

DYNAMIC HARMONIC REGRESSION

Peter C. Young, Diego J. Pedregal and Wlodek Tych

Appears in *Journal of Forecasting*, **18**, 369-394
CRES, Institute of Environmental and natural Sciences
Lancaster University, Lancaster LA1 4YQ, U.K.

Keywords: Time series, state-space models, optimal state estimation, optimal filtering, optimal smoothing, seasonal adjustment, dynamic harmonic regression

ABSTRACT

This paper describes in detail a flexible approach to nonstationary time series analysis based on a *Dynamic Harmonic Regression* (DHR) model of the *Unobserved Components* (UC) type, formulated with a stochastic state space setting. The model is particularly useful for adaptive seasonal adjustment, signal extraction and interpolation over gaps, as well as forecasting or backcasting. The *Kalman Filter* and *Fixed Interval Smoothing* algorithms are exploited for estimating the various components, with the *Noise Variance Ratio* and other hyper-parameters in the stochastic state space model estimated by a novel optimisation method in the frequency domain. Unlike other approaches of this general type, which normally exploit *Maximum Likelihood* methods, this optimisation procedure is based on a cost function defined in terms of the difference between the logarithmic pseudo-spectrum of the DHR model and the logarithmic autoregressive spectrum of the time series. This cost function not only seems to yield improved convergence characteristics when compared with the alternative ML cost function, but it also has much reduced numerical requirements.

1. INTRODUCTION

Many socio-economic and environmental time series exhibit periodic, seasonal or cyclical effects of various kinds. For instance, both climate and socio-economic phenomena are influenced by annual seasonal variations; life cycles display both diurnal and annual rhythms; and quasi-periodic business cycles influence economic observations. In order to investigate such periodic or quasi-periodic phenomena, as well as performing functions such as forecasting, interpolation over gaps or seasonal adjustment, one needs to estimate the periodic components in some manner. This estimation problem is complicated by the fact that the periodic variations which affect real time series are often nonstationary, in the sense that their amplitude and phase tend to change over the observation interval. Various identification and estimation procedures have been suggested for handling such nonstationary periodicity but some of the most powerful originate in the statistical and econometrics literature, where the estimation problem is often termed *signal extraction*.

One of the oldest and best known techniques for signal extraction is the Census X-11 method (Shiskin *et al.*, 1967) and its later extensions X-11 ARIMA, and X-12 ARIMA (e.g. Dagum, 1980,1988; Findley *et al.*, 1992, 1996). However, three other important approaches to signal extraction have been developed in recent years and have gained favour because of their greater inherent flexibility. These methods are based on: (a) the decomposition of *ARIMA* or 'reduced form' models (e.g. e.g. Box *et al.* (1978); Hillmer and Tiao (1982); Hillmer *et al.* (1983); Burman (1980); Gómez and Maravall (1996) and Maravall (1993); (b) optimal *Regularisation* (see e.g. Akaike, 1980; Jakeman and Young, 1979, 1984; Young, 1991; Young and Pedregal, 1999a); or (c) formulation of the problem within a stochastic *State Space* (SS) setting. This latter SS approach provides the most obvious formulation of Unobserved Component (UC) models of the kind discussed in the present paper. Here, the UC model is considered as the observation equation of a discrete time, stochastic SS model and the associated state equations are used to model each of the components in Gauss-Markov (GM) terms.

This SS formulation has its origin in the 1960's when control engineers realised that recursive estimation and, in particular, the *Kalman Filter* (KF), could be applied to the problem of estimating time variable parameters in regression-type models, usually within a dynamic systems context (see e.g. Young, 1969, 1974, 1984). More recent developments (see e.g.. Norton, 1975; Harrison and Stevens, 1976; Harvey, 1984, 1989; West and Harrison, 1989; Young, 1988, 1989, 1994; Young *et al.*, 1989; Ng and Young, 1990) have shown how this approach can be extended, in various ways, to problems of forecasting, backcasting, smoothing and signal extraction. Here, the most influential recent contributions are probably those of Harvey, whose *Structural Model* approach is now widely available in the successful STAMP (*Structural Time Series Analyser, Modeller and Predictor*: see Koopmans *et al.*, 1995) computer program. In addition to the Kalman filter, perhaps the most important of the methodological developments in this area of study have been the exploitation of various forms of the recursive *Fixed Interval Smoothing* techniques (e.g. Bryson and Ho, 1969), which provide an 'off-line', adaptive approach to time variable parameter and state estimation, so allowing for optimal signal extraction, smoothing, and interpolation over gaps in the data.

The *Dynamic Harmonic Regression* (DHR) model considered in the present paper is of the UC type and is formulated within the stochastic SS setting. It is important to stress, however, that this approach can yield asymptotically equivalent results to the other approaches mentioned above if the models on which they are all based are made compatible (see above references). Nevertheless, it seems to the authors (see Young and Pedregal, 1999a) that the SS formulation is not only a more natural and satisfying setting for UC models, but it is also more attractive computationally because it so nicely integrates the processes of forecasting, interpolation and seasonal adjustment into a single recursive framework. To paraphrase

Wiener (1949), it seems to provide a rather natural approach to the “extrapolation, interpolation and smoothing of *nonstationary* time series”. As we shall see, the main difference between the DHR approach and related techniques, such as Harvey’s Structural Modelling method, lies in the formulation of the UC model for the periodic components and the method of optimising the hyper-parameters in this model.

2. THE UNOBSERVED COMPONENTS MODEL

The DHR model is a special case of the univariate UC model which has the general form:

$$y_t = T_t + C_t + S_t + f(\mathbf{u}_t) + N_t + e_t \quad e_t \sim N\{0, \sigma^2\} \quad (1)$$

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); $f(\mathbf{u}_t)$ captures the influence of a vector of exogenous variables \mathbf{u}_t , if necessary including stochastic, nonlinear static or dynamic relationships; N_t is an stochastic perturbation model (i.e. coloured noise modelled as an AR or ARMA process); and, as shown, e_t is an ‘irregular’ component, normally defined for analytical convenience as a normally distributed Gaussian sequence with zero mean value and variance σ^2 (i.e. discrete-time white noise). In order to allow for nonstationarity in the time series y_t , the various components in (1), including the trend T_t , can be characterised by stochastic, Time Variable Parameters (TVP’s), with each TVP defined as a nonstationary stochastic variable, as discussed below.

In practice, not all the components in (1) are necessary: indeed, the simultaneous presence of all these components can induce identifiability problems and so use of the complete model (1) is not advisable in practice unless adequate precautions are taken. The DHR model is one such decomposition which contains only the trend, cyclical, seasonal and white noise components: i.e.,

$$y_t = T_t + C_t + S_t + e_t \quad (2)$$

Here, the most important unobserved components are the seasonal term S_t and the cyclical term C_t . Although these are both modelled in the same manner, it is convenient to define them separately, i.e.,

$$\text{Seasonal:} \quad S_t = \sum_{i=1}^{R_s} \{a_{i,t} \cos(\omega_i t) + b_{i,t} \sin(\omega_i t)\} \quad (2a)$$

where $a_{i,t}$ and $b_{i,t}$ are stochastic TVP’s and $\omega_i, i = 1, 2, \dots, R_s$ are the fundamental and harmonic frequencies associated with the seasonality in the series; and

$$\text{Cyclical:} \quad C_t = \sum_{i=1}^{R_c} \{\alpha_{i,t} \cos(f_i t) + \beta_{i,t} \sin(f_i t)\} \quad (2b)$$

where $\alpha_{i,t}$, $\beta_{i,t}$ are stochastic TVP's and f_i , $i = 1, 2, \dots, R_c$ are the frequencies associated with the (normally longer period) cyclical component. In both cases, the frequency values are chosen by reference to the spectral properties of the time series, as discussed later. The trend component T_t can also be considered as a stochastic, time variable 'intercept' parameter in the DHR and so it could be incorporated, if so desired, into the cyclical or seasonal components as a zero frequency term (i.e. by introducing an extra term in the summations at $i = 0$ with ω_0 or f_0 set equal to zero). This DHR model can be considered as a straightforward extension of the classical, constant parameter, harmonic regression (Fourier series or trigonometric) model, in which the gain and phase of the harmonic components can vary as a result of estimated temporal changes in the parameters $a_{i,t}$, $b_{i,t}$, $\alpha_{i,t}$ and $\beta_{i,t}$.

In the simplest case, each of these $a_{i,t}$, $b_{i,t}$, $\alpha_{i,t}$ and $\beta_{i,t}$ stochastically variable parameters, as well as the trend component T_t (see above comments), are defined by a two dimensional stochastic state vector $\mathbf{x}_{i,t} = [l_{i,t} \quad d_{i,t}]^T$, where $l_{i,t}$ and $d_{i,t}$ are, respectively, the changing level and slope of the associated trend or TVP. The stochastic evolution of each $\mathbf{x}_{i,t}$ is assumed to be described by a Generalised Random Walk (GRW) process of the form,

$$\mathbf{x}_{i,t} = \mathbf{F}_i \mathbf{x}_{i,t-1} + \mathbf{G}_i \eta_{i,t} \quad i = 1, 2, \dots, R \quad (2c)$$

where $R = 1 + R_c + R_s$ and,

$$\mathbf{F}_i = \begin{bmatrix} \alpha & \beta \\ 0 & \gamma \end{bmatrix}, \quad \mathbf{G}_i = \begin{bmatrix} \delta & 0 \\ 0 & 1 \end{bmatrix}$$

This general model comprises as special cases the Integrated Random Walk (IRW: $\alpha = \beta = \gamma = 1$; $\delta = 0$); the scalar Random Walk (RW: scalar but equivalent to (2c) with $\alpha = \beta = \delta = 0$; $\gamma = 1$); the intermediate case of Smoothed Random Walk (SRW: $0 < \alpha < 1$; $\beta = \gamma = 1$; and $\delta = 0$); and Harvey's 'Local Linear Trend' (LLT: $\alpha = \beta = \gamma = 1$; $\delta = 1$) and 'Damped Trend' (DT: $\alpha = \beta = \delta = 1$; $0 < \gamma < 1$), see Harvey (1989) and Koopmans *et al.* (1995). The GRW model was introduced in Jakeman and Young (1979, 1984); and is discussed further in Young *et al.* (1989).

Clearly other, more general and higher order GM processes could be used to model the stochastic TVP's and the trend T_t , if such models can be identified satisfactorily from the data.: for example, the higher order IRW's (DIRW, TIRW etc.), the Integrated or Double Integrated AutoRegressive (IAR, DIAR: see Young, 1994) model, and even more general processes (e.g. Pedregal and Young, 1998b). However, the more complex of these models introduce additional hyper-parameters into the model which would have to be well identified from the data and optimised, thus introducing potential practical difficulties.

The overall SS model is constructed by the aggregation of the subsystem matrices defined in (2c), with the observation equation defined by equation (2): i.e.,

$$\begin{aligned}
\text{Observation Equation:} \quad & y_t = \mathbf{H}_t \mathbf{x}_t + e_t & (i) \\
\text{State Equations} \quad & : \quad \mathbf{x}_t = \mathbf{F} \mathbf{x}_{t-1} + \mathbf{G} \boldsymbol{\eta}_t & (ii)
\end{aligned} \tag{3}$$

If $n = 2R$, then \mathbf{F} is an $n \times n$ block diagonal with blocks defined by the \mathbf{F}_i matrices in (2c); \mathbf{G} is an $n \times n$ matrix constructed by the concatenation of the corresponding subsystem matrices \mathbf{G}_i in (2c); \mathbf{H}_t is an appropriately defined $1 \times n$ vector which relates the scalar observation y_t to the state variables defined by (3)(ii), so that it represents the DHR model (2); and $\boldsymbol{\eta}_t$ is an n dimensional vector containing, in appropriate locations, the zero mean, white noise input vectors $\eta_{i,t}$ (system disturbances) to each of the GRW models in (2c). These white noise inputs are assumed to be independent of the observation noise e_t and have a covariance matrix \mathbf{Q} .

3. ESTIMATION OF THE TIME VARIABLE PARAMETERS

The SS formulation of unobserved components models is particularly well suited to estimation based on optimal Kalman filtering, often accompanied by optimal smoothing procedures such as Fixed Interval Smoothing (FIS). As in our previous publications, we use the two step (prediction-correction) version of KF followed by a version of FIS algorithm which is stable numerically and which has a structure allowing for variance intervention (Young and Ng, 1989) and handling of missing observations and outliers. More specifically, it is an adapted version combining Bryson and Ho's recursion for the Lagrange multipliers (Bryson and Ho, 1969) with the state update recursion of Norton (1986). In relation to the time series y_t , $t = 1, 2, \dots, N$, of N samples, the filtering algorithm has the form:

1. Forward Pass Filtering Equations (Kalman, 1960)

Prediction:

$$\begin{aligned}
\hat{\mathbf{x}}_{t|t-1} &= \mathbf{F} \hat{\mathbf{x}}_{t-1} \\
\mathbf{P}_{t|t-1} &= \mathbf{F} \mathbf{P}_{t-1} \mathbf{F}^T + \mathbf{G} \mathbf{Q}_r \mathbf{G}^T
\end{aligned} \tag{i} \tag{4a}$$

Correction:

$$\begin{aligned}
\hat{\mathbf{x}}_t &= \hat{\mathbf{x}}_{t|t-1} + \mathbf{P}_{t|t-1} \mathbf{H}_t^T \left[1 + \mathbf{H}_t \mathbf{P}_{t|t-1} \mathbf{H}_t^T \right]^{-1} \{ y_t - \mathbf{H}_t \hat{\mathbf{x}}_{t|t-1} \} \\
\mathbf{P}_t &= \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1} \mathbf{H}_t^T \left[1 + \mathbf{H}_t \mathbf{P}_{t|t-1} \mathbf{H}_t^T \right]^{-1} \mathbf{H}_t \mathbf{P}_{t|t-1}
\end{aligned} \tag{ii} \tag{4a}$$

while the FIS algorithm is in the form of a backward recursion operating from the end of the sample set to the beginning.

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

$$\begin{aligned}
\hat{\mathbf{x}}_{t|N} &= \mathbf{F}^{-1} \left[\hat{\mathbf{x}}_{t+1|N} + \mathbf{G} \mathbf{Q}_r \mathbf{G}^T \mathbf{L}_t \right] \\
\mathbf{L}_t &= \left[\mathbf{I} - \mathbf{P}_{t+1} \mathbf{H}_{t+1}^T \mathbf{H}_{t+1} \right]^T \left[\mathbf{F}^T \mathbf{L}_{t+1} - \mathbf{H}_{t+1}^T \{ y_{t+1} - \mathbf{H}_{t+1} \hat{\mathbf{x}}_{t+1} \} \right] \\
\mathbf{P}_{t|N} &= \mathbf{P}_t + \mathbf{P}_t \mathbf{F}^T \mathbf{P}_{t+1|t}^{-1} \left[\mathbf{P}_{t+1|N} - \mathbf{P}_{t+1|t} \right] \mathbf{P}_{t+1|t}^{-1} \mathbf{F} \mathbf{P}_t
\end{aligned} \tag{4b}$$

with $\mathbf{L}_N = \mathbf{0}$.

In these algorithms, the $n \times n$ Noise Variance Ratio (NVR) matrix \mathbf{Q}_r and the $n \times n$ matrix \mathbf{P}_t are defined as follows,

$$\mathbf{Q}_r = \frac{\mathbf{Q}}{\sigma^2} \quad \mathbf{P}_t = \frac{\mathbf{P}_t^*}{\sigma^2} \tag{4c}$$

where \mathbf{P}_t^* is the error covariance matrix associated with the state estimates. For simplicity, it will be assumed that the NVR matrix \mathbf{Q}_r is diagonal. Within the DHR context, these NVR parameters are unknown and need to be estimated before the filtering and smoothing algorithms can be utilised. Their estimation by optimisation in the frequency domain is discussed in the next section 4.

Note that forecasting, interpolation and backcasting are an inherent part of the filtering and smoothing algorithms (see Young, 1984; Harvey, 1989): if missing data anywhere within, or immediately outside, the data set are detected, then the filtering and smoothing algorithms simply replace the missing samples by their expectations, based on the DHR model and the data. In the case of the filtering algorithm (4a), this yields the multi-step ahead forecasts; while the smoothing equations (4b), provide the interpolation over gaps within the data set, or backcasts at the beginning of the data.

Finally, it is worth noting that, if necessary, these filtering/smoothing algorithms can be made computationally more efficient by taking advantage of factors such as the repeated computation of some terms common to the forward and backward pass, the evaluation of some forward pass terms prior to their storage for the backward pass, and exploiting the sparse structure of the matrices involved. Such an improved computational algorithm is implemented in microCAPTAIN (Young and Benner, 1991). See also Appendix 1 for a further more specific discussion on the computational aspects.

4. OPTIMISATION OF THE HYPER-PARAMETERS

The recursive algorithms in equations (4a) and (4b) require estimates of the NVR matrix \mathbf{Q}_r . The usual way of dealing with this estimation problem (see e.g. Harvey, 1989) is to formulate it in Maximum Likelihood (ML) terms and obtain estimates of the NVR parameters (and any other hyper-parameters, such as α in the SRW model) using numerical optimisation. While this method is generally accepted because of its strong theoretical basis, it has some practical disadvantages. In particular, the optimisation can be restricted in various ways due to fact

that, in this UC context where the ML function is defined by prediction error decomposition (Schweppe, 1965) in terms of the Kalman filter innovations (recursive residuals), the likelihood surface can be quite flat around its optimum. It is also a procedure that, in its normal form, is heavily dependent on the length of the series, since the recursive algorithms must be used to compute the ML function at each iteration in the numerical optimisation. In order to counter these and other limitations, we have devised an alternative approach which utilises a special form of optimisation in the frequency domain. Although this approach is not now based on ML and is not as generally applicable, it leads to a much better defined optimum in the objective function-hyper-parameter space, with consequent advantages to DHR modelling, both as regards convergence time and the number of parameters that can be optimised simultaneously.

4.1 Spectral Analysis of the DHR model

The new method of optimising the hyper-parameters in the DHR model is formulated in the frequency domain and based upon analytical expressions for the power spectra of the DHR model class. In order to obtain these analytical expressions, it is necessary to determine the spectral properties of both the TVP's, as defined by their assumed stochastic models, and the products of such parameters with the sine and cosine functions that characterise the DHR model. As mentioned earlier, these TVP's are normally described by any of the models belonging to the GRW family.

The Pseudo-Spectra of Generalised Random Walk models

Consider first a UC model consisting of only an IRW trend (i.e. equation (2c)) with $\alpha = \beta = \gamma = 1; \delta = 0$) plus a white noise irregular component. The resultant model can be expressed in transfer function (reduced form) terms as,

$$y_t = l_t + e_t = \frac{1}{(1-L)^2} \eta_{t-1} + e_t$$

where L denotes the backward shift operator (i.e. $Ly_t = y_{t-1}$). Initially, the inherent non-stationarity of the IRW process can be dealt with by taking second differences of y_t to yield,

$$\nabla^2 y_t = \eta_{t-1} + \nabla^2 e_t$$

where $\nabla = (1-L)$ is the difference operator. For this stationary process the power spectrum takes the form,

$$f_{\nabla^2 y}(\omega) = \frac{1}{2\pi} \left[\sigma_\eta^2 + \{2 - 2\cos(\omega)\}^2 \sigma^2 \right]; \quad \omega \in [0, \pi] \quad (5)$$

and it is straightforward to see that the pseudo-spectrum of y_t is given by,

$$f_y(\omega) = \frac{1}{2\pi} \left[\frac{\sigma_\eta^2}{\{2 - 2\cos(\omega)\}^2} + \sigma^2 \right]. \quad (6a)$$

Similarly, the pseudo-spectrum in the case of the simpler RW trend is given by,

$$f_y(\omega) = \frac{1}{2\pi} \left[\frac{\sigma_\eta^2}{2 - 2\cos(\omega)} + \sigma^2 \right] \quad (6b)$$

The resulting spectral properties of the IRW and RW-type filters are shown graphically in fig.1, for a selection of different NVR values.

(INSERT FIGURE 1)

In the case where the trend is modelled by an SRW process, the pseudo spectrum is somewhat more complex and takes the form.

$$f_y(\omega) = \frac{1}{2\pi} \left[\frac{\sigma_\eta^2}{\{1 + \alpha^2 - 2\alpha\cos(\omega)\}\{2 - 2\cos(\omega)\}} + \sigma^2 \right] \quad (6c)$$

Note how the introduction of the smoothing parameter α allows for a continuous transition between the RW model ($\alpha = 0$) and the IRW model ($\alpha = 1$). This advantageous property is illustrated in fig.2. From the filter design point of view, the additional α parameter allows for varying the shape (width or bandpass) of the filter's frequency response, while the NVR parameter shifts the response vertically.

(INSERT FIGURE 2)

Clearly, the pseudo-spectra of other more general GM processes (see previous discussion in section 2) can be analysed in a similar manner.

The Pseudo-Spectra of the Full DHR model

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

$$f_s(\omega) = \frac{1}{2} \left[f_A(\omega - \omega_j) + f_A(\omega + \omega_j) \right]$$

where $f_A(\omega)$ is the frequency response of a_t . Consider the case of a single frequency DHR term, $S_t = a_t \cos(\omega_j t) + b_t \sin(\omega_j t)$, in which the TVP associated with the sine and cosine terms are modelled as two IRW processes with equal variance parameters ($\sigma_{\omega_j}^2$). The pseudo-spectrum of S_t then takes the form:

$$f_{\omega_j}(\omega) = \frac{1}{2\pi} \left[\frac{\sigma_{\omega_j}^2}{4\{1 - \cos(\omega - \omega_j)\}^2} + \frac{\sigma_{\omega_j}^2}{4\{1 - \cos(\omega + \omega_j)\}^2} \right] \quad (7)$$

The RW and SRW cases may be obtained in the same way by replacing (6a) by (6b) or (6c) respectively.

Now, if $S(\omega, \omega_j)$ is defined as follows,

$$S(\omega, \omega_j) = \frac{1}{2\pi} \left[\frac{1}{4\{1 - \cos(\omega - \omega_j)\}^2} + \frac{1}{4\{1 - \cos(\omega + \omega_j)\}^2} \right]$$

then the pseudo-spectrum of the full DHR model (2) becomes:

$$f_y(\omega, \underline{\sigma}^2) = \sum_{j=0}^R \sigma_{\omega_j}^2 S(\omega, \omega_j) + \frac{\sigma^2}{2\pi} \quad \underline{\sigma}^2 = [\sigma^2 \quad \sigma_{\omega_0}^2 \quad \sigma_{\omega_1}^2 \quad \dots \quad \sigma_{\omega_R}^2] \quad (8)$$

where R is the total number of number of different frequency components included in the model (see equations (2a) and (2b)). It will be noted that the additional term at $j=0$ (i.e. $\omega_0 = 0$) represents the zero frequency (DC) component and so accounts for the IRW trend component which, as mentioned previously, has been absorbed rather conveniently into the definition of $S(\omega, \omega_j)$. Note also that this formula is linear in the variance parameters and, as we shall see, this facilitates the initial estimation of the hyper-parameters. The extension of $S(\omega, \omega_j)$ to accommodate more complex combinations of RW, IRW and SRW defined trends and parameters is obvious.

The dotted lines in fig.3 shows the pseudo-spectra of the components in a DHR model, with an IRW trend and the TVP's in the harmonic components modelled by RW processes. These harmonic components correspond to a periodic process with a fundamental frequency of 12 samples and harmonics at 6, 4, 3 and 2.4 samples (typical of a model for a monthly time series with annual seasonality). Note that, on the linear scale, the components are well separated spectrally and the ratio of maximum power to the minimum is of the order 10^3 .

(INSERT FIGURE 3)

The close relationship of the DHR to Discrete Fourier Transform (DFT) or, equivalently, the constant parameter Harmonic Regression (HR) model is clearly apparent from fig. 3. As the $\sigma_{\omega_j}^2$ (or more particularly as the ratio of this variance to the residual variance, σ^2 , as defined by the NVR) becomes smaller, which is equivalent to decreasing the potential variability of the related sine and cosine amplitude parameters, the spectral peak of the j th frequency component approaches the shape of spectral line, as in the case of the constant parameters HR model. In other words, the DHR model can be considered as a logical stochastic TVP (or nonstationary) version of HR and DFT models.

4.2 Estimation in the Frequency Domain

Equation (8) shows that the spectrum of the DHR process can be considered as a linear combination of known spectral terms, with the unknown variance parameters appearing as associated regression coefficients. Clearly, therefore, one way of estimating the variance parameter vector $\underline{\sigma}^2 = [\hat{\sigma}^2 \quad \hat{\sigma}_{\omega_0}^2 \quad \hat{\sigma}_{\omega_1}^2 \quad \dots \quad \hat{\sigma}_{\omega_R}^2]^T$ is to minimise a least squares objective function J of the form,

$$J(f_y, \hat{f}_y) = \sum_{i=0}^{T-1} [f_y(\omega_i) - \hat{f}_y(\omega_i, \underline{\hat{\sigma}}^2)]^2 \quad (9)$$

over T distinct frequencies, $\omega_i, i = 0, 1, 2, \dots, T-1$, in the range 0 to 0.5, where $f_y(\omega)$ is the empirical spectrum, while the model pseudo-spectrum $\hat{f}_y(\omega, \underline{\hat{\sigma}}^2)$ is defined as the following linear function of the $R+2$ unknown variance parameter estimates,

$$\hat{f}_y(\omega, \underline{\hat{\sigma}}^2) = \sum_{j=0}^R \hat{\sigma}_{\omega_j}^2 S(\omega, \omega_j) + \frac{\hat{\sigma}^2}{2\pi}$$

A visual illustration of the estimation problem posed in these terms can be obtained by referring again to fig.3, where the spectra of the IRW trend and the 5 harmonic components with RW parameters (dotted lines) can clearly be adjusted by the above optimisation approach to fit the Akaike Information Criterion (AIC) identified AR(14) spectrum (full line) of the well known Airline Passenger (AP) monthly series (Box and Jenkins, 1970), which are also plotted in fig.4.

While this linear least squares solution is attractive in its simplicity, practical experience has shown that an alternative objective function,

$$J_L(f_y, \hat{f}_y) = \sum_{i=0}^{T-1} [\log\{f_y(\omega_i)\} - \log\{\hat{f}_y(\omega_i, \underline{\hat{\sigma}}^2)\}]^2 \quad (10)$$

defined in logarithmic terms, yields improved estimation results, with a more clearly located optimum and better defined estimates of the variance parameters. This arises predominantly because the logarithmic spectra have clearer information on the shape of the spectral signatures, particularly the ‘shoulders’ that are so important in defining the magnitude of the variance parameters. This logarithmic measure is also intuitively rather natural in the signal processing area since power spectra are usually presented and measured on logarithmic scales such as decibels. The disadvantage of this approach is, of course, that J_L is a nonlinear function of the unknown variance parameters and so they must be estimated by nonlinear optimisation, as in the conventional ML optimisation case.

Fortunately, this nonlinear optimisation is quite straightforward and presents no difficulties. Not only is the location of the optimum normally well defined, but the linear least squares

solution (9) quickly and reliably provides very good initial conditions for the nonlinear minimisation problem (10). Consequently optimisation algorithms with quadratic convergence, that are otherwise sensitive to the initial conditions, can be used with confidence. For this purpose, the objective functions have to fulfil certain conditions (Deutsch, 1965) in order that the parameter estimates are mathematically acceptable, but it is easy to show that the objective function (10) satisfies these conditions and is both continuous and convex.

One final simplification is possible: if $f_y(\omega)$ is based on the estimated AutoRegressive (AR) spectrum (see below), then the estimate of the residual white noise (one step ahead prediction errors) from the AR model can be used to concentrate out the observation noise variance σ^2 from the model pseudo-spectrum (8), i.e.,

$$f_y(\omega, \underline{\sigma}^2) = \sigma^2 \left\{ \frac{f_y(\omega, \sigma^2)}{\sigma^2} \right\} = \sigma^2 \left\{ \sum_{j=0}^R \frac{\sigma_{\omega_j}^2}{\sigma^2} S(\omega, \omega_j) + \frac{1}{2\pi} \right\}$$

or equivalently

$$f_y(\omega, \mathbf{NVR}) = \sigma^2 \left\{ \sum_{j=0}^R \mathbf{NVR}_j S(\omega, \omega_j) + \frac{1}{2\pi} \right\} \quad (11)$$

where \mathbf{NVR} is the NVR vector with elements,

$$\mathbf{NVR}_j = \frac{\sigma_{\omega_j}^2}{\sigma^2}, j = 0, 1, \dots, R$$

In this manner, the model pseudo-spectrum is defined directly in terms of the unknown NVR values. This not only removes one parameter from the estimation problem but it also introduces a common scale for the remaining unknown parameters which has numerical advantages in the subsequent numerical optimisation. For instance, practical experience with the spectral forms (8) and (11) over the past five years has demonstrated that the latter formulation provides better defined solutions over a wide range of data series.

5. THE COMPLETE DHR ESTIMATION ALGORITHM

In the case where a cyclical component is not present, the complete DHR estimation algorithm consists of the following four steps.

1. Estimate an AR(n) spectrum $f_y(\omega)$ of the observation process $y_t, t = 1, 2, \dots, N$ and its associated residual variance $\hat{\sigma}^2$, with the AR order n normally identified by reference to the AIC. Note the R significant peaks that characterise the spectrum (these will normally include a fundamental frequency and several of its associated harmonics).

2. Find the Linear Least Squares estimate of the **NVR** parameter vector which minimises the linear least squares objective function:

$$J(f_y, \hat{f}_y) = \sum_{i=0}^{T-1} \left[f_y(\omega_i) - \hat{f}_y(\omega_i, \mathbf{NVR}) \right]^2$$

where $\hat{f}(\omega, \mathbf{NVR}) = \sigma^2 \sum_{j=0}^R \text{NVR}_j S(\omega, \omega_j)$; ω_0 and NVR_0 refer to the trend term; and the $\omega_j, j = 1, 2, \dots, R$, are the R significant frequencies identified from the AR(n) spectrum in step **1**.

3. Find the Nonlinear Least Squares estimate of the **NVR** parameter vector which minimises nonlinear least squares objective function :

$$J_L(f_y, \hat{f}_y) = \sum_{i=0}^{T-1} \left[\log\{f_y(\omega_i)\} - \log\{\hat{f}_y(\omega_i, \mathbf{NVR})\} \right]^2$$

using result from step **2**. to define the initial conditions.

4. Use the NVR estimates from step **3**. to obtain the recursive forward pass (Kalman filter) and backward pass (FIS algorithm) smoothed estimates of the components in the DHR model: i.e. the trend; the total cyclical and seasonal; the fundamental/harmonic components; and the irregulars. In this form, the algorithm can be used for forecasting, backcasting, and signal extraction (e.g. seasonal or cyclical adjustment). Allowance for interventions and outliers, as well as interpolation over gaps can be introduced in the normal manner, as required.

If a cyclical component is identified then the above procedure is simply expanded to accommodate this additional component (see the AP example in section 6.2 below).

This optimisation algorithm has been in continual use over the past five years, both in the *microCAPTAIN* time series analysis program, and in research studies using extended DHR models (see later) in the *Matlab* software system. In all cases, the algorithm has exhibited excellent convergence characteristics, with rapid convergence to well defined NVR and other hyper-parameter estimates.

6. A WELL KNOWN EMPIRICAL EXAMPLE REVISITED

The famous AP series shown in fig.4 is used here in order to both exemplify the results obtained with the present DHR method and to facilitate comparison with the many previous published results relating to this very well known and analysed series. In addition, the standard ML method is used separately to estimate the unknown variance parameters; and the resulting DHR and ML results are compared with those obtained using the STAMP program (Koopmans *et al*, 1995), which also exploits ML optimisation but, this time, applied to Harvey's alternative *Structural Model* (SM).

Since the percentage fluctuations of the AP series about its long-term trend 'might be expected to be comparable at different sales volumes' (Box and Jenkins, 1970), the AP series is most often subjected to a 'stationarity inducing' logarithmic transform before analysis. In order to demonstrate the power of the DHR model, however, we will consider also the modelling and forecasting of the basic data without any pre-transform at all, thus forcing the DHR model to explain the nonstationarity in both the mean and the variance of the original data. In avoiding any form of subjective pre-processing, this latter approach conforms with our Data-Based Mechanistic (DBM) modelling philosophy (see e.g. Young and Pedregal, 1997; Young, 1998) which suggests that a minimum of prior judgement should be imposed upon the nature of the series.

(INSERT FIGURE 4)

6.1 The Basic DHR Model

The clear seasonal pattern in the series is confirmed by the AR(14) spectral estimates shown already in fig.3, with visually significant peaks at periods of 12, 6, 4, 3 and 2.4 months, but with no significant power at the Nyquist frequency. The strong trend component does not mask these peaks in the AR(14) spectrum and so the frequencies used for the subsequent analysis are easily defined as 0, 1/12, 1/6, 1/4, 1/3 and 1/2.4. Fig.5 illustrates the optimised model fit in the frequency domain for the case when the series is logarithmically transformed: here the trend is modelled by an IRW process and the seasonal component parameters by RW processes.

(INSERT FIGURE 5)

Two typical sets of estimation results are shown in Table 1 (basic data with LLT trend and IRW models for the seasonal component parameters) and Table 2 (logarithmically transformed data with IRW trend and RW models for the seasonal component parameters). The first column of these Tables shows the values of the estimated NVR parameters obtained via the first stage, linear optimisation; while the final stage nonlinear optimisation results are presented in the second column. The third column of Table 1 and the third and fourth

columns of Table 2 report the ML results. Also shown at the bottom of the Tables are the variance of the innovations; the value of the log-likelihood for the optimised parameters in each case; the Ljung-Box autocorrelation test for 12 lags; the Jarque-Bera normality test; and the computational burden in millions of floating point operations (Mega-Flops).

Two versions of ML estimation are considered for comparison: the third column in Table 1 shows the ML results obtained with all the NVR's of the seasonal component constrained to be the same (this is a common assumption and is used in the STAMP and other similar computer programs: see e.g. Koopmans *et al.*, 1995; Pole *et al.*, 1994). Unfortunately, when all the NVR's for the seasonal component are unconstrained, the ML optimisation does not converge at all. In Table 2, the ML convergence is improved somewhat but it is still necessary to constrain two of the NVR's (for the harmonic components of periods 4 and 2.4 months) to be the same to obtain full convergence (see column 4: this was accomplished by trial and error and is clearly not a feasible practical approach). Moreover, it will be noted that, even when the ML optimisation procedure does converge, it takes much longer (particularly in the unconstrained situation: see the last 'Mega-Flops' row), even though optimisation is initiated from the converged NVR parameter estimates obtained by our frequency domain optimisation method. Of course, convergence time also depends on both the formulation of the ML optimisation routine and method of numerical optimisation. Here, the basic prediction error formulation was utilised, involving full Kalman filter processing at each iteration; and the 'leastsq' numerical optimisation method available in Matlab was used for both DHR and ML optimisation. We realise, therefore, that improved convergence speed might be possible with other numerical procedures but our frequency domain approach would still be considerably faster than ML because of its inherent simplicity.

(INSERT TABLES 1 & 2)

There are a number of conclusions that can be drawn from these estimation results for the AP data which confirm similar conclusions based on the DHR analysis of many other series over the last five years.

1. The proposed method of frequency domain optimisation, based on the non-linear cost function (10), provides a clear improvement over the simpler linear optimisation approach using the cost function (9). Not only is the nonlinear cost function justified on the basis of the spectral considerations discussed above, but it also leads to considerable improvement in the statistical properties of the innovations: the innovations variance and most of the statistical tests reported in the Tables are improved significantly after the nonlinear second stage in the optimisation, is completed.
2. In terms of the values of the likelihood function and the statistical properties of the innovations, the nonlinear optimisation produces a better solution than the constrained ML estimation (although, we have found that it is often not quite as good as that obtained

by unconstrained ML optimisation *in those cases when ML convergence is achieved*). This is not surprising because the ML criterion is defined as an explicit, time domain function of the normalised one-step-ahead prediction errors; whereas optimisation in the frequency domain, as suggested in this paper, is not based on ML optimisation at all and is primarily concerned with ensuring that the spectral properties of the estimated components match the empirical AR spectrum of the data. In general, as in this example, this tends to generate a solution which, *in terms of the variance of the innovations*, lies somewhere between that obtained by the ideal of unconstrained ML optimisation, which is difficult to consistently achieve in practice, and the reality of constrained ML optimisation, which is used in commercial software, such as STAMP.

3. From the computational standpoint, the proposed method represents an extremely fast alternative to the ML estimation. As pointed out previously, the likelihood surface for many UC-type models tends to be very flat around the optimum, so that convergence problems are quite common in the unconstrained case, where the number of hyper-parameters is quite large compared with the constrained situation. Quite naturally, this also leads to slow and often unacceptable convergence times when compared with the very rapid convergence of the method proposed in this paper. This is especially true in the completely unconstrained case, when ML optimisation can converge very slowly or may not occur at all.

To provide further material for comparison, Table 3 shows the results for a number of models estimated by our frequency domain method; while Table 4 shows the results obtained for similar models by the STAMP program (Koopmans *et al*, 1995) which utilises constrained ML estimation in the time domain. The first and second columns in both Tables present the results for the original, un-transformed data, using LLT and IRW trends, respectively; while the third and fourth columns are the equivalent results for the logarithmically transformed series.

(INSERT TABLES 3 & 4)

In Table 3, the value of the likelihood function for the DHR model optimised by our frequency domain method (which, it must be emphasised again, does not specifically seek to optimise the likelihood) can be compared with the value for the constrained model optimised by ML with the NVR parameters of the seasonal component constrained. These figures are given in the ‘Log-Likelihood’ and ‘Constrained Lik.’ rows at the bottom of the table. In the first 3 cases (columns 1 to 3), the likelihood is greater for the frequency domain approach; and it is only slightly worse for the IRW model in the fourth column based on logarithmically transformed data. In other words, even when the model is optimised in the frequency domain, without any direct likelihood considerations, the method performs well in likelihood terms. There are also concomitant advantages in forecasting terms, but these are discussed later.

Table 4 shows the results obtained by the STAMP program (Koopmans *et al*, 1995) for the same models as in Table 3. In all cases, the variance of the residuals is considerably larger than in the comparable cases in Table 3. Note that, since the STAMP program does not allow for the omission of any individual harmonics, all harmonics have to be present in the estimated model. In this AP example, however, the spectrum shows clearly that the harmonic of period 2 samples/cycle, at the Nyquist frequency, is not necessary. Together with the constraint that all the hyper-parameters have to be the same, this places restrictions on the estimation which probably account, in part, for the larger residual variances.

Of course, comparison of the results in tables 3 and 4 must be exercised with care since it is a comparison between two general approaches to time series, rather than a comparison of two different methods of estimating the same model, as in Tables 1 and 2. Indeed, the same care needs to be exercised in comparing the STAMP results in table 4 with our own version of the ML procedure. For example, the STAMP model used to obtain the results in Table 4 is naturally constrained by the model options available in the STAMP program. For example, it incorporates a ‘trigonometric seasonal’ component defined in by the state equations (see Koopmans *et al*, 1995), rather than the harmonic regression formulation of the DHR model. And the hyper-parameters are defined as white noise variances, rather than NVR’s as in the DHR case.

Also, it will be noted from Table 4 that both STAMP models estimated on the basis of the basic, un-transformed data produce zero estimates for the irregular components. This means that the whole series is modelled by STAMP as simply the sum of the trend and the seasonal component with zero residuals, so that all noise within the data has been distributed between these two components. This problem is automatically avoided in our DHR modelling strategy because of the definition of the NVR parameters: here, zero variance of the irregular component would mean infinite NVR parameters, which is an impossible outcome of the DHR modelling procedure. Of course, Koopmans *et al* (1995) would understandably argue that, in this example, STAMP users would be advised to logarithmically transform the data because of the changing variance associated with the seasonal component (see previous comments in this regard), so that these ‘raw data’ results, as such, would not normally be obtained. Indeed, we have included them here not as a comment on the performance of the STAMP program, which has many virtues, but for completeness and to demonstrate that the DHR model can handle such ‘badly behaved’ nonstationary data without the need for any stationarity-inducing pre-transformation.

The signal extraction exercise using the frequency domain optimised DHR model produces a pleasing decomposition of the series. For example, fig.6 shows the FIS estimated components when an IRW trend is used, with IRW models for the seasonal component parameters. Here, the trend and seasonal component have smoothly changing characteristics and the irregularity

in the AP series has been accommodated almost entirely by the irregular series, which is of very low and constant variance. In addition, the autocorrelation function of the innovations sequence shows that it is satisfactorily white, as required. These results suggest that the KF and FIS algorithms based on the optimised DHR model have successfully handled and explained the dramatically changing variance and heteroskedasticity present in the raw data.

Finally, this signal extraction exercise reveals rather nicely another feature in the series that is hardly distinguishable in the original raw data; namely the fact that the trend appears to include a medium period cycle of just over four years which is, presumably, related to the business cycle (see fig.6). This suggests that further analysis and refinement of the DHR model might improve the proposed model: in particular, if this business cycle effect is reasonably well defined, then it should be possible to include a quasi-cyclical DHR component C_t in the DHR model at the identified frequency (see equation (2b)) and so improve both the signal extraction and the forecasting performance. This possibility is considered in the next sub-section.

(INSERT FIGURE 6)

6.2 An Improved DHR Model

The addition of the medium term cyclical component into the DHR model fits very naturally into the DHR analysis. The first stage of this extended analysis is to obtain a better estimate of the spectrum at the low frequencies relating to the cycle (i.e. at periods of around four years), since the previous AR(14) spectrum does not reveal any clear evidence for the presence of such a cycle (otherwise it would have been incorporated at the start of the analysis). This improved spectrum is generated straightforwardly by defining a finer frequency axis and increasing the order of the AR spectrum from AR(14) to AR(54). Now, a clearly defined peak appears in the spectrum at a fundamental period of 51 months, with two other peaks at 25.5 and 17 months, which will be recognised as harmonics of this fundamental period

The main problem with this high order AR(54) spectrum is that, while it nicely exposes the lower frequency behaviour in the trend, it injects obviously spurious peaks and distortions over the rest of the frequency axis, making estimation of the NVR parameters more difficult for components defined in this 'seasonal' region. To overcome this problem, we simply concatenate the original AR(14) and the new AR(54) spectra, using the higher order AR spectrum to define the lower frequency cyclical band of the spectrum, and the lower order AR spectrum to specify the higher frequency seasonal behaviour. The concatenation is then

smoothed a little (using a simple IRW version of the FIS algorithm¹), in order to remove the minor distortion that arises at the boundary between the two component spectra because of the concatenation operation.

Fig.7 compares this concatenated empirical spectrum with the optimised DHR model spectrum achieved using the DHR model (2), with the cyclical component C_t defined as follows,

$$C_t = \sum_{i=1}^3 \{ \alpha_{i,t} \cos(f_i t) + \beta_{i,t} \sin(f_i t) \} \quad (12)$$

where the frequencies $f_i; i=1,2,3$ correspond to the identified cyclical component frequencies (i.e. at periods of 51, 25.5 and 17 months, respectively). The three peaks associated with the cycle and its harmonics are very obvious in the figure, and it is clear that this extended DHR model is explaining them very well.

(INSERT FIGURE 7)

This revised DHR model is an ingenious one: as far as we are aware, no previous analysis has identified clearly the presence of the longer, 51 month period cycle in the AP data and quantified its effect in a time series model form (although the possible presence of such a cycle has been referred to on a few occasions). It is also a very useful model in forecasting terms because the additional information provided by the added cyclical terms in the DHR model leads to significant improvements in forecasting accuracy. In order to demonstrate this improvement and to compare the forecasting results, both the original and revised DHR models are optimised by different procedures. In particular, the original DHR model is optimised by ML in the time domain using STAMP; while our frequency domain approach is used both for the original and the revised model. In the latter two cases, the analysis is based on the original un-transformed data (with IRW models used for the parameters in the seasonal component and IRW models used for the parameters in the cyclical components). In the STAMP analysis, on the other hand, the logarithmically transformed data is used, since this is the only way it can handle well the changing variance of the seasonal component (see previous discussion). However, in order to facilitate comparison, all the forecasting results

¹ This IRW version of the FIS algorithm, which we term IRWSMOOTH, has been used in the microCAPTAIN program for about twenty years and has proven very useful as a general tool for initial (non-optimised) trend estimation and interpolation over gaps, smoothing, and smoothed differentiation of data. In this form, it is related to the ‘Hodrick-Prescott’ filter with an adjustable NVR smoothing parameter (see the discussion in Young and Pedregal, 1999a)

refer to the original un-transformed data (with the STAMP results simply transformed back into the original data form).

In these forecasting exercises, the last three years of data, from December 1957 to December 1960, are reserved for the comparative study. At each sample, all three models are re-optimised and estimated on the basis of the past data; and they are then used to obtain up to 24 step-ahead true, ex-ante forecasts. This process is then repeated, expanding the sample by one observation at each such step. The results of this exercise are shown in fig.8, where the overall forecasting performance measures, plotted as a function of the forecast lead time, are: the Mean Absolute Percentage Error (MAPE, upper plot); Percentage Root Mean Square Error (PRMSE, middle plot) and the Mean Percentage Error (MPE, lower plot). The solid lines show the variations in these measures for the original, simpler DHR model; the dashed lines represent the STAMP results, for the standard STAMP model; and the dot-dashed lines show the results for the improved DHR model, including the estimated 'business cycle'.

(INSERT FIGURE 8)

The three plots in fig.8 present very interesting results. Firstly, it is clear that the two DHR model forecasts, with and without the 51 month period cyclical component, are both significantly better than those obtained using STAMP over most, but not all, of the range of forecast lead times. Only at very small lead times are the STAMP results marginally better than the improved DHR model. Moreover, the superiority of the DHR model results increases substantially as the forecast horizon increases, especially in the case of the improved DHR model. These results demonstrate that improved one-step-ahead forecasting performance does not always imply better performance for longer forecast horizons. Secondly, judged by the MPE results, the distribution of the forecast errors is superior for the DHR results than for the STAMP results and the biases tend to be of opposite sign: STAMP tends to over-predict, while DHR tend to under-predict to a lesser extent at these larger forecast lead times.

Finally, a specific ex-ante, 24 step-ahead forecasting example is shown in fig.9, where the solid line represents the actual data; the stars are the forecasts from the improved DHR model, including the cyclical component; and the plus signs are the STAMP forecasts, with the STAMP standard error band shown dashed. It should be emphasised that these are true, 2 year ahead ex-ante forecasts: they are based only on the data up to the forecast origin in December 1957. It is clear from this figure that the superior performance of the extended DHR model is due mainly to the contribution that the longer term cycle makes to the forecasts.

(INSERT FIGURE 9)

Before completing this example, two caveats are necessary. First, the usefulness of a model with a longer period cycle included, such as that discussed above, is dependent upon how

well the characteristics of the cycle are defined by the data and the degree to which the period and phase of the cycle remain stationary over time into the future. The good forecasting performance in this particular AP example is obviously a consequence of the stationarity of the 51 month period cycle over the duration of the AP series. If this cyclical component is, indeed, a consequence of quasi-periodic ‘business cycle’ effects in the series, then it seems unlikely that such stationarity would be maintained over a much longer time interval into the future. Second, the results presented in this section are specific to the AP data: although the same kind of comparative performance has been experienced with other series so far (see e.g. Pedregal, 1995), it is difficult to generalise completely and, in other circumstances, our DHR model-based approach may not prove superior to other alternatives, such as STAMP, in these same respects. The approach is, however, more flexible in its formulation and implementation than many current alternatives and we would expect it to always be very competitive, if not superior.

7. CONCLUSIONS

This paper describes a new and flexible approach to nonstationary time series analysis based on a *Dynamic Harmonic Regression* (DHR) model of the Unobserved Components (UC) type. The basic DHR model allows for the presence of trend, seasonal, cyclical and irregular components and is formulated in a state space framework, which makes it particularly useful for adaptive seasonal adjustment, signal extraction, interpolation over gaps, variance intervention and forecasting (or backcasting) of nonstationary time series. The approach exploits the powerful properties of the Kalman Filter and Fixed Interval Smoothing algorithms to recursively estimate the various unobserved components, with the hyper-parameters of the DHR model optimised in a novel manner using a cost function defined in terms of the difference between the logarithmic pseudo-spectrum of the model and the logarithmic autoregressive spectrum of the data.

It is important to emphasise that this new approach to the optimisation of hyper-parameters is *not* based on Maximum Likelihood (ML) concepts. In all of the applications considered over the last five years, however, the cost function has been better defined than the alternative ML cost function and has yielded both improved convergence properties and considerably reduced numerical requirements. It would appear that, as in the AP example discussed here, the main advantage of the new approach is its improved multi-step ahead forecasting performance. In those cases where the a fixed order AR spectrum does not adequately accommodate both the seasonal and cyclical components, because the periods of the latter are too long, different order AR spectra are used to describe the lower and upper frequency ranges, and these are then concatenated to synthesise a composite spectrum, which forms the basis for the hyper-parameter optimisation.

Various extensions of the DHR model are possible. For example, the basic approach described in this paper can be extended to the full UC model (1), including the (possibly dynamic and nonlinear) effects of one or more input or exogenous variables. This kind of UC model has been used in the forecasting of electricity demand (Pedregal and Young, 1996, 1998a), river flows (Young *et al*, 1997) and hourly telephone numbers for Barclaycard plc. In the case where there are seasonal and input variable effects, the optimisation involves an iterative ‘back-fitting’ procedure in which, alternately, the periodic effects are optimised by our frequency domain method and the input transfer functions (which can involve ‘intervention’ variables) are estimated using optimal Instrumental Variable (IV) methods (see Young, 1984). Although we concentrate on DHR models in this paper, the analysis carried out has been extended to other types of periodic models, such as the ‘trigonometric seasonality’ model used by Harvey (1989) and others, where the seasonal components are defined within the state equations, rather than the observation equation. Indeed, such models can be combined with the DHR model to allow for ‘modulated cycles’ (Pedregal and Young, 1996) in which, for example, the seasonal cycle is modulated by a cycle of another, longer period.

Finally, if the UC model involves static or dynamic nonlinear effects, then a general approach (e.g. Young, 1993, 1998) involves the use of FIS and TVP estimation for non-parametric estimation of the nonlinearity. This is then followed by parameterisation of the nonlinearity, usually in state-dependent terms, and final parametric optimisation of this identified nonlinear UC model. Examples of this approach are described in Young and Beven (1994), who exploit it to identify and estimate rainfall-flow models; Young and Pedregal (1997, 1999b), who show that nonlinearly defined relativistic investment variables (relative to GNP) are more useful for forecasting unemployment rate in the USA than the standard investment variables; and Young (1998) who applies it to a variety of different systems.

ACKNOWLEDGEMENTS

The authors are grateful for the support of the UK Engineering & Physical Science and the Research Council under grant GR/J 10136

REFERENCES

- Akaike, H., ‘Seasonal adjustment by a Bayesian modelling’, Journal of Time Series Analysis, **1** (1980), 1-13.
- Box, G. E. P., and Jenkins, G. M., Time Series Analysis: Forecasting and Control, 1970; revised edn., San Francisco: Holden-Day, 1976.
- Box, G. E. P., Hillmer, S. C. and Tiao, G. C., ‘Analysis and modelling of seasonal time series’, in Zellner, A. (ed.), Seasonal Analysis of Economic time Series, Washington D. C.: US Dept. of Commerce-Bureau of the Census, 1978, 309-334

- Bryson, A. E. and Ho, Y. C., Applied Optimal Control, Optimization, Estimation and Control, Waltham: Blaisdell Publishing Company, 1969.
- Burman, J. P., 'Seasonal adjustment by signal extraction', Journal of the Royal Statistical Society, Series A, **143** (1980), 321-337.
- Dagum, E. B., The X-11 ARIMA Seasonal adjustment method, Statistics Canada: Ottawa, 1980.
- Dagum, E. B., The X-11/88 ARIMA seasonal adjustment method, Statistics Canada: Ottawa, 1988.
- Deutsch, R., Estimation Theory, Prentice Hall, N.J., 1965.
- Findley, D. F., Monsell, B. C., Otto, M. C., Bell, W. R. and Pugh, M., Towards X-12 ARIMA, mimeo, U.S. Bureau of the Census, 1992.
- Findley, D. F., Monsell, B. C., Bell, W. R., Otto, M. C. and Chen, B. C., New capabilities and methods of the X-12 ARIMA seasonal adjustment program, U.S. Bureau of the Census, mimeo, May 16, 1996.
- Gómez, V. and Maravall, A., Programs TRAMO and SEATS, Instructions for the user (Beta version: September 1996), working paper No. 9628, Bank of Spain, 1996.
- Harrison, P. J. and Stevens, C. F., 'Bayesian forecasting', Journal Royal Statistical Society, Series B., **38** (1976), 205-247.
- Harvey, A. C., 'A unified view of statistical forecasting procedures' (with comments), Journal of Forecasting, **3** (1984), 245-283.
- Harvey, A. C., Forecasting Structural Time Series Models and the Kalman Filter, Cambridge: Cambridge University Press, 1989.
- Hillmer, S. C. and Tiao, G. C., 'An ARIMA-model based approach to seasonal adjustment', Journal of the American Statistical Association, **77** (1982), 63-70.
- Hillmer, S. C., Bell, W. R. and Tiao, G. C., 'Modelling considerations in the seasonal adjustment of economic time series' in Zellner, A. (ed.), Applied Time Series Analysis of Economic Data, Washington D. C.: US Dept. of Commerce-Bureau of the Census, 1983, 74-100.
- Jakeman, A. J. and Young, P. C., 'Recursive filtering and the inversion of ill-posed causal problems', Utilitas Mathematica, **35** (1979, 1984), 351-376, (originally CRES Report No. AS/R28/1979, Centre for Resource and Environmental Studies, Australian National University, 1979)
- Kalman, R. E., 'A new approach to linear filtering and prediction problems', ASME Transactions Journal Basic Engineering, **83-D** (1960), 95-108.
- Koopmans, S. J., 'Disturbance smoother for state-space models', Biometrika, **76** (1993), 65-79.
- Koopmans, S. J., Harvey, A. C., Doornik, J. A. and Shephard, N., STAMP 5.0: Structural Time Series Analyser, Modeller and Predictor, London: Chapman & Hall, 1995.
- Maravall, A., 'Stochastic linear trends. models and estimators', Journal of Econometrics, **56** (1993), 5-37.
- Ng, C. N. and Young, P. C., 'Recursive estimation and forecasting of non-stationary time series', Journal of Forecasting, **9** (1990), 173-204.
- Norton, J. P., 'Optimal smoothing in the identification of linear time-varying systems', Proceedings Institute Electrical Engineers, **122** (1975), 663-668.

- Norton, J. P., An Introduction to Identification, London and New York: Academic Press, 1986.
- Pedregal, D. J., Comparación Teórica, Estructural y Predictiva de Modelos de Componentes no Observables y Extensiones del Modelo de Young, Ph.d. Thesis. Universidad Autónoma de Madrid, 1995.
- Pedregal, D. J. and Young, P. C., 'Modulated cycles, a new approach to modelling seasonal/cyclical behaviour in unobserved component models', CRES Tech. Note No. TR145, Centre for Research on Environmental Systems and Statistics (1996), Lancaster University.
- Pedregal, D. J. and Young, P.C., 'A new adaptive approach to forecasting electricity demand', CRES Tech. Note No. TR155, Centre for Research on Environmental Systems and Statistics (1998a), Lancaster University.
- Pedregal, D. J. and Young, P. C., 'Extensions of Trend Models in Unobserved Component Models', CRES Tech. Note No. TR156, Centre for Research on Environmental Systems and Statistics (1998b), Lancaster University.
- Schweppe, F., 'Evaluation of likelihood function for Gaussian signals', I.E.E.E. Transaction Information Theory, **11** (1965), 61-70.
- Shiskin, J., Young, A. and Musgrave, J. C., 'The X-11 variant of the Census method II seasonal adjustment program', Technical Paper 15, Washington: Bureau of the Census, 1967.
- West, M. and Harrison, J., Bayesian Forecasting and Dynamic Models, New York: Springer-Verlag, 1989.
- Pole, A., West, M. and Harrison, J., Applied Bayesian Forecasting and Time Series Analysis, New York: Chapman & Hall, 1994.
- Wiener, N., Extrapolation, Interpolation and Smoothing of Stationary Time Series, New York: J. Wiley, 1949.
- Young, P. C., 'Applying Parameter Estimation to Dynamic Systems', parts 1, and 2 Control Engineering, **16** (1969), 10/11, 119-125/118-124.
- Young, P. C., 'Recursive Approaches to Time-Series Analysis', Bull. of Inst. Maths and its Applications, **10** (1974), 209-224.
- Young, P. C., Recursive Estimation and Time-Series Analysis, Berlin: Springer-Verlag, 1984.
- Young, P. C., 'Recursive extrapolation, interpolation and smoothing of non-stationary time series', in Chen, C.F. (ed.), Identification and System Parameter Estimation, Oxford: Pergamon Press, 1988, 33-44.
- Young, P. C., 'Recursive estimation, forecasting and adaptive control', in Leondes, C. T. (ed.), Control and Dynamic Systems, Vol. 30, Part 3, San Diego: Academic Press, 1989, 119-165.
- Young, P. C., 'Comments on likelihood and cost as path integrals', Journal Royal Statistical Society, Series B., **53** (1991), 529-531.
- Young, P. C., 'Time variable and state dependent modelling of nonstationary and nonlinear time series', in Subba Rao, T. (ed.), Developments in Time Series Analysis, London: Chapman and Hall, 1993, 374-413.

- Young, P. C., 'Time-variable parameter and trend estimation in non-stationary economic time series', Journal of Forecasting, **13** (1994), 179-210.
- Young, P. C., 'Data-based mechanistic modelling of environmental, ecological, economic and engineering systems', Environmental Modelling and Software, **13** (1998), 105-122.
- Young, P. C. and Benner, S., microCAPTAIN Handbook: Version 2.0, Centre for Research on Environmental Systems and Statistics, Lancaster University, 1991.
- Young, P. C. and Beven, K. J., 'Data-based mechanistic modelling and the rainfall-flow nonlinearity', Environmetrics, **5** (1994), 335-363.
- Young, P. C. and Ng, C. N., 'Variance Intervention', Journal of Forecasting, **8** (1989), 399-416.
- Young, P. C. and Pedregal, D. J., 'Data-based mechanistic modelling', in Heij, C., Hanzon, B. and Praagman, K. (eds.), System Dynamics in Economic and Financial Models, J. Wiley: Chichester, 1997, 169-213.
- Young, P. C. and Pedregal, D. J., 'Recursive and en-block approaches to signal extraction', Journal of Applied Statistics, **26** (1999a), 103-128.
- Young, P. C. and Pedregal, D. J., 'Macro-economic relativity: government spending, private investment and unemployment in the USA 1948-1998' (in press: *Structural Change and Economic Dynamics*, 1999b).
- Young, P. C., Jakeman, A. J. and Post, D., 'Recent Advances in Data-based Modelling and Analysis of Hydrological Systems', Water Science and Technology, **28**, (1997) 1573-1584.
- Young, P. C., Ng, C. N. and Armitage, P., '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** (1989), 481-501.

Appendix: Estimation of the State Space Model Disturbances

Following FIS estimation, the smoothed estimate $\hat{e}_{t|N}$ of observation noise is found trivially by subtracting the smoothed output estimate $\hat{y}_{t|N}$ from the observations y_t . A smoothed estimate $\hat{\eta}_{t|N}$ of the state disturbance vector is obtained by comparing,

$$\hat{\mathbf{x}}_{t+1|N} = \mathbf{F}\hat{\mathbf{x}}_{t|N} + \mathbf{G}\hat{\eta}_{t|N}$$

with the smoothed state update equation (4b),

$$\hat{\mathbf{x}}_{t+1|N} = \mathbf{F}\hat{\mathbf{x}}_{t|N} - \mathbf{G}\mathbf{Q}_r\mathbf{G}^T\mathbf{L}_t$$

This then yields the following relationship between $\hat{\eta}_{t|N}$ and the Lagrange Multiplier vector \mathbf{L}_t ,

$$\hat{\eta}_{t|N} = -\mathbf{Q}_r\mathbf{G}^T\mathbf{L}_t$$

which is easily computable.

It is interesting to compare our implementation of the FIS algorithm with that suggested by Koopmans (1993), which the author claims to be the most numerically efficient approach to FIS estimation. The main iteration in the Koopmans smoother is that on the vector \mathbf{r}_t , where,

$$\mathbf{r}_{t-1} = (\mathbf{F} - \mathbf{K}_t\mathbf{H}_t)^T\mathbf{r}_t + \mathbf{H}_t^T D_t^{-1}\tilde{y}_t$$

where $D_t = 1 + \mathbf{H}_t\mathbf{P}_{t|t-1}\mathbf{H}_t^T$ (scalar in this case), $\mathbf{F}\mathbf{K}_t$ (where \mathbf{K}_t is the Kalman gain) and \tilde{y}_t (the KF innovation) are calculated and stored during the forward KF pass. The smoothed estimate of the disturbance is then calculated as $\hat{\eta}_{t|N} = \mathbf{Q}_r\mathbf{r}_t$. However, if \mathbf{G} considered as an identity matrix, which is not a restrictive assumption in this context, then $\mathbf{r}_t = -\mathbf{L}_t$ and so the FIS algorithm used in the present paper, which is based on the algorithm suggested originally by Bryson and Ho (1969), turns out to be algebraically equivalent to the Koopmans algorithm. Moreover, the complexity of the main iteration in the Koopmans smoothing algorithm is similar to the main iteration (the Lagrangian update) of the algorithm employed here. In particular, the Koopmans algorithm, as published, requires a very similar number of floating point operations per iteration: $(2\dots3)n^2 + (5\dots8)n$, where n is the size of the state vector and the ranges (2...3) and (5..8) result from various possible structures of \mathbf{F} and \mathbf{Q}_r matrices. Consequently, the speed of execution will be very similar and will depend on programming trade-offs (e.g.. memory vs. speed) and similar technical modifications.

FIGURES

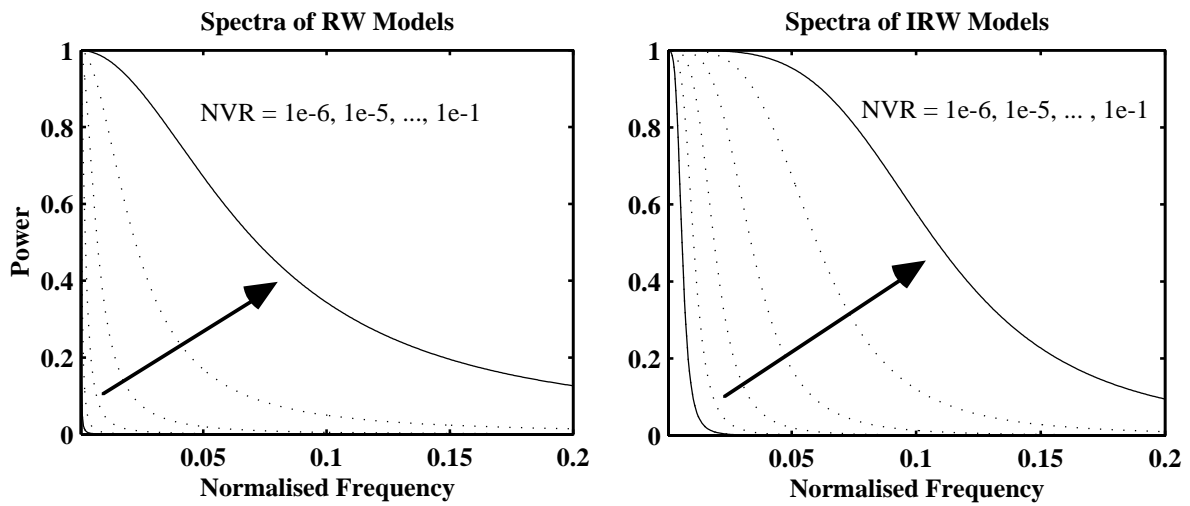


Figure 1: Spectral characteristic of RW and IRW filters for different values of the NVR parameter.

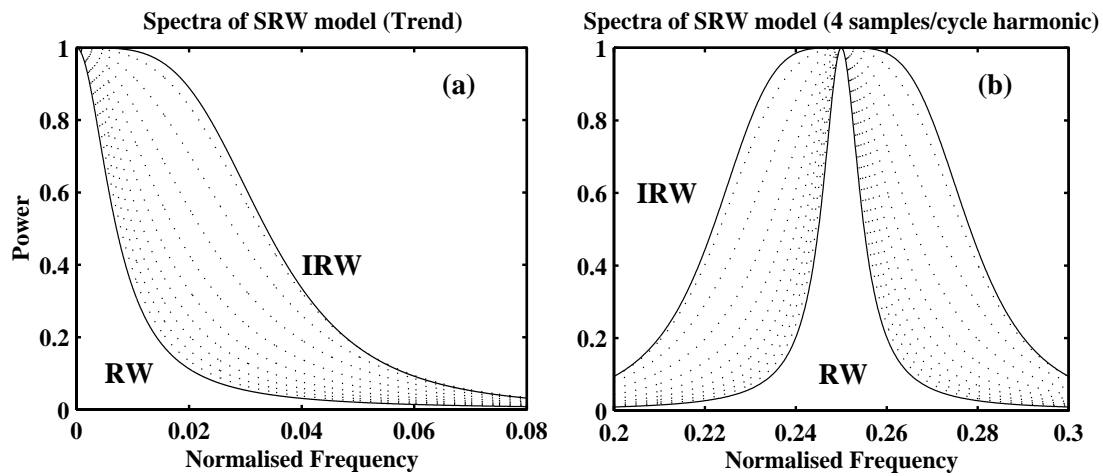


Figure 2: Spectral characteristic of SRW for different values of α for : (a) a trend described by an SRW process; and (b) a harmonic with period 4 samples/cycle modulated by a TVP parameter described by an SRW process.

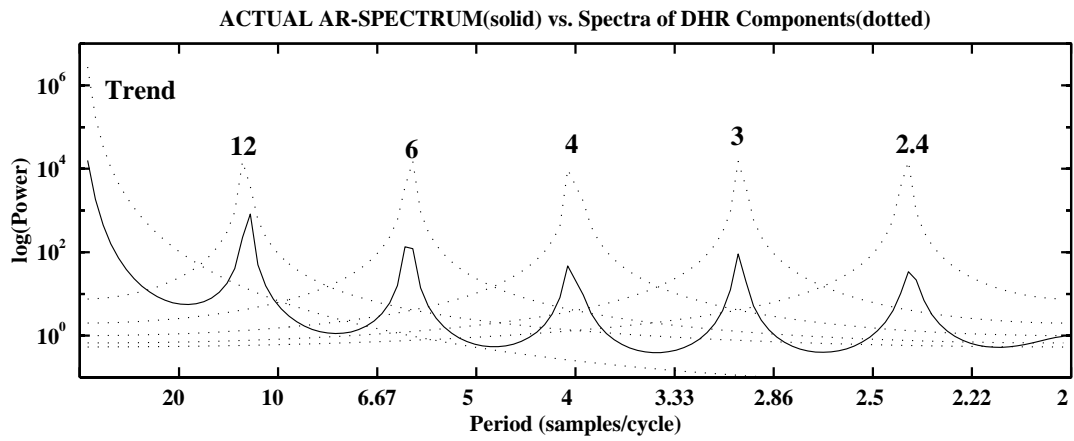


Figure 3: Empirical AR(14) spectrum (solid) and pseudo-spectra of DHR components (dotted) for the Airline Passenger (AP) series.

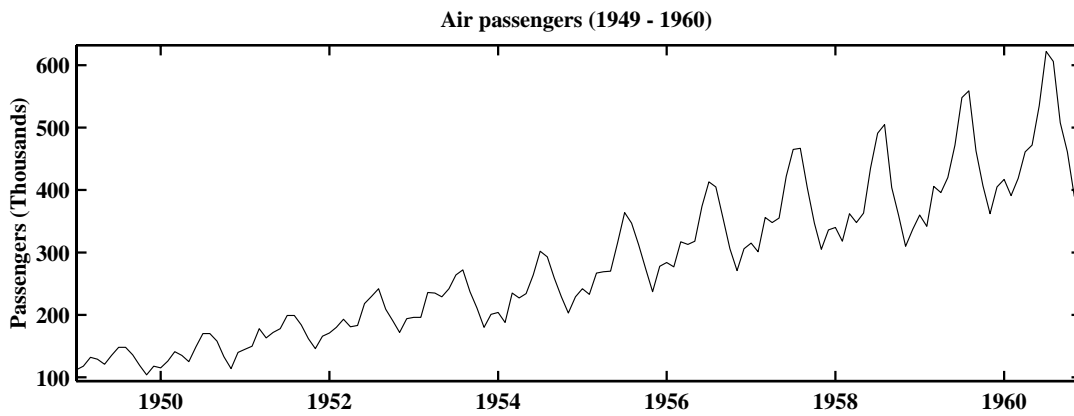


Figure 4: AP series taken from Box and Jenkins (1976).

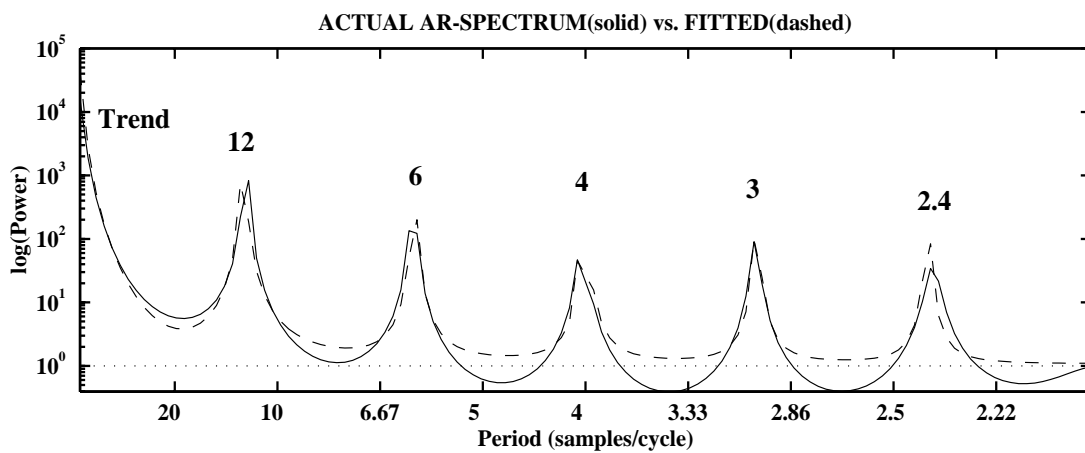


Figure 5: AR(14) spectrum (solid) vs. the DHR fitted spectrum (dashed) for the AP series.

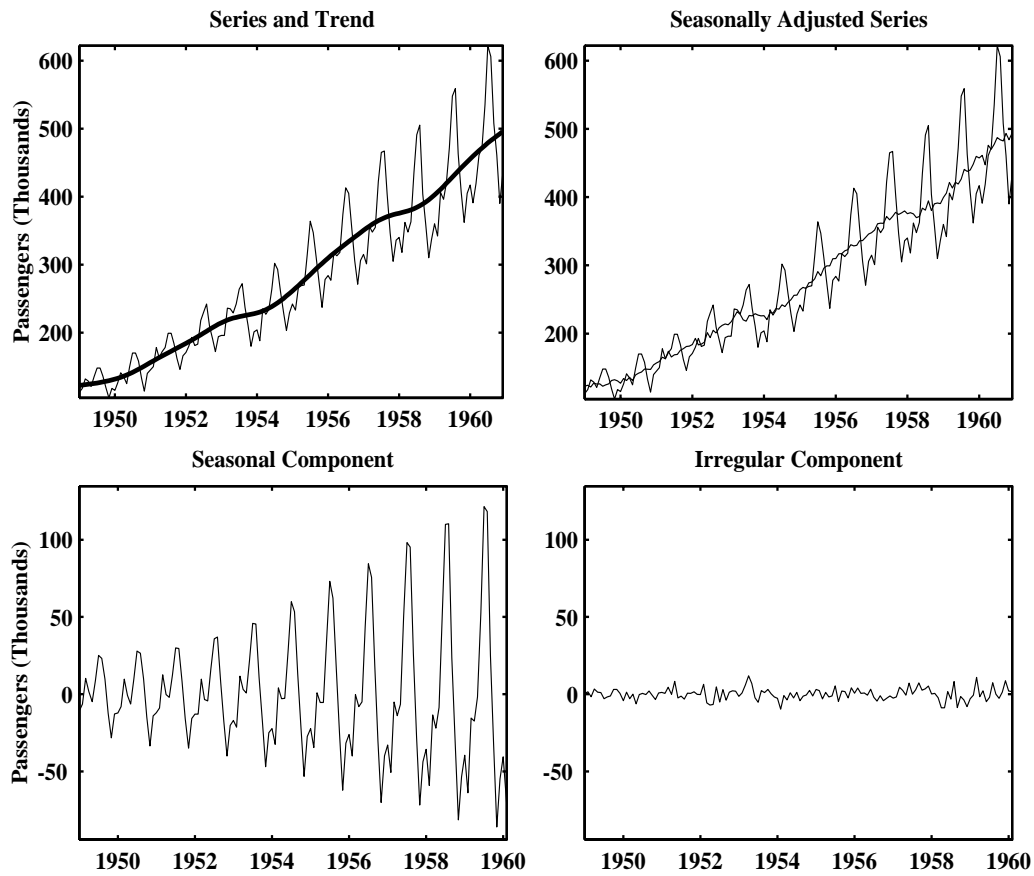


Figure 6: The estimated unobserved components for the AP series.

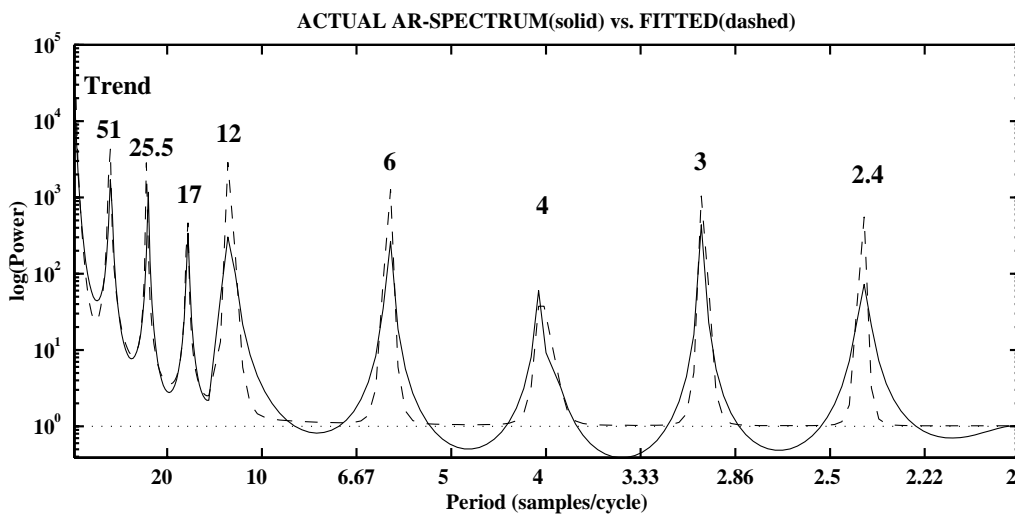


Figure 7: Optimised DHR model fit to the concatenated composite AR spectrum of the AP data: in addition to the standard seasonal components associated with the annual cycle, the model includes a medium term cycle of period 51 samples/cycle and two of its harmonics.

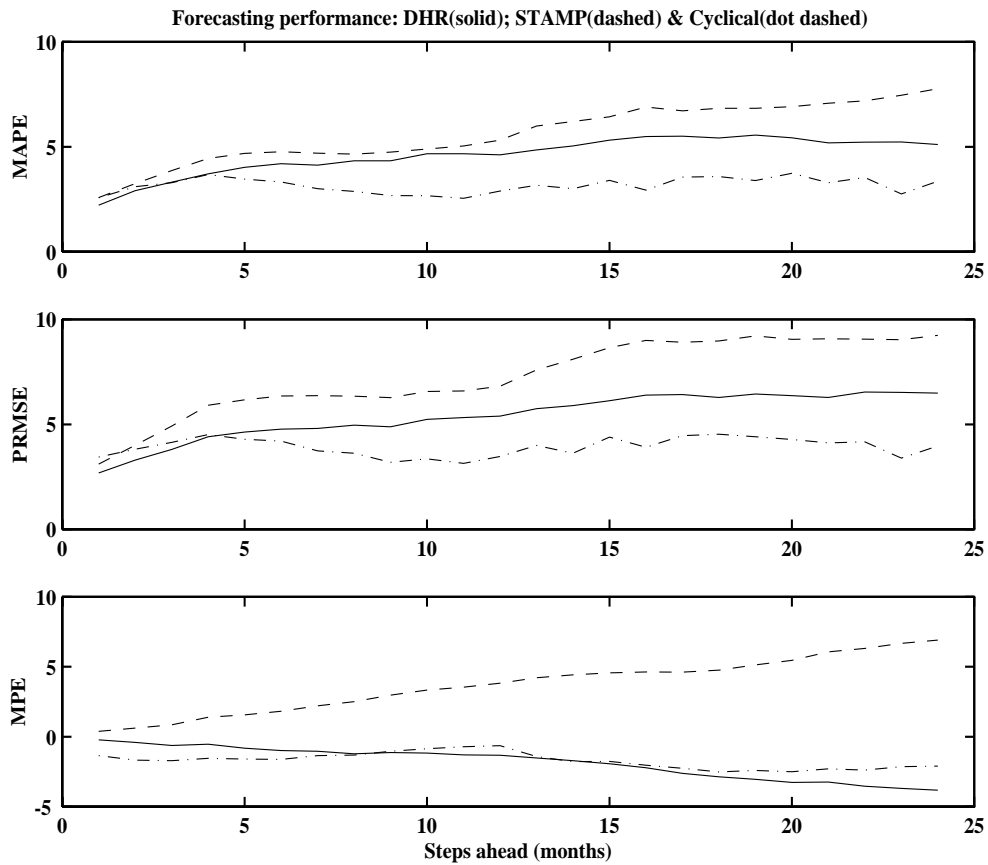


Figure 8: Relative forecasting performance for the AP series using different measures: Mean Absolute Percentage Error (MAPE, upper plot); Percentage Root Mean Square Error (PRMSE, middle plot) and the Mean Percentage Error (MPE, lower plot). Standard seasonal DHR model results (solid); STAMP results (dashed); and DHR seasonal plus cyclical model results (dot-dashed). The two DHR models have LLT; RW cyclical and IRW seasonal model parameters; while the STAMP results are obtained from the analysis of logarithmically transformed data using the Basic Structural Model (BSM: see Koopmans *et al*, 1995).

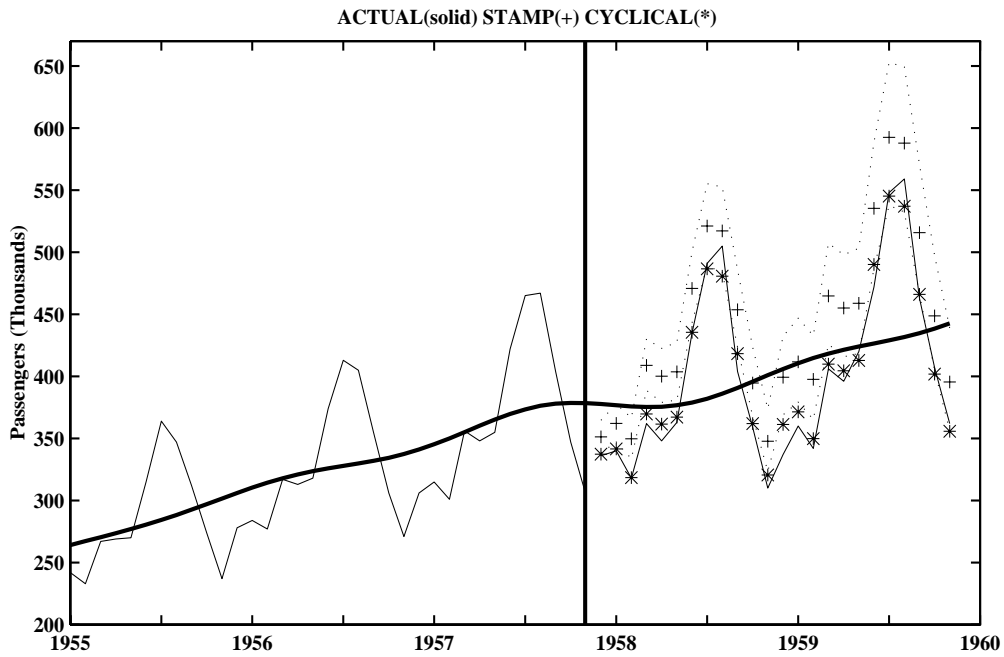


Figure 9: Relative 24 month ahead, ex-ante forecasting performance over the latter two years of the AP series: actual data (solid); STAMP results (+); and seasonal plus cyclical DHR model results (*). The graph also shows the standard errors of STAMP forecasts (dotted) and the DHR model trend-plus-cycle component (thick solid).

TABLES 1 to 4

Period	Linear	Non Linear	Constrained ML
Trend	4.41e-6 5.43e-1 (-) (4.27)	3.79e-17 5.64e-01 (-) (4.019)	8.99e-3 1.00e-3 (52.09) (40.82)
12	1.36e-6 (31.61)	9.349e-06 (27.119)	4.15e-6 (40.44)
6	1.18e-6 (34.00)	4.072e-06 (30.927)	»
4	1.21e-5 (31.88)	1.167e-05 (32.002)	»
3	1.95e-6 (35.55)	4.200e-06 (33.488)	»
2.4	3.25e-7 (31.46)	2.087e-06 (27.551)	»
$\hat{\sigma}_a^2$	127.39	123.66	140.73
Log-Likelihood	-324.126	-289.488	-293.781
Q(12)	13.99	13.22	18.33
Jarque-Bera	1.161	2.302	2.665
Mega-Flops		0.450	82.057

Table 1: Estimation results for AP series based on the analysis of the original un-transformed data and LLT trend: the numbers in brackets are the t-ratios; the ‘»’ symbol means that the NVR is the same for all other harmonic components as the harmonic of period 12 (constrained optimisation); $\hat{\sigma}_a^2$ is the variance of the innovations; Q(12) is the Ljung-Box statistics for whiteness of the innovations with 12 lags; H(50) is a standard ratio of variances test for the detection of heteroskedasticity (the distribution of the statistic is $F_{(50, N-50)}$); and Jarque-Bera is a Normality test. Note that convergence did not occur at all in the unconstrained ML case, so no results are shown.

Period	Linear	Non Linear	Constrained ML	ML
Trend	5.805e-03 (25.991)	1.453e-02 (27.052)	8.326e-03 (8.036)	8.248e-03 (7.923)
12	3.309e-02 (7.052)	4.220e-02 (21.853)	5.027e-03 (12.817)	1.675e-02 (6.395)
6	5.903e-02 (12.578)	1.482e-02 (24.239)	»	6.315e-03 (6.534)
4	2.212e-02 (0.745)	9.513e-03 (20.519)	»	7.663e-5** (7.115)
3	7.448e-03 (1.587)	7.093e-03 (23.738)	»	1.497e-03 (4.880)
2.4	1.878e-03 (0.400)	5.705e-03 (22.513)	»	7.663e-05** (7.115)
$\hat{\sigma}_a^2$	1.637e-03	1.481e-03	1.598e-03	1.450e-03
Log-Likelihood	361.353	363.161	360.934	368.093
Q(12)	35.289	18.196	22.124	14.399
Jarque-Bera	0.397	0.915	0.416	0.119
Mega-Flops		0.232	26.944	120.455

Table 2: Estimation results for AP series based on the analysis of the logarithmically transformed data and IRW trend (other information as for Table 1). Note that the ‘**’ symbol indicates constrained parameters: convergence did not occur in the completely unconstrained ML case but convergence was possible if the two NVR’s at periods of 4 and 2.4 months were constrained to be equal.

Period	Raw Data (LLT)	Raw Data (IRW)	Data in logs (LLT)	Data in logs (IRW)
Trend	3.79e-17 5.64e-01 (-) (4.019)	5.563e-03 (20.760)	4.69e-04 3.11e-01 (11.126) (6.640)	1.453e-02 (27.052)
12	9.349e-06 (27.119)	3.035e-05 (28.919)	3.891e-02 (20.778)	4.220e-02 (21.853)
6	4.072e-06 (30.927)	6.032e-06 (29.339)	1.483e-02 (24.420)	1.482e-02 (24.239)
4	1.167e-05 (32.002)	1.352e-05 (29.442)	9.071e-03 (21.106)	9.513e-03 (20.519)
3	4.200e-06 (33.488)	4.764e-06 (30.289)	8.511e-03 (23.540)	7.093e-03 (23.738)
2.4	2.087e-06 (27.551)	2.547e-06 (25.081)	5.898e-03 (22.642)	5.705e-03 (22.513)
$\hat{\sigma}_a^2$	123.66	139.39	1.379e-03	1.481e-03
Log-Likelihood	-289.488	-293.999	367.330	360.313
Constrained Lik.	-293.781	-294.180	365.800	360.934
Q(12)	14.27	24.67	9.512	18.19
Jarque-Bera	3.211	1.411	3.201	0.915

Table 3: Results for the AP series using the un-transformed and logarithmically transformed series, with different assumptions for the trend model. The row ‘Constrained Lik.’ denotes the value of the Log-Likelihood function when ML optimisation was performed with all of the NVR’s for the seasonal harmonics constrained to be the same and is included for comparison with the fully unconstrained results (other information as for Table 1).

Component	Raw Data (LLT)	Raw Data (IRW)	Data in logs (LLT)	Data in logs (IRW)
Level	15.579 (‘concentrated’)	0 (‘zero’)	2.982e-4 (0.364)	0 (‘zero’)
Slope	1.112e-02 (4.12)	0.701 (1.11)	0 (‘zero’)	8.757e-6 (6.206)
Seasonal	1.221 (2.80)	1.471 (‘concentrated’)	3.558e-6 (6.282)	3.828e-6 (6.013)
Irregular	0 (‘zero’)	0 (‘zero’)	2.343e-4 (‘concentrated’)	4.658e-4 (‘concentrated’)
$\hat{\sigma}_a^2$	157.94	173.47	1.477e-03	1.635e-03
Q(12)	45.136	45.66	9.622	22.641
Jarque-Bera	0.370	0.983	1.730	0.873

Table 4: STAMP program results for the AP series. The numbers in brackets are the t-ratios; other information as for Table 1.