

Optimal filtering

Paweł Kowal*

July 26, 2005

Abstract

This paper presents short summary about commonly used filters in macroeconomics. We consider Hodrick-Prescott filter, Butterworth filter, Ideal Bandpass filter, Baxter-King filter and Christiano-Fitzgerald filter.

1 The filtering problem

Given raw data, $\{x_t\}_{t=1}^T$, we are interested in isolating component of x_t , denoted by $\{y_t\}_{t=1}^T$, with period of oscillation between p_l and p_u , where $2 \leq p_l < p_u < \infty$.

In the rest of the paper we closely follow Schleicher (2003), Christiano and Fitzgerald (2003), and Iacobucci and Noullez (2004).

2 The Ideal Bandpass Filter

Consider the following decomposition of the stochastic process, x_t :

$$x_t = y_t + \bar{x}_t$$

The process, y_t , has power only in the frequencies belonging to the interval $\{(a, b) \cup (-b, -a)\} \in (-\pi, \pi)$. The process, \bar{x}_t , has power only in the complement of this interval in $(-\pi, \pi)$. Here, $0 < a < b < \pi$, and

$$a = \frac{2\pi}{p_u}, \quad b = \frac{2\pi}{p_l}$$

In case of infinite amount of data we can use the ideal bandpass filter

$$y_t = B(L)x_t$$

where the filter, $B(L)$, has the following structure:

$$B(L) = \sum_{j=-\infty}^{\infty} B_j L^j, \quad L^k x_t \equiv x_{t-k}$$

*Department of Economics, Warsaw School of Economics, email: pkowal3@sgh.waw.pl

where

$$B_j = \frac{\sin(jb) - \sin(ja)}{\pi j}, \quad j \geq 1$$

$$B_0 = \frac{b - a}{\pi}$$

With this specification we have

$$B(e^{-iw}) = 1, \text{ for } w \in (a, b) \cup (-b, -a)$$

$$= 0, \text{ otherwise.}$$

Thus, the assumption, $a > 0$, implies that $B(1) = 0$.

3 Approximations of the Ideal Bandpass Filter

Consider a general linear filter given by the linear transformation

$$y_t = B(L)x_t = \sum_{j=-\infty}^{\infty} B_j x_{t+j}$$

where $\{B_j\}_{j=-\infty}^{\infty}$ is the impulse response sequence of some linear filter. In finite samples this transformation is not feasible and it is necessary to use an alternative filter with finite impulse sequence $\{\hat{B}_j\}_{j=-n_1}^{n_2}$,

$$\hat{y}_t = \hat{B}_t(L)x_t = \sum_{j=-n_1}^{n_2} \hat{B}_{t,j} x_{t+j}$$

This finite sample filter does not have to be neither symmetric (i.e. $\hat{B}_{t,j} = \hat{B}_{t,-j}$) nor time-invariant (i.e. $\hat{B}_{t,j} = \hat{B}_j$). In line with Kolmogorov and Wiener we are looking for a sequence $\{\hat{B}_{t,j}\}$ that minimizes the mean squared error between y_t and \hat{y}_t

$$\{\hat{B}_{t,j}\} = \arg \min E\{(y_t - \hat{y}_t)^2\} \quad (1)$$

Theorem 3.1. *If $\{X_t\}$ follows a random walk*

$$\Delta X_t = \epsilon_t, \quad \epsilon \sim N(0, \sigma^2)$$

and the optimal filter $\{B_j\}_{j=-\infty}^{\infty}$ satisfies the conditions

1. $\sum_{j=-\infty}^0 |\sum_{k=-\infty}^j B_k|^2 < \infty$
2. $\sum_{j=1}^{\infty} |\sum_{k=j+1}^{\infty} B_k|^2 < \infty$
3. $B_j = B_{-j}$
4. $\sum_{j=-\infty}^{\infty} B_j = 0$

then

$$\hat{B}_j = B_j$$

for $j = -n_1 + 1, \dots, n_2 - 1$ and

$$\hat{B}_{-n_1} = \frac{B_0}{2} - \sum_{j=0}^{n_1-1} B_j \quad \hat{B}_{n_2} = \frac{B_0}{2} - \sum_{j=0}^{n_2-1} B_j$$

Proof. See Schleicher (2003)¹. □

3.1 The Christiano-Fitzgerald Filter

The Christiano-Fitzgerald (CF) Filter (Christiano and Fitzgerald (2003)) can be obtained using theorem 3.1 applying to the Ideal Bandpass Filter. The CF filter is neither symmetric nor time-invariant and thus the CF filter introduce phase shift between \hat{y}_t and y_t .

Simulations conducted by Iacobucci and Noullez applied to the process x_t generated by a random walk indicate that phase shift generated by the CF filter can be large. The absolute value of the phase reaches a maximum of approximately 1.6 quarters for certain components inside the passband. Some components can thus experience shifts up to ± 5 months, causing a maximum relative shift of almost one year. With the introduction of a spurious time- and frequency-dependent phase, timing and correlation properties among different frequency components within the series are irreversibly modified and cannot be recovered. The same happens to the correlation and causality relations among different series.

Simulations conducted by Christiano and Fitzgerald applied to macroeconomic time series indicate that \hat{y}_t appears to be reasonably stationary, except in an area in the tails. This tail area is fairly small (about 1.5 years) for the business cycle frequencies, but it grows for the lower frequencies. Additionally the CF filter seems to imply little phase shift between \hat{y}_t and y_t . There is essentially no phase shift, even at the lowest frequencies.

The CF filter is optimal only in case when the process x_t is a random walk. However Christiano and Fitzgerald (2003) find that there is relatively little gain in knowing the precise details of the time series representation generating the process x_t . In particular, the gain from using the true time series representation of x_t to compute y_t rather than proceeding as though x_t is a random walk, is minimal in practice.

3.2 The Baxter-King Filter

The method proposed in Baxter and King (1999) relies on the use of a symmetric finite odd-order $M = 2K + 1$ moving average so that

$$\hat{y}_t = \sum_{n=-K}^K \hat{B}_n x_{t-n} = \hat{B}_0 x_t + \sum_{n=1}^K \hat{B}_n (x_{t-n} + x_{t+n})$$

¹Schleicher (2003) considers also much more general situations, when the stochastic process $\{X_t\}$ is given by ARIMA model.

The set of M coefficients $\{\hat{B}_j\}$ is obtained by applying condition (1) to the ideal filter coefficients imposing symmetry and stationarity restrictions. The solution takes the form

$$\hat{B}_j = B_j - \frac{1}{M} \sum_{n=-K}^K B_n$$

The BK filter has many desirable properties. First, since it is real and symmetric, it does not introduce phase shifts and leaves the extracted components unaffected except for their amplitude. Second, being of constant finite length and time-invariant, the filter is stationary. The filter is insensitive to deterministic linear trends, provided that $M < N$, so that it is not used near the edges of the series.

On the other hand, filtering in time domain using moving averages, involves the loss of $2K$ data values. In Baxter and King (1999), a value of $K = 12$ for the passband $[1.5, 8]$ years is found to be basically equivalent to higher values, such as 16 or 20. As a consequence, the authors suggest putting $K = 12$ irrespective of the sample size, or the band to be extracted.

4 The Hodrick-Prescott Filter

The HP cyclical component y_t is defined as the difference between the original signal x_t and a smooth growth component g_t . The latter is the solution of the optimization problem

$$\arg \min_{g_t} \sum_{j=1}^N (u_j - g_j)^2 + \lambda \sum_{j=2}^{N-1} (g_{j+1} - 2g_j + g_{j-1})^2 \quad (2)$$

The first term is a measure of "goodness-of-fit" and the second term is a measure of the "degree-of-smoothness" which penalizes decelerations in the growth rate of the trend component. Variations in the smoothing parameter λ alters the trade-off between the goodness-of-fit and the degree-of-smoothness. The moving average representation of the cyclical component is

$$HP(L) = \frac{\lambda(1-L)^2(1-L^{-1})^2}{1 + \lambda(1-L)^2(1-L^{-1})^2}$$

The solution to (2) for $N \rightarrow \infty$ can be found explicitly in the frequency domain (King and Rebelo (1993)) and leads to the following expression for the frequency response function

$$H(\omega) = \frac{4\lambda(1 - \cos(\omega))^2}{1 + 4\lambda(1 - \cos(\omega))^2} = \frac{16\lambda \sin^4(\omega/2)}{1 + 16\lambda \sin^4(\omega/2)} \quad (3)$$

Hence, the impulse responses, B_j can be found by numerical integration of the inverse Fourier transform

$$B_j = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{4\lambda(1 - \cos(\omega))^2}{1 + 4\lambda(1 - \cos(\omega))^2} e^{i\omega j} d\omega$$

From (3) we can find a relation between the cutoff frequency – defined as the frequency for which the response is equal to 0.5 – and the parameter λ

$$\nu_c = \frac{\pi}{\arcsin(\frac{1}{2}\lambda^{-\frac{1}{4}})}$$

$$\lambda = \left[2 \sin\left(\frac{\pi}{\nu_c}\right)\right]^{-4}$$

where $\omega_c = 2\pi/\nu_c$ is the cutoff frequency. That is, $v_c = 39.7$ quarters when $\lambda = 1600$. Hence the HP filter, in the configuration suggested by the authors for quarterly data, selects periodicities shorter than approximately 10 years, but has the disadvantage of a wide transition band.

The transfer function of the HP filter has a smooth transition from the stopband to the passband. An important consequence of this gradual ascend is that a considerably large portion of low frequency components is passed through the highpass filter, a phenomenon that is especially pronounced when the underlying data series is integrated. In this case the HP filter generates strong cycles that lie to the left of the desired passband, similarly to case of approximate bandpass filters.

The optimal finite sample filter can be implemented according to the rules of theorem 3.1. In the middle of the sample both the standard filter as well as the optimal filter approximate the ideal filter very well. Since both filters are symmetric, there is no phase shift. However, at the end of the sample the standard HP approximation exaggerates the gain and experiences a large phase shift at low frequencies. In contrast the optimal finite sample filter underestimates the gain, but has a phase shift that is only half the size of that of the standard HP implementation².

5 The Butterworth Filter

Recently, Pollock (2000) propose the use of a digital Butterworth filter as an approximation to a square wave filter to detrend economic time-series. The Butterworth filter is characterized by a gain function that is maximally flat (in the sense of the best Taylor approximation) in the passband and monotone between pass- and stopbands. This monotonicity comes at the price of a decrease in steepness in the transfer function.

The digital version of the Butterworth highpass filter is described by the rational polynomial expression (the filter's z-transform)

$$\psi_H(z) = \frac{\lambda(1-z)^n(1-z^{-1})^n}{(1+z)^n(1+z^{-1})^n + \lambda(1-z)^n(1-z^{-1})^n}$$

Its time-domain representation (impulse response sequence) can be obtained by substituting z for the lag-operator L , while substituting z for $e^{-i\omega}$ gives the frequency response function. The parameter n is referred to as the order of the filter and determines the steepness of the ascend between the stopband and the passband. The parameter λ determines the cutoff frequency ω_c such that for

²See Schleicher (2003).

the highpass filter described above

$$\lambda = \left[\frac{1}{\tan \frac{\omega_c}{2}} \right]^{2n}$$

and

$$\psi_H(e^{-i\omega})|_{n \rightarrow \infty} = \begin{cases} 1, & \text{if } \omega > \omega_c; \\ 0, & \text{otherwise.} \end{cases}$$

Therefore, while the Butterworth filter shares some similarities with the Hodrick-Prescott filter (monotonicity and flatness), it is more flexible due the fact that the passband and steepness are controlled by two different parameters (for the HP filter both properties are controlled by λ). However increase in steepness of ascend comes at the price of a decrease in filter stability.

Due to the recursive nature of the Butterworth filter problems arise for short and nonstationary data series. The main difficulty is to provide plausible initial and end conditions. A common approach is to extend the sample by forecasting and backcasting, however a bad choice of starting values can affect the entire sample (a problem known as the 'transient effect').

Pollock derives a specialized finite-sample version of the Butterworth filter on the basis of signal extraction theory. This approach is optimal if the data is consistent with the statistical model upon which the filter is based. The model under consideration is given by

$$\begin{aligned} y_t &= s_t + c_t \\ &= \frac{(1+L)^n}{(1-L)^d} \nu_t + (1-L)^{n-d} \epsilon_t \end{aligned}$$

where $\nu \sim N(0, \sigma_\nu^2)$, $\epsilon \sim N(0, \sigma_\epsilon^2)$ and where s_t is the trend component extracted by the Butterworth lowpass filter and c_t is the cyclical component extracted by the Butterworth highpass filter.

The optimal finite sample filter for a random walk model is constructed by first computing the raw coefficients via numerical integration of the inverse Fourier transform and then applying the rule from theorem 3.1.

The standard Butterworth filter and its optimal finite sample approximation provide an almost perfect approximation to the asymptotic filter in the middle of the sample. This may be explained by the fact that the filter coefficients of the asymptotic filter die out fairly quickly. However, at the end of the sample, the filter, as implemented by Pollock, shows a rather strong deviation from its ideal transfer function, both in terms of the gain function as well as the phase shift³.

References

- [1] M. Baxter and R.G. King. Measuring business cycles: Approximate band-pass filters. *The Review of Economics and Statistics*, 81(4):575–93, 1999.
- [2] L. Christiano and T.J. Fitzgerald. The bandpass filter. *International Economic Review*, 44(2):435–65, 2003.

³See Schleicher (2003).

- [3] R.J. Hodrick and E.C. Prescott. Postwar US business cycles: an empirical investigation. *Journal of Money, Credit, and Banking*, 29(1):1–16, 1997.
- [4] R.G. King and S.T. Rebelo. Low frequency filtering and real business cycles. *Journal of Economic Dynamics and Control*, 17, 1993.
- [5] A. Noullez and A. Iacobucci. A frequency-selective filter for short-length time series. *Computing in Economics and Finance 2004* 128, Society for Computational Economics, 2004.
- [6] D.S.G. Pollock. Trend estimation and de-trending via rational square-wave filters. *Journal of Econometrics*, 99:317–334, 2000.
- [7] C. Schleicher. Essays on the decomposition of economic variables. unpublished dissertation, University of British Columbia, Vancouver, Canada, 2003.