

CYCLES, SYLLOGISMS AND SEMANTICS: EXAMINING THE IDEA OF SPURIOUS CYCLES

By D.S.G. Pollock

University of Leicester

Email: stephen_pollock@sigmapl.u-net.com

The claim that linear filters are liable to induce spurious fluctuations has been repeated many times of late. However, there are good reasons for asserting that this cannot be the case for the filters that, nowadays, are commonly employed by econometricians. If these filters cannot have the effects that have been attributed to them, then one must ask what effects the filters do have that could have led to the aspersions that have been made against them.

Introduction: The History of an Idea

The idea that fluctuations can be imparted to a data sequence by passing it through a linear filter has long been familiar to econometricians. It was associated with the discoveries of Slutsky and Yule in the early years of the 20th century.

Slutsky (1927, 1937) applied a moving-average filter to random numbers drawn from a public lottery to produce a sequence that had the characteristics of a macroeconomic business cycle. Yule (1927) demonstrated the manner in which a second-order autoregressive model, driven by a white-noise sequence of independently and identically distributed random variables, can give rise to an output that contains cycles of such regularity that one might imagine that they have a mechanical origin.

The danger of being misled by an inappropriate use of filters was emphasised by Howrey (1968), who discovered that the long-run economic cycles that Kuznets (1961) claimed to have detected were, in fact, the artefacts of his data processing. It seemed appropriate to describe these cycles as spurious.

A linear filter can have two effects. The first effect is to amplify or to attenuate the amplitudes of the cyclical elements of the data to an extent that varies with the frequencies of the elements. This is described as the gain effect of the filter. The second effect is to displace the elements in time, such that their peaks and troughs are advanced or retarded. This is the phase effect of the filter. The phase effect can be avoided if the filter coefficients are disposed symmetrically about a central point, so that the filter reaches equally forward and backwards in time.

The examples described so far entail a marked amplification of the amplitudes of sinusoidal elements within a narrow range of frequencies, accompanied by marked attenuations over the remaining frequencies. It is inaccurate to say that the cycles that have been amplified have been induced in the data, since they are already present; but, in these contexts at least, the abuse of language

is tolerable. However, such semantic issues will become important in the wider context of our enquiry.

The claim that linear filters are liable to induce spurious fluctuations has persisted; and it has been repeated often of late. However, there are good reasons for asserting that this cannot be the case for the filters that, nowadays, are commonly employed by econometricians. The purpose of this paper is to demonstrate conclusively that these filters cannot have the effects that have been attributed to them. Also, we need to ask what effects the filters do have that could have led to the aspersions that have been made against them.

The Linear Detrending of a Random Walk

The belief that an inappropriate processing of the data can induce spurious fluctuations has been reaffirmed in more recent times in connection with the filtering of data generated by random walk processes. Chan *et al.* (1977) and Nelson and Kang (1981) have described the effects of using linear and polynomial regressions to remove apparent trends from the data. They have observed that, regardless of the length of the data sequence, a random walk that has been subject to detrending exhibits major cycles that have a duration that is matched to the length of the sample.

The result can be explained in reference to the self-similarity of a Wiener process in continuous time, from which a discrete-time random walk can be obtained by a process of sampling. The self-similarity means that every segment of the Wiener process has a similar appearance and the same statistical characteristics, regardless of its duration.

A sample of n elements taken at regular intervals from a Wiener process, and scaled appropriately, has the same statistical properties as any other such sample of n elements, regardless of the rate of sampling. Let samples be taken at intervals of one and T time units. Then

$$T^{1/2}(x_1, x_2, \dots, x_n) \stackrel{D}{=} (x_{T1}, x_{T2}, \dots, x_{Tn}), \quad (1)$$

which is to say that the two sides have the same distribution. Therefore, the effects, in general, of fitting a linear trend according to a least-squares criterion should not vary with the length of time spanned by the data. Nor should the trend line be much affected by varying the rate of sampling within a given time span.

Figure 1 shows a random walk with an upward drift, through which a straight line has been interpolated by ordinary least-squares regression. The random walk, which has a minimal sampling interval, can be taken to represent a Wiener process seen with a limited visual acuity.

Nelson and Kang have presented the autocorrelation function of the residual sequence from the linear detrending of a random walk via a least-squares regression. The autocorrelations can be expressed as a function of the ratio of the time lags to the sample size T . This function tends to an asymptotic limit as T tends to infinity.

A spectral density function can be derived on the assumption that the autocorrelation function in question corresponds to a doubly-infinite stationary

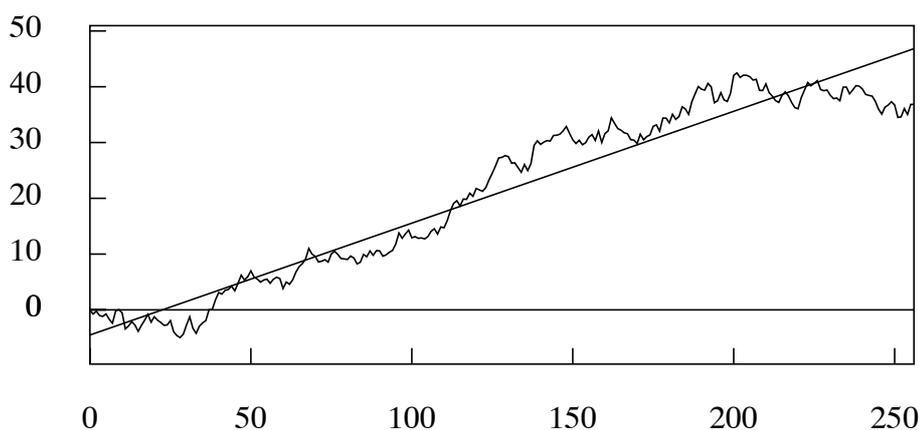


Figure 1. A random walk generated by the equation $y_t = y_{t-1} + \delta + \varepsilon_t$ together with an interpolated regression line. The variance of the white-noise disturbance is $V(\varepsilon_t) = 1$ and the drift parameter is $\delta = 0.2$.

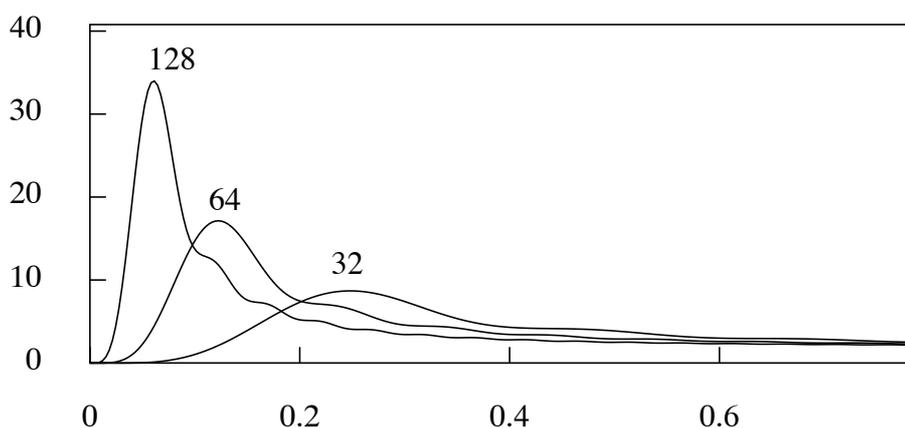


Figure 2. The spectral density function derived from the autocorrelation function of Nelson and Kang for sample sizes of 32, 64 and 128.

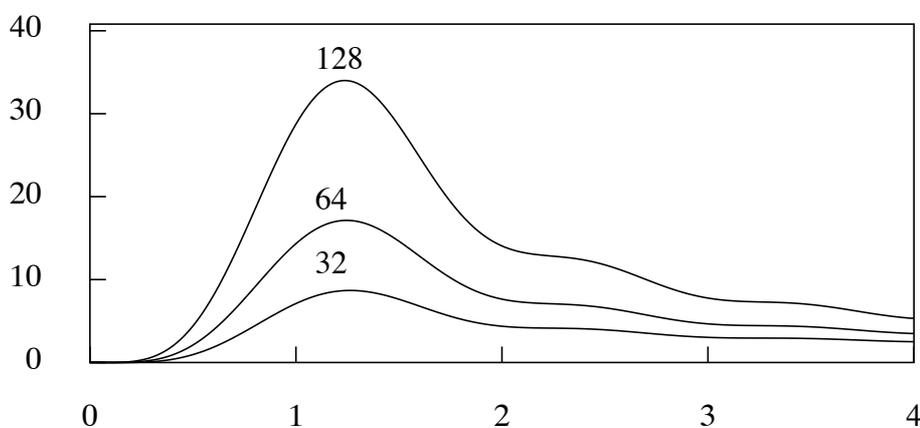


Figure 3. The spectral density functions for sample sizes of 32, 64 and 128, plotted as functions of cycles within the duration of the sample.

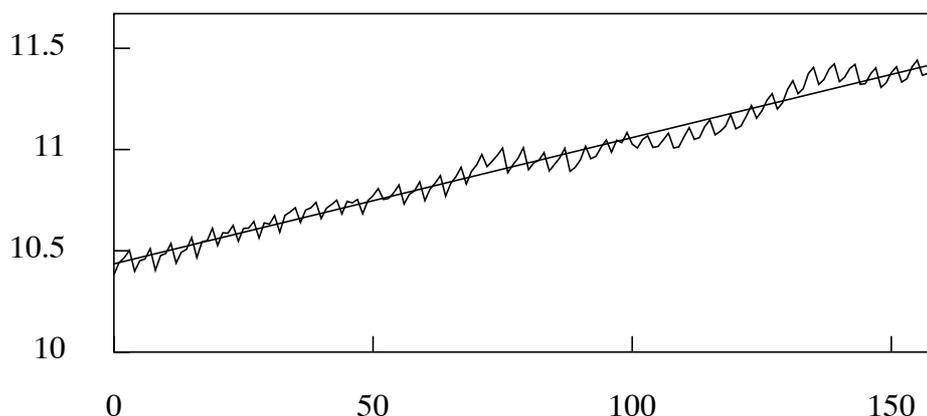


Figure 4. The quarterly series of the logarithms of consumption in the U.K., for the years 1955 to 1994, together with a linear trend interpolated by least-squares regression.

stochastic process. If the matrix of the autocovariances of the residuals from the linear detrending of a finite sample of a random walk is to serve as the basis for a spectral density function, then its form must be rectified, such that it becomes a Toeplitz matrix. The elements on the diagonals of the matrix, which will vary with the row or column index, must be replaced by constant values. This can be achieved by averaging the elements of each diagonal.

Figure 2 shows spectral density functions corresponding to the rectified autocorrelation functions for sample sizes of 32, 64, and 128. Here, the spectra are plotted against an axis that measures absolute frequency in radians per sample interval. These spectra are also plotted in Figure 3, where the horizontal axis measures the number of cycles within the length of the sample. In this Figure, the spectral peaks are aligned at a frequency value of 1.265 cycles per sample.

Nelson and Kang have not hesitated to describe the phenomenon that they have uncovered as one of spurious periodicity. This judgement may have been based on the perception that a random walk has no central tendency. It is presumed that, in the absence of a central tendency, there can be no cyclicity.

However, a random walk is generated by cumulating a white noise sequence, which contains cycles of every frequency in the interval running from zero to the Nyquist frequency π radians per period, which represents the limit of the frequencies that are observable in sampled data. Therefore, the idea that there is no cyclicity in the process should be treated with caution.

It is easy to see how, via a simple syllogism, a false conclusion concerning macroeconomic data sequences can arise. The major premise is that the macroeconomic data can be regarded as the product of random walk. The minor premise is that the detrending of a random walk gives rise to spurious cycles. The conclusion is that the detrending of a macroeconomic sequence induces spurious cycles.

One need not demur over the use of the word spurious in connection with the cycles resulting from the linear detrending of a random walk. It is the

complete identification of the macroeconomic process with a random walk (or with a random walk with drift) that is at fault.

In contrast to random walks, real economic processes are subject to evident constraints. They are driven by the buoyant forces of entrepreneurial endeavour and by consumer aspirations, and they are constrained by the more or less pliable limits of productive capacity and resource availability. In a thriving economy, they press alternately against the floors and the ceilings and they rebound from them in a manner that is undeniably cyclical.

A straight line interpolated through the logarithms of a macroeconomic data sequence represents a benchmark of constant exponential growth. The expectation that this should be the underlying trajectory of a well functioning economy became widespread amongst the citizens of affluent countries during 20th century; and the cyclical departures from such a trajectory have been characterised as booms and slumps.

Therefore, the analytic procedures that Nelson and Kang have warned against seem to be the natural ones to follow, at least in times that are not affected by major economic crises.

It should be born in mind that, when the number of major cycles within the (linearly) detrended macroeconomic data exceed two or three, then the analogy with a random walk begins to break down. Whether or not the analogy should be rejected at this stage depends entirely on the purpose for which it is being used.

Figure 4 shows the logarithms of a sequence of quarterly aggregate consumption in the U.K. through which a straight line has been interpolated by least-squares regression. The calculation of the trend line entails a matrix version of the twofold difference operator that takes the form of

$$Q' = \begin{bmatrix} 1 & -2 & 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 & -2 & 1 \end{bmatrix}. \quad (2)$$

From this matrix is formed the projection operator

$$P = Q(Q'Q)^{-1}Q', \quad (3)$$

which gives rise to the following decomposition of the data vector:

$$\begin{aligned} y &= (I - P)y + Py \\ &= f + e. \end{aligned} \quad (4)$$

Here, $f = (I - P)y$ contains the ordinates of the linear trend and $e = Py$ contains the residual deviations of the data from the trend.

The Line Crossings of a Random Walk

A great deal of effort has been devoted by econometricians in recent years to the matter of testing the hypothesis that the trajectory of an economic variable

has been generated by a random walk process; and many variations of the tests have been investigated. Usually, the question is posed of whether or not the process generating the data contains a unit root within an autoregressive operator. Despair has often arisen from the fact that the tests rarely provide a unequivocal answer to the question.

An alternative question that might be asked is whether or not a unit-root process provides an acceptable model for an economic process. Then, there should no expectation of an unequivocal answer, since the criteria for determining what is an appropriate model will vary according to the purpose of the investigation and the tastes of the investigator. Nevertheless, formal tests of a null hypothesis that the data have been generated by an equation of a simple parametric structure can provide an essential guide to mathematical modelling.

The null hypothesis that is commonly adopted is that the data have been generated by a first-order random walk with drift. The hypothesis can be represented by imposing the restriction that $\rho = 1$ within the equation

$$y_t = \delta + \rho y_{t-1} + \varepsilon_t, \tag{5}$$

wherein ε_t is an element from a white-noise process of independently and identically distributed random variables of zero mean and of a finite variance of $V(\varepsilon_t) = \sigma^2$, for all t . The process depicted by the equation is assumed to have begun at time $t = 0$ with a finite value y_0 .

The vast majority of the tests are addressed directly to the matter of whether $\rho = 1$ or whether, alternatively, $|\rho| < 1$. Such tests depend on measuring the rate of mean reversion or, equivalently, the strength of the central tendency of the data. It should be recognised that, the more rapidly the data are observed, the less will be the apparent rate of mean reversion, as measured from one point to the next. Therefore, it is reasonable to seek a test that is independent of the rate of sampling. Tests of this nature can be based on the number of times that the trajectory of the data crosses a line that represents the mean to which it is supposed to revert.

Such a test can be based on a result of Feller (1968) in the specialised case where $\delta = 0$ and where $\varepsilon_t = \pm 1$ is generated by a Bernoulli trial with equal probabilities for the two outcomes. Feller has demonstrated that, if N_T is the number of times that the resulting trajectory crosses the horizontal axis in the T periods covered by the data, then the probability $P(N_T/\sqrt{T} < z)$ that the scaled number does not exceed z will be given, in the limit as $T \rightarrow \infty$, by $2\Phi(2z) - 1$, where $\Phi(z)$ denotes the cumulative standard normal distribution.

This strange looking result does not provide an adequate prototype for the case where ε_t has a continuous distribution. As Burrige and Guerre (1996) have remarked, the number of trajectories that only touch the horizontal axis before bouncing back is equal, in the discrete case, to the number of trajectories that cross the axis. (This result is a consequence of the reflection principle, which indicates that, for any trajectory that crosses the axis, the segments that lie below the axis can be reflected upwards to create a trajectory that only touches the axis.) In the continuous case, by contrast, the trajectories that only touch the horizontal axis constitute a set of measure zero.

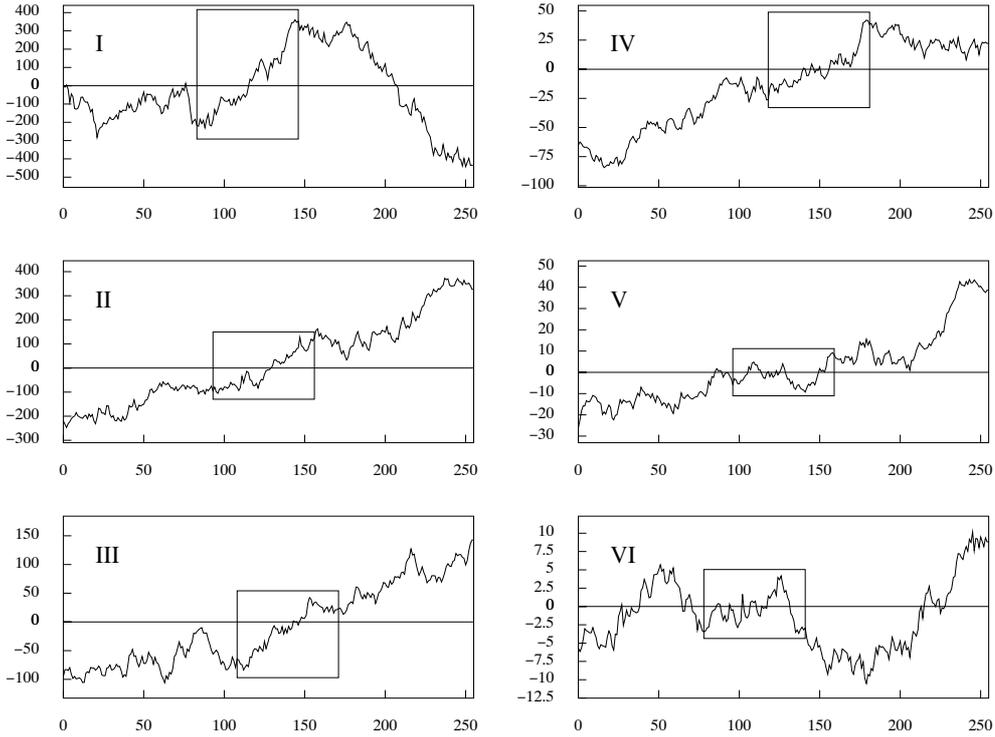


Figure 5. Nested segments of a trajectory of Brownian motion, sampled at rates that increase successively by factors of 4.

Burrige and Guerre have shown that, in the continuous case, the number of line crossings depend on the nature of the distribution of ε_t —distributions with fatter tails giving rise to fewer line crossings. They have established the result that

$$T^{-1/2}N_T \xrightarrow{D} \frac{E|\varepsilon_t|}{\sigma}|z|, \quad (6)$$

where $z \sim N(0, 1)$ is a standard normal variable, where $\sigma^2 = V(\varepsilon_t)$ and where \xrightarrow{D} denotes convergence in distribution. The half-normal distribution $|N(0, 1)|$ is the distribution of the absolute values of a standard normal variate; and it is obtained by folding the negative range of the $N(0, 1)$ distribution over the vertical axis in the manner of closing an open book.

In the case where ε_t is an element from a normal distribution, which corresponds to the hypothesis that the data have been sampled from a Wiener process, there is $E|\varepsilon_t|/\sigma = 2/\sqrt{2\pi}$.

The result given by (6) requires to be interpreted in the light of the familiar picture of a Wiener process. First, it should be recognised that, as $T \rightarrow \infty$, both the expected number of times that the sampled trajectory crosses the axis and the expected waiting times between successive crossings tend to infinity at the rate of $T^{1/2}$. By contrast, in a stationary mean-reverting process, the number of line crossings will increase at the rate of T and the expected duration between successive crossings will be a finite constant.

Next, since the infinite sampled sequence corresponds to a finite segment of the continuous Wiener process, one is bound to ask where one should look

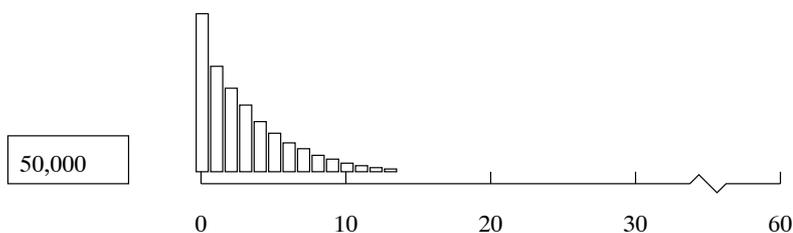


Figure 6. The histogram of the number of times a random walk of 60 steps crosses the horizontal axis, determined via 50,000 trials.

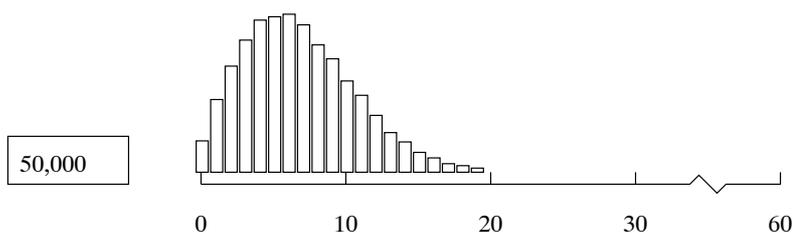


Figure 7. The histogram of the number of times a random walk of 60 steps crosses a line through the end points of the sample.

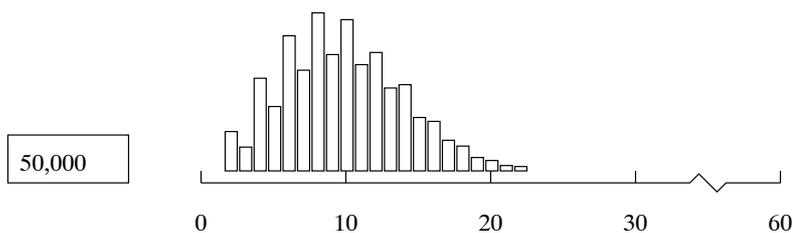


Figure 8. The histogram of the number of times a random walk of 60 steps crosses a line interpolated by least-squares regression.

to find an infinite number of line crossings. An answer is given via the succession pictures of Figure 5, which shows nested segments of a Wiener trajectory, sampled at rates that increase successively by factors of 4, which implies a magnification from first to last of $4^5 = 1024$.

The first picture reveals four line crossings, the first two of which can barely be distinguished from cases where the trajectory only touches the horizontal axis. We shall describe these four crossings that are associated with the major cycles as the major crossings.

Attention is focussed on the third crossing. The succession of pictures reveals that what seems, at the lowest resolution, to be a single crossing is a case of multiple crossings within a limited vicinity. For want of a better description, we shall call these the minor crossings. It will be understood that, as the resolution, or, equivalently, the rate of sampling, increases, the number of minor crossings that can be discerned is liable to increase indefinitely.

The implications of our analysis of a Wiener process and of the associated random walk appear to be at variance with the analysis of Nelson and Kang,

which points to a single dominant cycle of a duration that is comparable to the length of the data period. However, the latter cycle has been attributed to the linear detrending of a drifting random walk; and it would also arise in the case of an ordinary random walk with $\delta = 0$, if the horizontal axis were replaced by a straight line interpolated by least-squares regression.

Since the majority of macroeconomic data sequences show marked trends, it is inevitable that their degrees of central tendency should be measured relative to an interpolated trend. To fulfil this requirement, García and Sansó (2006) have proposed to generalise the test of Burridge and Guerre by replacing the horizontal axis at the level of y_0 by a line that passes through the first and the last of the data points of a finite sample, thereby creating what is known as a Brownian Bridge. Thus, they replace the observations y_0, y_1, \dots, y_{T-1} by the adjusted values x_0, x_1, \dots, x_{T-1} , where

$$x_t = y_t - y_0 - ct, \quad x_0 = y_0 \quad \text{and} \quad c = (y_{T-1} - y_0)/T. \quad (7)$$

García and Sansó have considered the number of sign changes N_T^B of the adjusted sequence, and they have established that

$$T^{-1/2} N_T^B \xrightarrow{D} \frac{E|\varepsilon_t|}{\sigma} R(z), \quad (8)$$

where $R(z) = ze^{-z^2/2}$ denotes a standard Raleigh distribution.

The distribution of the number of times that a drifting random walk crosses a line interpolated by least-squares regression does not seem to possess a simple analytic form. For any sample size, even numbers of crossing are, on average, more numerous than odd numbers of crossings.

In implementing tests based on the number of line crossings, it is appropriate, when the sample size is small, to rely on empirically determined distributions as opposed to asymptotic approximations.

Figures 6–8 show the empirical distributions of the number of line crossings for samples of 60 points generated by a random walk, determined, in each case via 50,000 trials. The three cases concern the number of times a random walk without drift crosses the horizontal axis, the number of times a random walk crosses a line interpolated through the first and the final points and the number of times that a random walk crosses a line interpolated by a least-squares regression.

The line-crossing tests have a limited ability to distinguish an interpolated random walk from a cyclical ARMA process of the sort that could be used to describe a business cycle. A more efficient test would take account not only of the number of line crossings but also of their locations. It would discount the minor crossings of the random walk that are to be found in the vicinity of the major crossings. Such minor crossings are due to the high-frequency contents of the random walk, which are absent from the ARMA process.

Aspersions against the Hodrick–Prescott Filter

The idea that filtering can induce spurious cycles has also been fostered by a succession of papers that have inveighed against the use of the filter of Hodrick

and Prescott (1980, 1997), which is also attributable to Leser (1961), as a device for extracting trends from economic data. See, for example, King and Rebelo (1993), Harvey and Jaeger (1993), Jaeger (1994), Cogley and Nason (1995), Schenk-Hoppe (2001) and Ivanov (2005). These critics tend to regard the random walk as an appropriate model for an economic process; and their analysis typically concerns the interaction of the frequency response of the filter with the pseudo spectrum of the random-walk process, which is defined on a doubly infinite set of indices.

Such a random walk is truly an unimaginable process; and its values have a zero probability of being found within a finite distance of the origin. Also, the pseudo spectrum is unbounded in the vicinity of the zero frequency. It is observed that, when the pseudo spectrum is modulated by the frequency response function of the highpass Hodrick–Prescott filter, a spectral density function is produced that has a prominent peak in the low-frequency region. This spectral density function, which corresponds to the output of the filter, is identified with a cyclical process. It is commonly asserted, on this basis, that the filter is liable to induce spurious cycles.

In Figure 9, the curve labelled *A* represents the pseudo spectrum of a random walk and the curve labelled *B* is the frequency response function of the Hodrick–Prescott filter, which shows the squared gain of the filter over the range of frequencies. These run from zero to the Nyquist frequency π radians per period. The curve *C*, which is the product of *A* and *B*, represents the spectral density function of the output of the filter.

The matter can be approached from two sides—that of the filter and that of the pseudo-spectrum of process. As regards the filter, it will be observed that, over the entire range of frequencies, its gain never exceeds unity. Its gain is close to unity over the range of frequencies that are described as the pass band. Elsewhere, the gain makes a transition from a value close to unity to zero, which is reached at the zero frequency.

The effect of the filter is, therefore, to nullify or to attenuate strongly some elements of the data that are in the vicinity of the zero frequency while preserving other elements that are of higher frequencies. Therefore, one can declare emphatically that nothing is induced or amplified by the filter and that nothing that is to be found in its output can be described as spurious. Opinions to this effect have been voiced by Pedersen (2001), Pollock (1997, 2000) and Kaiser, and Maravall (1999), amongst others.

Next, it can be declared that the pseudo-spectrum is a doubtful concept that ought to be handled with caution. It relates to a nonstationary process and, as such, it has a doubtful role within the context of an analysis which, ostensibly, is appropriate only to stationary processes of the sort that can be represented as weighted sums of trigonometrical or complex exponential function defined over the entire set of positive and negative integers.

The Typical Spectral Shape of an Economic Variable

The pseudo spectrum of a random walk has been regarded, occasionally, as an appropriate surrogate for the typical spectrum of an economic process of

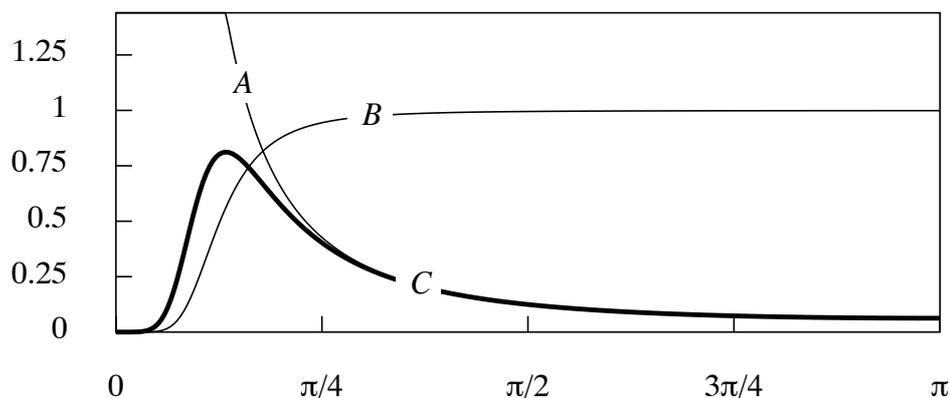


Figure 9. The pseudo-spectrum of a random walk, labelled *A*, together with the squared gain of the highpass Hodrick–Prescott filter with a smoothing parameter of $\lambda = 100$, labelled *B*. The curve labelled *C* represents the spectrum of the filtered process.

the sort that was identified by Granger (1966). This is a so-called “one-over- f ” spectrum of which the power declines monotonically as the frequency f increases. Such spectra are common to a wide variety of physical and biological processes including, for example, ocean waves and electroencephalographs.

However, there is a strong supposition that, in econometrics, this spectral shape is associated, primarily, with a failure to reduce the data to stationarity on account of an inadequate detrending. The subject of a spectral analysis of a finite data sequence is the indefinite periodic extension of the sequence. The periodic extension of a finite trended data sequence will give rise to a saw tooth function of which the spectrum or periodogram has a typical “one-over- f ” profile.

However, a “one-over- f ” spectrum that achieves a maximum value—albeit a finite value—at zero frequency may also correspond to a regular stationary stochastic process. In such a context, one can reasonably analyse the effects of the Hodrick–Prescott filter. Then, a possible aspersion against the filter is that it allows low-frequency elements to be transmitted when they ought to be stopped.

Filters and Trended Data

Any misgivings regarding the pseudo spectrum should not be taken to imply that a linear filter cannot be applied to trended data. It is proposed only that a conventional spectral analysis is inappropriate to cases of nonstationary processes and to their pseudo-spectra.

In applying a filter directly to trended data, one must take care to supply the appropriate pre-sample and post-sample values to allow it to be run up or down the sample in a manner that avoids creating inappropriate end effects. Erroneous end effects can easily contaminate all of the filtered data sequence.

An appropriate recourse in the case of trended data, which avoids the difficulty of the end effects, is to apply the filter to residuals that have been

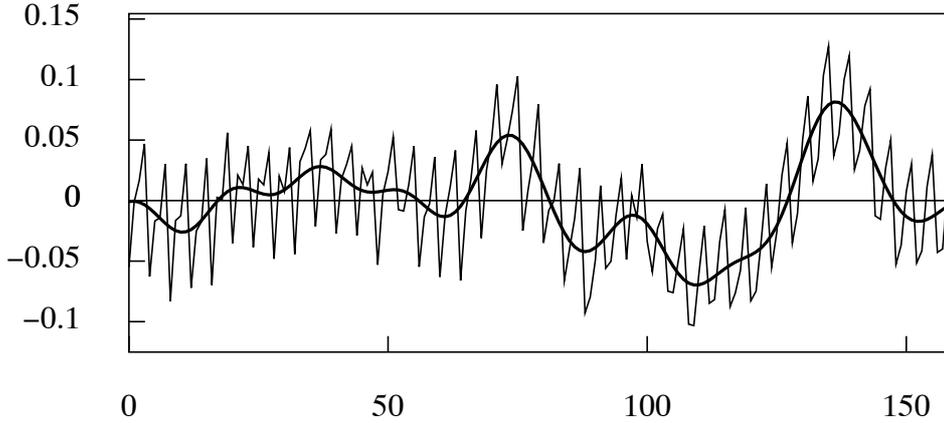


Figure 10. The residual sequence $e = Py$ from fitting a linear trend to the logarithmic consumption data with an heavy interpolated line De representing the business cycle, obtained by the frequency-domain method.

obtained from fitting a polynomial function to the data. The residuals can be filtered to separate their low-frequency elements from their high-frequency elements. The low-frequency elements can be added back to the polynomial trend to generate a more variable trend. The high-frequency elements, which correspond to the detrended data, may be subjected to further filtering which could be designed, for example, to remove the seasonal fluctuations.

Figure 10 shows the effects of filtering the residual sequence obtained by fitting a straight line to the data of Figure 4. The smooth curve represented by the heavy line has been obtained via a synthesis based on the Fourier ordinates of the residual sequence that lie in the frequency interval $[0, \pi/8]$. The purpose of this filtering is to remove the powerful seasonal fluctuations from the sequence $e = Py$ and to eliminate some minor high-frequency elements.

A frequency-domain filtering requires a Fourier transform to be applied to the data vector to carry it into the frequency domain. Then, the Fourier ordinates, which are the product of the transformation, can be modified in the desired manner before being carried back to the time domain via an inverse Fourier transform to become the filtered values.

For a matrix representation of these operations, one may define

$$\begin{aligned} U &= T^{-1/2}[\exp\{-i2\pi tj/T\}; t, j = 0, \dots, T-1], \\ \bar{U} &= T^{-1/2}[\exp\{i2\pi tj/T\}; t, j = 0, \dots, T-1], \end{aligned} \quad (9)$$

which are unitary complex matrices such that $U\bar{U} = \bar{U}U = I_T$. Then,

$$\zeta = T^{-1/2}Uz \quad \longleftrightarrow \quad z = T^{1/2}\bar{U}\zeta, \quad (10)$$

where $z = [z_0, z_1, \dots, z_{T-1}]'$ and $\zeta = [\zeta_0, \zeta_1, \dots, \zeta_{T-1}]'$ are the vectors of the data and of their Fourier ordinates, respectively.

Let $\Lambda = \text{diag}\{\lambda_0, \lambda_1, \dots, \lambda_{T-1}\}$ be a diagonal matrix of the weights. Then, the modified Fourier ordinates are in the vector

$$\Lambda\zeta = T^{-1/2}\Lambda Uz. \quad (11)$$

Subjecting this vector to the inverse Fourier transform gives the filtered output

$$x = T^{1/2}\bar{U}\Lambda\zeta = \{\bar{U}\Lambda U\}z = \Phi^\circ z, \quad (12)$$

where $\Phi^\circ = \bar{U}\Lambda U$ is the matrix of the filtering operation in the time domain. This is a circulant matrix; and the filtering of the data in the time domain would entail the circular convolution of the data with the filter coefficients that are to be found at successive displacements within successive rows of the matrix.

Notwithstanding the fact that the filtering is performed more efficiently in the frequency domain in the manner that has been described, the filtering of the residual vector $e = Py$ will be represented, hereafter, by the time domain equation $h = De$.

In the case of the ideal filter that selects only the Fourier ordinates that lie in the frequency interval $[0, \pi/8]$, the element of the matrix D are the coefficients of a Dirichelet kernel. These matters have been elucidated in Pollock (2009), where some devices are described for avoiding the disjunctions that may occur in the periodic extension of the data sequence where the end of one iteration of the data joins the beginning of the next iteration.

The smooth trajectory of Figure 10 might be regarded as a good representation of the business cycle in the U.K. over the period 1955–1994.

Spurious Regularisation

The version of the highpass Hodrick–Prescott filter that is appropriate to a finite data sequence entails the following matrix transformation:

$$H = Q(\lambda^{-1}I + Q'Q)^{-1}Q'. \quad (13)$$

This is derived in an appendix of the paper. The complementary lowpass filter that determines the trend has the matrix $I - H$. The flexibility of the trend line is determined by the so-called smoothing parameter λ . When $\lambda \rightarrow \infty$, then $H \rightarrow P = Q(Q'Q)^{-1}Q'$, and the trend becomes a straight line.

It will be seen that $HP = H$. From this identity, it follows that

$$(I - H)y = (I - P)y + (I - H)Py, \quad (14)$$

This shows that the output of the lowpass filter can be expressed as the sum of the ordinates $(I - P)y$ of a linear trend and those of the filtered version $(I - H)Py = (I - H)e$ of the residual vector from the linear detrending. Thus it will be recognised that the trend line of Figure 12 can be obtained by adding the trajectory represented by the heavy line of Figure 11 to the linear trend of Figure 4.

Figure 13 represents the residual sequence $Hy = He$ generated by the highpass version of the Hodrick–Prescott filter. The sequence is strongly affected by seasonal fluctuations; and these can be eliminated by applying the frequency-domain filter, represented in the time domain by the matrix D . The resulting sequence $DHy = DHe$ is represented by the heavy line in Figure 13.

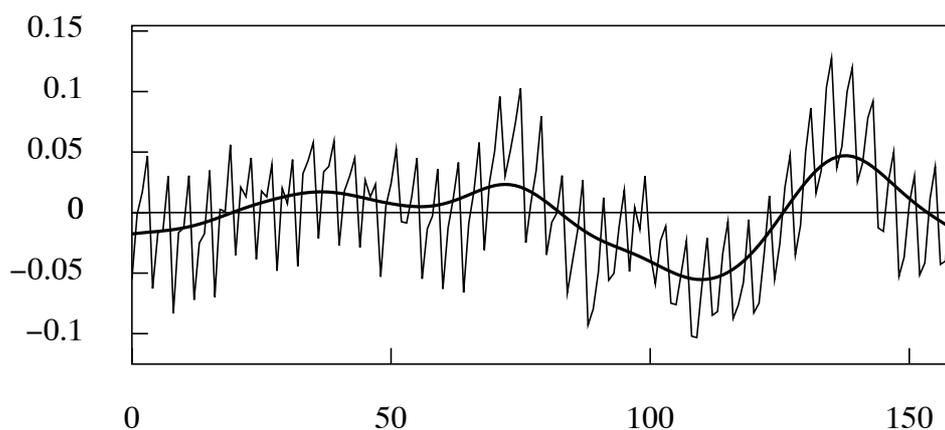


Figure 11. The residual sequence $e = Py$ obtained by extracting a linear trend from the logarithmic consumption data, together with a low-frequency trajectory $(I - H)e$, represented by the heavy line, which has been obtained via the lowpass Hodrick–Prescott filter.

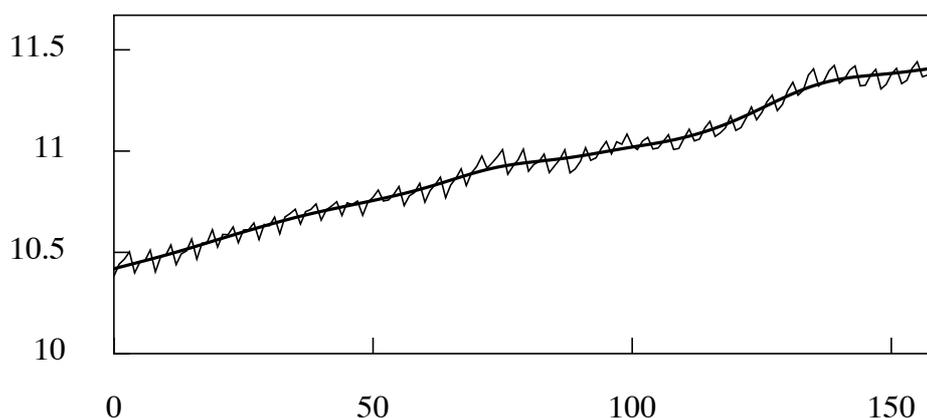


Figure 12. The quarterly logarithmic consumption data together with a trend $(I - H)y = (I - P)y + (I - H)Py$ interpolated by the lowpass Hodrick–Prescott filter with the smoothing parameter set to $\lambda = 1,600$.

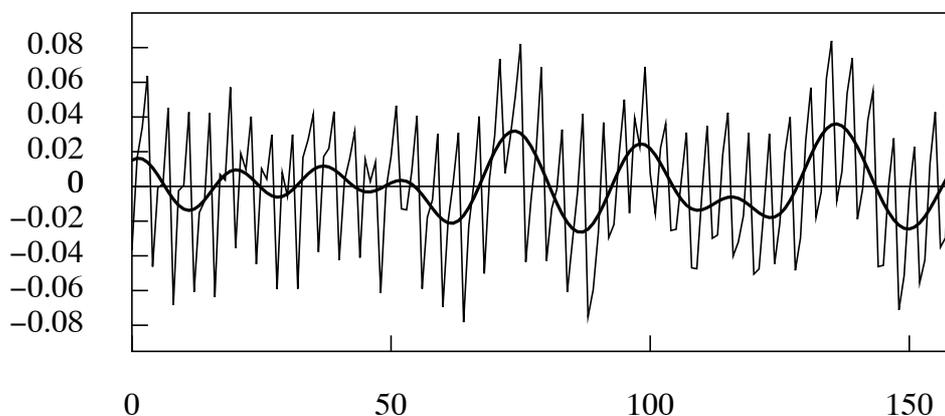


Figure 13. The residual sequence $Hy = He$ obtained by using the lowpass Hodrick–Prescott filter to extract the trend, together with a fluctuating component $DHy = DHe$ obtained by subjecting the sequence to a lowpass frequency-domain filter with a cut-off point at $\pi/8$ radians.

According to the common assertion, this product of the Hodrick–Prescott filter is liable to be affected by spurious fluctuations. Given that the Hodrick–Prescott filter has a gain that never exceeds unity, it cannot amplify elements that are already in the data. Nor can it add anything to the data. Therefore, one must look for other reasons that might justify the aspersions that have been made against the filter.

One effect of the filter that may be problematic is its tendency to regularise the amplitudes of the fluctuations that are present in the data. Fluctuations of large amplitudes appear to be attenuated to a greater extent than those of smaller amplitudes. Therefore, the effect of passing a linearly detrended data sequence through the filter may be to increase significantly the number of times that its trajectory crosses the horizontal axis. This is evident in the comparison of Figures 10 and 13.

It could be said that the effect of the Hodrick–Prescott filter has been to impart a spurious regularity to the fluctuations. However, it cannot be said that the fluctuations of $DHy = DHe$ are spurious or that they have been induced by the filter. The effects of the filter can be explained in reference either to the time domain or to the frequency domain.

For the explanation in the time domain, one can make reference to the least-squares criterion function from which the Hodrick–Prescott filter can be derived. Large deviations from the fitted function are penalised to a greater extent than are the smaller deviations, with the effect that the fluctuations acquire similar amplitudes. Also, the more flexible is the trend function, the greater will be the regularisation of the amplitudes of the residual deviations.

To explain the effect from the point of view of the frequency domain, one can observe that the highpass filter H serves to attenuate the low-frequency motions within y and e , which carry the fluctuations away from the horizontal axis. Therefore, the filtered sequence DHe is liable to cross the axis more often than De does.

These explanations suggest that the effects of the Hodrick–Prescott filter are bound to be shared with other filters that attenuate or remove the low-frequency elements of the data.

It should be observed that, in common with Hodrick–Prescott filter, the ideal frequency-domain filter fulfils a least-square criterion. The Fourier synthesis that constitutes the output of the filter corresponds to the trigonometrical polynomial that provides the least-squares approximation to the data sequence. This result is proved by Baxter and King (1999) in their appendix and by Pollock (1999, p. 375).

Appendix: A Least-Squares Derivation of the H-P Filter

The Hodrick–Prescott filter may be derived as a solution to the problem of finding a smooth curve that adheres closely to the data.

The vector of the data may be denoted by y and that of the ordinates of the curve by x . The goodness of fit is measured by the sum of squares of the deviations of the ordinates from the data. The smoothness of the curve is measured by the sum of squares of the second differences of x , which is given by

the vector $d = Q'x$, where Q' is the matrix of the twofold difference operator. The filter is, therefore, derived via the following criterion:

$$\text{Minimise } (y - x)'(y - x) + \lambda x'QQ'x. \quad (1)$$

Here, λ is the so-called smoothing parameter, which governs the trade-off between the two objectives.

Differentiating the criterion function in respect of x and setting the result to zero gives

$$(y - x) = \lambda QQ'x = \lambda Qd, \quad (2)$$

Premultiplying by $\lambda^{-1}Q'$ gives

$$\lambda^{-1}Q'(y - x) = \lambda^{-1}(Q'y - d) = Q'Qd, \quad (3)$$

whence

$$\begin{aligned} \lambda^{-1}Q'y &= \lambda^{-1}d + Q'Qd \\ &= (\lambda^{-1}I + Q'Q)d, \end{aligned} \quad (4)$$

which gives

$$\lambda d = (\lambda^{-1}I + Q'Q)^{-1}Q'y. \quad (5)$$

Putting this into

$$x = y - \lambda Qd, \quad (6)$$

which comes from (2), gives

$$\begin{aligned} x &= y - Q(\lambda^{-1}I + Q'Q)^{-1}Q'y \\ &= (I - H)y. \end{aligned} \quad (7)$$

Letting $\lambda \rightarrow \infty$ gives

$$\begin{aligned} f &= y - Q(Q'Q)^{-1}Q'y \\ &= (I - P)y. \end{aligned} \quad (8)$$

which is an expression for the vector of the ordinates of a straight line interpolated into the data by least-squares regression. The vector of the residuals is

$$e = Q(Q'Q)^{-1}Q'y = Py. \quad (9)$$

One should observe that

$$\begin{aligned} h &= Hy = Q(\lambda^{-1}I + Q'Q)^{-1}Q'y \\ &= Q(\lambda^{-1}I + Q'Q)^{-1}Q'e, \end{aligned} \quad (10)$$

which is a consequence of the fact that $HP = PH = H$. Thus, the vector $h = y - x = Hy = He$ of the deviations from the Hodrick–Prescott trend can be obtained equally by applying the filter to the data vector y or the residual vector e of a linear regression.

If the lowpass filter is applied to $e = Py$ instead of to y , then the resulting vector $(I - H)e$ can be added to the ordinates of the interpolated regression line $f = (I - P)y$ to create the estimate of the trend $x = (I - H)y$.

References

- Baxter, M., and R.G. King, (1999), Measuring Business Cycles: Approximate Band-Pass Filters for Economic Time Series, *Review of Economics and Statistics*, 81, 575–593.
- Burridge, P., and E. Guerre, (1996), The Limit Distribution of Level Crossings of a Random Walk, *Econometric Theory*, 12, 705–723.
- Chan, K.H., J.C. Hayya and J.K. Ord, (1971), A Note on Trend Removal Methods: The Case of Polynomial Regression versus Variate Differencing, *Econometrica*, 45, 737–744.
- Cogley, T., and J.M. Nason, (1995), Effects of the Hodrick–Prescott Filter on Trend and Difference Stationary Time Series: Implications for Business Cycle Research, *Journal of Economic Dynamics and Control*, 19, 253–278.
- Feller, W., (1968), *An Introduction to Probability Theory and its Applications: Third Edition*, John Wiley and Sons, New York.
- García, A., and A. Sansó, (2006), A Generalization of the Burridge–Guerre Nonparametric Unit Root Test, *Econometric Theory*, 22, 756–761.
- Granger, C.W.J, (1966), The Typical Spectral Shape of an Economic Variable, *Econometrica*, 34, 150–161.
- Harvey, A.C., and A. Jaeger, (1993), Detrending, Stylized Facts and the Business Cycle, *Journal of Applied Econometrics*, 8, 231–247.
- Hodrick, R.J., and E.C. Prescott, (1980), *Postwar U.S. Business Cycles: An Empirical Investigation, Working Paper*, Carnegie–Mellon University, Pittsburgh, Pennsylvania.
- Hodrick, R.J., and E.C. Prescott, (1997), Postwar U.S. Business Cycles: An Empirical Investigation, *Journal of Money, Credit and Banking*, 29, 1–16.
- Howrey, E., (1968), A Spectrum Analysis of the Long Swing Hypothesis, *International Economic Review*, 9, 228–252.
- Ivanov, L., (2005), Is the Ideal Filter Really Ideal: The Usage of Frequency Filtering and Spurious Cycles, *South Eastern Journal of Economics*, 1, 79–96.
- Jaeger, A., (1994), Mechanical Detrending by Hodrick–Prescott Filtering: A Note, *Empirical Economics*, 19, 493–500.
- King, R.G., and S.T. Rebelo, (1993), Low Frequency Filtering and Real Business Cycles, *Journal of Economic Dynamics and Control*, 17, 207–231.
- Kuznets, S.S., (1961), *Capital and the American Economy: Its Formation and Financing*, National Bureau of Economic Research, New York.
- Kaiser, R., and A. Maravall, (1999), Estimation of the Business Cycle: A Modified Hodrick–Prescott Filter, *Spanish Economic Review*, 1, 175–206.
- Leser, C.E.V., (1961), A Simple Method of Trend Construction, *Journal of the Royal Statistical Society, Series B*, 23, 91–107.

- Nelson, C.R., and H. Kang, (1981), Spurious Periodicity in Inappropriately Detrended Time Series, *Econometrica*, 49, 741–751.
- Pedersen, M.T., (2001), The Hodrick–Prescott Filter, the Slutsky Effect, and the Distortionary Effect of Filters, *Journal of Economic Dynamics and Control*, 25, 1081–1101.
- Pollock, D.S.G., (1997), Data Transformation and De-trending in Econometrics, Chapter 11 (pps. 327–362) in Christian Heij et al. (eds.), 1997, *System Dynamics in Economic and Financial Models*, John Wiley and Sons, Chichester.
- Pollock, D.S.G., (1999), *A Handbook of Time-Series Analysis, Signal Processing and Dynamics*, Academic Press, London.
- Pollock, D.S.G., (2000), Trend Estimation and Detrending via Rational Square Wave Filters, *Journal of Econometrics*, 99, 317–334.
- Pollock, D.S.G., (2009), Realisations of Finite-sample Frequency-selective Filters, *Journal of Statistical Planning and Inference*, 139, 1541–1558.
- Schenk-Hoppe, K.R., (2001), Economic Growth and Business Cycles: A Critical Comment on Detrending Time Series, *Studies in Nonlinear Dynamics and Econometrics*, 5, 75–86.
- Slutsky, E.E., (1927), The Summation of Random Causes as the Source of Cyclic Processes, *The Problem of Economic Conditions*, 3:1, 34–64 (English summary, 156–161), Moscow.
- Slutzky, E.E., (1937), The Summation of Random Causes as the Source of Cyclic Processes, *Econometrica*, 37, 105–146 (in Russian 1927).
- Yule, G.U., (1927), On a Method of Investigating Periodicities in Disturbed Series with Special Reference to Wolfer’s Sunspot Numbers, *Philosophical Transactions of the Royal Society, Series A*, 226, 267–298.