ALTERNATIVE METHODS FOR FILTERING ECONOMIC DATA

by D.S.G. POLLOCK

University of Leicester

Email: stephen_pollock@sigmapi.u-net.com

Website: https://www.le.ac.uk/users/dsgp1/

This brief note, which is intended to accompany a presentation at University of the Balearic Islands, Mallorca, describes the principal methods for filtering economic data in the time domain and the frequency domain. The methods are illustrated within the computer program IDEOLOG, which is available on the author's website at the University of Leicester. The website can be reached most readily by typing dsgp1 into a web browser.

Date: May, 2019

Wiener–Kolmogorov Filtering of Finite Stationary Sequences

It is assumed that the data vector y is the sum of two statistically independent components, which are the signal vector ξ and the noise vector η:

$$y = \xi + \eta. \tag{1}$$

Their first and second moments are

$$\begin{aligned} E(\xi) &= 0, \qquad D(\xi) = \Omega_\xi, \\ E(\eta) &= 0, \qquad D(\eta) = \Omega_\eta, \\ \text{and} \quad & C(\xi, \eta) = 0. \end{aligned} \tag{2}$$

If the data are generated by a Gaussian processes, then the process will be characterised completely by these moments.

Minimum-mean-sqare-error estimates

The conditional expectations of the components, which are their minimum-mean-sqare-error estimates and also their so-called Wiener–Kolmogorov estimates, are

$$\begin{aligned} E(\xi|y) &= E(\xi) + C(\xi, y)D^{-1}(y)\{y - E(y)\} \\ &= \Omega_\xi(\Omega_\xi + \Omega_\eta)^{-1}y = x, \end{aligned} \tag{3}$$

$$\begin{aligned} E(\eta|y) &= E(\eta) + C(\eta, y)D^{-1}(y)\{y - E(y)\} \\ &= \Omega_\eta(\Omega_\xi + \Omega_\eta)^{-1}y = h. \end{aligned} \tag{4}$$

The filtering of trended data

We assume that the intention is to decompose trended data into two complementary components, which can be recombined to recreate the original data. These are a low-frequency tend-cycle component and a high-frequency residual or noise component. The decomposition can be achieved in various ways.

Filtering in the time domain

The trended data can be processed directly by a low pass FIR (finite impulse response) filter. This will require the provision of some extra-sample data to allow for the fact that, otherwise, a filter of $2m + 1$ coefficients cannot process the first m and last m elements of the data. Alternatively, the filter can be collapsed as it nears the ends of the sample and the filter coefficients can be modified accordingly.

The Henderson FIR filter

The filter of Henderson (1916, 1924) provides a good example of such a device that can be applied directly to the data. Adaptations that allow the filter to reach the ends of the sample have be provided by Musgrave (1964a, b).

Differencing the data

The data can be rendered stationary by differencing (usually by double differencing). Thereafter, the differenced data is filtered and the resulting sequence is re-inflated by anti-differencing or summation. This requires the provision of some of starting values equal in number to the degree of differencing.

If the process of differencing and re-inflation is applied in the context of a finite-sample Wiener–Kolmogorov estimation, then it can be shown that the re-inflation of the low pass component with estimated initial conditions is equivalent to subtracting the estimate of the high pass component from the data.

Let the differenced data vector be denoted by $g = Q'y$. Then, applying Q' to the equation $y = \xi + \eta$ gives

$$\begin{aligned} Q'y &= Q'\xi + Q'\eta \\ &= \delta + \kappa = g. \end{aligned} \tag{5}$$

The expectations and the dispersion matrices of the component vectors are

$$\begin{aligned} E(\delta) &= 0, \qquad D(\delta) = \Omega_\delta, \\ E(\kappa) &= 0, \qquad D(\kappa) = \sigma_\nu^2 Q'\Omega_\eta Q. \end{aligned} \tag{6}$$

Trended data and Wiener–Kolmogorov estimates

The conditional expectation of the high-frequency component, which is its Wiener–Kolmogorov estimate, is

$$\begin{aligned} h = E(\eta|g) &= E(\eta) + C(\eta, g)D^{-1}(g)\{g - E(g)\} \\ &= C(\eta, g)D^{-1}(g)g, \end{aligned} \tag{7}$$

and, in view of (6), this becomes

$$h = \Omega_\eta Q(\Omega_\delta + Q'\Omega_\eta Q)^{-1}Q'y, \tag{8}$$

whence the estimate of the trend-cycle component is

$$x = y - h = \{I - \Omega_\eta Q(\Omega_\delta + Q'\Omega_\eta Q)^{-1}Q'\}y. \tag{9}$$

The data can also be reduced to stationarity by extracting a polynomial trend function. The polynomial residuals are subjected to a low pass filter. To create the trend-cycle, the filtered sequence can be added to the polynomial.

Filtering in the Frequency Domain

If the data sequence is stationary, which it might be in consequence of differencing or of trend extraction, then the filtering can take place in the frequency domain. The data are translated to the frequency domain by a Fourier transform. Then, the Fourier ordinates are modulated by factors that are indicated by the desired frequency response function. The modified ordinates are translated back to the time domain by an inverse Fourier transform.

The Fourier transform and the Fourier synthesis

The relationship between the data sequence $\{y_t; t = 0, 1, \ldots, T-1\}$ and the Fourier ordinates $\{\zeta_j; j = 0, 1, \ldots, T-1\}$ is represented by

$$y_t = \sum_{j=0}^{T-1} \zeta_j e^{\mathrm{i}\omega_j t} \longleftrightarrow \zeta_j = \frac{1}{T}\sum_{t=0}^{T-1} y_t e^{-\mathrm{i}\omega_j t}, \qquad \text{where} \qquad \omega_j = \frac{2\pi j}{T}. \tag{10}$$

The first of these equations, which depicts the inverse Fourier transform, represents the Fourier synthesis of the data, whereas the second equation depicts the direct Fourier transform of the data.

The data can also be expressed in terms of a set of mutually orthogonal trigonometric functions:

$$y_t = \sum_{j=0}^{[T/2]} (\alpha_j \cos\omega_j t + \beta_j \sin\omega_j t), \tag{11}$$

where $[T/2]$ is the quotient (i.e. the integral part) of $T/2$. The coefficients of this equation are

$$\alpha_j = \zeta_j + \zeta_{T-j} \qquad \text{and} \qquad \beta_j = \mathrm{i}(\zeta_j - \zeta_{T-j}), \tag{12}$$

whereas, according to Euler's equations, there are

$$\cos\theta = \frac{e^{\mathrm{i}\theta} + e^{-\mathrm{i}\theta}}{2} \qquad \text{and} \qquad \sin\theta = \frac{-\mathrm{i}}{2}(e^{\mathrm{i}\theta} - e^{-\mathrm{i}\theta}) = \frac{1}{2\mathrm{i}}(e^{\mathrm{i}\theta} - e^{-\mathrm{i}\theta}). \tag{13}$$

The coefficients of equation (11) are obtained by projecting the data onto the trigonometrical functions, which gives

$$\alpha_0 = \frac{1}{T}\sum_t y_t = \bar{y}, \quad \alpha_j = \frac{2}{T}\sum_t y_t \cos\omega_j t \quad \text{and} \quad \beta_j = \frac{2}{T}\sum_t y_t \sin\omega_j t. \tag{14}$$

In choosing an appropriate frequency response by which to modulate the ordinates or the coefficients, there is more flexibility in the frequency-domain approach than there is in the time-domain approach

Circular and periodic data

In a Fourier analysis, the data sequence is regarded as a single cycle of a periodic or circular process. For the end-of-sample problem, this means that the pre-sample and post-sample values are proxied by values from the other end of the sequence, which seems like a bad idea.

Therefore, it is proposed to interpolate a segment of dummy data between the end and the beginning of the circular data sequence. This can be created by combing segments from the beginning and the end of the data. The weights of the combination, which must add to one, are governed by a logistic curve, provided by a raised cosine function.

Inflating the Fourier ordinates

The Fourier method can be applied to data that have been reduced to stationarity, either by taking differences or by extracting a polynomial trend function.

In the case of a high pass or band pass filtering of differenced data, the necessary re-inflation can take place in the frequency domain before translating the results back to the time domain.

In that case, the first step is to apply a centralised difference operator to the original data in the time domain. If L represents the time-domain lag operator, then the centralised second difference of the data sequence $y(t)$ is

$$\Gamma(L)y(t) = \{2 - (L + L^{-1})\}y(t) = -y(t-1) + 2y(t) - y(t+1). \tag{15}$$

The end effects associated with the operation concern only single points from either end of the data.

The time-domain operator $\Gamma(L)$ translates to the following frequency-domain operator:

$$\Gamma(\omega) = 2 - 2\cos(\omega) = 4\sin^2(\omega/2). \tag{16}$$

The re-inflation of the high pass or band pass component can be achieved by dividing the filtered Fourier ordinates by this function before carrying them back to the time domain.

References

Henderson, R., (1916), Note on Graduation by Adjusted Average, *Transactions of the Actuarial Society of America,* 17, 43–48.

Henderson, R., (1924), A New Method of Graduation, *Transactions of the Actuarial Society of America,* 25, 29–40.

Musgrave, J.C., (1964a), A Set of End Weights to End all End Weights. *Unpublished Working Paper of the U.S. Bureau of Commerce.*

Musgrave, J.C., (1964b), Alternative Sets of Weights Proposed for X-11 Seasonal Factor Curve Moving Averages, *Unpublished Working Paper of the U.S. Bureau of Commerce.*

$$
\begin{aligned}
\alpha_j \cos\omega_j t + \beta_j \sin\omega_j t &= (\zeta_j + \zeta_{T-j})\left(\frac{e^{\mathrm{i}\omega_j t} + e^{-\mathrm{i}\omega_j t}}{2}\right) \\
&\quad + \mathrm{i}(\zeta_j - \zeta_{T-j})\left(\frac{e^{\mathrm{i}\omega_j t} - e^{-\mathrm{i}\omega_j t}}{2\mathrm{i}}\right) \qquad (*) \\
&= \zeta_j e^{\mathrm{i}\omega_j t} + \zeta_{T-j} e^{-\mathrm{i}\omega_j t}.
\end{aligned}
$$