The Band Pass Filter*

Lawrence J. Christiano † and Terry J. Fitzgerald ‡

July, 2001

Abstract

The 'ideal' band pass filter can be used to isolate the component of a time series that lies within a specified band of frequencies. However, applying this filter requires a dataset of infinite length. In practice, an approximation is needed. We derive formulas for the least squares optimal approximation. These formulas exploit the properties of the time series representation of the data being filtered. One approximation that is particularly simple to implement is based on the assumption that the data are generated by a random walk. We show that this approximation works well for many US macroeconomic data series, even ones that appear not to be random walks.

To illustrate the use of this approximation, we document the substantial, statistically significant, shift in the money-inflation relationship before and after the 1960s. In the early period, the money-inflation relationship was strong and positive at all frequencies. In the later period, the relationship turned negative in frequencies 20 years and higher, although it remained positive in the very low frequencies.

Keywords: Band pass filter; projection; time series; frequency domain; unit roots.

We thank Eduard Pelz and Jeff Schwarz for their outstanding research assistance. The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Cleveland, or the Federal Reserve System. We are grateful for the comments of an anonymous referee.

† Northwestern University and NBER. Christiano is grateful for a National Science Foundation Grant to the National Bureau of Economic Research.

‡ St. Olaf College and Federal Reserve Bank of Cleveland.

1. Introduction

Economists have long been interested in the different frequency components of data. For example, business cycle theory is primarily concerned with understanding fluctuations in the range of 1.5 to 8 years while growth theory focuses on the longer run. In addition, some economic hypotheses are naturally formulated in the frequency domain, such as Milton Friedman's hypothesis that the long-run Phillips curve is positively sloped, while the short run Phillips curve is negatively sloped. Another example is the proposition that money growth and inflation are highly correlated in the long run, and less correlated in the short run.[1] Finally, certain frequency components of the data are important as inputs into macroeconomic stabilization policy. For instance, a policy maker who observes a recent change in output is interested in knowing whether that change reflects a shift in trend (i.e., the lower frequency component of the data) or whether it is just a transitory 'blip' (i.e., part of the higher frequency component).

The theory of the spectral analysis of time series provides a rigorous foundation for the notion that there are different frequency components of the data. According to the Spectral Representation Theorem, *any* time series within a broad class can be decomposed into different frequency components.[2] The theory also supplies a tool for extracting those components. That tool is the *ideal band pass filter*. It is a linear transformation of the data, which leaves intact the components of the data within a specified band of frequencies and eliminates all other components. The adjective, ideal, on this filter reflects an important practical limitation. Literally, application of the ideal band pass filter requires an infinite amount of data. Some sort of approximation is required.

In this paper, we characterize and study optimal linear approximations and compare these with alternative approaches developed in the literature. To explain what we mean by an optimal linear approximation, let y_t denote the data generated by applying the ideal, though infeasible, band pass filter to the raw data, x_t. We approximate y_t by $\hat{y}_t$, a linear function, or filter, of the observed sample x_t's. We select the filter weights to make $\hat{y}_t$ as close as possible to the object of interest, y_t, in the sense of minimizing the mean square error criterion:

$$E\left[(y_t - \hat{y}_t)^2 |x\right], \ x \equiv [x_1, ..., x_T]. \tag{1.1}$$

Thus, $\hat{y}_t$ is the linear projection of y_t onto every element in the data set, x, and there is a different projection problem for each date t. Since the first order condition associated with the minimization problem in (1.1) is linear in the unknown filter weights, they can be obtained by straightforward matrix manipulations.

The optimal approximation to the band pass filter requires knowing the true time series representation of x_t. In practice, this is not known and must be estimated. It turns out, however, that for standard macroeconomic time series, a more straightforward approach that does not

[1] For early work that explored this hypothesis using tools closely related to the ones explored in this paper, see Engle (1974).

[2] For a formal analysis of the Spectral Decomposition Theorem, see Cramer and Leadbetter (1967) or Lippi (2001). A simplified exposition appears in Christiano and Fitzgerald (1998, Appendix).

involve first estimating a time series model works well. That approach uses the approximation that is optimal under the (most likely, false) assumption that the data are generated by a pure random walk.[3] The procedure is nearly optimal for the type of time series representations that fit US data on interest rates, unemployment, inflation, and output. The filter, which we call the Random Walk filter, is easy to implement. Suppose one wishes to isolate the component of x_t with period of oscillation between p_l and p_u, where $2 \leq p_l < p_u < \infty$.[4] The Random Walk filter approximation of this component, $\hat{y}_t$, is computed as follows:

$$\begin{aligned}\hat{y}_t &= B_0 x_t + B_1 x_{t+1} + ... + B_{T-1-t} x_{T-1} + \tilde{B}_{T-t} x_T \\ &\quad + B_1 x_{t-1} + ... + B_{t-2} x_2 + \tilde{B}_{t-1} x_1,\end{aligned} \tag{1.2}$$

for $t = 3, 4, ..., T-2$. In (1.2),

$$\begin{aligned}B_j &= \frac{\sin(jb) - \sin(ja)}{\pi j},\ j \geq 1 \\ B_0 &= \frac{b-a}{\pi},\ a = \frac{2\pi}{p_u},\ b = \frac{2\pi}{p_l}.\end{aligned} \tag{1.3}$$

and $\tilde{B}_{T-t}$, $\tilde{B}_{t-1}$ are simple linear functions of the B_j's.[5] The formulas for $\hat{y}_t$ when $t = 2$ and $T-1$ are straightforward adaptations on the above expressions. The formulas for $\hat{y}_1$ and $\hat{y}_T$ are also of interest. For example,

$$\hat{y}_T = \left(\frac{1}{2} B_0\right) x_T + B_1 x_{T-1} + ... + B_{T-2} x_2 + \tilde{B}_{T-1} x_1, \tag{1.4}$$

where $\tilde{B}_{T-1}$ is constructed using the analog of the formulas underlying the $\tilde{B}_j$'s in (1.2).[6] The

[3] Our formulas assume there is no drift in the random walk. If there is a drift in the raw data, we assume it has been removed prior to analysis. For more details, see footnote 11 below.

[4] If the data are quarterly and $p_l = 6$, $p_u = 32$, then y_t is the component of x_t with periodicities between 1.5 and 8 years.

[5] In particular, $\tilde{B}_{T-t}$ is the sum of the B_j's over $j = T-t, T-t+1, ...$ and $\tilde{B}_{t-1}$ is the sum of the B_j's over $j = t-1, t, ...$. Exploiting the fact that $B_0 + 2\sum_{i=1}^{\infty} B_i = 0$,

$$\tilde{B}_{T-t} = -\frac{1}{2} B_0 - \sum_{j=1}^{T-t-1} B_j,\ \text{for } t = 3, ..., T-2.$$

Also, $\tilde{B}_{t-1}$ solves

$$0 = B_0 + B_1 + ... + B_{T-1-t} + \tilde{B}_{T-t} + B_1 + ... + B_{t-2} + \tilde{B}_{t-1}.$$

[6] Here $\tilde{B}_{T-1} = -\frac{1}{2} B_0 - \sum_{j=1}^{T-2} B_j$.

expression for $\hat{y}_T$ is useful in circumstances when an estimate of y_T is required in real time, in which case only a one-sided filter is feasible.

Note from (1.2) that the weights in the Random Walk filter vary with time. Also, except for t in the middle of the data set, the weights are not symmetric in terms of future and past x_t's. It is easy to adjust the Random Walk filter weights to impose stationarity and symmetry, if these features are deemed to be absolutely necessary. Simply construct (1.2) so that $\hat{y}_t$ is a function of a fixed number, p, of leads and lags of x_t and compute the weights on the highest lead and lag using simple functions of the B_j's.[7] This is the solution to our projection problem when x_t is a random walk, and $\hat{y}_t$ is restricted to be a linear function of $\{x_t, x_{t\pm1}, ..., x_{t\pm p}\}$ only. With this approach, it is not possible to estimate y_t for the first and last p observations in the data set. In practice, this means restricting p to be relatively small, to say three years of data. This filter induces stationarity in time series which have up to two unit roots, or which have a quadratic trend.

It is important to emphasize a caveat regarding the Random Walk filter, (1.2)-(1.3). That filter does not closely approximate the optimal filter in all conceivable circumstances. For cases in which the appropriateness of the Random Walk filter is questionable, we conjecture that the following strategy is a good one. First, estimate the time series representation of the data to be filtered. Second, use the formulas derived below to compute the optimal filter based on the assumption that the estimated time series representation is the true one.[8] The formulas derived below apply for a large class of time series models. It is straightforward to adapt the formulas so that they apply to an even larger class.

To illustrate the value of the filtering methodology studied here, we present an empirical application to money growth and inflation. We document a substantial, statistically significant, shift in the money-inflation relationship before and after 1960. In the early period, the money-inflation relationship was strong and positive at all frequencies. In the later period, the relationship turned negative in frequencies 20 years and higher, although it remained positive in the very low frequencies. To our knowledge, this intriguing change in the money-growth relationship has not been documented before.[9] The example complements others in the literature

[7] The weights on $x_t, x_{t\pm1}, ..., x_{t\pm(p-1)}$ are $B_0, ..., B_{p-1}$, respectively. The weight on x_{t-p} and x_{t+p}, $\tilde{B}_p$, is obtained using

$$\tilde{B}_p = -\frac{1}{2}\left[B_0 + 2\sum_{j=1}^{p-1} B_j\right].$$

It is easy to verify that in this case there is no need to drift-adjust the raw data because the output of the formula is invariant to drift. The reason is that the optimal symmetric filter when the raw data are a random walk has two unit roots. The first makes x_t stationary and the second eliminates any drift. In contrast, the output of the potentially asymmetric filter just discussed in the text is *not* invarient to drift. When $p \neq f$, that filter has just one unit root.

[8] Software for computing the filters in GAUSS, MATLAB, STATA, EVIEWS and RATS can be obtained from the authors' homepages. The default option in this software takes as input a raw time series, removes its drift, and then filters it using our recommended Random Walk filter. Alternatively, one can input a time series belonging to the class considered below and in Christiano and Fitzgerald (1999), and the software returns the relevant optimal filter approximation.

[9] Our findings are consistent with the results in Backus and Kehoe (1992, Table 8). However, they do not discuss the money growth - inflation relationship in the text, and the results in their table only pertain to

by illustrating the value of the band pass filter in isolating economically interesting features of the data.[10] In addition, the bootrap methodology we apply is used to show that statistics based on the different frequency bands - even bands as low as 8-20 years - can be estimated with precision.

The outline of the paper is as follows. Section 2 discusses the ideal band pass filter and our formula for approximating it. Section 3 studies the importance of several key features of our optimal filter approximations. For example, we quantify the value of allowing filter weights to be asymmetric and time-varying. Section 4 presents our empirical application. Section 5 relates our analysis to the relevant literature. We stress in particular, the important papers by Baxter and King (1999) (BK) and Hodrick and Prescott (1997) (HP). The HP filter is sometimes used in a policymaking framework to develop a real-time estimate of the trend component of aggregate output. Among other things, this section compares the real-time performance of our filters with the HP filter. Section 6 concludes.

2. Optimal Approximation to the Band Pass Filter

We begin this section by precisely defining the object that we seek: the component of x_t that lies in a particular frequency range. We then present formulas for computing the optimal approximation. Several examples are presented which highlight features of the optimal approximation.

Our approximation formulas can accommodate two types of x_t processes. In one, x_t has a zero mean, and is covariance stationary. If the raw data have a non-zero mean, we assume it has been removed prior to analysis. If the raw data are covariance stationary about a trend, then we assume that trend has been removed. We also consider the unit root case, in which $x_t - x_{t-1}$ is a zero mean, covariance stationary process. If in the raw data this mean is non-zero, then we suppose that it has been removed prior to analysis.[11] As we will see, the latter is actually only necessary when we consider asymmetric filters.

business cycle frequencies.

[10] For other interesting analyses see Baxter (1994), Hornstein (1998), King and Watson (1994) and Stock and Watson (1999).

[11] Removing this mean corresponds to 'drift adjusting' the x_t process. We elaborate on this briefly here. Suppose the raw data are denoted w_t, and they have the representation, $w_t = \mu + w_{t-1} + u_t$, where u_t is a zero mean, covariance stationary process. Then, w_t can equivalently be expressed as $w_t = (t-j)\mu + x_t$, where $x_t = x_{t-1} + u_t$ for all t and j is a fixed integer, which we normalize to unity for concreteness. The variable, x_t, is the 'drift-adjusted' version of w_t, and can be recovered from observations on w_t as follows: $x_1 = w_1$, $x_2 = w_2 - \mu$, $x_3 = w_3 - 2\mu$, ... In practice, μ must be estimated, with $\hat{\mu} = (w_T - w_1)/(T-1)$. Though we set $j = 1$, it is readily confirmed that the output of our filter is invariant to the value of j chosen. In sum, in the unit root case, we assume x_t is the result of removing a trend line from the raw data, where the slope of the line is the drift in the raw data and the level is arbitrary.

2.1. The Ideal Band Pass Filter

Consider the following orthogonal decomposition of the stochastic process, x_t:

$$x_t = y_t + \tilde{x}_t. \tag{2.1}$$

The process, y_t, has power only in frequencies belonging to the interval $\{(a,b) \cup (-b,-a)\} \in (-\pi,\pi)$. The process, $\tilde{x}_t$, has power only in the complement of this interval in $(-\pi,\pi)$.[12] Here, $0 < a \leq b \leq \pi$. It is well known (see, for example, Sargent (1987, page 259)), that,

$$y_t = B(L)x_t, \tag{2.2}$$

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

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

where the B_j's are given by (1.3). With this specification of the B_j's, we have

$$\begin{aligned} B(e^{-iw}) &= 1, \text{ for } w \in (a,b) \cup (-b,-a) \\ &= 0, \text{ otherwise.} \end{aligned} \tag{2.3}$$

Our assumption, $a > 0$, implies, together with (2.3), that $B(1) = 0$. Note from (2.2) that to compute y_t using $B(L)$ requires an infinite number of observations on x_t. Moreover, it is not clear that simply truncating the B_j's will produce good results.

This can be seen in two ways. First, consider Figure 1a, which displays B_j for $j = 0, ..., 200$, when $a = 2\pi/96$ and $b = 2\pi/18$. These frequencies, in monthly data, correspond to the business cycle, i.e., periods of fluctuation between 1.5 and 8 years. Note how the B_j's die out only for high values of j. Even after $j = 120$, i.e., 10 years, the B_j's remain noticeably different from zero. Second, Figures 1b - 1d show that truncation has a substantial impact on $B(e^{-iw})$. They display the Fourier transform of filter coefficients obtained by truncating the B_j's for $j > p$ and $j < -p$ for $p = 12$, 24, 36 (i.e., 1 to 3 years). These differ noticeably from $B(e^{-iw})$.

[12]The notion that $\tilde{x}_t$ and y_t are orthogonal is problematic in the case where x_t has one (or more) unit roots. In this case, we interpret the orthogonality property as applying to an arbitrarily small perturbation of the x_t process in which the unit root is replaced by a root that is inside the unit circle.

2.2. A Projection Problem

Suppose we have a finite set of observations, $x = [x_1, ..., x_T]$ and that we know the population second moment properties of $\{x_t\}$. Our estimate of $y = [y_1, ..., y_T]$ is $\hat{y}$, the projection of y onto the available data:

$$\hat{y} = P[y|x].$$

This corresponds to the following set of projection problems:

$$\hat{y}_t = P[y_t|x], \ t = 1, ..., T. \tag{2.4}$$

For each t, the solution to the projection problem is a linear function of the available data:

$$\hat{y}_t = \sum_{j=-f}^{p} \hat{B}_j^{p,f} x_{t-j}, \tag{2.5}$$

where $f = T - t$ and $p = t - 1$, and the $\hat{B}_j^{p,f}$'s solve

$$\min_{\hat{B}_j^{p,f}, j=-f,...,p} E\left[(y_t - \hat{y}_t)^2 |x\right]. \tag{2.6}$$

We can express this problem in the frequency domain by exploiting the standard frequency domain representation for a variance:[13]

$$\min_{\hat{B}_j^{p,f}, j=-f,...,p} \int_{-\pi}^{\pi} |B(e^{-i\omega}) - \hat{B}^{p,f}(e^{-i\omega})|^2 f_x(\omega) d\omega. \tag{2.7}$$

Here, $f_x(\omega)$ is the spectral density of x_t and

$$\hat{B}^{p,f}(L) = \sum_{j=-f}^{p} \hat{B}_j^{p,f} L^j, \ L^h x_t \equiv x_{t-h}.$$

We wish to stress three aspects of the $\hat{B}_j^{p,f}$'s which solve (2.7). First, the presence of f_x in (2.7) indicates that the solution to the minimization problem depends on the properties of the time series representation of x_t. This stands in contrast with the weights in the ideal band pass filter, which do not depend on the time series properties of the data.

[13] For a closely related discussion, see Sims (1972).

Second, since the minimization problem depends on t, this strategy for estimating y_1, $y_2, ..., y_T$ uses T different filters, one for each date. In particular, the filters are not stationary with respect to t, and for each t they weight future and past observations on x_t asymmetrically.[14] In practice, one could impose stationarity and symmetry on (2.7). Stationarity may have econometric advantages. Symmetry ensures that there is no phase shift between $\hat{y}_t$ and y_t. Still, stationarity and symmetry come at cost. In general these properties represent binding restrictions on (2.7), so that imposing them on the filter approximation results in a less precise estimate of y_t. One of our objectives is to quantify the severity of this trade-off in settings that are of practical interest.

Third, in practice one does not know the true spectral density for x_t. Presumably, the solution to (2.7) would be different if one explicitly took into account uncertainty in f_x. Doing so is beyond the scope of this paper. In any case our results suggest that, for typical macroeconomic data series, reasonable approximations to the solution can be obtained without knowing the details of the time series representation of x_t.

2.3. Solution to the Projection Problem

The quadratic nature of (2.7), together with linearity in (2.5) guarantee that the solution to (2.7) has a simple representation. In particular, the $\hat{B}_j^{p,f}$'s solve a system of linear equations. This section derives this system of equations for a particular class of spectral densities, f_x.

We consider spectral densities corresponding to x_t processes which have the following time series representation:

$$x_t = x_{t-1} + \theta(L)\varepsilon_t, \; E\varepsilon_t^2 = 1, \tag{2.8}$$

where $\theta(L)$ is a q^{th}-ordered polynomial in the lag operator, L. The corresponding spectral density is:

$$f_x(\omega) = \frac{g(\omega)}{(1-e^{-i\omega})(1-e^{i\omega})},$$

where

$$\begin{aligned} g(\omega) &= \theta(e^{-i\omega})\theta(e^{i\omega}) \\ &= c_0 + c_1\left(e^{-i\omega} + e^{i\omega}\right) + ... + c_q\left(e^{-i\omega q} + e^{i\omega q}\right). \end{aligned}$$

The class of time series representations in (2.8) encompasses the case where x_t is stationary (i.e., $\theta(1) = 0$), possibly because a trend has already been removed from the raw data, and where x_t is difference stationary (i.e., set $\theta(1) \neq 0$). In the stationary case, $x_t = [\theta(L)/(1-L)]\,\varepsilon_t = \tilde{\theta}(L)\varepsilon_t$, where $\tilde{\theta}(L)$ is a $(q-1)$ - ordered polynomial in L.

We treat the difference stationary case below. A straightforward adaption of our argument can be used to address the stationary case. See Christiano and Fitzgerald (1999) (CF) for

[14] If T is odd, then there is one filter that is symmetric, namely the one associated with date $t = (T+1)/2$.

details. The solution to (2.7) when x_t is a random walk (i.e., $q = 0$) was presented in the introduction.[15] We now consider $q > 0$.

A necessary condition for an optimum is $\hat{B}^{f,p}(1) = 0$, for otherwise the criterion in (2.7) would be infinite. This implies that $b(z)$ is a finite ordered polynomial, where

$$b(z) = \frac{\hat{B}^{f,p}(z)}{1-z}$$

and

$$b(z) = b_{p-1}z^{p-1} + b_{p-2}z^{p-2} + ... + b_0 + ... + b_{-f+1}z^{-f+1} + b_{-f}z^{-f}.$$

The link between the b_j's and the $\hat{B}_j^{f,p}$'s is expressed in matrix form as follows:

$$Q\hat{B}^{f,p} = b, \tag{2.9}$$

where Q is a $(p+f) \times (p+f+1)$ matrix and b and $\hat{B}^{f,p}$ are $(p+f) \times 1$ and $(p+f+1) \times 1$ column vectors:

$$Q = \begin{bmatrix} -1 & 0 & 0 & \cdots & 0 & 0 \\ -1 & -1 & 0 & \cdots & 0 & 0 \\ -1 & -1 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ -1 & -1 & -1 & \cdots & -1 & 0 \end{bmatrix}, \; b = \begin{bmatrix} b_{p-1} \\ b_{p-2} \\ b_{p-3} \\ \vdots \\ b_{-f} \end{bmatrix}, \; \hat{B}^{f,p} = \begin{bmatrix} \hat{B}_p^{f,p} \\ \hat{B}_{p-1}^{f,p} \\ \hat{B}_{p-2}^{f,p} \\ \vdots \\ \hat{B}_{-f}^{f,p} \end{bmatrix}. \tag{2.10}$$

We suppose that $p+f > 0$, $p \geq 0$, $f \geq 0$ and q is small, in the sense that $p + f \geq 2q$.[16]

[15]The random walk case can be established with a simple time-domain argument. The problem is that not all the observations on x_t are available to evaluate y_t in (2.2). The missing data are the x_t's after the end of the data set and before the beginning. The time domain version of the least squares approach taken in this paper replaces the missing observations with the least squares optimal guess based on the observed data. In the Random Walk case, the best estimate of each presample observation is just the first data point and the best estimate of every post sample observation is the last data point. The weights in Random Walk approximate filter are computed by pursuing the implications of this observation. This time domain strategy for solving our problem corresponds to the one implemented by Stock and Watson (1999) in a business cycle context, and by Geweke (1978) and Wallis (1983) in a seasonal adjustment context.

[16]It is straightforward to adapt the argument to accommodate larger q. It is also straightforward to accommodate the case in which $x_t - x_{t-1}$ is a mixed autoregressive moving average process. The specification

We rewrite (2.7) as a optimization problem in the b_j's:

$$\min_{b_j, j=p-1,...,-f} \int_{-\pi}^{\pi} |\tilde{B}(e^{-i\omega}) - b(e^{-i\omega})|^2 g(\omega) d\omega, \tag{2.11}$$

where

$$\tilde{B}(z) = \frac{B(z)}{1-z}.$$

The first order conditions for this problem are (see CF for details):

$$\int_{-\pi}^{\pi} \tilde{B}(e^{-i\omega}) g(\omega) e^{i\omega j} d\omega = \int_{-\pi}^{\pi} b(e^{-i\omega}) g(\omega) e^{i\omega j} d\omega, \tag{2.12}$$

$j = p-1, ..., -f.$

Expression (2.12) is a system of $p+f$ linear equations in the $\hat{B}_j^{f,p}$'s. The $p+f+1^{st}$ equation is obtained from $\hat{B}^{f,p}(1) = 0$. To solve for the $p+f+1$ unknown $\hat{B}_j^{f,p}$'s, it is convenient to express (2.12) in matrix form:[17]

$$\int_{-\pi}^{\pi} \tilde{B}(e^{-i\omega}) g(\omega) e^{i\omega j} d\omega = 2\pi F_j Q \hat{B}^{f,p}. \tag{2.13}$$

Here, F_j is a $1 \times (p+f)$ - dimensional row vector, $j = p-1, ..., -f$. For $p-q-1 \geq j \geq q-f$:

$$F_j = \left[\underbrace{0, ..., 0}_{1\times(p-q-1-j)}, c, \underbrace{0, ..., 0}_{1\times(j-q+f)} \right], \tag{2.14}$$

where $c = [c_q, c_{q-1}, ..., c_0, ..., c_{q-1}, c_q]$. When $j = p-q-1$, the first set of zeros is absent in F_j, and when $j = q-f$, the second set of zeros is absent. When $j > p-q-1$, the first set of zeros is absent in F_j, and the first $j-(p-q-1)$ elements of c are absent too. When $j < q-f$, the last set of zeros is absent in F_j, and the last $q-f-j$ elements of c are absent too.

---

we work with was chosen because it seems adequate for standard macroeconomic time series.

[17]Here, and elsewhere, we make use of the following well-known result:

$$\begin{aligned} \int_{-\pi}^{\pi} e^{i\omega h} d\omega &= 0, \text{ for } h = \pm 1, \pm 2, ... \\ &= 2\pi, \text{ for } h = 0. \end{aligned}$$

Combining the $p+f$ equations in (2.13) with $\hat{B}^{f,p}(1)=0$, we obtain:

$$d = A\hat{B}^{f,p}, \tag{2.15}$$

where

$$d = \begin{bmatrix} \int_{-\pi}^{\pi} \tilde{B}(e^{-i\omega})g(\omega)e^{i\omega(p-1)}d\omega \\ \vdots \\ \int_{-\pi}^{\pi} \tilde{B}(e^{-i\omega})g(\omega)e^{i\omega(-f+1)}d\omega \\ \int_{-\pi}^{\pi} \tilde{B}(e^{-i\omega})g(\omega)e^{i\omega(-f)}d\omega \\ 0 \end{bmatrix}, \quad A = 2\pi \begin{bmatrix} F_{p-1}Q \\ \vdots \\ F_{-f+1}Q \\ F_{-f}Q \\ 1 \quad \cdots \quad 1 \end{bmatrix}, \tag{2.16}$$

and the square matrix, A, has dimension $p+f+1$. It is straightforward to compute the objects in A. We now discuss the computation of the integrals that make up d. Note

$$\begin{aligned} \int_{-\pi}^{\pi} \tilde{B}(e^{-i\omega})g(\omega)e^{-i\omega j}d\omega &= \int_{0}^{\pi} \left[\tilde{B}(e^{-i\omega})e^{-i\omega j} + \tilde{B}(e^{i\omega})e^{i\omega j}\right] g(\omega)d\omega \\ &= \int_{a}^{b} \left[\frac{e^{-i\omega j}}{1-e^{-i\omega}} + \frac{e^{i\omega j}}{1-e^{i\omega}}\right] g(\omega)d\omega \\ &= R(j), \end{aligned} \tag{2.17}$$

say, for $j = 1-p,\ 2-p,\ ....,\ f$. We compute $R(j)$ using a particular recursive procedure based on three relations. The first is an expression for $R(0)$:[18]

$$R(0) = \int_{-\pi}^{\pi} \tilde{B}(e^{-i\omega})g(\omega)d\omega = \int_{-\pi}^{\pi} B(e^{-i\omega})g(\omega)d\omega. \tag{2.18}$$

The second relation is:

$$R(j) - R(j+1) = \int_{a}^{b} \left[\left(\frac{e^{-i\omega j}}{1-e^{-i\omega}} + \frac{e^{i\omega j}}{1-e^{i\omega}}\right) - \left(\frac{e^{-i\omega(j+1)}}{1-e^{-i\omega}} + \frac{e^{i\omega(j+1)}}{1-e^{i\omega}}\right)\right] g(\omega)d\omega$$

[18] The second equality uses

$$\frac{1}{1-e^{-i\omega}} + \frac{1}{1-e^{i\omega}} = \frac{1-e^{i\omega}+1-e^{-i\omega}}{(1-e^{-i\omega})(1-e^{i\omega})} = 1$$

$$
\begin{aligned}
&= \int_a^b \left[e^{-i\omega j} + e^{i\omega j}\right] g(\omega) d\omega \\
&= \int_{-\pi}^{\pi} B(e^{-i\omega}) g(\omega) e^{-i\omega j} d\omega. \qquad (2.19)
\end{aligned}
$$

The third relation is:

$$\int_{-\pi}^{\pi} B(e^{-i\omega}) g(\omega) e^{i\omega j} d\omega = 2\pi \left(B_j c_0 + \sum_{i=1}^{q} \left[B_{|j|+i} + B_{||j|-i|}\right] c_i \right). \qquad (2.20)$$

Expressions (2.18)-(2.20) can be used in an obvious way to compute $R(j)$ for $j = 1 - p, ..., f$. The vector d is then obtained from (2.16) and (2.17).

Finally, we solve for the $\hat{B}_j^{f,p}$'s using

$$\hat{B}^{f,p} = A^{-1} d. \qquad (2.21)$$

It is easy to verify that when $q = 0$, this solution coincides with the solution reported in the introduction. In addition, when $p = f$, $\hat{B}^{p,p}(L)$ is a symmetric polynomial.

2.4. Examples

The key to understanding the solution to (2.7) is to note that for finite p and f, it is not possible to construct $\hat{B}^{p,f}(e^{-i\omega})$ so that $\hat{B}^{p,f}(e^{-i\omega}) = B(e^{-i\omega})$ for all ω. The two functions can be made close over some subintervals, but only at the cost of sacrificing accuracy over other subintervals. The existence of this trade-off implies that some weighting scheme is needed to determine which intervals to emphasize in constructing $\hat{B}^{p,f}(e^{-i\omega})$. The weighting scheme implicit in our optimization criterion is the spectral density of x_t. The reason for this is that the optimization problem seeks to make y_t and $\hat{y}_t$ as close as possible, and this translates into making the *product* of B and f_x similar to the product of $\hat{B}^{p,f}$ and f_x. Thus, the optimization criterion picks the $\hat{B}_j^{p,f}$'s so that $\hat{B}^{p,f}(e^{-i\omega})$ resembles $B(e^{-i\omega})$ closely for values of ω where $f_x(\omega)$ is large and places less emphasis on regions where f_x is relatively small. We illustrate this principle using three examples with $p = f = 12$, $p_u = 24$, $p_l = 4$. They tilt the graph of $f_x(\omega)$, $\omega \in (0, \pi)$ in different ways.

In our first example, the 'IID case', f_x is constant for all ω. This is a natural benchmark, because in this case, the $\hat{B}_j^{p,f}$'s are just the ideal filter weigths, B_j, truncated. The objects, $\hat{B}^{p,f}(e^{-i\omega})$ and $B(e^{-i\omega})$ are displayed in Figure 2a. The next two examples allocate increasing power towards the low frequencies. Accordingly, for our second example, the 'Near IID case', we perturb f_x from the first example by placing a spike at frequency zero.[19] In part, this is of

[19] In this case, $\theta(z) = 1 - (1 - \eta)z$, $\eta > 0$, η small. In later sections we set $\eta = 0.01$.

interest because, as we discuss in a later section, it rationalizes the widely-cited filter proposed in BK. In this case, the $\hat{B}^{p,f}$'s are obtained by adjusting all the $\hat{B}^{p,f}$'s in the IID case by a constant to ensure $\hat{B}^{p,f}(1) = 0$. Note in Figure 2a how, relative to the IID case, the $Near\ IID$ case 'lifts' up $\hat{B}^{p,f}(e^{-i\omega})$ in the neighborhood of $\omega = 0$.

The third example is the Random Walk case discussed in the introduction. Now, optimality dictates truncating the ideal band pass filter, and then only adjusting the highest order terms to ensure $\hat{B}^{p,f}(1) = 0$. In this case, f_x assigns high weight in a larger neighborhood of $\omega = 0$ than in the $Near\ IID$ case. Figure 2b shows that the Random Walk $\hat{B}^{p,f}(e^{-i\omega})$ is more accurate in a larger region of $\omega = 0$, at the cost of doing relatively poorly for higher ω. These examples show how shifting power towards a particular frequency range causes the optimal filter to become more accurate in that range, at the expense of doing poorly elsewhere.

3. Properties of the Optimal Approximation

The purpose of this section is to explore several aspects of the solution to the projection problem, (2.7). We examine the role of asymmetry and time nonstationarity of the $\hat{B}^{p,f}$'s. We do this by solving the projection problem under various constraints. In the most constrained version we impose stationarity and symmetry by setting $p = f$ and holding p constant for all t. We then relax stationarity by letting p be as large as possible for each t. After that, we relax symmetry by allowing $p \neq f$, and letting both p and f be as large as possible for each t. Finally, we ask how crucial it is to know the details of the time series representation of x_t. In particular, we investigate whether the filter described in the introduction - the one which assumes a random walk x_t - works well for non-random walk processes.

Clearly, the results of our analysis depend on the actual time series properties of x_t, as summarized by f_x. To make the analysis interesting, we consider difference stationary and trend stationary time series representations that fit standard US macroeconomic data.

Our findings are as follows. First, we find that nonstationarity and asymmetry are useful in minimizing (1.1), with nonstationarity being relatively more important. These gains reflect that allowing nonstationarity and asymmetry substantially increases the amount of information in x that can be used in estimating y_t. Second, we show that the degree of nonstationarity in the optimal filter approximation is quantitatively small.[20] We do this by demonstrating that second moment statistics of $\hat{y}_t$ do not vary much with t. Third, we show that the degree of asymmetry in the optimal filter is small. We do this by showing that the dynamic cross correlation function between $\hat{y}_t$ and y_t is nearly symmetric about zero, even for t near the beginning of the data set.[21] Finally, we find that the gain from using the true time series representation of x_t to compute $\hat{y}_t$ rather than proceeding as though x_t is a random walk, is minimal in practice. These findings underlie our view that an adequate, though perhaps not optimal, procedure

[20] This is similar to results obtained for the Hodrick-Prescott filter, which is also nonstationary and asymmetric. Christiano and den Haan (1996) show that, apart from data at the very beginning and end of the data set, the degree of nonstationarity in this filter is quantitatively small.

[21] For a detailed discussion of the link between asymmetry in $\hat{B}^{p,f}$ and asymmetry in the dynamic cross correlation between y_t and $\hat{y}_t$, see CF, page 14.

for isolating frequency bands in macroeconomic time series is to proceed as if the data were a random walk and use filters that are optimal in that case.

We begin by describing the time series models used in our analysis. We then study the properties of the solution to (2.7) for these models. The models are difference stationary. In CF we show that our essential results hold, even if we adopt trend stationary time series representations that fit US macroeconomic data well. In particular, we find that our Random Walk filter is nearly optimal.

We estimated time series models of the form (2.8) for four data sets often studied in macroeconomic analysis. In each case, we fit a model to the monthly, quarterly and annual data. The variables, x_t, considered are: inflation, output (GDP for the annual and quarterly frequencies and industrial production for monthly), the rate of interest (measured by the three-month return on US Treasury bills) and the unemployment rate. Inflation is measured as the first difference of the log of the consumer price index (CPI); the rate of interest is measured as the logarithm of the net rate; and output was transformed using the logarithm. The data set covers the period 1960-1997. Since the results based on these time series representations are similar, we do not reproduce them all here (see CF for details). Instead, we just present results based on the time series model estimated using monthly inflation data. We chose this representation because it provides the weakest support for our contention that the Random Walk filter is close to optimal. The estimated representation is:

$$(1-L)x_t = \varepsilon_t - 0.75\varepsilon_{t-1},\ E\varepsilon_t^2 = 0.0021^2,$$

where $x_t = log(CPI_t/CPI_{t-1})$.

Table 1: Band Pass Filter Approximation Procedures Considered

Name	Definition
Optimal	Optimal
Random Walk	Optimal, assuming random walk x_t ((1.2)-(1.3))
Optimal, Symmetric	Optimal, subject to $p = f$
Optimal, Fixed	Optimal, subject to $p = f = 36$
Random Walk, Fixed	Optimal, subject to $p = f = 36$, assuming random walk x_t

Note: (i) The various procedures optimize (1.1) subject to the indicated constraints. Where the time series representation of x_t is not indicated, it will be clear from the context. (ii) We use $p = 36$ because this is recommended by BK.

We evaluate various procedures for computing $\hat{y}_t$ under a variety of specifications of the time series representation for x_t, various data sampling intervals and frequency bands. We compare $\hat{y}_t$ and y_t using $corr_t(\hat{y}_t, y_{t-\tau})$ and $(Var_t(\hat{y}_t)/Var(y_t))^{1/2}$, for various t and τ. Here, $corr$ and Var correspond to the correlation and variance of the indicated variables.[22] These statistics are closely related to our optimization criterion, and for $\tau \neq 0$ they allow us to assess the degree of asymmetry in filters. The procedures we consider are listed in Table 1. Comparison of Optimal Symmetric and Optimal Fixed permits us to assess the importance of time stationarity in the filter. Comparison of Optimal and Optimal Symmetric permits us to assess the importance of symmetry. Comparison of Optimal and Random Walk permits us to assess the importance of getting the details of the time series representation of x_t just right.

We now summarize the results obtained using the inflation representation presented above, and using the *Near IID* representation described in the previous section. We consider three frequency bands: 1.5 to 8 years, 8 to 20 years, and 20 to 40 years.

Figure 3 presents the results based on the monthly inflation time series representation. The first, second and third columns of the figure provide information on $corr_t(\hat{y}_t, y_t)$, $[Var_t(\hat{y}_t)/Var(y_t)]^{1/2}$, and $corr_t(\hat{y}_t, y_{t-k})$, respectively. We report results for $t = 1, ..., 240$, since statistics are symmetric across the first and second halves of the sample. The first, second and third rows in the figure correspond to three different frequency bands: 1.5 to 8 years; 8 to 20 years; and 20 to 40 years, respectively. Each panel in the first column contains four curves, differentiated according to the procedure used to compute $\hat{y}_t$: Optimal, Random Walk, Optimal Symmetric, or Optimal Fixed. Results for Optimal and Random Walk are presented for $t = 1, ..., T/2$. Results for Optimal Symmetric and Optimal Fixed are presented for $t = 37, ..., T/2$, since we set $p = f = 36$. Here, $T = 480$, which is slightly more than the number of monthly observations used to estimate the time series representation for inflation. The second column contains results for Optimal and Random Walk alone. Here, results for Optimal are simply repeated for convenience from the first column. For filters that solve a projection problem, the correlation and relative standard deviation coincide. Finally, the third column reports $corr_t(\hat{y}_t, y_{t-k})$ for five different values of t: $t = 1$, 31, 61, 121, 240. In each case, k ranges from -24 to 24. Also, the location of $k = 0$ is indicated by a '+'.

The main findings in Figure 3 are as follows. First, the efficiency differences between Random Walk and Optimal are very small (column 1). A minor exception to this can be found in the business cycle frequencies for $t = 4, ..., 11$. For these dates, the difference between $corr_t(\hat{y}_t, y_t)$ based on Optimal and Random Walk is between 0.08 and 0.12. Although these differences are noticeable, they do not seem quantitatively large. Moreover, the differences between Random Walk and Optimal are barely visible when the analysis is based on time series representations fit to the other macroeconomic time series discussed above.

Second, imposing symmetry (see Optimal Symmetric) results in a relatively small loss of efficiency in the center of the data set, but that loss grows in the tails. Third, imposing

[22] Formally, these statistics are defined as follows. Suppose $[x_t, z_t]$ is a vector stochastic process and consider the various realizations of this stochastic process at date t. Then, $corr_t(x_t, z_t)$ is the correlation between x_t and z_t across realizations at t. Similarly, $Var_t(z_t)$ is the variance, across realizations, at t. Since y_t is covariance stationary, $Var_t(y_t)$ is the same for all t, and so we drop the t subscript on the variance operator in this case. For details about how we computed these statistics, see CF.

stationarity in addition to symmetry (Optimal Fixed) results in a noticeable loss of efficiency throughout the data set. However, the efficiency losses due to the imposition of symmetry and stationarity are comparatively small in the business cycle frequencies. They are dramatic in the lowest frequencies.

Fourth, $\hat{y}_t$ based on Optimal and Random Walk appears to be reasonably stationary, except in an area near the tails. This tail area is fairly small (about 1.5 years) for the business cycle frequencies, but it grows for the lower frequencies (columns 1 and 2). It is interesting to note that, for Optimal, $Var_t(\hat{y}_t)$ falls in the tail area. This pattern holds for all optimal estimates of $\hat{y}_t$ reported in this paper. The intuition is simple. Recall that $\hat{y}_t$ is the solution to the projection of y_t onto the data. By the smoothing properties of projections, we expect the variance of $\hat{y}_t$ to be smaller than that of y_t. These two variables differ to the extent that the pre- and post-sample observations on x_t play an important role in y_t. This implies that in the middle of a fairly large data set, y_t and $\hat{y}_t$ will be quite similar. So, in this case we expect the variance of $\hat{y}_t$ to be only a little smaller than the variance of y_t. However, at the beginning and the end of a data set, there is a substantial 'missing data' problem. For data points like this we expect the smoothing properties of projections to be particularly important, implying that the variance of $\hat{y}_t$ is substantially below that of y_t.

Fifth, Random Walk seems to imply very little asymmetry in the cross correlations between $\hat{y}_t$ and y_t (Column 3). The degree of symmetry in these cross correlations is notable. It is particularly surprising in the case, $t = 1$, when $\hat{B}^{p,f}$ is one-sided.

We conclude from these results that the noticeable efficiency gains obtained by filters that use all the data come at little cost in terms of nonstationarity and asymmetry. However, the gains of going from a simple procedure like Random Walk to Optimal are quite small.

These findings apply to the time series representations fit to the four macroeconomic time series discussed above. The conclusions obviously do not hold up in all conceivable circumstances. When we analyzed the *Near IID* case, we found some evidence against the proposition that Random Walk is nearly optimal and roughly stationary (see CF for details). However, Random Walk continues to dominate Optimal Fixed and it is nearly optimal in the business cycle frequencies, outside of tail areas. In addition, our other conclusions continue to hold: (i) optimal is still nearly stationary outside of tail areas; (ii) imposing symmetry and time-stationarity on the filter results in noticeable efficiency costs in the low frequencies.

4. Application to Inflation and Money Growth

We illustrate the use of our recommended filter using data on money growth and inflation. We apply a bootstrap methodology to assess statistical significance of correlations based on filtered data. We divide the annual data on CPI inflation and M2 money growth for the period 1900 to 1997 into two parts: data covering the period 1900-1960 and data covering 1961-1997. The data are broken into three sets of frequencies: those corresponding to 2-8 years (the business cycle), 8-20 years and 20-40 years. Figures 4a-d displays results for the pre-1960 period and Figures 5a-d displays results for the post-1960 period.

We find that in the pre-1960 period, the two variables move together closely in all frequency bands. The relationship remains positive in the low frequencies in the post-1960 data. However, in the business cycle and 8-20 year frequencies, there is a substantial and statistically significant change. The positive relationship between inflation and money growth in the early period can be seen in Figure 4a, which displays the raw data. Figures 4b-d indicate that this positive relationship holds in all frequency components. This is confirmed by the results in Table 2, which show that the correlations are positive and statistically significantly different from zero in each frequency band.[23]

Table 2: Money Growth-Inflation Correlations

Sample	Business Cycle Frequency	8-20 years	20-40 years
1900 - 1960	0.45 (0.00)	0.59 (0.04)	0.95 (0.01)
1961 - 1997	-0.72 (0.00) [0.00]	-0.77 (0.02)[0.00]	0.90 (0.10)[0.37]

Notes: (i) contemporaneous correlation between indicated two variables, over indicated sample periods and frequencies. (ii) Numbers in parentheses are p-values, in decimals, against the null hypothesis of zero correlation at all frequencies. (iii) Numbers in square brackets are p-values, in decimals, against the null hypothesis that the post-1960 correlations are the same as the pre-1960 correlstions. For details, see footnote 23.

Figures 5a-d depict the same data over the post-1960 period. Note first that the point estimate of the correlation between inflation and money growth in the 20-40 year frequency band is quite similar across the two periods. This is consistent with the notion that the relationship has not changed in the very lowest frequencies. However, the correlation in the later period is imprecisely determined and so this failure to reject could reflect low power. The difficulty in pinning down the correlation in the later period is not surprising, given the

[23]This footnote discusses the p-values that appear in Table 2. The ones that appear in parentheses were computed using a bootstrap procedure under the null hypothesis that inflation and money growth are unrelated. We fit separate q-lag scalar autoregressive representations to of inflation (first difference, log CPI) and to money growth (first difference, log $M2$). We use the fitted disturbances and actual historical initial conditions to simulate 2,000 artificial data sets on inflation and money growth. For both the early and late sample, the amount of data simulated corresponds to the amount of data in the sample. For the pre-1960s annual data, $q = 3$; and for the post-1960's monthly data, $q = 12$. In each artificial data set we compute correlations between the various frequency components, using the same procedure applied in the actual data. In the data and the simulations, we dropped the first and last three years of the filtered data before computing sample correlations. The numbers in parentheses in Table 2 are the frequency of times that the simulated correlation is greater (less) than the positive (negative) estimated correlation.

The p-values that appear in square brackets are the fraction of times, in 2000 artificial post-1960 data sets generated by a pre-1960 data generating mechanism (DGM), that the contemporaneous correlation between the indicated frequency components of inflation and money growth exceeds, in absolute value, the corresponding post 1960 empirical estimate. The DGM used in these simulations is a 3-lag, bivariate vector autoregression fit to pre-1960 data.

relatively short span of the post-1960 data set. Still, we are impressed by the similarity in the low frequencies between the two data sets.

Now consider the correlations in the higher frequencies, which are estimated precisely. Note that there is a striking change in the relationship between the variables at these frequencies. Inflation and money growth are now strongly and statistically significantly negatively correlated. Table 2 offers another way to see this. The numbers in square brackets are $p-$values for a test of the null hypothesis that the business cycle and 8-20 year correlations in the post-1960 coincide with the corresponding pre-1960 correlation. That null hypothesis is strongly rejected. An interpretation of the change is that in these frequencies, inflation now lags money growth by a few years. We think that the change in the dynamic relationship between money growth and inflation is an interesting phenomenon to explain. One possibility is that sticky prices and wages became a more significant phenomenon in the later period. This in turn may be a consequence of the lower volatility of that period, something that is evident in the figures. Endogenous models of sticky prices, such as those of Dotsey, King and Wolman (1999) and Ireland (1997) suggest that prices are more flexible in volatile environments.

5. Comparison with Other Filters

This section compares our band-pass filter approximations with other filtering approaches used in the literature. These alternatives include the HP filter and the band-pass filtering approach recommended by BK. In addition, we consider the band-pass approximation based on regressing data on sine and cosine functions, as described in Christiano and Fitzgerald (1998, Appendix) and Hamilton (1994, pages 158-163). We call this last filter the Trigonometric Regression filter. Our analysis is based on time series representations fit to the four macroeconomic data series discussed in section 3.

We find that, in terms of our optimality criterion, the Random Walk filter dominates the BK and Trigonometric Regression filters. The differences are most pronounced for filter approximations designed to extract frequencies lower than the business cycle.

Our comparison of the Random Walk and HP filters is inspired by one interpretation of the HP filter, according to which it approximates a particular high-pass filter.[24] We show that our Random Walk filter delivers a better approximation to that high-pass filter than the HP filter does.

Although the Random Walk filter is a better approximation to a high pass filter than is HP, CF shows that the improvement is not large enough to produce quantitatively large differences in the sort of statistics business cycle analysts are interested in. In this sense, we view our results as confirming the value of the HP filter as a device for extracting the business cycle and higher frequency components in quarterly data. So why might a researcher who already uses the HP filter be interested in the band pass filter? The key advantage of the band pass filter is that it expands the range of questions one can explore. By suitably 'tuning' a band pass filter,

[24]See King and Rebelo (1993), and Singleton (1988). A high pass filter is a band pass filter with $p_l = 2$. That is, it permits all frequencies above a specified one (i.e., the one associated with period $p_u > 2$) to pass.

one can ask questions involving different frequency components of the data.[25] This feature of the band pass filter is what allowed us to do the analysis in the previous section. This type of analysis is not feasible with the HP filter. Another advantage of the band pass filter is that the adjustments necessary for handling monthly or annual data are quite natural. Dealing with alternative data sampling intervals is problematic for the HP filter.[26]

Finally, this section also evaluates the filters from the perspective of the covariance stationarity properties of $\hat{y}_t$. For both HP and Random Walk the second moment properties of $\hat{y}_t$ appear to be approximately constant with respect to t, outside of tail areas. On this dimension the performance of the HP and Random Walk filters is similar to BK, which produces a $\hat{y}_t$ that is exactly covariance stationary. In contrast, the $\hat{y}_t$ associated with Trigonometric Regression exhibits significant departures from covariance stationarity.

5.1. The Baxter-King Filter

We first summarize the differences between the BK filter approximation strategy and our strategy. We then compare the performance of BK with filters obtained using our strategy.

5.1.1. The Baxter-King Approximation Strategy

The filter proposed in BK is the fixed-lag, symmetric filter defined in section 2.4.[27] They arrive at this approximation by choosing the filter weights to minimize the unweighted integral of the squared approximation error, $|\hat{B}^{p,p}(e^{-i\omega}) - B(e^{-i\omega})|^2$, over $\omega \in (-\pi, \pi)$, subject to the constraint, $\hat{B}^{p,p}(1) = 0$. They impose the constraint on the grounds that if it were not imposed, then if $\hat{B}^{p,p}(L)$ is applied to a trending series, the result has a trend too.

The BK approximation strategy differs from ours in several respects. First, as noted in section 2.4, our strategy is to choose the approximation so that $\hat{y}_t$ and y_t are as close as possible. This leads to the criterion in (2.7), in which the squared approximation errors, $|\hat{B}^{p,f}(e^{-i\omega}) - B(e^{-i\omega})|^2$, are weighted by the spectral density of the data being filtered. Second, in our approach $\hat{B}^{p,f}(1) = 0$ is never imposed as a constraint. It emerges as a feature of the solution

[25] The perspective adopted in this paper does suggest one strategy for designing the HP filter to isolate alternative frequency bands: optimize, by choice of λ, the version of (2.7) with $B^{p,f}$ replaced by the Hodrick-Prescott filter. This strategy produces a value of λ that is time-dependent and dependent upon the properties of the true time series representation. We doubt that this strategy for filtering the data is a good one. First, implementing it is likely to be computationally burdensome. Second, as this paper shows, identifying the optimal band-pass filter approximation is straightforward.

[26] In part, this is due to the fact that there is not complete agreement as to what precisely one is trying to extract from the data with the HP filter. Some say the filter simply draws a smooth line (Marcet and Ravn (2000), Prescott (1986)), others (King and Rebelo (1993), Ravn and Uhlig (1997)) say it approximates a high pass filter, others (Hodrick and Prescott (1997)) say it extracts the trend component in a particular trend-cycle statistical model of the data. Given this lack of agreement, there is no natural, single way to adapt the HP filter for monthly or annual data. For example, Marcet and Ravn (2000) and Ravn and Uhlig (1997) both address this problem and come up with different solutions.

[27] Early applications of this filter can be found in Baxter (1994), King and Watson (1994) and King, Stock and Watson (1995).

if the data contain a unit root. Third, not surprisingly, our approximation strategy in general leads to a different filter than does BK's strategy. Only in exceptional cases will the two produce the same result. For example, they do so if x_t is generated by the $Near\ IID$ representation discussed in section 2.4. Finally, because our strategy uses all the data for each t, we allow p and f to vary with t and to be different from each other.

5.1.2. Filter Performance

We found that our Random Walk filter dominates the BK filter for the four macroeconomic time series models described in section 3, and for the $Near\ IID$ case. The primary reason for this is that the Random Walk filter fully exploits the entire data set. For example, the Random Walk filter dominates even in the $Near\ IID$ case when the BK filter is the optimal fixed lag and symmetric filter (recall the discussion in section 3). Despite this finding, the differences we found are quantitatively small for standard statistics based on the business cycle frequencies (see CF for additional evidence on this). However, they are large for statistics based on lower frequency components of the data. For example, when data are generated using a time series model that fits the monthly US time series on inflation, then BK does a poor job of extracting the 8 to 20 year component of inflation. To see this, recall that BK is worse than Optimal Fixed, which understates the standard deviation of the 8 to 20 year component of inflation by one-half.[28] In the $Near\ IID$ case, when BK is Optimal Fixed, CF report that BK understates this standard deviation by around two-thirds. This poor performance of BK in the lower frequencies makes it ill-suited for the type of application studied in the previous section.

Of course, BK would probably work better if p and f were both increased when extracting lower frequency components of the data. But, this introduces its own practical problems. First, increasing p and f requires throwing away more data at the beginning and the end of a data set. Second, the proper criterion for choosing p and f is not clear. These complications are completely sidestepped by a procedure like our Random Walk, which uses all the data all the time.

5.2. Hodrick-Prescott Filter

We evaluate the HP filter as a high-pass filter that isolates frequencies 8 years and higher in data.[29] We pay particular attention to the relative performance of Random Walk and HP near the end of a data sample. This allows us to evaluate the real-time performance of the two filters. We use time series representations fit to quarterly data on log GDP, the unemployment rate and inflation.[30] We use quarterly data because this is the frequency often used in practice.

[28] This can be seen in the 2,1 entry in Figure 3. Although that reports $corr_t(\hat{y}_t, y_t)$, we know that when $\hat{y}_t$ is the solution to a projection problem, as it is for Optimal Fixed, then $corr_t(\hat{y}_t, y_t) = [Var_t(\hat{y}_t)/Var(y_t)]^{1/2}$.

[29] The HP filter parameter, λ, is set to 1600, as is typical in applications using quarterly data.

[30] The time series model estimated using the unemployment rate, x_t, is:

$$(1-L)x_t = \varepsilon_t + 0.65\varepsilon_{t-1} + 0.48\varepsilon_{t-2} + 0.41\varepsilon_{t-3},\ \sigma_\varepsilon = 0.27$$

The discussion that follows is divided into two parts. We begin with the evidence on $\hat{y}_t$ in Columns 1 and 2 of Figure 6, ignoring the first 2-5 years' observations. We then focus separately on the latter observations and on the information in Column 3, because this allows us to assess the real-time performance of the filters. The HP filter performs very poorly on this dimension. Indeed, in the case of GDP it is useless. However, even the optimal procedure seems relatively unreliable in real-time.[31]

5.2.1. Performance Outside Tail Areas

The first column in Figure 6 displays $corr_t(\hat{y}_t, y_t)$ associated with the HP, Random Walk, and Optimal Fixed filters for $t = 1, ..., 80$ and for the indicated three quarterly time series models. We do not display these statistics for Optimal, because they are virtually indistinguishable from Random Walk. Even though HP uses all the data, it nevertheless performs slightly less well than Optimal Fixed. It also performs less well than Random Walk, particularly for the time series representations associated Unemployment and GDP. However, the differences are not dramatic. For the Random Walk filter, $corr_t(\hat{y}_t, y_t)$ exceeds 0.95 and is often in the neighborhood of 0.99. The corresponding magnitude for the HP filter is a little below 0.90.

The results in Column 2 indicate very little difference between the two filters. The HP filter slightly overstates the variance of y_t, while the Random Walk filter slightly understates it.

Note how, apart from tail areas, the curves in Columns 1 and 2 are reasonably flat. This is consistent with the notion that both filters produce data that are reasonably consistent with covariance stationarity.

For later purposes, it is convenient to introduce another statistic for comparing HP filter and Random Walk. Let R_t the absolute size of the typical estimation error, $\hat{y}_t - y_t$ (measured by its standard deviation), to the absolute size of the typical value of y_t (measured by its standard deviation):[32]

$$R_t = \left[\frac{Var_t(\hat{y}_t - y_t)}{Var(y_t)}\right]^{1/2}.$$

This variable is measured in percentage points. The time series model estimated using log GDP data is

$$(1-L)x_t = \varepsilon_t + 0.25\varepsilon_{t-1} + 0.16\varepsilon_{t-2} + 0.10\varepsilon_{t-3} + 0.12\varepsilon_{t-4},\ \sigma_\varepsilon = 0.0088$$

The time series model estimated using $\log(P_t/P_{t-1})$, where P_t is the CPI, is:

$$(1-L)x_t = \varepsilon_t - 0.23\varepsilon_{t-1} - 0.27\varepsilon_{t-2} + 0.32\varepsilon_{t-3},\ \sigma_\varepsilon = 0.0042$$

[31] We have abstracted from several real-time issues which could make the HP filter, Optimal and Random Walk seem even worse at estimating y_t in real time. We abstract from possible breaks in the underlying time series representation, and from data revisions. A more complete analysis would also take these factors into account in characterizing the accuracy of real time estimates of the business cycle and higher frequency components of the data. For further discussion, see Orphanides (1999) and Orphanides and van Norden (1999).

[32] When $\hat{y}_t$ solves (2.7), R_t and $corr_t(\hat{y}_t, y_t)$ have a monotone relationship. Neither filter discussed in this paragraph satisfies this condition.

A large value of R_t indicates a poor filter approximation. In the extreme case when R_t is greater than or equal to unity, then the filter approximation is literally useless. In this case one can do just as well, or better, estimating y_t by its mean with $\hat{y}_t \equiv 0$. When we apply the Random Walk filter to the time series representations fit to Unemployment, GDP and Inflation, we find that R_t is no greater than 0.31 if we ignore the first two years' data. In the case of the HP filter, this number is 0.49 for Unemployment and GDP and around 0.37 for Inflation. These results complement the findings above. Outside of tail areas, Random Walk outperforms the HP filter, though not by a lot.

5.2.2. Real-Time Performance

We now investigate the effectiveness of the HP, Random Walk and Optimal filters in the tail area of the data. This is of interest when real-time estimates of y_t are desired. One example is stabilization policy, when real-time estimates of the output and unemployment gaps are of interest.[33] In practice, some analysts estimate these gaps using the HP filter. One interpretation is that they define gaps as the sum of the business cycle and higher frequency components of the data.[34]

At the outset, it should be clear that estimating the current value of y_t is likely to be a difficult task. In practice, it is very hard to say without the light of hindsight whether a given change in a variable is temporary (i.e., part of y_t) or more persistent (i.e., part of $\tilde{x}_t$). So, we can expect that even our best real-time estimates of y_t will be disappointing. This is indeed the case for the Random Walk and HP filters. However, we show that the estimate based on HP has lower quality than the one based on Random Walk.

We are interested in the accuracy of $\hat{y}_T$ in estimating y_T. However, the results in Columns 1 and 2 of Figure 6 only pertain to the first half of the data set. By symmetry, the second half is the mirror image of the first half. So, statistics for $\hat{y}_T$ correspond to the ones reported for $\hat{y}_1$.[35] Note that the correlation between $\hat{y}_t$ and y_t is relatively low for $t = T$ for both the Random Walk and HP filters. For example, in the case of Random Walk, the correlation is roughly 0.65 for each data series. This implies that $\hat{y}_T$ accounts for only about 40 percent of the variation in y_T. This is a substantial deterioration relative to the results one obtains for t closer to the middle of the data set.

Column 2 turns up some evidence of differences in the performance of the HP and Random Walk filters. Consistent with the fact that Random Walk is nearly optimal, we see that $Var_T(\hat{y}_T)$ is small relative to $Var_t(\hat{y}_t)$ for $1 < t < T$. As discussed in section 3, we interpret this as reflecting the smoothing properties of projections and the fact that there is relatively little information about y_t at the end of the data sample. The relatively low variance of $\hat{y}_T$ indicates that the trend implied by $\hat{y}_t$ moves more closely with the raw data for observations near the

[33] See Orphanides (1999), who argues that the output gap plays a role in the Federal Reserve's monetary policy strategy.

[34] Consistent with this interpretation, Orphanides and van Norden (1999) treat the output gap and the business cycle as synonyms. For example, according to them (p. 1), 'The difference between [actual output and potential output] is commonly referred to as the *business cycle or the output gap* (italics added).'

[35] In the case of Random Walk, $\hat{y}_T$ is computed using the one-sided filter, (1.4).

end of a sample than for observations in the middle.[36] In contrast, $Var_T(\hat{y}_T)$ implied by HP is large relative to $Var_t(\hat{y}_t)$ for $1 < t < T$, when the data are generated by the output and the inflation time series representations.[37] So, in the case of these time series representations, the trend implicit in HP appears to move less closely with the raw data at the end of the data sample than in the middle.[38]

We now compare Random Walk and HP filter using the R_t statistic for $t = T$. Using the Random Walk filter, $R_T = 0.77$, 0.78, and 0.69 for GDP, unemployment, and inflation respectively. Note that these numbers are substantially larger than what they are for data points closer to the middle of the sample. Still, they indicate Random Walk provides at least *some* information about y_T. Now consider HP filter. For GDP, $R_T = 1.01$. For unemployment and inflation, R_T is 1.03 and 0.80, respectively. Evidently, these statistics indicate that Random Walk dominates HP filter in real time. Moreover, for purposes of estimating the GDP and unemployment gaps in real time, HP filter is worse than useless.[39] The estimate, $\hat{y}_T = 0$, produces a smaller error than using the HP filter estimate, $\hat{y}_T$.

The statistics on the real-time properties of the filters that we have just considered abstract from scale. The evidence in Column 3 exhibits the magnitude of the error in real-time gap estimates for our variables. We consider the standard deviation of the error, $y_t - \hat{y}_t$, for a fixed date, $t = 160$. Specifically, we study $[Var_T(\hat{y}_{160} - y_{160})]^{1/2}$ for $T = 160, 161, ..., 200$, based on the Random Walk, Optimal and HP filters.[40] These results allow us to quantify the value of hindsight when estimating y_t.

There are several things worth emphasizing in the third column of Figure 6. First, Random Walk and Optimal essentially coincide, and both dominate the HP filter. Second the error in estimating y_{160} declines by roughly one-half in the first year (i.e., four observations) after $t = 160$. Thereafter, further declines in the error come more slowly. Third, the error of the HP filter asymptotes to a relatively high level. The reason is that, as the size of the data set grows, the HP filter does not asymptote to the ideal band pass filter. By contrast, both Random Walk and Optimal do. If T were allowed to grow indefinitely, $[Var_T(\hat{y}_{160} - y_{160})]^{1/2}$ would shrink to zero for Random Walk and Optimal. The information in Figure 6 suggests that this requires a very large value of T. These results are the basis for our conclusion that Random Walk is nearly optimal and outperforms the HP filter in real time.[41]

[36] The trend implicit in $\hat{y}_t$ is $x_t - \hat{y}_t$. In the text, we follow convention in adopting the variance as the measure of distance between two random variables. Thus, $Var_t(\hat{y}_t)$ is the distance between the trend implicit in $\hat{y}_t$ and the raw data, x_t.

[37] The same pattern for $Var_t(\hat{y}_t)$ is reported in Figure 3, page 316 of Christiano and den Haan (1996).

[38] These are counterexamples to the conjectures by Barrell and Sefton (1995, p. 68) and St-Amant and van Norden (1997, p. 11).

[39] These observations on the HP filter complement those obtained using different methods by others, including Laxton and Tetlow (1992), Orphanides (1999) and St-Amant and van Norden (1997).

[40] The subscript convention adopted here is slightly inconsistent with the convention used elsewhere in the paper. Before, the subscript on Var indicated the specific date that the variance corresponds to. Here, the subscript refers to the data set used to construct $\hat{y}_{160}$. We adopt this notation to avoid proliferating notation, and hope that it will not lead to confusion.

[41] Our results for the unemployment gap can be compared with those reported, using a very different conceptual and econometric framework, by Staiger, Stock and Watson (1997). Their estimated standard deviations of this gap range from 0.46 to 1.25 percentage points, depending on the data used in the analysis.

5.3. Trigonometric Regression

We now discuss the Trigonometric Regression filter. This filter makes use of the entire data set, $x_1, ..., x_T$, to estimate each y_t, as follows:

$$\hat{y}_t = B_t(L)x_t, \ t = 1, ..., T,$$

where

$$\begin{aligned} B_t(L)x_t &= \sum_{l=t-T}^{t-1} \left\{ \frac{2}{T} \sum_{j \in J} \cos(\omega_j l) \right\} x_{t-l}, \text{ if } \frac{T}{2} \notin J \qquad (5.1) \\ &= \sum_{l=t-T}^{t-1} \left\{ \frac{2}{T} \sum_{j \in J, j \neq \frac{T}{2}} \cos(\omega_j l) + \frac{1}{T} \cos(\pi(t-l)) \cos(\pi t) \right\} x_{t-l}, \text{if } \frac{T}{2} \in J \\ t &= 1, ..., T, \ \omega_j = \frac{2\pi}{T} j. \end{aligned}$$

Here, J indexes the set of frequencies we wish to isolate, and is a subset of the integers $1, ..., T/2$.[42] It is easy to see that $B_t(1) = 0$, so that $B_t(L)$ has a unit root for $t = 1, 2, ..., T$.[43] Evidently, $B_t(L)$ only has a second unit root for t in the middle of the data set, when $B_t(L)$ is

---

Because (as they emphasize) this range is so wide, it is not surprising that our estimates fall inside it.

[42] We assume T is even. Also, J is the set of integers between j_1 and j_2, where $j_1 = T/p_u$ and $j_2 = T/p_l$. The representation of $\hat{y}_t$ given in the text, while convenient for our purposes, is not the conventional one. The conventional representation is based on the following relation:

$$\hat{y}_t = \sum_{j \in J} \{a_j \cos(\omega_j t) + b_j \sin(\omega_j t)\},$$

where the a_j's and b_j's are coefficients computed by ordinary least squares regression of x_t on the indicated sine and cosine functions. The regression coefficients are:

$$a_j = \begin{cases} \frac{2}{T} \sum_{t=1}^{T} \cos(\omega_j t) x_t, \ j = 1, ..., T/2 - 1 \\ \frac{1}{T} \sum_{t=1}^{T} \cos(\pi t) x_t, \ j = T/2, \end{cases}, b_j = \begin{cases} \frac{2}{T} \sum_{t=1}^{T} \sin(\omega_j t) x_t, \ j = 1, ..., T/2 - 1 \\ \frac{1}{T} \sum_{t=1}^{T} x_t, \ j = T/2 \end{cases}.$$

The expression in the text is obtained by collecting terms in x_t and making use of the trigonometric identity, $\cos(x)\cos(y) + \sin(x)\sin(y) = \cos(x - y)$.

[43] To see that $B_t(1) = 0$ when $T/2 \notin J$, simply evaluate the sum of the coefficients on $x_1, x_2, ..., x_T$ for each t :

$$\frac{1}{T} \sum_{j \in J} \sum_{l=t-1}^{t-T} 2\cos(\omega_j l) = \frac{1}{T} \sum_{j \in J} \sum_{l=t-1}^{t-T} \left[e^{i\omega_j l} + e^{-i\omega_j l}\right] = \frac{1}{T} \sum_{j \in J} \left[e^{-i\omega_j (t-1)} \frac{1 - e^{i\omega_j T}}{1 - e^{i\omega_j}} + e^{i\omega_j (t-1)} \frac{1 - e^{-i\omega_j T}}{1 - e^{-i\omega_j}}\right] = 0,$$

because $1 - e^{i\omega_j T} = 1 - e^{-i\omega_j T} = 1 - \cos(2\pi j) + \sin(2\pi j) = 1$ for all integers, j.

When $T/2 \in J$, the expression for $B_t(1)$ includes $\sum_{l=t-T}^{t-1} \left\{\frac{1}{T} \cos(\pi(t-l)) \cos(\pi t)\right\}$. This expression is simply the sum of an even number of 1's and -1's, so it sums to 0.

symmetric.[44] For this reason, it is important to drift adjust x_t prior to filtering.

Our basic finding is that when the data are generated by the time series representations discussed in section 3, the performance of Trigonometric Regression is worse than that of Random Walk. Since the results based on these time series representations are fairly similar, we present only those based on the data generating mechanism for inflation. These are displayed in Figure 7, which has the same format as Figure 3. The results for Random Walk and Optimal in Figure 7 correspond to those reported in Figure 3, and are reproduced here for convenience. In Column 1, we see that in terms of $corr_t(\hat{y}_t, y_t)$, Trigonometric Regression is outperformed in all frequency ranges by Random Walk, which is nearly optimal. Column 2 shows that the estimates of y_t based on Trigonometric Regression overshoot $Var(y_t)$, sometimes by a great deal, and performs worse on this dimension than either Random Walk or Optimal. The relative performance of Trigonometric Regression is particularly poor in the lower frequencies. Column 3 displays the dynamic cross correlations between $\hat{y}_t$ and y_t when the former are computed by Trigonometric Regression. The evidence suggests that there is little asymmetry in the band pass filter approximation implied by Trigonometric Regression, but there appears to be a substantial departure from covariance stationarity. The correlations in the tails of the data set are notably smaller than they are in the middle.

Although Trigonometric Regression appears in Figure 7 to be substantially worse than Random Walk, for some purposes the poor performance may not be quantitatively important. For example, in CF we report that, for standard business cycle statistics, Trigonometric Regression produces results quite similar to Random Walk.

6. Conclusion

The filtering methodology outlined here is not for everybody. For analysts exclusively interested in statistics based on business cycle and higher frequency components of quarterly data, the HP filter appears to do just fine. However, researchers who are also interested in (i) other frequency components of the data, (ii) daily, weekly, monthly or annual data, or (iii) real-time trend estimates, may prefer to use the band pass filter. This filter offers a simple, consistent framework doing what the HP filter is good at, plus it can also handle (i)-(iii).

This paper derives the optimal approximation to the band pass filter. We compare this approximation with several alternatives, including BK, Trigonometric Regression and Random Walk. We find that for applications that involve the business cycle and higher frequency components of the data, it makes little difference which of these methods is applied. However, for lower frequency components of the data Random Walk outperforms BK and Trigonometric Regression, and is nearly optimal. We established this for time series representations that fit US macroeconomic data on unemployment, output, inflation and interest rates. An advantage of

[44]When T is even, then there cannot be an *exact* second unit root since it rules out the existence of a date precisely in the middle of the dataset. By $B_t(L)$ having n unit roots we mean that it can be expressed as $\tilde{B}_t(L)(1-L)^n$, where $\tilde{B}_t(L)$ is a finite ordered polynomial. The discussion of two unit roots in the text exploits the fact that the roots of a symmetric polynomial come in pairs.

Random Walk over the optimal approximation is that it is very easy to implement. The latter requires first estimating a time series model for the data to be filtered. These considerations are the basis for our recommendation that Random Walk be the band pass filter approximation used in macroeconomic data.

Since the case for Random Walk relies on its superior performance in lower frequencies, an important question is: should we care about lower frequencies and can we measure them reliably? We presented an empirical example which suggests that the answer to both questions is 'yes'.

References

[1] Backus, David K., and Patrick J. Kehoe, 1992, 'International Evidence on the Historical Properties of Business Cycles,' *American Economic Review*, vol. 82, no. 4, September, pages 864-888.

[2] Barrell, Ray and James Sefton, 1995, 'Output Gaps. Some Evidence from the UK, France and Germany,' *National Institute Economic Review*, February.

[3] Baxter, Marianne, 1994, 'Real Exchange Rates and Real Interest Differentials,' *Journal of Monetary Economics*, 33, pp. 5-37.

[4] Baxter, Marianne, Robert G. King, 1999, 'Measuring Business Cycles: Approximate Band-Pass Filters for Economic Time Series,' *Review of Economics and Statistics*, 81:4, November, pp. 575-593.

[5] Christiano, Lawrence J., and Wouter J. den Haan, 1996, 'Small-Sample Properties of GMM for Business Cycle Analysis ,' *Journal of Business and Economic Statistics*, vol. 14, no. 3, July, pp. 309-327.

[6] Christiano, Lawrence J., and Terry J. Fitzgerald, 1998, 'The Business Cycle: It's Still a Puzzle,' *Economic Perspectives*, 4th Quarter, vol. xxii, issue 4, Federal Reserve Bank of Chicago.

[7] Christiano, Lawrence J., and Terry J. Fitzgerald, 1999, 'The Band Pass Filter,' National Bureau of Economic Research Working Paper number 7257.

[8] Cramer, Harald, M. R. Leadbetter, 1967, *Stationary and Related Stochastic Processes,* New York: Wiley.

[9] Dotsey, Michael, King, Robert, G. and Alexander L. Wolman, 1999, 'State-Dependent Pricing and the General Equilibrium Dynamics of Money and Output', *Quarterly Journal of Economics*, 114(2), May, pages 655-90.

[10] Engle, Robert F., 1974, 'Band Spectrum Regression,' *International Economic Review*, 15 (1), February, pp. 1-11.

[11] Geweke, John, 1978, 'The Revision of Seasonally Adjusted Time Series,' Proceedings of the Business and Economic Statistics Section, American Statistical Association, pp. 320-325.

[12] Hamilton, James, 1994, *Time Series Analysis*, Princeton University Press.

[13] Hodrick, Robert, and Edward Prescott, 1997, 'Post-War Business Cycles: An Empirical Investigation,' *Journal of Money, Credit and Banking,* vol. 29, no. 1, February, pp. 1-16.

[14] Hornstein, Andreas, 1998, 'Inventory Investment and the Business Cycle,' Richmond Federal Reserve Bank *Economic Quarterly*, vol. 84, no. 2, spring, pp. 49-71.

[15] Ireland, Peter N., 1997, 'Stopping Inflations, Big and Small,' *Journal of Money, Credit and Banking*, Part II, November 1997, vol. 29, issue 4, pp. 759-775.

[16] King, Robert, and Sergio Rebelo, 1993, 'Low Frequency Filtering and Real Business Cycles,' *Journal of Economic Dynamics and Control*, vol. 17, pp. 251-231.

[17] King, Robert, James Stock and Mark Watson, 1995, 'Temporal Instability of the Unemployment-Inflation Relationship,' Federal Reserve Bank of Chicago *Economic Perspectives*, May/June Volume XIX, issue 3.

[18] King, Robert, and Mark Watson, 1994, 'The Post-War US Phillips Curve: A Revisionist Econometric History,' Carnegie-Rochester Conference Series on Public Policy, December, vol. 41, no. 0, pp. 152-219.

[19] Laxton, Douglas, and R. Tetlow, 1992, 'A Simple Multivariate Filter for the Measurement of Potential Output,' Technical Report No. 59, Bank of Canada, Ottowa.

[20] Lippi, Marco, 2001, 'Convergence of Stochastic Processes and the Spectral Representation Theorem,' manuscript, Dipartimento di Scienze Economiche, Universita di Roma.

[21] Marcet, Abert, and Morten O. Ravn, 2000, 'The HP-Filter in Cross-Country Comparisons,' unpublished manuscript, Universitat Pompeu Fabra.

[22] Orphanides, Athanasios, 1999, 'The Quest for Prosperity Without Inflation,' unpublished manuscript, Board of Governors of the Federal Reserve.

[23] Orphanides, Athanasios, and Simon van Norden, 1999, 'The Reliability of Output Gap Estimates in Real Time,' manuscript, Federal Reserve Board of Governors, July.

[24] Prescott, Edward, 1986, 'Theory Ahead of Business Cycle Measurement,' Federal Reserve Bank of Minneapolis *Quarterly Review*, vol. 10, no. 4, fall, pp. 9-22.

[25] Ravn, M. O. and Harald Uhlig, 1997, 'On Adjusting the HP Filter for the Frequency of Observations,' unpublished manuscript, Tilburg University.

[26] St-Amant, Pierre, and Simon van Norden, 1997, 'Measurement of the Output Gap: A Discussion of Recent Research and the Bank of Canada,' manuscript, Bank of Canada.

[27] Sargent, Thomas J., 1987, *Macroeconomic Theory*, second edition, Academic Press.

[28] Sims, Christopher, 1972, 'Approximate Prior Restrictions in Distributed Lag Estimation,' *Journal of the American Statistical Association*, vol. 67, no. 337, pp. 169-175.

[29] Singleton, Kenneth, 1988, 'Econometric Issues in the Analysis of Equilibrium Business-Cycle Models,' *Journal of Monetary Economics* 21:361-86.

[30] Staiger, Douglas, James H. Stock and Mark W. Watson, 1997, 'The NAIRU, Unemployment, and Monetary Policy,' *Journal of Economic Perspectives*, vol. 11, no. 1, winter, pages 33-49.

[31] Stock, James, and Mark Watson, 1999, 'Business Cycle Fluctuations in US Macroeconomic Time Series,' *Handbook of Macroeconomics,* Vol. 1A, eds.Michael Woodford and John Taylor, Amsterdam; New York and Oxford: Elsevier Science, North-Holland.

[32] Wallis, K.F., 1983, 'Models for X-11 and X-11-FORECAST Procedures for Preliminary and Revised Seasonal Adjustments,' in A. Zellner, editor, *Applied Time Series Analysis of Economic Data*, ASA-CENSUS-NBER Conference proceeding, U.S. Department of Commerce.

Figure 1a: Ideal Filter Weights

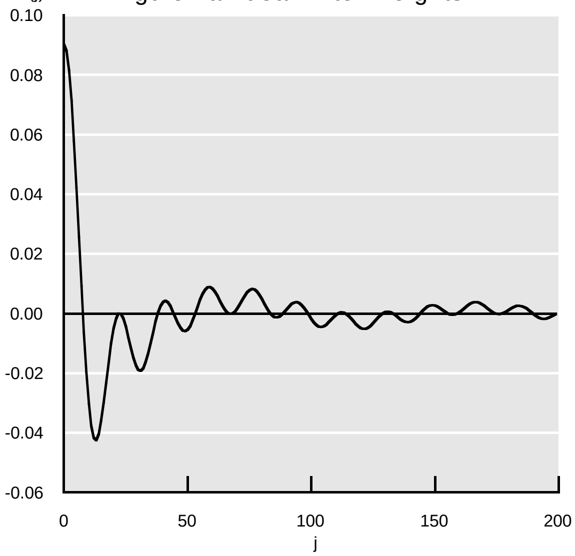


Figure 1b: p=12

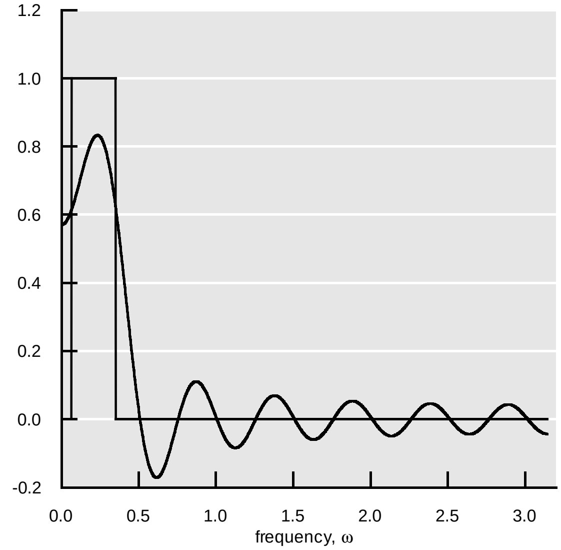


Figure 1c: p=24

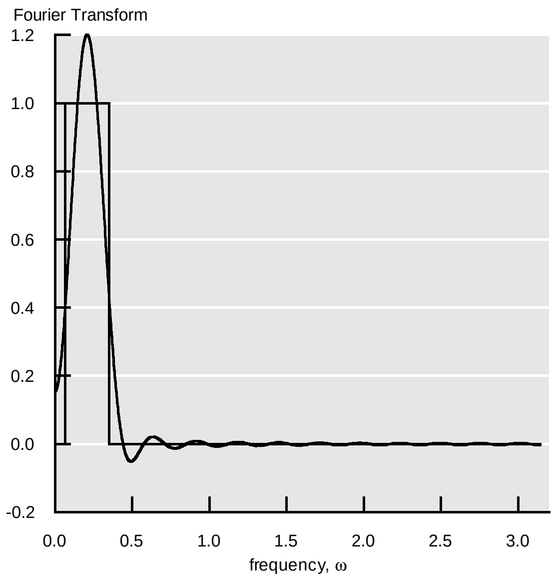


Figure 1d: p=36

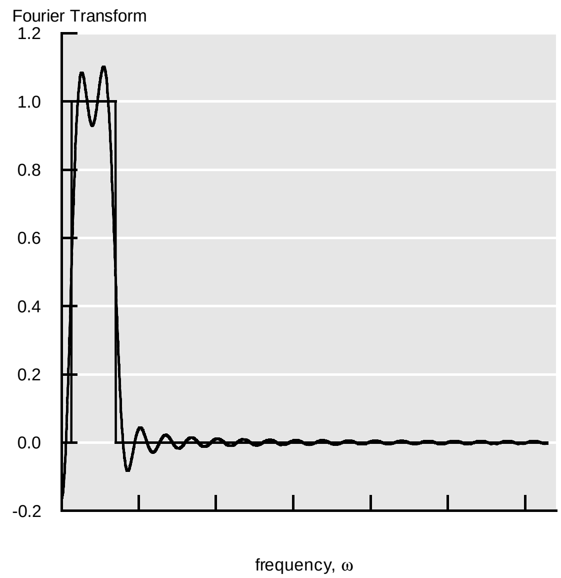


Figure 2a: IID Case

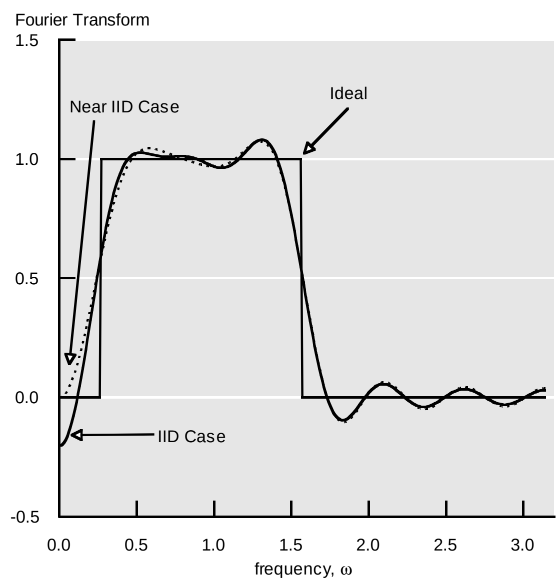


Figure 2b: Random Walk & Near IID Cases

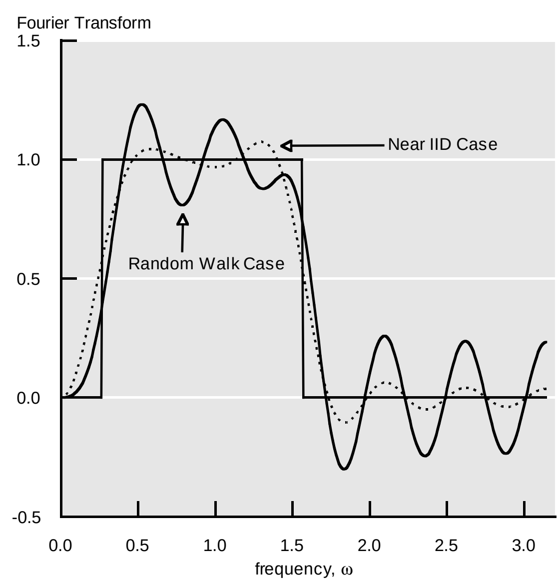
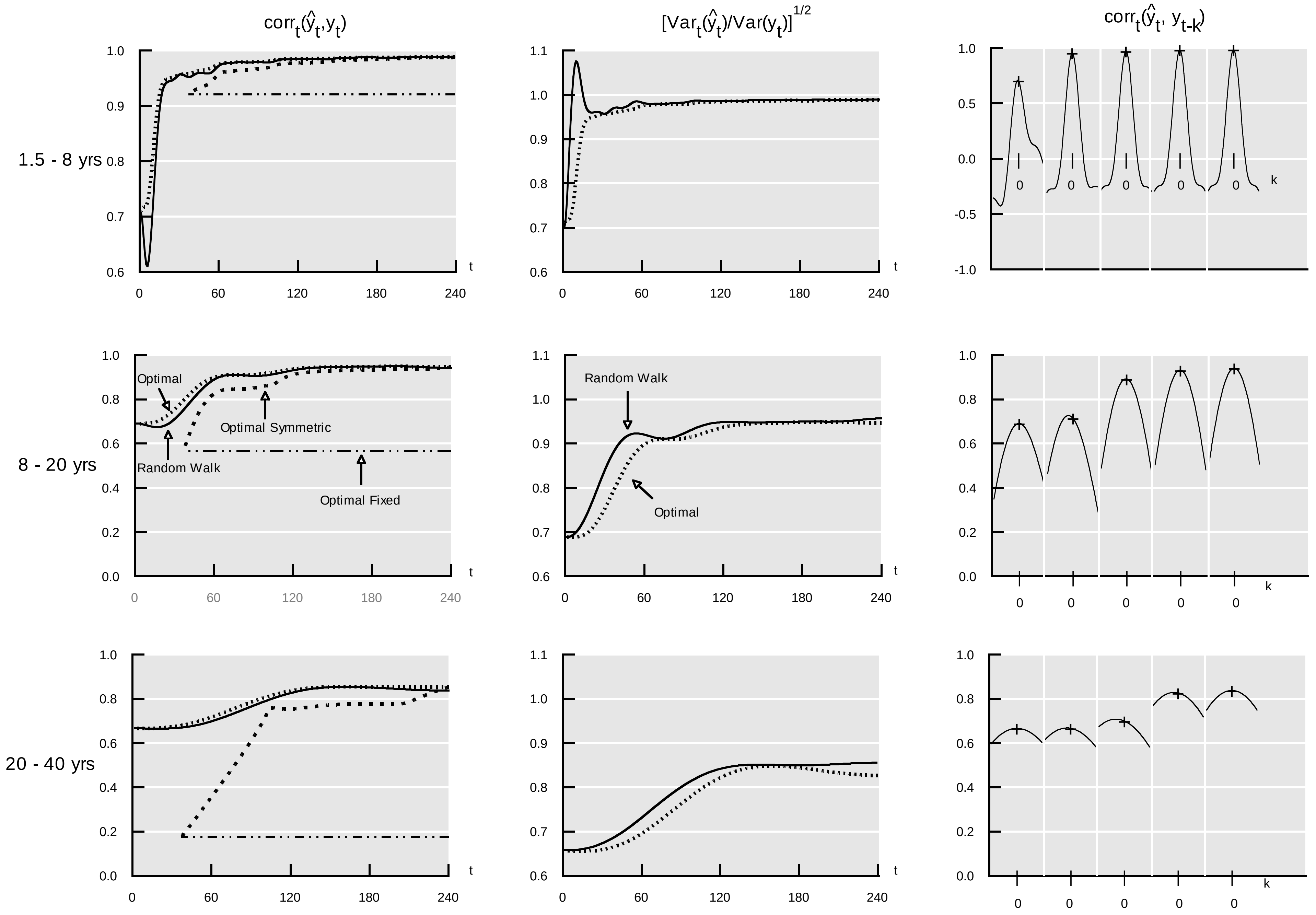


Figure 3: Comparing Filter Approximations - Monthly U.S. Inflation Case



The panels in the first two columns contain curves differentiated according to the underlying procedure used to compute $\hat{y}_t$. See Table 2 for definitions of these procedures. While the sample size is T = 480, the statistics are shown only for t = 1, 2, ..., 240 since the statistics are symmetric across the first and second halves of the sample. The third column displays five correlation functions, $corr_t(\hat{y}_t, y_{t-k})$, associated with the Random Walk filter for t = 1, 31, 61, 121, 240 (1/12, 2 1/2, 5, 10, 20 years). Each correlation function is displayed for k = -24, -23, ..., 0, ..., 23, 24. The plus on the correlation function indicates the location of k = 0.

Figure 4a: M2 Growth and Inflation

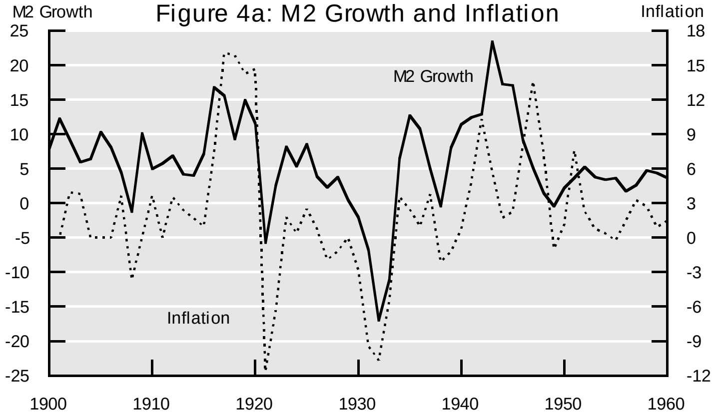


Figure 4b: Component, 2-8 years

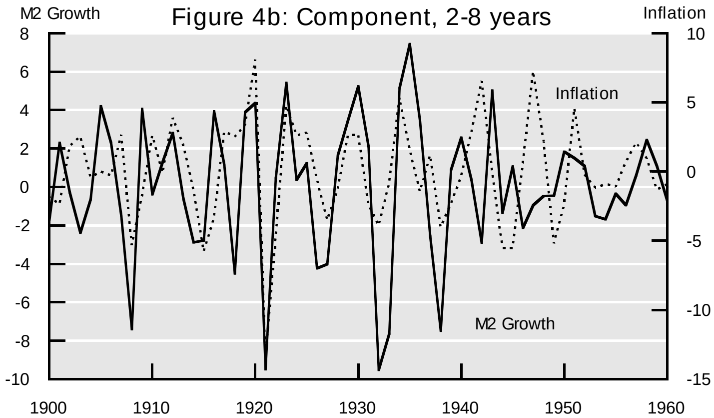


Figure 4c: Component, 8-20 years

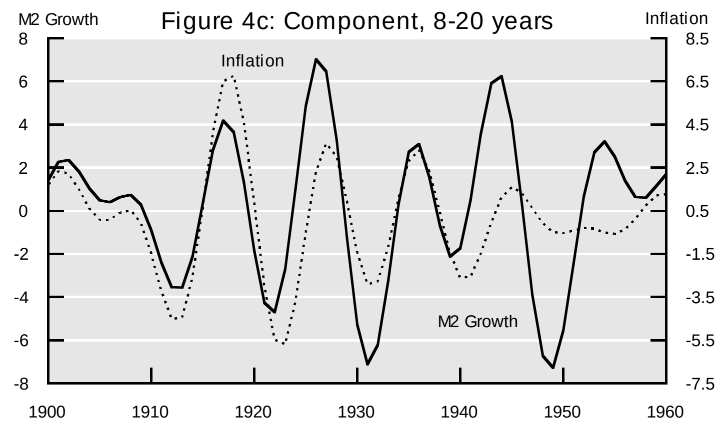


Figure 4d: Component, 20-40 years

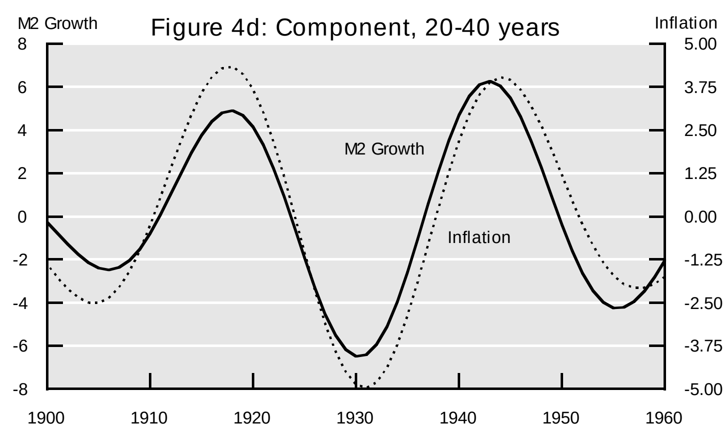


Figure 5a: M2 Growth and Inflation

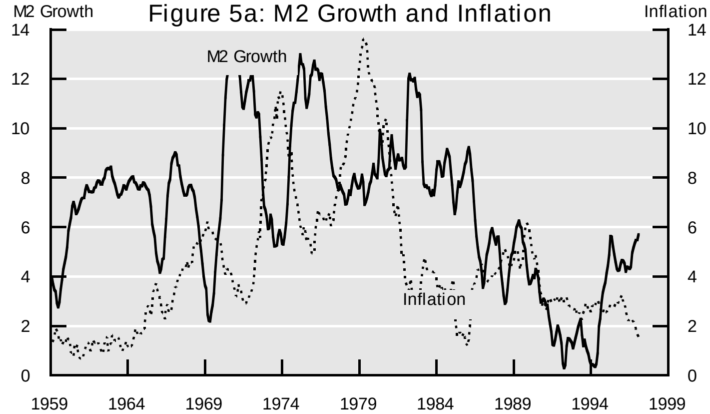


Figure 5b: Component, 1.5-8 years

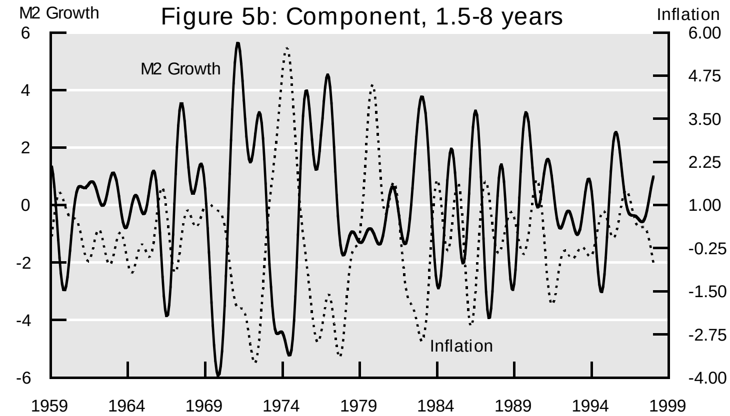


Figure 5c: Component, 8-20 years

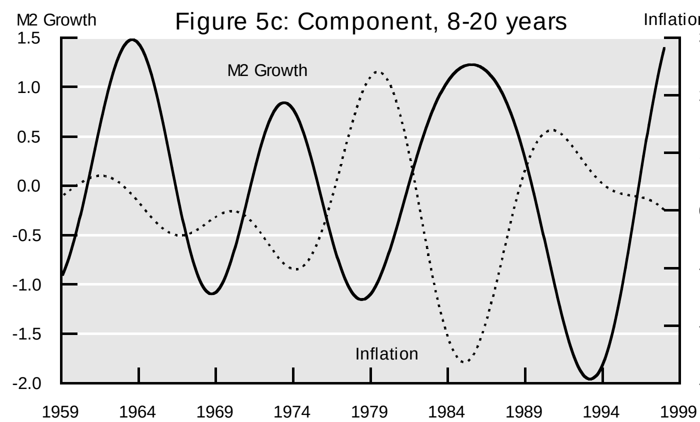


Figure 5d: Component, 20-40 years

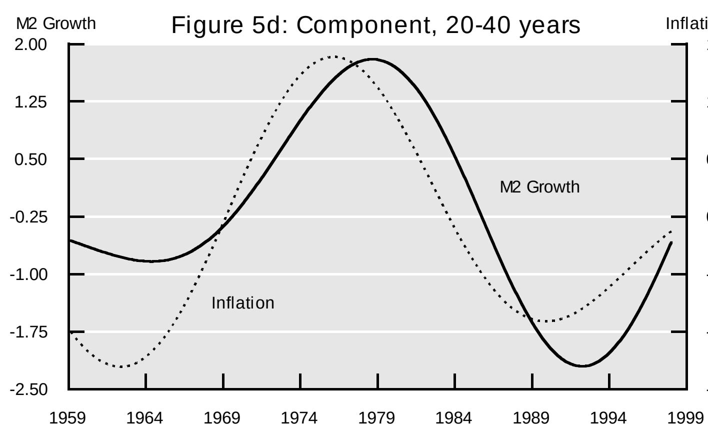
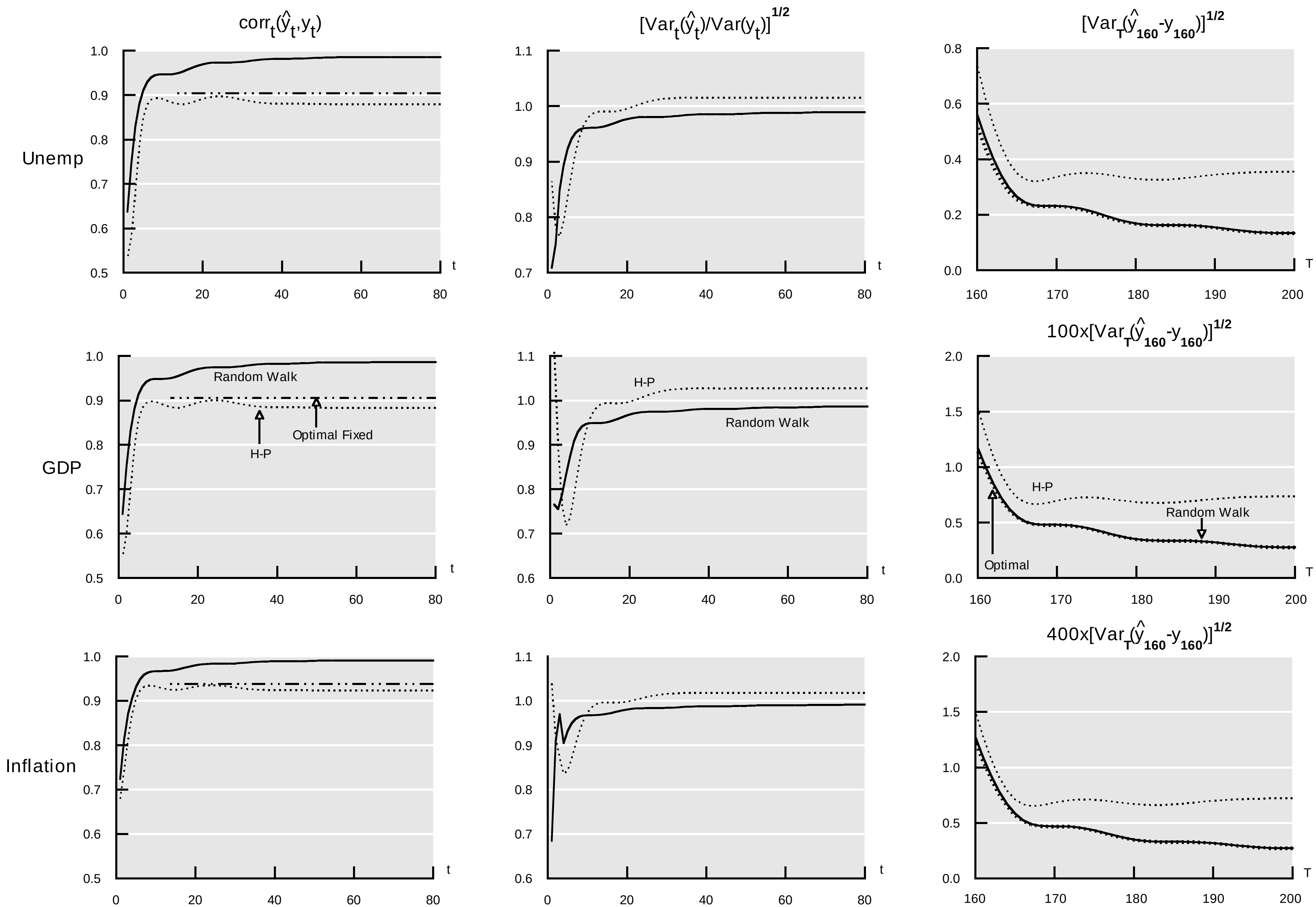
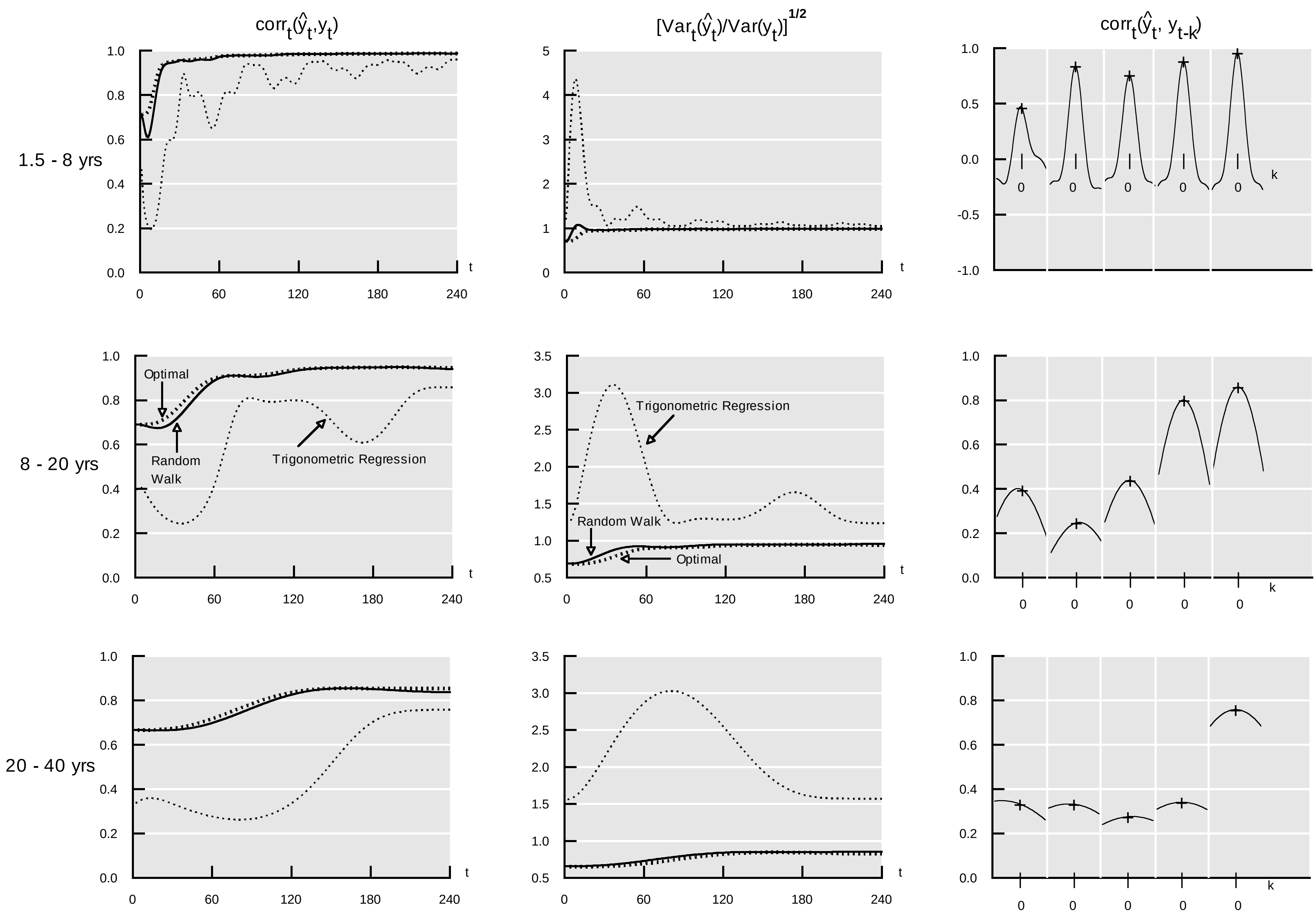


Figure 6: Comparing Filter Approximations with the Hodrick-Prescott Filter



The panels in each column contain curves differentiated according to the underlying procedure used to compute $\hat{y}_t$. The band pass filter approximations extract the frequency band between 2 quarters and 8 years. The Hodrick-Prescott filter has the parameter lambda set to 1600. The time series representations are the quarterly models described in Table 1. While the sample size is T = 160, the statistics are shown only for t = 1, 2, ..., 80 since the statistics are symmetric across the first and second halves of the sample. The third column contains the standard deviation of $\hat{y}_t - y_t$ for t = 160 and T equals 160, 161, ..., 200.

Figure 7: Comparing the Random Walk Filter with Trigonometric Regressions - Monthly U.S. Inflation Case



The panels in the first two columns contain curves differentiated according to the underlying procedure used to compute $\hat{y}_t$. While the sample size is T = 480, the statistics are shown only for t = 1, 2, ..., 240 since the statistics are symmetric across the first and second halves of the sample. The third column displays five correlation functions, $corr_t(\hat{y}_t, y_{t-k})$, associated with the Trigonometric Regressions for t = 1, 31, 61, 121, 240 (1/12, 2 1/2, 5, 10, 20 years). Each correlation function is displayed for k = -24, -23, ..., 0, ..., 23, 24. The plus on the correlation function indicates the location of k = 0