Discussion Paper 7

Institute for Empirical Macroeconomics
Federal Reserve Bank of Minneapolis
250 Marquette Avenue
Minneapolis, Minnesota 55480

February 1989

RECURSIVE ESTIMATION AND MODELLING OF NONSTATIONARY AND NONLINEAR TIME-SERIES

Peter C. Young

Director, Centre for Research on
Environmental Systems, Institute of
Environmental and Biological Sciences
University of Lancaster
Lancaster, England

David E. Runkle

Institute of Empirical
Macroeconomics
Federal Reserve Bank
of Minneapolis

ABSTRACT

This paper presents a unified approach to nonlinear and nonstationary time-series analysis for a fairly wide class of linear time variable parameter (TVP) or nonlinear systems. The method theory exploits recursive filtering and fixed interval smoothing algorithms to derive TVP linear model approximations to the nonlinear or nonstationary stochastic system, on the basis of data obtained from the system during planned experiments or passive monitoring exercises. This TVP model includes the State Dependent type of Model (SDM) as a special case, and two particular SDM forms, due to Priestley and Young, are discussed in detail. The paper concludes with three practical examples: the first based on the modelling of data from a simulated nonlinear growth equation; the second concerned with the adaptive forecasting and smoothing of the Box-Jenkins Airline Passenger data; and the third providing a critical appraisal of state dependent modelling applied to the famous Sunspot time-series.

Keywords: Nonlinear and nonstationary time-series; recursive estimation, time variable parameter models; state dependent parameter models; adaptive extrapolation, interpolation and smoothing.

RECURSIVE ESTIMATION AND MODELLING OF NONSTATIONARY AND NONLINEAR TIME-SERIES

Peter C. Young

Director, Centre for Research on Environmental Systems,
Institute of Environmental and Biological Sciences,
University of Lancaster, Lancaster, England.

David E. Runkle

Institute of Empirical Macroeconomics,
Federal Reserve Bank of Minneapolis,
Minneapolis, Minnesota, U.S.A.

Abstract. This paper presents a unified approach to nonlinear and nonstationary time-series analysis for a fairly wide class of linear time variable parameter (TVP) or nonlinear systems. The method theory exploits recursive filtering and fixed interval smoothing algorithms to derive TVP linear model approximations to the nonlinear or nonstationary stochastic system, on the basis of data obtained from the system during planned experiments or passive monitoring exercises. This TVP model includes the State Dependent type of Model (SDM) as a special case, and two particular SDM forms, due to Priestley and Young, are discussed in detail. The paper concludes with three practical examples: the first based on the modelling of data from a simulated nonlinear growth equation ; the second concerned with the adaptive forecasting and smoothing of the Box-Jenkins Airline Passenger data; and the third providing a critical appraisal of state dependent modelling applied to the famous Sunspot time-series.

Keywords. Nonlinear and nonstationary time-series; recursive estimation; time variable parameter models; state dependent parameter models; adaptive extrapolation, interpolation and smoothing.

1. INTRODUCTION

This paper presents a unified approach to nonstationary and nonlinear time-series analysis based on time-variable and state-dependent parameter estimation. The approach exploits recursive filtering and smoothing algorithms which derive directly from that best known of all recursive estimation algorithms, the Kalman filter.

The factor which most differentiates the Kalman filter from the prior recursive estimation algorithms of Gauss and Plackett (see e.g. Young,1984) is its inherent ability to handle **nonstationary** systems; i.e. systems where any, or all, of the parameters in the dynamic model of the stochastic dynamic system being investigated may exhibit temporal variation over the observation interval. But, as Kalman admitted in his seminal paper (1960), this state space model and its variable parameters would need to be known *exactly* by the analyst, prior to the application of the filter, in order to exploit this advantage of the formulation.

In the light of this limitation in the Kalman filter, there has been much interest in the development of a more general procedure for the extrapolation, interpolation and smoothing of nonstationary or nonlinear time series; a procedure which inherently incorporates time-variable parameter estimation and allows the user to proceed directly from the time-series data to recursive filtering and smoothing without the need to define the model parameters prior to the analysis. Clearly, such a procedure needs to utilize estimation techniques which are inherently able to handle models with parameters that may vary over time. This was, indeed, one of the major motivations for the development of recursive techniques for **Time Variable Parameter** (TVP) estimation, in which the object is to "model the parameter variations" (Young,1969a,b,1984) by some form of stochastic state-space model.

Such TVP models have been in almost continual use in the control and system's field since the early 1960's, when Kopp and Orford (1963) and Lee (1964) pioneered their use in the wake of the Kalman (1960) and Kalman and Bucy (1961) papers. Interestingly, these two early but important contributions demonstrate rather different approaches to TVP estimation. Kopp and Orford recognised the nonlinearity of the state-parameter estimation problem caused by the multiplication of the state variables by the unknown parameters, and introduced a method which is now universally known as the **Extended Kalman Filter**. Here the unknown parameters are also considered as state variables and are adjoined to the state vector to form a composite state-parameter vector. This composite state vector is then estimated by a suboptimal linearisation procedure applied at each recursion.

Lee, on the other hand, realised that, by allowing the system model to appear only in the "observation equation" of the state space system, with the parameter variations alone being described by the Gauss-Markov state equations, it was possible to estimate the parameters using a time variable version of the recursive linear least squares estimation algorithm. In other words, Lee reversed the roles of the states and the parameters, with the states appearing only in an "observation space" spanned by the measured variables in the model and with the parameters defining a "parametric" state space of dimension greater than, or equal to, the number of unknown parameters. This introduced some limitations on the approach, however, since the dynamic model for the system had to be of a type which would allow it to be considered from the standpoint of the observation equation alone. We shall have much more to say on this approach to TVP estimation later in the paper.

One of Lee's proposals was that the parameter variation should be characterised by a first order vector Random Walk (RW) model which, because of its unity roots, would allow for wide temporal variability in the parameters over any finite observation interval. The present first author made liberal use of this same device in the 1960's within the context of self adaptive control design (Young, 1969a, 1970, 1971,1981), and proposed an extension to the idea if *a priori* information was available about the nature of the parametric time variability (Young, 1969b). Later, in the early 1970's, he also reminded a statistical audience of the extensive system's literature on recursive estimation and its application to TVP estimation (see Young,1971b,1975; also the comments of W.D.Ray on the paper by Harrison and Stevens (1976).

Since the early 1970's, time varying parameter models have also also been proposed and studied extensively in the statistical and econometrics literatures. For example, a major line of development has been linked to the well known "structural" or "component" time-series model[1] (e.g Harrison and Stevens

1. The term "structural" has been used in other connections in both the statistical and economics literatures and so we will employ the former term.

1971,1976; Kitagawa,1981; Harvey,1984). Here, the approach is an extension of the Lee procedure (although this is not overtly acknowledged by the authors), in which the parameter variations are described by a higher dimensional, vector random-walk type model termed the "linear growth equation" by Harrison and Stevens. In some of these references, the potential importance of recursive smoothing is also highlighted and the methodology can be compared with that proposed in the systems literature by Norton (1975) and pursued in more detail by Jakeman and Young (1979,1984).

The latter reference also shows how the recursive state-space algorithms are closely related and, in some cases yield equivalent results, to other smoothing procedures based on the optimisation technique known as "regularisation", in which the smoothed estimate is obtained by minimising (non-recursively) a least squares criterion function which includes constraints on the rates of change of the estimated variables. Recent research in the economic literature (e.g. Kalaba and Tesfatsion, 1988), which terms this approach "flexible least squares" also uses this kind of optimisation technique. However, we feel that the state-space smoothing procedures used in the present paper provide a more elegant and flexible method of smoothing estimation.

In the wider econometrics literature, there have been numerous contributions involving the concept of TVP estimation and Engle *et al* (1988), for example, present a recent brief review of this topic and discuss an interesting application to electricity sales forecasting, in which the model is a time variable parameter regression plus an adaptive trend described by an RW model. Of considerable importance, particularly in the economics context, is the work of Sims and his co-workers (e,g. Doan et al, 1984) on **Bayesian Vector AutoRegressive Modelling and Forecasting** (BVAR). Here the Vector AutoRegressive (VAR) model is extended so that its potentially time-variable parameters are each assumed to be described by random walk models. The model is then considered within a Bayesian framework, somewhat similar to that used by Harrison and Stevens, but with the Bayesian hyper parameters estimated via maximum likelihood using special methods of numerical optimisation.

Recent research by the present first author and his collaborators (e.g. Young,1988a,b,c; Young and Ng,1988; Ng and Young,1988; Ng et al,1988) has also been concerned with the component type of time-series model and, like the earlier contributions in this context, employs the standard Kalman filter-type recursive filtering and smoothing algorithms. Except in the final forecasting and smoothing stages of the analysis, however, the justification for using these algorithms is not based on either a Bayesian interpretation (Harrison and Stevens,1976) or "optimality" in a prediction error or maximum likelihood (ML) sense (Harvey,1984). Rather, the algorithms are utilised in a manner which allows for straightforward and effective **sequential spectral decomposition** of the time series into **quasi-orthogonal** components. A unifying element in this analysis is the modelling of nonstationary state variables *and* time variable parameters by a class of **second order random walk models** which are able to handle abrupt changes, or even discontinuities, in the states or parameters, so extending its range of applicability.

Finally, two previous papers (Young,1978; Priestley,1980) have attempted to consider the use of TVP estimation in a more general context; namely the identification and estimation of nonlinear stochastic, dynamic systems. The present first author (Young;1978) approached this problem from an engineering standpoint, noting that normal Taylor series linearisation of nonlinear dynamic systems usually produced time variable coefficient linearised models which could be estimated by TVP versions of the various recursive parameter estimation algorithms that have been developed in recent years. In this manner, the nature of the nonlinearity could be inferred and the model could either be useful in its own right, or as a prelude to nonlinear estimation based on the identified nonlinear structure and using techniques such as maximum likelihood.

Priestley (1980) used a more formal approach in which he developed various linearised forms of the nonlinear models, including Volterra series expansions. However, his basic approach, as demonstrated in a later paper (Haggan et al,1984), is very similar to that of Young: he also uses a Taylor series expansion of a particular nonlinear model form similar to that used by Young, and he develops recursive algorithms to estimate the time variable parameters in this linearised representation.

The major difference between the Young and Priestley methods lies in the assumptions made about the time-variability of the parameters. Based on the nature of the first order terms in the linearisation expansion, Priestley notes that the parameters will be "state dependent" and he uses this information to define the form of the stochastic model for the parameter variations. Young, on the other hand recognises this state dependency but widens the utility of his approach by utilising more general stochastic models for parameter variation. In this manner, he allows for dependency on other variables; variables that are not necessarily "states" in the more limited definition of the state-space employed by Priestley.

In the present paper, we will explore further the concepts put forward by Young and Priestley and show how they can be cast within a general recursive estimation and smoothing context.

2. THE TIME-SERIES MODEL

In order to simplify the presentation, we will first consider a scalar time-series $y(k)$ which can be described by a nonlinear stochastic, dynamic equation of the form,

$$y(k) = f\{\, y(k-1), \ldots, y(k-n), u(k), \ldots, u(k-m), \ldots, U(k), \ldots, U(k-q), e(k-1), \ldots, e(k-p)\,\} + e(k) \quad (1)$$

where $f\{.\}$ is a reasonably behaved, nonlinear function dependent upon past values of $y(k)$, as well as present and past values of a deterministic input (or exogenous) variable vector $\mathbf{u}(k)$ with elements $u_i(k)$, $i=1,2,\ldots,r$; the present and past values of a vector $U(k)$ of other exogenous variables $U_j(k)$, $j=1,2,\ldots,s$; and a white noise process $e(k)$. The vector $U(k)$ represents any other variables which may affect the system nonlinearly but whose importance in this regard is not clear prior to time-series analysis.

In this setting, $e(k)$ can be considered as an "innovations" process, with the nonlinear function acting as a "nonlinear predictor" or conditional expectation of the $y(k)$ given all information and data on the system up to the kth sample, i.e.,

$$f\{\chi(k)\} = E\{y(k)|k\}$$

where $\chi(k)$ is, in general, a non-minimal state-space (NMSS) vector (see Priestley,1980; Young et al,1987; Burridge and Wallis,1988) for the system with elements $y(k-i)$, $i=1,2,\ldots n$; $u_i(k-j)$, $i=1,2,\ldots,r$; $j=0,1,\ldots,m$; $U_j(k-h)$, $j=1,\ldots,s$; $h=0,1,\ldots,q$; and $e(k-l)$, $l=1,2,\ldots p$.

Using the normal systems approach to linearisation and, for simplicity, considering only a single exogenous variable $u(k)$, we can now expand the RHS of equation (1) in a Taylor series about $f\{x(k_0)\}$ at some sampling instant k_0, i.e.,

$$\begin{aligned} y(k) = {} & f\{\chi(k_0) + \sum_{i=1}^{n}\left[\frac{\partial f\{\chi(k)\}}{\partial y(k-i)}\right]_{k=k_0}\{y(k-i) - y(k_0-i)\} \\ & + \sum_{j=1}^{m}\left[\frac{\partial f\{\chi(k)\}}{\partial u(k-j)}\right]_{k=k_0}\{u(k-j) - u(k_0-j)\} + \\ & + \sum_{l=1}^{p}\left[\frac{\partial f\{\chi(k)\}}{\partial e(k-l)}\right]_{k=k_0}\{e(k-l) - e(k_0-l)\} + \\ & + e(k) \\ & + \text{first order terms in } U(k-h),\ h=1,2,\ldots,q + \\ & + \text{higher order terms } \ldots\ldots \end{aligned} \quad (2)$$

At this point, we assume that the first order sensitivity with respect to the $U(k)$ variables is small enough for us to ignore them, in addition to the usual higher order terms in the other variables. Note that this does *not* mean that these variables are unimportant: clearly, the partial derivatives of $f\{\chi(k)\}$ with respect to the other variables may well be functions of the $U(k)$ variables. In particular, we might expect these variables to influence the low frequency, wide ranging changes in these derivatives and, therefore, in the resulting time variable parameters of the linearised model.

With some manipulation of equation (2), y(k) can be represented in the form,

$$y(k) + \sum_{i=1}^{n} a_i[\chi(k)]\, y(k-i) = T[\chi(k)] + \sum_{i=1}^{r}\sum_{j=0}^{m} b_{ij}[\chi(k)]\, u_i(k-j) + \sum_{l=1}^{p} c_l[\chi(k)]\, e(k-l) + e(k) \qquad (3)$$

In this equation, $a_i[\chi(k)]$, $b_{ij}[\chi(k)]$, $c_l[\chi(k)]$ and $T[\chi(k)]$ are coefficients in the model which are functions of the NMSS vector and the sampling index k. As a result, they can be considered both as "state dependent" (Priestley,1980) or "time variable" (Young,1978) parameters, depending upon the perspective of the analyst. The Time Variable Parameter (TVP) models introduced by Young in this context are a wider class of models than the State Dependent Models (SDM) of Priestley since they not only include the SDM as a special case, but are also based on a wider definition of the state which acknowledges the potential presence and importance of the additional variables U(k). These additional variables, if present and found to be significant, may lead to the requirement for additional terms in the expansion (2) and will almost certainly affect the time variable nature of the coefficients $a_i(k)$, $b_{ij}(k)$, $c_l(k)$ and T(k). We believe that the TVP concept posed in these terms is more natural in a systems context than the SDM since it allows for a wider variety of situations encountered in applied systems analysis.

As a concrete example of the TVP model, consider an aerospace vehicle designed to fly over an extended flight envelope. The dynamic behaviour of the vehicle, at any particular flight condition, will be characterised by the perturbations of those variables which describe the motion relative to the local reference frame. Over a complete flight mission, however, the coefficients of the linearised equations of motion for these variables (the "stability and control derivatives") will also be functions of those other "flight condition" variables (playing the role of the U(k) variables in our formulation of the general model), such as dynamic pressure and altitude, which define the changing environment and significantly affect the dynamic characteristics of the vehicle. Indeed, it is well known that local linearisation at a given flight condition often results in a linearised model such as (3), or its deterministic equivalent, with parameters that can be assumed sensibly constant at that flight condition for purposes such as control system design. Thus, in the present context, the SDM would be linear with constant parameters, while the TVP model would acknowledge the "nonlinearity-in-the-large" and provide a time varying parameter linear model for the whole flight envelope. Indeed, this is the motivation behind the self adaptive control system of Young (1969a,b;1981) to which we shall refer later.

Another example from an entirely different area is the modelling of socioeconomic time-series. Here, the U(k) variables could represent overall, long-term economic and political factors which give rise to such phenomena as quasi-periodic "economic" or "trade" cycles. As a result, we would expect that the sensitivity of the nonlinear function to the major model variables would be functions of these factors. This would lead to long term variations in all the coefficients of the linearised model: for example, the T(k) parameter would then be associated with long-term "trends" in the measured time-series, the variations of which would reflect these trade cycle effects. This kind of thinking lies behind the Bayesian Vector AutoRegression (BVAR) approach to data-based economic modelling and forecasting used by the Minnesota School of economists (e.g. Sims,1986).

It is now convenient to write equation (3) in the following vector form,

$$y(k) = z(k)^T a(k) + e(k) \qquad (4)$$

where,

$$z(k)^T = [\, 1, y(k-1), \ldots, y(k-n), u^T(k), \ldots, u^T(k-m), e(k-1), \ldots, e(k-l) \,]$$

$$a(k)^T = [\, T(k), a_1(k), \ldots, a_n(k), b_{10}(k), \ldots, b_{1m}(k), \ldots, b_{im}, \ldots, b_{sm}(k), c_1(k), \ldots, c_l(k) \,]$$

and where the TVP nature of the model is denoted by the temporal dependence of the parameters in the a vector. This temporal dependence could, of course, be due to state dependence in the sense of Priestley and the later model identification and estimation procedures will acknowledge this possibility. For simplicity of exposition, however, we will drop the state dependent argument and proceed under the assumption that the parameters will, for various reasons, be dependent upon the time index k. Note that it is tempting, at this time, to compare this form of the model with the well known, constant parameter ARMAX model. In fact, the model (4) has much wider significance than the ARMAX model, as we shall see in later sections of the paper.

In order to complete the model description, it is now necessary to introduce some form of mathematical description for the temporal variation in the parameters of model (4). There are a number of different ways of approaching this problem, but here we will choose to "model the parameter variations" (see Young,1978,1984) by the following Gauss-Markov process,

$$x(k) = F(k-1)\, x(k-1) + G(k-1)\, \eta(k-1) \qquad (5)$$

where x(k) is a "state" vector representing the parameters in a(k) as well as any other elements required in the complete state description of their evolution through time. The dimension of x(k) will be equal to or greater than that of a. The matrices F(k-1) and G(k-1) are, respectively, appropriately dimensioned transition and input matrices whose elements may also vary over time; while $\eta(k)$ is a white noise vector with zero mean and (possibly time-variable) covariance matrix Q(k), i.e.,

$$E\{\eta(k)\,\eta(k)^T\} = Q(k)\delta_{k,j} \quad ; \quad \delta_{k,j} = \begin{cases} 1 \text{ for } k=j \\ 0 \text{ for } k\neq j \end{cases}$$

The nature of the matrices F(k-1), G(k-1), Q(k) and the state vector x(k) (including various possible forms for their temporal dependence) will become clearer in Section 5. of the paper, when we discuss special examples of the general model. For the moment, it will suffice to note that this model, in one form or another, has been employed on many occasions over the past 30 years as a device for modelling parameter variations. For example, with F and G both equal to the identity matrix, the model is simply the well known and used vector Random Walk (RW).

From the inherent complexity of the model (1) with its ill-defined, nonlinear functional dependence, we have evolved the fairly straightforward TVP representation described by equations (4) and (5). The major estimation problem associated with the model arises from the presence of the unobservable stochastic terms e(k-1) to e(k-p) in z(k). However, the model can be simplified further to a linear TVP relationship if it is possible to assume that the stochastic influences in equation (4) reside completely in the additive white noise term e(k), so that z(k) does not depend on the past values of this variable. The stochastic disturbance vector $\eta(k)$ in the parameter variation equation (5) then constitutes the only other other stochastic input to the system and, as we shall see, this can be associated directly with the constraints we choose to impose on the nature of the variable parameters in equation (4).

4. IDENTIFICATION and ESTIMATION of the TVP MODEL

The model described by equations (4) and (5) can be represented in the following state-space setting,

$$x(k) = F(k-1)\, x(k-1) + G(k-1)\, \eta(k-1) \qquad (6)$$

$$y(k) = H(k-1)\, x(k) + e(k) \qquad (7)$$

where H(k-1) is an observation vector chosen so that the observation equation (7) represents the TVP model (4). The specific form of H will, of course, depend upon the application but the specific examples discussed in Section 6 will help to clarify the nature of this vector

Nominally, this model presents a quite formidable estimation problem since it involves the estimation of a combination of unknown time-variable parameters and states appearing in nonlinear relation to each other. From a theoretical standpoint, the most obvious approach is to formulate the problem in Maximum Likelihood (ML) terms. If the stochastic disturbances in the model are normally distributed, the likelihood function for the observations may then be obtained from the Kalman Filter via "prediction error decomposition" (Schweppe,1965). For a suitably identified model, therefore, it is possible to maximise the likelihood with respect to any or all of the unknown parameters in the state-space model, using some form of numerical optimisation.

This kind of maximum likelihood approach has been tried by a number of research workers but their results (e.g. Harvey and Peters (1984), which provides a good review of competing methods of optimisation) suggest that it can be quite complex, even if particularly simple structural models are utilised (e.g. those containing trend and seasonal models, in which the only unknown parameters are the variances of the stochastic disturbances). In addition it is not easy to solve the ML problem in practically useful and completely recursive terms; i.e. with the parameters being estimated recursively, as well as the states. If we consider first the simpler linear TVP representation, where z(k) is assumed to be independent of the past values of e(k), however, then the recursive estimation of the parameters in a(k) is fairly straightforward.

4.1 The Linear TVP Model

In this case, the Recursive Least Squares (RLS) algorithm, suitably modified to allow for time variable parameters described by a Gauss-Markov model such as (5), can be applied directly to the model. For this to be successful, however, the analyst must be able to specify the "system" matrices F(k) and G(k), for all k, together with information on the statistical characteristics of the stochastic disturbances e(k) and $\eta(k)$ (see Young,1984).

This latter requirement is eased somewhat by the scalar form of equation (4), which we now recognise as the "observation" equation in a state-space model. It is easy to show that, for the purposes of estimation, it is not the absolute values of σ^2 and Q(k) that are important, but their relative values. As a result, without any loss of generality, we can consider $\sigma^2=1$ and define a "Noise Variance Ratio" (NVR) matrix $Q_r(k)$, i.e.,

$$Q_r(k)=Q(k)/\sigma^2 \qquad (8)$$

For simplicity, it is normally assumed that $Q_r(k)$ is a diagonal matrix with elements (the NVR values) $q_{ii}(k)$, i=1,2, ... ,n+m+2, that are associated with the time variable nature of the parameters $a_i(k)$, i=1,2,...,n; $b_{ij}(k)$, i=1,2, ..., r; j=0,1,...,m; and T(k).

The RLS algorithm, with the TVP modification and the introduction of the NVR matrix, takes the following prediction-correction form,

Algorithm I

Prediction :

$$\hat{x}(k/k-1) = F(k-1)\hat{x}(k-1)$$

$$P(k/k-1) = F(k-1)\,P(k-1)\,F(k-1)^T + G(k-1)\,[Q_r(k-1)]\,G(k-1)^T \qquad (9)$$

Correction :

$$\hat{x}(k) = \hat{x}(k/k-1) + P(k/k-1)H(k-1)^T \ldots$$
$$[1 + H(k-1)P(k/k-1)H(k-1)^T]^{-1}\{\, y(k) - H(k-1)\hat{x}(k/k-1)\} \qquad (10)$$

$$P(k) = P(k/k-1) - P(k/k-1)H(k-1)^T \ldots$$
$$[1 + H(k-1)P(k/k-1)H(k-1)^T]^{-1}\, H(k-1)\, P(k/k-1)$$

Here, $\hat{x}(k)$ denotes the recursive estimate of x(k) at the kth sampling instant, while $\hat{x}(k/k-1)$ is the recursive estimate of x(k) at k conditional on data up to and including (k-1)th sample. It can be shown that $P(k)^*=P(k)/\sigma^2$ provides an estimate of the covariance matrix for the estimate vector x(k) and so, with this statistical interpretation, $P(k/k-1)/\sigma^2$ is an estimate of the covariance at k conditional on the information processed up to the (k -1)th instant.

The algorithm I is, of course, identical in form to the Kalman filter algorithm. We choose to describe it within the RLS parameter estimation context because the vector H(k-1), which plays the role of the observation vector in conventional Kalman filter terms is, in part, composed here of stochastic variables measured in the presence of noise. Formally, the Kalman filter requires that the elements of this vector should be exactly known, deterministic variables. While this formal requirement is not critical to the success of the present algorithm in estimation terms, it is important that we recognise the differences between the present formulation and the more conventional Kalman filter. In this manner, it should be possible to ensure that these differences do not cause estimation problems or that we do not read more into the statistical properties of the estimates than is justified.

Bearing these caveats in mind, it is possible to proceed one step further in the estimation of x(k); namely the generation of a "smoothed estimate" for the TVP vector. The algorithm I provides an estimate of x(k) at the kth sampling instant which is based on the data up to and including the kth sample, i.e. $\hat{x}(k)=\hat{x}(k/k)$. If we are pursuing off-line analysis and are confronted with a data set with N>k samples, however, it is an advantage, in this TVP situation, to obtain an estimate $\hat{x}(k/N)$ at the kth instant conditional on all of the available data over the observation interval. This smoothed estimate will not then be affected by the phase lag which is inherent on the filtered estimate x(k) and it will have lower estimation error variance. This argument suggests the generation of such a smoothed estimate by the use of a "fixed interval" smoothing algorithm (see e.g. Bryson and Ho (1969); Gelb (1974))

There are a variety of algorithms for off-line, fixed interval smoothing but the one we will consider here utilises the following backwards recursive algorithm, subsequent to application of the above Kalman filtering forwards recursion (see e.g. Norton,1975; Young,1984),

Algorithm II

$$\hat{x}(k/N) = F(k)^{-1}\,[\,\hat{x}(k+1/N) + G(k)Q_r(k)G(k)L(k)] \qquad (11)$$

where,

$$L(N) = 0\,;$$

N is the total number of observations (the "fixed interval"); and

$$L(k) = [\,I - P(k+1)H(k)^TH(k)] \ldots$$
$$[\,F(k+1)^TL(k+1) - H(k)^T]\,\{y(k+1) - H(k)F(k)\hat{x}(k)\} \qquad (11A)$$

, is an associated backwards recursion for the "Lagrange Multiplier" vector L(k) required in the solution of this two point boundary value problem.

Finally, the covariance matrix $P^*(k/N) = \sigma^2P(k/N)$ for the smoothed estimate is obtained by reference to P(k/N) generated by the following matrix recursion,

$$P(k/N) = P(k) + P(k)F(k)^T[P(k+1/k)]^{-1} \ldots$$
$$\{P(k+1/N)-P(k+1/k)\}[P(k+1/k)]^{-1}F(k)P(k) \qquad (12)$$

while the smoothed estimate of original series y(k) is given simply by,

$$\hat{y}(k/N) = H\hat{x}(k/N) \qquad (13)$$

i.e., the appropriate linear combination of the smoothed state variables.

As in the forward filtering pass, these recursions are only formally applicable if z(k) is a purely deterministic vector. Indeed, the problem here is rather more problematic than in the filtering case and the smoothing estimates obtained in this manner are sub-optimal in strict maximum likelihood or Bayesian sense. However, as we shall see, this sub-optimality is not of major practical significance in the present context.

4.2 The Pseudo-linear Time-Series Model

Strictly, the filtering and smoothing algorithms I and II are not directly applicable in the more general case of equation (4) where the z(k) vector is a function of past values of the unobserved e(k) variable. Nevertheless, an approximate recursive solution can be evolved using a device first proposed by Young (1968) and Panuska (1969) where, at each recursion, the e(k-i) elements in z(k) are replaced by their estimates $\hat{e}(k-i)$ obtained recursively from the following equation,

$$\hat{e}(k) = y(k) - H(k-1)\hat{x}(k) = y(k) - z(k)^T\,\hat{a}(k) \qquad (14)$$

The resulting filtering algorithm has been termed either the Approximate Maximum Likelihood (AML) or Extended Least Squares (ELS) estimation procedure: it is an intuitively appealing approximation which allows us to develop a fairly simple recursive solution to the estimation problem posed by equation (4) using linear-like estimation procedures.

It is not obvious, of course, that this "pseudo-linear" RLS algorithm will converge under all conditions. However, it has been used successfully in many practical applications. Moreover, Solo (1980) has considered its convergence from a theoretical standpoint and shown that it possesses reasonable characteristics in this regard.

The smoothing algorithm in this pseudo-linear case is less well known. As far as we are aware, only Norton (1975) has previously used the smoothing algorithm II in this context and, although we can confirm the generally good performance he reports, there is clearly need for further research on this topic.

4.3 The Transfer Function (TF) Time-Series Model

Finally, we should mention another approach to the estimation problem posed by the TVP model of equations (4) and (5) which we will not exploit in the present paper but which has been used very successfully for TVP estimation. The approach emerges if we consider first the transfer function (TF) form of equation (3), i.e.,

$$y(k) = T(k) + \sum_{i=1}^{m} \frac{B_i(z^{-1})}{A(z^{-1})} u_i(k) + \frac{D(z^{-1})}{A(z^{-1})} e(k) \qquad (15)$$

where $A(z^{-1})=A(k, z^{-1})$, $B(z^{-1})=B_i(k, z^{-1})$ and $D(z^{-1})=D(k, z^{-1})$ are *time variable coefficient* polynomials in the backward shift operator z^{-1}, i.e. $z^{-i}y(k) = y(k-i)$, each characterised, respectively, by the time variable parameters $a_i(k)$, $b_{ij}(k)$ and $c_i(k)$.

In this model, the system and noise transfer functions are both characterised by the same denominator polynomial $A(z^{-1})$. However, if we choose to separate out the effect of $u_i(k)$ into a second nonlinear function g{.} when formulating the original nonlinear model, i.e.,

$$y(k) = f\{y(k-1), \ldots, y(k-n), e(k-1), \ldots, e(k-p)\} + g\{y(k-1), \ldots, y(k-n), u(k), \ldots, u(k-m)\} + e(k)$$

then equation (15) would be transformed into the following alternative form,

$$y(k) = T(k) + \sum_{i=1}^{m} \frac{B_i(z^{-1})}{A(z^{-1})} u_i(k) + \frac{D(z^{-1})}{C(z^{-1})} e(k) \qquad (16)$$

In the constant parameter situation, Jakeman and Young (1981,1983) have shown that there are some advantages to considering this second "Box-Jenkins" model form rather than the common denominator "ARMAX" form of equation (15); see also the discussion in Section 5.5.

In this alternative setting, the recursive filtered estimates of the TF model parameters can be obtained by application of the recursive Instrumental Variable (IV) or Instrumental Variable-Maximum Likelihood (IV-AML) algorithms (e.g. Young,1984); while the smoothed estimates can be obtained from an IV modification of the fixed interval smoothing algorithm II (see Kaldor,1978).

4.4 Parametric Nonstationarity and Variance Intervention

In practical time-series analysis and modelling, the exact nature of the parametric variation in TVP models is difficult to predict: while the changes in the behavioural characteristics of dynamic systems are often relatively slow and smooth, more rapid and violent changes do occur from time-to-time and lead to similarly rapid changes, or even discontinuities, in the nature of the related time series. One limited approach to this kind of problem in a conventional time-series context is the method of "intervention analysis" (see e.g. Box and Tiao, 1975), where dummy inputs are introduced to account for the observed rapid, and otherwise unexplained changes in the time-series. However, the GM model for the parameter variations allows for more flexibility in this regard and is able to account for rapid changes in the levels and slopes of either the low frequency components of the time series *or* the parameters of the models representing the other components.

If the parameters in the GM model (5) are assumed constant then, as we shall see, the model can be used to describe a relatively wide range of smooth variation in the associated trend or model parameters. Moreover, if we allow the parameters to change over time, then an even wider range of behaviour can be accommodated. For example, large but otherwise arbitrary, *instantaneous* changes in the diagonal elements of the Q_r matrix (e.g.increases to values $\geq 10^2$) introduced at selected "intervention" points, can signal to the associated estimation algorithm the possibility of significant changes in the associated parameter at these same points. The sample number associated with such intervention points can be identified either objectively, using statistical detection methods (e.g. Tsay,1988); or more subjectively by the analyst (see Young and Ng, 1988).

It is interesting to note that this same device, which we term **Variance Intervention** (Young and Ng,1988) can be applied to any state-space or TVP model: Young (1969a,b; 1971, 1981), for example, has used a similar approach to track the significant and rapid changes in the level of the model parameters of an aerospace vehicle during a rocket boost phase. Note that less severe changes in the model parameters can normally be tracked by selection of a suitable GM model, but it may sometimes be useful to introduce more moderate variations in the NVR values applied over a number of samples, rather than instantaneously (see T.J. Young et al,1988)

5. SPECIAL EXAMPLES of the TVP MODEL

Although equation (4) is not completely general in nonlinear terms, it is capable of representing a wide variety of nonlinear and nonstationary phenomena. In order to demonstrate its efficacy in this regard, let us consider some special examples which have particular practical significance and are related to well known constant parameter, linear models. In all these models, estimation of the time-variable parameters will, of course, be controlled by the nature of the GM process (5) selected by the analyst, which provides information to the estimation algorithm on the expected nature of the parameter variations. Specific examples of the GM model that can be used with any of the models outlined in this Section are discussed in Section 6.

5.1 The Dynamic Trend (DT) Model

The simplest example of the TVP model described in equations (4) and (5) is obtained if we wish to examine only the low frequency or trend behaviour of a time-series, as represented by T(k) in equation (4). If, for the moment, the other terms on the RHS of (4), except e(k), are considered negligible, then the model becomes,

$$y(k) = T(k) + e(k) \qquad (17)$$

The estimation of T(k), suitably modelled by some form of GM process, can be considered as a problem of trend estimation, with the GM model merely providing prior information on the expected nature of the variations in the scalar trend component of the time series.

This simple model is very useful in practice and recursive smoothing algorithms, based on an associated GM model in the form of an Integrated Random Walk (see next Section 6), have been used for many years in the *micro*CAPTAIN program (Young and Benner,1988) for both trend estimation and removal in data pre-processing. In this particular application context, the algorithms are normally used in a suboptimal manner, in the sense that the residuals will not normally be simply the white noise e(k) but also functions of the other variables on the RHS of equation (4). Nevertheless, the algorithm works very well in this sub-optimal setting and its use can be justified in alternative spectral terms (see Young,1988a).

5.2 The Dynamic Linear Regression (DLR) Model

By "dynamic linear regression", we mean here a simple, nonstationary linear regression model with regression coefficients that may vary over time. In terms of equation (4), the DLR model is obtained if, in addition to e(k), the only terms of importance on the RHS of the equation are the trend T(k) and the contemporaneous values of the exogenous variables $u_i(k)$, i=1,2, ..., r; i.e.,

$$y(k) = T(k) + \sum_{i=1}^{r} b_{i0}(k)\, u_i(k) + e(k) \qquad (18)$$

Here, y(k) is the "dependent variable"; $u_i(k)$, i=1,2, ..., r are the "independent variables" or "regressors"; and the time-variable regression coefficients are T(k) and $b_{i0}(k)$, i=1,2, ..., r. The model is, in other words, a "static" version of equation (4) but is "dynamic" in the sense that its coefficients are time variable.

5.3 The Dynamic Harmonic Regression (DHR) Model

The DHR model is similar to the DLR except that the $u_i(k)$ variables are chosen in the special form required by the harmonic regression (or Fourier) relationship, i.e.,

$$y(k) = T(k) + \sum_{i=1}^{r} b_{i01}(k) \sin(2\pi f_i k) + b_{i02}(k) \cos(2\pi f_i k) + e(k) \qquad (19)$$

where the additional third subscripts (1,2) are required since each term in the regression is now characterised by both sine and cosine functions in the frequency f_i, i=1,2, ..., r. This model is particularly useful for the estimation of nonstationary periodic or seasonal components in time-series and it has been be used as the basis for adaptive forecasting and seasonal adjustment of periodic time-series (see Young,1988a,b,c; Ng and Young,1988; T.J.Young et al,1988; Todd and Young,1989).

5.4 The Dynamic AutoRegression (DAR) Model

Here, in addition to the e(k) term, only the trend T(k) and the past values of y(k) are considered of importance on the RHS of equation (4), i.e.,

$$y(k) = T(k) + \sum_{i=1}^{n} a_i(k)\, y(k-i) + e(k) \qquad (20)$$

This model is particularly useful for adaptive forecasting and for estimating the changing spectral characteristics of nonstationary time-series. Its estimation in this TVP form allows for the definition of time-variable "maximum entropy" spectra which evolve over time and can be presented either in a three dimensional representation or a contour plot (see T.J. Young, 1987).

5.5 The Dynamic ARMA (DARMA) and Dynamic ARMAX (DARMAX) Models

These models follow directly from the DAR model and are obtained simply by retaining more terms on the RHS of equation (4). In the DARMA case, the y(k-i), i=1,2, ..., n, and e(k-l), l=1,2, ..., p are the primary variables; while in the DARMAX case, the $u_{ij}(k-j)$, i=1,2, ..., r; j=1,2, ..., m are also present. Unlike the previous four cases, these models are nonlinear in form since the unknown $c_l(k)$ parameters are multiplied by the unobserved e(k-l) white noise variables. In both cases, however, they can be considered as pseudo-linear regressions if we recursively estimate the e(k) terms and substitute them back into the equation, as discussed in Section 4.2. Equivalently, we could replace these terms in the model by their conditional expectations. Once again, these models can be useful for forecasting or evolutionary spectral analysis. The DARMAX model is also popular for adaptive and self-tuning control, where it is sometimes termed the Extended Least Squares model (ELS).

5.6 The General Transfer Function (GTF) Model

This is, of course, the model given by equation (16), which we referred to previously in Section 4.3. Like the DARMAX model it is particularly useful in adaptive control system design. In estimation terms, the TF model has the advantage (see Jakeman and Young, 1981, 1983) that, in the constant parameter case, the ML estimates of the system TF (i.e the coefficients of the polynomials $A(z^{-1})$ and $B(z^{-1})$) are asymptotically independent of the ML estimates in the noise TF (i.e. the coefficients of $C(z^{-1})$ and $D(z^{-1})$). In certain circumstances, these advantages will carry over to the TVP situation and can simplify the estimation problem by allowing for the separation of the system and noise parameter estimation using instrumental variable (IV) methods.

5.7 The Component Model

The component model (see Young,1988a) is really a selective combination of the other models mentioned above in Sections 5.1 to 5.6, the exact nature of which will be dependent upon the nature of the application. For example, the DT model could provide the trend component if such behaviour was identified in the data; the DLR could account for any contemporaneous effects from other exogenous variables; the DHR model could provide a description of any markedly periodic characteristics; and the GTF in its fully stochastic form (i.e. featuring no deterministic input variable TF) could account for the "coloured noise" effects. This composite model is, however, described in detail elsewhere (Young,1988a,b,c; Ng and Young,1988) and it will suffice here to point out that the model is particularly useful when the various components can be separated sufficiently in spectral terms to be considered "quasi-orthogonal". The procedure of "sequential spectral decomposition" (e.g. Young,1988a) is then applicable and can yield a model which exhibits excellent performance in applications such as forecasting and seasonal adjustment.

5.8 Dynamic Vector (Multivariable) Models

Each of the models described in Sections 5.1 to 5.7 have vector equivalents and these multivariable models can also be considered in TVP terms using the approach presented in this paper (see Ng et al, 1988). Perhaps the best known and used multivariable time-series model is the Vector AutoRegression (VAR). The dynamic version of this is the following DVAR model,

$$y(k) = T(k) - \sum_{i=1}^{n} A_i(k)\, y(k-i) + e(k) \qquad (21)$$

where now T(k) and $A_i(k)$, i=1,2, ..., n, are appropriately dimensioned TVP matrices.

Clearly this model and its vector DVARMA, DVARMAX and DVTF relatives present a much higher dimensional challenge in estimation terms. Also the problems of identifying a suitable structure for the parameter matrices should not be underestimated (see e.g. Young and Wang,1986; Mittnik,1988).

A notably successful example of this vector TVP model is the **Bayesian Vector AutoRegression** (BVAR) procedures developed for economic forecasting at the Federal Reserve Bank of Minneapolis (Doan et al,1984; Sims,1986). Here, the GM model (5) for the parameter variations is selected in its simplest vector Random Walk (RW) form (see next Section 6), and constraints are applied on the estimated variation of the parameters by the definition, using ML estimation, of the prior statistics (i.e. $\hat{x}(0)$, $P(0)$) and the statistical properties of the stochastic inputs (i.e. the NVR matrix Q_r).

6. SPECIAL EXAMPLES of the GM for the PARAMETER VARIATIONS

The GM model is clearly very general in form and can spawn a wide variety of special cases. Here, we will concentrate on those special model forms that have either proved particularly useful in practice, or are important in conceptual terms. We will consider first the simplest GM models, where the elements of the F(k) and G(k) matrices have constant elements. These have very wide application potential because they are so straightforward and easy to apply. The more complicated GM models are potentially capable of yielding more precise and detailed inferences on the nonlinear characteristics of the system under investigation. But their complexity, their greater need for *a priori* assumptions and information, and the subsequent constraints these assumptions apply on the analysis, may make them less attractive in practical application. There is, of course, a rich opportunity to develop other models within this very general model form. We will leave the reader, however, to consider other possibilities, hopefully stimulated by the examples given here.

6.1 The Generalised Random Walk (GRW)

In the GRW, the GM model takes the specific form,

$$x_t(k) = F_t\, x_t(k-1) + G_t\, \eta_t(k-1) \qquad (22)$$

where,

$$x_t(k) = [\; t(k)\; d(k)\;]^T \;\text{ and }\; \eta_t(k) = [\; \eta_{t1}(k)\; \eta_{t2}(k)\;]^T$$

and,

$$F_t = \begin{bmatrix} \alpha & \beta \\ 0 & \gamma \end{bmatrix} \;;\; G_t = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Here, the subscript t is used merely to differentiate the matrices in this specific GM process from the general GM matrices F and G and does not relate to time; while $\eta_{t1}(k)$ and $\eta_{t2}(k)$ represent zero mean, serially uncorrelated, discrete white noise inputs, with the vector $\eta_t(k)$ normally characterised by a covariance matrix Q_t, i.e.,

$$E\{\, \eta_\tau(k)\, \eta_t(k)^T \} = Q_t\, \delta_{k,j} \quad ; \quad \delta_{k,j} = \begin{cases} 1 \text{ for } k=j \\ 0 \text{ for } k\neq j \end{cases}$$

where, $\delta_{k,j}$ is the Kronecker delta function. Unless there is evidence to the contrary, Q_t is assumed to be diagonal in form with unknown elements q_{t11} and q_{t22}, respectively.

This GRW model subsumes, as special cases (see e.g. Young,1984): the very well known and used **Random Walk** (RW: α=1; β=γ=0; $\eta_{t2}(k)$=0); the **Smoothed Random Walk** (SRW: β=γ=1; 0< α<1.0; $\eta_{t1}(k)$=0); and, most importantly in the present paper, the **Integrated Random Walk** (IRW: α=β=γ=1; $\eta_{t1}(k)$=0). In the case of the IRW, we see that t(k) and d(k) can be interpreted as level and slope variables associated with the variations of the trend, with the random disturbance entering only through the d(k) equation. If $\eta_{t1}(k)$ is non-zero, however, then both the level and slope equations can have random fluctuations defined by $\eta_{t1}(k)$ and $\eta_{t2}(k)$, respectively. This variant has been termed the "Linear Growth Model" by Harrison and Stevens (1971,1976).

The advantage of these random walk models is that they allow, in a very simple manner, for the introduction of nonstationarity into the time series models. By introducing a simple GM model of this type for each of the unknown parameters, we are assuming that they can be characterised by a variable mean value with stochastically variable level and/or slope. The nature of this variability will depend upon the specific form of the GRW chosen: for instance, the IRW model is particularly useful for describing large smooth changes in the parameters; while the RW model (in which the slope is not separately defined) provides for smaller scale, less smooth variations (Young,1984). In all cases, the variance intervention procedure described in Section 4.4 can be used to allow for any abrupt changes in the level or slope at specified points over the observation interval.

6.2 The Periodic Random Walk (PRW)

The simplest formulation of the PRW model is given by the equation,

$$p(k) = p(k-1) + \eta_p(k-1)$$

where $\eta_p(k)$ is a white noise input sequence with the usual statistical properties. The associated GM process takes the form,

$$x_p(k) = F_p x_p(k-1) + G_p \eta_p(k-1) \qquad (23)$$

where,

$$x_p = [\; p(k)\; p_1(k), \ldots\ldots,\; p_w(k)]^T$$

Here p(k) is the periodic function and $p_i(k)$, i=1,2, ..., w, are additional state variables which are introduced to span the seasonal period. The state transition and input matrices take the following form,

$$F_p = \begin{bmatrix} 0 & 0 & \ldots & 0 & 1 \\ 1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & \ldots & 0 & 0 \\ . & . & \ldots & . & . \\ . & . & \ldots & . & . \\ 0 & 0 & \ldots & 1 & 0 \end{bmatrix} \;;\; G_p = \begin{bmatrix} 1 \\ 0 \\ 0 \\ . \\ . \\ 0 \end{bmatrix}$$

and $\eta_p(k)$ is the white noise input sequence with the usual statistical properties.

This model allows for general periodicity in the parameters and will be most useful in those cases where there is known to be some reason for periodically varying parameter values. It has also been used for time-series forecasting and smoothing, when it is employed to model seasonal components (see Young, 1988a; Ng and Young, 1988).

6.3 The State Dependent Model of Young (SDM1)

The Taylor series linearisation approach used in Section 2 suggests that the variations in the linearised parameters will be time-dependent functions of the state $\chi(k)$. Probably the simplest general assumption which acknowledges this state dependency is that a(k) is linearly related to functions of $\chi(k)$, i.e.,

$$a(k) = M[\chi(k)]\; \alpha(k) \qquad (24)$$

or,

$$a_i(k) = m_i(k)^T \alpha(k) \;\; ; \;\; i=1,2, \ldots, n+m+l+2 \qquad (25)$$

where $M[\chi(k)]$, which we will denote below simply as $M(k)$, is a transformation matrix functionally dependent upon $\chi(k)$; $m_i(k)^T$ is the ith row of $M(k)$; $a_i(k)$ is the ith element of $a(k)$; and $\alpha(k)$ is a transformed parameter vector which, in certain, ideal circumstances, could have time invariant elements.

Given the generality of the model (4), however, it seems unlikely if such an ideal situation will apply in practice and it is necessary to add a statistical degree of freedom to the relationship by assuming that $\alpha(k)$ can be modelled as a GM process. And, in the simplest case which certainly seems the most appropriate in general applications, we might assume that this GM process is a vector RW, e.g.,

$$\alpha(k) = \alpha(k-1) + \eta_\alpha(k-1) \tag{26}$$

with the usual assumptions about the white noise vector $\eta_\alpha(k)$.

If M is a square, nonsingular matrix, then we can substitute from equation (24) into (25) and obtain a GM model for the variations of $a(k)$ which is similar in form to equation (5) with,

$$F(k-1) = M(k)\,M(k-1)^{-1} \;;\; G(k-1) = M(k) \tag{27}$$

and, in this case, $x(k)=a(k)$.

This particular approach to the modelling of parameter variations was first used as a device for tracking the rapid variations in the coefficients of a linearised model of an airborne vehicle for the purposes of adaptive control (Young,1969a,b; 1971, 1981). In this example, $M(k)$ was chosen to be diagonal in form with diagonal elements $m_{ii}(k)$ defined as physically motivated functions of certain "air data" variables, such as dynamic pressure and altitude. These variables can be interpreted as "extended" state variables and are associated with the elements of the $U(k)$ vector in equation (1). In other words, the functional dependence is restricted to these other variables and a tighter state dependence in terms of the primary state variables of the system (i.e. $y(k)$,$u_i(k)$ and $e(k)$ in the present context) was not found to be necessary.

If $M(k)$ is diagonal, then we see that $F(k-1)$ is also diagonal with elements $f_{ii}(k-1)=m_{ii}(k)/m_{ii}(k-1)$; in other words, this model has a particularly simple effect on the recursive estimation algorithm, with the ith parameter estimate $\hat{a}(k)$ being updated via a prediction equation (cf Algorithm I, equation (9)),

$$\hat{a}(k/k-1) = \{m_{ii}(k)/m_{ii}(k-1)\}\,\hat{a}(k-1)$$

In this manner, a large increase (decrease) in $m_{ii}(k)$, in relation to its prior value at the previous sampling instant $m_{ii}(k-1)$, will lead to a similar proportionate increase (decrease) in the inter-sample predicted value of the parameter, which will then be updated on receipt of the next data sample by the correction equation (10).

Of course, $M(k)$ need not be diagonal, in which case the simple interpretation of the model given above does not apply and the resulting state-space representation would be more complicated. Also, we might wish to consider other, more complicated forms of state dependence, such as those discussed in the next Section 6.4..

6.4 The State Dependent Model of Priestley (SDM2)

In several important papers on nonlinear and nonstationary time-series analysis and in a recent book, Priestley and his collaborators (1980,1984,1988a,b) have presented an SDM approach to nonlinear modelling which is similar to the procedures discussed in the present paper. Within the present context, Priestley's approach yields models which can be considered as a subset of the models proposed here. The major differences in the models and the associated algorithms lies in the definition of the GM model form for the parameter variations. Also, Priestley employs only the recursive filtering algorithm for parameter estimation and does not utilise fixed interval smoothing, as proposed here.

Priestley's starts the most practically relevant part of his analysis from a nonlinear model similar to equation (1), but without the $U(k)$ variables included in the nonlinear function. He then employs Taylor series expansion, similar to equation (2), but explicitly uses the form of the resultant first order terms in the expansion to define the parameter variation law directly in terms of the primary state variables[2] $y(k)$, $e(k)$ and $u_i(k)$. In particular, he assumes that *each* unknown parameter evolves in time according to an equation of the form,

$$a_i(k) = a_i(k-1) + \Delta z(k)^T \alpha_i(k) \;;\; i=1,2,\ldots,n+m+l+2 \tag{28}$$

where $\Delta z(k) = z(k)-z(k-1)$ is the incremental change in the state vector $\mathbf{z}$ over the sampling interval; while $\alpha(k) = [\alpha_1(k), \ldots, \alpha_{n+m+l+2}(k)]^T$ is a vector of unknown "gradient" parameters assumed to vary as a vector RW process, i.e.,

$$\alpha_i(k) = \alpha_i(k-1) + \eta_\alpha(k-1) \tag{29}$$

with $\eta_\alpha(k)$ a white noise input vector defined in the usual manner. This model can be put in the normal GM form of equation (5) with,

$$F(k) = \begin{bmatrix} 1 & \Delta z^T \\ 0 & I_{n+m+l+2} \end{bmatrix} \;;\; G(k)^T = \begin{bmatrix} 0 & 1 & 1 & \ldots & 1 \end{bmatrix} \tag{30}$$

$$x(k) = x_i(k) = [\, a_i(k) \;\; \alpha_i(k)^T \,]^T$$

where $I_{n+m+l+2}$ is the (n+m+l+2)th order identity matrix. This GM can be compared directly with the IRW model: for example, in the case of a first order AR(1) model with $T(k)=0$, we see that the identity matrix is reduced to a scalar of unity and so the only difference between the model (30) and the IRW is that the $f_{12}(k)$ element of $F(k)$ is now defined as the change $y(k-1)-y(k-2)$, rather than unity.

Of course, for higher order equations, the GM model for each parameter is considerably more complex. And the complete GM model for the vector $x(k)$, as obtained by combining the individual models (30) into a composite state-space form, is of quite large dimension[3]. As a result, the filtering and smoothing algorithms are relatively expensive in relation to the other GM models discussed previously. Also, the selection of this particular GM places quite heavy constraints on the nature of the parameter variations. This is, of course, an advantage if the linearisation assumptions are appropriate to the nonlinear system under investigation. However, it could yield poor performance in prediction (forecasting) terms if the linearisation assumptions are not appropriate.

Finally, two comments on the SDM2 approach are in order. Firstly, we see from equation (28) that,

$$\Delta a_i(k) = \Delta z(k)^T \alpha_i(k) \;;\; i=1,2,\ldots,n+m+l+2$$

If this is compared with equation (25) of the SDM1 approach, we see that the major difference in the assumptions are that here, in SDM2, the *changes* in the unknown parameters are related linearly to the *changes* in the state variables; while in SDM1 it is the *levels* that are related. Also, equations (24) and (25) permit nonlinear functions of $\chi(k)$ in the $M(\chi(k)$ matrix. Secondly, we might question on practical grounds the insertion of differenced stochastic variables in the $F(k)$ matrix, since it is well known that such differencing can cause high frequency noise amplification which, in turn, could lead to problems in the implementation of the filtering and smoothing algorithms.

This latter point certainly justifies the use of fixed interval smoothing, which should help to suppress some of the noise amplification effects. However, it may be better to look for other solutions such as replacing these differenced state elements in $F(k-1)$ by their conditional expectations. For example, .

2. Initially, Priestley considered only univariate processes, but he introduced deterministic input variables in later work (1988a,b)
3. Note that, for clarity, we have concentrated here on the model at the individual parameter level. Priestley (1980, 1988a) presents the complete model in a block form with F(k) and G(k) defined accordingly.

since only F(k-1) is required at the kth instant and this matrix depends only on y(k-2) and y(k-3), we could consider replacing these variables by their fixed lag smoothed estimates y(k-2/k) and y(k-3/k), respectively. We are examining this and other related possibilities.

7. EXAMPLES

The general approach to nonstationary and nonlinear time-series analysis and modelling presented in previous sections of this paper has significance in many different areas where the adaptive extrapolation, interpolation and smoothing of nonstationary or nonlinear time-series is important. These areas include: digital signal and image[4] processing; forecasting and seasonal adjustment of socioeconomic, business, ecological and environmental data; geophysical, biological and medical data processing; and adaptive, learning, or self-tuning control.

The results of such analysis in some of these areas are given in a number of recent papers (Young, 1988a,b,c; Young et al, 1988; Ng and Young, 1988; ng et al, 1988; and T.J. Young et al, 1988). Because of space restrictions, therefore, we will consider here only three examples; one based on simulated data, and the other two on real data. Other examples will be presented at the Conference. The results in the first example were obtained using the recursive smoothing option in Version 2.0 of *micro*CAPTAIN (Young and Benner, 1988); those in the second from a prototype forecasting/smoothing program developed by Dr C.N. Ng and the first author, which is planned for incorporation in later versions of *micro*CAPTAIN; and those in the last example, from a general smoothing program for TVP/SDM modelling developed by the authors using the GAUSS programming language.

7.1 A Simple Simulation Example: TVP Estimation of the Nonlinear Growth Equation

It is useful to consider a simulation example since our knowledge of the exact nature of the signal generation process then allows us to assess the performance of the analytical procedures. Following our earlier general paper on the TVP estimation (1978), we will consider the well known "logistic" growth equation, i.e., in continuous-time,

$$\frac{dx(t)}{dt} = [\,C - x(t)\,]\,x(t)$$
$$y(t) = x(t) + e(t)$$

where x(t) is the growth variable (the "population"), C is the "carrying capacity"; and e(t) is white observation noise which we assume here will be state dependent (i.e. dependent on x(t)).

The discrete-time version of this model, as obtained by the simplest discretisation procedure for a sampling interval Δt takes the form,

$$x(k) = x(k-1) + \alpha(k)\,x(k-1)$$
$$\alpha(k) = [C - x(k-1)]\,\Delta t \qquad (31)$$
$$y(k) = x(k) + x(k)\,e(k)$$

where e(k) is a zero mean, serially uncorrelated sequence with variance σ^2.

Fig.1 shows the variation of y(k), k=1,2, ..., 100, as obtained by simulating the model with the following settings for the parameters in (31),

$$C = 10;\ \Delta t = 0.015;\ x(0) = 0.01;\ \sigma^2 = 0.001$$

The full line shows the deterministic forecast of the estimated TVP model from a forecasting origin at k=50. This model is in the form of a first order DAR model, in which the unknown $a_1(k)$ coefficient can be associated with $1+\alpha(k)$, and the GM model for the parameter variations is chosen as an IRW process without intervention. Fig.2 compares the actual variation in the parameter a(k) with the smoothed estimate obtained with an NVR=0.001, and we see that very good parameter tracking has been achieved. For these results and those discussed below, the recursive algorithm I was initiated using the "diffuse prior assumptions, i.e., with $P(0)=\gamma I$ and $\hat{x}(0)=0$, where γ, in this case was set to unity.

Fig.3 shows the variation of y(k) and the deterministic model forecast when the growth rate is changed abruptly from the normal logistic form to a constant rate of 0.15 between k=49 to k=59; i.e.,

$$\alpha(k) = [C - x(k-1)]\,\Delta t \qquad \text{for } 50>k>59$$
$$\alpha(k) = 0.15 \qquad \text{for } 49<k<60$$

Figs.4 to 6 illustrate various TVP estimation results in this case. First, in Fig.4, we see the recursive estimate of $a_1(k)$ obtained when we model the parameter variations as an IRW without interventions and with an NVR=0.1. This higher NVR is necessary to accommodate the required sharp changes in the DAR(1) parameter; but we see that raising the NVR in this manner has a deleterious effect on the estimation results during the periods of relatively smooth change.

To address this problem, we can resort to variance intervention as discussed in Section 4.4. But if this were a real problem, how would we choose the intervention point locations? One possible subjective approach is shown in Figs.5 and 6: in Fig.5 interventions are introduced every 10 samples, using an RW model with NVR=0.001. When introducing multiple interventions like this it is advisable to use the RW model for parameter variations since little change is expected between interventions. On the basis of these results, a third run is carried out in Fig.6, where interventions are only applied at the 50th and 60th samples, now with an IRW model and NVR=0.001. The resulting estimation results are quite acceptable and the TVP model with these estimates was used to generate the deterministic forecasting results in Fig.3.

Of course other, less subjective, methods for inferring where the sharp changes occur could be devised but the results here will suffice for illustrative purposes. The main point to note is the importance of techniques such as variance intervention if the benefits of recursive smoothing are going to be exploited in this kind of rapid TVP situation. Note also that the forward pass filtered estimates in this case would not be able to track the parameters so well because of the inherent lag and higher variance of the filtered estimates.

Finally, we could use an SDM approach in this example. However, we will consider SDM modelling in the final example.

7.2 Adaptive Forecasting and Smoothing of the Airline Passenger Data

In this example, we consider the well known monthly airline passenger data of Box and Jenkins (1970). However, unlike Box and Jenkins, who analysed the logarithmically transformed data, we will deal with the original, untransformed series. This will help to illustrate better how the TVP approach is able to handle the marked nonstationarity which characterises this particular data set.

The periodogram of the series suggests that the periodicity can be represented quite well by the first two harmonics, with periods of 12 and 6 months, respectively. This suggests a DHR model such as (19) with r=2, $f_1=1/12$, and $f_2=1/6$. Fig.7 shows the adaptive estimation and forecasting results obtained using this DHR, in which the GM is defined with T(k) represented as an IRW process with NVR=0.00001, and the 4 coefficients are also each modelled as IRW processes with NVR=0.0001. The choice of these NVR values is based on the spectral properties of the smoothing algorithms, as discussed in the references by Young et al. cited earlier.

The TVP estimation algorithms were initiated with the diffuse prior, as in the previous example, with $\gamma=10^6$. Since this implies little confidence in the initial estimates, we see that the algorithm requires about 2 years of data to converge to sensible values. Thereafter, however, the one-step-ahead forecasts (shown dashed) are quite good, as we can see from the lower plot in Fig.7, which shows the forecasting errors and their associated standard error bounds (also shown dashed). After 108

4. research is proceeding on the extension of the techniques to 2D signal and image processing, where the smoothing power, with its variance intervention capability could be particularly useful.

samples, multi-step ahead forecasting is initiated and, once again, the forecasting performance is good, with the forecasting residuals very little different from the one-step-ahead results up to sample 108.

Fig.8 shows the estimate of the seasonal component obtained during the analysis and, below this, we see the recursive estimates of the two amplitude parameters, A(k), associated with each harmonic, i.e.,

$$\hat{A}(k) = \sqrt{\hat{b}_{i01}^2 + \hat{b}_{i02}^2}$$

This illustrates quite well the initial convergence of the estimates; the one-step-ahead tracking performance up to k=108; and the multi-step ahead forecast of the two amplitudes over the final three years.

Figs. 9 and 10 illustrate the fixed interval smoothing results obtained by backwards recursion using algorithm II applied from sample 108 to the beginning of the data set. The upper plot in Fig.9 shows the smoothed estimate of the series and the estimated trend component; while the lower plot shows the smoothing residuals, which can be compared with the larger forecasting residuals in Fig.7 and demonstrate well the advantages of fixed interval smoothing. Finally, the smoothed estimate of the seasonal component is given in the upper plot of Fig.10; while the lower plot shows the smoothed estimate of the A(k) variations.

7.3 State-Dependent and TVP Modelling of the Sunspot Data

Our third example is concerned with the well known and much analysed annual sunspot data (1700 to 1945) shown in Fig.11. We will consider SDM2 analysis but recursive smoothing is used here to improve the estimates. The SDM model used is that suggested by Haggan et al (1984) for this same data. This is an AR(2) model of the form,

$$y(k) = T(k) + a_1(k)\, y(k-1) + a_2(k)\, y(k-2) + e(k)$$

where each parameter is modelled by the GM model (30); e.g. in the case of T(k) the model takes the form,

$$T(k) = T(k-1) + \alpha_1(k)[y(k-1)-y(k-2)] + \alpha_1(k)[y(k-2)-y(k-3)]$$

$$\alpha_1(k) = \alpha_1(k-1) + \eta_{\alpha 1}(k-1)$$

$$\alpha_2(k) = \alpha_2(k-1) + \eta_{\alpha 2}(k-1)$$

In an attempt to compare our results with Haggan et al, we use a similar initiation procedure to them (although the diffuse prior approach used in the two previous examples could be used if γ is chosen judiciously). This involves computing the estimates of the a constant parameter AR(2) model based on the first 20 observations and using the results to prime $\hat{x}(0)$ and P(0), and to define the NVR[5]. However, we found that their effective choice of NVR=0.01 was too large, in the sense that a very good fit to the data (see Fig.11) can still be obtained with NVR=0.001. Indeed, lack of time prevented us from investigating this further and it is possible that this value is also too high (see discussion below).

An excellent fit to the data such as that shown in Fig.11 can be deceptive when using TVP or SDM methods. In this kind of estimation, it is important to choose the NVR in order to obtain a sensible balance between parameter tracking and smoothing (see Young,1984). If the NVR is chosen too large then the statistical degrees of freedom will be too great and the estimates will tend to fluctuate rapidly in order to 'force' the model to fit the data. Consequently, we need to examine the estimated time-variable parameters in order to ensure that this is not happening. If we look at Figs.12 and 13, which show the smoothed estimate of T(k) plotted against k and y(k-1), respectively, then we might conclude that this, indeed, happening in this case.

5. Note that Haggan et al use the basic form of the recursive filtering algorithm rather than the NVR form preferred in this paper. However, their "smoothing factor" α is, in effect, the same as the NVR. Their use of "smoothing" in this context is ambiguous; they only use forward pass filtering in their applications.

The results in Figs. 12 and 13 are typical of those obtained for the other two parameters $a_1(k)$ and $a_2(k)$, and it is clear that the estimates obtained with NVR=0.001 are being allowed to vary rapidly and over a wide range. And, of course, the filtered estimates used by Haggan et al vary to an even greater extent than these. More importantly, the associated gradient parameters $\hat{\alpha}_1(k)$ and $\hat{\alpha}_2(k)$ also vary to a large extent as we see from the plot of $\hat{\alpha}_1(k)$ associated with the $a_1(k)$ parameter in Fig.14.

Unfortunately, it is difficult to compare these results any further with those of Haggan et al, since they do not provide full information on their results. In particular, they do not supply plots of the estimated parameter variations against k or y(k-1) and choose, instead, to show only a smoothed version of the scatter plots of the estimates against y(k-1), as obtained using a "Gaussian Smoothing kernel", which we were unable to reproduce. In addition, during this smoothing (again not to be confused with fixed interval recursive smoothing), they reject certain "outliers" caused by "end effects", but without giving precise details of the points so rejected. All that can be concluded, therefore, is that our estimated parameters exhibit the same general pattern of variation revealed by the Haggan et al analysis.

So what can we conclude from our results and, by inference, those of Haggan et al? First, there seems to be good evidence that too much freedom is being allowed for TVP/SDM estimation and the good fit to the data shown in Fig.11 is deceptive. Second, because it employs the highly erratic differenced data terms in the GM model, the SDM procedure may be suspect in its present form. We feel that this aspect of the procedure serves not only to amplify the effects of noise on the data, but also tends, in certain circumstances, to place too much restriction on the nature of the parameter variations. For example, we have chosen to display the T(k) estimates since it is our belief that T(k) being, in effect, a "trend" parameter, should follow the lower frequency fluctuations in the sunspot data caused by the long term amplitude modulation, which is so obvious in the data (see T.J. Young, 1987). By constraining T(k) to be a function of the higher frequency content, differenced data, its estimate has strong higher frequency components associated with the more famous "11 year" cycle. Might it not be better to model it, as in the TVP approach, by a simpler, non state-dependent, IRW process? We are examining this and other possibilities and hope to report this in future papers.

8. CONCLUSIONS

This paper has introduced an approach to nonlinear and nonstationary time-series analysis for a fairly wide class of linear time variable parameter (TVP) or nonlinear systems. The method theory presented here exploits recursive filtering and fixed interval smoothing algorithms to derive TVP linear model approximations to the nonlinear or nonstationary stochastic system, on the basis of data obtained from the system during planned experiments or passive monitoring exercises. This TVP model includes the State Dependent type of Model (SDM) as a special case, and two particular SDM forms due to Priestley (1980) and Young (1969b; 1978) are discussed in detail. The procedure used here to estimate the Priestley type of SDM is, however, somewhat different to that proposed by Priestley, since fixed interval smoothing is used both to improve the noise rejection qualities of the state dependent model and to remove the inherent lag in single-pass, filtering algorithms.

The methodology presented here has wide application potential. It will prove useful as an off-line method for the detection and identification of nonlinearities in time-series. In this context, it can be considered as a pre-processing procedure, in which the recursive filtering and smoothing algorithms are used for identifying the nonlinear model structure, prior to more efficient parameter estimation based on these identification results. Because of its fully recursive formulation, however, the off-line identification analysis can be a prelude to the development of on-line procedures for the adaptive estimation, forecasting and control of nonlinear and nonstationary dynamic systems. Here, the recursive filtering algorithms used in the earlier off-line studies can form the basis for adaptive system design, so integrating the processes of systems analysis and synthesis in a rather useful and elegant manner.

REFERENCES

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

Box, G.E.P., and Tiao, G.C. (1975). Intervention analysis with application to economic and environmental problems. *Jnl. American Stat. Ass.*, **70**, 70-79.

Bryson, A.E., and Ho, Y.C. (1969). *Applied Optimal Control.* Blaisdell: Mass.

Doan, T., Litterman, R., and Sims, C.A. (1984). Forecasting and conditional projection using realistic prior distributions. *Econometrics Reviews*, **3**, 1-100.

Gelb, A., Kasper, J.F., Nash, R.A., Price, C.F., and Sutherland, A.A. (1974). *Applied Optimal Estimation*. MIT Press: Mass.

Engle, R.F., Brown, S.J., and Stern, G. (1988). A comparison of adaptive structural forecasting models for electricity sales. *Jnl. of Forecasting*, **7**, 149-172.

Haggan, V., Heravi, S.M., and Priestley, M.B. (1984). A study of the application of state-dependent models in nonlinear time-series analysis. *Journal of Time Series Analysis*, **5**, 69-102.

Harrison, P.J., and Stevens, C.F. (1971). A Bayesian approach to short-term forecasting. *Operational Res. Quarterly*, **22**, 341-362.

Harrison, P.J., and Stevens, C.F. (1976). Bayesian forecasting. *Jnl. Royal Stat. Soc., Series B.*, 38, 205-247.

Harvey, A.C. (1984). A unified view of statistical forecasting procedures. *Jnl. of Forecasting*, **3**, 245-275.

Harvey, A.C., and Peters, S. (1984). Estimation procedures for structural time-series models. *London School of Economics, Discussion Paper No. A28.*

Jakeman, A.J., and Young, P.C. (1979,1984). Recursive filtering and the inversion of ill-posed causal problems. *Utilitas Mathematica*, **35**, 351-376, 1984 (appeared originally as *Report No. AS/R28/1979, Centre for Resource and Environmental Studies, Australian National University, 1979*)

Jakeman, A.J., and Young, P.C. (1981). On the decoupling of system and noise model parameter estimation in time-series analysis. *Int. Jnl. of Control*, 34, 423-431.

Jakeman, A.J., and Young, P.C. (1983). Advanced methods of recursive time-series analysis. *Int. Jnl. of Control*, **37**, 1291-1310.

Kalaba, R., and Tesfatsion, L. (1988). The flexible least squares approach to time-varying linear regression. *Jnl. of Econ. Dynamics and Control*. **12**, 43-48.

Kaldor, J. (1978). *The Estimation of Parametric Change in Time-Series Models.* M.A. Thesis, Australian National University.

Kalman, R.E.(1960). A new approach to linear filtering and prediction problems. *ASME Trans.,Jnl. Basic Eng.*, **83-D**, 95-108.

Kalman, R.E., and Bucy, R.S. (1961). New results in linear filtering and prediction theory. *ASME Trans., Jnl. Basic Eng.*, **83-D**, 95-108.

Kitagawa, G. (1981). A non-stationary time-series model and its fitting by a recursive filter. *Jnl. of Time Series*, **2**, 103-116.

Kopp, R.E, and Orford, R.J. (1963). Linear regression applied to system identification for adaptive control systems, *AIAA Journal*, **1**, 2300-2306.

Lee, R.C.K.(1964) *Optimal Identification, Estimation and Control*. M.I.T. Press: Cambridge Mass.

Ng, C.N. (1987). *Recursive Identification, Estimation and Forecasting of Non-Stationary Time-Series.* Ph.D Thesis, Centre for Research on Environmental Systems, Univ. of Lancaster, England.

Ng,C.N., and Young,P.C. (1988). Recursive estimation and forecasting of nonstationary time-series, to appear *Jnl. of Forecasting*, 1989.

Ng, C.N., Young, P.C., and Wang, C.L. (1988). Recursive identification, estimation and forecasting of multivariate time-series, to appear in *Proc. IFAC Symposium on Identification and System Parameter Estimation, Beijing, China, 1988*

Norton, J.P. (1975). Optimal smoothing in the identification of linear time-varying systems. *Proc. I.E.E.*, **122**, 663-668.

Panuska, V. (1969). An adaptive recursive least squares identification algorithm. *Proc. 8th I.E.E.E Symposium on Adaptive Processes*.

Priestley, M.B. (1981). *Spectral Analysis and Time Series.* Academic Press: London.

Priestley, M.B. (1980). State-Dependent Models: A general approach to nonlinear time-series analysis. *Journal of Time Series Analysis*, **1**, 47-71.

Priestley, M.B. (1988a). *Nonlinear and Nonstationary Time-Series Analysis*. Academic Press: London.

Priestley, M.B. (1988b). Current developments in time-series modelling. *Jnl. of Econometrics*, **37**, 67-86.

Schweppe, F. (1965). Evaluation of likelihood function for Gaussian signals. *I.E.E.E. Trans. on Inf. Theory*, **11**, 61-70.

Sims, C.A. (1986) Are forecasting models usable for policy analysis? *Federal Reserve Bank of Minneapolis Quarterly Review*, **10**, 2-16.

Solo, V. (1980) Some aspects of recursive parameter estimation. *Int. Jnl. of Control*, **32**, 395-410.

Todd, R.M, and Young, P.C. (1989) Seasonal bias in preliminary seasonal adjustment, in preparation.

Tsay, R.S. (1988). Outliers, level shifts and variance changes in time series. *Jnl of Forecasting*, **7**, 1-20. Whittle, P. (1952). Tests of fit in time series. *Biometrica*, **39**, 309-318.

Young, P.C. (1968) The use of linear regression and related procedures for the identification of dynamic processes. *Proc. 7th I.E.E.E. Symposium on Adaptive Processes*.

Young, P.C. (1969a) *The Differential Equation Error Method of Process Parameter Estimation*. Ph.D. Thesis, University of Cambridge, England.

Young, P.C. (1969b) The use of *a priori* parameter variation information to enhance the performance of a recursive least squares estimator, *Tech. Note 404-90, Naval Weapons Center, China Lake, California,*, 53 pages.

Young, P.C. (1970) An instrumental variable method for real time identification of a noisy process. *Automatica*, **6**, 271-287.

Young, P.C.(1971) A second generation adaptive pitch autostabilisation system for a missile or aircraft. *Tech. Note 404-109,Naval Weapons Centre, China Lake, California (100 pages)*.

Young, P.C. (1971b). Comments on dynamic equations for economic forecasting. *Jnl. Royal Stat. Soc., Series A*, **134**, 220-223.

Young, P.C. (1975). Comments on techniques for testing the constancy of regression relationships over time. *Jnl. Royal Stat. Soc.*, 37, 149-192.

Young, P.C. (1976). Some observations on instrumental variable methods of time-series analysis. *Int. Jnl. of Control*, **23**, 593-612.

Young, P.C. (1978). A general theory of modelling for badly defined dynamic systems. Appears in G.C.Vansteenkiste (ed.), *Modeling, Identification and Control in Environmental Systems*, North-Holland: Amsterdam, 103-135.

Young, P.C. (1981). A second generation adaptive autostabilisation system for airborne vehicles. *Automatica*, **17**, 459-469.

Young, .C. (1984). *Recursive Estimation and Time-Series Analysis*. Springer-Verlag: Berlin.

Young, P.C. (1988a). Recursive extrapolation, interpolation and smoothing of nonstationary time-series. *Proc. IFAC Symposium on Identification and System Parameter Estimation, Beijing, China, 1988*

Young, P.C. (1988b). Recursive estimation, forecasting and adaptive control, chapter to appear in C.T.Leondes (ed) *Control and Dynamic Systems, Vol. XXXII*, Academic Press: Florida.

Young, P.C., Ng, C.N., and Armitage, P. (1988c). A systems approach to economic forecasting and seasonal adjustment. To appear in a special issue of the *International Journal on Computers and Mathematics with Applications* on"System Theoretic Methods in Economic Modelling".

Young, P.C., Ng, C.N., Lane, K., and Parker, D. (1988). Recursive forecasting, smoothing and seasonal adjustment on nonstationary environmental data. to appear in special issue of the *Jnl. of Forecasting*, 1989.

Young, P.C., and Benner,S. (1988). *microCAPTAIN Handbook : Version 2.0*.

Young, P.C, and Ng,C.N. (1988). Variance intervention, *Jnl. of Forecasting*, to appear, 1989.

Young, P.C., and Wang, C.L. (1986). Identification and estimation of multivariable dynamic systems. Appears in J. O'Reilly (ed.) *Multivariable Control for Industrial Applications*, Peter Perigrinus: London, 244-279.

Young, T.J. (1987). *Recursive Methods in the Analysis of Long Time Series in Meteorology and Climatology*. Ph.D. Thesis, Centre for Research on Environmental Systems, Univ. of Lancaster, England.

Young, T.J., Ng, C.N., and Young, P.C.(1988). Seasonal adjustment by optimally smoothed time-variable parameter estimation. *Proc. IFAC Symposium on Identification and System Parameter Estimation, Beijing, China, 1988*.

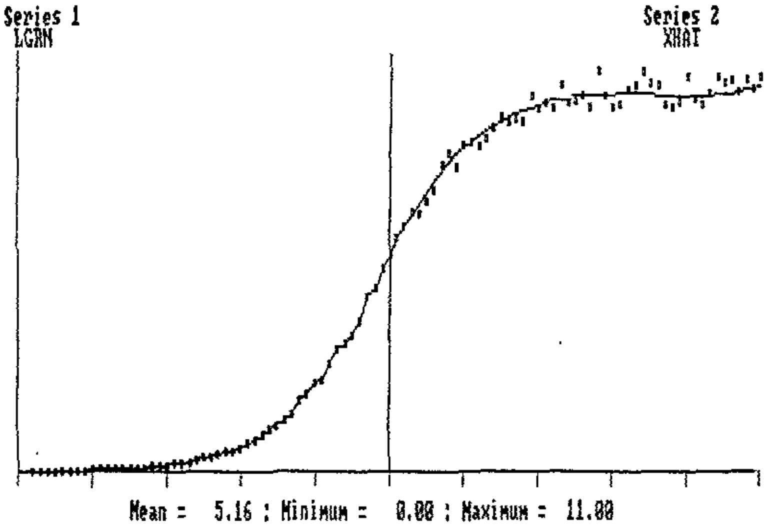

Fig.1 The Logistic Growth Equation: Data and Model Output

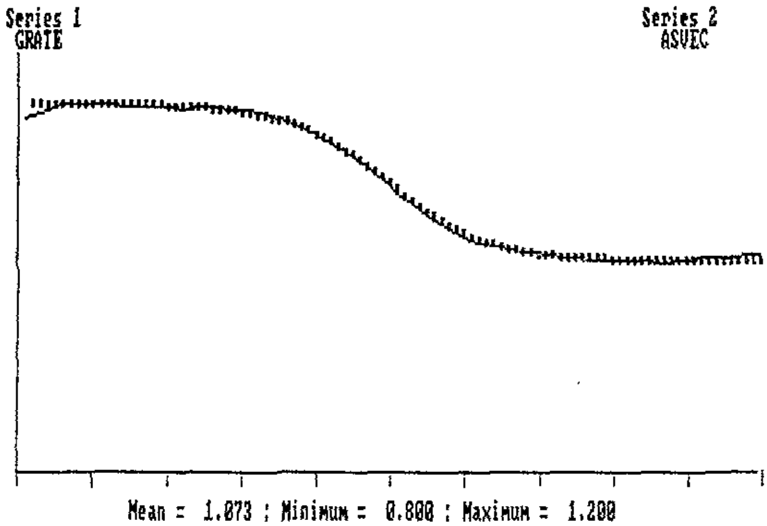

Fig.2 The Logistic Growth Equation: TVP Estimate (full line) Compared with Actual Parameter (points).

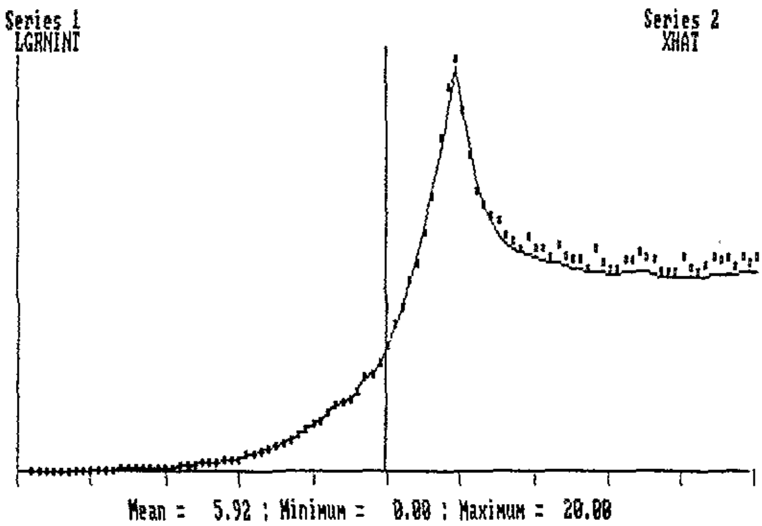

Fig.3 The Modified Logistic Growth Equation: Data and Model Output.

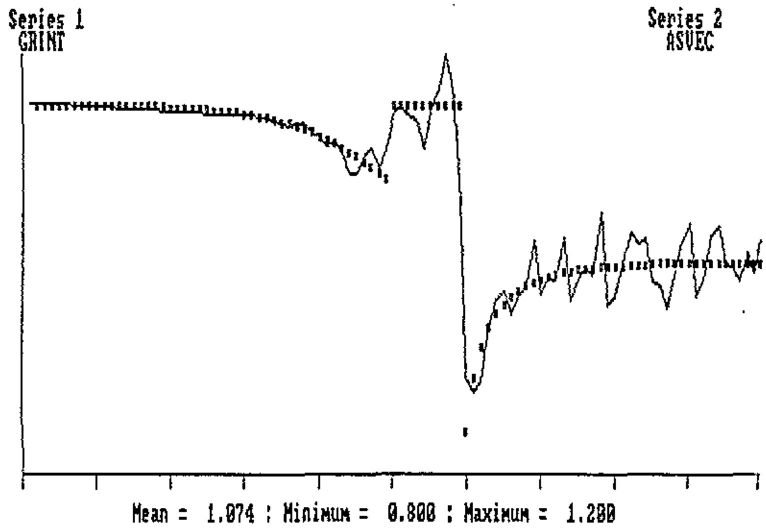

Fig.4 The Modified Logistic Growth Equation: TVP Estimate Using IRW Model with NVR=0.1 (full line) Compared with Actual Parameter (points).

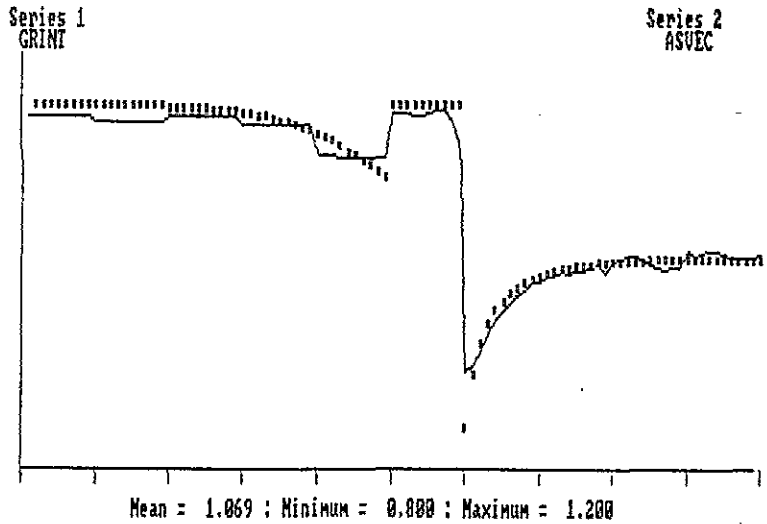

Fig.5 The Modified Logistic Growth Equation: TVP Estimate Using RW Model with 9 Interventions (full line) Compared with Actual Data (points).

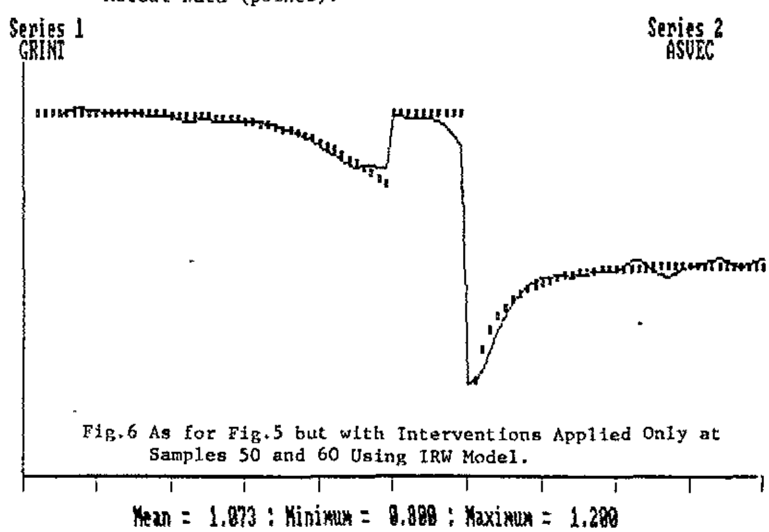

Fig.6 As for Fig.5 but with Interventions Applied Only at Samples 50 and 60 Using IRW Model.

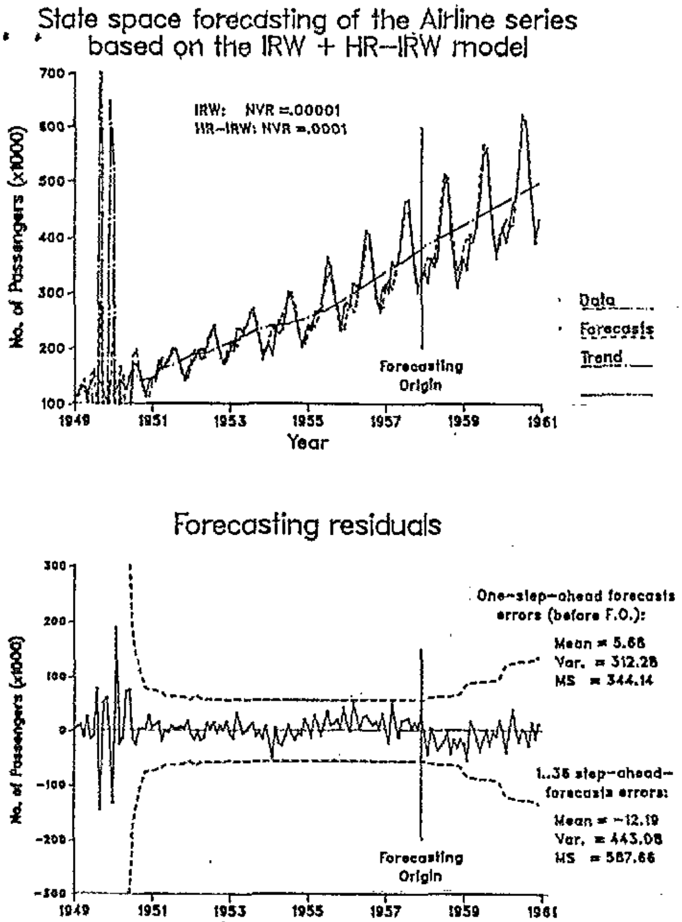

Fig.7 Airline Passenger Data: Adaptive Forecasting.

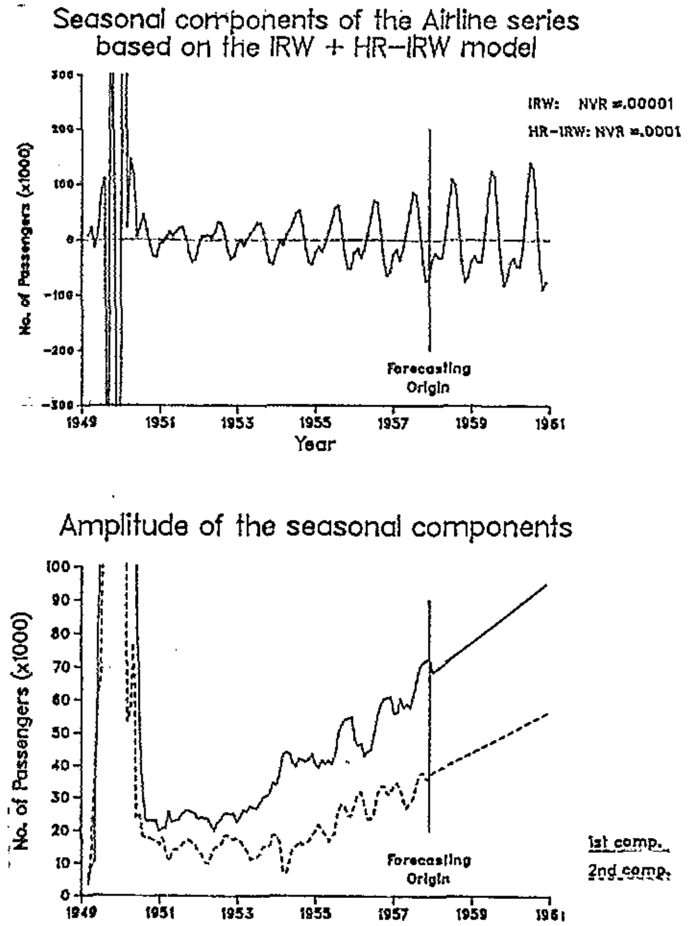

Fig.8 Airline Passenger Data: Forecasting Results Cont

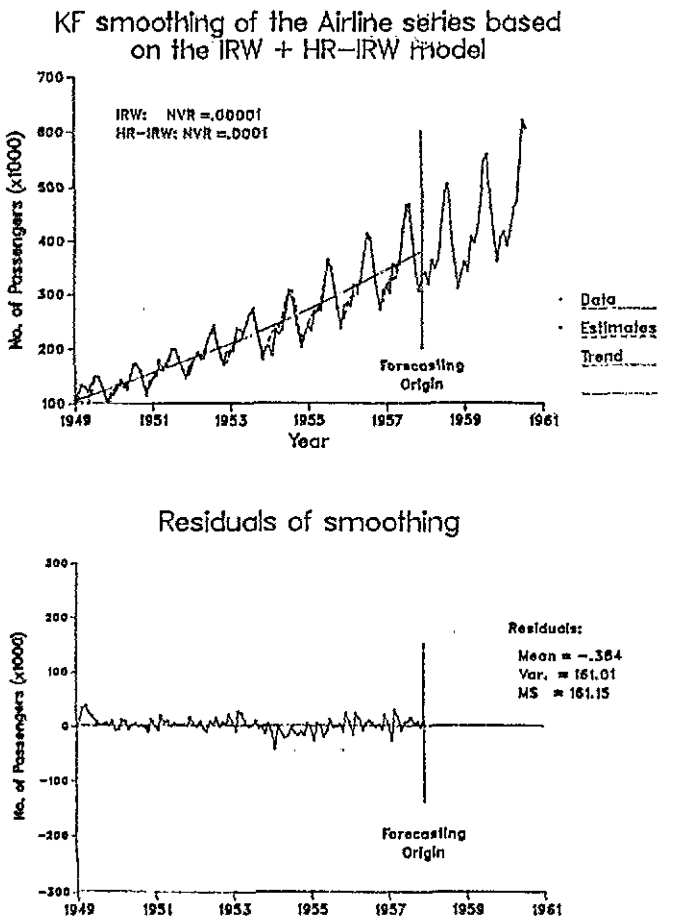

Fig.9 Airline Passenger Data: Smoothing Results
Over first 108 Data Points.

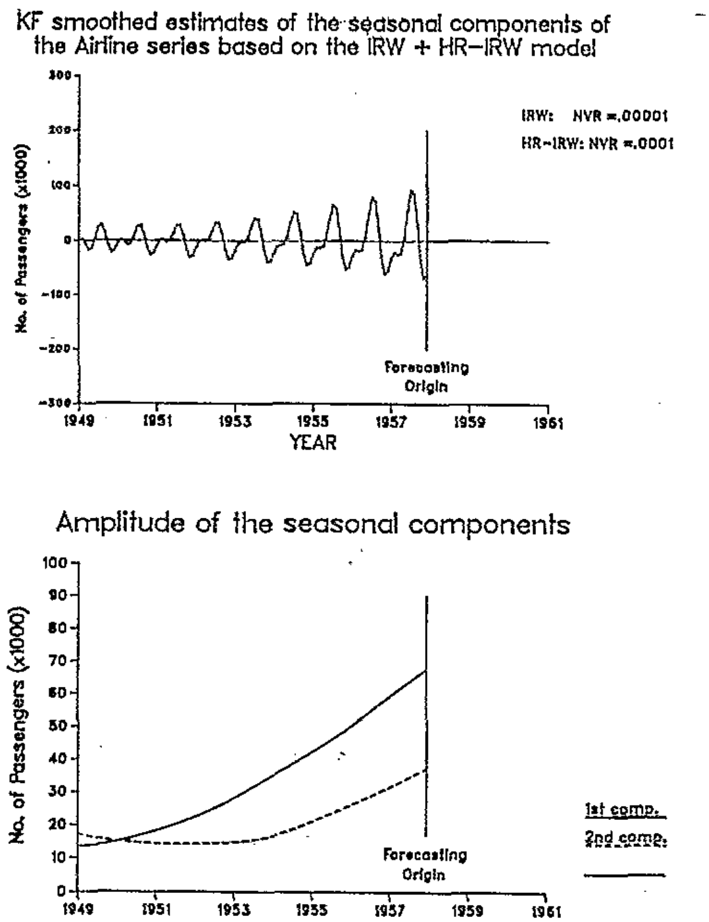

Fig.10 Airline Passenger Data: Smoothing Results Cont

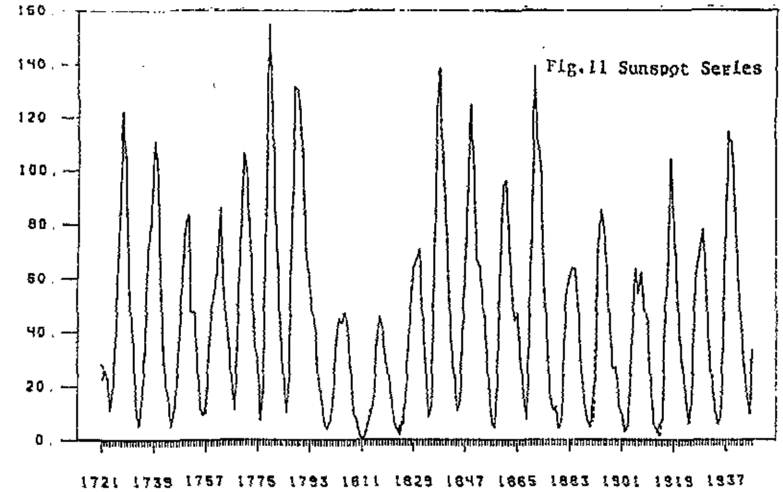

Fig.11 Sunspot Series

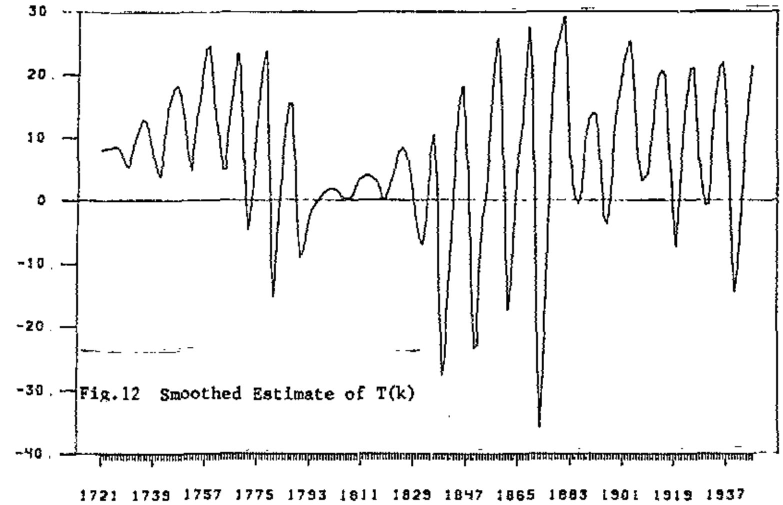

Fig.12 Smoothed Estimate of T(k)

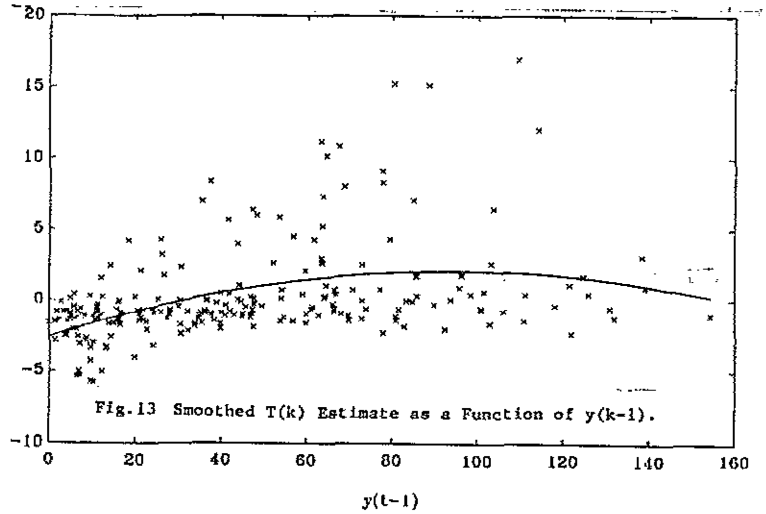

Fig.13 Smoothed T(k) Estimate as a Function of y(k-1).

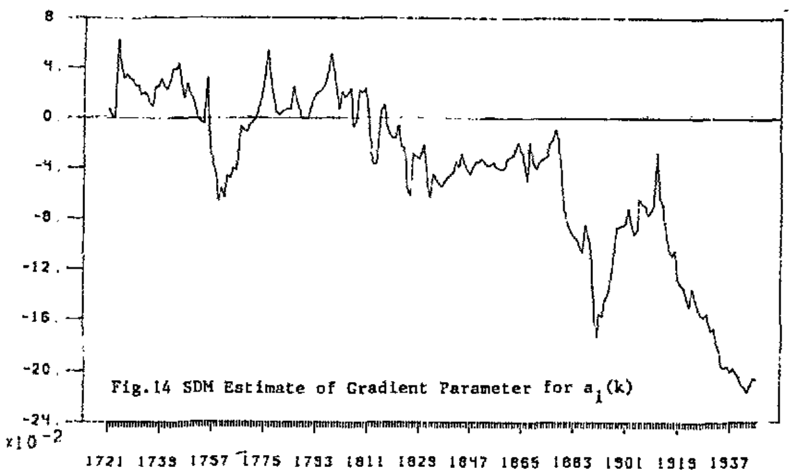

Fig.14 SDM Estimate of Gradient Parameter for $a_1(k)$