of Jumps", The Statistician 32, pp. 207-213.

Martin, R. D., and Raftery, A. (1987), "Comments on 'Non-Gaussian State-Space Modeling of Nonstationary Time Series'", JASA 82, pp. 1044-1050.

Masrieliez, C. J., and Martin, R. D., (1977), "Robust Bayesian Estimation for the Linear Model and Robustifying the Kalman Filter", IEEE Transactions on Automatic Control 22, pp. 361-371.

Meinhold, L. R., and Singapurawalla, N. (1983). "Understanding the Kalman Filter", The American Statistician vol. 37, pp. 207-213.

Nelder, J. A., and Wedderburn, R. W. M. (1972). "Generalized Linear Models", Journal of the Royal Statistical Society, Series A 35, pp. 370-384.

Smith, A. F. M., West, M., (1983), "An Application of Multiprocess Kalman Filter to Renal Transplant Data", Biometrics 39, pp. 867-878.

Smith, A. F. M., West, M., Gordon, K., Knapp, M., and Trimble, I. C. (1983), "Monitoring Kidney Transplant Patients", Statistician 32, pp. 46-54.

West, M. (1985). "Generalized Linear Models; Scale Parameters, Outlier Accomodation and Prior Distributions", in Bayesian Statistics 2, eds J. M. Bernardo, M. H. DeGroot, D. V. Lindley, and A. F. M. Smith, Amsterdam: North Holland .

West, M., Harrison, P. J., and Migon, H. S. (1985). "Dynamic Generalized Linear Models and Bayesian Forecasting (With Discussion)", JASA 80, pp. 73-97.

Willsky, A. S., (1976). "A Survey of Design Methods for Failure Detection in Dynamic Systems", Automatica 12, pp. 601-611.

Received by Editorial Board member November, 1987;
Revised July 1988.

Recommended by Lyle Broemeling, Virginia Polytechnic Institute, Falls Church, VA.

Refereed by James C. Spall, Johns Hopkins University, Laurel, MD.

```
C THIS PROGRAM ESTIMATES A DYNAMIC PROPORTION USING THE MULTIPROCESS
C DYNAMIC GENERALIZED LINEAR MODEL
      implicit double precision (a-h,o-z)
      dimension betai(4,3),xbeta(4,4),varebeta(4,4,3,3),evarbeta(4,4,3,3)
      DIMENSION beta(4,4,3),c(4,4,3,3),UMU(4,4,3),q(4),se(3),w3(4,4,3)
      DIMENSION pi(4),po(4),po2(4)
      DIMENSION P(4,4),prob(99),y(99),BEST(3),BPRED(3),VPRED(3,3),W1(4,4,3)
      DIMENSION A(3,3),r(4,3,3),X(3),AT(3,3),w(4,4,3,3),xc(4,4,3),cx(4,4,3)
      dimension eta(4,4),veta(4,4),vetainv(4,4),etheta(4,4),vtheta(4,4)
      dimension eeta(4,4),gamma(4,4),delta(4,4),v(4,4,3,3),vi(4,3,3)
      dimension betapred(4,3),cpred(4,4,3,3),betahat(3),chat(3,3),we(4,4)
      dimension wv(4,4),betaest(3),VEST(3,3)
      INTEGER I,J,K,l,N,ichk,ii,iny,iy,nn
      REAL em,ev,OBS,u,err,etahat,vetahat3,sum,eetahat,thethat,vthethat
      real etaest,vetaest
C beta CONTAINS THE STATE VECTORS -THIRD INDEX
c betai contains the condensed state vector -second index
C c CONTAINS THE STATE VECTOR COVARIANCE MATRICES-THIRD & FOURTH INDICES
C pi CONTAINS THE PRIOR PROBABILITIES
C po CONTAINS THE POSTERIOR PROBABILITIES
C A CONTAINS THE STATE TRANSISTION EQUATION
C   R contains the perturbation covariance matrix second & Third indices
C X CONTAINS THE OBSERVATION EQUATION
C Y IS THE TIME SERIES
C eta contains the linear predictor
C veta contains the variance of linear predictor
c etaest contains the combined linear predictor
c vetaest contains the combined linear predictor variance
c thetaest contains the combined estimated sampling parameter
c vthetaest contains the combined variance of the sampling parameter
c etahat contains the predicted linear predictor
c vetahat contains the variance of the prediction
c thetahat contains the predicted sampling parameter
c vthetahat contains the variance of the predicted sampling parameter
      ichk=1
      open(unit=4,file='bino.dat',status='old')
      do 50 n=1,50
   50 read(4,60)  prob(n),y(n)
   60 format(f7.3,f3.0)       do 63 i=1,4

      do 63 k=1,3

   63 vi(i,k,k)=99
      R(2,3,3)=.25
      R(3,1,1)=.25
      R(4,2,2)=.0025
      pi(1)=.9
      pi(2)=.08
      pi(3)=.015
      pi(4)=.005
      po(1)=.9
      po(2)=.045
      po(3)=.045
      po(4)=.01
      phi=1
      n=100
      call assign(8,'mpbino.out')
      CALL ASSIGN(6,'mpdglm.OUT')
C FOR pi AND po, 1=NO CHANGE,2=TRANSIENT,3=STEP CHANGE,4=SLOPE CHANGE
      x(1)=1
      X(2)=0
      X(3)=1
      DO 89 K=1,3
      DO 89 L=1,3
      A(K,L)=0
   89 CONTINUE
```

```
      A(1,2)=1
      A(2,2)=1
      DO 90 K=1,3
      DO 90 L=1,3
 90   AT(K,L)=A(L,K)
      if (ichk .eq. 1)  write(6,91) ((K,L,A(K,L),L=1,3),K=1,3)
 91   FORMAT(3(3(' A(',I1,',',I1,')=',F10.6)/))
C FOR A,1=MEAN,2=SLOPE,3=TRANSIENT
      do 300 nn=1,50
c the section up to statement 110 computes the updated (conditional) mean and
c variance of the linearly dynamic parameters
      write(6,60) prob(nn),y(n)
      if (ichk .eq. 1) write(6,271) ((i,k,betai(i,k),k=1,3),i=1,4)
      if (ichk .eq. 1) write(6,272) (((j,k,l,vi(j,k,l),l=1,3)
     c  ,k=1,3),j=1,4)
      do 100 i=1,4
      do 100 j=1,4
      do 100 k=1,3
      beta(i,j,k)=0
      do 100 l=1,3
 100  beta(i,j,k)=beta(i,j,k)+a(k,l)*betai(i,l)
      if (ichk .eq. 1) write(6,102) (((i,j,k,beta(i,j,k),k=1,3),j=1,4),i=1,4)
 102  format(4(4(3(' beta(',i1,',',i1,',',i1,')=',f10.6)/)/))
      do 110 i=1,4
      do 110 j=1,4
      do 110 k=1,3
      do 110 l=1,3
      w(i,j,k,l)=0
      do 105 k1=1,3
      do 105 l1=1,3
 105  w(i,j,k,l)=w(i,j,k,l)+a(k,k1)*vi(i,k1,l1)*at(l1,l)
 110  c(i,j,k,l)=w(i,j,k,l)+R(j,k,l)
      if (ichk .eq. 1) write(6,111) ((((i,j,k,l,w(i,j,k,l),l=1,3)
     c  ,k=1,3),j=1,4),i=1,4)
 111  format(4(4(3(3(' w(',i1,',',i1,',',i1,',',i1,')=',f10.6)/)/)/))
      if (ichk .eq. 1) write(6,112) ((((i,j,k,l,c(i,j,k,l),l=1,3)
     c  ,k=1,3),j=1,4),i=1,4)
 112  format(4(4(3(3(' c(',i1,',',i1,',',i1,',',i1,')=',f10.6)/)/)/))
      do 193 i=1,4
      do 193 j=1,4
      eta(i,j)=0
      do 193 k=1,3
 193  eta(i,j)=eta(i,j)+x(k)*beta(i,j,k)
      do 194 i=1,4
      do 194 j=1,4
      veta(i,j)=0
      do 194 k=1,3
      xc(i,j,k)=0
      do 194 l=1,3
      xc(i,j,k)=xc(i,j,k) + x(l)*c(i,j,l,k)
 194  veta(i,j)=veta(i,j)+x(k)*C(i,j,k,l)*x(l)
      if (ichk .eq. 1) write(6,195) ((i,j,eta(i,j),j=1,4),i=1,4)
 195  format(4(4(' eta(',I1,',',I1,')=',F10.6)/))
      if (ichk .eq. 1) write(6,196) ((i,j,veta(i,j),j=1,4),i=1,4)
 196  format(4(4(' veta(',I1,',',I1,')=',F10.6)/))
      if (ichk .eq. 1) write(6,197) (((i,j,k,xc(i,j,k),k=1,3),j=1,4),i=1,4)
 197  format(4(4(3(' xc(',i1,',',i1,',',i1,')=',f10.6)/)/))
      do 198 i=1,4
      do 198 j=1,4
 198  vetainv(i,j)=1/veta(i,j)
c the section up to statement 201 computes the mean and variance of the prior
c distribution of the sampling parameter.
      do 199 i=1,4
      do 199 j=1,4
      eeta(i,j)=exp(eta(i,j))
```

```
      etheta(i,j)=eeta(i,j)/(1-eeta(i,j))
  199 vtheta(i,j)=(etheta(i,j)*(1-etheta(i,j)))**2 * veta(i,j)
      if (ichk .eq. 1) write(6,200) ((i,j,etheta(i,j),j=1,4),i=1,4)
  200 format(4(4(' etheta(',I1,',',I1,')=',F10.6)/))
      if (ichk .eq. 1) write(6,201) ((i,j,vtheta(i,j),j=1,4),i=1,4)
  201 format(4(4(' vtheta(',I1,',',I1,')=',F10.6)/))
      do 203 i=1,4
      do 203 j=1,4
      em=etheta(i,j)
      ev=vtheta(i,j)
      we(i,j)=etheta(i,j)
      wv(i,j)=vtheta(i,j)
      gamma(i,j)=em**2 *(1-em)/ev -em
      delta(i,j)=gamma(i,j)*(1-em)/em
  203 continue
      if (ichk .eq. 1) write(6,205) ((i,j,gamma(i,j),j=1,4),i=1,4)
  205 format(4(4(' gamma(',I1,',',I1,')=',F12.6)/))
      if (ichk .eq. 1) write(6,207) ((i,j,delta(i,j),j=1,4),i=1,4)
  207 format(4(4(' delta(',I1,',',I1,')=',F12.6)/))
c this section up to statement 209 calculates the parameters of the posterior
c distribution of the sampling parameter
      do 209 i=1,4
      do 209 j=1,4
      gamma(i,j)=gamma(i,j) + y(nn)
  209 delta(i,j)=delta(i,j) + n - y(nn)
      if (ichk .eq. 1) write(6,205) ((i,j,gamma(i,j),j=1,4),i=1,4)
      if (ichk .eq. 1) write(6,207) ((i,j,delta(i,j),j=1,4),i=1,4)
c this section up to statement 210 calculates the parameters of the updated
c linear predictor distribution
      do 210 i=1,4
      do 210 j=1,4
      em=gamma(i,j)/(gamma(i,j)+delta(i,j))
      ev=em*(1-em)/(gamma(i,j)+delta(i,j)+1)
      etheta(i,j)=em
      vtheta(i,j)=ev
      eta(i,j)=log(em/(1-em))
  210 veta(i,j)= ev/((em)*(1-em))**2
      if (ichk .eq. 1) write(6,200) ((i,j,etheta(i,j),j=1,4),i=1,4)
      if (ichk .eq. 1) write(6,201) ((i,j,vtheta(i,j),j=1,4),i=1,4)
      if (ichk .eq. 1) write(6,195) ((i,j,eta(i,j),j=1,4),i=1,4)
      if (ichk .eq. 1) write(6,196) ((i,j,veta(i,j),j=1,4),i=1,4)
c the section up until statement 220 calculates the corrected mean of the
c dynamic parameter (conditional)
      do 215 i=1,4
      do 215 j=1,4
      xbeta(i,j)=0
      do 213 k=1,3
  213 xbeta(i,j)=xbeta(i,j) + x(k)*beta(i,j,k)
      do 215 k=1,3
      w1(i,j,k)=0
      do 214 l=1,3
  214 cx(i,j,l)=xc(i,j,l)
  215 w1(i,j,k)=w1(i,j,k)+cx(i,j,k)*(eta(i,j)-xbeta(i,j))*vetainv(i,j)
      if (ichk .eq. 1) write(6,217) (((i,j,k,w1(i,j,k),k=1,3),j=1,4),i=1,4)
  217 format(4(4(3(' w1(',i1,',',i1,',',i1,')=',f10.6)/)/))
      if (ichk .eq. 1) write(6,102) (((i,j,k,beta(i,j,k),k=1,3),j=1,4),i=1,4)
      do 220 i=1,4
      do 220 j=1,4
      do 220 k=1,3
  220 beta(i,j,k)=beta(i,j,k)+w1(i,j,k)
      if (ichk .eq. 1) write(6,102) (((i,j,k,beta(i,j,k),k=1,3),j=1,4),i=1,4)
c the section up to statement 240 calculates the corrected covariance
c matrix of the dynamic parameter (conditional)
      do 230 i=1,4
      do 230 j=1,4
      do 230 k=1,3
```

```
      do 230 l=1,3
  230 evarbeta(i,j,k,l)=c(i,j,k,l)-cx(i,j,k)*cx(i,j,l)*vetainv(i,j)
      do 234 i=1,4
      do 234 j=1,4
      do 234 k=1,3
  234 w3(i,j,k)=cx(i,j,k)*vetainv(i,j)
      do 240 i=1,4
      do 240 j=1,4
      do 240 k=1,3
      do 240 l=1,3
      varebeta(i,j,k,l)=w3(i,j,k)*veta(i,j)*w3(i,j,l)
  240 v(i,j,k,l)=evarbeta(i,j ,k,l)+varebeta(i,j,k,l)
      if (ichk .eq. 1) write(6,243) ((((i,j,k,l,evarbeta(i,j,k,l),l=1,3)
     c ,k=1,3),j=1,4),i=1,4)
      if (ichk .eq. 1) write(6,244) ((((i,j,k,l,varebeta(i,j,k,l),l=1,3)
     c ,k=1,3),j=1,4),i=1,4)
      if (ichk .eq. 1) write(6,242) ((((i,j,k,l,v(i,j,k,l),l=1,3)
     c ,k=1,3),j=1,4),i=1,4)
  242 format(4(4(3(3(' v(',il,',',il,',',il,',',il,')=',f10.6)/)/)/))
  243 format(4(4(3(3(' evarbeta(',il,',',il,',',il,',',il,')=',f10.6)
     c /)/)/))
  244 format(4(4(3(3(' varebeta(',il,',',il,',',il,',',il,')=',f10.6)
     c /)/)/))
c the section up to statement 260 is determining the posterior probabilities
c of the perturbation index variable
      sum=0
      do 258 i=1,4
      do 258 j=1,4
      em=we(i,j)
      ev=wv(i,j)
      err=Y(nn)/n -em
      u=ev
      eta(i,j)=log(em/(1-em)) if (ichk .eq. 1) write(6,254) y(nn),em,ev,err,u
  254 Format (5f10.5)
  258 p(i,j)=po(i)*pi(j)*.3989422803 /(sqrt(u)) * exp(-.5*err**2/u)
      if (ichk .eq. 1) write(6,261) ((i,j,p(i,j),j=1,4),i=1,4)
      do 259 i=1,4
      do 259 j=1,4
  259 sum=sum+p(i,j)
      do 260 i=1,4
      do 260 j=1,4
  260 p(i,j)=p(i,j)/sum
      if (ichk .eq. 1) write(6,261) ((i,j,p(i,j),j=1,4),i=1,4)
  261 format(4(4(' p(',I1,',',I1,')=',F10.6)/))       c the section up to state
c covariance matrices
      do 262 j=1,4
      po(j)=0
      po2(j)=0
      do 262 i=1,4
      po2(j)=po2(j)+p(j,i)
  262 po(j)=po(j)+p(i,j)
      if (ichk .eq. 1) write(6,263) (j,po(j),j=1,4),(i,po2(i),i=1,4)
  263 format(4(' po(',I1,')=',F10.6)/,4(' po2(',i1,')=',f10.6))
      do 266 j=1,4
      do 264 k=1,3
  264 betai(j,k)=0
      do 266 k=1,3
      do 266 i=1,4
  266 if (po(j) .ne. 0.) betai(j,k)=betai(j,k)+p(i,j)/po(j)*beta(i,j,k)
      do 270 j=1,4
      do 268 k=1,3
      do 268 l=1,3
  268 vi(j,k,l)=0
      do 270 k=1,3
      do 270 l=1,3
      do 270 i=1,4
```