Trend Models for the Prediction of Economic Cycles

DIEGO J. PEDREGAL, *Escuela Técnica Superior de Ingenieros Industriales, Universidad de Castilla-La Mancha, Spain.*

SUMMARY *Many different approaches have been proposed to deal with the signal extraction problem in general. In line with this problem, trend estimation has also received a great deal of attention in the time series literature, especially when the interest is focused on forecasting turning points. In spite of all the differences among methods, one common feature remains in most of them. This is that trends tend to extrapolate themselves into the future as a line with a slope that depends on the recent past information. Although this is an optimal (e.g. in a Mean Square Error sense) and a sensible way to do it, it can be systematically erroneous when turning points are at hand. Nor those trend changes could be detected. In those situations, the main source of forecast errors is due to the trend. In this paper two linear trend models with a non-linear like forecast function are explored, namely the Smoothed Random Walk and the Double Integrated Autoregressive model. A combination of two frequency domain methods are explored as the procedure for the identification and estimation of these models. The trend models are compared with standard ones and its forecast performance tested on several real time series examples.*

KEY WORDS Generalised Random Walk, Kalman Filter, Fixed Interval Smoothing, Noise Variance Ratio, Unobserved Components Models, State Space Models.

1. INTRODUCTION

It is difficult to overemphasise the importance of the trend concept and all related to it in both theoretical and practical grounds. In our modern societies most people are aware of the importance of concepts like recession, expansion, economic cycle, turning points, etc.

From a formal point of view, the Unobserved Components Models (UCM) offers a natural framework to deal with all these important concepts. Although there are a great number of different techniques, two of them have remained as the main streams in the literature of UCM, at least among economists.

Firstly, *Structural Methods* propose *a priori* models for the components that posses the properties assumed for each of them. The estimation is then performed by

Correspondence: Diego J. Pedregal, Escuela Técnica Superior de Ingenieros Industriales, University of Castilla-La Mancha, Edificio Politécnica, campus universitario s/n, 13071 Ciudad Real, Spain. Tel: +34 902 204 100 3811; Fax: +34 926 295361; E-mail: pedregal@uclm.es

smoothing the series by using the Kalman Filter and Fixed Interval Smoothing algorithms in a State Space framework (see e.g. Harrison and Stevens, 1976; Nerlove *et al*., 1979; Gersch and Kitagawa, 1983, 1984; West and Harrison, 1989; Harvey, 1989; Ng and Young, 1990; Young *et al*., 1999).

Secondly, the *Reduced Form Methods* begin with the assumption that the series follows an ARIMA model (reduced form) and find the structural models for the components by a process of identification. Since there exist an infinite number of structural forms compatible with each given reduced form, such identification lies on the imposition of a number of (arbitrary) constrains to ensure existence and uniqueness of the decomposition. Many different ways to achieve this goal have been proposed. The most popular is the 'canonical decomposition' (Box *et al*., 1978; Hillmer and Tiao, 1982; Hillmer *et al*., 1983; Burman, 1980; Gómez and Maravall, 1998).

Despite all the differences between these two approaches one common characteristic remains in most of the practical implementations of them. It is that the eventual forecast function of trends is linear. It may be an optimal (e.g. in a Mean Squared Error sense) and sensible way to do it, but it can be systematically erroneous when turning points are at hand. At least, other trend models could outperform standard ones. What is more important is that those turning points would be detected much later with usual trend models. In those situations the main source of forecast errors is due to the trend.

In that regard, it would be very useful to use models that, though linear in mathematical terms, have non-linear like forecast eventual functions that allow a linear extrapolative behaviour as a particular case. Such models would be ideal for the prediction of the Economic Cycle, especially in those time series so common in Economics where the length of the cycles is not constant over time. These models are supposed to outperform models that incorporate explicitly an Economic Cycle with constant length. Pedregal (2001) shows how UC models may accommodate naturally constant length cycles and important improvements in forecasting terms are achieved. That reference also proposes one possibility to model non-constant length Economic Cycles, namely the Hodrick-Prescott filter (Hodrick and Prescott, 1980, 1997) or its equivalent formulation more in line with the methodology used in this paper, the Integrated Random Walk (IRW; a formal demonstration of the relation between these models may be seen in Pedregal, 1995 and Young and Pedregal, 1996). Care should be exercised in handling such filter (see Pedregal and Young, 2001) and Pedregal (2001) emphasises its limitation when forecasts of a time series with a non-constant length cycle is required.

This paper presents one way to overcome the difficulties outlined before. It is achieved by means of an UCM approach that has proven successful for a number of years and two models for the trend with non-linear forecasting properties are explored. These are the Smoothed Random Walk (SRW) and the Double Integrated Autoregressive (DIAR) model. The models are set up in a State Space framework and takes advantages of the exceptional spectral properties of the Dynamic Harmonic Regression (DHR) model. The procedures for the identification and estimation of these models exploited by Ng and Young (1990) and Young *et al*. (1999) is used here. The methodology is illustrated working with some real examples and its forecasting performance is assessed.

2. THE UNOBSERVED COMPONENTS TIME SERIES MODEL

There are many different ways of decomposing a time series in Unobservable Components. A general form widely accepted is

$$y_t = T_t + C_t + S_t + f(\boldsymbol{u}_t) + e_t \qquad (1)$$

where y_t is the observed time series; T_t is a trend or low frequency component; C_t is a sustained cyclical or quasi-cyclical component (e.g. an economic cycle) with period different from that of any seasonality in the data; S_t is a seasonal component (e.g. annual seasonality); $f(\boldsymbol{u}_t)$ captures the influence of exogenous variables; and e_t is an 'irregular' component, normally defined for analytical convenience as a normally distributed Gaussian sequence with zero mean value and variance s^2. In order to allow for nonstationarity in the time series y_t, the various components in the model, including the trend T_t, can be characterised by stochastic, Time Variable Parameters (TVP's), with each TVP defined as a nonstationary stochastic variable. Under certain assumptions a non-linear stochastic system can be approximated by a linear TVP model (see e.g. Young, 1998, 1999).

Not all the components in equation (1) are necessary in every application. For example, it is quite common that the decomposition in trend, seasonal and irregular is good enough for a fair amount of monthly and quarterly economic time series. Sometimes components are aggregated in a new one called in a different way, like the 'cyclical trend' (trend plus cycle). Multivariate formulations of model (1) are also possible using the procedures presented in this paper. However we will restrict the discussion, for simplicity of exposition, to the univariate case.

In the UC framework, equation (1) is considered as the observation equation in a discrete-time Non Minimal State Space (NMSS) model, which describes the stochastic evolution of state variables, associated with the UC's in (1). The SS form is an extraordinarily powerful and flexible tool for the time series analysis. It is also a framework in which the UC type of models fit quite naturally. In spite of this, there are other alternative formulations, like the Wiener Kolmogorov (WK) classical filter (Whittle, 1983; Bell, 1984), which, though it has advantages as an analytical tool (see e.g. Pedregal, 1995), its practical implementation is rather much more complex.

In the SS formulation, the SS description of the whole UC model is synthesised by assembling all the individual SS forms of all the components (see e.g. Harvey, 1989; Ng and Young, 1990; Young *et al*., 1999). Therefore, in order to formulate this overall SS form of the model, specific assumptions about the statistical nature of every component have to be made. The adequacy of these assumptions may be checked afterwards, by standard testing procedures.

Trend models

There are a number of possibilities available in the literature. One general formulation commonly used is the Generalised Random Walk (GRW). Less common but very useful models in practice are also possible, like the Double Integrated Autoregressive trend (DIAR, see e.g. Young, 1994).

1. GRW: this trend is one of the simplest stochastic representations of trend components. The form is

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}_t = \begin{pmatrix} a & b \\ 0 & g \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}_{t-1} + \begin{pmatrix} h_1 \\ h_2 \end{pmatrix}_t \qquad T_t = \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}_t \qquad (2)$$

Here, a, b, and g are constant parameters; T_t is the trend component (i.e. the first state, x_{1t}); x_{2t} is a second state variable (generally known as the 'slope'); h_{1t} and h_{2t} are zero mean, serially uncorrelated white noises with constant block diagonal covariance matrix $\boldsymbol{Q}$.

This model includes, as special cases, the Random Walk (RW: $a = 1$; $b = g = 0$; $h_{2t} = 0$); Smoothed Random Walk ($0 < a < 1$; $b = g = 1$; $h_{1t} = 0$); the Integrated Random Walk (IRW: $a = b = g = 1$; $h_{1t} = 0$); the Local Linear Trend (LLT: $a = b = g = 1$); and the Damped Trend ($a = b = 1$; $0 < g < 1$). In the case of the IRW, x_{1t} and x_{2t} can be interpreted as level and slope variables associated with the variations of the trend, with the random disturbance entering only via the x_{2t} equation.

The IRW model has been used with success during a long period (see e.g. Young *et al*., 1989; Young, 1994; Ng and Young, 1990; Young *et al*., 1999) and is particularly useful for describing large smooth changes in the trend or TVP; the RW model provides less smooth variations; and the SRW allows for a whole range of intermediate behaviour between IRW and RW extremes depending on the value of the parameter a ($a = 0$ for RW and $a = 1$ for IRW). It is also well known that LLT is in general less smooth than IRW, given the fact that the former contains one additional source of noise than the latter.

The forecasting functions of these models (RW, IRW, SRW and LLT) are, however, quite different, as it is illustrated in figure 1. The RW model predictions are a horizontal line fixed at the last observation before the forecast origin; the IRW (and LLT) extrapolates in a linear fashion with the slope fixed at its level in the forecast origin; the SRW provides again intermediate possibilities in a non-linear fashion. It is remarkable that the SRW model produces a kind of predictions that are in line with the actua1 observed behaviour of many economic series. Neither they tend to grow in an exponential way for long periods of time, nor even as linear trends. That means that the use of SRW trend will be adequate in forecasting terms for a high number of series, or at least in situations close to turning points. Besides, the estimation of the a parameters proposed below can be considered as a formal way of testing the adequacy of IRW vs. RW models.

(FIGURE 1)

2. <u>DIAR</u>: When modelling IRW trends it is assumed that its second difference (h_{2t} in equation (2)) is white noise. While this is true for the theoretical components, it is not the case for the estimated components. It is quite common to find correlation in this residual in practical situations. This raises the interesting point of trying to take advantage of that correlation and build one model for that residual in order to improve the overall forecasting performance. The observation about correlation of such residuals must be considered cautiously since we know that the FIS algorithm induces that kind of correlation, and even it is possible to demonstrate theoretically that its correlation structure depends on the autocorrelation of the original time series itself[1]. However, if

[1] In particular, Pedregal (1995) demonstrates that the fourth difference of the FIS estimated IRW trend is exactly equal to the estimated perturbations, lagged by two samples and re-scaled by a factor that is exactly the ratio of the variances of the noises in the state equation to the observed noise in (2).

the second difference shows some predictable behaviour with some physical meaning (e.g. related to the business cycle), it is worthwhile to try to forecast it.

A straightforward and simple possibility is modelling the second difference series as an AR model, that is

$$
\begin{aligned}
T_t &= T_{t-1} + D_{t-1} \\
D_t &= D_{t-1} + x_{1t} \\
x_{1t} &= \frac{h_{3t}}{1 + f_1 B + f_2 B^2 + \ldots + f_p B^p}
\end{aligned}
$$

where B is the backward shift operator, such that $B^l x_t = x_{t-l}$ and a_t and the observation equation noise are serially uncorrelated white noises. In SS form the model is described by the following equations

$$
\begin{bmatrix} T \\ D \\ \hline x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_p \end{bmatrix}_t =
\left[\begin{array}{cc|cccccc}
1 & 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\
0 & 1 & 1 & 0 & 0 & \cdots & 0 & 0 \\ \hline
0 & 0 & -f_1 & -f_2 & -f_3 & \cdots & -f_{p-1} & -f_p \\
0 & 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & 0 & 0 & \cdots & 1 & 0
\end{array}\right]
\begin{bmatrix} T \\ D \\ \hline x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_p \end{bmatrix}_{t-1} +
\begin{bmatrix} 0 \\ 0 \\ \hline h_3 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}_t
$$

$$
T_t = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix} \boldsymbol{x}_t \qquad \text{with } \boldsymbol{x}_t = \begin{bmatrix} T & D \mid x_{1t} & x_{2t} & \cdots & x_{pt} \end{bmatrix}^T
$$

The model is then fully defined by the variance of h_{3t} and the coefficients of the AR polynomial. This is a much more complex model than GRW, but it has the capability of providing non-linear like forecasts of the trend, very useful in situations near turning points.

Cyclical and Seasonal component models

Although these two types of components are named differently, both can be treated in the same way from a modelling standpoint, since both reflect a periodic type of behaviour. The difference between them lies only on the period considered, longer than one year for cyclical components, while ‘seasonal’ is often reserved just for the annual cycle.

Many possibilities are available. The most common are:

1. General Transfer Function (GTF) model (Ng and Young, 1990): In this option, the periodic components are represented by a GTF model, similar to the well-known ARMA model employed in Box-Jenkins forecasting (Box and Jenkins, 1970; 1976; Box *et al*., 1994), but the term ‘general’ is used to show that no prior stationarity restrictions are assumed. This may create theoretical problems, but their application in practice has proven very useful in past years and has been implemented in the CAPTAIN package[2].

[2] Information about this software is available in http://cres1.lancs.ac.uk/captain/ and a beta-test version is available from the author.

Obviously, care should be taken in evaluating the results obtained when the models are used for forecasting over very long periods.

Similar type of models is used in some ‘reduced form’ methods (see e.g. Gómez and Maravall, 1998). There is, however, a fundamental difference: the GTF model proposed here is identified and estimated directly from the data; while in ‘reduced-form’ methods, the ARIMA mode1 for the time series is estimated from the data, and the separate models for the components are identified (not estimated) from this with the help of the e.g. canonical decomposition.

In order to consider the SS form, it is convenient to assume that the sum of the periodic component and the irregular component constitutes and ARMA process with the same white noise input, i.e.

$$S_t + e_t = \frac{J(B)}{f(B)} e_t$$

where $J(B) = \left(1 + J_1 B + J_2 B^2 + \ldots + J_p B^p\right)$ and $f(B) = \left(1 + f_1 B + f_2 B^2 + \ldots + f_p B^p\right)$ are both polynomials in the backward shift operator of the same order[3]. One SS form of such a model is

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix}_t = \begin{bmatrix} -f_1 & 1 & 0 & \cdots & 0 \\ -f_2 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -f_{p-1} & 0 & 0 & \cdots & 1 \\ -f_p & 0 & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix}_{t-1} + \begin{bmatrix} J_1 - f_1 \\ J_2 - f_2 \\ \vdots \\ J_p - f_p \end{bmatrix} e_t$$

$$S_t + e_t = x_{1t}$$

2. Dummy seasonality (Harvey, 1989): The form of this model is

$$\sum_{j=0}^{s-1} S_{t-j} = h_t$$

where s is the seasonal period and h_t is white noise. This formulation follows from the standard dummy variable methods of modelling a fixed seasonal pattern in regression, i.e.

$$S_t = \sum_{j=1}^{s} b_j D_{jt} + h_t \qquad D_{jt} = \begin{cases} 1, t = j,\ j+s,\ j+2s, \ldots \\ 0, t \neq j,\ j+s,\ j+2s, \ldots \\ \quad -1, t = s, 2s, 3s, \ldots \end{cases} \quad j = 1, 2, \ldots, s-1$$

where b_j are constant parameters and D_{jt} are s dummy variables defined above. This model can be done more flexible by replacing the constant parameters by TVP, for which the pattern of variation may be, once more, GRW models.

The Dummy variable seasonality may be seen as a constrained version of the previous GTF model, in which the MA order is zero, the AR order is $s-1$ and all the

[3] Note that this assumption does not constrain the model in any way, because any trailing coefficients could be assumed zero. In the same way, specification of ‘subset’ ARMA models are possible.

AR parameters are constrained to one. It is also equivalent to the set of trigonometric terms at seasonal frequencies, i.e.

$$S_t = \sum_{f=1}^{[s/2]} \left[a_f \cos(w_f t) + b_f \sin(w_f t) \right] \qquad w_f = 2pf/s,\ f = 1, \dots, [s/2]$$

being then a particular case of the following model.

3. Dynamic Harmonic Regression (DHR) (Ng and Young, 1990; Young *et al*. 1999): This option is represented by the following state and observation equations:

$$\begin{pmatrix} a_f \\ a^*_f \end{pmatrix}_t = \begin{pmatrix} a_f & b_f \\ 0 & g_f \end{pmatrix} \begin{pmatrix} a_f \\ a^*_f \end{pmatrix}_{t-1} + \begin{pmatrix} h_f \\ h^*_f \end{pmatrix}_t$$

$$S_t = \sum_{f=1}^{R} \left[a_{ft} \cos(w_f t) + b_{ft} \sin(w_f t) \right]$$

and b_{ft} parameters are defined in the same way as a_{ft}.

In other words, the DHR model has a linear regression form with deterministic periodic functions of time as inputs with TVP that follow GRW models, in principle. The rigidity introduced by the deterministic functions is compensated by the TVP giving the model a great flexibility, capable of represent many types of different seasonal patterns. This is extremely useful for series that exhibit the kind of non-stationarity behaviour in the seasonal component so commonly observed in economic time series.

4. Trigonometric Cycle or Seasonal (Harvey 1989; West and Harrison, 1989): The state representation of the model is

$$\begin{pmatrix} a_f \\ a^*_f \end{pmatrix}_t = r \begin{pmatrix} \cos(w_f t) & \sin(w_f t) \\ -\sin(w_f t) & \cos(w_f t) \end{pmatrix} \begin{pmatrix} a_f \\ a^*_f \end{pmatrix}_{t-1} + \begin{pmatrix} h_f \\ h_f \end{pmatrix}_t \qquad (3)$$

$$S_t = \sum_{f=1}^{R} a_{ft}$$

in which the state disturbances are the same in both equations. This is a usual *a priori* constraint, in order to avoid the proliferation of parameters in the model. The cyclical behaviour is introduced via the state equation, while the observation equation simply add up all the states defining the periodic component.

5. Modulated Cycle or Seasonal (Pedregal and Young, 1996): As can be seen in the previous paragraphs, different formulation of periodic components models are possible, including the periodic behaviour either in the state equations or in the observation equation. Such duality allows for a mixture of periodic behaviours by combining state and observation equations that have periodic components for different periods simultaneously.

For example, mixing up state equation in (3) and a DHR model as the observation equation, we have the modulated cycle model for one single frequency in the observation equation f_1, i.e.

$$\begin{pmatrix} a_f \\ a_f^* \end{pmatrix}_t = r_f \begin{pmatrix} \cos(w_f t) & \sin(w_f t) \\ -\sin(w_f t) & \cos(w_f t) \end{pmatrix} \begin{pmatrix} a_f \\ a_f^* \end{pmatrix}_{t-1} + \begin{pmatrix} h_f \\ h_f \end{pmatrix}_t \qquad (4)$$

$$S_t = \sum_{f=1}^{R} \left[a_{ft} \cos(w_{f1} t) + b_{ft} \sin(w_{f1} t) \right]$$

where any $w_f \neq w_{f1}$. The model can be expanded naturally for several frequencies in the observation equation in an obvious way. This model introduces multiplicative cycles since in the observation equation periodic functions of a given frequency are multiplied by parameters that are themselves periodic functions of a different frequency. This is known as 'modulated signals' in the signal processing literature. The peculiarity of this new model is that the amplitude of the shortest period cycle is modulated by the longest one giving a different pattern for each realisation of the short period cycle.

Exogenous Variables Component Model

This part of the model allows for the possibility that the variable is affected by other time series $\boldsymbol{u}_t = [u_{1t}, u_{2t}, \ldots, u_{nt}]^T$. These variables are called 'inputs', 'explanatory variables', 'leading indicators', depending on the context. If the input variables are of the type of dummy variables, then intervention analysis can be carried out.

As it happens with the rest of components, several possibilities exist, from which we will highlight three for them:

1. <u>Dynamic Linear Regression (DLR)</u>: Here all the variables enter in a linear form, but with TVP following some GRW scheme of variation, i.e.

$$f(\boldsymbol{u}_t) = b_{1t} u_{1t} + b_{2t} u_{2t} + \ldots + b_{nt} u_{nt}$$

2. <u>Multiple Transfer Function</u>:

$$f(\boldsymbol{u}_t) = \frac{w_1(B)}{d_1(B)} u_{1t} + \frac{w_2(B)}{d_2(B)} u_{2t} + \ldots + \frac{w_n(B)}{d_n(B)} u_{nt}$$

where typically

$$w_i(B) = w_{0i} + w_{1i} B + \ldots + w_{qi} B^q$$

$$d_i(B) = 1 + d_{1i} B + \ldots + d_{pi} B^p$$

and parameters may be varying in time, if necessary.

3. <u>Non-linear Relationships</u>:

$$f(\boldsymbol{u}_t) = \Im(u_{1t}, u_{2t}, \ldots, u_{nt})$$

where $\Im(\bullet)$ is a non-linear function of the exogenous variables (and possible the other state variables). To define the properties of the non-linearity, theory could be used or complementarily, *Data Based Mechanistic* approaches (Young, 1993, 1998, 1999; Young and Pedregal, 1997, 1999).

3. IDENTIFICATION AND ESTIMATION OF UC MODELS

Having defined SS structures for all the components of the UC model, it is straightforward to assemble them in an aggregate, non-minimal, SS model described by a state and observation equations (see e.g. Harvey, 1989; Young *et al*., 1999).

Given such form for the overall model, the *Kalman Filter* (Kalman, 1960) and the associate recursive *Fixed Interva1 Smoothing* algorithm (Bryson and Ho, 1969; Norton, 1986; Young, 1988), provide the basis for forecasting, interpolating and smoothing (estimate the unobserved components). For a data set of N samples, the former algorithm yields a 'filtered' estimate of the state vector at every sample t, based on the time series data up to sample t. The later produces a 'smoothed' estimate of the states which, at every sample t, is based on all N samples of the data. This means that, as more information is used in the later estimate, its *Mean Square Error* cannot be greater than the former. As these algorithms are discussed in detail in the previous references, we will not pursue the topic further.

Previous to the application of this algorithms, all parameters on which the model depend on, including the variances of all the noises, must be known or estimated in some way. In the present approach, a 'noise variance ratio' matrix (NVR) is estimated, instead of the variances themselves. This NVR matrix is defined as the ratio of the state noise covariance matrix to the variance of the observation equation noise i.e. $\boldsymbol{NVR} = \boldsymbol{Q}/s^{2}$. Since $\boldsymbol{Q}$ is constrained to be diagonal, it is common to speak about the *NVR* parameters, referring to the diagonal elements of the $\boldsymbol{NVR}$ matrix.

Estimation of hyper-parameters

It seems clear that the problems of identification and subsequent parameter estimation for the complete state space form are non-trivial. A usua1 way to deal with the problem is to formulate it in Maximum Likelihood (ML) terms (e.g. Harvey, 1989; Durbin and Koopman, 2001). Assuming that all the disturbances in the state space form are norma1ly distributed, the ML function can be computed using the *Kalman Filter* via 'prediction error decomposition'. This is the generally accepted method, because of his well-known theoretical basis. However, the optimisation can be very complex even for relative simple models due to the flatness around the optimum (Young *et al*., 1999). There are several alternatives to ML, from which we will explore two: Sequential Spectral Decomposition and an alternative optimisation in frequency domain. Both methods are based on the spectral properties of the models proposed.

1. <u>Sequential Spectral Decomposition (Young, 1988; Ng and Young, 1990)</u>: This approach consists of the decomposition of the original series in quasi-orthogonal components, taking advantage of the exceptiona1 spectral properties of the smoothing algorithms mentioned above. The overall non-linear problem is decomposed in several linear or quasi-linear steps, each solved in fully recursive terms. This gives a simple solution, with some loss from the ML viewpoint, but has proven to be very successful in practise. As a final step, filtering and smoothing are repeated using the whole SS formulation based on the analysis done in the previous steps.

2. <u>Model Optimisation in the Frequency Domain</u>: The parameters are estimated so that the logarithm of the model spectrum fits the logarithm of the empirical pseudo-spectrum (either an AR-spectrum or periodogram) in a least squares sense. A full

description of this algorithm can be found in Young *et al*. (1999). Here a summary is reported.

Let's consider a DHR model given by

$$y_t = \sum_{f=1}^{R} \left[a_{ft} \cos\left(w_f t\right) + b_{ft} \sin\left(w_f t\right)\right] + e_t \tag{5}$$

where each of the parameters a_{ft}, b_{ft} are modelled as GRW or Trigonometric Equation Cycles.

The model spectrum then can be built in two steps: (a) derivation of the spectrum of the TVP models and (b) derivation of the spectrum of the sinusoidal components modulated by the TVP.

In order to derive the spectrum of the TVP models it is necessary to obtain first the Transfer Function (TF) form (or 'reduced form ') of its SS description. For example, in the case of a SRW model (equation (2) with $0 < a < 1$; $b = g = 1$; $h_{1t} = 0$) the TF form is given by

$$y_t = \frac{h_{2t}}{(1-B)(1-aB)} + e_t$$

For this process the power pseudo-spectrum[4] can be found as

$$f_y(w) = \frac{1}{2p}\left[\frac{s^2_{h_2}}{(2-2\cos w)(1+a^2-2a\cos w)} + s^2\right]; \quad w \in [0,p] \tag{6}$$

Spectra for other cases in the GRW family of models can be found just constraining the a parameter (e.g. 0 or 1 for RW or IRW models, respectively). In figure 2 the spectral characteristic of this type of filters is represented for different values of the a parameter at frequencies 0 (trend) and 4 cycles/time units. There, it can be seen that this parameter tunes the width of the spectral band defined by the model between the extremes models RW and IRW.

From the basic Fourier transform properties, the frequency response of amplitude modulated signals of the form $S_t = a_t \cos\left(w_j t\right)$, is known to be:

$$f_s(w) = \frac{1}{2}\left[f_A\left(w - w_j\right) + f_A\left(w + w_j\right)\right] \tag{7}$$

where $f_A(w)$ is the frequency response of a_t. Let's consider the case of a single frequency DHR term, $S_t = a_t \cos(w_j t) + b_t \sin(w_j t)$ in which the TVP associated with the sine and cosine terms are modelled as two IRW processes with equal variance parameters ($s^2_{w_j}$). The pseudo-spectrum of S_t may be found by combining equations (6) for the IRW case and (7), i.e.

$$f_{w_j}(w) = \frac{1}{2p}\left[\frac{s^2_{w_j}}{4(1-\cos(w-w_j))^2} + \frac{s^2_{w_j}}{4(1-\cos(w+w_j))^2}\right].$$

(FIGURE 2)

---

4 'Pseudo' because the process is non-stationary (see Harvey, 1989).

Defining

$$S(w, w_j) = \left[\frac{s^2_{w_j}}{4(1-\cos(w - w_j))^2} + \frac{s^2_{w_j}}{4(1-\cos(w + w_j))^2} \right]$$

the spectrum for the whole DHR model is then formed by the addition of the spectra of the individual terms (fundamental frequency and harmonics), and it can also be expressed in terms of the NVR parameters $NVR_f = s^2_{w_f} / s^2$, i.e.

$$f_y^*(w, \boldsymbol{NVR}) = \frac{s^2}{2p} \left\{ \sum_{f=0}^{R} NVR_{w_f} S(w, w_f) + 1 \right\} \qquad \boldsymbol{NVR} = \left\lfloor NVR_{w_0} \quad NVR_{w_1} \quad \dots \quad NVR_{w_R} \right\rfloor$$

The problem then is to find the set of parameters $\boldsymbol{NVR}$ (and all the rest of possible *hyper-parameters* present in the model, like a 's in SRW models) that give the optimal fit to the empirical estimated spectra. Although a linear least-squares fit is the most obvious, substantial advantages can be found when a non-linear objective function is used without increasing the computation burden too much. The proposed cost function is (see Young *et al*., 1999)

$$\Im\left(f_y, \hat{f}_y^*\right) = \sum_{k=0}^{T-1} \left[\log\left\{ f_y(w_k) \right\} - \log\left\{ \hat{f}_y^*(w_k, \boldsymbol{NVR}) \right\} \right]^2$$

Here, $f_y(w_k)$ is either the sample periodogram or AR spectrum of the time series. Using the log-transformed spectra yields much better defined NVR estimates since it concentrates attention on the most important shape of the 'shoulders' associated with the harmonic peaks in the AR Spectrum. The linear solution, however, constitutes a good start point for this non-linear optimisation.

There exist practical advantages in matching the AR spectrum and this is the approach used in the CAPTAIN program (see footnote number 2), where DHR model estimation is an option for use in forecasting and seasonal adjustment. One of the most interesting advantages of the AR spectrum is the similarity it presents with the DHR-type periodic functions, making this option more attractive than the periodogram.

With respect to identification and estimation of DIAR trends, especial considerations are necessary. Following the developments in this section, it is straightforward to extend the optimisation algorithm to deal with the optimisations of the *NVR* parameters, a 's (for SRW components), and f 's of the AR polynomial all together.

This would be the full optimal solution. However, the AR order for the trend model should be identified in some way previous to its estimation, and a high number of parameters are involved in such a model. These two facts make advisable the use of a procedure in several steps, in which both the Sequential Spectral Decomposition and the automatic Estimation in the Frequency Domain outlined above are combined. These steps are:

1. Estimation of the $\boldsymbol{NVR}$ vector of parameters in a DHR model, using GRW or State Equation Cycle type of parameters and an IRW model for the trend.
2. Obtain the components based on the estimates in step 1 and find state equation residuals by

$$\hat{h}_{2t} = \frac{1}{(1-B)^2}\hat{T}_{t|N}$$

where $\hat{T}_{t|N}$ is the estimated (smoothed) trend.

3. Identify and estimate an AR model for $\hat{h}_{2t}$ using the AIC criterion or any other standard identification tool. Estimate and carry out adequacy tests on residuals.
4. Given the $\boldsymbol{NVR}$ set in step 1, and f's in step 3, build the final state space model and use it to forecast, backcast, interpolate as required. Do final adequacy tests.

Steps 1 and 2 could be replaced by other kind of trends creating different trend models. As an example, a Smoothed AR trend could be easily built just using a SRW trend in step 1 and replacing the estimated residuals in 2 by

$$\hat{h}_{2t} = \frac{1}{(1-\hat{a}B)(1-B)}\hat{T}_{t|N}$$

where $\hat{a}$ was estimated in step 1.

4. EMPIRICAL EXAMPLES

Figure 3 shows two quarterly energy demand variables in the UK (taken from Harvey, 1989) and the monthly Industrial Production Index (IPI) of both the American and the British economies. Vertical lines in the plots indicate some of the forecast origins selected later on for the evaluation of forecast performance. It is assumed that all the information previous to such points is known and used to estimate the models. The values after the lines are assumed unknown and used to evaluate the forecast performance. It is worth noting that situations of change of trend have been selected on purpose just to proof whether the new trend models are able to detect and forecast those changes correctly.

Three different trend models were estimated for each series, IRW, SRW and DIAR. In all cases the identification and estimation procedures discussed in previous sections have been used.

AR spectrum orders were selected using the AIC. AR spectra showed systematically a trend component plus a seasonal one formed by the main period (4 samples/cycle for quarterly series and 12 for monthly ones) and all its harmonics (2 s/c in quarterly data and 6, 4, 3, 2.4 and 2 s/c in monthly data).

(FIGURE 3)

When estimating IRW trends, the models for all the rest of components were IRW. On the other hand, SRW models were used for all the components when the SRW trend was estimated. This change does not make any important difference in general since the estimated a's for the seasonal components tend to be very close to 1. This seems logical, since many series show a growing or sustained amplitude seasonal pattern rather than a damped one, and that behaviour is replicated by IRW seasonal models. The estimation results are summarised in tables 1 and 2.

IRW trends were used afterwards for the DIAR identification, again using AIC. The orders selected appear in the last row of table 1. These are low enough to ensure parsimony but high enough to get all the correlation information in the series. Autocorrelation functions indicated adequacy of the models chosen.

Period of Component	Coal	Gas	IPI (USA)	IPI (UK)
Trend	1.66e-3	4.90e-4	6.99e-3	8.29e-3
12			4.37e-2	2.12e-2
6			8.67e-3	4.05e-3
4	1.61e-1	1.25e-1	4.15e-3	4.74e-3
3			3.46e-3	7.88e-3
2.4			2.06e-3	5.97e-3
2	1.38e-2	6.15e-2	6.89e-4	5.89e-3
Objective Function	3.443	3.659	4.715	4.443
AR order of DIAR trends	8	8	12	13

Table 1: Estimated NVR's for IRW-type harmonics.

Period of Component	Coal	Gas	IPI (USA)	IPI (UK)
Trend	1.89e-2 (0.73)	2.69e-3 (0.86)	2.04e-1 (0.81)	4.46e-2 (0.79)
12			4.57e-2 (0.97)	1.67e-2 (0.99)
6			8.02e-3 (1)	3.99e-3 (0.99)
4	1.50e-1 (0.99)	1.18e-1 (0.99)	4.05e-3 (1)	4.68e-3 (0.99)
3			3.42e-3 (1)	7.60e-3 (0.99)
2.4			2.05e-3 (1)	1.23e-2 (0.97)
2	4.35e-2 (0.89)	9.40e-2 (0.93)	6.84e-4 (1)	5.54e-3 (0.99)
Objective Function	1.877	3.326	4.375	4.059

Table 2: Estimated NVR's for SRW-type harmonics (number in parenthesis are the a parameters).

By comparing tables 1 and 2 it is easy to discover that the closer the a 's are to 1 in SRW seasonals, the closer are their respective NVR's to their IRW counterparts. That sounds sensible since the a parameters tend to be one in most cases. The biggest differences appeared in the trend parameter where a are significantly less than one and NVR's are quite distant from the ones in table 1. That means that all the differences in forecasting performance are due almost exclusively to the trend model, not to the seasonal pattern.

This kind of results have been found systematically for a wide range of series, only on a few cases SRW models were necessary for seasonal components. This indicates that simpler models in which SRW are specified for the trend fit the spectrum remarkably well and the number of parameters to estimate is reduced considerably.

As models in table 1 are constrained versions of models in table 2, it is important to underline that the objective functions are always smaller in the unconstrained case meaning that the optimisation algorithm converged to a real optimum and that there is information enough to estimate all the parameters. Since most a 's are close to 1, the fit and forecast improvements for SRW models (see below) are almost exclusively due to the trend models.

Figure 4 shows some forecasting results. Only trend forecasts are included to avoid confusion in the plot. Some important conclusions are drawn from this information. As we stated before, the IRW forecasts (that is, linear-type) are the worst of the three because we selected forecast origins where there are trend changes on purpose. DIAR trends are more risky in general than SRW ones because of the formulation of the model, though sometimes both are very similar and none of SRW or DIAR is systematically the best of all.

(INSERT FIGURE 4)

Time Series	Forecast origin	IRW	SRW	DIAR
Coal	64	20.13	11.08	9.77
	68	26.25	6.81	7.92
	72	14.88	6.53	7.57
Gas	80	1.91	1.62	1.67
	84	1.49	1.10	0.87
	88	1.02	0.59	0.39
IPI (USA)	186	2.97	2.55	2.52
	192	1.80	1.29	1.36
	198	2.88	1.11	0.38
IPI (UK)	198	2.59	0.62	0.86
	204	2.06	0.74	1.59
	210	1.94	1.65	1.76

Table 3: One year ahead MAPE(%) of IRW, SRW and DIAR trend models.

The Mean Absolute Percentage Errors (MAPE) one year ahead of the three types of models estimated at different forecast origins are reported in table 3. Results oscillate from no practical improvement for some forecast origins to quite big ones in cases where the trend change is correctly predicted. In any case, both non-linear trends produce results at least as good as the linear one.

Apart from the numerical accuracy, it is quite important the qualitative information this trends supply as a mean of detecting changes of trend behaviour. In that regard, it is useful to look at all the models at the same time, instead of relying on one of them alone when there is some evidence of a trend change. For example, when all models indicate a trend change, it is quite probable that such change is going to occur.

5. CONCLUSIONS

In this paper we have presented two models for trend components, the Smoothed Random Walk (SRW) and the Double Integrated Autoregressive (DIAR) model in a general framework of an Unobserved Component Model exploited with success during many years (see e.g. Ng and Young, 1990; Young *et al.*, 1999; Pedregal, 2001). The especial feature of these models that makes them attractive is the non-linear type of extrapolative predictions. This allows for a better forecasting performance in situations near turning points.

The approach takes advantage of the flexibility and excellent spectral properties of the Dynamic Harmonic Regression (DHR) model in a State Space set up. The parameters in the models are estimated using the same procedure in Young *et al.*

(1999) for the SRW model, while a mixture of such procedure and a Sequential Spectrum Decomposition is used for DIAR models.

Both types of trend models are tested on real examples in both normal situations and situations where a trend change is occurring. SRW and DIAR models outperform the IRW in all cases, but none of them is the best in all cases from a forecasting point of view.

6. References

Bell, W.R. (1984), 'Signal Extraction for Nonstationary Time Series', Annals of Statistics, 12, 646-664.

Box, G.E.P., and Jenkins, G.M., (1970, 1976), *Time Series Analysis, Forecasting and Control*, Holden-Day: San Francisco.

Box, G.E.P., Jenkins, G.M. and Reinsel, G.C. (1994), *Time Series Analysis, Forecasting and Control*, Prentice Hall International: Englewood Cliffs, New Jersey.

Box, G.E.P., Hillmer, S.C. and Tiao, G.C. (1978), 'Analysis and Modelling of Seasonal Time Series', in A. Zellner (ed.), *Seasonal Analysis of Economic time Series*, Washington D.C.: US Dept. of Commerce-Bureau of the Census, 309-334.

Bryson, A.E. and Ho, Y.C. (1969), *Applied Optimal Control, Optimization, Estimation and Control*, Blaisdell Publising Company: Waltham, Mass.

Burman, J.P. (1980), 'Seasonal Adjustment by Signal Extraction'. *Journal of the Royal Stat. Soc. A*, No.143, part 3, 321-337.

Durbin, J. and Koopman, S.J. (2001), *Time Series Analysis by State Space Methods*, Oxford University Press: New York.

Gersch, W. and Kitagawa, G. (1983), 'The prediction of Time Series with Trends and Seasonalities', *Journal of Buss. & Econ. Statist.*, Vol. 1, No.3, 253-264.

Gersch, W. and Kitagawa, G. (1984), 'A Smoothness Priors State Space Modelling of Time Series with Tren and Seasonality', *Journal of the American Statist. Assoc.*, 79, 378-389.

Gómez, V. and Maravall, A. (1998), *Programs TRAMO and SEATS, Instructions for the User (Beta Version: June 1998)*, Bank of Spain: Madrid.

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

Harvey, A.C. (1989), *Forecasting Structural Time Series Models and the Kalman Filter*, Cambridge University Press: Cambridge.

Hillmer, S.C. and Tiao, G.C. (1982), 'An ARIMA-Model Based Approach to Seasonal Adjustment', *Journal of the Amer. Assoc.*, Vol. 77, No.377 , 63-70.

Hillmer, S.C., Bell, W.R. and 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*, Washington D.C.: US Dept. of Commerce-Bureau of the Census, 74-100.

Hodrick, T. and Prescott, E. (1980), *Post-War US Business Cycles: An Empirical Investigation*, Carnegie Mellon University, manuscript.

Hodrick, T. and Prescott, E. (1997), 'Post-War US Business Cycles: An Empirical Investigation', *Journal of Money, Credit, and Banking*, 29, 1-16.

Ka1man, R.E. (1960), 'A New Approach to Linear Filtering and Prediction Problems', *ASME Trans., Journal Basic Eng.*, 83-D, 95-108.

Nerlove, M., Grether, D.M. and Carvalho, J.L. (1979), *Analysis of Economic Time Series: a Synthesis*, Academic Press: New York.

Ng, C.N. and Young, P.C. (1990), 'Recursive Estimation and Forecasting of Nonstationary Time Series', *Journal of Forecasting*, 9, 173-204.

Norton, J.P. (1986), *An Introduction to Identification*, Academic Press.

Pedregal, D.J. (1995), *Comparación Teórica, Estructural y Predictiva de Modelos de Componentes no Observables y Extensiones del Modelo de Young*, Phd Thesis. Universidad Autónoma de Madrid: Madrid.

Pedregal, D.J. and 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), Technical Note No. 145.

Pedregal, D.J. (2001), 'Analysis of Economic Cycles Using Unobserved Components Models', *Review on Economic Cycles*, II, 77-92.

Pedregal, D.J. and Young, P.C. (2001), 'Some comments on the use and abuse of the Hodrick-Prescott Filter', *Review on Economic Cycles*, II, 93-104.

West, M. and Harrison, J. (1989), *Bayesian Forecasting and Dynamic Models*, Springer-Verlag: New York.

Whittle, P. (1983), *Prediction and Regulation*, English Universities press: London.

Young, P.C. (1988), 'Recursive Extrapolation, Interpolation and Smoothing of Nonstationary Time Series', in Chen, C. F. (ed.), *Identification and System Parameter Estimation*, Pergamon Press: Oxford, 33-44.

Young, P.C. (1993), 'Time Variable and State Dependent Modelling of Nonstationary and Nonlinear Time Series' in Subba Rao, T. (ed.) *Developments in Time Series, volume in honour of Maurice Priestley*, Chapman and Hall: London.

Young, P.C. (1994), 'Time-variable Parameter and Trend Estimation in Non-stationary Economic Time Series', *Journal of Forecasting*, 13, 179-210.

Young, P.C. (1998), 'Data-based mechanistic modelling of engineering systems', *Journal of Vibration and Control*, 4, 5-28.

Young, P.C. (1999), 'Data-based mechanistic modelling , generalised sensitivity and dominant mode analysis', *Computer Physics Communications*, 115, 1-17.

Young, P.C. and Pedregal, D.J., (1996), 'Recursive and En-Block Approaches to Signal Extraction', *Journal of Applied Statistics*, 26, 103-128.

Young, P.C. and Pedregal, D.J. (1997), 'Data-Based Mechanistic Modelling', C. Heij *et al.* (eds.), *System Dynamics in Economic and Financial models*, J. Wiley: Chichester.

Young, P.C. and Pedregal, D.J. (1999), 'Macro-Economic Relativity: Government Spending, Private Investment and Unemployment in the USA', *Structural Change and Economic Dynamics*, 10, 359-380.

Young, P.C., Ng, C.N. and 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, 481-501.

Young, P.C., Pedregal, D.J. and Tych, W. (1999), 'Dynamic Harmonic Regression', *Journal of Forecasting*, 18, 369-394.

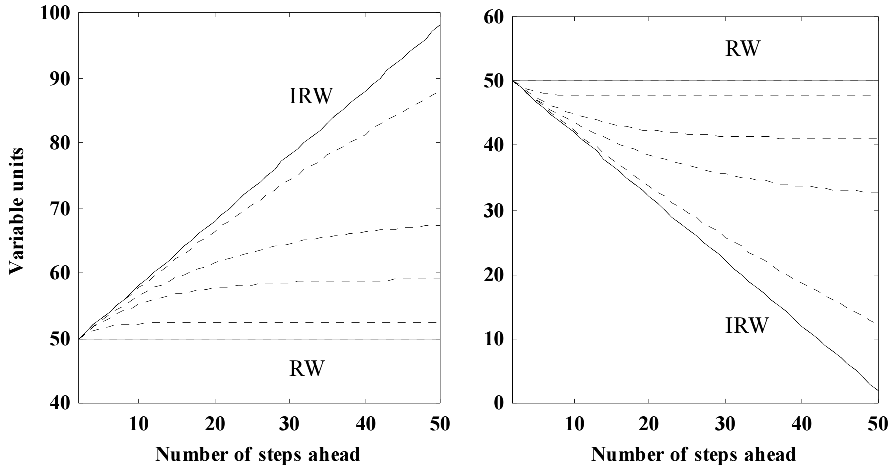

Figure 1: Examples of SRW trend forecasting for different values of the α parameter (0, 0.7, 0.9, 0.95, 0.99 and 1, respectively, from bottom to top on left hand box and 1, 0.99, 0.95, 0.9, 0.7 and 0, respectively, from bottom to top on right hand box).

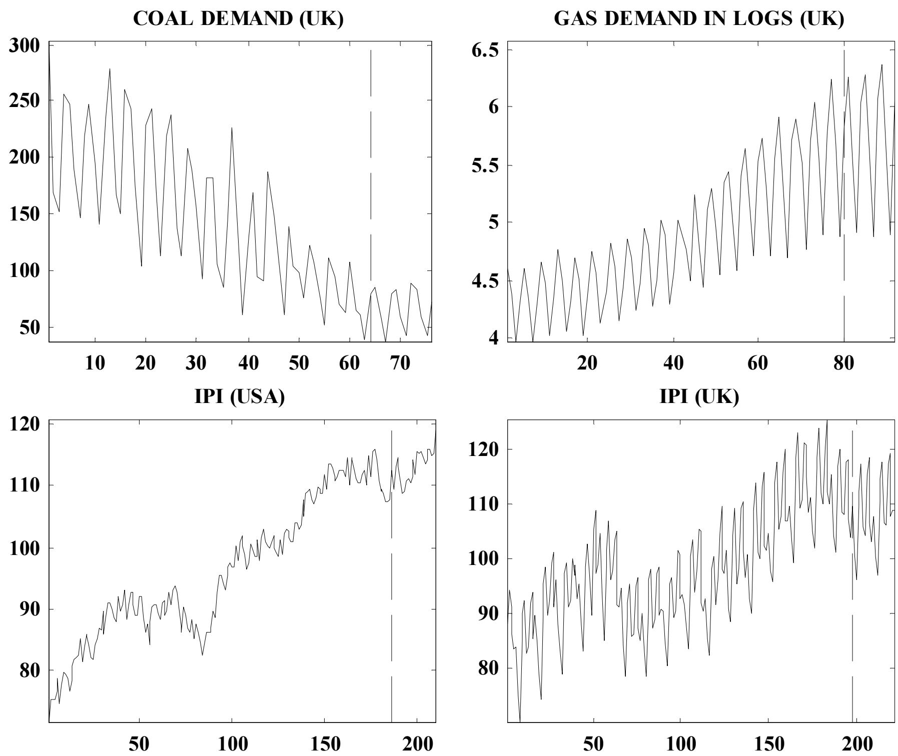

Figure 2: Band-pass of SRW filters for (a) trend and (b) one seasonal periodic component (4 samples/cycle). α values vary from 0 to 1 by 0.1 increments.

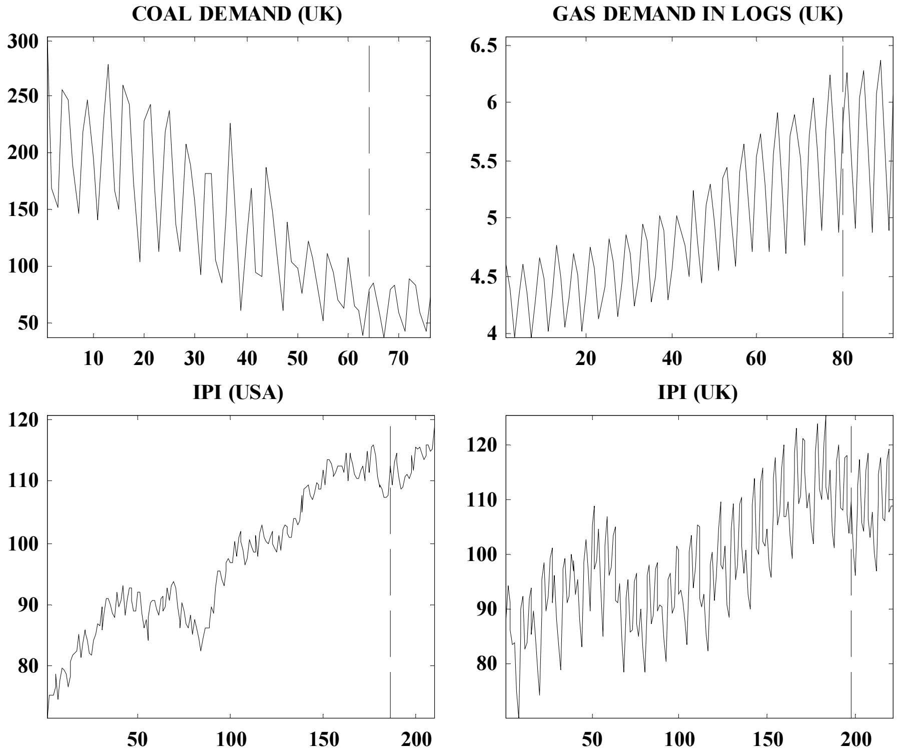

Figure 3: Time series examples.

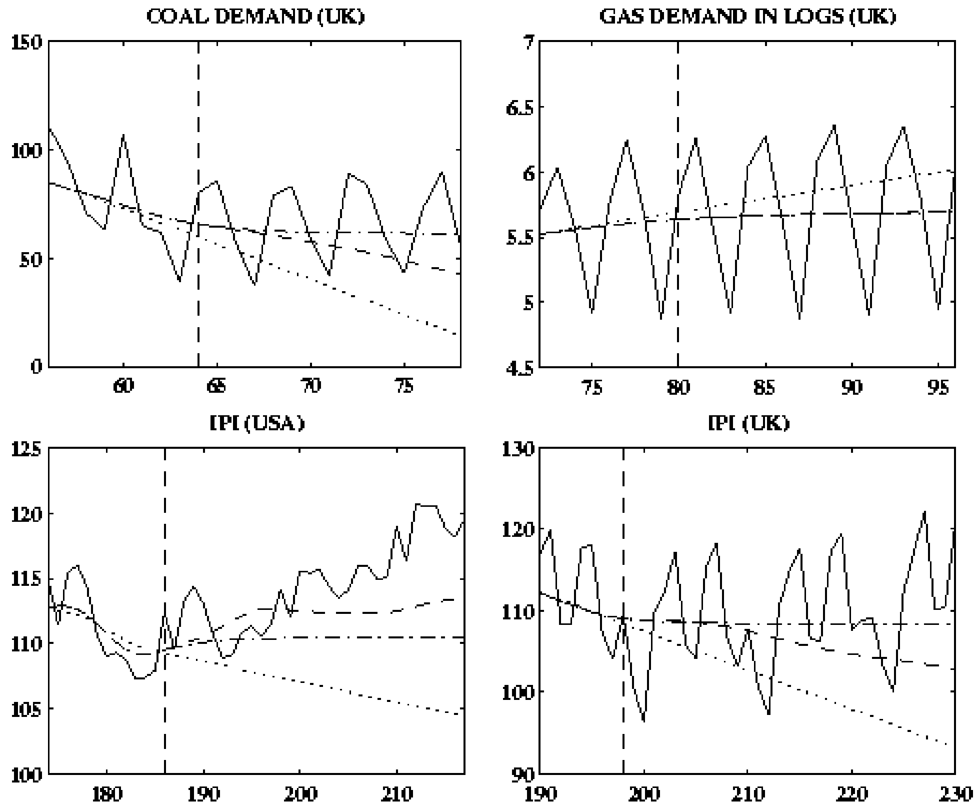

Figure 4: Actual series (solid); IRW trend (dotted); SRW trend (dashed-dot) and DIAR trend (dashed).