Statistical Signal Extraction and Filtering: Structual Time Series Models

D.S.G. Pollock
University of Leicester
Email: stephen_pollock@sigmapi.u-net.com

Structural Time Series Models

In economics, it is traditional to decompose time series into a variety of components, some or all of which may be present in a particular instance.

One is liable to assume that the relative proportions of the components of an aggregate index are maintained, approximately, in spite of the variations in their levels. Therefore, the basic model of an economic index is a multiplicative one; and, if $Y(t)$ is the sequence of values of an economic index, then it can be expressed as

$$Y(t) = L(t) \times C(t) \times S(t) \times H(t), \tag{1}$$

where

$L(t)$	is the global trend,
$C(t)$	is a secular cycle,
$S(t)$	is the seasonal variation and
$H(t)$	is an irregular component.

Many of the more prominent macroeconomic indicators are amenable to a decomposition of this sort. One can imagine, for example, a quarterly index of Gross Domestic Product which appears to be following an exponential growth trend $L(t)$.

The trend might be obscured, to some extent, by a superimposed cycle $C(t)$ with a period of roughly four and a half years, which happens to correspond, more or less, to the average lifetime of the legislative assembly. The reasons for this curious coincidence need not concern us here.

The ghost of an annual cycle $S(t)$ might also be apparent in the index; and this could be a reflection of the fact that some economic activities, such as building construction, are affected significantly by the weather and by the duration of sunlight.

When the foregoing components—the trend, the secular cycle and the seasonal cycle—have been extracted from the index, the residue should correspond to an irregular component $H(t)$ for which no unique explanation can be offered.

The logarithms $y(t) = \ln Y(t)$ of the aggregate index are amenable to an additive decomposition. Thus, equation (1) gives rise to

$$\begin{aligned} y(t) &= \{\lambda(t) + \gamma(t)\} + \sigma(t) + \eta(t) \\ &= \tau(t) + \sigma(t) + \eta(t), \end{aligned} \tag{2}$$

where $\lambda(t) = \ln L(y)$, $\gamma(t) = \ln C(t)$, $\sigma(t) = \ln S(t)$ and $\eta(t) = \ln H(t)$. Since the trend and the cycles are not easily separable, there is a case for combining them in a component $T(t) = L(t) \times C(t)$, of which the logarithm is $\ln T(t) = \tau(t)$.

In the structural time-series model, the additive components are modelled by independent ARMA or ARIMA process. Thus

$$\begin{aligned} y(z) &= \tau(z) + \sigma(z) + \eta(z) \\ &= \frac{\theta_\tau(z)}{\phi_\tau(z)}\zeta_\tau(z) + \frac{\theta_\sigma(z)}{\phi_\sigma(z)}\zeta_\sigma(z) + \eta(z), \end{aligned} \tag{3}$$

where $\zeta_\tau(z)$, $\zeta_\sigma(z)$ and $\eta(z)$ are the z-transforms of statistically independent white-noise processes. Within the autoregressive polynomial $\phi_\tau(z)$ of the trend component will be found the unit-root factor $(1 - z)^p$, whereas the autoregressive polynomial $\phi_\sigma(z)$ of the seasonal component will contain the factor $(1 + z + \cdots + z^{s-1})^D$, wherein s stands for the number of periods in a seasonal cycle.

The sum of a set of ARIMA processes is itself and ARIMA process. Therefore, $y(t)$ can be expressed as a univariate ARIMA process which is described as the reduced form of the time-series model:

$$y(z) = \frac{\theta(z)}{\phi(z)}\varepsilon(z) = \frac{\theta(z)}{\phi_\sigma(z)\phi_\tau(z)}\varepsilon(z). \tag{4}$$

Here, $\varepsilon(z)$ stands for the z-transform of a synthetic white-noise process.

There are two alternative approaches to the business of estimating the structural model and of extracting its components. The first approach, which is described as the canonical approach, is to estimate the parameters of the reduced-form ARIMA model. From these parameters, the Wiener–Kolmogorov filters that are appropriate for extracting the components can be constructed.

Canonical Decompositions

On the assumption that the degree of the moving-average polynomial $\theta(z)$ is at least equal to that of the autoregressive polynomial $\phi(z)$, there is a partial-fraction decomposition of the autocovariance generating function of the model into three components, which correspond to the trend effect, the seasonal effect and an irregular influence. Thus

$$\frac{\theta(z)\theta(z^{-1})}{\phi_\sigma(z)\phi_\tau(z)\phi_\tau(z^{-1})\phi_\sigma(z^{-1})} = \frac{Q_\tau(z)}{\phi_\tau(z)\phi_\tau(z^{-1})} + \frac{Q_\sigma(z)}{\phi_\sigma(z)\phi_\sigma(z^{-1})} + R(z). \tag{5}$$

Here, the first two components on the RHS represent proper rational fractions, whereas the irregular component $R(z)$ is an ordinary polynomial. If the degree of the moving-average polynomial in the reduced form is less than that of the autoregressive polynomial, then the irregular component is missing from the decomposition in the first instance.

To obtain the spectral density function $f(\omega)$ of $y(t)$ and of its components, we set $z = e^{-i\omega}$ in (5). (This function is more properly described as a pseudo-spectrum in view of the singularities occasioned by the unit roots in the denominators of the first two components.) The spectral decomposition can be written as

$$f(\omega) = f_\tau(\omega) + f_\sigma(\omega) + f_R(\omega). \tag{6}$$

Let $\nu_\tau = \min\{f_\tau(\omega)\}$ and $\nu_\sigma = \min\{f_\sigma(\omega)\}$. These are the elements of white noise embedded in $f_\tau(\omega)$ and $f_\sigma(\omega)$. The principle of canonical decomposition is that the white-noise elements should be reassigned to the residual component. (The principle of canonical decompositions has been expounded, for example, by Hillmer and Tiao (1982), Maravall and Pierce (1987), and, more recently, Kaiser and Maravall (2001).) On defining

$$\begin{aligned} \gamma_\tau(z)\gamma_\tau(z^{-1}) &= Q_\tau(z) - \nu_\tau\phi_\tau(z)\phi_\tau(z^{-1}), \\ \gamma_\sigma(z)\gamma_\sigma(z^{-1}) &= Q_\sigma(z) - \nu_\sigma\phi_\sigma(z)\phi_\sigma(z^{-1}), \\ \text{and} \quad \rho(z)\rho(z^{-1}) &= R(z) + \nu_\tau + \nu_\sigma, \end{aligned} \tag{7}$$

the canonical decomposition of the generating function can be represented by

$$\frac{\theta(z)\theta(z^{-1})}{\phi(z)\phi(z^{-1})} = \frac{\gamma_\tau(z)\gamma_\tau(z^{-1})}{\phi_\tau(z)\phi_\tau(z^{-1})} + \frac{\gamma_\sigma(z)\gamma_\sigma(z^{-1})}{\phi_\sigma(z)\phi_\sigma(z^{-1})} + \rho(z)\rho(z^{-1}). \tag{8}$$

There are now two improper rational functions on the RHS, which have equal degrees in their numerators and denominators.

According to Wiener–Kolmogorov theory, the optimal signal-extraction filter for the trend component is

$$\begin{aligned} \beta_\tau(z) &= \frac{\gamma_\tau(z)\gamma_\tau(z^{-1})}{\phi_\tau(z)\phi_\tau(z^{-1})} \times \frac{\phi_\sigma(z)\phi_\tau(z)\phi_\tau(z^{-1})\phi_\sigma(z^{-1})}{\theta(z)\theta(z^{-1})} \\ &= \frac{\gamma_\tau(z)\gamma_\tau(z^{-1})\phi_\sigma(z)\phi_\sigma(z^{-1})}{\theta(z)\theta(z^{-1})}. \end{aligned} \tag{9}$$

This has the form of the ratio of the autocovariance generating function of the trend component to the autocovariance generating function of the process $y(t)$.

Observe that, in the process of forming this filter, the factor $\phi_\tau(z)\phi_\tau(z^{-1})$ is cancelled out. With the consequent removal of the unit-root factor $(1 - z)^p(1 - z^{-1})^p$ from the denominator, the basis of a stable filter is created which, with the provision of appropriate starting values, can be applied to nonstationary data. This filter would also serve to extract a differenced version of the component $\tau(t)$ from the differenced data. The filter that serves to extract the seasonal component is of a similar construction.

These formulations presuppose a doubly-infinite data sequence; and they must be translated into a form that can be implemented with finite sequences. The various ways of achieving this have been described in he accompanying paper; In the TRAMO–SEATS program of Gómez and Maravall (1996) and of Caporello and Maravall (2004), a contragrade method of Burman (1980) has been adopted, which entails a unique treatment of the start-up problem.

The alternative method of estimating the parameters of the structural model and of extracting the unobserved components makes use of the fact that a univariate autoregressive moving-average model can be expressed as a first-order multivariate Markov model, which constitutes a state-space representation of the model. This allows the structural parameters to be estimated directly, as opposed to being inferred indirectly from the parameters of the reduced-form model.

The state-space approach to the structural time-series model was pioneered by Harrison and Stevens (1971, 1976). An extensive account of the approach has been provided by Harvey (1989). Other important references are the books of West and Harrison (1997) and Kitagawa and Gersch (1996). Proietti (2002) has also provided a brief but thorough account. A brief introductory survey has been provided by West (1997), and an interesting biomedical application has been demonstrated by West *et al.* (1999).

The methods may be illustrated by considering the so-called basic structural model, which has been popularised by Harvey (1989). The model, which lacks a non-seasonal cyclical component, can be subsumed under the second of the equations of (2).

The trend or levels component $\tau(t)$ of this model is described by a stochastic process that generates a trajectory that is approximately linear within a limited locality. Thus

$$\tau(t) = \tau(t-1) + \beta(t-1) + \upsilon(t) \quad \text{or, equivalently,} \qquad (10)$$
$$\nabla(z)\tau(z) = z\beta(z) + \upsilon(z),$$

where $\nabla(z) = 1 - z$ is the difference operator. That is to say, the change in the level of the trend is compounded from the slope parameter $\beta(t-1)$, generated in the previous period, and a small white-noise disturbance $\upsilon(t)$. The slope parameter follows a random walk. Thus

$$\beta(t) = \beta(t-1) + \zeta(t) \quad \text{or, equivalently,} \quad \nabla(z)\beta(z) = \zeta(z), \qquad (11)$$

where $\zeta(t)$ denotes a white-noise process that is independent of the disturbance process $\upsilon(t)$. By applying the difference operator to equation (10) and substituting from (11), we find that

$$\begin{aligned}\nabla^2(z)\tau(z) &= \nabla(z)z\beta(z) + \nabla(z)\upsilon(z) \\ &= z\zeta(z) + \nabla(z)\upsilon(z).\end{aligned} \qquad (12)$$

The two terms of the RHS can be combined to form a first-order moving-average process, whereupon the process generating $\tau(t)$ can be described by an integrated moving-average IMA(2, 1) model. Thus

$$\begin{aligned}\nabla^2(z)\tau(z) &= z\zeta(z) + \nabla(z)\upsilon(z) \\ &= (1-\mu z)\varepsilon(z).\end{aligned} \tag{13}$$

A limiting case arises when the variance of the white-noise process $\zeta(t)$ in equation (11) tends to zero. Then, the slope parameter tends to a constant β, and the process by which the trend is generated, which has been identified as an IMA(2,1) process, becomes a random walk with drift.

Another limiting case arises when the variance of $\upsilon(t)$ in equation (10) tends to zero. Then, the overall process generating the trend becomes a second-order random walk, and the resulting trends are liable to be described as smooth trends. When the variances of $\zeta(t)$ and $\upsilon(t)$ are both zero, then the process $\tau(t)$ degenerates to a simple linear time trend.

The seasonal component of the structural time-series model is described by the equation

$$\sigma(t) + \sigma(t-1) + \cdots + \sigma(t-s+1) = \omega(t) \tag{14}$$

or, equivalently,

$$S(z)\sigma(z) = \omega(z),$$

where $S(z) = 1 + z + z^2 + \cdots + z^{s-1}$ is the seasonal summation operator, s is the number of observation per annum and $\omega(t)$ is a white-noise process.

The equation implies that the sum of s consecutive values of this component will be a random variable distributed about a mean of zero. To understand this construction, we should note that, if the seasonal pattern were perfectly regular and invariant, then the sum of the consecutive values would be identically zero. Since the sum is a random variable with a zero mean, some variability can occur in the seasonal pattern.

By substituting equations (12) and (14) into equation (2), we seen that the structural model can be represented by the equation

$$\nabla^2(z)S(z)y(z) = S(z)z\zeta(z) + \nabla(z)S(z)\upsilon(z) + \nabla^2(z)\omega(z) + \nabla^2(z)S\eta(z),$$

or, equivalently, (15)

$$\nabla(z)\nabla_s(z)y(z) = S(z)z\zeta(z) + \nabla_s(z)\upsilon(z) + \nabla^2(z)\omega(z) + \nabla(z)\nabla_s(z)\eta(z),$$

where $\zeta(t)$, $\upsilon(t)$, $\omega(t)$ and $\eta(t)$ are mutually independent white-noise processes. Here, the alternative expression comes from using the identity

$$\nabla(z)S(z) = (1-z)(1+z+\cdots+z^{s-1}) = (1-z^s) = \nabla_s(z).$$

We should observe that the RHS or equation (15) corresponds to a moving average of degree $s+1$, which is typically subject to a number of restriction on its parameters. The restrictions arise from the fact there are only four parameters in the model of (15), which are the white-noise variances $V\{\zeta(t)\}$,

$V\{\upsilon(t)\}$, $V\{\omega(t)\}$ and $V\{\eta(t)\}$, whereas there are $s+1$ moving-average parameters and a variance parameter in the unrestricted reduced-form of the seasonal ARIMA model.

The basic structural model can be represented is a state-space form which comprises a transition equation, which constitutes a first-order vector autoregressive process, and an accompanying measurement equation. For notational convenience, let $s = 4$, which corresponds to the case of quarterly observations. Then, the transition equation, which gathers together equations (10), (11) and (14), is

$$\begin{bmatrix} \tau(t) \\ \beta(t) \\ \sigma(t) \\ \sigma(t-1) \\ \sigma(t-2) \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & -1 & -1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \tau(t-1) \\ \beta(t-1) \\ \sigma(t-1) \\ \sigma(t-2) \\ \sigma(t-3) \end{bmatrix} + \begin{bmatrix} \upsilon(t) \\ \zeta(t) \\ \omega(t) \\ 0 \\ 0 \end{bmatrix}. \tag{16}$$

The observation equation, which corresponds to (2), is

$$y(t) = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} \tau(t) \\ \beta(t) \\ \sigma(t) \\ \sigma(t-1) \\ \sigma(t-2) \end{bmatrix} + \eta(t). \tag{17}$$

The state-space model is amenable to the Kalman filter and the associated smoothing algorithms, which can be used in estimating the parameters of the model and in extracting estimates of the unobserved components $\tau(t)$, $\sigma(t)$.

There are various ways of handling, within the context of the Kalman filter, the start-up problem that is associated with filtering of nonstationary data sequences. These will be touched upon at the end of the next section.

Example. Figure 1 shows the logarithms of a monthly sequence of 132 observations of the U.S. money supply, through which a quadratic function has been interpolated. This provides a simple way of characterising the growth over the period in question.

However, it is doubtful whether such an analytic function can provide an adequate representation of a trend that is subject to irregular variations; and we prefer to estimate the trend more flexibly by applying a linear filter to the data. In order to devise an effective filter, it is helpful to know the extent of the frequency band in which the spectral effects of the trend are located.

It is difficult to discern the spectral structure of the data in the periodogram of the trended sequence y. This is dominated by the effects of the disjunctions in the periodic extension of the data that occur where the end of one replication of the data sequence joins the beginning of the next. In fact, the periodic extension of a segment of a linear trend will generate a sawtooth function, of which the periodogram will have the form of a rectangular hyperbola, within which any finer spectral detail will be concealed.

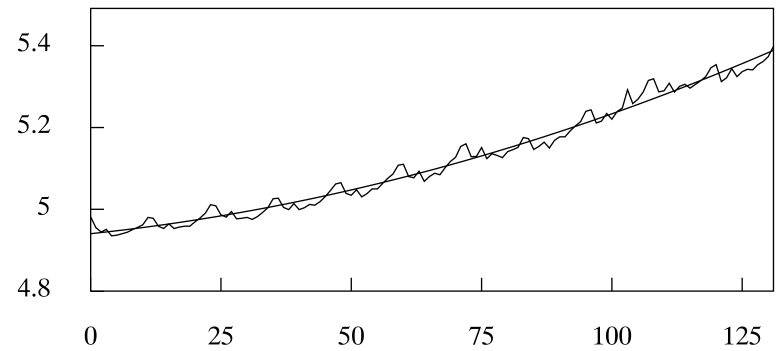

Figure 1: The plot of 132 monthly observations on the U.S. money supply, beginning in January 1960. A quadratic function has been interpolated through the data.

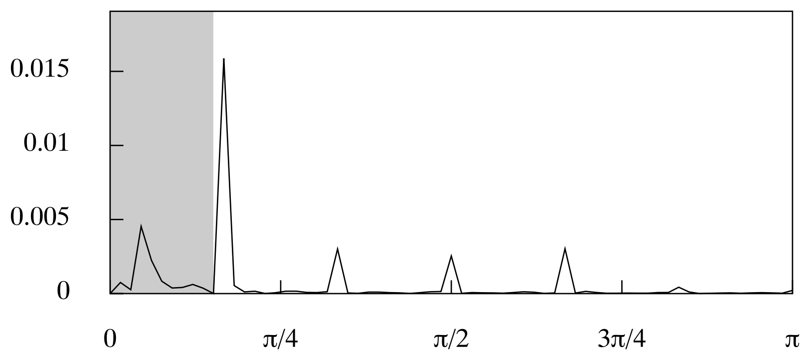

Figure 2: The periodogram of the residuals of the logarithmic money-supply data.

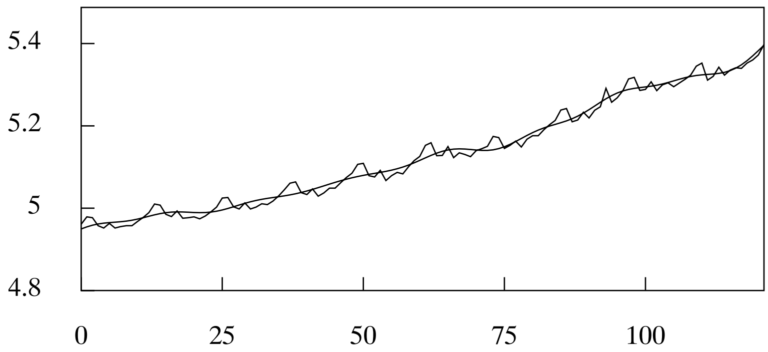

Figure 3: logarithms of 132 monthly observations on the U.S. money supply, beginning in January 1960. A trend, estimated by the Fourier method, has been interpolated through the data.

On the other hand, if a d-fold differencing operation is used to reduce the data to stationarity to produce $g = Qy$, then one may find that the low-frequency spectral ordinates have been diminished to such an extent that the structure of the trend has become invisible. The problem will to be exacerbated when the data contain a strong seasonal component, which may be amplified by the differencing operation to become the dominant feature of the periodogram.

An effective way of discerning the spectral structure of the data is to examine the periodograms of the residuals obtained by fitting polynomials of various degrees to the data. The residual sequence from fitting a polynomial of degree d, can expressed as

$$r = Q(Q'Q)^{-1}Q'y, \tag{18}$$

where Q' is the aforementioned differencing operator. This sequence contains the same information as the differenced sequence $g = Q'y$, but its periodogram renders the spectral structure visible over the entire frequency range.

Figure 2 which shows the periodogram of the residuals from the quadratic detrending of Figure 1. There is a significant spectral mass within the frequency range $[0, \pi/6)$, of which the upper bound is the fundamental frequency of the seasonal fluctuations. This mass properly belongs to the trend and, if the trend had been adequately estimated, it would not be present in the periodogram of the residuals.

To construct a better estimate of the trend, an ideal lowpass filter, with a sharp cut-off frequency a little short of $\pi/6$, has been applied to the twice differenced data and the filtered sequence has been reinflated with initial conditions that are supplied by equation (**??**). The result is the trend that is shown in Figure 3. The pass band of the ideal lowpass filter has been superimposed upon the periodogram of Figure 2 as a shaded area.

Figure 4 shows the gains of the trend estimation filters that have been obtained by applying two of the model-based procedures to the data. The

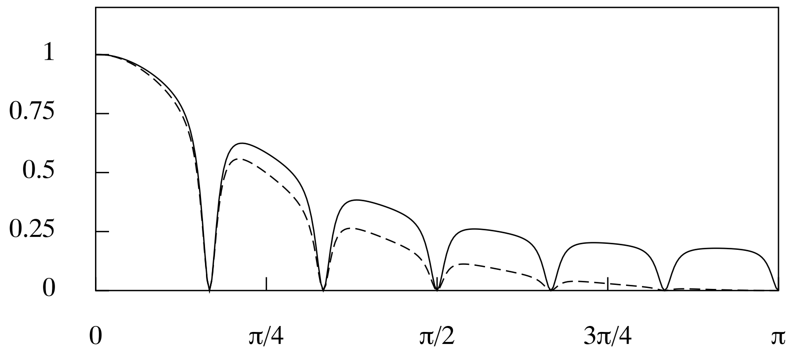

Figure 4: The gain function of the trend-extraction filter obtained from the STAMP program (solid line) together with that of the canonical trend-extraction filter (broken line) obtained from the TRAMO–SEATS program.

outer envelope is the gain of a trend extraction filter obtained in the process of using the STAMP program to estimate the components of the data. The inner envelope represents the gain of the analogous filter from the TRAMO–SEATS program. The indentations in the gain functions of both filters at the frequencies $\pi j/6; j = 1, \ldots, 6$ have the effect of nullifying the seasonal elements and of preventing them from entering the trend.

The two model-based filters differ greatly from the ideal filter. Disregarding the indentations, one can see how the gain of the filters is reduced only gradually as the frequency value increases. The trend component extracted by the STAMP filter would contain a substantial proportion of the non-seasonal high-frequency components that are present in the original data.

In practice, however, the trends that are estimated by the ideal filter and by the two model-based filters are virtually indistinguishable in the case of the money supply data. The reason for this is that, after the elimination of the seasonal components, whether it be by nullifying all elements of frequencies in excess of $\pi/6$ or only by eliminating the elements in the vicinities of the seasonal frequencies of $\pi j/6; j = 1, \ldots, 6$, there is virtually nothing remaining in the data but the trend. Therefore, in this case, the potential of the two model-based filters to transmit high-frequency components can do no harm.

In other cases, it has been observed that the STAMP filter produces a trend estimate that has a profile which is noticeably rougher than the one produced by the TRAMO–SEATS program—see Pollock (2002), for example—and this is a testimony to fact that the latter program, which observes the so-called canonical principle, suppresses the high-frequency noise more emphatically.

1 The Kalman Filter and the Smoothing Algorithm

One of the reasons for setting a structural time-series model in a state-space form is to make it amenable to the application the Kalman filter, which may be used both for estimating the parameters of the model and for extracting the unobserved components. To obtain estimates that take full advantage of all of the sampled data, a smoothing algorithm must also be deployed. These algorithms are described in the present section.

The state-space model, which underlies the Kalman filter, consists of two equations

$$y_t = H\xi_t + \eta_t, \qquad \textit{Observation Equation} \tag{19}$$

$$\xi_t = \Phi\xi_{t-1} + \nu_t, \qquad \textit{Transition Equation} \tag{20}$$

where y_t is the observation on the system and ξ_t is the state vector. The observation error η_t and the state disturbance ν_t are mutually uncorrelated random vectors of zero mean with dispersion matrices

$$D(\eta_t) = \Omega \qquad \text{and} \qquad D(\nu_t) = \Psi. \tag{21}$$

It is assumed that the matrices H, Φ, Ω and Ψ are known and that an initial estimate x_0 is available for the state vector ξ_0 at time $t = 0$ together with a dispersion matrix $D(\xi_0) = P_0$. This set of initial information is denoted by $\mathcal{I}_0$. (In a more general formulation, the parameter matrices would be allowed to vary with time, but here they are constant.) The information available at time t is $\mathcal{I}_t = \{y_t, \ldots, y_1, \mathcal{I}_0\} = \{y_t, \mathcal{I}_{t-1}\}$.

The Kalman-filter equations determine the state-vector estimates $x_{t|t-1} = E(\xi_t|\mathcal{I}_{t-1})$ and $x_t = E(\xi_t|\mathcal{I}_t)$ and their associated dispersion matrices $P_{t|t-1}$ and P_t from the values x_{t-1}, P_{t-1} of the previous period. From $x_{t|t-1}$, the prediction $\hat{y}_{t|t-1} = Hx_{t|t-1}$ is formed which has a dispersion matrix F_t. A summary of these equations is as follows:

$$x_{t|t-1} = \Phi x_{t-1}, \qquad \textit{State Prediction} \tag{22}$$

$$P_{t|t-1} = \Phi P_{t-1}\Phi' + \Psi, \qquad \textit{Prediction Dispersion} \tag{23}$$

$$e_t = y_t - Hx_{t|t-1}, \qquad \textit{Prediction Error} \tag{24}$$

$$F_t = HP_{t|t-1}H' + \Omega, \qquad \textit{Error Dispersion} \tag{25}$$

$$K_t = P_{t|t-1}H'F_t^{-1}, \qquad \textit{Kalman Gain} \tag{26}$$

$$x_t = x_{t|t-1} + K_t e_t, \qquad \textit{State Estimate} \tag{27}$$

$$P_t = (I - K_t H)P_{t|t-1}. \qquad \textit{Estimate Dispersion} \tag{28}$$

The equations of the Kalman filter may be derived using the ordinary algebra of conditional expectations which indicates that, if x, y are jointly distributed variables which bear the linear relationship $E(y|x) = \alpha + B\{x -$

$E(x)\}$, then

$$E(y|x) = E(y) + C(y,x)D^{-1}(x)\{x - E(x)\}, \tag{29}$$

$$D(y|x) = D(y) - C(y,x)D^{-1}(x)C(x,y), \tag{30}$$

$$E\{E(y|x)\} = E(y), \tag{31}$$

$$D\{E(y|x)\} = C(y,x)D^{-1}(x)C(x,y), \tag{32}$$

$$D(y) = D(y|x) + D\{E(y|x)\}, \tag{33}$$

$$C\{y - E(y|x), x\} = 0. \tag{34}$$

Of the equations listed under (22)—(28), those under (24) and (26) are merely definitions.

To demonstrate equation (22), we use (31) to show that

$$\begin{aligned} E(\xi_t|\mathcal{I}_{t-1}) &= E\{E(\xi_t|\xi_{t-1})|\mathcal{I}_{t-1}\} \\ &= E\{\Phi\xi_{t-1}|\mathcal{I}_{t-1}\} \\ &= \Phi x_{t-1}. \end{aligned} \tag{35}$$

We use (33) to demonstrate equation (23):

$$\begin{aligned} D(\xi_t|\mathcal{I}_{t-1}) &= D(\xi_t|\xi_{t-1}) + D\{E(\xi_t|\xi_{t-1})|\mathcal{I}_{t-1}\} \\ &= \Psi + D\{\Phi\xi_{t-1}|\mathcal{I}_{t-1}\} \\ &= \Psi + \Phi P_{t-1}\Phi'. \end{aligned} \tag{36}$$

To obtain equation (25), we substitute (19) into (24) to give $e_t = H(\xi_t - x_{t|t-1}) + \eta_t$. Then, in view of the statistical independence of the terms on the RHS, we have

$$\begin{aligned} D(e_t) &= D\{H(\xi_t - x_{t|t-1})\} + D(\eta_t) \\ &= HP_{t|t-1}H' + \Omega = D(y_t|\mathcal{I}_{t-1}). \end{aligned} \tag{37}$$

To demonstrate the updating equation (27), we begin by noting that

$$\begin{aligned} C(\xi_t, y_t|\mathcal{I}_{t-1}) &= E\{(\xi_t - x_{t|t-1})y_t'\} \\ &= E\{(\xi_t - x_{t|t-1})(H\xi_t + \eta_t)'\} \\ &= P_{t|t-1}H'. \end{aligned} \tag{38}$$

It follows from (29) that

$$\begin{aligned} E(\xi_t|\mathcal{I}_t) &= E(\xi_t|\mathcal{I}_{t-1}) + C(\xi_t, y_t|\mathcal{I}_{t-1})D^{-1}(y_t|\mathcal{I}_{t-1})\{y_t - E(y_t|\mathcal{I}_{t-1})\} \\ &= x_{t|t-1} + P_{t|t-1}H_t'F_t^{-1}e_t. \end{aligned} \tag{39}$$

The dispersion matrix under (28) for the updated estimate is obtained via equation (30):

$$\begin{aligned} D(\xi_t|\mathcal{I}_t) &= D(\xi_t|\mathcal{I}_{t-1}) - C(\xi_t, y_t|\mathcal{I}_{t-1})D^{-1}(y_t|\mathcal{I}_{t-1})C(y_t, \xi_t|\mathcal{I}_{t-1}) \\ &= P_{t|t-1} - P_{t|t-1}H_t'F_t^{-1}H_tP_{t|t-1}. \end{aligned} \tag{40}$$

The set of information $\mathcal{I}_t = \{y_t, \ldots, y_1, \mathcal{I}_t\}$, on which the Kalman filter estimates are based, can be represented, equivalently, by replacing the sequence $\{y_t, \ldots, y_1\}$ of observations by the sequence $\{e_t, \ldots, e_1\}$ of the prediction errors, which are mutually uncorrelated.

The equivalence can be demonstrated by showing that, given the initial information of $\mathcal{I}_0$, there is a one-to-one correspondence between the two sequences, which depends only on the known parameters of equations (19), (20) and (21). The result is intuitively intelligible, for, at each instant t, the prediction error e_t contains only the additional information of y_t that is not predictable from the information in the set $\mathcal{I}_{t-1}$; which is to say that $\mathcal{I}_t = \{e_t, \mathcal{I}_{t-1}\}$.

The prediction errors provide a useful formulation of the likelihood function from which the parameters that are assumed to be know to the Kalman filter can be estimated from the data. Under the assumption that the disturbances are normally distributed, the likelihood function is given by

$$\ln L = -\frac{kT}{2}\ln 2\pi - \frac{1}{2}\sum_{t=1}^{T}\ln|F_t| - \frac{1}{2}\sum_{t=1}^{T} e_t' F_t^{-1} e_t. \tag{41}$$

This form was proposed originally by Schweppe (1965). It tractability, which is a partial compensation for the complexity of the Kalman filter, has contributed significantly to the popularity of the state-space formulation of the structural time-series models.

There are various ways in which the value of the initial condition in $\mathcal{I}_0 = \{\xi_0, P_0\}$ may be obtained. If the processes are stationary, then the eigenvalues of the transition matrix Φ must lie within unit circle, which implies that $\lim(n \to \infty)\Phi^n = 0$. Then, there is $E(\xi_0) = x_0 = 0$ and $D(\xi_0) = P_0 = \Phi P_0 \Phi' + \Psi$; and the latter equation may be solved by analytic or iterative means for the value of P_0.

In the nonstationary case, the initial conditions require to be determined in the light of the data. To allow the information of the data rapidly to assert itself, one may set $P_0 = \lambda I$, where λ is given a large value. This will associate a large dispersion to the initial state estimate x_0 to signify a lack of confidence in its value, which will allow the estimate to be enhanced rapidly by the information of the data points. Using the terminology of Bayesian estimation, this recourse may be described as the method of the diffuse prior.

Data-dependent methods for initialising the Kalman filter of a more sophisticated nature, which make amends for, or which circumvent, the arbitrary choices of x_0 and P_0, have been proposed by Ansley and Kohn (1982) and by de Jong (1991), amongst others. These methods have been surveyed by Pollock (2003). Another account of the method of Ansley and Kohn, which is more accessible than the original one, has also been provided by Durbin and Koopman (2001).

The method of the diffuse prior bequeaths some pseudo information to the Kalman filter, in the form of arbitrary initial conditions, which remains in the system indefinitely, albeit that its significance is reduced as the sample information is accumulated. The technique of Ansley and Kohn is designed to remove the pseudo information at the earliest opportunity, which is when there is enough sample information to support the estimation of the state vector.

In their exposition of the technique, Ansley and Kohn described a transformation of the likelihood function that would eliminate its dependence on the initial conditions. This transformation was a purely theoretical device without any practical implementation. However, it is notable that the method of handling the start-up problem that has been expounded tje accompanying paper, which employs a differencing operation to reduce the data sequence to stationarity, has exactly the effect of eliminating the dependence upon initial conditions.

The Smoothing Algorithms

The Kalman filter generates an estimate $x_t = E(\xi_t|\mathcal{I}_t)$ of the current state of the system using information from the past and the present. To derive a more efficient estimate, we should take account of information that arises subsequently up to the end of the sample. Such an estimate, which may be denoted by $x_{t|T} = E(\xi_t|\mathcal{I}_T)$, is described as a fixed-interval estimate; and the various algorithms that provide the estimate are described as a fixed-interval smoothers.

It is laborious to derive the smoothing algorithms, of which there exist a fair variety. The matter is treated at length in the survey article of Merkus, Pollock and de Vos (1993) and in the monograph of Weinert (2001). Econometricians and others have derived a collection of algorithms which are, in some respects, more efficient in computation than the classical fixed-interval smoothing algorithm that is due to Rauch (1963), of which a derivation can be found in Anderson and Moore (1979), amongst other sources. A variant of the classical algorithm has been employed by Young *et al.* (2004) in the CAPTAIN MatLab toolbox, which provides facilities for estimating structural time-series models.

The classical algorithm may be derived via a sleight of hand. Consider enhancing the estimate $x_t = E(\xi_t|\mathcal{I}_t)$ in the light of the information afforded by an exact knowledge of the subsequent state vector ξ_{t+1}. The information would be conveyed by

$$h_{t+1} = \xi_{t+1} - E(\xi_{t+1}|\mathcal{I}_t), \tag{42}$$

which would enable us to find

$$E(\xi_t|\mathcal{I}_t, h_{t+1}) = E(\xi_t|\mathcal{I}_t) + C(\xi_t, h_{t+1}|\mathcal{I}_t)D^{-1}(h_{t+1}|\mathcal{I}_t)h_{t+1}. \tag{43}$$

Here there are

$$\begin{aligned} C(\xi_t, h_{t+1}|\mathcal{I}_t) &= E\{\xi_t(\xi_t - x_t)'\Phi' + \xi_t\nu_t'|\mathcal{I}_t\} = P_t\Phi' \quad \text{and} \\ D(h_{t+1}|\mathcal{I}_t) &= P_{t+1|t}. \end{aligned} \tag{44}$$

It follows that

$$E(\xi_t|\mathcal{I}_t, h_{t+1}) = E(\xi_t|\mathcal{I}_t) + P_t\Phi' P_{t+1|t}^{-1}\Big\{\xi_{t+1} - E(\xi_{t+1}|\mathcal{I}_t)\Big\}. \tag{45}$$

Of course, the value of ξ_{t+1} in the RHS of this equation is not observable. However, if we take the expectation of the equation conditional upon the available information of the set $\mathcal{I}_T$, then ξ_{t+1} is replaced by $E(\xi_{t+1}|\mathcal{I}_T)$ and we get a formula that can be rendered as

$$x_{t|T} = x_t + P_t\Phi' P_{t+1|t}^{-1}\{x_{t+1|T} - x_{t+1|t}\}. \tag{46}$$

The dispersion of the estimate is given by

$$P_{t|T} = P_t - P_t\Phi' P_{t+1|t}^{-1}\{P_{t+1|t} - P_{t+1|T}\}P_{t+1|t}^{-1}\Phi P_t. \tag{47}$$

This derivation was published by Ansley and Kohn (1982). It highlights the notion that the information that is used in enhancing the estimate of ξ_t is contained entirely within the smoothed estimate of ξ_{t+1}.

The smoothing algorithm runs backwards through the sequence of estimates generated by the Kalman filter, using a first-order feedback in respect of the smoothed estimates. The estimate $x_t = E(\xi_t|\mathcal{I}_t)$ is enhanced in the light of the "prediction error" $x_{t+1|T} - x_{t+1|t}$, which is the difference between the smoothed and the unsmoothed estimates of the state vector ξ_{t+1}.

In circumstances where the factor $P_t\Phi' P_{t+1|t}^{-1}$ can be represented by a constant matrix, the classical algorithm is efficient and easy to implement. This would be the case if there were a constant transition matrix Φ and if the filter gain K_t had converged to a constant. In all other circumstances, where it is required recompute the factor at each iteration of the index t, the algorithm is liable to cost time and to invite numerical inaccuracies. The problem, which lies with the inversion of $P_{t+1|t}$, can be avoided at the expense of generating a supplementary sequence to accompany the smoothing process.

Equivalent and Alternative Procedures

The derivations of the Kalman filter and the fixed-interval smoothing algorithm are both predicated upon the minimum-mean-square-error estimation criterion. Therefore, when the filter is joined with the smoothing algorithm, the resulting estimates of the data components should satisfy this criterion. However, its fulfilment will also depend upon an appropriate choice of the initial conditions for the filter. For this, one may use the method of Ansley and Kohn (1985).

The same criterion of minimum-mean-square-error estimation underlies the derivation of the finite-sample Wiener–Kolmogorov filter that has been presented in the accompanying paper. Therefore, when they are applied to a common model, the Wiener–Kolmogorov filter and the combined Kalman filter and smoother are expected to deliver the same estimates.

The handling of the initial-value problem does appear to be simpler in the Wiener–Kolmogorov method than in the method of Ansley and Kohn.

However, the finite-sample Wiener–Kolmogorov method is an instance of the transformation approach that Ansley and Kohn have shown to be equivalent to their method.

It should be noted that the minimum-mean-square-error estimates can also be obtained using a time-invariant version of the Wiener–Kolmogorov filter, provided that the finite data sequence can be extended by estimates of the presample and post-sample elements. However, this requires that the filter should relate to a well-specified ARMA or ARIMA model that is capable of generating the requisite forecasts and backcasts. If this is the case, then a cogent procedure for generating the extra-sample elements is the one that has been been described by Burman (1980) and which is incorporated in the TRAMO–SEATS program.

The upshot is that several routes lead to the same ends, any of which may be taken. Nevertheless, there have been some heated debates amongst econometrics who are the proponents of alternative approaches. However, the only significant issue is the practical relevance of the alternative models that are intended to represent the processes that underlie the data or to provide heuristic devices for generating the relevant filters.

An agnostic stance has been adopted in this chapter; and no firm pronouncements have been made concerning the nature of economic realities. Nevertheless, it has been proposed that the concept of a band-limited process, which had been largely overlooked in the past, is particularly relevant to this area of econometric analysis.

This concept encourages consideration of the Fourier methods of filtering presented in the accompanying paper, which are capable of separating components of the data that lie in closely adjacent frequency bands, as is the case in Figure 2, where the fundamental seasonal component abuts the low-frequency structure of the trend-cycle component. Such methods have been explored in greater detail in a paper of Pollock (2008); and they have been implemented in a program that is available from a website at the address

`http://www.le.ac.uk/users/dsgp1/`

References

Anderson BDO and Moor JB 1979 *Optimal Filtering.* Prentice–Hall, Englewood Cliffs, New Jersey.

Ansley CF and Kohn R 1982 A geometrical derivation of the fixed interval smoothing equations. *Biometrika* **69**, 486–7.

Ansley CF and R Kohn 1985 Estimation, Filtering and Smoothing in State Space Models with Incompletely Specified Initial Conditions. *The Annals of Statistics* **13**, 1286–1316.

Burman JP 1980 Seasonal adjustment by signal extraction. *Journal of the Royal Statistical Society, Series A* **143**, 321–337.

Caporello G and Maravall A 2004 *Program TSW: Revised Reference Manual.* Working Paper, Research Department, Banco de España.

de Jong P 1991 The diffuse Kalman filter. *The Annals of Statistics* **19**, 1073–1083.

Durbin J and Koopman SJ 2001 *Time Series Analysis by State Space Methods.* Oxford University Press.

Eurostat, 2002 *Demetra 2.0 User Manual: Seasonal Adjustment Interface for TRAMO–SEATS and X12-ARIMA,* The Statistical Office of the European Communities.
`http://circa.europa.eu/irc/dsis/eurosam/info/data/demetra.htm.`

Gómez V and Maravall A 1996 *Programs TRAMO and SEATS: Instructions for the user, (with some updates).* Working Paper 9628, Servicio de Estudios, Banco de España.

Harrison PJ and Stevens CF 1971 A Bayesian approach to short-term forecasting. *Operational Research Quarterly* **22**, 341–362.

Harrison PJ and Stevens CF 1976 Bayesian forecasting (with discussion). *Journal of the Royal Statistical Society, Series B* **38**, 205–247.

Harvey AC 1989 *Forecasting, Structural Time Series Models and the Kalman Filter.* Cambridge University Press, Cambridge.

Hillmer SC and Tiao GC 1982 An ARIMA-model-based approach to seasonal adjustment. *Journal of the American Statistical Association* **77**, 63–70.

Kaiser Regina and Maravall A 2001 *Measuring Business Cycles in Economic Time Series.* Springer Lecture Notes in Statistics, Springer-Verlag, New York.

Koopman SJ, Harvey AC, Doornik JA and Shephard N 2000 *STAMP 6.0: Structural Time Series Analyser, Modeller and Predictor.* Timberlake Consultants Press, London.

Kitagawa G and Gersch W 1996 *Smoothness Priors Analysis of Time Series.* Springer-Verlag, New York.

Kolmogorov AN 1941 Interpolation and extrapolation. *Bulletin de l'Academie des Sciences de U.S.S.R* Ser. Math **5**, 3–14.

Maravall A and Pierce DA 1987 A prototypical seasonal adjustment model. *Journal of Time Series Analysis* **8,** 177–193.

Merkus HR, Pollock DSG and de Vos AF 1993 A synopsis of the smoothing formulae associated with the Kalman filter. *Computational Economics* **6**, 177–200.

Pollock DSG 2002 A review of TSW: the Windows version of the TRAMO-SEATS program. *Journal of Applied Econometrics* **17**,291–299.

Pollock DSG 2003 Recursive estimation in econometrics. *Journal of Computational Statistics and Data Analysis* **44**, 37–75. .

Pollock DSG 2008 Realisations of finite-sample frequency-selective filters. *Journal of Statistical Planning and Inference* doi: 10.1016/j.jspi.2008.08.020.

Proietti T 2002 Forecasting with Structural Time Series Models. In *A Companion to Economic Forecasting,* (ed. Clements M). Blackwell Publishers, Oxford.

Rauch HE 1963 Solutions to the linear smoothing problem. *IEEE Transactions on Automatic Control* **AC-8**, 371–372.

Schweppe FC 1965 Evaluation of likelihood functions for Gaussian signals. *IEEE Transactions on Information Theory,* **11**, 61–70.

Simonoff JS 1996 *Smoothing Methods in Statistics.* Springer Series in Statistics Springer, New York.

Weinert HL 2001 *Fixed Interval Smoothing for State Space Models.* Kluwer International Series in Engineering and Computer Science, Kluwer Academic Publishers, Dordrecht.

West M 1997 Time series decomposition. *Biometrika* **84**, 489–494.

West M and Harrison J 1997 *Bayesian Forecasting and Dynamic Models, 2nd edition.* Springer-Verlag, New York.

West M, Prado R and Krystal AD 1999 Evaluation and comparison of EEG traces: latent structure in nonstationary time series. *Journal of the American Statistical Association* **94**, 1083–1095.

Young PC, Taylor CJ, Tych W, Pedregal DJ and McKenna CJ 2004 The Captain Toolbox, Centre for Research on Environmental Systems and Statistics, Lancaster University.
(`www.es.lancs.ac.uk/cres/captain`).