Federal Reserve Bank of New York
Staff Reports

Extracting Business Cycle Fluctuations: What Do Time Series Filters Really Do?

Arturo Estrella

Staff Report no. 289
June 2007

Extracting Business Cycle Fluctuations: What Do Time Series Filters Really Do?
Arturo Estrella
Federal Reserve Bank of New York Staff Reports, no. 289
June 2007
JEL classification: C22, E32

Abstract

Various methods are available to extract the "business cycle component" of a given time series variable. These methods may be derived as solutions to frequency extraction or signal extraction problems and differ in both their handling of trends and noise and their assumptions about the ideal time-series properties of a business cycle component. The filters are frequently illustrated by application to white noise, but applications to other processes may have very different and possibly unintended effects. This paper examines several frequently used filters as they apply to a range of dynamic process specifications and derives some guidelines for the use of such techniques.

Key words: frequency domain, spectral analysis, signal extraction

---

Estrella: Federal Reserve Bank of New York (e-mail: arturo.estrella@ny.frb.org). The views expressed in this paper are those of the author and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System.

1. Introduction

The analysis of how a typical macroeconomic time series behaves over the business cycle is complicated by the fact that its movements may contain low-frequency trends and high-frequency noise. Various methods are available to extract “the business cycle component” of a given time series variable. These methods differ in their handling of trends and noise, and in their assumptions about the time-series properties of a business cycle component.

Mechanical use of these filters without careful consideration of the characteristics of the particular problem or setting may lead to inferior results for at least two reasons. First, the objectives of the exercise may vary. For example, in some cases the ideal result will only contain variation in certain segments of the frequency spectrum, whereas in other cases the ideal result will contain variation across the full spectrum.

Second, the consequences of applying a given filter may vary substantially, depending on the time series properties of the process to which it is applied. The effects of various filters are often illustrated by plotting their gains, which essentially amounts to showing the results of applying the filters to white noise. Application to other processes may produce results that are very different, both qualitatively and quantitatively.

To address these issues, this paper examines two general approaches to the construction of business cycle components. The approaches are based on tools derived as solutions to two different statistical problems: frequency extraction and signal extraction. A filter designed for one of these purposes may or may not be acceptable when applied to the other. Moreover, the analysis shows that the consequences of applying a given filter to processes other than white noise may be quite different, with possible unintended effects.

No one filter emerges as the best solution across the board, but it is possible to derive certain guidelines for the selection of filters in various settings. As suggested above, the criteria depend on the exact goals of the exercise and on the time series properties of the variables involved.

Sections 2 and 3 describe the frequency and signal extraction problems, respectively, and show how various filters are solutions to these problems. Section 4 applies the filters to a range of processes to see how they interact. Section 5 discusses the issue of trends in the data. Section 6 provides some empirical illustrations and Section 7 offers some conclusions and guidelines.

2. The frequency extraction problem

2.1 Objective of frequency extraction

Let y_t be a time series, t=1,…,T, such that

$$\left(1-L\right)^d y_t = A(L)\varepsilon_t \,, \qquad (1)$$

where d is a non-negative integer, $A(L)=1+a_1L+a_2L^2+\ldots$ is a lag polynomial with $\sum_{j=1}^{\infty} a_j^2 < \infty$, and ε_t is white noise. Thus, $A(L)\varepsilon_t$ is the moving average Wold representation of a variance-stationary purely non-deterministic random process.

The objective in frequency extraction problems is to estimate the component of y_t that fluctuates cyclically at frequencies in a range that corresponds to some notion of the business cycle. For instance, we may think of the business cycle fluctuations of a given variable as containing components with a range of specific frequencies, say from 1 to 8 years per cycle. Sargent (1987) proposes that this range consists of the "frequencies at which most aggregates have most of their spectral power if they have 'typical' spectral shapes."

Focusing on variation at individual frequencies or ranges of frequencies is accomplished most efficiently by operating in the frequency domain, rather than in the time domain.[1] Time series variation may be decomposed into orthogonal components corresponding to individual frequencies. Since the relevant functions are periodic, we may restrict attention to frequencies ranging from $-\pi$ to π. Moreover, symmetry permits focusing on frequency values in [0, π].

For simplicity, assume initially that we associate the business cycle with frequencies $\{\omega : 0 < \omega_0 \le \omega \le \pi\}$ and that we take lower frequencies to be associated with trends in y_t.[2] The frequency extraction problem is then to retain only cycles of length $2\pi/\omega_0$ or less with minimum possible distortion of the variability of the included individual frequency components. We consider three alternative approaches: frequency domain, Baxter-King, and Christiano-Fitzgerald.

[1] Useful surveys of frequency domain or spectral analysis techniques are Brillinger (1981), Koopmans (1995) and Sargent (1987).

[2] In the literature, the business cycle is most often associated with a range of frequencies that is also bounded above by a frequency strictly less than π, in order to censor high frequency noise. See the discussion of band-pass filters in Section 2.5. An exception is the "medium-term business cycle" as defined by Comin and Gerter (2006).

2.2 Frequency domain filter (FD)

The FD filter is the ideal solution to the frequency extraction problem and is defined as follows. For $-\pi \le \omega \le \pi$, let $f(\omega)$ be a function such that

$$f(\omega) = \begin{cases} 0 & if\ |\omega| < \omega_0 \\ 1 & if\ \omega_0 \le |\omega| \le \pi \end{cases} \quad (2)$$

and let $F(L)$ be its inverse Fourier transform. The cyclical component of y_t is then defined as

$$c_t = F(L)y_t. \quad (3)$$

It contains frequency components only in the range $\omega_0 \le |\omega| \le \pi$, with the same weights that those components have in y_t. $F(L)$ is the time-domain representation of the high-pass frequency-domain filter $f(\omega)$, and its coefficients are given by

$$F(L) = 1 - \frac{\omega_0}{\pi} - \sum_{h=1}^{\infty} \frac{\sin(\omega_0 h)}{\pi h}\left(L^h + L^{-h}\right). \quad (4)$$

Rather than applying the time domain filter (4), which in principle requires an infinite sample, a straightforward way of estimating c_t is to take the Fourier transform of y_t, say $\tilde{y}(\omega)$ and to calculate c_t as the inverse Fourier transform of the product $f(\omega)\tilde{y}(\omega)$. Note that this filter may be applied to integrated processes because it annihilates the spectrum in a neighborhood of $\omega = 0$, where the spectrum is singular when $d > 0$.

2.3 Baxter-King filter (BK)

Baxter and King (1999) propose a time-domain approximation to the frequency domain filter $f(\omega)$. They define an approximation $B(L)$ that is optimal in the sense that it minimizes the equally-weighted average square modulus of the difference between the frequency domain representation of the FD filter, $f(\omega)$, and the frequency domain representation of the approximation, $b(\omega)$. Specifically, the objective is to minimize

$$\int_{-\pi}^{\pi} \left| f(\omega) - b(\omega) \right|^2 d\omega \quad (5)$$

subject to restrictions that (1) $B(L)$ be symmetrical

$$B(L) = b_0 + \sum_{h=1}^{K} b_h \left(L^h + L^{-h} \right) \qquad (6)$$

and that (2) its coefficients sum to zero

$$B(1) = b_0 + 2\sum_{h=1}^{K} b_h = 0\,. \qquad (7)$$

The solution to the optimization problem is

$$b_0 = 1 - \omega_0 / \pi - \theta \qquad (8)$$

$$b_h = -\sin(h\omega_0)/(h\pi) - \theta \qquad (9)$$

for h = 1, … , K, where θ is chosen to satisfy condition (7). The cyclical component of y_t is estimated as

$$\hat{c}_t = B(L) y_t\,. \qquad (10)$$

The BK filter is optimal under the stated conditions in the sense that it minimizes the mean squared error when the series y_t is white noise. The two restrictions (6) and (7) imply that the BK filter may be applied to integrated series up to I(2) and still produce stationary results, a feature shared with filters designed for signal extraction, as discussed in Section 3 below. If the zero-sum restriction (7) is not imposed, $\theta = 0$ and the coefficients b_h, h = 0, … , K, are the same as in the time domain representation of the ideal FD filter $F(L)$.

2.4 Christiano-Fitzgerald filter (CF)

Christiano and Fitzgerald (2003) propose a filter that solves an optimization problem similar to the one that leads to the BK filter, with a few notable differences. The objective function for the CF filter is also a mean squared error, but it differs from the one used by Baxter and King (1999) in that squared deviations between the approximate filter $c(\omega)$ and the ideal filter $f(\omega)$ are weighted by $s_{yy}(\omega)$, the spectrum of y_t:

$$\int_{-\pi}^{\pi} \left| f(\omega) - c(\omega) \right|^2 s_{yy}(\omega) d\omega\,. \qquad (11)$$

The two criteria are the same if y_t is white noise, which has a constant spectrum. The CF filter, however, is derived under the assumption that y_t follows a random walk. As in the BK filter, the CF time domain coefficients are constrained to sum to zero, so the filter can deal with the single unit root implicit in the random walk assumption. In contrast to the BK filter, the symmetry restriction is not imposed on the coefficients of y_t, t=1,…,T, all of which may be nonzero.

A simple way to think about the calculation of the CF filter is to extend the data sample {y_t, t=1,…,T} infinitely in both directions by taking $y_t = y_1$ for $t < 1$ and $y_t = y_T$ for $t > T$. This extension is motivated by the predictive properties of the random walk assumption. The ideal weights (4) are then applied to the extended sample. It follows directly from this description of the filter that it is asymptotically ideal in the sense that it approaches the ideal filter $F(L)$ as the sample size approaches infinity in both directions.

2.5 Band-pass filters

The business cycle is frequently associated in the literature with a range of frequencies that is bounded both below and above, in order to censor both low frequency trends and high frequency noise.[3] The upper and lower bounds may be implemented using band-pass filters. Let $\Phi_{\bar{\omega}}(L)$ be a high-pass filter (FD, BK or CF), calibrated to retain frequencies $\bar{\omega}$ and above. Then

$$\Phi_{\omega_0}(L) - \Phi_{\omega_1}(L) \qquad (12)$$

is a band-pass filter that extracts (retains) frequencies ω in the range $0 < \omega_0 \le \omega < \omega_1 < \pi$.

The bulk of the analysis that follows focuses on high-pass – rather than band-pass – versions of filters for two principal reasons. First, comparisons across filters are clearer if we focus on the effects of a single application of the high-pass filter rather than the two applications implicit in the band-pass. Second, the signal extraction filters of Section 3 below do not have an explicit band-pass form. They may be nevertheless interpreted as approximate high-pass filters, as is common in the literature. Thus, comparison of both frequency extraction and signal extraction filters is better carried out by focusing on high-pass formats.

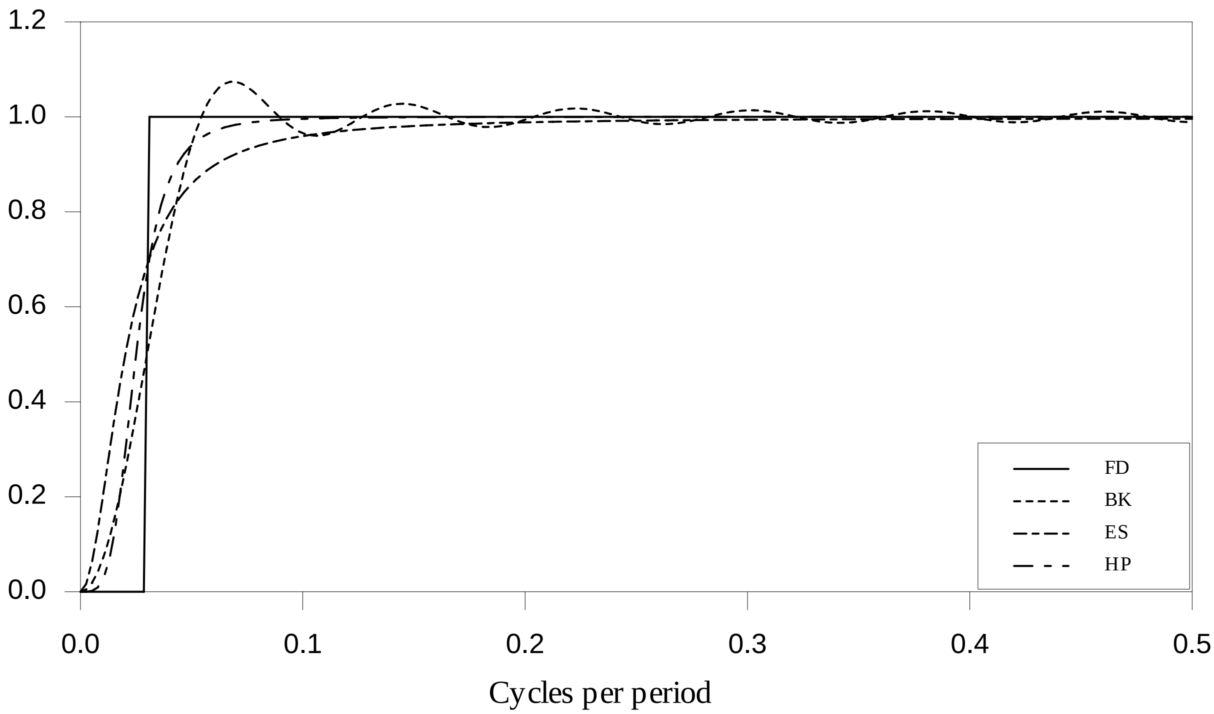
2.6 Graphical comparisons

We illustrate the relative effects of the filters discussed so far by presenting the gain of each filter, that is, the multiplier that affects the variability of each frequency component in the interval [0, π]. Figure 1 shows the gains of the high-pass FD and BK filters for cycles up to 8

[3] See, e.g., Engle (1974) and Sargent (1987).

years in length ($\omega_0 = 2\pi/32$), assuming for simplicity that the sample is infinite.[4] ES and HP filters are discussed in Section 3.

Figure 1. Gain of high-pass filters

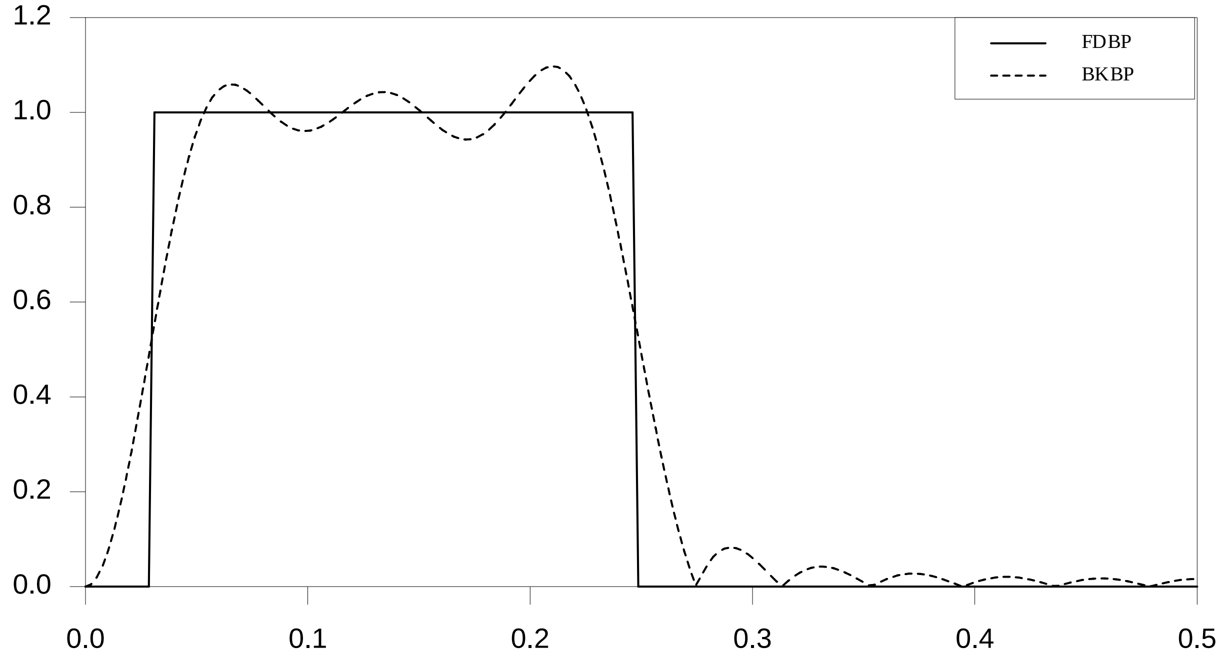


When examined in isolation from the processes to which they are applied, the filters seem relatively similar. The BK filter is least accurate in the neighborhood of the step at $\omega_0 = 2\pi/32$, but it tends to dampen frequencies below that level and is close to unity for higher frequencies. Section 4 will show that looking at the filters in isolation from the process to which they are applied may underestimate the potential for inaccuracy.

Figure 2 shows the gain for the band-pass filters corresponding to the frequency extraction problem, with an upper frequency bound of $\omega_1 = 2\pi/4$ or four-quarter cycles. The qualitative features are similar to Figure 1, with the caveat that the BK approximation has to deal with two steps here instead of one.

[4] With an infinite sample, the CF filter coincides with the FD, as noted earlier, and is not shown separately in the figure.

Figure 2. Gain of band-pass filters



3. The signal extraction problem

3.1 General principles

Let y_t be a time series, t=1,…,T, such that

$$y_t = g_t + c_t \tag{13}$$

where

$$(1-L)^d g_t = A(L)\varepsilon_t \tag{14}$$

$$c_t = A(L)\eta_t \tag{15}$$

Where ε_t and η_t are mutually independent white noise series, $A(L) = 1 + a_1 L + a_2 L^2 + \ldots$ is a lag polynomial with $\sum_{j=1}^{\infty} a_j^2 < \infty$, and d is a positive integer. Note that the stationary series on the right hand sides of (14) and (15) share the same lag polynomial, though they are driven by orthogonal white noise processes.

In the frequency extraction problem, y_t was defined as an I(d) process. In the signal extraction problem, y_t has an I(d) component g_t defined in the same way as before, but here y_t has an additional stationary component c_t, which is the object of investigation.

Whittle (1983, Section 5.1) shows that the least-squares estimate of the trend component is

$$\hat{g}_t = \frac{1}{1+\lambda(1-L)^d(1-L^{-1})^d}\, y_t \tag{16}$$

and that the corresponding estimate of the cyclical component is thus

$$\hat{c}_t = \frac{(1-L)^d(1-L^{-1})^d}{1/\lambda+(1-L)^d(1-L^{-1})^d}\, y_t \tag{17}$$

where $\lambda = \sigma_\eta^2 / \sigma_\varepsilon^2$.[5] As in the frequency extraction problem, the estimate $\hat{c}_t$ minimizes the mean squared error, though the characteristics of the benchmark c_t are different here.

3.2 Exponential smoothing filter (ES).

When $d=1$, we obtain from (17) an estimate $\hat{c}_t = S(L)y_t$ of the cyclical component by applying the filter

$$S(L) = \frac{(1-L)(1-L^{-1})}{1/\lambda+(1-L)(1-L^{-1})} \tag{18}$$

which corresponds to exponential smoothing (ES). The ES filter may also be obtained by minimizing a function of the form

$$\sum\left[(y_t-\hat{g}_t)^2+\lambda(\Delta\hat{g}_t)^2\right] \tag{19}$$

which penalizes both deviations of the estimated trend from the observed value and changes in the trend. The second term produces smoothness in the trend. See King and Rebelo (1993).

3.3 Hodrick-Prescott filter (HP).

When $d=2$, we obtain from (17) an estimate $\hat{c}_t = H(L)y_t$ of the cyclical component by applying the filter

[5] Similar derivations, at various levels of generality, are provided in King and Rebelo (1993), Harvey and Jaeger (1993), and Harvey and Trimbur (2003).

$$H(L) = \frac{(1-L)^2(1-L^{-1})^2}{1/\lambda + (1-L)^2(1-L^{-1})^2} \tag{20}$$

which corresponds to the HP filter of Hodrick and Prescott (1997). The HP filter may also be obtained by minimizing a function of the form

$$\sum \left[(y_t - \hat{g}_t)^2 + \lambda(\Delta^2 \hat{g}_t)^2 \right] \tag{21}$$

which penalizes both deviations of the estimated trend from the observed value and the second difference in the trend. The second term produces more smoothness in the trend than the corresponding term for the ES filter. See King and Rebelo (1993), Hodrick and Prescott (1997) and Ehlgen (1998).

Note that the cyclical component c_t is defined here in a fundamentally different way than in the frequency extraction problem, though the language used in connection with both types of filter may be similar. The objective of the ES and HP filters is to extract a stationary variable that may contain a large amount of low frequency variation. For instance, when $A(L) = 1$, the ideal result is white noise, which contains all frequencies with equal weights. Thus, in that case, the filter would ideally give the same weight to all frequencies, including low frequencies that would be eliminated as trends in the frequency extraction problem. In practice, the ES and HP filters dampen low frequencies, but not to the same extent as the ideal high-pass filter. Refer once again to Figure 1 for illustrations of the gain of the ES and HP filters.

4. Effects of filtering on specific processes

4.1 Strategy for comparing filters

It was noted earlier that comparison of the gains of the individual filters in isolation may be misleading because the interaction with the spectral characteristics of some processes may play a large role in the final result. In this section, the filters are applied to a range of specific processes that vary as to two aspects: the lag polynomial $A(L)$ of the Wold representations of the stationary components of the series and the degree of integration of the series.

4.1.1 Wold representations

For the Wold representations, we follow Ehlgen (1998) in selecting the following four processes, which contain a reasonable degree of spectral diversity.

White noise, WN: $A(L)=1$

Moving average, MA(1): $A(L)=1+.5L$

Autoregressive process, AR(1,.5): $A(L)=\left(1-.5L\right)^{-1}$

Persistent autoregressive process, AR(1,.9): $A(L)=\left(1-.9L\right)^{-1}$

4.1.2 Degree of integration

By construction, the BK, ES and HP filters are applicable to I(2) processes, so we allow the degree of integration to be as high as 2.[6] In the context of the frequency extraction problem, this calls for I(0), I(1) and I(2) cases. The variance of the innovation is taken to be unity: $\sigma_{\varepsilon}^{2}=1$ in all cases.

In the context of the signal extraction process, there is also a stationary component, and we denote the corresponding processes as I(1)+I(0) and I(2)+I(0). The variance of the innovation in the integrated component is taken to be unity: $\sigma_{\varepsilon}^{2}=1$ in all cases. The signal extraction setup also requires the specification of the parameter λ, which represents the ratio of the variances of the two noise processes. For the I(2)+I(0) case, we use $\lambda=1600$, which is the figure proposed by Hodrick and Prescott (1997) for quarterly data. For the I(1)+I(0) case, we use $\lambda=3200\left(1-\cos\left(\pi/16\right)\right)$, which is the ES filter parameter value suggested by King and Rebelo (1993) so that the gain at $\omega=2\pi/32$ (eight year cycle) is the same as that of the HP filter with $\lambda=1600$.

These five cases of degree of integration are combined with each of the four Wold processes defined above to produce a total of twenty combinations. The four filters FD, BK, ES and HP are then applied to each of the combinations. We first illustrate a few cases graphically and then examine all cases numerically to gauge the accuracy of the filters in various conditions.

4.2 Graphical illustrations

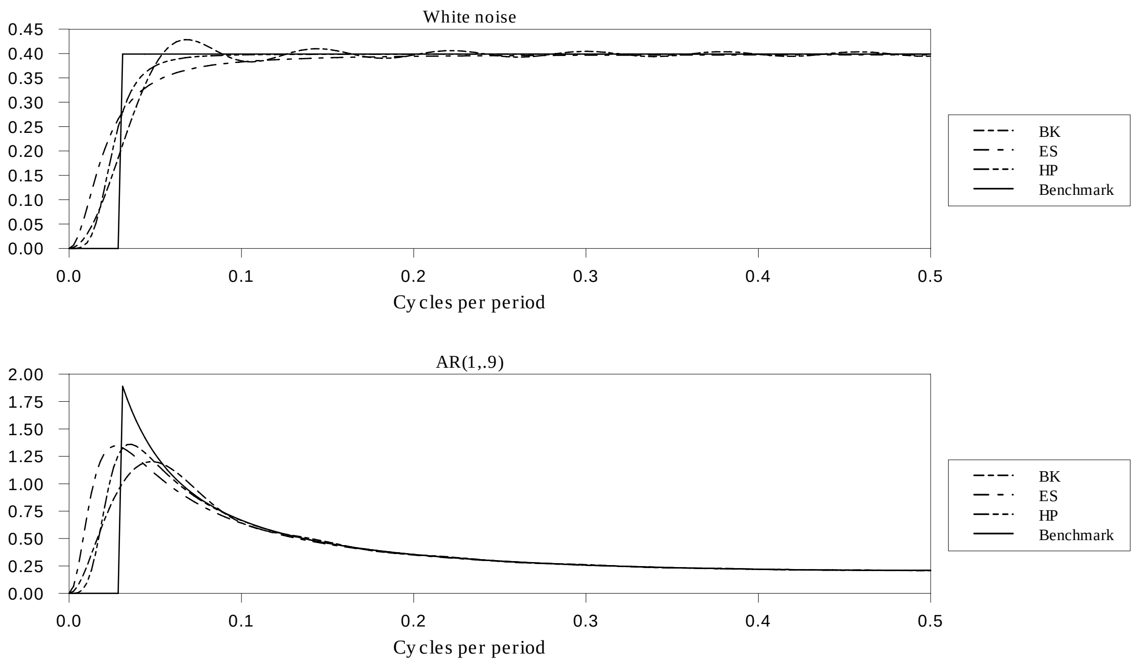
In the graphical illustrations, we focus on the white noise and AR(1,.9) processes, which best represent the range of results. The MA(1) and AR(1,.5) specifications are qualitatively

[6] The HP filter may be applied to integrated processes up to I(4) as well.

similar to the white noise case in terms of their low frequency performance, and all the filters tend to perform fairly well at high frequencies. In general, the most interesting test case turns out to be the persistent AR(1,.9) process.

In the illustrations based on the frequency extraction problem, the benchmark of performance is the ideal FD filter, which is shown in each figure as a solid line.

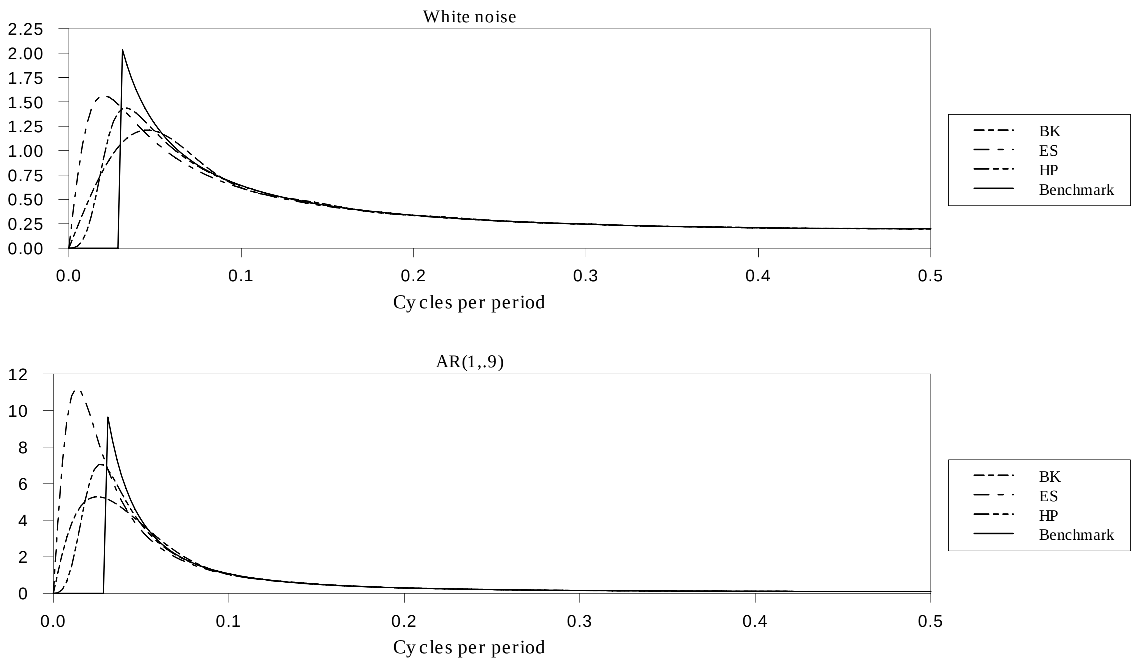
Figure 3. Frequency extraction problem with I(0) process



Since the spectrum of white noise is constant, the top panel of Figure 3 is essentially the same as Figure 1, except that the vertical axis is rescaled by a factor of $1/\sqrt{2\pi}$. The BK, ES and HP filter overestimate the low frequency components below the 32 period frequency and tend to underestimate above that frequency. In general, however, the performance for all filters seems reasonably good.

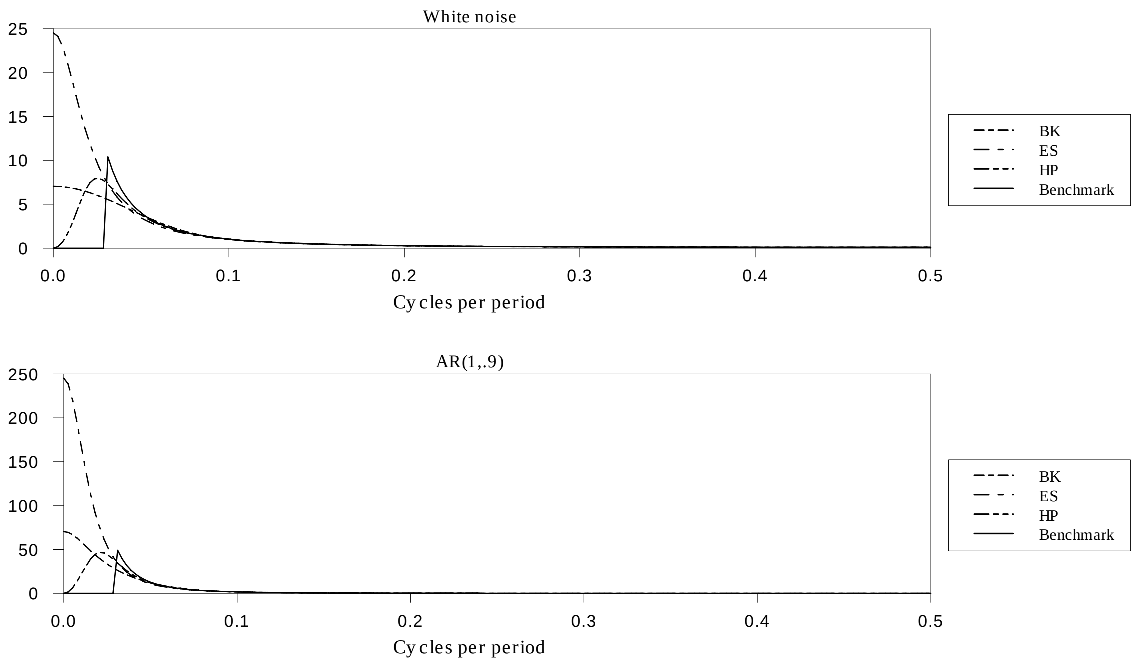
The characteristics of the AR(1,.9) case in the lower panel are similar, but the low frequencies play a larger role here and the overestimation of low frequency trends seems to be more of a problem in relative terms.

Figure 4. Frequency extraction problem with I(1) process



The I(1) processes in Figure 4 show the same low frequency problem as the AR case in the previous figure. The white noise case, a random walk, is similar to the stationary but persistent AR case. The AR case here is even worse than before, particularly for the ES filter.

Figure 5. Frequency extraction problem with I(2) process

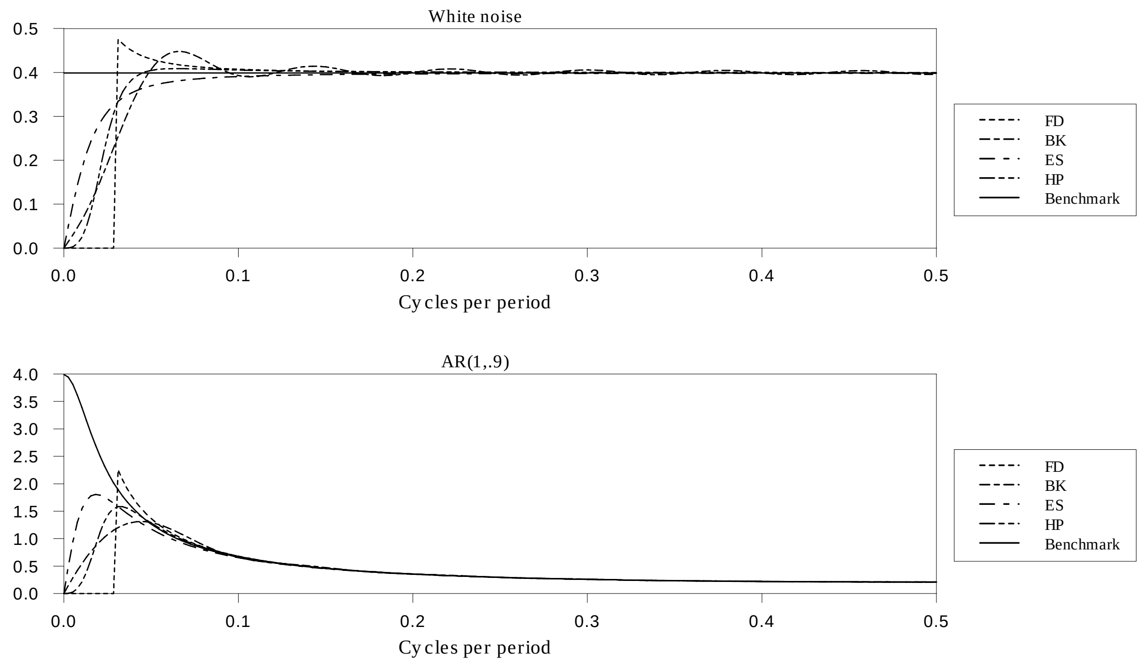


When the process is I(2), Figure 5 shows that all the approximate filters fail dramatically at low frequencies, even though the are designed to extract implicitly up to two unit roots. The BK and ES filters produce finite but very large values at $\omega = 0$. The HP filter, which can extract up to four unit roots, still produces a zero value at $\omega = 0$, but overestimates substantially other low frequency values of the spectrum up to the 32-period threshold.[7]

For cases based on the signal extraction problem, recall that the objective is to produce a cyclical component corresponding to the particular $A(L)$ or Wold representation, which in general contains non-zero low frequency components. This benchmark process appears in the Figures 6 and 7 as a solid line.

[7] Ehlgen (1998) has examined a different but related aspect of the interaction between the HP filter and the various time series processes.

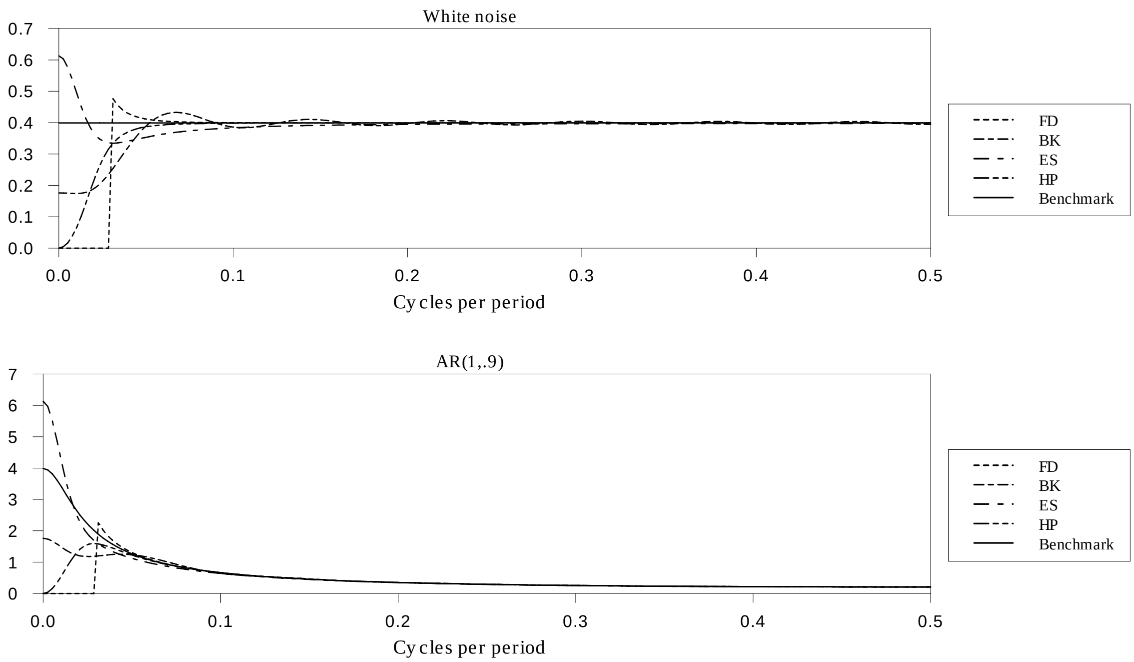
Figure 6. Signal extraction problem with I(1)+I(0) process



With an I(1)+I(0) process, all filters underestimate the low-frequency components of the spectrum. In Figure 6, the numerical illustrations use $\lambda = 3200\left(1-\cos\left(\pi/16\right)\right)$, as in the ES filter. The low-frequency values produced by the BK, ES and HP filters, which were too high in the frequency extraction problem are in this case not high enough. The FD filter, which assigns no weight to low frequencies, fails dramatically here in that range. We see that the ES filter produces the best low-frequency results, as expected from theory.

Results for the I(2)+I(0) processes are presented in Figure 7, with $\lambda = 1600$. The BK and ES filters have finite non-zero values at the zero frequency. For low frequencies more generally, the BK filter tends to approximate the benchmark, while the ES filter tends to overshoot on the positive side. It is not altogether clear from the graphics that the HP filter is best in this case, as theory suggests. The one-dimensional representation of the gain fails to capture the ability of the HP filter to strike a balance in its effects on the two components of the I(2)+I(0) process. The appropriate relationships emerge clearly when exact mean squared errors are calculated numerically in the next section.

Figure 7. Signal extraction problem with I(2)+I(0) process



4.3 Frequency domain errors (RMSEs)

Visual representations are helpful, but a quantitative measure of goodness of fit allows for more precise comparisons of the accuracy of the estimates produced by the filters. In each case, we use the root mean square error of the estimate of the cyclical component, which is equivalent to the optimization criteria proposed by Christiano and Fitzgerald (2003) and Whittle (1983).

Let $u_t = c_t - \hat{c}_t$ be the estimation error, where the cyclical component is estimated by $\hat{c}_t = \hat{F}(L)y_t$ and $\hat{f}(\omega) = \hat{F}\left(e^{-i\omega}\right)$. The frequency domain representation of the RMSE is

$$\sigma_u = \left(\int_{-\pi}^{\pi} \left| f(\omega) - \hat{f}(\omega) \right|^2 s_{yy}(\omega) d\omega \right)^{1/2} \qquad (22)$$

in the frequency extraction problem and

$$\sigma_u = \left(\int_{-\pi}^{\pi} \left\{ \left| 1 - \hat{f}(\omega) \right|^2 s_{cc}(\omega) + \left| \hat{f}(\omega) \right|^2 s_{gg}(\omega) \right\} d\omega \right)^{1/2} . \qquad (23)$$

in the signal extraction problem. In the tables that follow, RMSEs are scaled by $\sigma_c = \left(\int_{-\pi}^{\pi} s_{cc}(\omega) d\omega \right)^{1/2}$, the volatility of the benchmark for each problem. Filter rankings for a given process are not affected, but the interpretation of results expressed this way is scale-independent.

Table 1. Frequency extraction problem: Frequency domain root-mean-squared errors

	I(0)				I(1)				I(2)			
	WN	MA(1)	AR(1,.5)	AR(1,.9)	WN	MA(1)	AR(1,.5)	AR(1,.9)	WN	MA(1)	AR(1,.5)	AR(1,.9)
FD	0	0	0	0	0	0	0	0	0	0	0	0
BK	0.088	0.120	0.156	0.332	0.390	0.417	0.466	0.780	1.119	1.131	1.197	2.258
ES	0.121	0.166	0.216	0.512	0.661	0.707	0.797	1.532	2.801	2.835	3.019	6.379
HP	0.090	0.124	0.162	0.360	0.415	0.444	0.497	0.801	0.930	0.940	0.987	1.485

Results for the frequency extraction problem framework, presented in Table 1, exhibit some clear general patterns. The benchmark in the frequency domain problem is the ideal FD filter and, by definition, it has the best performance in Table 1. However, the table helps ascertain whether the other filters provide good approximations and, if so, under what circumstances. In particular, what is the second best filter under each of the various conditions?

Several patterns are manifest in the table. First, the degree of integration plays an important role in the accuracy of the approximate filters. The results for I(1) processes are markedly worse than in the stationary cases, and the I(2) figures are worse by an order of magnitude. Clearly, the fact that the approximate filters annihilate two to four unit roots is no comfort as far as the accuracy of results in the frequency extraction problem is concerned.

Among the four Wold specifications, results for the approximate filters in the AR(1,.9) case are clearly inferior to the others. Even when the process is I(0), the RMSEs are about four times as large as for WN. Since many economic series exhibit substantial autocorrelation, this pattern suggests caution when applying approximate filters to economic variables.

As to the second best filter, the results of the BK and HP filters are very similar for I(0) and I(1) processes, but the BK has a slight edge in these cases. In the I(2) cases, the HP filter is clearly aided by its ability to deal with up to four unit roots and is better than the other approximations. The differences from the ideal filter, however, are still quite large. Barring some powerful reason to avoid frequency domain calculations, the table suggests that the FD filter is to be preferred.

Table 2. Signal extraction problem: Frequency domain root-mean-squared errors

	I(1)+I(0)				I(2)+I(0)			
	WN	MA(1)	AR(1,.5)	AR(1,.9)	WN	MA(1)	AR(1,.5)	AR(1,.9)
FD	0.298	0.392	0.485	0.854	0.267	0.358	0.454	0.844
BK	0.265	0.346	0.426	0.771	0.247	0.331	0.420	0.824
ES	0.252	0.329	0.405	0.741	0.328	0.439	0.561	1.195
HP	0.263	0.344	0.423	0.778	0.237	0.317	0.402	0.786

In the signal extraction problem, the goal is to obtain the stationary series corresponding to the given $A(L)$ specification. Thus, all four filters are approximations and have positive RMSEs. For each process, the most accurate filter is known by construction, since the ES filter is derived to be optimal in the I(1)+I(0) cases, whereas the HP filter is optimal in the I(2)+I(0) cases.

In contrast to the frequency extraction problem, Table 2 shows that the FD filter is outperformed here by all the others in the I(1)+I(0) case, and by all but ES in the I(2)+I(0) case. Another contrast with the previous table is that the results in Table 2 are not very sensitive to the degree of integration of the series. Although all the cases examined contain an integrated component, they all contain also an I(0) component whose innovation is more variable. This stationary component is clearly very influential in the comparative results. As in Table 1, the AR(1,.9) results are clearly worse than in the other cases.

4.4 Volatility distortion

Another measure of the error involved in estimating the cyclical component is a possible distortion of the overall cyclical variability of the process. Thus, we can look at the variance ratio $\sigma_{\hat{c}}^2 / \sigma_c^2$ as an indicator of this overall distortion in variability, where c_t is the benchmark cyclical process and $\hat{c}_t$ is a particular estimate. Results are shown in Tables 3 and 4 for the frequency and signal extraction problems, respectively.

There are essentially two types of situations, not unrelated, in which the overall volatility tends to be distorted. One is in the frequency extraction problem when the filters are applied to

I(2) processes, as seen in the last four columns of Table 3. The results of applying the ES filter to the autoregressive I(1) processes are similar but less pronounced. The other situation is in the signal extraction problem when the Wold representation is AR(1,.9). In these cases, the estimates fail to capture the large trend-like components of the persistent AR(1,.9) process and, with one exception, they underestimate the volatility by more than 30%.

Table 3. Frequency extraction problem: Volatility relative to benchmark

	I(0)				I(1)				I(2)			
	WN	MA(1)	AR(1,.5)	AR(1,.9)	WN	MA(1)	AR(1,.5)	AR(1,.9)	WN	MA(1)	AR(1,.5)	AR(1,.9)
FD	1	1	1	1	1	1	1	1	1	1	1	1
BK	0.996	0.993	0.987	0.968	0.979	0.976	0.974	1.086	1.341	1.349	1.396	2.347
ES	0.982	0.973	0.968	1.022	1.098	1.114	1.158	1.723	2.908	2.940	3.116	6.422
HP	0.996	0.993	0.989	1.001	1.016	1.019	1.029	1.169	1.257	1.262	1.293	1.687

Table 4. Signal extraction problem: Volatility relative to benchmark

	I(1)+I(0)				I(2)+I(0)			
	WN	MA(1)	AR(1,.5)	AR(1,.9)	WN	MA(1)	AR(1,.5)	AR(1,.9)
FD	1	1	1	1	1	1	1	1
BK	0.996	0.993	0.987	0.968	0.979	0.976	0.974	1.086
ES	0.982	0.973	0.968	1.022	1.098	1.114	1.158	1.723
HP	0.996	0.993	0.989	1.001	1.016	1.019	1.029	1.169

5. Differencing, over-differencing, and deterministic trends

5.1 Pre-differencing integrated processes

As before, let y_t be a time series, t=1,…,T, such that

$$(1-L)^d\, y_t = A(L)\varepsilon_t \tag{24}$$

Frequency extraction approaches generally work best when the observable series is stationary. In principle, the FD filter annihilates the spectrum for frequencies in a neighborhood of zero. Similarly, the CF filter may be applied to I(1) processes, the BK and ES filters may be used with I(2) processes, and the HP filter may even be applied to I(4) processes. However, the damping effects of these filters for low frequencies may be limited in practical applications, and the retained low frequency components may contain substantial trend-like properties. Thus, a standard first step, particularly if the filter is applied in the frequency domain, is to extract all unit roots from y_t and to focus on the stationary variable

$$(1-L)^d\, y_t \tag{25}$$

whose spectrum is finite as $\omega \to 0$.

The difference operator

$$\Delta^d(L) = (1-L)^d \tag{26}$$

has the frequency domain representation

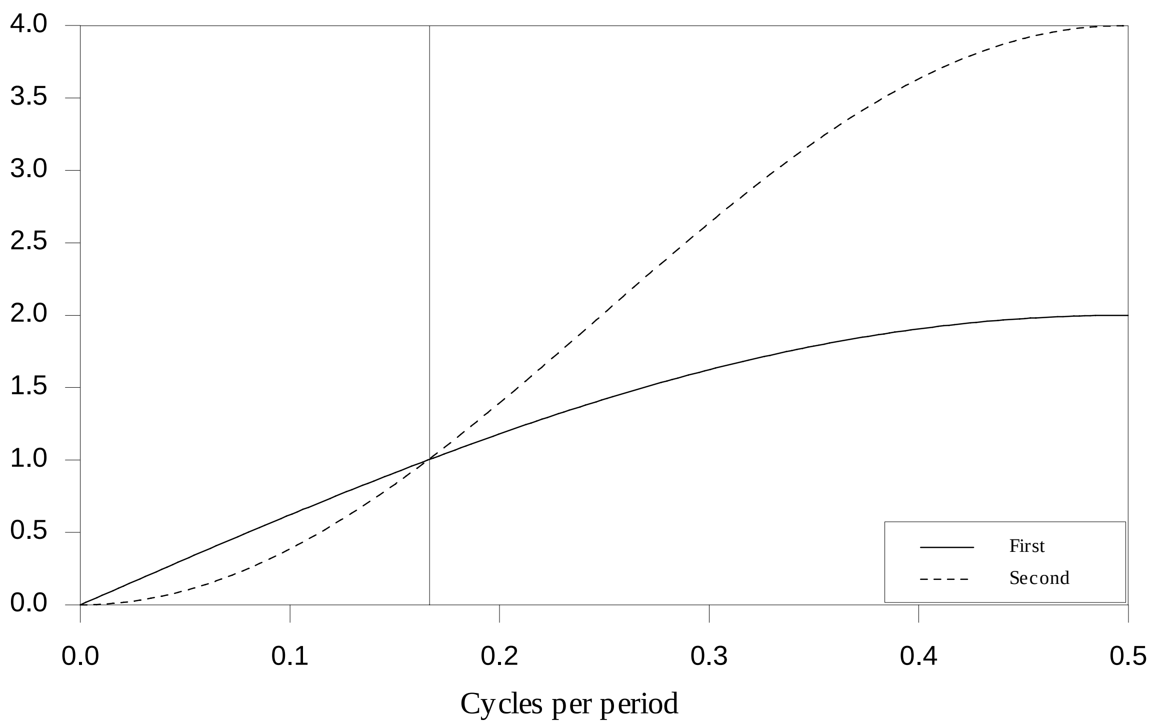
$$\delta_d(\omega) = \left(1-e^{-i\omega}\right)^d \tag{27}$$

and its gain is given by

$$\left|\delta_d(\omega)\right| = \left[2(1-\cos\omega)\right]^{d/2} \tag{28}$$

This gain is plotted in Figure 8 for d = 1 and 2.

Figure 8. Effects of first and second differencing (gain)



Note that the gain implies that the filter dampens variation in frequencies $0 \le \omega < \pi/3$, particularly frequencies close to zero, and that it amplifies variation for $\pi/3 < \omega \le \pi$. The frequency $\pi/3$ corresponds to 6 periods per cycle. With quarterly data, 6 quarters per cycle is often taken as the high frequency bound of the business cycle spectrum (see, eg, Baxter and King (1999), Stock and Watson (1999)). With monthly data, 6 months is well outside the normal range of business cycle frequencies.

Hence, the operator $\Delta^d(L)$ tends to amplify only frequencies that are generally considered to correspond to short-term noise and that are often censored in frequency domain analyses of business cycle fluctuations.

Clearly, appropriate application of the differencing filter to variables with unit roots (e.g., application of $\Delta^d(L)$ to the process y_t in (24)) produces stationary variables amenable to the application of frequency extraction techniques. Overdifferencing (the application of $\Delta^{d+j}(L)$ with $j > 0$ to y_t in (24)) in general produces stationary series, but may lead to undesirable results. It may excessively dampen low frequencies and amplify high frequencies if the latter are retained.

5.2 Testing for overdifferencing

Granger and Hatanaka (1964) and Granger (1966) identify a "typical spectral shape" for economic variables. They refer to the fact that the spectrum of an economic series tends to be decreasing as a function of the frequency from 0 to π. The typical spectral shape is consistent with the presence in many series of autoregressive features, with real roots greater than 1 and possibly unit or near-unit roots. For example, for the AR(1) series x_t defined by

$$(1-\rho L)x_t = \varepsilon_t \qquad (29)$$

with $0 < \rho < 1$ and ε_t white noise, the spectrum is given by

$$s_{xx}(\omega) = \sigma_\varepsilon^2 / (1 - 2\rho\cos\omega + \rho^2) \qquad (30)$$

In this case, $s'_{xx}(\omega) < 0$ for ω from 0 to π.

However, if the first difference operator Δx_t is applied to this series, the slope of the spectrum is reversed. The spectrum is then

$$s_{\Delta x \Delta x}(\omega) = 2\sigma_\varepsilon^2(1-\cos\omega)\big/(1-2\rho\cos\omega+\rho^2) \tag{31}$$

and $s'_{\Delta x \Delta x}(\omega) > 0$ for ω from 0 to π.

Note that a decreasing spectrum is not a feature of every possible ARIMA specification. For instance, the spectrum of x_t in (29) is increasing in ω if $-1 < \rho < 0$. However, many standard empirical specifications share the “typical shape.” Another example is an MA(1) process

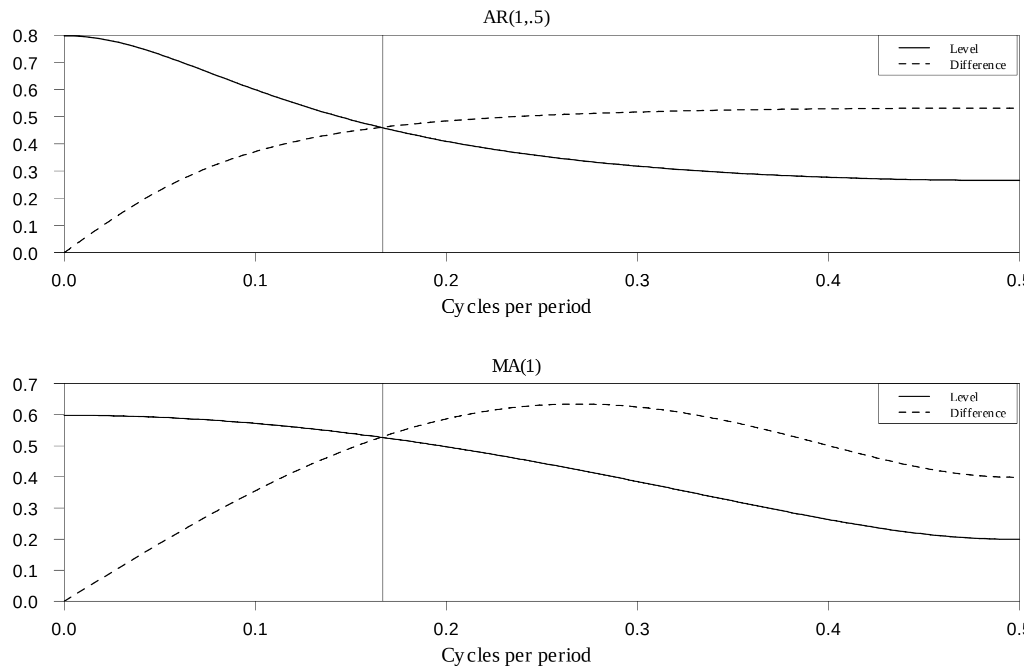
$$x_t = (1 + aL)\varepsilon_t \tag{32}$$

with $a > 0$, which also has a monotonically decreasing spectrum. This case is a bit different from the AR(1) case in that the spectrum of Δx_t is not monotonically increasing. Recall, however, that the effect of Δ switches from dampening to amplifying at $\omega = \pi/3$. For the first-differenced MA(1) series, the spectrum is increasing on average, in the sense that the average spectrum for $\omega < \pi/3$ is less than the average spectrum for $\omega > \pi/3$. More precisely,

$$\left(\pi/3\right)^{-1}\int_0^{\pi/3} s_{\Delta x \Delta x}(\omega)d\omega < \left(2\pi/3\right)^{-1}\int_{\pi/3}^{\pi} s_{\Delta x \Delta x}(\omega)d\omega \tag{33}$$

Figure 9 illustrates the spectra of the level and first difference of the AR(1) and MA(1) processes with $\rho = a = .5$.

Figure 9. Spectrum of AR(1) and MA(1) processes and their first differences



The foregoing patterns suggest a strategy for testing for overdifferencing of economic series. If the series x_t has a bounded spectrum that is higher on average for $0 \le \omega < \pi/3$ than for $\pi/3 < \omega \le \pi$, but the relative magnitudes are reversed for Δx_t, the shape of the resulting spectrum is dominated by the difference operator rather than the original series, a sign of overdifferencing.

This hypothesis may be tested empirically using spectral methods. Operating in the frequency domain, we take advantage of statistical sampling results available for spectral functions. Suppose the spectrum of x_t, t=1,…,T, is estimated using the periodogram

$$I_{xx}(\lambda_j) = \frac{1}{2\pi T}\left|\sum_{t=1}^{T} x_t e^{-i\lambda_j t}\right|^2, \qquad (34)$$

where $\lambda_j = 2\pi j/T$, j=1,…,T. Brillinger (1981, Section 5.2) shows that asymptotically (as $T \to \infty$)

$$E\, I_{xx}(\lambda_j) = s_{xx}(\lambda_j) \qquad (35)$$

$$Var\, I_{xx}(\lambda_j) = s_{xx}(\lambda_j)^2 \quad (36)$$

$$Cov\left\{I_{xx}(\lambda_j), I_{xx}(\lambda_k)\right\} = 0 \text{ for } j \neq k . \quad (37)$$

Moreover, Brillinger (1981, Theorem 5.6.3) implies that

$$\frac{1}{n}\sum_{j=j_0}^{j_0+n-1} I_{xx}(\lambda_j) \quad (38)$$

is asymptotically normal when $n \to \infty$ as $T \to \infty$.

Let n_1 be the integer such that $\lambda_{n_1} \leq \pi/3$ and $\lambda_{n_1+1} > \pi/3$. Then

$$S_1 = \frac{1}{n_1}\sum_{j=1}^{n_1} I_{xx}(\lambda_j) \text{ and } S_2 = \frac{1}{T-n_1}\sum_{j=n_1+1}^{T} I_{xx}(\lambda_j) \quad (39)$$

are asymptotically normal and independent with

$$E\, S_1 = \left(\pi/3\right)^{-1}\int_0^{\pi/3} s_{xx}(\omega)d\omega \text{ and } E\, S_2 = \left(2\pi/3\right)^{-1}\int_{\pi/3}^{\pi} s_{xx}(\omega)d\omega \quad (40)$$

By (35)-(37), we have also

$$Var\left(S_2 - S_1\right) = \frac{1}{n_1^2}\sum_{j=1}^{n_1} s_{xx}(\lambda_j)^2 + \frac{1}{\left(T-n_1\right)^2}\sum_{j=n_1+1}^{T} s_{xx}(\lambda_j)^2 \quad (41)$$

Hence, the hypothesis

$$H_0 : \left(\pi/3\right)^{-1}\int_0^{\pi/3} f_{xx}(\omega)d\omega < \left(2\pi/3\right)^{-1}\int_{\pi/3}^{\pi} f_{xx}(\omega)d\omega \quad (42)$$

is rejected with confidence level $1-\alpha$ if

$$z = \frac{S_2 - S_1}{\left[Var\left(S_2 - S_1\right)\right]^{1/2}} > z_\alpha \quad (43)$$

where *z* is asymptotically standard normal and $N(z > z_\alpha) = \alpha$ (*N* is the standard normal cdf).

5.3 Deterministic trends

So far we have assumed that there are no constant terms in the equations for the observable non-observable variables, but such constants are likely to appear in most empirical applications. For instance, consider a simple form of equation (1) in which $d = 1$ and $A(L) = 1$. An empirical model would include a constant α such that

$$\left(1-L\right)y_t = \alpha + \varepsilon_t \quad (44)$$

Summing from $t = 0$ we obtain

$$y_t = y_0 + \alpha t + \sum_{j=1}^{t} \varepsilon_j \tag{45}$$

which clearly contains a deterministic trend. Since the assumption in frequency domain methods is that the series is purely non-deterministic, this trend must be extracted before applying spectral methods.

If analysis of an empirical time series as in equation (44) suggests that a first difference filter should be applied, detrending is accomplished simply by subtracting the sample mean of $(1-L)\,y_t$. In this case, preliminary extraction of the mean is helpful in obtaining good empirical estimates of the frequency domain properties of time series, particularly at low frequencies. This is particularly clear when the spectrum is estimated using the periodogram, as described earlier. For a stationary variable with mean $\bar{x}$, the estimate of the periodogram at $\omega = 0$ is $(T/2\pi)\bar{x}^2$ so that this point contains information about the first moment of the series, rather than the second.

If it is important to work with the I(1) process y_t directly, detrending may be accomplished by removing a linear trend from the series, that is, by regressing y_t on a constant and *t*, and using the residuals as an estimate of $\sum_{j=1}^{t} \varepsilon_j$. This procedure is necessary for the application of the FD and ES filters to series that exhibit significant linear trends. The BK and HP filters are constructed in ways that incorporate the extraction of linear trends directly into the filter, so preliminary detrending is not necessary.

6. Empirical illustrations

6.1 Application of various filters to GDP and GDP deflator

A few practical issues arise when applying the time series filters to actual data. First, it is normally convenient to extract the mean from the raw series. The mean has a direct effect on the zero frequency component of the periodogram, and also on other low frequencies, particularly if smoothing windows are used.

Second, for series that contain a linear trend, or a component that looks in practice like a deterministic linear trend, that component should be extracted as well. This step is less important with the BK and HP filters, which incorporate at least two levels of differencing and implicitly extract linear terms. It is much more important when using the FD and ES filters, which do not automatically pre-difference the data.

Third, should the data be pre-differenced? This issue may be addressed by the computation of the *z* statistic defined in Section 5.2 or by application of unit root tests in the time domain, such as those of Dickey and Fuller (1979) or Phillips and Perron (1988).

The first row of Table 5 shows the results of applying the *z* statistic test to log levels of GDP and the GDP deflator. The negative results indicate that the spectrum is generally decreasing, as in the typical spectral shape, with significance at the 10% level for both variables. The second row applies the same test to first differences and shows very similar results, suggesting that this degree of differencing is appropriate.[8] Application of a second difference in the last row leads to large positive values, suggesting that the difference operator dominates the results and that this step may constitute over-differencing.[9]

Table 5. GDP and GDP deflator: z tests for over-differencing
Quarterly data, 1959Q1 to 2006Q2

	GDP		GDP deflator	
	z statistic	*p* value	*z* statistic	*p* value
Log level	-1.62	.053	-1.34	.090
First difference	-1.98	.024	-1.51	.065
Second difference	4.41	1.000	3.58	1.000

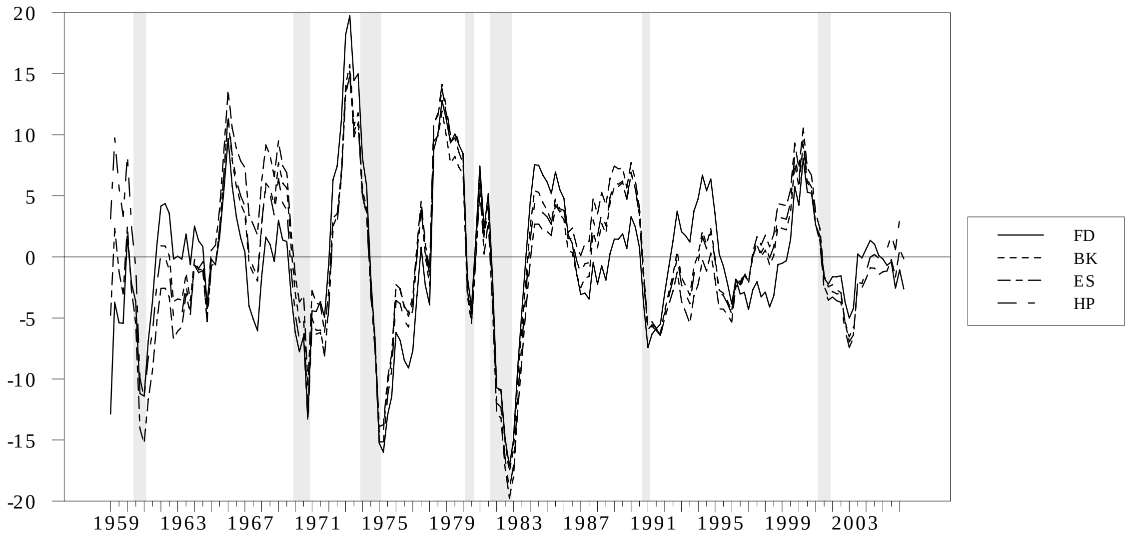
The foregoing results indicate that we should apply the filters to either log levels or first differences of the variables. Figure 10 presents the application of the filters to log levels of GDP. In the case of the FD and ES filters, a simple linear trend is extracted for reasons given earlier. The figure shows the inverse Fourier transform after the application of each filter in the frequency domain. Results are fairly consistent across filters, though the lack of differencing in

[8] Using alternative methods, also in the frequency-domain, Müller and Watson (2006) similarly conclude that a unit root model is often consistent with the observed low-frequency variability of twenty U.S. macroeconomic and financial time series.

[9] Standard unit root tests lead to similar conclusions. Using a *t*-test for log levels with constant and trend, a unit root cannot be rejected in either variable. Dickey-Fuller (1979) *p* values (0 lags) are .392 for GDP and 1.000 for the deflator, whereas and Phillips-Perron (1988) *p* values (4 lags) are .218 and .998, respectively. A test for first differences with a constant term rejects a second unit root with *p* values of .000 and .016 (Dickey-Fuller) and .000 and .052 (Phillips-Perron). Computation of *p* values is as in MacKinnon (1996).

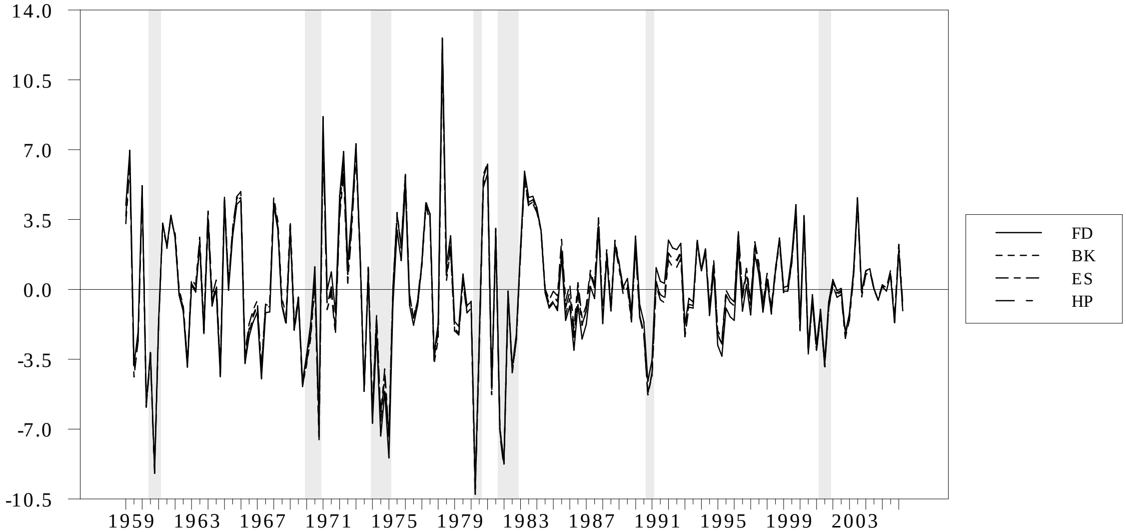
the FD and ES filters produces some visible discrepancies, particularly toward the ends of the sample.

Figure 10. High-pass filters applied to GDP in log levels



The same approach is applied to first (log) differences of GDP in Figure 11. No prior detrending is necessary in this case for the FD and ES filters. The homogeneity of the results suggests that all the filters produce very reasonable approximations to the theoretical FD filter if the appropriate level of differencing is first applied.

Figure 11. High-pass filters applied to GDP in first differences



With the GDP deflator, the results are qualitatively similar, though the stronger trends in this variable lead to greater deviations across filters. Because of the persistence of the series, results for FD and ES in Figure 12 are quite clearly outliers in the absence of any differencing. Once again, however, the application to first differences leads to fairly homogeneous results.

Figure 12. High-pass filters applied to GDP deflator in log levels

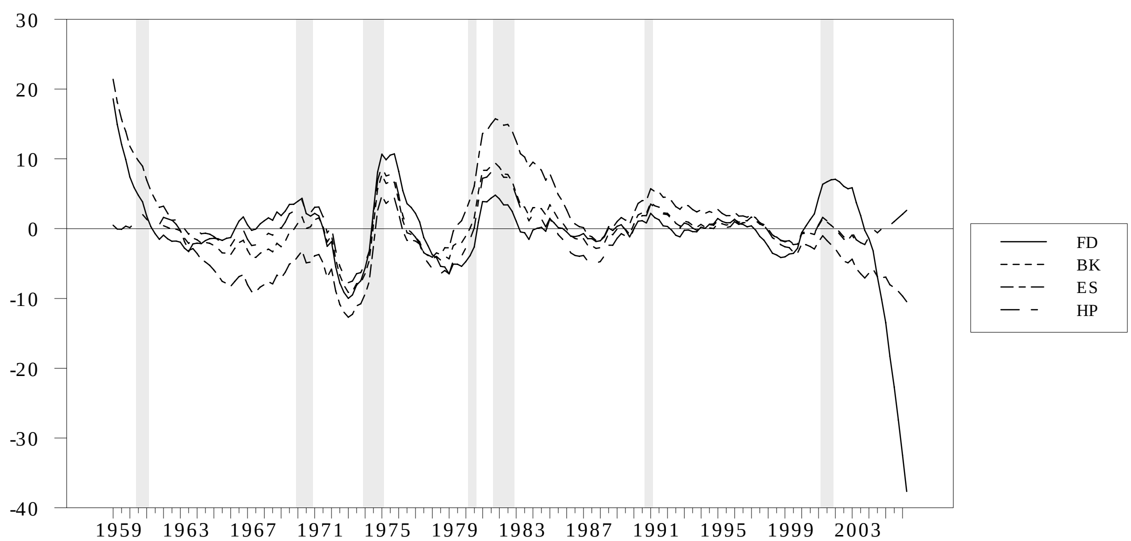
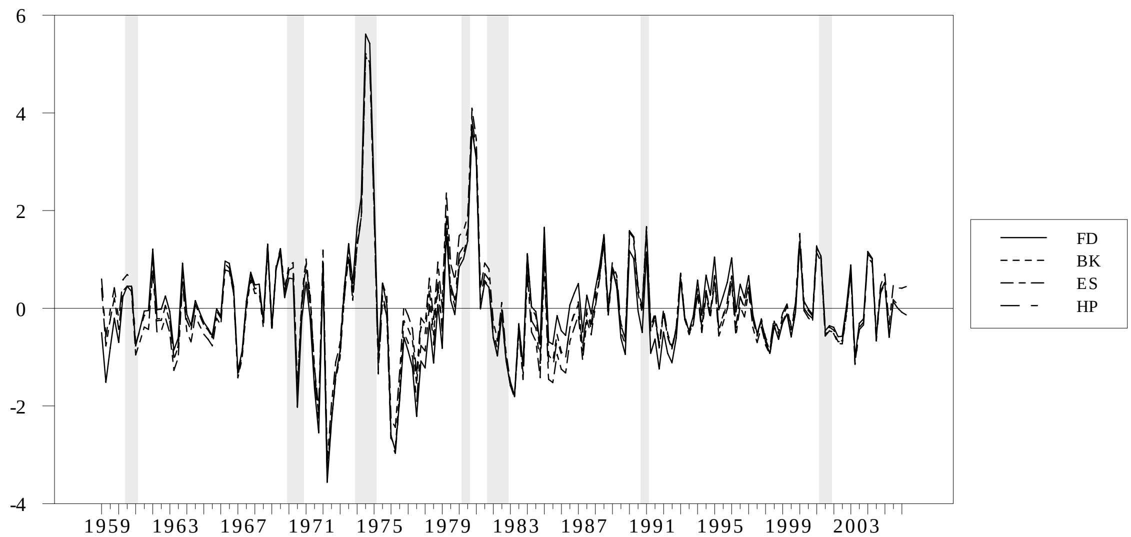


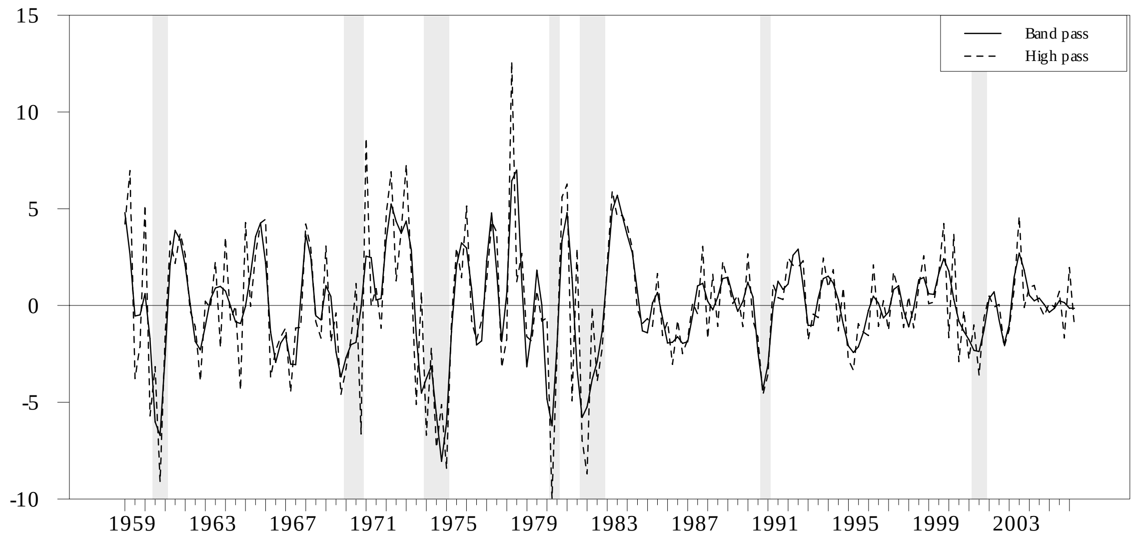
Figure 13. High-pass filters applied to GDP deflator in first differences



The series in the foregoing figures are substantially smoother than the raw data, but greater smoothness consistent with some notions of the business cycle may be achieved by censoring high-frequency movements in a band-pass filter. As noted, band-pass filters are straightforward extensions of the FD and BK high-pass filters. Figure 14 compares the FD versions of the band-pass and high-pass filters, both applied to the first difference of GDP.

Figure 14. Band-pass filter compared with high-pass filter:

FD filter (1 to 8 years and less than 8 years) applied to GDP in first differences



Quantitative measures of the features of the foregoing graphical illustrations may be obtained by calculating RMSEs of the resulting series, using the FD filter as a benchmark as in Section 3. Table 6 confirms that differences across filters are much larger for levels than for first differenced series, though given the visual results of Figures 10 and 12, the use of the FD filter as a benchmark must be taken with a grain of salt.

The results for first differences are more easily interpreted. We see that all the RMSEs are relatively small. The HP filter provides the best approximation for GDP but the BK filter does slightly better for the GDP deflator. These differential results are indicative of the interactions between the filters and the process to which they are applied.

Table 6. RMSE relative to FD filter: GDP and the GDP deflator

Quarterly data, 1959Q1 to 2006Q2

	GDP		GDP deflator	
	Level	First difference	Level	First difference
FD	0	0	0	0
BK*	2.61	.44	5.07	.31
ES	4.19	.54	9.63	.43
HP	3.30	.39	11.15	.33

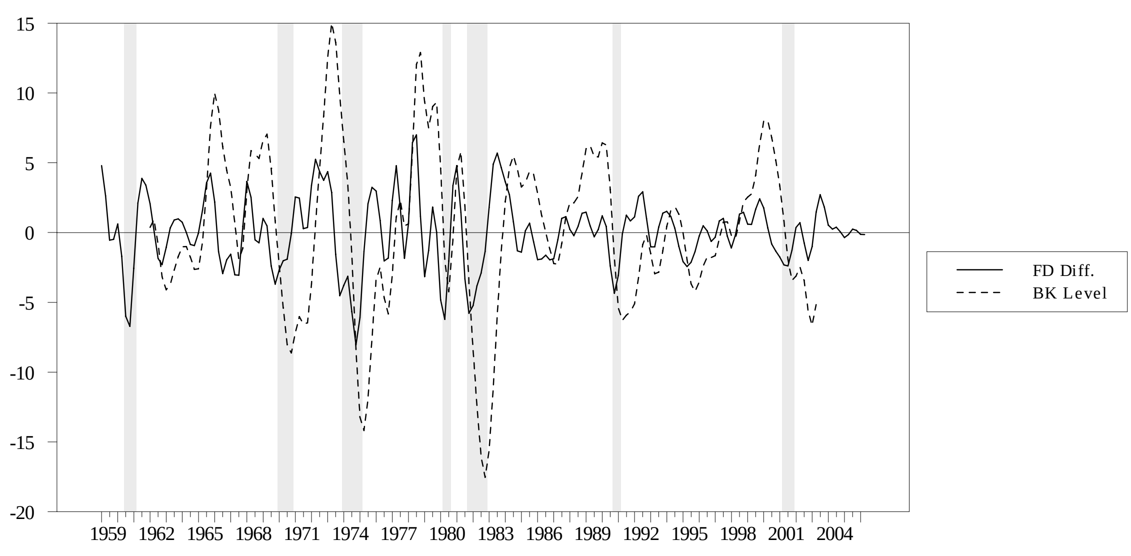
* Sample for BK is 1962Q1 to 2003Q2, since *K* observations must be dropped at either end.

6.2 Correspondence of filtered GDP to NBER recessions

One possible benchmark for the business cycle components derived from the various filters is how well they match the dating of recessions from the NBER. Figure 15 illustrates how different filters have different qualitative characteristics in relation to NBER recessions. Consider the application of band-pass FD and BK filters to GDP. The FD filter is applied to first differences to avoid distortions from the trend, whereas the BK filter is applied to log levels, since Baxter and King (1999) show that it can annihilate up to two degrees of integration.

The results are visually quite different. The BK filter, implicit differencing notwithstanding, produces a series that has the flavor of a level, as far as recessions are concerned. Note that this series tends to peak before the start of each shaded recession and fall sharply to a trough after the end of each recession, suggesting the presence of a substantial residual low-frequency component (Cf., BK results in Figure 4). In contrast, the FD filtered series tends to be negative during the course of each recession.[10] In addition, volatility seems overstated, which Table 3 suggests is a feature of BK with highly-persistent processes.

Figure 15. BK levels versus FD first differences for GDP



[10] Murray (2003) provides evidence that the BK filter allows the first difference of a stochastic trend to pass through with U.S. real GDP. In the terminology of the present paper, the BK filter applied to the log level is a suboptimal solution to the signal extraction problem.

Which representation is more accurate? One test is to include each filtered series in a probit equation in which the binary dependent variable is the recession indicator. Table 7 provides the pseudo-R^2 for each such experiment,[11] using the high-pass versions of the four filters, as well as band-pass versions of the FD and BK filters. The filters are applied to both levels and first differences, and the unfiltered series are included as well. In the unfiltered, FD and ES cases, a simple linear trend is extracted from the levels.

If the filters are applied to log levels of GDP, the HP filter produces the best relative fit, and the unfiltered series is a distant last. However, the results in general suggest that first-differencing is entirely appropriate, given the much more significant results obtained. With first differences, all the high-pass filters are inferior to the unfiltered series. However, the band-pass versions of both FD and BK are somewhat better than the unfiltered series. The BK band-pass filter produces the best results, though note that the sample period is shorter by six years because of the need to drop observations at both ends.

[11] Estrella (1998) shows that this pseudo-R^2, in addition to a measure of fit, is a monotonic transformation of the likelihood ratio test statistic for exclusion of the single explanatory variable. If the sample period is held constant, the pseudo-R^2 and the likelihood ratio test produce the same rankings of models.

Table 7. Probit equations for NBER recession indicator: pseudo-R^2 for filtered GDP
Quarterly data, 1959Q1 to 2006Q2

	Level	First difference
Unfiltered*	.023	.378
FD*	.134	.302
BK**	.148	.332
ES*	.148	.325
HP	.166	.328
FD band*	.130	.402
BK band**	.139	.429

* Detrended level.

** Sample for BK is 1962Q1 to 2003Q2, since *K* observations must be dropped at either end.

7. Conclusions

This paper shows that the features of individual time-series filters designed to extract business cycle fluctuations interact systematically with the characteristics of the processes to which they are applied. The exact nature of this interaction may not always be straightforward and its implications may differ dramatically from illustrations based on application to white noise.

In frequency extraction problems, the ideal solution involves the application of the FD filter to stationary data. If the data are in fact stationary, the BK, HP and ES filters also produce good results, though they are somewhat less accurate.

If the data process is integrated, all filters benefit from preliminary extraction of unit roots, even if the filters produce finite spectra without differencing. The implicit differencing incorporated in the BK and HP filters helps dampen low frequency components, but the effects of these components are not altogether eliminated and tend to distort results when applied to highly persistent processes. Preliminary application of the appropriate level of differencing to integrated processes, without over-differencing, leads to fairly similar results across filters. The FD filter emerges as somewhat preferable, however, particularly on theoretical grounds.

The FD and BK filters have the additional advantage that band-pass versions are easily computed, though the latter has the drawback that observations are lost at either end of the sample in either high- or band-pass versions.

In signal extraction problems, the ideal solution differs systematically from that of frequency extraction problems in that it may include large low-frequency components. In contrast to the frequency extraction problem, the cyclical component is always estimated with error, even asymptotically.

The ES and HP filters are the best performers in the cases in which they are theoretically optimal. In the signal extraction problem with either I(1)+I(0) or I(2)+I(0) data, the appropriate choice of these two filters is the best course of action. If it is unclear whether the trend component is I(1) or I(2), the HP is the safer choice. Errors with the HP filter in the I(1)+I(0) case are relatively moderate, whereas errors with ES in the I(2)+I(0) case are the largest of any filter, even those not expressly designed for signal extraction.

In general, important differences between the frequency and signal extraction problems and the diverse interactions between filters and processes suggest that filters must be carefully selected for any particular application. No single method can accommodate all circumstances well.

References

Baxter, Marianne and Robert G. King, 1999, Measuring business cycles: Approximate band-pass filters for economic time series, The Review of Economics and Statistics 81: 575-593.

Brillinger, David R., 1981, Time series: Data analysis and theory, San Francisco: Holden-Day.

Christiano, Lawrence J. and Terry J. Fitzgerald, 2003, The band pass filter, International Economic Review 44: 435-465.

Comin, Diego and Mark Gertler, 2006, Medium-term business cycles, American Economic Review 96: 523-551.

Dickey, David A. and Wayne A. Fuller, 1979, Distribution of the Estimators for Autoregressive Time Series With a Unit Root, Journal of the American Statistical Association 74: 427-431.

Ehlgen, Jürgen, 1998, Distortionary effects of the optimal Hodrick-Prescott filter, Economics Letters 61: 345-349.

Engle, Robert F., 1974, Band spectrum regression, International Economic Review 15: 1-11.

Estrella, Arturo, 1998, A new measure of fit for equations with dichotomous dependent variables, Journal of Economic and Business Statistics 16: 198-205.

Granger, C.W.J., 1966, The typical spectral shape of an economic variable, Econometrica 34: 150-161.

Granger, C.W.J. and M. Hatanaka, 1964, Spectral analysis of economic time series, Princeton: Princeton University Press.

Harvey, Andrew C. and Thomas M. Trimbur, 2003, General model-based filters for extracting cycles and trends in economic time series, The Review of Economics and Statistics 85: 244-255.

Harvey, Andrew C. and A. Jaeger, 1993, Detrending, stylized facts and the business cycle, Journal of Applied Econometrics 8: 231-247.

Hodrick, Robert J. and Edward C. Prescott, 1997, Postwar U.S. business cycles: An empirical intestigation, Journal of Money, Credit, and Banking 29: 1-16.

King, Robert G. and Sergio T. Rebelo, 1993, Low frequency filtering and real business cycles, Journal of Economic Dynamics and Control 17: 207-231.

Koopmans, Lambert H., 1998, The spectral analysis of time series, San Diego: Academic Press.

MacKinnon, James G., 1996, Numerical distribution functions for unit root and cointegration tests, Journal of Applied Econometrics 11: 601-618.

Müller, Ulrich K. and Mark W. Watson, 2006, Testing models of low-frequency variability, NBER Working Paper No. 12671.

Murray, Christian J., 2003, Cyclical properties of Baxter-King filtered time series, The Review of Economics and Statistics 85: 472-476.

Phillips, Peter C.B. and Pierre Perron, 1988, Testing for a Unit Root in Time Series Regression, Biometrika 75: 335-346.

Sargent, Thomas J., 1987, Macroeconomic theory, 2nd edition, Boston: Academic Press.

Stock, James H. and Mark W. Watson, 1999, Business cycle fluctuations in U.S. macroeconomic time series, in John B. Taylor and Michael Woodford, eds, Handbook of Macroeconomics, vol. 1A, 2-64.

Whittle, Peter, 1983, Prediction and regulation by linear least-square methods, 2nd edition, Minneapolis: University of Minnesota Press.