

RECURSIVE AND EN-BLOC APPROACHES TO SIGNAL EXTRACTION

(Published in *Journal of Applied Statistics*, 1999, **26**, 1, 103-128)

Peter Young* and Diego Pedregal

Centre of Research on Environmental Systems and Statistics (CRES)

Lancaster University

ABSTRACT

In the literature on *Unobservable Component Models*, three main statistical instruments have been used for signal extraction: *Fixed Interval Smoothing* (FIS) which derives from Kalman's seminal work on optimal state-space filter theory in the time domain; *Wiener-Kolmogorov-Whittle Optimal Signal Extraction* (OSE) theory, which is normally set in the frequency domain and dominates the field of classical statistics; and *Regularisation*, which was developed mainly by numerical analysts but is referred to as *Smoothing* in the statistical literature (e.g. smoothing splines, kernel smoothers and local regression). Although some recognition of the inter-relationship between these methods can be discerned from the literature, no clear discussion of their equivalence has appeared. This paper exposes clearly the inter-relationships between the three methods; highlights important properties of the smoothing filters used in signal extraction; and stresses the advantages of the FIS algorithms as a practical solution to the signal extraction problem. It also emphasises the importance of the classical OSE theory as an analytical tool for a better understanding of the problem of signal extraction.

Key words: Fixed Interval Smoothing; Wiener-Kolmogorov-Whittle filter/smoothener; Deterministic Regularisation; Unobserved Component Models; State Space Models.

* Peter C Young
Environmental Science Division, Lancaster University, Lancaster LA1 4YQ, UK
Tel: +44 1524 59 3966 FAX: +44 1524 593985
E-Mail: p.young@lancaster.ac.uk

1. INTRODUCTION

For many years, the problem of filtering, or the extraction of signals from noisy data, has occupied the minds of mathematicians and statisticians from a variety of disciplines. Indeed, three of the most notable mathematicians of the Twentieth Century - Wiener, Kolmogorov, and Kalman - have all made central contributions to the development of filter or 'signal extraction' theory. A useful concept in signal extraction is the Unobserved Component (UC) model, where the observed variable is related to a number of components which represent different perceived features of the data, usually differentiated by their characteristic spectral properties. Thus a typical discrete time series y_t is considered as the sum of a low frequency *trend component*, T_t ; a *seasonal component*, S_t , often defined by a fundamental frequency and several harmonics; and an irregular component e_t , usually represented by a zero mean sequence of serially uncorrelated random variables (discrete white noise). Of course, these components are not exclusive nor do they actually exist: they simply reside in the eye of the beholder and represent one possible decomposition of the series into components that appear reasonable based on a variety of factors, such as the physical nature of the series, visual observation of its major characteristics, and its temporal and spectral properties obtained by time-series analysis.

In this paper, we consider first one important aspect of UC modelling: namely the identification and estimation of low frequency trend components. Before considering trend estimation in specific terms, however, it is worthwhile reviewing some of the most important contributions to signal extraction, most of which involve trend or low frequency component estimation in one way or another. One of the oldest and best known time domain techniques for signal extraction is Census X-11 and its later extensions X-11 ARIMA and X-12 ARIMA (Shiskin *et al.*, 1967; Dagum, 1980,1988; Findley *et al.*, 1992, 1996) which, for over three decades, has been the main method of trend estimation and seasonal adjustment used by Government Agencies all over the World. In X-11, the trend estimation methods of Henderson (1916), which are based on the exploitation of *Centralised Moving Average* (CMA) filters, have played a significant roll. Such CMA filters are clearly aimed at extracting smooth components defined within different frequency bands in the spectrum of the data (e.g., low frequency trends and seasonal variations). Alternative filtering approaches are defined more specifically by frequency domain considerations. For example, spectral methods that involve the synthesis of low-pass or band-pass filters for the trend or other components have been widely studied and used by engineers for signal processing; whilst, in the economic literature, Sims (1974) and Melis (1989) present two examples of such an approach.

More recently, time domain methodology has been dominated by techniques which exploit Kalman's seminal work on optimal filter and prediction theory using stochastic state space methodology (e.g., Harrison and Stevens, 1976; West and Harrison, 1989; Jakeman and Young, 1979, 1984; Nerlove *et al.*, 1979; Gersch and Kitagawa, 1983, 1984; Harvey, 1984, 1989; Young, 1988; Young *et al.*, 1989, 1991; Ng and Young, 1990; Harvey and Peters, 1990; Young, 1994). In this approach, models for the various components, including the low frequency trend, are formulated in discrete-time, stochastic state space terms and the recursive Kalman Filter (KF) and Fixed Interval Smoothing (FIS) algorithms (see below) are then used for estimation, signal extraction and seasonal adjustment.

An alternative approach to signal extraction, which was stimulated by the popularity of Box-Jenkins methods of time series analysis (Box and Jenkins, 1970), assumes that the time series can be modelled as an ARIMA model. The various unobserved components are then obtained from the estimated ARIMA model through a process of statistical identification, with the classical Wiener-Kolmogorov formulae (e.g. Whittle, 1963, 1983) used to extract the components from the original time series. One procedure which ensures the existence and uniqueness of the decomposition is the so called 'canonical decomposition' (Box *et al.*, 1978; Burman, 1980; Hillmer and Tiao, 1982; Hillmer *et al.*, 1983; Maravall and Gómez, 1992, 1994; Gómez and Maravall, 1996; Maravall, 1993).

Finally, a variety of deterministic optimisation methods have been proposed for signal extraction. These are considered from a variety of different standpoints such as: 'regularisation' (e.g., Jakeman and Young, 1979, 1984; Young, 1991); 'smoothing splines' (e.g., Wahba, 1990); 'pseudosplines' (Hastie, 1996); 'smoothing kernels' (e.g., Wand and Jones, 1995); and 'wavelet' methods (e.g., Daubechies, 1988). In general, the optimisation problem in these cases can be posed as one of minimising the variance of residuals, subject to a given degree of smoothness on the signal components. The techniques that have been suggested range from overly simplistic approaches (e.g., Hodrick and Prescott, 1980), which only involve trend estimation with a single, fixed smoothing parameter; to more complex formulations involving multiple components and numerous, optimally selected smoothing parameters (e.g., Akaike, 1980).

This present paper focuses on the major mathematical tools used by these signal extraction methods: *Fixed Interval Smoothing* (e.g., Bryson and Ho, 1969; Jakeman and Young, 1979, 1984; Norton, 1986; Young, 1988); *Wiener-Kolmogorov-Whittle Optimal Signal Extraction* (OSE) theory (e.g., Kolmogorov, 1941; Wiener, 1949; Whittle, 1963, 1983; Bell, 1984); and *Deterministic Regularisation* (DR) in its various forms, including smoothing splines (e.g., Wahba, 1990) and kernel smoothing (e.g., Wand and

Jones, 1995)¹. The inter-relationships between these methods are explained and the practical advantages of the FIS approach are emphasised. In addition, the paper argues that OSE theory provides the most appropriate analytical tool for understanding the smoothing problem.

For simplicity and clarity of presentation, the main methodological aspects of FIS, DR and OSE are described first within the context of a simple but important time series problem: namely the estimation of the low frequency trend associated with a nonstationary time series consisting of two additive components: the trend itself and an irregular component. In particular, the time series y_t is assumed to be described by the following univariate UC model,

$$y_t = T_t + e_t \quad : \quad t = 1, 2, \dots, N \quad (1)$$

where T_t is a low frequency trend component; e_t is a zero mean, serially uncorrelated sequence of random variables (discrete-time ‘white noise’) with variance σ_e^2 , which represents the deviations about the trend; and N is the total sample size. In general, the various simple, second order, smoothing algorithms discussed here can be used to estimate the trend T_t even if e_t exhibits serial correlation. In this situation, however, they will not perform optimally in a statistical sense unless additional components are added to account for the serial correlation, as discussed in section 6. of the paper.

2. FIXED INTERVAL SMOOTHING (FIS)

The State Space (SS) formulation of the UC model in equation (1) consists of following state and observation equations, i.e.

$$\begin{aligned} \text{Observation Equation:} & \quad y_t = \mathbf{H}_t \mathbf{x}_t + e_t \\ \text{State Equations} & \quad : \quad \mathbf{x}_t = \mathbf{F}_{t-1} \mathbf{x}_{t-1} + \boldsymbol{\eta}_t \end{aligned} \quad (2)$$

where \mathbf{x}_t is an n dimensional state vector; y_t is the scalar observed variable; $\boldsymbol{\eta}_t$ is an n dimensional vector of zero mean, white noise inputs (system disturbances) with diagonal covariance matrix \mathbf{Q} ; and e_t is the white noise variable in equation (1), which is assumed to be independent of $\boldsymbol{\eta}_t$. \mathbf{F}_t and \mathbf{H}_t are, respectively, the $n \times n$ state transition matrix and

¹Kernel smoothing is not considered in detail in this paper since the CMA kernel tends to be selected in a more *ad hoc* manner than in the case of the FIS, DR and OSE methods: it is normally formulated in a non-recursive, local least squares regression form, with the cost function modulated by the kernel (e.g Gaussian, quadratic, Epanechnikov) whose bandwidth defines the smoothness. In simple situations, it can produce results very similar or identical to the other methods (see e.g. Young and Pedregal, 1996).

the $1 \times n$ observation vector, which relates the state vector \mathbf{x}_t to the scalar observation y_t . A typical example of this SS model, and one which is particularly important in the present context, is the Integrated Random Walk (IRW) model², where:

$$\mathbf{F} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}; \quad \mathbf{H} = [1 \quad 0]; \quad \mathbf{Q} = \begin{bmatrix} 0 & 0 \\ 0 & \sigma_\eta^2 \end{bmatrix} \quad (3)$$

or, in expanded form,

$$\begin{aligned} y_t &= x_{1,t} + e_t \\ x_{1,t} &= x_{1,t-1} + x_{2,t-1} \\ x_{2,t} &= x_{2,t-1} + \eta_t \end{aligned}$$

where the first state variable $x_{1,t} = T_t$ is the trend; the second state variable is the differenced trend $d_t = T_{t+1} - T_t = x_{1,t+1} - x_{1,t}$; and σ_η^2 is the variance of the scalar (in this case) white noise input η_t . In other words, the trend is assumed to evolve stochastically such that its rate of change is a random walk process with the variance of the increments specified by σ_η^2 .

Whatever trend model is utilised, however, the following FIS algorithm provides the optimally smoothed estimate $\hat{\mathbf{x}}_{t|N}$ of the state vector \mathbf{x}_t , together with an estimate of its associated covariance matrix $\mathbf{P}_{t|N}^* = \sigma_e^2 \mathbf{P}_{t|N}$:

1. Forward Pass Filtering Equations (Kalman, 1960)

Prediction:

$$\begin{aligned} \hat{\mathbf{x}}_{t|t-1} &= \mathbf{F}_{t-1} \hat{\mathbf{x}}_{t-1} \\ \mathbf{P}_{t|t-1} &= \mathbf{F}_{t-1} \mathbf{P}_{t-1} \mathbf{F}_{t-1}^T + \mathbf{Q}_n \end{aligned}$$

Correction:

$$\begin{aligned} \hat{\mathbf{x}}_t &= \hat{\mathbf{x}}_{t|t-1} + \mathbf{P}_{t|t-1} \mathbf{H}_t^T \left[1 + \mathbf{H}_t \mathbf{P}_{t|t-1} \mathbf{H}_t^T \right]^{-1} \{ y_t - \mathbf{H}_t \hat{\mathbf{x}}_{t|t-1} \} \\ \mathbf{P}_t &= \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1} \mathbf{H}_t^T \left[1 + \mathbf{H}_t \mathbf{P}_{t|t-1} \mathbf{H}_t^T \right]^{-1} \mathbf{H}_t \mathbf{P}_{t|t-1} \end{aligned} \quad (4.1)$$

2. Backward Pass Smoothing Equations (e.g., Bryson and Ho, 1969)

The FIS algorithm has a recursive form running backwards from the end of the sample to the beginning, with the Lagrange multiplier vector $\mathbf{L}_N = 0$ at the start of the backward pass:

²Harvey (1989) would refer to this as a Local Linear Trend (LLT) model with the variance of the noise on the trend state constrained to zero. However, the IRW terminology has been in use for a considerably longer time (see e.g. Norton, 1975; Young, 1984) and will be retained here.

$$\begin{aligned}
\hat{\mathbf{x}}_{t|N} &= \mathbf{F}_t^{-1} [\hat{\mathbf{x}}_{t+1|N} + \mathbf{Q}_t \mathbf{L}_t] \\
\mathbf{L}_t &= \left[\mathbf{I} - \mathbf{P}_{t+1} \mathbf{H}_{t+1}^T \mathbf{H}_{t+1} \right]^T \left[\mathbf{F}_{t+1}^T \mathbf{L}_{t+1} - \mathbf{H}_{t+1}^T \{y_{t+1} - \mathbf{H}_{t+1} \hat{\mathbf{x}}_{t+1}\} \right] \\
\mathbf{P}_{t|N} &= \mathbf{P}_t + \mathbf{P}_t \mathbf{F}_t^T \mathbf{P}_{t+1|t}^{-1} [\mathbf{P}_{t+1|N} - \mathbf{P}_{t+1|t}] \mathbf{P}_{t+1|t}^{-1} \mathbf{F}_t \mathbf{P}_t
\end{aligned} \tag{4.2}$$

Remarks

1. The above equations are not in their most common form: they have been manipulated into a computationally and theoretically more convenient normalised form, where $\mathbf{P}_{t|N} = \mathbf{P}_{t|N}^* / \sigma_e^2$ and $\mathbf{Q}_n = \mathbf{Q} / \sigma_e^2$, the latter representing the diagonal Noise Variance Ratio (*NVR*) matrix. In the case of the IRW model (3), the only non-zero element of \mathbf{Q}_n is the scalar *NVR* associated with the white noise input η_t ; i.e. $\mathbf{Q}_n(2,2) = \sigma_\eta^2 / \sigma_e^2$ (see below). The second order (i.e. $n = 2$) smoothing algorithm obtained in this manner has been used for more than twenty years as a data pre-processing tool in the commercial CAPTAIN and microCAPTAIN computer programs (e.g. Young and Benner, 1991), where it is called the IRWSMOOTH filter.
2. This is not the only form of the FIS algorithm (see e.g., Young, 1984, page 96 *et seq.*) but it is the one originally suggested in the 1960's (e.g., Bryson and Ho, 1969), which we have found to be the most useful and reliable form over many years. It must be emphasised, however, that it is *not* the same form as that proposed by Anderson and Moore (1979), which is often quoted in the statistical and econometrics literatures (e.g., Harvey, 1984, 1989) but little used in the 'Systems' applications. Indeed, the algorithm (4.1)-(4.2) is closely related to the algorithm proposed recently by Koopman (1993), as discussed by Young and Tych (1996).
3. Note that the full FIS equations constitute an *off-line* smoother, in the sense that the smoothed estimate $\hat{\mathbf{x}}_{t|N}$ of the state vector \mathbf{x}_t is obtained only by access to the *whole* data set of N samples and the algorithm cannot completely process the data as they are received on-line. This is in contrast to the forward pass KF part of the algorithm, which yields estimates $\hat{\mathbf{x}}_t = \hat{\mathbf{x}}_{t|t}$ of the state vector \mathbf{x}_t that can be processed *on-line* and in real-time, as the data are received, if this is desired by the user. In order to differentiate these operations, it is convenient to refer to an algorithm that has the potential for on-line use as a 'filter'; and one that can only be exploited off-line, as 'smoother' or 'smoothing filter'. Clearly, the filter, in this sense, is often more convenient for practical application; but the smoother will always yield a lower variance estimate of the state at any sampling instant. In the present

context, therefore, where we are concerned with the best estimate of the smooth trend T_t in equation (1), it makes sense to consider only smoothing solutions obtained by off-line analysis of the time series y_t , $t = 1, 2, \dots, N$.

4. Because it is formulated in stochastic state space terms, the FIS approach is extremely flexible and can be extended straightforwardly to handle much more complex problems than simple trend estimation. Such extensions are discussed later in section 6 of the paper.

3. DETERMINISTIC REGULARISATION (DR)

The non-recursive approach, known in numerical analysis as regularisation and in the statistical literature as spline smoothing, is based on the computation of an estimate $\hat{T}_{t|N}$ which minimises a deterministic cost function of the general form (see e.g., Jakeman and Young, 1979, 1984; Young, 1993),

$$J = \sum_{t=1}^N \left\{ (y_t - \hat{T}_{t|N})^2 + \lambda (\nabla^2 \hat{T}_{t|N})^2 \right\} \quad (5)$$

In other words, the trend component is computed so that it minimises the sum of the squares of the difference between the trend and the series, with the constraint that it should have a given level of smoothness, as measured by the square of the 2nd difference of the estimated trend. This constraint is introduced by the Lagrange Multiplier, or smoothing parameter, λ , which is selected in some manner, as defined by the user.

The solution to the optimisation problem (5) is obtained in the standard manner as,

$$\hat{T}_{t|N} = \left[\mathbf{I} + \lambda \mathbf{U}^T \mathbf{U} \right]^{-1} \mathbf{y} = \mathbf{W} \mathbf{y} \quad (6)$$

where \mathbf{I} is the $N \times N$ unity matrix; \mathbf{y} is the $1 \times N$ vector of data; and \mathbf{U} is the following $N \times N$ matrix:

$$\mathbf{U} = \begin{bmatrix} 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 1 & -2 & 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & -2 & 1 \end{bmatrix}$$

The $N \times N$ matrix \mathbf{W} has rows which define the nature of the moving average weighting applied to the data to yield the estimate of the trend. Figure 1 shows a plot of every

fourth row in the matrix for a data set with $N=110$ and $\lambda = 90$. The weighting is symmetric in the centre of the data but becomes asymmetric at the ends to handle the end effects in this 2-point boundary value problem. An enlarged version of the plot at the left hand side shows this asymmetry in more detail (note that the area under each curve is always maintained at unity).

Remarks

1. It is clear from equation (6) that DR requires the inversion of a matrix with the same dimension N as the data, so that it can be very computationally intensive and requires a large memory size for long series unless some device for handling large matrices is introduced at the computational stage. This is in contrast to the recursive FIS algorithm which, in the case of the present trend estimation problem, requires no matrix inversion at all.
2. As in the case of FIS, it is straightforward to generalise the above DR formulation. For example, the Lagrange multiplier term could be defined in terms of the j th difference of the trend estimate, where $j = 1, 2, \dots, d$ (see the comments relating to this in the next section 4.). And even more complex generalisations can be developed straightforwardly, as discussed later in section 6.
3. The so-called Hodrick-Prescott (HP) detrending method (Hodrick-Prescott, 1980) is a special case of this DR approach obtained when $\lambda = 1600$ (see later discussion in sections 5 and 6).

4. WIENER-KOLMOGOROV-WHITTLE OPTIMAL SIGNAL EXTRACTION

The original derivations of optimal filter theory appeared in Kolmogorov (1941) and Wiener (1949). More recently, optimal filter and signal extraction theory for stationary time series has been presented by Whittle (1963, 1983); and for nonstationary time series by Bell (1984). In contrast to FIS estimation, OSE is normally formulated in the frequency domain based on a transfer function (TF) model of the stochastic process. For example, the TF (or ‘reduced’) form of the IRW trend model introduced in section 2. is given by,

$$y_t = T_t + e_t \quad ; \quad T_t = \frac{1}{\nabla^2} \eta_t \quad (7)$$

where $\nabla = 1 - L$ is the difference operator and L is the lag (or ‘backward shift’) operator, i.e. $Ly_t = y_{t-1}$. The OSE estimate $\hat{T}_{t|N}$ of the trend is then given by the expression,

$$\hat{T}_{t|N} = \frac{g(T_t)}{g(T_t + e_t)} y_t = \frac{\frac{\sigma_\eta^2}{(1-L)^2(1-F)^2}}{\frac{\sigma_\eta^2}{(1-L)^2(1-F)^2} + \sigma_e^2} y_t = \frac{NVR}{NVR + (1-L)^2(1-F)^2} y_t \quad (8)$$

where $g(T_t)$ and $g(T_t + e_t)$ are, respectively, the spectral generating functions of the ‘signal’ T_t and the ‘signal+noise’ $T_t + e_t$; $NVR = \sigma_\eta^2 / \sigma_e^2$ is the signal-noise ratio (or ‘Noise Variance Ratio’: see previous section), and F is the lead (or ‘forward shift’) operator, i.e. $Fy_t = y_{t+1}$.

Remarks

1. The IRW smoothing filter defined by equation (3) is a low-pass filter, in which the spectral characteristics are defined as follows by the NVR parameter,

$$f_T(\omega) = \frac{NVR}{NVR + \{2(1 - \cos \omega)\}^2}$$

2. More general formulations than that given above for the IRW trend model are clearly possible. For example, the OSE smoother for the whole family of multiple IRW smoothing filters (i.e. IRW, Double Integrated Random Walk (DIRW), Triple Integrated Random Walk (TIRW), etc.) is obtained simply by replacing the 2nd power in the IRW trend definition by a general i th power, where $i = 2, 3, 4 \dots$ etc. The spectral characteristics of the OSE smoother defined in this more general manner are given by,

$$f_T(\omega) = \frac{NVR}{NVR + \{2(1 - \cos \omega)\}^i} \quad (9)$$

where i is the order of integration (see figure 3). From this formula, the ‘cut-off’ frequency that corresponds to a given power of the spectrum is given by (Pedregal, 1995)

$$\omega = \arccos \left[1 - \left\{ \frac{NVR(1 - \alpha)}{4\alpha} \right\}^{\frac{1}{i}} \right] \quad (10)$$

in which α is the specified percentage cut-off (usually 50% in signal processing literature). Note that the special case of this filter when $i = 4$ (i.e. a TIRW modelled trend) is interesting since it can be compared directly with the well known Henderson family of CMA filters used in the X-11 seasonal adjustment program. This is discussed further in Appendix 1.

4. Clearly, as in the case of FIS and DR, it is straightforward to develop still more general formulations of OSE, as discussed later in section 6.
5. Finally, it must be stressed that, unlike the FIS and DR approaches, the *primary* function of the above OSE analysis is to provide a theoretical basis for smoothing, rather than a computational methodology. Although it is possible to use the theory to develop a smoothing algorithm from the theoretical results, this is not entirely satisfactory, since the theory applies only to infinitely long series. Consequently, it is necessary to introduce approximations to handle end effects (see later, section 5). Whittle (1983) concludes that, for most applications, the state space smoothing algorithms have greater generality and are computationally much more attractive.

5. THE INTER-RELATIONSHIPS

Jakeman and Young (1979, 1984) demonstrated the equivalence between the FIS and DR approaches by considering FIS within the context of Maximum Likelihood estimation (see also Young, 1991). Bearing in mind the properties of Markov processes and the state space system (1), the density function defining the log-likelihood is given by

$$\log p(\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_N | \mathbf{Y}_N) = \log \{ p(\mathbf{Y}_N | \mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_N) p(\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_N) \} - \log p(\mathbf{Y}_N).$$

where \mathbf{Y}_i stands for the information contained in the variable up to the i th sample; while y_i represents only the information at the i th sampling instant. Because the term $\log p(\mathbf{Y}_N)$ does not depend on the states, it can be dropped from the log-likelihood, giving as a result

$$\begin{aligned} \log p(\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_N | \mathbf{Y}_N) &= \log \{ p(\mathbf{Y}_N | \mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_N) p(\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_N) \} = \\ &= \log \{ p(\mathbf{Y}_N | \mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_N) p(\mathbf{x}_N | \mathbf{x}_{N-1}) p(\mathbf{x}_{N-1} | \mathbf{x}_{N-2}) \dots p(\mathbf{x}_0) \} = \\ &= \log \left\{ p(y_N | \mathbf{x}_N) p(\mathbf{Y}_{N-1} | \mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_{N-1}) \prod_{t=1}^N p(\mathbf{x}_t | \mathbf{x}_{t-1}) p(\mathbf{x}_0) \right\} = \\ &= \log \left\{ \prod_{t=1}^N p(y_t | \mathbf{x}_t) \prod_{t=1}^N p(\mathbf{x}_t | \mathbf{x}_{t-1}) p(\mathbf{x}_0) \right\} \end{aligned}$$

Assuming Gaussian distributions and recognising that

$$E(y_t|\mathbf{x}_t) = \mathbf{H}_t\mathbf{x}_t; V(y_t|\mathbf{x}_t) = E(y_t - \mathbf{H}_t\mathbf{x}_t)(y_t - \mathbf{H}_t\mathbf{x}_t)^T = \sigma_e^2; E(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathbf{F}_{t-1}\mathbf{x}_{t-1}$$

$$V(\mathbf{x}_t|\mathbf{x}_{t-1}) = E(\mathbf{x}_t - \mathbf{F}_{t-1}\mathbf{x}_{t-1})(\mathbf{x}_t - \mathbf{F}_{t-1}\mathbf{x}_{t-1})^T = \mathbf{Q}$$

we have that the maximisation of the log-likelihood function is the same than minimising the following expression

$$J = \sum_{t=1}^N \|y_t - \mathbf{H}_t\mathbf{x}_t\|_{1/\sigma_e^2}^2 + \sum_{t=1}^N \|\mathbf{x}_t - \mathbf{F}_{t-1}\mathbf{x}_{t-1}\|_{\mathbf{Q}^{-1}}^2 \quad (11)$$

Comparing this cost function with the DR cost function in (5), it is clear that, if \mathbf{F}_t and \mathbf{H}_t are defined as for the IRW model in equation (3), they are identical if σ_e^2 is normalised to unity and

$$\mathbf{Q}^{-1} = \begin{pmatrix} 0 & 0 \\ 0 & \lambda \end{pmatrix}.$$

This is confirmed by practical implementation, as shown in figure 2, which compares the FIS and DR results obtained when the algorithms so defined are applied to the airline passenger data in Box and Jenkins (1970). Here, the DR estimation is carried out with the value of the $\lambda = 1600$ used in the HP filter; while the FIS algorithm utilised the equivalent $NVR = 1/\lambda = 0.000625$.

It is clear from the identical nature of these results that the algorithms, although very different in computational form, are entirely equivalent in their smoothing effect on the time series. In particular, the CMA operation that is explicit in the non-recursive DR algorithm is implicit in the recursive FIS algorithm. Moreover, this equivalence is quite general: for any FIS algorithm defined by a selected SS model, it is possible to formulate an equivalent DR cost function that will yield identical results. An example of this is general equivalence is revealed in the work of Akaike (1980) who uses a DR formulation which includes a seasonal component, in addition to the trend. In this case the cost function takes the following form, where the nomenclature is that used by Akaike,

$$J = \sum_{t=1}^N \left[\left(y_t - \hat{T}_{t|N} - \hat{S}_{t|N} \right)^2 + d^2 \left\{ \left(\nabla^i \hat{T}_{t|N} \right)^2 + r^2 \left(\nabla_s \hat{S}_{t|N} \right)^2 + z^2 \left(\hat{S}_{t|N} + \hat{S}_{t-1|N} + \dots + \hat{S}_{t-11|N} \right)^2 \right\} \right]$$

Here $\hat{S}_{t|N}$ is the estimate of the seasonal component; s is the seasonal period; $\nabla_s = (1 - L^s)$ is the seasonal difference operator; and $\{d, r, z\}$ are Lagrange multipliers which Akaike estimates within a Bayesian framework. If the term $r^2 \left(\nabla_s \hat{S}_{t|N} \right)^2$ is

dropped from the previous equation, the problem formulated in these terms yields a similar result to that obtained using the ‘Basic Structural Model’ with seasonal dummies as defined by Harvey (1984).

The equivalence between the DR approach and OSE is also easily established by considering the latter from the standpoint of its implied CMA operation. Considering once more the IRW model for the trend in equation (3), the symmetric, infinite dimensional CMA must satisfy the following equation,

$$\frac{NVR}{NVR + (1-L)^2(1-F)^2} = (\dots + \varphi_2 L^2 + \varphi_1 L + \varphi_0 + \varphi_1 F + \varphi_2 F^2 + \dots)$$

or, alternatively

$$NVR = \{L^2 - 4L + (NVR + 6) - 4F + F^2\}(\dots + \varphi_2 L^2 + \varphi_1 L + \varphi_0 + \varphi_1 F + \varphi_2 F^2 + \dots)$$

Clearly the complete solution of this equation is not possible because of the infinite dimension, which follows directly from the nature of OSE theory. However, an approximate solution can be obtained by truncating the infinite polynomial to a finite one and equating coefficients by like powers in L on both sides of this equation. This yields the following set of linear equations

$$\mathbf{Z}\boldsymbol{\psi} = \mathbf{a}$$

where

$$\mathbf{Z} = \begin{pmatrix} NVR+6 & -8 & 2 & 0 & 0 & \dots & 0 & 0 & 0 \\ -4 & NVR+7 & -4 & 1 & 0 & \dots & 0 & 0 & 0 \\ 1 & -4 & NVR+6 & -4 & 1 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \dots & NVR+6 & -4 & 1 \\ 0 & 0 & 0 & 0 & 0 & \dots & -4 & NVR+6 & -4 \\ 0 & 0 & 0 & 0 & 0 & \dots & 1 & -4 & NVR+6 \end{pmatrix} \quad (12)$$

$$\boldsymbol{\psi} = (\varphi_0 \quad \varphi_1 \quad \varphi_2 \quad \dots \quad \varphi_N)^T \quad \mathbf{a} = (NVR \quad 0 \quad \dots \quad 0)^T$$

The vector of weights $\boldsymbol{\psi}$, which are then obtained from the equation,

$$\boldsymbol{\psi} = \mathbf{Z}^{-1}\mathbf{a} \quad (13)$$

are exactly the same as those in the rows of the \mathbf{W} matrix in the DR approach. As in the comparison between the FIS and DR approaches, this equivalence applies for any kind of model and any value of the NVR parameter.

Unfortunately, this equivalence between OSE and DR and, therefore, between OSE and FIS, is only approximate, since the purely theoretical origins of the OSE smoother mean

that it is only applicable to infinitely long time series. Thus, while the CMA weights that underlie both the DR and FIS approaches are automatically modified at the beginning and end of the sample in order to allow for end-effects, as shown in figure 1, the notional weights in OSE remain the same over the whole sample. Consequently, any approximate implementation of the OSE smoother (which is clearly not advisable in practice), requires the series to be extended artificially in some manner (e.g., by forecasting and backcasting, as used in X-11 seasonal adjustment) in order to obtain satisfactorily smoothed estimates over the whole sample. This is not only an inconvenient and inelegant requirement, but it also means that the smoothing results will be overly dependent upon the efficacy of the forecast/backcasting procedures used to extend the series. Consequently, it is clearly better to regard the FIS and DR solutions as practically implementable versions of the OSE smoother and consider OSE theory itself as one of the fundamental theoretical bases for these solutions (the other being Kalman filter/smoothing theory). In this sense, the inter-relationships exposed here also reveal graphically, and in a novel manner, the well known link between the steady state Kalman filter and the Wiener filter.

Finally, since the results obtained from all the three smoothing filters are virtually identical, it is worth noting one, at first quite surprising, property which they all possess: namely, as shown in Appendix 1, that all the estimated components can be expressed as filtered versions of only one of them, showing that there is only one true source of noise in the model. For example, in a UC model with an IRW modelled trend, the fourth difference of the estimated trend is *exactly* equal to the estimated perturbations (i.e. the detrended data), lagged by two samples and re-scaled by a factor that is *exactly* equal to the *NVR* parameter used in the estimation.

6. MORE COMPLEX PROBLEMS AND FORMULATIONS

For simplicity and clarity of presentation, previous sections of the paper have concentrated on a particularly simple problem; namely the estimation of the low frequency trend in a nonstationary time series that can be modelled by equation (1). But all of these approaches can be applied to much more complex processes than this. To illustrate this point, let us consider only the recursive FIS approach since, as we shall argue in the next section 7., this has certain advantages, in both theory and practice, over the non-recursive methods. While the DR and OSE approaches could be extended in a similar manner to that discussed here, neither their formulation nor the implementational aspects are as simple or attractive as FIS.

Previous research by the authors and their colleagues on the extrapolation, interpolation and smoothing of nonstationary time series (Young, 1988; Young *et al.*, 1989, 1991; Ng and Young, 1990; Pedregal and Young, 1996) have utilised models which can be generalised to the following UC model form,

$$y_t = T_t + C_t + S_t + N_t + f(\mathbf{u}_t) + e_t \quad (14)$$

where, as before, T_t is a low frequency trend and e_t is a white noise term. The additional components represent the kind of features one tends to see in nonstationary time series: S_t is an undamped but variable amplitude seasonal or periodic component, which could itself be composed of many harmonic frequency components (e.g., in the case of the highly periodic hourly demand for electricity, there are up to 84 harmonics); C_t is a damped periodic (AR or cyclical) component; N_t is a stochastic perturbation (AR/ARMA or coloured noise component); and $f(\mathbf{u}_t)$ represents the (possibly nonlinear) effect of an ‘exogenous’ or input vector $\mathbf{u}_t = [u_1 \ u_2 \ \dots \ u_p]^T$, where the dimension p will depend upon the number of significant exogenous variables and $f(\bullet)$ is a general nonlinear function. This latter component also allows for the many kinds of nonlinear model that are available in the statistical literature: from non-parametric regression, through locally weighted kernel regression, to radial basis functions.

By considering equation (14) as the observation equation in a SS model and formulating SS models for any of the components on the right hand side of the equation that require this, it is straightforward to use the FIS algorithm for extrapolation, interpolation, and smoothing, as well as for operations that require the extraction of estimated components, such as seasonal adjustment. Indeed, this model can be extended much further to include vector (multivariable) observations, although this would considerably increase the complexity of the FIS solution, except in the simplest, highly constrained situations (see e.g. Harvey, 1989).

A good example of a UC model, such as equation (14), is the *Dynamic Harmonic Regression* (DHR: Ng and Young, 1990; Young and Tych, 1996). Its application to the problem of signal extraction for the famous monthly time series of atmospheric CO₂ concentration at Mauna Loa in Hawaii from 1970 to 1985 is illustrated in figure 6. Here, the DHR model includes a trend and a monthly seasonal component. In addition, all the samples for the year 1978 have been missed out intentionally to show the capabilities of the methodology in dealing with a large number of consecutive missing samples. The analysis shows clearly that not only is the trend increasing rapidly, but there is also a slight increase in the amplitude of the monthly oscillations. The derivative of the trend is often a valuable instrument for detecting possible cycles present in the data: for instance, in this case, it is possible to see some irregular, small amplitude oscillations in the estimated derivative as shown in figure 6. This means that, although the model seems

satisfactory, further analysis taking into account these estimated changes could yield marginal improvements and insights into the changing nature of the atmospheric CO₂ concentration over this period.

Clearly, it is wiser to see the trend estimation problem in (1) as a special case of the UC model (14), since the former applies only if the low frequency trend and white noise components alone are able to fully explain the series. In all other cases, additional terms, such as those on the right hand side of the equation, should be introduced. This more generalised formulation of the trend estimation problem illustrates the need for great care when a simple second order smoother, such as the HP or IRWSMOOTH algorithm, is applied to a series where there is clear evidence of other significant components at intermediate and high frequencies. In this situation, objective maximum likelihood optimisation of the smoothing parameter (NVR or λ) will yield a trend estimate which includes the other components and so cannot really be considered as a ‘trend’ in the usual meaning of the term. A more suitable trend estimate can be obtained from the simple second order smoothing algorithms by selecting the NVR value on the basis of spectral considerations, so that the algorithm acts as a sensible low-pass filter and extracts only the low frequency components of interest. But then caution is necessary since the end effects in this situation can be considerable and the trend estimate can be deleteriously influenced by the perturbations of the coloured residual (i.e. the detrended component $y_t - \hat{T}_t$) at either end of the series.

These latter considerations show how careful one must be in using the simple second order FIS and DR algorithms, such as the HP and IRWSMOOTH smoothing filters, for general smoothing applications. On the other hand, if such simple filters are used carefully, with a judicious choice of the smoothing parameter³, then meaningful results can be obtained, despite the dire warnings that have appeared recently in the literature (see e.g., Cogley and Nason, 1995; as well as the comments by the present authors on this paper, 1996). After all, these algorithms perform simply as zero lag, low pass filters and *if used carefully as such*, they can perform a quick and useful operation that may well reveal interesting aspects of the data deserving of further investigation (see e.g., Young, 1994). The fact that the filter preferentially amplifies certain frequency components in relation to others is the *whole point* of a signal processing filter and the output of the filter simply represents the information in the series over the defined pass-band of the filter. There is no danger in performing such low-pass filtering operations

³ The specification of a fixed value of the smoothing parameter (e.g. the $\lambda = 1600$ proposed for the HP filter) is clearly not justified in these general terms and is not recommended: the smoothing parameter must be selected by reference to the time series, its spectral characteristics, its sampling interval and the nature of the problem at hand.

provided the nature of the filter is taken fully into account when evaluating the filtered output.

7. THE ADVANTAGES OF FIS

Previous sections of the paper have shown how two well known methods of smoothing and trend estimation, FIS and DR, are equivalent and can be considered to have a common theoretical link via Wiener-Kolmogorov-Whittle-Kalman filter-smoothing theory considered in CMA terms. But while these methods are equivalent and can, with suitable choice of parameters, produce identical results, they are not equally useful in practical and computational terms. Indeed, the aim of this section is to argue that the FIS approach has significant advantages over DR (and related procedures) in terms of flexibility, combined with computational elegance and efficiency. In this we echo and extend the remarks of Whittle (1983, page *xi*).

The first advantage of FIS lies in the stochastic state space formulation, which allows for an almost infinite variety of different model formulations, some of which have been outlined in the previous section 6. While it is clear that DR could be extended to obtain similar solutions, as demonstrated by Akaike's work in this area (Akaike, 1980), it seems much more straightforward to formulate such extensions within the convenient stochastic SS framework and exploit recursive estimation. Indeed, this framework even allows for the introduction of mechanistic (physically meaningful) models, since most mechanistic models can be formulated most naturally in such SS terms. As such, the resulting algorithms should be more attractive to scientists (and econometricians), who tend to be sceptical about purely 'black-box' methods.

The second advantage of FIS is its elegant and flexible computational nature, deriving mainly from the inherent recursive formulation. As the first author has pointed out (Young, 1984), most existing, non-recursive, (*en bloc*) methods of time series analysis can be derived or converted into recursive form, where they inherit numerous advantages. These include: an inherent Bayesian interpretation; automatic handling of missing observations or gaps in the data; abrupt changes in level or slope; forecasting and backcasting outside the data; and on-line utilisation in applications such as adaptive signal processing, forecasting and control. The algorithms can also be extended easily to detect and allow for outliers, including automatic 'robustness' modifications, which are particularly simple in recursive processing. In contrast, when using non-recursive approaches, such as DR or OSE, it is necessary to introduce additional and more *ad hoc* statistical processing tools to handle these problems before proceeding to the estimation of components.

Finally, there is the question of algorithmic optimisation. All of the algorithms discussed here involve the specification of parameters, such as the NVR values in the FIS algorithms, the Lagrange multiplier parameters (e.g. λ) in the DR algorithms, and the ‘bandwidth’ parameters required in the specification of smoothing kernels (see e.g. Hart, 1991). In order to obtain optimum smoothing performance in an objective manner, it is necessary to optimise these ‘hyper-parameters’ in some manner. Within the DR and smoothing kernel context, for example, the favoured approach is often some form of cross-validation (see e.g. Bowman, 1984; Rudemo, 1982). But, once again, the FIS approach has an inherent advantage because of its recursive formulation: within this recursive context, it seems very natural to optimise the hyper-parameters by either maximum likelihood based on *Prediction Error Decomposition* (Schweppe, 1965); or some related method, such as the minimisation of the sum-of-squares of the recursive one- or multiple-step-ahead prediction errors⁴.

7. CONCLUSIONS

This paper has considered the close inter-relationships that exist between three, apparently different, statistical instruments used for signal extraction in nonstationary time series analysis: *Fixed Interval Smoothing* (FIS) based on Kalman filter/smoothing theory; *Deterministic Regularisation* (DR) which exploits a special form of Lagrange multiplier-constrained numerical optimisation; and *Wiener-Kolmogorov-Whittle Optimal Signal Extraction* (OSE) theory. Although passing references to the links between these methods can be discerned from the literature, no clear statement of the fact that they are virtually identical has appeared before.

The equivalence of the results obtained from these methods has been illustrated by considering them all in terms of their explicit (DR and OSE) or implicit (FIS) use of closely related Centralised Moving Average (CMA) smoothing operations involving smoothing kernels. In particular, it is shown that the FIS and DR algorithms can be made identical by appropriate formulation of the smoothing problem in each case; while the similarly formulated OSE smoother involves the same basic CMA operation but without inherent adjustment for end effects.

⁴Other, less obvious approaches are also possible. For example, in the case of the DHR model discussed in section 6, we optimise the hyper-parameters by fitting the logarithm of the theoretical DHR pseudo-spectrum to the logarithm of the AR spectrum, with the AR order defined by AIC optimisation.

The paper has also revealed an interesting property of the various smoothing algorithms: namely that all the estimated components in a UC model can be obtained by filtering one of them, showing that the UC model is only affected by a unique source of noise. A related result shows that, given an estimated UC model for two series and a known linear relation between the two series, then the relationships among all the other components can be derived straightforwardly.

Finally, despite the underlying equivalence between the different approaches to signal extraction revealed by the paper, we have argued strongly that the FIS algorithm has many advantages, in both computational and practical terms, that make it superior to DR and OSE. As a general approach to signal extraction, the recursive state space formulation of FIS is an extraordinary powerful general framework into which many problems, from diverse disciplines, can be introduced quite naturally and then solved in an elegant and computationally efficient manner.

APPENDIX 1 HENDERSON DETRENDING AND TIRW TREND ESTIMATION

CMA filtering techniques were developed by actuaries in the early part of the 20th century to smooth out irregularities in data and allow for interpolation in measures such as mortality rates. In 1916, Henderson developed certain CMA filters that have since been used in a number of application areas, most importantly (since the 1950's) in the Census X-11 seasonal adjustment procedure used by numerous government agencies all over the World. This Appendix, which derives from Young (1993) shows the close relationship between the CMA weighting patterns that define the Henderson smoothing filter and the implicit CMA weighting used by the FIS algorithm when the trend is modelled as a TIRW. These results confirm the recent theoretical analysis of Wallis (1993).

Consider the following UC model, where the series is represented as trend plus white noise, with the trend is modelled as a TIRW process:

$$\begin{aligned} y_t &= \mathbf{H}_t \mathbf{x}_t + e_t \\ \mathbf{x}_t &= \mathbf{F} \mathbf{x}_{t-1} + \eta_t \end{aligned} \tag{A1.1}$$

where,

$$\mathbf{F} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \mathbf{H} = [1 \quad 0 \quad 0 \quad 0] \quad \mathbf{Q} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_\eta^2 \end{bmatrix}$$

and \mathbf{Q} is the covariance matrix of the white noise input vector η_t . We will refer to the FIS algorithm based on this model as the TIRWSMOOTH filter, since it is a higher order relation of the IRWSMOOTH filter mentioned in the main text.

Figure 4 compares implicit CMA weighting patterns of such TIRWSMOOTH filters defined by various NVR values with the weighting patterns used by the Henderson 17, 13 and 5 term CMA filters. The similarities between the CMA weighting are obvious on inspection: the only difference of note is that the Henderson CMA filter is only finite dimensional, while the TIRWSMOOTH filter is nominally infinite dimensional. In practice, of course, the TIRWSMOOTH filter has the advantage that it has inherent end effect compensation, whereas the Henderson filters need to make additional allowance for end effects: in the X-11 context, for example, this involves forecasting and backcasting operations as a prelude to Henderson filtering over the artificially enlarged series.

APPENDIX 2 THE RELATIONSHIP BETWEEN THE COMPONENTS IN THE UNOBSERVED COMPONENT MODEL ESTIMATED BY FIS

This Appendix reveals how the trend estimate obtained from the FIS algorithm (or the similarly defined DR and OSE algorithms, since the paper shows how these are equivalent) contains all the information on the detrended residual component; and that the detrended component (i.e. the perturbations about the trend) can be recovered from the trend by appropriate operations. It goes on to point out that this result can be generalised so that, in any UC model estimated in the manner considered here, all the estimated components can be expressed as filtered versions of only one of them, implying that there is only one true source of noise in the model.

Consider the i th order generalisation of equation (8) in the main text, i.e.,

$$\hat{T}_{t|N} = \frac{NVR}{NVR + (1-L)^i(1-F)^i} y_t$$

If this equation is differenced j times, then

$$\nabla^j \hat{T}_{t|N} = \frac{\nabla^j NVR}{NVR + (1-L)^i(1-F)^i} y_t$$

The FIS estimate $\hat{P}(t|N)$ of the perturbations about the trend can then be found either by using the OSE smoother again or by computing the perturbation estimate directly as $\hat{P}_{t|N} = y_t - \hat{T}_{t|N}$, i.e.,

$$\hat{P}_{t|N} = \frac{(1-L)^i(1-F)^i}{NVR + (1-L)^i(1-F)^i} y_t$$

Combining the two previous equations we obtain,

$$\nabla^j \hat{T}_{t|N} = NVR \frac{\nabla^{j-i}}{(1-F)^i} \hat{P}_{t|N} \quad (\text{A2.1})$$

This equation shows how the j th difference of the estimated trend can be recovered from the estimated perturbations. Similar formulae can be obtained for far more general UC model's, with any number of components and any definition of these components within the FIS context (Pedregal, 1995).

Let us now consider a particular case of equation (A2.1) to highlight the importance of the above result. When $i=2$ (IRW trend) and $j=4$ (fourth difference), we have

$$\nabla^4 \hat{T}_{t|N} = NVR \frac{\nabla^2}{(1-F)^2} \hat{P}_{t|N}$$

Noting that $\nabla^2/(1-F)^2 = L^2$, this equation can be written in the following, more convenient form

$$\nabla^4 \hat{T}_{t/N} = NVR L^2 \hat{P}_{t/N} \quad (\text{A2.2})$$

which reveals that, for UC model's with IRW modelled trends, the fourth difference of the FIS estimated trend is *exactly* equal to the estimated perturbations, lagged by two samples and re-scaled by a factor that is *exactly* equal to the *NVR* parameter used in the estimation. It is important to realise that this result has been obtained analytically using the formal expression of the OSE smoothing filter. Empirically, the result is corroborated using the FIS algorithm for any value of the *NVR* and for any value of *i* and *j*. A typical example is presented in figure 5.

Another interesting result follows from the above. Assume two series are modelled as the following UC models:

$$y = T_y + P_y \quad x = T_x + P_x$$

where the time subscript is suppressed to avoid confusion. The estimates of components are given by,

$$\begin{aligned} \hat{T}_y &= \frac{g(T_y)}{g(y)} y & \hat{P}_y &= \frac{g(P_y)}{g(y)} y \\ \hat{T}_x &= \frac{g(T_x)}{g(x)} x & \hat{P}_x &= \frac{g(P_x)}{g(x)} x \end{aligned}$$

so that the estimates of the trends can be expressed as a function of the perturbations

$$\hat{T}_y = \frac{g(T_y)}{g(P_y)} \hat{P}_y \quad \hat{T}_x = \frac{g(T_x)}{g(P_x)} \hat{P}_x$$

Now, if the trends are related by a linear TF

$$\hat{T}_y = \frac{B(L)}{A(L)} \hat{T}_x + e$$

where $B(L) = b_0 + b_1L + \dots + b_mL^m$ and $A(L) = 1 + a_1L + \dots + a_nL^n$, then the relationship between the estimated perturbations is given by

$$\hat{P}_y = \frac{B(L)}{A(L)} \frac{g(T_x)}{g(P_x)} \frac{g(P_y)}{g(T_y)} \hat{P}_x + \frac{g(P_y)}{g(T_y)} e$$

For example, in the case of IRW trend processes, we have

$$\hat{P}_y = \frac{B(L)}{A(L)} \frac{NVR_x}{NVR_y} \hat{P}_x + \frac{\nabla^4}{NVR_y} e$$

showing that if two IRW trends are related by a linear TF, the perturbations around them are also related by the same TF. The reverse is not necessarily true, however, since integration is not unique (because it depends on initial conditions).

ACKNOWLEDGEMENTS

The authors are most grateful to the UK Engineering and Physical and Science Research Council (EPSRC: Grant ESA7088) and the Spanish Education Ministry (CICYT, program PB-90-0188) for their support.

REFERENCES

- AKAIKE, H. (1980) Seasonal Adjustment by a Bayesian Modelling, *Journal of Time Series Analysis*, 1, pp. 1-13.
- ANDERSON, B. D. O. & MOORE, J. B. (1979) *Optimal Filtering* (Englewood Cliffs, New Jersey, Prentice-Hall).
- BELL, W. R. (1984) Signal Extraction for Nonstationary Time Series, *Annals of Statistics*, 12, pp. 646-664.
- BOWMAN, A. W. (1984) An alternative Method of Cross Validation for the Smoothing of Density Estimates, *Biometrika*, 71, pp. 353-360.
- BOX, G. E. P. & JENKINS, G. M. (1970) *Time Series Analysis: Forecasting and Control* (San Francisco, Holden-Day).
- BOX, G. E. P., HILLMER, S. C. & TIAO, G. C. (1978) Analysis and Modelling of Seasonal Time Series. In: A. ZELLNER (Ed.), *Seasonal Analysis of Economic time Series*, pp. 309-334 (Washington D. C.: US Dept. of Commerce-Bureau of the Census).
- BRYSON, A. E. & HO, Y. C. (1969) *Applied Optimal Control, Optimization, Estimation and Control* (Waltham, Blaisdell Publishing Company).
- BURMAN, J. P. (1980) Seasonal Adjustment by Signal Extraction, *Journal of the Royal Statistical Society A*, 143, pp. 321-337.
- COGLEY, T. & NASON, J. M. (1995) Effects of the Hodrick-Prescott filter on trend and difference stationary time series. Implications for business cycle research, *Journal of Economic Dynamics and Control*, 19, pp. 253-278.
- DAGUM, E. B. (1980) The X-11 ARIMA Seasonal Adjustment Method, *Statistics Canada*, Ottawa.
- DAGUM, E. B. (1988) The X-11/88 ARIMA Seasonal Adjustment Method, *Statistics Canada*, Ottawa.
- DAUBECHIES, I. (1988) Orthonormal Bases of Compactly Supported Wavelets, *Communications on Pure Applied Mathematics*, 41, pp. 906-966.

- FINDLEY, D. F., MONSELL, B. C., OTTO, M. C., BELL, W. R. & PUGH, M. (1992) Towards X-12 ARIMA, mimeo, Bureau of the Census.
- FINDLEY, D. F., MONSELL, B. C., BELL, W. R., OTTO, M. C. & CHEN, B.C. (1996) New Capabilities and Methods of the X-12 ARIMA Seasonal Adjustment Program, U.S. Bureau of the Census, mimeo, May 16.
- GERSCH, W. & KITAGAWA, G. (1983) The Prediction of Time Series with Trends and Seasonalities, *Journal of Business & Economic Statistics*, 1, pp. 253-264.
- GERSCH, W. & KITAGAWA, G. (1984) A Smoothness Priors State Space Modelling of Time Series with Trend and Seasonality, *Journal of the American Statistical Association*, 79, pp. 378-389.
- GÓMEZ, V. & MARAVALL, A. (1996) Programs TRAMO and SEATS, instructions for the user. Documento de Trabajo N. 9628, Servicio de Estudios del Banco de España, Madrid.
- HARRISON, P. J. & STEVENS, C. F. (1976) Bayesian Forecasting, *Journal Royal Statistical Society, Series B.*, 38, pp. 205-247.
- HART, J. D. (1991) Kernel Regression Estimation with Time Series Errors, *Journal of the Royal Statistical Society, B*, 53, pp. 173-187.
- HARVEY, A. C. (1984) A Unified View of Statistical Forecasting Procedures (with comments), *Journal of Forecasting*, 3, pp. 245-283.
- HARVEY, A. C. (1989) *Forecasting Structural Time Series Models and the Kalman Filter*. (Cambridge, Cambridge University Press).
- HARVEY, A. C. & PETERS, S. (1990) Estimation Procedures for Structural Time Series Models, *Journal of Forecasting*, 9, pp. 173-204.
- HASTIE, T. (1996) Pseudosplines, *Journal of the Royal Statistical Society, Series B.*, 58, pp. 379-396.
- HENDERSON, R. (1916) Note on Graduation by Adjusted Average, *Transactions of the Actuarial Society of America*, 17, pp. 43-48.
- HILLMER, S. C. & TIAO, G. C. (1982) An ARIMA-Model Based Approach to Seasonal Adjustment, *Journal of the American Statistical Association*, 77, pp. 63-70.
- HILLMER, S. C., BELL, W. R. & TIAO, G. C. (1983) Modelling Considerations in the Seasonal Adjustment of Economic Time Series. In: A. ZELLNER (Ed.), *Applied Time Series Analysis of Economic Data*, pp. 74-100 (Washington D. C.: US Dept. of Commerce-Bureau of the Census).

- HODRICK, T. & PRESCOTT, E. (1980) Post-War US Business Cycles: an Empirical Investigation, Carnegie Mellon University, manuscript.
- JAKEMAN, A. J. & YOUNG, P. C. (1979, 1984) Recursive Filtering and the Inversion of Ill-Posed Causal Problems. *Utilitas Math.*, 35, pp. 351-376.
- KALMAN, R. E. (1960) A New Approach to Linear Filtering and Prediction Problems, *ASME Transactions Journal Basic Engineering*, 83-D, pp. 95-108.
- KOLMOGOROV, A. (1941) Interpolation and Extrapolation von Stationären Zufälligen Folgen, *Science Academy Bulletin, URSS, Series Mat.*, 5, pp. 3-14.
- KOOPMAN, S. J. (1993) Disturbance smoother for state-space models, *Biometrika*, 76, pp. 65-79.
- MARAVALL, A. (1993) Stochastic Linear Trends. Models and Estimators, *Journal of Econometrics*, 56, pp. 5-37.
- MARAVALL, A. & GÓMEZ, V. (1992) Signal Extraction in Arima Time Series. EUI Working Paper ECO no. 92/65. Department of Economics, European University Institute, Florence.
- MARAVALL, A. & GÓMEZ, V. (1994) The program SEATS, Signal Extraction in Arima Time Series, instruction for the user. EUI Working Paper ECO no. 94/28. Department of Economics, European University Institute, Florence.
- MELIS, F. (1989) Sobre la Hipótesis de Componentes y la Extracción de la Señal de Coyuntura sin Previa Desestacionalización, *Revista Española de Economía*, 6, pp. 131-163.
- NERLOVE, M., GREYER, D. M. & CARVALHO, J. L. (1979) *Analysis of Economic Time Series: a Synthesis* (New York, Academic Press).
- NG, C. N. & YOUNG, P. C. (1990) Recursive Estimation and Forecasting of Nonstationary Time Series, *Journal of Forecasting*, 9, pp. 173-204.
- NORTON, J. P. (1975) Optimal smoothing in the identification of linear time-varying systems, *Proceedings Institute Electrical Engineers*, 122, pp. 663-668.
- NORTON, J. P. (1986) *An Introduction to Identification* (Academis Press).
- PEDREGAL, D. J. (1995) Comparación Teórica, Estructural y Predictiva de Modelos de Componentes no Observables y Extesiones del Modelo de Young. Phd Thesis. Universidad Autónoma de Madrid.

- PEDREGAL, D. J. & YOUNG, P. C. (1996) Modulated Cycles, a New Approach to Modelling Seasonal/Cyclical Behaviour in Unobserved Component Models, Centre for Research on Environmental Systems and Statistics (CRES), Tech. Note No. TR??. To be submitted for publication.
- RUDEMO, M. (1982) Empirical Choice of Histogram and Kernel Density Estimators, *Scandinavian Journal of Statistics*, 9, pp. 65-78.
- SCHWEPPE, F. (1965) Evaluation of likelihood function for Gaussian signals, *I.E.E.E. Transaction Information Theory*, 11, pp. 61-70.
- SHISKIN, J., YOUNG, A. & MUSGRAVE, J. C. (1967) The X-11 Variant of the Census Method II Seasonal Adjustment Program. Technical Paper 15, Bureau of the Census, Washington.
- SIMS, C. (1974) Seasonality in Regression, *Journal of the American Statistical Association*, 69, pp. 618-626.
- WAHBA, G. (1990) *Spline Models for Observational Data* (Philadelphia, SIAM).
- WALLIS, K. F. (1993) Henderson detrending, symmetric and non-symmetric, reduces I(4) series to stationarity. Department of Economics, University of Warwick.
- WAND, M. & JONES, C. (1995) *Kernel Smoothing* (London, Chapman and Hall).
- WEST, M. AND HARRISON, J. (1989) *Bayesian Forecasting and Dynamic Models*, (New York, Springer-Verlag).
- WHITTLE, P. (1963) *Prediction and Regulation*, (London, English Universities press). 2nd edn. (revised 1983) (Oxford, Blackwell).
- WIENER, N. (1949) *Extrapolation, Interpolation and Smoothing of Stationary Time Series* (New York, J. Wiley).
- YOUNG, P. C. (1984) *Recursive Estimation and Time-Series Analysis* (Berlin, Springer-Verlag).
- YOUNG, P. C. (1988) Recursive Extrapolation, Interpolation and Smoothing of Nonstationary Time Series. In: C.F. CHEN (Ed.), *Identification and System Parameter Estimation* , pp. 33-44 (Oxford, Pergamon Press).
- YOUNG, P. C. (1991) Comments on likelihood and cost as path integrals, *Journal Royal Statistical Society, Series B.*, 53, pp. 529-531.

- YOUNG, P. C. (1993) Fixed Interval Smoothing, Regularisation and Henderson Detrending. Centre for Research on Environmental Systems and Statistics (CRES), Tech. Report No. TR97.
- YOUNG, P. C. (1994) Time-variable Parameter and Trend Estimation in Non-stationary Economic Time Series, *Journal of Forecasting*, 13, pp. 179-210.
- YOUNG, P. C. AND BENNER, S. (1991) *microCAPTAIN Handbook: Version 2.0*. Centre for Research on Environmental Systems and Statistics, Lancaster University
- YOUNG, P. C. & PEDREGAL, D. J. (1996) A comment on 'Effects of the Hodrick-Prescott filter on trend and difference stationary time series. Implications for business cycle research' by Cogley and Nason, submitted for publication to the *Journal of Economic Dynamics & Control*.
- YOUNG, P. C. & PEDREGAL, D. J. (1996) Recursive Fixed Interval Smoothing and the Evaluation of LIDAR Measurements: Comment on the Paper by Holst et al. *Environmetrics*, 7, pp. 417-427.
- YOUNG, P. C., TYCH, W. & PEDREGAL, D. J. (1996) Dynamic Harmonic Regression. Centre for Research on Environmental Systems and Statistics (CRES), Tech. Note No. TR96. To be submitted for publication.
- YOUNG, P. C., NG, C. N. & ARMITAGE, P. (1989) A Systems Approach to Economic Forecasting and Seasonal Adjustment, *International Journal on Computers and Mathematics with Applications*, special issue on System Theoretic Methods in Economic Modelling, 18, pp. 481-501.
- YOUNG, P. C., LANE, K., NG, C. N. & PARKER, D. (1991) Recursive Forecasting, Smoothing and Seasonal Adjustment of Nonstationary Environmental Data, *Journal of Forecasting*, 10, pp. 57-89.

FIGURES

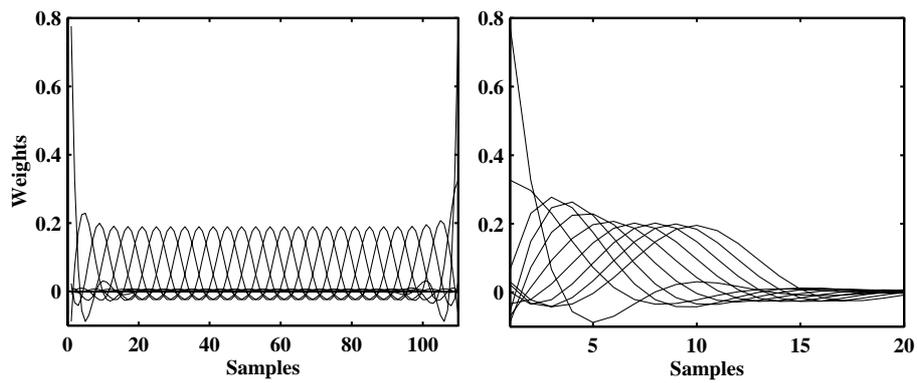


Figure 1. The W matrix (left) and an enlarged version of the left hand side (right).

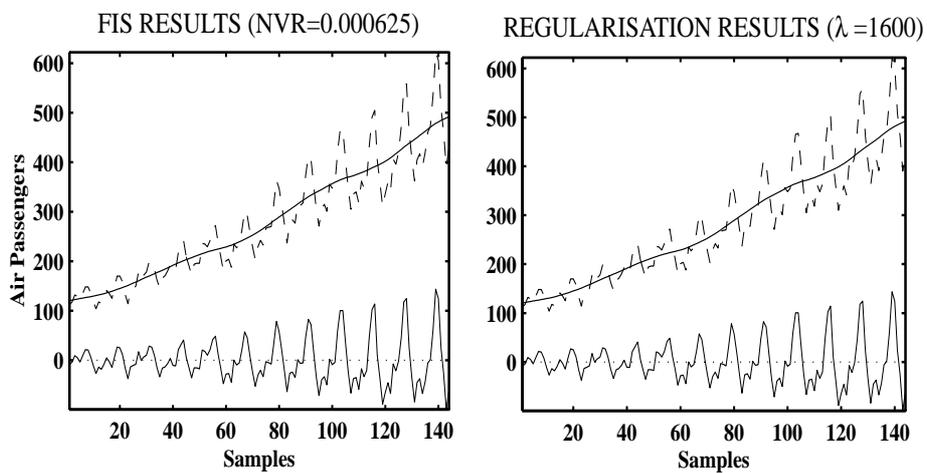


Figure 2. Trend and perturbations (detrended component) of the airline passenger data using the FIS algorithm (IRWSMOOTH; left) and DR algorithm (HP filter; right) with the same smoothing parameter.

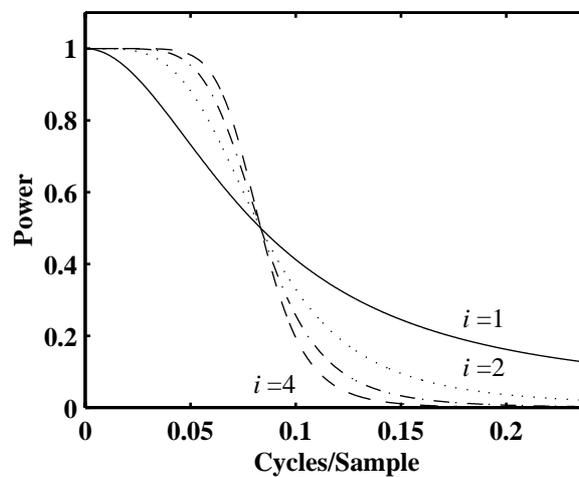


Figure 3. Spectral characteristic of the RW family of smoothing filters.

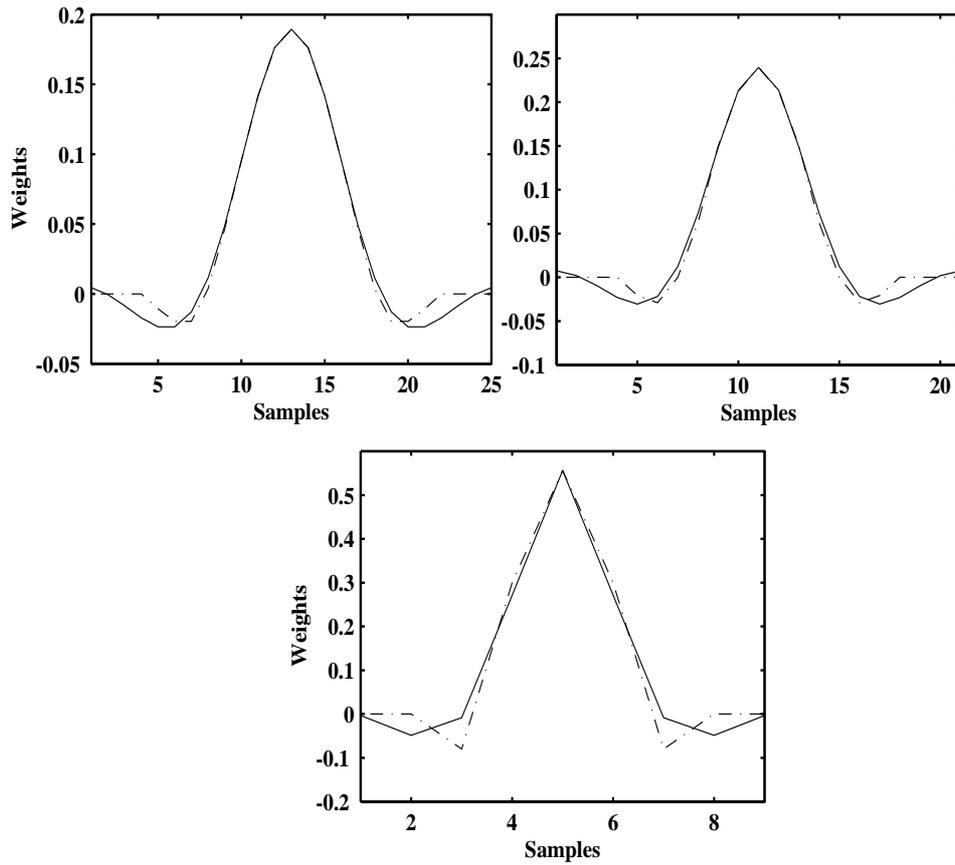


Figure 4. Henderson filter weighting patterns for 17, 13, and 5 terms (dash-dot) compared with the implicit weighting of the equivalent TIRW filter (solid).

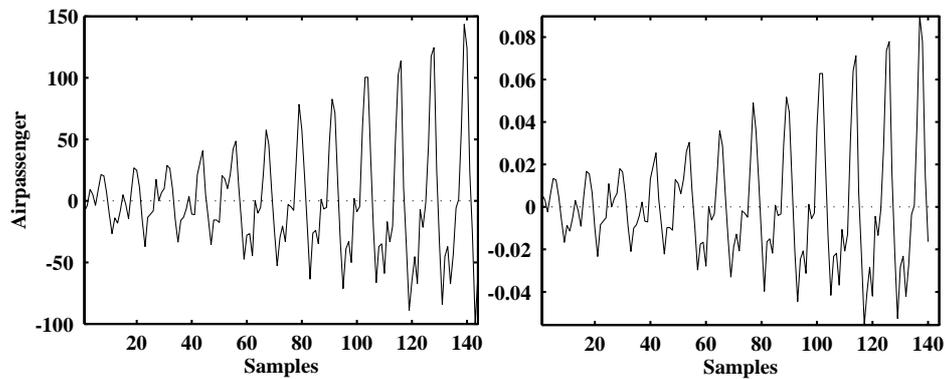


Figure 5. Estimated perturbations of the airline passenger data around the IRWSMOOTH trend (left plot) compared with the fourth difference of the trend (right) for $NVR=0.000625$: if the latter plot is lagged by two samples and multiplied by the NVR parameter, then it is *exactly* equal to the perturbation curve in the left hand plot.

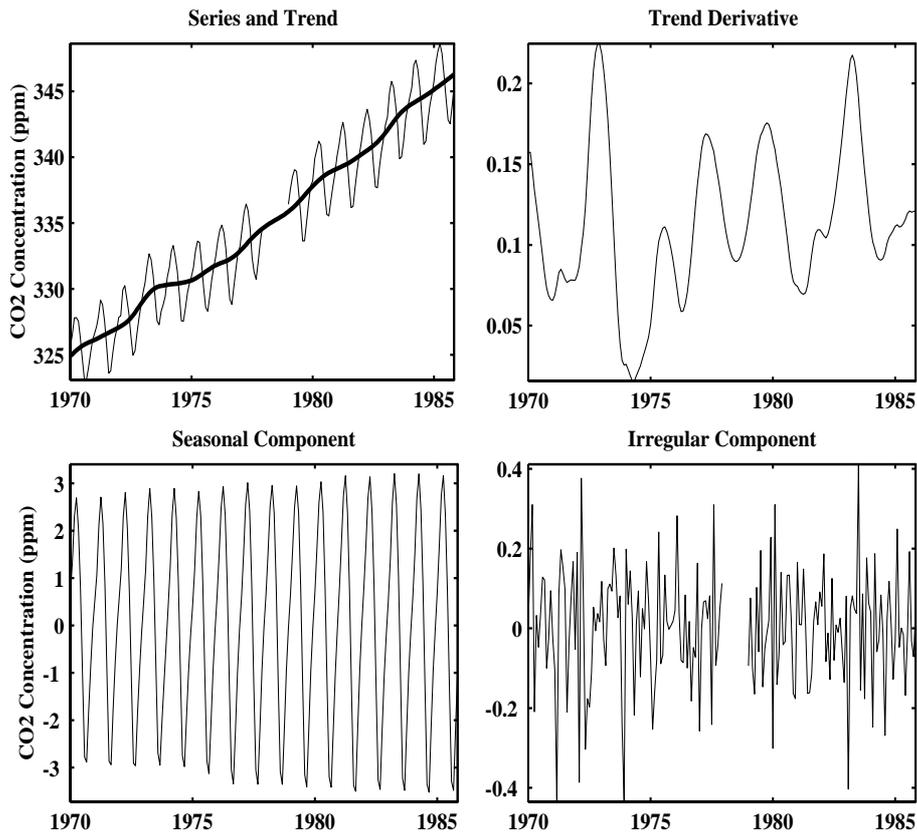


Figure 6. DHR estimated components of the monthly atmospheric CO₂ concentration time series from 1970 to 1985 (note the automatic interpolation of missing data, see text).