# The Band Pass Filter*

Lawrence J. Christiano † and Terry J. Fitzgerald ‡

July, 1999

**Abstract**

The 'ideal' band pass filter can be used to isolate the component of a time series that lies within a particular band of frequencies. However, applying this filter requires a dataset of infinite length. In practice, some sort of approximation is needed. Using projections, we derive approximations that are optimal when the time series representations underlying the raw data have a unit root, or are stationary about a trend. We identify one approximation which, though it is only optimal for one particular time series representation, nevertheless works well for standard macroeconomic time series.

To illustrate the use of this approximation, we use it to characterize the change in the nature of the Phillips curve and the money-inflation relation before and after the 1960s. We find that there is surprisingly little change in the Phillips curve and substantial change in money growth-inflation relation.

Keywords: Band pass filter; projection; time series; frequency domain; unit roots; bootstrap; business cycle; Phillips curve.

*We thank Jeff Schwarz for his 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.

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

‡Federal Reserve Bank of Cleveland.

## 1. Introduction

Economists have long been interested in the different frequency components of the 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 components of the data. 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, and the proposition that money growth and inflation are highly correlated in the long run, and less correlated in the short run. As a final example, 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 is part of a trend (i.e., part of the lower frequency component of the data) or is more transitory (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. An advantage of this theory, relative to other perspectives on decomposing time series, is that it does not require a commitment to any particular statistical model of the data. Instead, it relies on the Spectral Representation Theorem, according to which *any* time series within a broad class can be decomposed into different frequency components. 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, assess alternative approaches developed in the literature and provide empirical illustrations. 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:

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

where the expectation operator is evaluated using the time series properties of $x_t$. 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 $t$. We derive closed form formulas for the filter weights in these projections.

We illustrate the use of $\hat{y}_t$ in two empirical applications: one concerns the relationship between inflation and unemployment and the other, the relationship between money growth and inflation. We use the filtering technology developed here to characterize the change in the dynamics of these variables before and after 1960. A bootstrap procedure is applied for conducting the relevant statistical inference.

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 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.[1] 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 is easy to implement, and is described as follows. To isolate the component of $x_t$ with period of oscillation between $p_l$ and $p_u$, where $2 \leq p_l < p_u < \infty$, our recommended approximation of $y_t$, $\hat{y}_t$, is computed as follows:[2]

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

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.[3] 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.

[1] 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 5 below.

[2] 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.

[3] 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

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

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)[4]. The 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. As we discuss below, the need for real time estimates of $y_t$ arises in a macroeconomic stabilization context.

Note from (1.2) that our recommended filter varies with time, and is not symmetric in terms of future and past $x_t$'s. It is easy to adjust our recommended filter weights to impose stationarity and symmetry, if these features were 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.[5] 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 renders stationary time series which have up to two unit roots, or which have a polynomial trend up to the power of two.

It is important to emphasize a caveat regarding our recommended filter, (1.2)-(1.3). That filter does not closely approximate the optimal filter in *all* circumstances. To illustrate this, we display an example in which the first difference of the data displays substantial negative autocorrelation, and our recommended filter does not work well. For cases in which the appropriateness of our recommended filter is questionable, it makes sense to estimate the time series representation of the data to be filtered, and then compute, using the formulas we provide, the optimal filter based on the estimated time series representation.[6] Our formulas work for a fairly large class of time series models. Still, the class can be extended even further by applying the kind of algebraic manipulations we do in the appendix.

The outline of the paper is as follows. In section two we describe precisely the component of the

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

[5]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.

[6]Software for computing the filters in MATLAB is available from the authors on request. 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, any other formula in this paper can also be implemented simply by overriding the default.

data we seek to identify. We then discuss how we estimate it in a finite data set. Section 3 does a quantitative evaluation of alternative approaches to this estimation problem. Section 4 relates our analysis to the relevant literature. We stress in particular, the important papers by Baxter and King (1999) and Hodrick and Prescott (1997). Section 5 presents two empirical applications to illustrate the use of the filtering technology analyzed here. These examples complement the ones presented in Baxter (1994), Baxter and King (1999), Hornstein (1998), King and Watson (1994), King, Stock and Watson (1995), and Stock and Watson (1998). Section 6 concludes.

## 2. The Problem and Its Solution

Our analysis accommodates 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.[7] As we will see, the latter is actually only necessary when we consider asymmetric filters. Any sampling uncertainty in the parameters needed obtain $x_t$ from the raw data is ignored in our analysis.

We begin this section by defining precisely the object that we seek: the component of $x_t$ that lies in a particular frequency range. We then explain why this is difficult to compute directly in a finite data set and that some sort of approximation is necessary. We describe a method for obtaining an optimal approximation. Several examples are presented which convey the intuition about the formulas we develop to implement this.

### 2.1. The Ideal Band Pass Filter

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

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

[7] 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, $\mu$ 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.

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 $(-\pi,\pi)$. Here, $0<a\leq b\leq\pi$. It is well known (see, for example, Sargent (1987, p. 259)), that,

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

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

$$B(L)=\sum_{j=-\infty}^{\infty}B_jL^j,\ L^lx_t\equiv x_{t-l},$$

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 $x_t$.[8] 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, e.g., 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

[8] As already noted, we want to consider not just cases in which $x_t$ is covariance stationary, but also cases in which it has a unit root. In the unit root case, the covariance function and hence, the spectral density of $x_t$, are not well defined. This creates a problem of how to interpret the notion that (2.1) represents an orthogonal decomposition. We do so as follows. Let $x_t(\lambda)$ for $|\lambda|<1$ be a covariance stationary stochastic process in which the unit root of $x_t$ is replaced by $\lambda$. Denote the spectral density of this process, which *is* well defined, by $f_{x,\lambda}(\omega)$. Then, (see Sargent (1987, p. 268)) the cross-spectrum between $y_t$ and $\tilde{x}_t$, $g_{y,\tilde{x}}(\omega;\lambda)$, is

$$g_{y,\tilde{x}}(\omega;\lambda)=B(e^{-i\omega})\left[1-B(e^{i\omega})\right]f_{x,\lambda}(\omega),\ |\lambda|<1.$$

By the definition of $B$ (see (2.3)), this cross spectrum is zero for all $\omega$, for each $\lambda$. Since the covariance between $y$ and $\tilde{x}$ at any lag is just the suitably weighted integral of this spectrum, it follows that $y_t$ and $\tilde{x}_t$ are uncorrelated at all leads and lags, for each $\lambda$. In the unit root case, our interpretation of (2.1) as an orthogonal decomposition reflects that we define

$$g_{y\tilde{x}}(\omega;1)=\lim_{\lambda\to1}g_{y\tilde{x}}(\omega;\lambda)=0,$$

for all $\omega\in(-\pi,\pi)$.

years).[9] These differ noticeably from $B(e^{-iw})$.

### 2.2. Optimal Approximations

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

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

Presumably, the solution to this problem would be different if instead of having the true second moment properties of $\{x_t\}$, we had to rely on sample estimates. Studying the solution to (2.6) in this latter case is beyond the scope of this paper. In any case, we report results below which suggest that in practice, reasonable approximations to the solution can be obtained without knowing the details of the time series representation of $x_t$.

To discuss the solution to (2.6), it is useful to first define the following filter:

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

The strategy for estimating $y_1$, $y_2, ..., y_T$ described above in effect uses a different filter for each $t$. The filters differ according to the values of $p$ and $f$, which vary with $t$.

Alternative strategies for estimating the $y_t$'s impose various restrictions. For example, one might impose stationarity, *i.e.*, constancy of $p$ and $f$, and/or symmetry, *i.e.*, $p = f$, on the sequence of

[9] The figure reports $B_0 + 2\sum_{j=1}^{p} B_j \cos(\omega j)$ for various values of $p$.

projection problems defined above. A feature of these alternatives is that they do not use all the available observations on $x_t$ to estimate the $y_t$'s. There may be advantages to restricting one's attention to stationary and/or symmetric filters. For example, stationarity may have econometric advantages. Symmetry has the advantage of ensuring that there is no phase shift between $\hat{y}_t$ and $y_t$.[10] Thus, there is a trade-off between stationarity and symmetry and obtaining the best possible estimate of $y_t$ in the sense of (2.6). One of our objectives is to quantify the severity of this trade-off in settings that are of practical interest.

For given $t$, we can obtain a closed form solution to the projection problem, (2.4), by formulating it in the frequency domain:

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

where $f_x(\omega)$ is the spectral density of $x_t$.[11] This formulation of the problem emphasizes that the solution to the projection problem, (2.4), depends on the time series properties of the data being filtered. This is true, despite the fact that the ideal band pass filter is not dependent on the time series properties of the data.

### 2.3. Time Series Representations

We consider the following class of time series representations for $x_t$:

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

so that

$$f_x(\omega) = \frac{\theta(e^{-i\omega})\theta(e^{i\omega})}{(1-e^{-i\omega})(1-e^{i\omega})}. \tag{2.9}$$

[10] This observation follows from standard results. The phase-shift between $\hat{y}_t$ and $y_t$ is determined by the phase of the $z-$transform of the cross-covariance function between $\hat{y}_t$ and $y_t$, $g_{\hat{y}y}(z)$. This is given by:

$$g_{\hat{y}y}(z) = \hat{B}^{p,f}(z)B(z^{-1})f_x(z),$$

where $f_x(z)$ is the $z-$transform of the covariance function of the data being filtered. The phase in $g_{\hat{y}y}(e^{-i\omega})$ is related to the complex part of $g_{\hat{y}y}(e^{-i\omega})$. Given that $B(e^{-i\omega})$ and $f_x(e^{-i\omega})$ are real, this is zero if, and only if, $\hat{B}^{p,f}(e^{-i\omega})$ is real. But, this is equivalent to the requirement, $\hat{B}^{p,f}(e^{-i\omega}) = \hat{B}^{p,f}(e^{i\omega})$, i.e., that $\hat{B}_j^{p,f} = \hat{B}_{-j}^{p,f}$.

[11] This formula corresponds to one derived under similar circumstances by Sims (1972). Sims (1972) posits a situation in which some variable, $y_t$, is related to $x_t$ by a distributed lag, $y_t = B(L)x_t + \varepsilon_t$, while the econometrician imposes that the lag structure has the form, $\hat{B}^{p,f}(L)$. He assumes that the restrictions imposed by the econometrician on $\hat{B}^{p,f}(L)$ exclude the true distributed lag, $B(L)$. Sims showed that, in population, the econometrician's estimate of $\hat{B}^{p,f}(L)$, obtained by a least squares regression of $y_t$ on $x_t$'s, solves (2.7).

We place a simple structure on $\theta(L)$, by assuming it is a finite-ordered polynomial[12]:

$$\theta(z) = \theta_0 + \theta_1 z + ... + \theta_q z^q,\ q \geq 0. \tag{2.10}$$

We suppose that $p$ and $f$ are large by comparison with $q$.[13] Specifically,

$$p \geq 0,\ f \geq 0,\ p + f > 0,\ p + f \geq 2q.$$

Define

$$g(z) = \theta(z)\theta(z^{-1}) = c_0 + c_1(z + z^{-1}) + ... + c_q(z^q + z^{-q}), \tag{2.11}$$

with $c_{-\tau} = c_\tau$ for all $\tau$, and $c_\tau = 0$ for $\tau > q$.

The case where $x_t$ is difference stationary corresponds to $\theta(1) \neq 0$. The covariance stationary case corresponds to $\theta(1) = 0$, when the time series representation of $x_t$ is $x_t = [\theta(L)/(1-L)]\,\varepsilon_t = \tilde{\theta}(L)\varepsilon_t$, where $\tilde{\theta}(L)$ is a finite-ordered polynomial in $L$.

Formulas for the solution to (2.7) when $x_t$ has the time series representation, (2.8), are derived in Appendix A. The solution when $x_t$ is a random walk (i.e., $\theta(L) \equiv 1$) was presented in the introduction.[14] Other special cases are discussed in the next subsection.

### 2.4. The Optimal Approximation and $f_x$

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 $\omega$. 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

[12]The case where $(1-e^{-i\omega})(1-e^{i\omega})f_x(\omega)$ is a rational polynomial is a conceptually straightforward, though algebraically intensive, extension of what we do here. Our analysis suggests that the extension to the rational polynomial case may not be necessary to achieve good results in practice.

[13]This is only imposed to simplify the projection formulas in the Appendix. The requirement could be dropped by a suitable extension of the formulas derived there.

[14]The random walk case is simple, and can be established with the following 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. We can use the assumed time series model for $x_t$ to forecast and backcast these missing observations based on the actual data. We obtain $\hat{y}_t$ by incorporating the actual data and the estimates for the missing observations into the formula for $y_t$. In the random walk case, the formulas for the missing observations are particularly simple: for $j < t - T$, $P[x_{t-j}|x_1, ..., x_T] = x_T$ and for $j > t - 1$, $P[x_{t-j}|x_1, ..., x_T] = x_1$. With these observations, it is easy to verify (1.2)-(1.4), where $\tilde{B}_{T-t}$ is the sum of the $B_j$'s over all $j \geq T - t$ and $\tilde{B}_{t-1}$ is the sum of the $B_j$'s over all $j \geq t - 1$. This time domain strategy for solving our problem follows the one implement by Stock and Watson (1998) in a business cycle context and by Geweke (1978) and Wallis (1981) in a seasonal adjustment context.

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 $\omega$ where $f_x(\omega)$ is large and places less emphasis on regions where $f_x$ is relatively small. We illustrate this principle using four examples. The four examples tilt the graph of $f_x(\omega)$, $\omega \in (0, \pi)$ in various ways and the effects on the $\hat{B}_j^{p,f}$'s are displayed.

In what we call the '$IID$ case', $\theta(z) = 1-z$, so that $f_x(\omega) = 1$ for all $\omega$. In this case, the optimization criterion assigns equal weight to all frequencies. In the appendix, it is shown that the $\hat{B}_j^{p,f}$'s that solve (2.7) in this case are given by:

$$IID \text{ case: } \hat{B}_j^{p,f} = B_j, \text{ for } j = -f, -f+1, ..., 0, ...., p. \tag{2.12}$$

For $p = f = 12$, $p_u = 24$, $p_l = 4$ (see (1.3)), $\hat{B}^{p,f}(e^{-i\omega})$ and $B(e^{-i\omega})$ are displayed in Figure 2a.[15]

In our other three examples, $x_t$ has a single unit root. That (2.12) is *not* the solution to (2.7) in this case can be seen using a simple argument by contradiction. Thus, suppose (2.12) *is* the solution to (2.7) in the unit root case. Since $B(1) = 0$, it follows that $y_t$ is covariance stationary. But, since most likely $\hat{B}^{p,f}(1) \neq 0$, $\hat{y}_t$ will not be covariance stationary.[16] This implies that $\hat{y}_t$ and, hence, $y_t - \hat{y}_t$, has infinite variance. This variance can be made finite by just altering the filter weights so that $\hat{B}^{p,f}(1) = 0$. This is a contradiction to the assumption that the filter weights solve (2.7).

The preceding considerations suggest that a necessary condition for $\hat{B}^{p,f}(L)$ to solve (2.7) when $x_t$ has a unit root is that $\hat{B}^{p,f}(1) = 0$. We impose this in the formulas derived in the appendix.

To illustrate these formulas, we analyze the following additional examples. The case where $x_t$ is a random walk is discussed in the introduction. In the 'Near $IID$ case',

$$\theta(z) = 1 - (1-\eta)z, \ \eta > 0, \ \eta \text{ small}. \tag{2.13}$$

[15] That (2.12) is the solution to (2.7) in the $IID$ case can be verified trivially using the time domain strategy in footnote 12. The frequency domain strategy is also straightforward in this case. The first order condition of (2.7) associated with the optimal choice of $\hat{B}_j^{p,f}$ is

$$\int_{-\pi}^{\pi} B(e^{-i\omega})e^{-i\omega j}d\omega = \int_{-\pi}^{\pi} \hat{B}^{p,f}(e^{-i\omega})e^{-i\omega j}d\omega,$$

for $j = -f, -f+1, ..., 0, ..., p$. Equation (2.12) follows by evaluating these integrals using the well known fact

$$\int_{-\pi}^{\pi} e^{-i\omega h}d\omega = \begin{cases} 2\pi, \ h = 0 \\ 0, \ h \neq 0 \end{cases}.$$

[16] We have not ruled out the possibility that there is *some* combination of $p$ and $f$ for which $\hat{B}^{p,f}(1) = 0$, when the $\hat{B}_j^{p,f}$'s are constructed as in (2.12).

Note that if $\eta = 0$, then this reduces to the $IID$ case just discussed. However, with $\eta$ positive, no matter how small, $f_x(\omega)$ diverges to infinity for $\omega \to 0$. Outside of a small region around $\omega = 0$, the spectrum of $x_t$ in the Near $IID$ case looks like the one in the $IID$ case: constant at unity. In this case, the solution to (2.7), for $p = f$, is

$$\text{Near } IID \text{ case: } \hat{B}_j^{p,f} = B_j + \frac{\Delta}{1+2p}, \; j = 0, \pm 1, ..., \pm p, \tag{2.14}$$

where

$$\Delta = -\left[B_0 + 2\sum_{j=1}^{p} B_j\right],$$

and the $B_j$'s are defined in (1.3).[17] In this case, it is optimal to truncate the ideal band pass filter and then adjust all the weights by a constant to ensure $\hat{B}^{p,f}(1) = 0$. This differs from the Random Walk case discussed in the introduction. In the latter, optimality dictates truncating the ideal band pass filter, and then only adjusting the highest order terms to ensure $\hat{B}^{p,f}(1) = 0$.

Figure 2 allows us to assess the impact on the optimal filter approximations of the three specifications of $f_x$ already described and of the fourth one that will be described shortly. We begin by comparing $\hat{B}^{12,12}(e^{-i\omega})$ in the Near $IID$ and $IID$ cases (see Figures 2a and 2b). We can think of the Near $IID$ case as adapting $\hat{B}^{12,12}(e^{-i\omega})$ from the $IID$ case by shifting it up towards zero in the neighborhood of $\omega = 0$. This improvement is achieved at the cost of making $\hat{B}^{12,12}(e^{-i\omega})$ diverge from $B(e^{-i\omega})$ for values of $\omega$ slightly above zero. In the $IID$ case, when $f_x(\omega) = 1$ for all $\omega$, this cost outweighs the improvement. In the Near $IID$ case, when $f_x(0) = \infty$, the situation is reversed. For frequencies well above $\omega = 0$, $\hat{B}^{12,12}(e^{-i\omega})$ for the two cases essentially coincide. The example shows that shifting power towards one frequency range causes the optimal filter to become more accurate in that range, at the expense of doing poorly in another range.

We now compare $\hat{B}^{12,12}(e^{-i\omega})$ in the Near $IID$ and Random Walk cases (see Figure 2b). The spectrum of $x_t$ in the Random Walk case resembles the one in the Near $IID$ case in that it diverges to infinity for $\omega \to 0$. However, it declines more slowly as $\omega$ increases above zero.[18] As a result,

[17] Verifying this is only a little more complicated than the problem of verifying (2.12), discussed in footnote 12. Consider the case, $p = f$, so that the solution to (2.7) is symmetric with $\hat{B}_j^{p,f} = \hat{B}_{-j}^{p,f}$ for $j = 0, ..., p$ (this symmetry property is an unsurprising feature of the $p = f$ case, and is established rigorously in Appendix A.) The $p+1$ unknown $\hat{B}_j^{p,f}$'s may be found by solving the condition, $\hat{B}^{p,f}(1) = 0$, and the $p$ first order conditions associated with (2.7):

$$\int_{-\pi}^{\pi} B(e^{-i\omega})e^{-i\omega j} f_x(\omega)d\omega = \int_{-\pi}^{\pi} \hat{B}^{p,f}(e^{-i\omega})e^{-i\omega j} f_x(\omega)d\omega,$$

for $j = 0, ..., p-1$. That these indeed are the first order conditions can be verified with some algebra, but in any case is established carefully in Appendix A. It is easily verified that the $\hat{B}_j^{p,f}$'s in (2.14) satisfy these conditions.

[18] In the random walk case, $f_x(\omega) = 1/\left[2(1-\cos(\omega))\right]$.

by comparison with the $IID$ case, the Random Walk case assigns relatively heavy weight to the low frequencies and relatively low weight to the others. The effects on the optimal filter are evident in Figure 2b: the Random Walk case does better than the Near $IID$ case in the region of $\omega = 0$, but at the cost of performing relatively poorly at the higher frequencies.

A final case, the Persistent case, is studied in Figure 2c. There,

$$\theta(z) = 1 + z + ... + z^{13}. \tag{2.15}$$

Taking the Random Walk case as a benchmark, this case goes to the opposite extreme relative to the Near $IID$ case. The Persistent case tilts power in $x_t$ even more towards frequency zero and away from the higher frequencies. It almost corresponds to the case in which it is not the first difference, but the second difference of $x_t$, that is a white noise. We compare $\hat{B}^{12,12}(e^{-i\omega})$ for these two cases in Figure 2c. Note how $\hat{B}^{12,12}(e^{-i\omega})$ for the Persistent case does better in a neighborhood of $\omega = 0$ than the Random Walk case. This is achieved in exchange for a very pronounced deterioration in performance in the higher frequencies.

## 3. Quantitative Assessment of Various Filters

A key purpose of this section is to explore the quantitative importance of three factors in the solution to (2.7). We examine the role of asymmetry and time nonstationarity of the $\hat{B}^{p,f}$'s. We also assess the importance of knowing the details of the time series representation of $x_t$. Clearly, the importance of these three factors depends on the actual time series properties of $x_t$. To make the analysis interesting, we use difference stationary and trend stationary time series representations that fit standard US macroeconomic data.

Our findings are as follows. First, in minimizing (1.1), the biggest gains come from allowing the filter weights to vary over time. These gains reflect that allowing nonstationarity substantially increases the amount of information in $x$ that can be used in estimating $\hat{y}_t$. Second, allowing the filter weights to be asymmetric further increases the amount of information in $x$ that can be used in constructing $\hat{y}_t$, though by a lesser amount. So, we find that nonstationarity and asymmetry are valuable in minimizing the distance metric. It turns out that the cost of these features is relatively minor. We display evidence that the degree of asymmetry and nonstationary in the optimally filtered data is quantitatively small.[19] Finally, we find that there is relatively little gain in knowing the precise

[19]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

details of the time series representation generating the $x_t$'s. In particular, 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 are the findings that lead us to the view that an adequate, though not optimal, procedure 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.

The second purpose of this section is to evaluate other filtering approaches used in the literature. These include the band-pass filtering approach recommended by Baxter and King (1999) and the filter proposed by Hodrick and Prescott (1997) (HP). 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 procedure the Trigonometric Regression procedure.

The first subsection describes the statistics that we use in our analysis. The next subsection studies the properties of the solution to (2.7). The third subsection compares these properties with those of alternative filters that are used in the literature. The second and third subsections are based on unit root representations of the data. The final subsection considers robustness of our results by considering trend stationary representations instead.

### 3.1. Some Useful Statistics

We evaluate the filtering procedures by studying the dynamic correlations, $corr_t(\hat{y}_t, y_{t-\tau})$, and the relative standard deviations, $(var_t(\hat{y}_t)/var(y_t))^{1/2}$, for various $t$.[20] This section explains why we look at these statistics, and how we compute them.

These statistics have three features which make them useful for our purposes. First, when the $\hat{y}_t'$s are based on variable lag filters, they are non-trivial functions of time. We use the quantitative magnitude of this variation to assess the degree of non-stationarity in $\hat{y}_t$. Of course, statistics based on $y_t$ alone, or statistics based on $\hat{y}_t$ when fixed lag filters are used, are invariant with respect to $t$, given the data generating mechanisms that we consider. Second, for filters that solve a projection problem,

nonstationarity in this filter is quantitatively small.

[20]Following convention, a statistic for a particular date is defined across the ensemble of realizations for the underlying stochastic process at that date.

the projection criterion, $E[(y_t - \hat{y}_t)^2|\text{-}_t]$, is a simple function of the correlation between $y_t$ and $\hat{y}_t$:[21]

$$E[(y_t - \hat{y}_t)^2|\text{-}_t] = \left[1 - corr_t(\hat{y}_t, y_t)^2\right] var(y_t).$$

Here, $\text{-}_t$ is the relevant information set in the projection.[22] This relation indicates that we are free to think of the projection criterion in terms of $corr_t(\hat{y}_t, y_t)$, an object that we find easier to interpret than $E[(y_t - \hat{y}_t)^2|\text{-}_t]$ itself. In evaluating our results, it is useful to be aware of two other results that are true when $\hat{y}_t$ solves a projection problem:

$$corr_t(\hat{y}_t, y_t) = \left[\frac{var_t(\hat{y}_t)}{var(y_t)}\right]^{1/2}, \; var_t(\hat{y}_t) \leq var(y_t).$$

Third, one of our concerns is to quantify the phase shifts that exist between $y_t$ and $\hat{y}_t$ when the latter is computed using non-symmetric filters. As explained below, the dynamic correlations between $y_t$ and $\hat{y}_t$ are useful for this.

We now briefly discuss how we compute these statistics. We evaluate the variance of $y_t$ using the following relation:[23]

$$var(y_t) = \frac{1}{\pi}\int_a^b f_x(\omega)d\omega.$$

[21] This is a standard result. To see it, note that by the orthogonality property of projections, we can write $y_t = \hat{y}_t + \varepsilon_t$, where $\varepsilon_t$ is orthogonal to $\text{-}_t$ and, hence, $\hat{y}_t$ itself. Then,

$$var_t(y_t - \hat{y}_t) = var(y_t) - var_t(\hat{y}_t) = var(y_t)\left[1 - \frac{var_t(\hat{y}_t)}{var(y_t)}\right] = var(y_t)\left[1 - \rho_t(0)^2\right].$$

The last equality follows from the following observation:

$$corr_t(\hat{y}_t, y_t) = \frac{cov_t(\hat{y}_t, y_t)}{[var_t(\hat{y}_t)var_t(y_t)]^{1/2}} = \frac{var_t(\hat{y}_t)}{[var_t(\hat{y}_t)var_t(y_t)]^{1/2}} = \left(\frac{Var_t(\hat{y}_t)}{Var_t(y_t)}\right)^{1/2}.$$

[22] That is, $\text{-}_t = \{x_1, ..., x_T\}$ for all $t$ in the case of Optimal, $x_{t-12}, ..., x_0, ..., x_{t+12}$ in the case of Optimal Fixed, $p = f = 12$, etc.

[23] In practice, we approximate an integral like this by the Riemann sum:

$$\frac{1}{\pi}\int_a^b f_x(\omega)d\omega \approx \frac{b-a}{\pi}\frac{1}{N}\sum_{j=1}^{N} f_x(\omega_j) = \frac{b-a}{\pi}\frac{1}{N}\sum_{j=1}^{N}\frac{\theta(e^{-i\omega_j})\theta(e^{i\omega_j})}{2\left(1-\cos(\omega_j)\right)}, \; \omega_j = a + \frac{b-a}{N}j, \; j = 1, ..., N$$

To guarantee that $var(y_t)$ is accurately computed, and because it only needs to be computed a few times, we set $N$ to a very high value, $10,000$.

We evaluate the variance of $\hat{y}_t$ using the following expression:[24]

$$var_t(\hat{y}_t) = \frac{1}{\pi}\int_0^\pi |\hat{B}^{p,f}(e^{-i\omega})|^2 f_x(\omega)d\omega.$$

Here, *var* is a non-trivial function of time when $p$ and $f$ vary with $t$.

The dynamic correlations between $y_t$ and $\hat{y}_t$ are computed using the following the expression:

$$corr_t(\hat{y}_t, y_{t-\tau}) = \frac{cov_t(\hat{y}_t, y_{t-\tau})}{\sqrt{var_t(\hat{y}_t)var(y_t)}} = \frac{\frac{1}{\pi}\int_a^b \text{real}\left[r^{p,f}(e^{-i\omega})e^{i[\theta(\omega;p,f)+\omega k]}f_x(\omega)\right]d\omega}{\sqrt{var_t(\hat{y}_t)var(y_t)}}.$$

For convenience, we have written $\hat{B}^{p,f}$ in polar form:

$$\hat{B}^{p,f}(e^{-i\omega}) = r^{p,f}(e^{-i\omega})e^{i\theta(\omega;p,f)},$$

where $r^{p,f}(e^{-i\omega})$ is real and the function $\theta(\omega; p, f)$ is the phase of $\hat{B}^{p,f}(e^{-i\omega})$ (see Sargent (1987, chapter XI).[25] We use the location of the peak of this correlation function to determine whether there is a phase shift between $y_t$ and $\hat{y}_t$. To see the rationale for this, consider the following special cases. When $\hat{B}^{p,f}$ is symmetric, so that $\theta(\omega; p, f) \equiv 0$ and there is no phase shift between $y_t$ and $\hat{y}_t$, then it is easily verified that $corr_t(\hat{y}_t, y_{t-\tau})$ is symmetric about $\tau = 0$.[26] When $\hat{B}^{p,f}$ is not symmetric, so that there is a non-zero phase shift between $y_t$ and $\hat{y}_t$, then this symmetry property of the correlation function fails. Consider, for example, the extreme case in which $\hat{B}^{p,f}(L) = L^h$ and $h \neq 0$, so that $r^{p,f}(e^{-i\omega}) = 1$ and $\theta(\omega; p, f) = -h\omega$. In this case, it is easy to see that $corr_t(\hat{y}_t, y_{t-\tau})$ is symmetric about $\tau = h$. In general, when symmetry fails the phase of $\hat{B}^{p,f}(e^{-i\omega})$ does not satisfy this proportionality property,

[24]The integrand in the expression for $var_t(\hat{y}_t)$ exhibits substantial variation, particularly in the region about $\omega = a$ and $\omega = b$. To ensure that these integrals are computed accurately, we divided the interval, $\omega \in (0, \pi)$ into five regions: $(0, a - \Delta_1)$, $(a - \Delta_1, a + \Delta_1)$, $(a + \Delta_1, b - \Delta_2)$, $(b - \Delta_2, b + \Delta_2)$, $(b + \Delta_2, \pi)$, with $\Delta_1$, $\Delta_2 > 0$. The integral in each region was evaluated using the Riemann approximation formula described in the previous footnote, with $N = 100$ in each interval.

[25]To verify our covariance formula, note first from the results in Sargent (1987, chapter XI), that

$$cov(\hat{y}_t, y_{t-k}) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \hat{B}^{p,f}(e^{-i\omega})B(e^{i\omega})e^{i\omega k}f_x(\omega)d\omega.$$

Then, note that the integral from 0 to $\pi$ is the complex conjugate of the integral from $-\pi$ to 0. Finally, we have taken into account of the definition of $B$ to limit the range of integration.

[26]The connection between the symmetry of $\hat{B}^{p,f}$ and its phase was pointed out in a previous footnote. To see why symmetry of $\hat{B}^{p,f}$ also implies symmetry of the covariance function, note that when $\theta \equiv 0$, the covariance can be written

$$\frac{1}{\pi}\int_0^\pi B(e^{-i\omega})r^{p,f}(e^{-i\omega})e^{i\omega k}f_x(\omega)d\omega$$

which is the covariance function of a stochastic process whose spectral density is $B(e^{-i\omega})r^{p,f}(e^{-i\omega})f_x(\omega)$ . The result follows from the symmetry about $k = 0$ of the covariance function of a stochastic process.

and so the asymmetry in $\hat{B}^{p,f}(e^{-i\omega})$ is manifested in more exotic forms of asymmetry in $\rho_t(\tau)$.

### 3.2. Difference Stationary Data Generating Mechanisms

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. The estimation results are presented in Table 1. With one exception, the Box-Pierce statistic ('Q') indicates little evidence of serial correlation in the errors, as indicated by the $p-$values in the table. In the exceptional case, monthly inflation, the addition of higher order lags does not raise the $p-$value.

The spectral density of $(1-L)x_t$ for each model is graphed in Figure 3. Note that the spectral densities of output, interest rates and unemployment are very similar. They exhibit relatively high power in the low frequencies. Quarterly and monthly inflation deviates from this pattern in having relatively more power in the higher frequencies.

| Table 1: Time Series Representations for Selected US Data | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $(1-L)x_t = \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \theta_3\varepsilon_{t-3} + \theta_4\varepsilon_{t-4} + \theta_5\varepsilon_{t-5}$, $E\varepsilon_t^2 = \sigma_\varepsilon^2$ | | | | | | | | |
| variable, $x_t$ | $\sigma_\varepsilon$ | $\theta_1$ | $\theta_2$ | $\theta_3$ | $\theta_4$ | $\theta_5$ | Q | Frequency |
| $\log(CPI_t/CPI_{t-1})$ | 0.0021 | -0.75 | | | | | 0.00 (36-1) | M |
| | 0.0042 | -0.23 | -0.27 | 0.32 | | | 0.11 (30-3) | Q |
| | 0.015 | 0.81 | | | | | 0.34 (9-1) | A |
| $log(GDP_t)$ | 0.0075 | 0.28 | 0.21 | 0.20 | 0.16 | | 0.20 (36-4) | M |
| | 0.0088 | 0.25 | 0.16 | 0.10 | 0.12 | | 0.09 (30-4) | Q |
| | 0.021 | 0.34 | | | | | 0.89 (9-1) | A |
| $log(Tbill_t)$ | 0.064 | 0.26 | | | | | 0.05 (36-1) | M |
| | 0.105 | 0.45 | -.16 | | | | 0.24 (30-2) | Q |
| | 0.21 | 0.58 | | | | | 0.54 (9-1) | A |
| $Unemployment_t$ | 0.18 | -0.02 | 0.19 | 0.17 | 0.20 | 0.05 | 0.021 (36-5) | M |
| | 0.27 | 0.65 | 0.48 | 0.41 | | | 0.66 (30-3) | Q |
| | 0.96 | 0.28 | | | | | 0.80 (9-1) | A |
| Note: All time series models estimated using the RATS command, boxjenk. Data cover the period 1960-1997. $Q$ denotes the $p-$value of the Box-Pierce test of the null hypothesis of zero autocorrelation in the residuals. In $(n, m)$, $n$ denotes number of autocorrelations in residuals used, and $m$ denotes degrees of freedom of chi-square test statistic. | | | | | | | | |

### 3.3. Properties of the Optimal Filter Approximation

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. The procedures we consider are listed in Table 2. 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.

| Table 2: Band Pass Filter Approximation Procedures Considered | |
|---|---|
| Name | Definition |
| Optimal | Optimal |
| Random Walk | Optimal, assuming random walk $x_t$ ((1.2)-(**??**)) |
| 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 Baxter and King (1999).

This section reports our results, based on the unit root representations in Table 1, and on the Near $IID$ ((2.13) with $\eta = .01$) and Persistent, (2.15), representations considered in the previous section. Given the evidence in Figure 3, it is not surprising that the findings for many of the time series representations in Table 1 are very similar. Therefore, of the time series representations in Table 1, we only present results for inflation. We selected this time series representation because it provides the weakest support for our contention that the Random Walk filter is close to optimal. Results based on the Persistent representation are similar to those for the other representations in Table 1, and so we do not present these. Results based on annual, monthly and quarterly data sampling intervals are also very similar. Consequently, we only present the findings based on the monthly sampling interval. We consider three frequency bands: 1.5 to 8 years, 8 to 20 years, and 20 to 40 years.

The results we present are based on two monthly time series representations: the one estimated using inflation (see Figure 4) and the $Near\ IID$ representation (Figure 5). Consider Figure 4 first. 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; $8-20$ years; and $20-40$ 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 in the estimation period for the monthly time series representations in Table 1. The second column contains results for Optimal and Random Walk alone. Here, results for Optimal

are simply repeated for convenience from the first column. As discussed above, 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 4 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 the other time series representations in Table 1.

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 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 in the tails. This tail area is fairly small (about 1.5 years) for the business cycle frequencies, but it grows for the lower frequencies (columns 1 and 2). Fifth, Random Walk seems to imply little phase shift between $\hat{y}_t$ and $y_t$ (column 3). There is essentially no phase shift, even at the lowest frequencies.

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 phase shift. However, the gains of going from a simple procedure like Random Walk to Optimal are quite small.

These findings apply to all the time series models in Table 1, as well as to the Persistent time series representation. One time series representation where these findings do not hold up is the $Near\ IID$ case, which is reported in Figure 5. In particular, the finding that Random Walk is nearly optimal and roughly stationary no longer holds (columns 1 and 2). However, the other conclusions continue to hold. For example, Optimal is still nearly stationary outside of tail areas. Also, imposing symmetry and time-stationarity on the filter imposes substantial efficiency costs, especially in the low frequencies.

### 3.4. Properties of Alternative Filter Approximations

Next we discuss three alternative band pass filter approximations that have been used in the literature. The first of these, the one due to Baxter and King (1999), is the fixed-lag, symmetric filter defined in (2.14). This filter can do no better than Optimal Fixed, discussed in the previous section. That filter

is dominated by Random Walk, which is nearly optimal for the time series models reported in Table 1. Below, we provide a detailed analysis of two other filters: the HP filter and Trigonometric Regression. Again, we find that Random Walk is nearly optimal and dominates the HP filter and Trigonometric Regression.

### 3.4.1. Hodrick-Prescott Filter

Proponents of the HP filter offer several interpretations of what it does (see section 4 below for an elaboration). The interpretation that we adopt here is that it approximates a band pass filter designed to isolate, in quarterly data, frequencies corresponding to the business cycle and higher.[27] We identify these with cycles of period 8 years and less. We display evidence which suggests that, for data like US GDP, unemployment and the CPI, our Random Walk filter is nearly optimal and outperforms the HP filter. This finding also applies when these filters are used to obtain real-time estimates of $y_t$. We stress this here because the HP filter is sometimes used in a macroeconomic stabilization context, when estimates of current $y_t$ are needed.

Our discussion in this section is divided into two parts. We first discuss 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 those observations and on the information in Column 3 because this allows us to assess the value of the filters for obtaining real-time estimates of $y_t$.

The first column in Figure 6 displays $corr_t(\hat{y}_t, y_t)$ associated with the HP filter, Random Walk, and Optimal Fixed 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. Interestingly, despite the fact that HP uses all the data, it nevertheless performs less well than Optimal Fixed. Moreover, it does notably less well than Random Walk, particularly for Unemployment and GDP. Outside of a tail area of roughly two years, $corr_t(\hat{y}_t, y_t)$ exceeds 0.95 and it is closer to 0.99 towards the middle of the data set for Random Walk. For GDP and Unemployment, this statistic based on the HP filter never exceeds 0.90.

Column 2 shows that, outside of a five year tail area, the performance of Random Walk is similar to that of the HP filter, in terms of the standard deviation of $\hat{y}_t$: Random Walk undershoots $Var(y_t)$ somewhat while HP overshoots.

Another way to compare HP filter and Random Walk focuses on the metric, (1.1), that we use to construct optimal filters. In particular, we consider $R_t$, the absolute size of the typical estimation error,

[27] The HP filter parameter, $\lambda$, is set to 1600, as is typical in applications using quarterly data.

$\hat{y}_t - y_t$ (measured by its standard deviation), to the absolute size of the typical value of $y_t$ (measured by its standard deviation): $R_t = [Var_t(\hat{y}_t - y_t)/Var(y_t)]^{1/2}$.[28] 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 $\hat{y}_t \equiv 0$. In the case of Random Walk, $R_t$ is 0.14 towards the middle of the data set and $R_t$ 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 are the basis for our conclusion that, outside of the tail area, Random Walk outperforms the HP filter.

We now turn to the implications of the results in Figure 6 for the real time performance of the filters. We do this because it is of interest in the context of stabilization policy, when current estimates of the output and unemployment gaps are used. For example, John Taylor has argued persuasively that monetary policy makers set current policy as a function of the current output gap, among other things.[29] The output gap is the log difference between actual GDP and 'potential' GDP, and the unemployment gap is the difference between actual unemployment and the 'natural' rate of unemployment. One interpretation of potential GDP and the natural rate of unemployment is that they correspond to the HP trend in actual GDP and unemployment, respectively.[30] For this reason, we now investigate how effective HP filter, Random Walk and Optimal are in estimating $y_t$ in real time.

At the outset, it should be clear that estimating the current value of $y_t$ is likely to be a difficult task. Suppose, for example, that the variable, $x_t$, has recently changed value. How is one to decide whether that change is temporary (i.e., part of $y_t$) or more persistent (i.e., part of $\tilde{x}_t$)? No doubt we can confidently answer this question with a sufficient passage of time, with more data. In the light of hindsight, it is relatively easy to determine whether a given change in a variable was just a blip or a

[28] Here,

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

As noted before, when $\hat{y}_t$ is the result of projecting $y_t$ on $x$, then $corr_t(\hat{y}_t, y_t) = Var_t(\hat{y}_t)/Var(y_t)$, so that $R_t$ reduces to $(1 - corr_t(\hat{y}_t, y_t)^2)^{1/2}$. This formula does not hold for HP filter, since it is not the result of the indicated projection. It only holds approximately for Random Walk. Although that filter *is* the result of a projection of $y_t$ on $x$, that projection incorporates a specification error by falsely assuming the data are generated by a random walk.

[29] See Taylor (1999). See also Orphanides (1999), who discusses the pitfalls of conditioning policy actions on current gap measures.

[30] For example, according to Orphanides and van Norden (1999, p. 1), 'The difference between [actual output and potential output] is commonly referred to as the *business cycle or the output gap* (italics added).' In practice, the concept of output and unemployment gaps varies among practioners and does not always coincide with the one used here: the deviation between the actual data series and its 'trend', defined as the component with fluctuations having period 8 years and longer. Alternative gap concepts are based on different notions of trend. For them, the trend corresponds to the 'nonaccelerating inflation' level of the variable: the level which, if it obtained, would produce a forecast of zero change in the rate of inflation in the near future. Gap concepts like this are fundamentally multivariate. To see how the HP filter can be adapted to correspond more closely to this alternative gap concept, see Laxton and Tetlow (1992) and St-Amant and van Norden (1997).

movement in the trend. But, in real time without the advantage of hindsight, we can expect that even our best estimates of $y_t$ will be noisy. That is the case for Random Walk and HP Filter. However, we show that the estimate based on HP filter is noisier than that based on Random Walk.

Recall the symmetry properties of the statistics in Columns 1 and 2 in Figure 6: their properties in the second half of the sample are the mirror image of their properties in the first half. As a result, the statistics at the left end-point of these two columns provide evidence on the real-time performance of the filters being considered. That is because, by symmetry, these statistics can be viewed as being based upon the estimate, $\hat{y}_T$, of 'current' $y_T$, where $T = 160$. (In the case of Random Walk, $\hat{y}_T$ is computed using the one-sided filter, (1.4).) Note that the correlation between $\hat{y}_t$ and $y_t$ is at its lowest for $t = T$ ($t = 1$) for both Random Walk and HP filter. That is, the real time correlation between the estimated and actual gaps is lower than what it is after the arrival of new data.[31] The drop is fairly substantial.

The evidence in Column 2 shows that, for the three time series representations considered, the real-time variance of $\hat{y}_t$ computed using Random Walk is less than what it is after new data arrives. In this sense, the estimate of the trend implicit in this filtering procedure (and Optimal) tends to follow the actual data more closely in real time than after the arrival of new data.[32] We do not know if this is a general property of Random Walk, true for all time series representations. Evidently, it is *not* a general property of the HP filter. With the data generating mechanisms based on GDP and inflation, we see that the real time variance of $\hat{y}_t$ is at its global *maximum* for $t = T$ and $t = 1$.[33]

We now compare Random Walk and HP filter using the $R_t$ statistic described above, for $t = T$. Using Random Walk, $R_T = 0.77$, 0.78, and 0.69 for GNP, unemployment, and inflation respectively. Note that these numbers are substantially larger than what they are for data points closer to the middle. 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. 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

[31] The evidence in Columns 1 and 2 actually do not allow us to literally infer what happens with the arrival of new data. The data set underlying the experiments in these columns are based on a *fixed* sample of length 160.

[32] In our context, the estimated trend is $d_t$, where $d_t$ equals $\hat{x}_t$ plus the drift in the data, where $\hat{x}_t = x_t - \hat{y}_t$ (recall our decomposition, (2.1).)

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

variables. We consider the standard deviation of the error , $y_t - \hat{y}_t$, for a fixed date, $t = 160$. We display this statistic for the time when $t$ is the current date and continuing as new data become available and the data set becomes correspondingly larger. These results allow us to quantify the value of hindsight when estimating $y_t$.

We study $[var_{160}(\hat{y}_{160} - y_{160})]^{1/2}$ for $T = 160, 161, ..., 200$, based on Random Walk, Optimal and HP filter. For Random Walk and Optimal, the standard deviations are based on $\hat{y}_{160} = P[y_{160}|x_1, ..., x_T]$, for $T = 160, ..., 200$. In the case of $HP$, they are based on $\hat{y}_{160}$, the $160^{th}$ observation in the HP filtered $x_1, ..., x_T$, for $T = 160, 161, ..., 200$. 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 after $t = 160$. Thereafter, further declines in the error come more slowly. Third, after initially falling with the arrival of the first two years' data, 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 a band pass filter. By contrast, both Random Walk and Optimal do. If $T$ were allowed to grow indefinitely, $[var_{160}(\hat{y}_{160} - y_{160})]^{1/2}$ 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 terms of real-time performance.[34]

In this discussion, we have emphasized the differences in the performance of the Hodrick-Prescott filter and Random Walk. We note, however, that the quantitative magnitude of the differences is not very great for some purposes. For example, in Appendix B we display business cycle statistics based on US data using various filter approximations and the HP filter, and there is little quantitative difference between them. The differences seem relatively larger when we consider real-time estimation of $y_t$. The HP filter performs very poorly on this dimension. Indeed, in the case of GDP it is useless.[35] However, even the optimal procedure seems relatively unreliable in this case.[36]

[34]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. They note how wide this range is and so it is not surprising that our estimates fall inside it.

[35]These conclusions about the real time performance of 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).

[36]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 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).

### 3.4.2. Trigonometric Regression

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

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

where

$$\begin{aligned} B_t(L)x_t &= \sum_{l=t-T}^{t-1} \left\{ \frac{2}{T} \sum_{j \in J} \cos(\omega_j l) \right\} x_{t-l},\ \text{if } \frac{T}{2} \notin J \\ &= \sum_{l=t-T}^{t-1} \left\{ \frac{2}{T} \sum_{j \in J, j \neq \frac{T}{2}} \cos(\omega_j l) + \frac{1}{T} \cos(\pi(t-l)) \cos(\pi t) \right\} x_{t-l}, \text{if } \frac{T}{2} \in J \\ t &= 1, ..., T,\ \omega_j = \frac{2\pi}{T} j. \end{aligned} \tag{3.1}$$

Here, $J$ indexes the set of frequencies we wish to isolate, and is a subset of the integers $1, ..., T/2$.[37] It is easy to see that $B_t(1) = 0$, so that $B_t(L)$ has a unit root for $t = 1, 2, ..., T$.[38] Evidently, $B_t(L)$ only has a second unit root for $t$ in the middle of the data set, when $B_t(L)$ is symmetric.[39] For this reason, it is important to drift adjust $x_t$ prior to filtering.

---

[37] 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:

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

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

[38] 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.

[39] 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.

Our basic finding is that when the when the data are generated by the time series representations in Table 1, 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 Figures 4 and 5. The results for Random Walk and Optimal in Figure 7 correspond to those reported in Figure 4, 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. Trigonometric Regression also performs poorly in real time. For example, $R_T = 1.04$, which is even worse than HP filter. Column 3 displays the dynamic cross correlations between $\hat{y}_t$ and $y_t$ when the former are computed by Trigonometric Regression. The evidence shows that there is very little phase shift between the variables, 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 Appendix B we find that, for standard business cycle statistics, Trigonometric Regression produces results quite similar to Random Walk.

### 3.5. Robustness of Analysis to Trend Stationarity

The quantitative analysis of our filters has been based, up to now, on the assumption that the data are generated by a difference stationary process estimated for one of the standard macroeconomic data series. Our basic finding is that Random Walk is nearly Optimal. Here, we show that - with a slight qualification - the same basic conclusion holds, even if we assume the data are trend stationary.

Our first step in the analysis was to redo the calculations in Table 2, by fitting moving average representations to the residuals from regressions of log GDP, inflation, the interest rate and unemployment on a constant and time trend. We then repeated the preceding calculations, using Optimal, Random Walk, HP filter and Trigonometric regression. Although the results are fairly similar across different time series representations, the model for inflation poses a modestly greater challenge for Random Walk. This is why we decided to only present results based on the time series representation for monthly inflation.

Results are presented in Figure 8, which is in the same format as Figures 4, 5, and 7. There are

two things worth emphasizing here. First, consider the results for Random Walk. As in the difference stationary case (see Figure 4), Random Walk is nearly optimal in the business cycle frequencies, if we ignore the first and last two years' observations. However, some discrepancies between Random Walk and Optimal are apparent in the lower frequencies. Still, in the 8-20 year frequencies, there is no noticeable difference if we ignore the first and last five years' observations. Second, note that Trigonometric Regression still performs worse than Random Walk. Any concerns about Random Walk that might be raised by the results in Figure 8 do not warrant adopting the Trigonometric Regression procedure. If anything, they suggest adopting something closer to Optimal.

## 4. Related Literature

We now briefly relate our analysis to the existing literature. Early work by Engle (1974) argued forcefully the case that economic hypotheses are usefully cast in the frequency domain. One of the examples studied in the next section of this paper, an analysis of the relationship between money growth and inflation, is inspired in part by Engle's work on the same subject.

On a methodological level, the closest work to ours is that of Baxter and King (1999). They emphasize the potential usefulness of the band pass filter for constructing statistics to characterize the dynamics in the data. They confront the same practical problem that we do. Namely, to apply the band pass filter in a finite data set requires approximating it in some way. As noted above, the approximation that they recommend is the filter defined in (2.14). Together with Baxter (1994), Hornstein (1998), King and Watson (1994), and Stock and Watson (1998). They show how the Baxter-King recommended filter can be used to address interesting empirical questions.

Our analysis can be compared with Baxter and King's in three ways. First, our approach to approximating the band pass filter differs from theirs. We select the approximation which minimizes (1.1). Baxter and King adopt a different optimization criterion. They require that the approximating filter optimize (2.7) with $f_x \equiv 1$, subject to the requirement, $\hat{B}(1) = 0$. This is equivalent to optimizing (2.7) under the assumption that $x_t$ has a Near $IID$ time series representation. This is the representation analyzed in the previous section, in which the spectral density is flat over most frequencies, and then rises sharply in a small neighborhood of zero. This observation is useful because it clarifies the set of circumstances in which the Baxter-King filter is expected to work well, in the sense of (1.1).[40] Second, we supply formulas for the optimal approximation to the band pass filter that apply in a broad class of time series representations. This provides alternatives to the Baxter-King filter, which is suitable for

[40] We have not established that the Near $IID$ representation is the only one that rationalizes the Baxter-King filter as the optimal one.

cases when the Near $IID$ assumption is a poor one. We do not expect the Near $IID$ representation to be well suited for many macroeconomic variables. As noted by Granger (1966) (see also Sargent (1987, pp. 280-281)), the 'typical spectral shape' of macroeconomic time series is one in which there is substantial power in a significant range of low frequencies. Third, the Baxter-King approach works with symmetric, fixed lag filters. In the previous section we presented experiments which suggest that adopting filters which use *all* the data, and which are therefore asymmetric and time-varying, improves the estimate of $y_t$. There is a valid basis for concern that filters like this might result in nonstationary $\hat{y}_t$'s and/or might induce phase shifts between $\hat{y}_t$ and $y_t$. However, we find that with time series representations like those that fit several key postwar macroeconomic data series, these effects are not quantitatively large.

Our approach can also be compared to that of Hodrick and Prescott (1997). They provide a linear filter of the data, which has three interpretations. Under the first interpretation, their filter is viewed as solving a particular projection problem: extract a signal from a data series that is the sum of a signal and an orthogonal noise. As is evident from (2.1) and the discussion thereafter, optimally extracting a particular band of frequencies from the data requires solving the same kind of problem. The difference between the two approaches is that, under this interpretation of the HP filter, it is based on a particular statistical model of the data while ours is not.[41] Under this interpretation, the HP filter has two shortcomings: (i) The underlying model has the same difficulty of interpretation as do other trend-cycle decompositions: the concepts of 'signal' and 'noise' they seek to extract from the data do not correspond to meaningful economic objects in standard business cycle models. The data generated by these business cycle models do contain components that could perhaps loosely be characterized as trend and noise. But, they do not satisfy the orthogonality conditions posited by typical trend-cycle decompositions, including the one underlying the HP filter (see Christiano and Eichenbaum (1990)). (ii) Under the first interpretation of the HP filter, one has to take the underlying statistical model seriously. For example, a parameter $\lambda$, which corresponds to the relative variance of the signal and noise in the underlying model, needs to be estimated. This gives rise to numerous estimation and model evaluation issues that, from our perspective, are tangential. The decomposition we focus on is guaranteed to exist under very general conditions by the Spectral Representation Theorem.[42] In our approach, the decomposition selected simply reflects the research interests of the analyst. Focusing on

[41] Lucas (1980) also adopts a structural model like that of Hodrick-Prescott. Both assume that the data, $x_t$, are the sum of a signal, $s_t$, and a noise, $n_t$. The signal and noise are assumed to be orthogonal, and the noise is uncorrelated over time. Lucas assumes the signal has the representation, $s_t = \rho s_{t-1} + v_t$, where $v_t$ is a white noise and $|\rho| < 1$. Hodrick and Prescott assume $s_t = 2s_{t-1} - s_{t-2} + v_t$.

[42] For a formal analysis of the Spectral Decomposition Theorem, see Koopmans (1974). A simplified discussion appears in Christiano and Fitzgerald (1998, Appendix).

a particular decomposition does not require a commitment to any particular model of the data, beyond the relatively weak assumptions needed for the Spectral Representation Theorem to hold.

A second interpretation of the HP filter is that it is a particular band pass filter. For $\lambda = 1600$, it has been argued that the filter does well at isolating frequencies 8 years and higher (see Prescott (1986), King and Rebelo (1993), and Singleton (1988)). No claim has ever been made that the HP filter is an *optimal* approximation to this band pass filter in any sense. This is consistent with our findings in section 3.4.1 which show that the Random Walk and Optimal filters dominate HP, especially for computing real-time estimates of $y_t$. Still, it is shown in the Appendix that these differences are not quantitatively large enough to produce substantial differences in the computation of standard business cycle statistics. A problem is that when one wants to isolate different frequency bands or use monthly or annual data, it is not clear how to proceed under this interpretation of the Hodrick-Prescott approach.[43] The framework of this paper, which focuses on developing optimal approximations to the band pass filter, provides a straightforward way to proceed in these cases.

A third interpretation of the HP filter is that it is a precisely stated algorithm which simply draws a smooth line through the data. All the other filters discussed in this paper do this too.

We find the second two interpretations of the HP filter appealing. Under these interpretations, the band pass filter represents a natural extension of the work of Hodrick and Prescott. The band pass filter performs similarly to the HP filter in situations for which the latter was designed: extracting the business cycle component from quarterly time series data (see Appendix B). However, the band pass filter - particularly the approach advocated here - can be used to construct filters that are effective in isolating other frequency bands as well. The next subsection describes empirical applications where filters of this type are of interest.

## 5. Applications

We illustrate the use of our recommended filter, presented in the introduction, using two examples. The first focuses on the relationship between unemployment and inflation, and the second examines the relationship between money growth and inflation.

We divide the annual data available for the period 1900 to 1997 into two parts: 1900-1960 and

[43]The perspective adopted in this paper does offer one strategy: optimize, by choice of $\lambda$, the version of (2.7) with $B^{p,f}$ replaced by the Hodrick-Prescott filter. This strategy produces a value of $\lambda$ that is time-dependent and dependent upon the properties of the true time series representation. We suspect that closed form solutions for the values of $\lambda$ that solve these optimization problems do not exist. We think that this strategy for filtering the data is not 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.

1961-1997. In each case, 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. With respect to the Phillips curve, we find that in the pre-1960 sample, the correlation between inflation and unemployment is negative at all frequencies. Loosely, we characterize this as reflecting that all frequencies of the data exhibit a Phillips curve trade-off.. The most significant change in the post-1960 sample is that the 20-40 year correlation appears to have switched to positive, with unemployment lagging inflation by several years. We assess this evidence in light of the Barro-Gordon model of inflation.

We then turn to the inflation-money growth relation. We find that in the pre-1960 period, the two variables move closely together in all frequencies. The relationship remains positive in the low frequencies in the post-1960s data. However, at the business cycle and 8-20 year frequencies, there is a substantial change.

### 5.1. The Phillips Curve

We consider annual data for the 1900-1960 and 1960-1997 periods separately. Figure 9a displays the raw unemployment and inflation data for the first sample. For convenience, the figure also depicts the NBER business cycle peak and trough dates. Casual inspection suggests a negative relationship (i.e., a 'Phillips curve') at all frequencies.[44] The various frequency components of the data are displayed in Figures 9b - 9d, and they confirm this impression. The contemporaneous correlations between filtered inflation and unemployment are reported in Table 4, Panel A. The table also reports $p-$values under the null hypothesis that there is no relationship between the variables in any frequency band.[45] The negative correlation in the business cycle frequencies is particularly significant, statistically. In light of these observations, it is not surprising that the scatter plot of inflation and unemployment, exhibited in Figure 9b, also shows a negative relationship. This is just the classic Phillips curve, of textbook fame.[46]

[44]It is worth emphasizing that, by 'Phillips Curve', we mean a *statistical* relationship, and not necessarily a relationship exploitable by policy.

[45]These are computed by fitting separate $q$-lag scalar autoregressive representations to the level of inflation (first difference, log $CPI$) and to the level of the unemployment rate, and using the fitted disturbances and actual historical initial conditions to simulate 2,000 artificial data sets on inflation and unemployment. For annual data, $q = 3$; for monthly, $q = 12$; and for quarterly, $q = 8$. The data sets on unemployment and inflation are independent by construction. In each artificial data set we compute correlations between the various frequency components, as we did 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 4 are the frequency of times that the simulated correlation is greater (less) than the positive (negative) estimated correlation. These are p-values under the null hypothesis that there is no relationship between the inflation and unemployment data.

[46]The slope of the regression line drawn through the scatter plot of points in Figure 9b is $-0.42$, with a $t-$statistic of 3.77 and an $R^2$ of 0.20.

| Table 4: Phillips Curve and Money Growth-Inflation Correlations | | | | |
|---|---|---|---|---|
| Sample | H. Frequency | Bus. Cyc. Frequency | 8-20 years | 20-40 years |
| Panel A: CPI Inflation and Unemployment | | | | |
| 1900-1960 (A) | | -0.57 (0.00) | -0.32 (0.19) | -0.51 (0.23) |
| 1961-1997 (A) | | -0.38 (0.11) | -0.16 (0.41) | 0.45 (0.32) |
| 1961, Q2 - 1997, Q4 (Q) | -0.37 (0.00) | -0.65 (0.00) | -0.30 (0.29) | 0.25 (0.34) |
| 1961, Jan. - 1997, Dec. (M) | -0.24 (0.00) | -0.69 (0.00) | -0.27 (0.30) | 0.23 (0.40) |
| Panel B: CPI Inflation and M2 Growth | | | | |
| 1900-1960 (A) | | 0.45 (0.00) | 0.59 (0.04) | 0.95 (0.01) |
| 1961 - 1997 (A) | | -0.72 (0.00) | -0.77 (0.02) | 0.90 (0.10) |
| 1961, Q2 - 1997, Q4 (Q) | -0.34 (0.00) | -0.67 (0.00) | -0.71 (0.05) | 0.87 (0.09) |
| 1961, Jan. - 1997, Dec. (M) | -0.13 (0.02) | -0.66 (0.00) | -0.74 (0.04) | 0.87 (0.10) |

Note: Contemporaneous correlation between indicated two variables, over indicated sample periods and frequencies. Numbers in parentheses are $p-$ values, in decimals, against the null hypothesis of zero correlation at all frequencies. For further details, see the text and footnote 25.

The post 1960 monthly inflation and unemployment data are analyzed in Figures 10a-f.[47] There is a surprising degree of similarity between the pre and post 1960s data. For example, it is plain from the raw data in Figure 10a that for frequencies in the neighborhood of the business cycle, inflation and unemployment covary negatively. That is, the Phillips curve seems to be a pronounced feature of the higher frequency component of the data. At the same time, the Phillips curve appears to have vanished in the very lowest frequencies. The data in Figure 10a show a slow trend rise in unemployment throughout the 1960s and 1970s, which is reversed starting in early 1983. A similar pattern occurs in inflation, though the turnaround in inflation begins in April 1980, roughly three years before the turnaround in unemployment. The low frequency component of the data dominates in the scatter plot of inflation versus unemployment, exhibited in Figure 10b. That figure suggests that the relationship between inflation and unemployment is positive, in contrast with the pre-1960's data, which suggest otherwise (see Figure 9b). Clearly, this scatter plot exaggerates the degree to which the inflation-

[47]Figure 10 exhibits monthly observations on inflation and unemployment. To reduce the high frequency fluctuations in inflation, Figure 10a exhibits the annual average of inflation, rather than monthly inflation rate. The scatter plot in Figure 10b is based on the same data used in Figure 10a. Figures 10c-10f are based on monthly inflation, i.e., $1200log(CPI_t/CPI_{t-1})$ and unemployment. The line in Figure 10b represents regression line drawn through the scatter plot. The slope of that line, based on monthly data covering the period 1959:2 - 1998:1, is 0.47 with a $t-$statistic of 5.2.

unemployment dynamics have changed in the post-1960s data.[48]

Impressions based on casual inspection of the raw data can be formalized and quantified using the band pass filtered data reported in Figures 10c - 10f. Thus, the frequency range from two months to 20 years (see Figures 10c and 10e) is characterized by a noticeable Phillips curve. Table 4 shows that the correlation in the range of high frequencies (when available) and in the business cycle frequencies is significantly negative. The correlation between inflation and unemployment is also negative in the 8-20 year range, but it is not statistically significantly different from zero in this case. Presumably, this reflects the relative paucity of information about these frequencies in the post-1960s data. Finally, Figure 10f indicates that the correlation between 20-40 year components is now positive, with unemployment lagging inflation. These results are consistent with the hypothesis that the Phillips curve changed relatively little in the 2-20 year frequency range, and that the changes that did occur are primarily concentrated in the very low frequencies.

Formal tests of this hypothesis, presented in Panel A of Table 5, fail to reject it. The table displays $p-$values for the null hypothesis that the post-1960s data on inflation and unemployment are generated by the bivariate vector autoregression (VAR) that generated the pre-1960s data. We implement the test using 2,000 artificial post-1960s data sets obtained by simulating a three-lag VAR and its fitted residuals estimated using the pre-1960s unemployment and inflation data.[49] In each artificial data set we compute correlations between filtered inflation and unemployment just like we did in the actual post 1960s data. Table 5 indicates that 9 percent of correlations between the business cycle component of inflation and unemployment exceed the $-0.38$ value reported in Table 4 for the post-1960s data, so that the null hypothesis fails to be rejected at the 5 percent level. The $p-$value for the $8-20$ year correlation is quite large, and is consistent with the null hypothesis at any significance level.

The statistical evidence against the null hypothesis that there has been no change in the $20-40$ year component of the data is also not strong. This may in part reflect a lack of power stemming from the relatively small amount of information in the sample about the $20-40$ year frequency component of the data. But, the $p-$value may also be overstated for bias reasons. The table indicates that there is a small sample bias in this correlation, since the small sample mean, $-0.35$, is substantially larger than the corresponding probability limit of $-0.45$. This bias may, at least in part, be the reason the VAR's

[48] Consistent with these observations, when inflation and unemployment are detrended using a linear trend with a break in slope (not level) in 1980:4 for inflation and 1983:1 for unemployment, the scatter plot of the detrended variables show a negative relationship. The regression of detrended inflation on detrended unemployment has a coefficient of $-0.31$, with $t$ statistic of -4.24 and $R^2 = 0.037$. The slope coefficient is similar to what was obtained in an earlier footnote for the pre-1960s period, but the $R^2$ is considerably smaller.

[49] We redid the calculations in both panels of Table 4 using a 5-lag VAR and found that the results were essentially unchanged. The only notable differences in the results are that the $p-$value for the business cycle correlations between inflation and unemployment is 0.06 and the $p-$value for these correlations in the $20-40$ year range is 0.11.

small sample mean overstates the pre-1960s sample mean of $-0.51$. A bias-adjustment procedure would adjust the coefficients of the estimated pre-1960s VAR so that the implied small sample mean lines up better with the pre-1960s empirical estimate. Presumably, such an adjustment procedure would shift the simulated correlations to the *left*, reducing the $p-$value. It is beyond the scope of our analysis to develop a suitable bias adjustment method.[50] However we suspect that, given the large magnitude of the bias, the bias-corrected $p-$value would be substantially smaller than the 14 percent value reported in the table.[51]

[50] One could be developed along the lines pursued by Kilian (1998).

[51] To get a feel for the likely quantitative magnitude of the effects of bias adjustment, we redid the bootstrap simulations by adjusting the variance-covariance matrix of the VAR disturbances used in the bootstrap simulations. Let $V = [V]_{ij}$ denote the variance-covariance matrix. In the pre-1960s estimation results, $V_{1,2} = -0.1024$, $V_{1,1} = 0.0018$, $V_{2,2} = 6.0653$. When we set the value of $V_{1,2}$ to $-0.0588$ and recomputed the entries in Table 4, we found that the mean correlations were as follows: business cycle, $-0.75$ (0.01); 8-20 year: $-0.54$ (0.09); 20-40 year: $-0.51$ (0.06). The numbers in parentheses are the analogs of the $p-$values in Table 4. Note how the mean correlation in the 20-40 year frequency coincides with the empirical estimate reported in the first row of Panel A of Table 4, and that the $p-$value has dropped substantially, from 0.23 to 0.06. This confirms our conjecture that bias adjustment may have an important impact on the $p-$value for the 20-40 year correlation. However, the other numbers indicate that the bias adjustment procedure that we applied, by varying $V_{1,2}$ only, is not a good one. Developing a superior bias adjustment method is clearly beyond the scope of this paper.

| Table 5: Testing Null Hypothesis That Post-1960s Equal pre-1960s Correlations | | | | |
|---|---|---|---|---|
| | | Small Sample | Std. Dev., Small | |
| Frequency | Plim | Mean | Sample Mean | $p-$value |
| Panel A: DGM, Bivariate VAR with CPI Inflation ($x$) and Unemployment ($y$) | | | | |
| 2-8 year | -0.66 | -0.61 | $0.0036\times\sqrt{2000}$ | 0.09 |
| 8-20 year | -0.36 | -0.38 | $0.0079\times\sqrt{2000}$ | 0.25 |
| 20-40 year | -0.45 | -0.35 | $0.0129\times\sqrt{2000}$ | 0.14 |
| Panel B: DGM, Bivariate VAR with CPI Inflation ($x$) and M2 Growth ($y$) | | | | |
| 2-8 year | 0.49 | 0.48 | $0.0044\times\sqrt{2000}$ | 0.00 |
| 8-20 year | 0.73 | 0.64 | $0.0062\times\sqrt{2000}$ | 0.00 |
| 20-40 year | 0.78 | 0.64 | $0.0099\times\sqrt{2000}$ | 0.37 |

Notes: (i) Data Generating Mechanism (DGM) in all cases is 3-lag, Bivariate VAR fit to pre-1960s data. (ii) P-value: frequency, in 2000 artificial post-1960s data sets, that contemporaneous correlation between indicated frequency components of $x$ and $y$ exceeds, in absolute value, corresponding post 1960s estimate. (iii) Plim: mean, over 1,000 artificial samples of length 2,000 observations each, of correlation.
(iv) Small Sample Mean: mean of correlation, across 2000 artificial post-1960s data sets.
(v) Std. dev., small sample (product of Monte Carlo std. error for mean and $\sqrt{2000}$) : standard deviation, across 2000 artificial post-1960s data sets, of correlations.

The low-frequency observations on unemployment have been documented using other methods (see, for example, Barro (1987, Chapter 16)). Similarly, using the Baxter-King band-pass filtering approach, the presence of the Phillips curve in the business cycle frequency range has been documented by King and Watson (1994) and Stock and Watson (1998) (see also Sargent (1999, p.12)). What has not been documented is how far the Phillips curve extends into the low frequencies of the post war data. In addition, we draw attention to the evidence that inflation leads unemployment in the low frequency range. We think these frequency domain properties of inflation and unemployment are interesting because we suspect that they pose a challenge for at least simple versions of conventional interpretations of the post war data.

One such interpretation is articulated by the theory developed in Kydland and Prescott (1977) and Barro and Gordon (1983), which builds on the idea of the expectations-augmented Phillips curve

initially suggested by Friedman (1968) and Phelps (1968).[52] The basic idea is that institutional impediments that had previously constrained US monetary policy, such as the gold standard, ceased to do so by the early 1960s. With institutional constraints impeding discretionary monetary policy minimized, a new monetary regime was initiated, in which the monetary authority set out to improve economic performance by attempting to actively exploit the Phillips curve. But, according to this view, the Phillips curve is merely a statistical relation, and does not represent a menu of inflation, unemployment levels that can be selected by policy makers. The simplest version of the Barro-Gordon theory predicts that the new monetary regime should have caused it to disappear, and be replaced by a deterministic, positive relationship between inflation and unemployment.[53]

Our scatter diagrams show that before the 1960s inflation and unemployment are negatively related, while they are positively related after the 1960s. At first glance, this appears to be a resounding victory for the Barro-Gordon theory. However, evidence that is plain from casual inspection of the raw time series and is formalized using the band pass filter suggests a different picture. The simple version of the Barro-Gordon theory implies that, after 1960, the Phillips curve relationship should have evaporated, not just at the low frequencies, but *at all frequencies*. It also implies that inflation and unemployment should be perfectly in phase. Both these implications appear to be contradicted by the data.

Barro and Gordon (1983, pp. 601-602) recognize these extreme implications of their theory and describe an extension to their model that is designed to deal with them. The extension involves making the (plausible) assumption that the monetary authority's control over inflation is imperfect, perhaps because of lags in the effects of monetary actions. Under these conditions, inflation surprises occur, thus opening up the possibility that the forecast errors between inflation and unemployment might be

[52] Our discussion builds on Ireland (1998), who emphasizes the importance of evaluating the time series implications of the Barro-Gordon model.

[53] For convenience, we briefly remind the reader of the Barro-Gordon model. The loss function of the monetary authority is $\left[(u-ku^N)^2+\gamma\pi^2\right]/2$, where $\gamma>0$ is a parameter that specifies the relative social cost associated with inflation, $\pi$; $u$ is the actual unemployment rate; and $u^N$ is the natural rate of unemployment, i.e., the rate that would prevail if inflation were at its expected level. Also, $0<k<1$. Inflation surprises generate a fall in the unemployment rate according to the 'expectations augmented Phillips curve': $u-u^N=-\alpha\left(\pi-\pi^e\right),\ \alpha>0$, where $\pi^e$ is expected inflation. At the time the monetary authority chooses its action, $\pi^e$ is a state variable. The policy maker's objective is to minimize: $L(\pi|\pi^e,u^N)=\left\{\left[\alpha\left(\pi-\pi^e\right)-(1-k)\,u^N\right]^2+\gamma\pi^2\right\}/2$. Let $\pi\left[\pi^e,u^N\right]=\arg\min L(\pi|\pi^e,u^N)$. Conventionally, the Markov equilibrium concept is used, in which equilibrium $\pi$ and $\pi^e$ depend only on $u^N$, and not the past history of central bank actions. This equilibrium concept has the advantage that, in the Barro-Gordon model, it is unique. The equilibrium is an expectation function, $\pi^*(u^N)$, with the property $\pi^*(u^N)=\pi\left[\pi^*(u^N),u^N\right]$, so that, in a Markov equilibrium, $u=u^N$. Examining the first order condition associated with the monetary authority's problem and imposing the fixed point condition, implies

$$\pi=\psi u,\ \psi=\frac{\alpha(1-k)}{\gamma}>0.$$

The microeconomic environment in the Barro-Gordon theory is specified at the reduced form level. Environments in which preferences and technology are stated explicitly, and which capture elements of the Barro-Gordon analysis may be found in Ireland (1997), and Chari, Christiano and Eichenbaum (1998).

negatively correlated. In turn, this may put the model into better conformity with the observation that there is a Phillips curve in the high frequencies, but not in the low frequencies.

Perhaps it is possible to reconcile the last observation with plausible implementation delays.[54] However, we suspect that it will be difficult to identify variations on the model that make it consistent with the observation that inflation leads unemployment in the low frequency range.[55] This phase relationship calls into question a basic feature of the Barro-Gordon model, according to which policy makers set inflation in *response* to developments in unemployment. A fruitful approach may be to explore modifications of the theory, in which policymakers' incentives to inflate are driven by variables that exhibit a more plausible timing relationship to inflation than unemployment.

We hope that this discussion demonstrates that statistics based on approximations to the band pass filter have a potentially useful role to play in helping us think about the inflation, unemployment dynamics. This is true, whether they are used for evaluating a Barro-Gordon type conception of the inflation process, or some other conception.[56]

---

[54] Ireland (1998) introduces a one period lag, and he shows that this is this is strongly rejected by the data. In addition to considering longer implementation lags, another potentially useful modification would introduce rigidities that lead inflation surprises to have long-lasting effects. In the standard Barro-Gordon model, the effect of such a surprise is only contemporaneous.

[55] Our skepticism is based in part on analysis of an extension of Ireland (1998)'s version of the Barro-Gordon model in which a one-period implementation lag is replaced by a $p > 0$ period implementation lag in policy. Thus, we replace the policymaker's objective in footnote 37 with $E_{-p}\left[(u-u^N)^2+\gamma\pi^2\right]/2$. Actual inflation, $\pi$, is $\pi = \hat{\pi} + \theta * \eta$, where $\hat{\pi}$ is a variable chosen $p$ periods in the past by the policy maker, and $\theta * \eta$ captures the shocks that impact on $\pi$ between the time $\hat{\pi}$ is set and $\pi$ is realized. Here,

$$\theta * \eta_t = \left(\theta_0 + \theta_1 L + ... + \theta_{p-1}L^{p-1}\right)\eta_t,$$

where $\eta_t$ is a white noise and $L$ is the lag operator. The policymaker's problem is optimized by setting $\hat{\pi} = \psi\hat{u}^N$, where $\hat{u}^N$ is the forecast of the natural rate of unemployment available at the time $\hat{\pi}$ is selected. As in footnote 37, we suppose that $u^N$ is a random walk with innovation $\nu$, so that $u^N = \hat{u}^N + g * \nu$, where

$$g * \nu_t = (1 + L + ... + L^{p-1})\nu_t.$$

Suppose the Lucas supply curve is $u = u^N - \alpha(\pi - \pi^e) + \varepsilon$, where $\varepsilon$ is a shock. We impose rational expectations, which implies $\pi^e = \hat{\pi}$. Then, it is easy to verify that inflation and unemployment evolve according to

$$\pi_t = \psi\frac{L^p}{1-L}\nu_t + \theta(L)\eta_t,\ u_t = \frac{1}{1-L}\nu_t - \alpha\theta(L)\eta_t + \varepsilon_t,$$

respectively. We make the simplifying assumption that all shocks are uncorrelated with each other. Then, the cross spectrum between these variables is:

$$g_{\pi u}(\omega) = \psi\sigma_\nu^2\left[\frac{e^{-ip\omega}}{(1-e^{-i\omega})(1-e^{i\omega})} - \frac{\alpha}{\psi}\theta(e^{-i\omega})\theta(e^{-i\omega})\lambda\right],$$

where $\lambda = \sigma_\eta^2/\sigma_\nu^2$. The phase angle, $\theta(\omega)$, between inflation and unemployment is:

$$\theta(\omega) = \tan^{-1}\left[\frac{\text{imaginary}\left[g_{\pi u}(\omega)\right]}{\text{real}\left[g_{\pi u}(\omega)\right]}\right] = \tan^{-1}\left[\frac{-\sin(p\omega)}{\cos(p\omega) - 2(1-\cos(\omega))\frac{\alpha}{\psi}\theta(e^{-i\omega})\theta(e^{i\omega})\lambda}\right],$$

subject to $-\pi \leq \theta(\omega) \leq \pi$. It is easy to verify that for $\omega$ sufficiently small, $\theta(\omega)/\omega = p$, so that inflation lags unemployment by $p$ periods at these frequencies.

[56] An alternative approach, based on the assumption that policy makers have bounded rationality, is pursued in Sargent

### 5.2. Money Growth and Inflation

Figure 11a displays annual M2 growth and CPI inflation for in the pre-1960s period. The figure suggests, and the results in Table 4 confirm, that the two variables move very closely with each other. The primary exception appears to be the period of the second world war, when M2 growth was strong and inflation rose relatively less. But, this presumably reflects the effects of wage and price controls. The impression of the strong contemporaneous correlation at all frequencies is confirmed by the evidence in Figures 11b - 11d. Now consider Figures 12a-d. These depict the same data over the post-1960 period. Note first that there is essentially no difference in the 20-40 year frequency. In this band, money growth and inflation are highly correlated and there is very little phase shift between the two. However, there is a striking change in the relationship at the higher frequencies. The two variables are now strongly negatively correlated. According to Table 5, the null hypothesis that the business cycle and 8-20 year correlations in the post-1960s coincide with the corresponding pre-1960s correlation is strongly rejected. The change appears to be that in these frequencies, inflation now lags money growth by a few years. We think these are interesting statistics which a good model of money ought to confront.[57]

## 6. Conclusion

The theory of spectral analysis informs us that data can be viewed as the sum of uncorrelated components with different frequencies of fluctuation. This way of thinking about the data makes rigorous a perspective that is standard in macroeconomics: that the data are the sum of low frequency components, a business cycle component and a high frequency component. The various components can in principle be extracted by a suitably chosen 'ideal' band-pass filter.

Unfortunately, application of the ideal band pass filter requires substantially more data than is available in typical macroeconomic time series. This is not surprising. A subinterval of frequencies in a time series is composed of a continuum of objects, and in general there is no way that a discrete set of observations can pin these down. However, if something is known about the time series representation that generated the data, then it is possible to use projection theory to extrapolate from the observed time series to the frequency components of interest. We derive a set of formulas for this that are valid

(1999) and Sims (1988). Sargent (1999, chapter 9) argues that this approach also has empirical problems too. In particular, it has difficulty accounting for the burst of inflation in the late 1970s.

[57] Lucas (1980) was also interested in examining the relationship between the low frequency components of inflation and money growth. However, rather than using an approximation to the band pass filter, he used an early precursor to what later became known as the Hodrick-Prescott filter. Our comments on the Hodrick-Prescott filter, stated in the previous section, also apply to Lucas' filter. We think the band pass filter represents an advance over that filter. In a comment on Lucas (1980), Whiteman (1984) presents a useful reminder of the pitfalls of trying to interpret statistics like those discussed in this paper without the assistance of explicit economic theory.

under standard time series representations, and are easy to implement. We identify one approximation which, though it is only optimal for one particular time series representation, nevertheless works well for standard macroeconomic time series. It is displayed in the introduction.

To illustrate the use of our recommended filter approximation, we use it to characterize the change in the nature of the Phillips curve and the money-inflation relation before and after the 1960s. We find that there is surprisingly little change in the Phillips curve and substantial change in money growth-inflation relation. In this analysis, we display a bootstrap methodology for conducting statistical inference on statistics computed using filtered data.

## A. Derivation of the Filter

This Appendix derives the optimal filtering formulas analyzed in the body of the paper.

### A.1. The Problem

In our discussion we feature the unit root case (i.e., $\theta(1) \neq 0$ in (2.8)) and indicate later how to adjust things to accommodate covariance stationarity. The problem is to optimize (2.7):

$$\min_{\hat{B}_j^{p,f}, j=-f,...,p} \int_{-\pi}^{\pi} \delta(\omega)\delta(-\omega) f_x(\omega) d\omega,$$

where $f_x(\omega)$ is defined in (2.9) and,

$$\delta(\omega) = B(e^{-i\omega}) - \hat{B}^{f,p}(e^{-i\omega}).$$

A necessary condition for an optimum is $\hat{B}^{f,p}(1) = 0$. Without this, the criterion is infinite. With this condition,

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

is a finite-ordered polynomial in $z$:

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

with

$$b_j = -\sum_{i=j+1}^{p} \hat{B}_i^{f,p}, \; j = p-1, ..., -f. \tag{A.1}$$

We suppose that $p \geq 0$ and $f \geq 0$ are 'large' in the sense that $p + f > 0$ and $p + f \geq 2q$.

Write the criterion as follows:

$$\min_{b_j, j=p-1,...,-f} \int_{-\pi}^{\pi} \tilde{\delta}(\omega)\tilde{\delta}(-\omega) g(e^{-i\omega}) d\omega,$$

where

$$\tilde{\delta}(\omega) = \tilde{B}(e^{-i\omega}) - b(e^{-i\omega}),$$

and

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

To obtain the first order conditions, note first that

$$\frac{d\tilde{\delta}(\omega)}{db_j} = -e^{-i\omega j},\ j = p-1, ..., -f,$$

so that,

$$\begin{aligned}\frac{d\tilde{\delta}(\omega)\tilde{\delta}(-\omega)}{db_j} &= \tilde{\delta}(\omega)\frac{d\tilde{\delta}(-\omega)}{db_j} + \tilde{\delta}(-\omega)\frac{d\tilde{\delta}(\omega)}{db_j} \\ &= -\left[\tilde{\delta}(\omega)e^{i\omega j} + \tilde{\delta}(-\omega)e^{-i\omega j}\right],\ j = p-1, ..., -f.\end{aligned}$$

Then, the first order conditions are:

$$\int_{-\pi}^{\pi}\left[\tilde{\delta}(\omega)e^{i\omega j} + \tilde{\delta}(-\omega)e^{-i\omega j}\right] g(e^{-i\omega})d\omega = 0,\ j = p-1, ..., -f,$$

or,

$$\int_{-\pi}^{\pi}\left[\tilde{B}(e^{-i\omega})e^{i\omega j} + \tilde{B}(e^{i\omega})e^{-i\omega j}\right] g(e^{-i\omega})d\omega = \int_{-\pi}^{\pi}\left[b(e^{-i\omega})e^{i\omega j} + b(e^{i\omega})e^{-i\omega j}\right] g(e^{-i\omega})d\omega,$$

$j = p-1, ..., -f$. This expression reduces further to:

$$\int_{-\pi}^{\pi}\tilde{B}(e^{-i\omega})g(e^{-i\omega})e^{i\omega j}d\omega = \int_{-\pi}^{\pi}b(e^{-i\omega})g(e^{-i\omega})e^{i\omega j}d\omega, \tag{A.2}$$

$j = p-1, ..., -f$.[1] Our strategy for computing the $\hat{B}_j^{f,p}$'s is to translate the $p+f$ equations, (A.2) into a system of linear equations in $\hat{B}_j^{f,p}$, $j = p, ..., -f$. We obtain the $p+f+1^{st}$ equation in the $\hat{B}_j^{f,p}$'s from the fact, $\hat{B}^{f,p}(1) = 0$.

We begin by replacing the system of $p+f$ equations, (A.2), with another system composed of the equation with $j = -f$ and the $p+f-1$ equations formed by subtracting the $j-1^{th}$ equation in (A.2) from the $j^{th}$, $j = p-1, ..., -f+1$.

---

[1] To see this, note that

$$\begin{aligned}R &\equiv \int_{-\pi}^{\pi}f(\omega)d\omega = \int_{0}^{\pi}f(\omega)d\omega + \int_{-\pi}^{0}f(\omega)d\omega \\ &= \int_{0}^{\pi}\left[f(\omega) + f(-\omega)\right]d\omega,\end{aligned}$$

so that $R$ is real. Consequently,

$$\int_{-\pi}^{\pi}f(\omega)d\omega = \int_{-\pi}^{\pi}f(-\omega)d\omega,$$

and

$$2\int_{-\pi}^{\pi}f(\omega)d\omega = \int_{-\pi}^{\pi}\left[f(\omega) + f(-\omega)\right]d\omega.$$

### A.2. Representing the Solution as a Solution to a System of Linear Equations

Consider the term on the left of (A.2). That term, evaluated at $j$, minus itself at $j-1$, for $j = p-1, ..., -f+1$, is:

$$\begin{aligned} & \int_{-\pi}^{\pi} \{\tilde{B}(e^{-i\omega})e^{i\omega j} - \tilde{B}(e^{-i\omega})e^{i\omega(j-1)}\} g(e^{-i\omega}) d\omega \\ = & \int_{-\pi}^{\pi} \tilde{B}(e^{-i\omega})(1-e^{-i\omega})e^{i\omega j} g(e^{-i\omega}) d\omega \\ = & \int_{-\pi}^{\pi} B(e^{-i\omega}) g(e^{-i\omega}) e^{i\omega j} d\omega. \end{aligned} \tag{A.3}$$

To evaluate this integral and others like it, we find it convenient to apply the well-known results:

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

With this, it is easy to see that the integral in (A.3) is the product of $2\pi$ and the constant (i.e., the coefficient on $z^0$) in the polynomial in $z$, $B(z)g(z)z^{-j}$.

To evaluate the integral in (A.3), consider first the case $q = 0$, so that $g(e^{-i\omega}) = \theta_0^2$. In this case, (A.3) reduces to $2\pi B_j \theta_0^2$, where $B_j$ is defined in (1.3). When $q > 0$,

$$\begin{aligned} & \int_{-\pi}^{\pi} B(e^{-i\omega}) g(e^{-i\omega}) e^{i\omega j} d\omega \\ = & \ 2\pi \left( B_j c_0 + \sum_{i=1}^{q} \left[ B_{|j|+i} + B_{||j|-i|} \right] c_i \right) \end{aligned} \tag{A.4}$$

The reason why the absolute value of $j$, $|j|$, appears in this expression is that the constant term in $B(z)g(z)z^{-j}$ corresponding to a given value of $j$, coincides with the constant term associated with $-j$. This is because $B(z)g(z) = B(z^{-1})g(z^{-1})$.

Now consider the term on the right of (A.2). For $j = p-1, ..., -f+1$, subtract that term minus itself at $j-1$ to obtain:

$$\begin{aligned} & \int_{-\pi}^{\pi} \left[ b(e^{-i\omega})e^{i\omega j} - b(e^{-i\omega})e^{i\omega(j-1)} \right] g(e^{-i\omega}) d\omega \\ = & \int_{-\pi}^{\pi} b(e^{-i\omega})(1-e^{-i\omega})e^{i\omega j} g(e^{-i\omega}) d\omega \\ = & \int_{-\pi}^{\pi} \hat{B}^{f,p}(e^{-i\omega}) g(e^{-i\omega}) e^{i\omega j} d\omega. \end{aligned} \tag{A.5}$$

This integral is the product of $2\pi$ and the constant term in the polynomial in $z$, $\hat{B}^{f,p}(z)g(z)z^{-j}$, for

$j = p-1, ..., -f+1$. When $q = 0$, this is $2\pi \hat{B}_j^{f,p} \theta_0^2$. When $q > 0$,

$$\int_{-\pi}^{\pi} \hat{B}^{f,p}(e^{-i\omega}) e^{i\omega j} g(e^{-i\omega}) d\omega = 2\pi A_j \hat{B}^{f,p}, \tag{A.6}$$

where $\hat{B}^{f,p}$ is a $p+f+1$ column vector:

$$\hat{B}^{f,p} = \left[\hat{B}_p^{f,p}, \hat{B}_{p-1}^{f,p}, ..., \hat{B}_0^{f,p}, ...\hat{B}_{-f+1}^{f,p}, \hat{B}_{-f}^{f,p}\right]',$$

and $A_j$ is a $p+f+1$ row vector, $j = p-1, ..., -f+1$. For $p-q \geq j \geq q-f$:

$$A_j = \left[\underbrace{0, ..., 0}_{1\times(p-j-q)}, c, \underbrace{0, ..., 0}_{1\times(j-q+f)}\right], \tag{A.7}$$

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

We now consider equation (A.2) for $j = -f$ :

$$\int_{-\pi}^{\pi} \tilde{B}(e^{-i\omega}) g(e^{-i\omega}) e^{-i\omega f} d\omega = \int_{-\pi}^{\pi} b(e^{-i\omega}) e^{-i\omega f} g(e^{-i\omega}) d\omega.$$

This equation can be written

$$\int_a^b \left[\frac{e^{-i\omega f}}{1-e^{-i\omega}} + \frac{e^{i\omega f}}{1-e^{i\omega}}\right] g(e^{-i\omega}) d\omega = 2\pi F b, \tag{A.8}$$

where $b = [b_{p-1}, b_{p-2}, ..., b_0, ...., b_{-f}]'$ is a $p+f-$ dimensional column vector and $F$ is the $p+f$ dimensional row vector:

$$F = [\underbrace{0, ..., 0}_{1\times(p+f-q-1)}, c_q, c_{q-1}, ..., c_0]. \tag{A.9}$$

If $p+f-q-1 = 0$, the row vector is as above, with the zeros deleted. We want to express the right hand side of (A.8) in terms of $\hat{B}^{f,p}$ instead of $b$. We do this using the matrix representation of the relation, (A.1):

$$Q\hat{B}^{f,p} = b, \tag{A.10}$$

where $Q$ is the following $(p+f)\times(p+f+1)$ matrix:

$$Q=\begin{bmatrix} -1 & 0 & 0 & \cdots & 0 & 0 \\ -1 & -1 & 0 & \cdots & 0 & 0 \\ -1 & -1 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ -1 & -1 & -1 & \cdots & -1 & 0 \end{bmatrix}. \tag{A.11}$$

Substitute (A.10) into (A.6) to get:

$$\int_a^b \left[\frac{e^{-i\omega f}}{1-e^{-i\omega}}+\frac{e^{i\omega f}}{1-e^{i\omega}}\right] g(e^{-i\omega})d\omega = 2\pi FQ\hat{B}^{f,p} = 2\pi A_{-f}\hat{B}^{f,p},$$

say, where $A_{-f}=FQ$.

We have now achieved what we set out to do. We have a linear system of $p+f+1$ equations in the $p+f+1$ unknown $\hat{B}_j^{f,p}$'s. We summarize this as follows:

$$d = A\hat{B}^{f,p}, \tag{A.12}$$

where

$$d=\begin{bmatrix} \int_{-\pi}^{\pi} B(e^{-i\omega})g(e^{-i\omega})e^{i\omega(p-1)}d\omega \\ \vdots \\ \int_{-\pi}^{\pi} B(e^{-i\omega})g(e^{-i\omega})e^{i\omega(-f+1)}d\omega \\ \int_a^b \left[\frac{e^{-i\omega f}}{1-e^{-i\omega}}+\frac{e^{i\omega f}}{1-e^{i\omega}}\right] g(e^{-i\omega})d\omega \\ 0 \end{bmatrix},\ A=2\pi\begin{bmatrix} A_{p-1} \\ \vdots \\ A_{-f+1} \\ A_{-f} \\ 1 \quad \cdots \quad 1 \end{bmatrix}, \tag{A.13}$$

where $A_j$, $j=p-1,...,-f+1$ is defined in (A.7) and $A_{-f}=FQ$, where $Q$ is defined in (A.11) and $F$ is defined in (A.9).

With a small adjustment to $d$ and $A$, the filter weights are also determined by an expression like (A.12) when $x_t$ is covariance stationary. To indicate the adjustment, it is convenient to adopt a notational convention that is different from the one used in section 2 of the paper. In the covariance stationary case, we specify that $x_t$ has the representation

$$x_t=\theta(L)\varepsilon_t,\ E\varepsilon_t^2=1,$$

where

$$\theta(z)=\theta_0+\theta_1 z+...+\theta_q z^q,\ q\geq 0,$$

and the spectral density of $\theta(L)\varepsilon_t$ is $g(e^{-i\omega})$, with

$$g(e^{-i\omega}) = \theta(e^{-i\omega})\theta(e^{i\omega}) = c_0 + c_1(e^{-i\omega} + e^{i\omega}) + ... + c_q(e^{-i\omega q} + e^{i\omega q}).$$

Here, $c_\tau$ is the covariance function of $\theta(L)\varepsilon_t$. As before, we wish to solve (2.6) subject to (2.5). The unknown filter weights satisfy the linear system, (A.12), in which all but the last two rows of $A$ and $d$ are unchanged. The last two rows of $d$ contain $\int_{-\pi}^{\pi} B(e^{-i\omega})g(e^{-i\omega})e^{i\omega l}d\omega$ for $l = p$ and $-f$, respectively. The last two rows of $A$ contain $A_p$ and $A_{-f}$, where these are evaluated using (A.7).

### A.3. The Solution

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

$$\hat{B}^{f,p} = A^{-1}d. \tag{A.14}$$

When $q = 0$, this problem is very simple. In this case,

$$\begin{aligned} \hat{B}_j^{f,p} &= B_j,\ j = p-1, ..., -f+1 \\ \hat{B}_{-f}^{f,p} &= \frac{1}{2\pi}\int_a^b \left[\frac{e^{-i\omega f}}{1-e^{-i\omega}} + \frac{e^{i\omega f}}{1-e^{i\omega}}\right] d\omega \\ \hat{B}_p^{f,p} &= -\sum_{j=p-1}^{-f} \hat{B}_j^{f,p}. \end{aligned} \tag{A.15}$$

It is of interest to note that, when $p = f$, $\hat{B}^{p,p}(L)$ is a symmetric polynomial. To see this, note first that $\hat{B}_j^{p,p} = B_j$ for $j = p-1, ..., 0, ..., 1-p$ and recall that the $B_j$'s are themselves symmetric. But, is it the case that $\hat{B}_p^{p,p} = \hat{B}_{-p}^{p,p}$? To see that this is indeed the case, note first from (A.1) that $\hat{B}_p^{p,p} = -b_{p-1}$. But, after evaluating (A.2) at $j = p-1$, we conclude:

$$\hat{B}_p^{p,p} = -\frac{1}{2\pi}\int_a^b \left[\frac{e^{-i\omega(p-1)}}{1-e^{-i\omega}} + \frac{e^{i\omega(p-1)}}{1-e^{i\omega}}\right] d\omega.$$

Combining this with the middle expression in (A.15), evaluated at $f = p$, we find that $\hat{B}_p^{p,p} = \hat{B}_{-p}^{p,p}$ if, and only if

$$\int_a^b \left[\frac{e^{-i\omega p}}{1-e^{-i\omega}} + \frac{e^{i\omega p}}{1-e^{i\omega}} + \frac{e^{i\omega(p-1)}}{1-e^{-i\omega}} + \frac{e^{-i\omega(p-1)}}{1-e^{i\omega}}\right] d\omega = 0.$$

To see that this relationship indeed is satisfied, first collect terms in $e^{i\omega p}$ :

$$e^{i\omega p}\left[\frac{1}{1-e^{i\omega}} + \frac{e^{-i\omega}}{1-e^{-i\omega}}\right] = e^{i\omega p}\frac{1-e^{-i\omega} + \left(1-e^{i\omega}\right)e^{-i\omega}}{\left(1-e^{i\omega}\right)\left(1-e^{-i\omega}\right)} = 0.$$

The same is true for the term in $e^{-i\omega p}$. We conclude,

$$\hat{B}_p^{p,p} = \hat{B}_{-p}^{p,p},$$

so that when $p = f$ and $q = 0$, the filter produced by this projection is symmetric.

### A.4. Computational Note

Note from (A.7), (A.4), (A.13), that, with one exception, we have closed form expressions for all the objects- that are needed to solve our problem. The exception occurs in the unit root case, with the expression

$$R(f) = \int_a^b \left[ \frac{e^{-i\omega f}}{1 - e^{-i\omega}} + \frac{e^{i\omega f}}{1 - e^{i\omega}} \right] g(e^{-i\omega}) d\omega.$$

This expression could be solved using a numerical integration procedure. The drawback is that such computations take time, and, as emphasized in the paper, in practice the expression needs to be evaluated for many different values of $f$. For this reason, we now describe a quick way to evaluate this integral.

Note first that when $f = 0$,

$$\begin{aligned} R(0) &= \int_a^b \left[ \frac{1}{1 - e^{-i\omega}} + \frac{1}{1 - e^{i\omega}} \right] g(e^{-i\omega}) d\omega \\ &= \frac{1}{2} \int_{-\pi}^{\pi} B(e^{-i\omega}) g(e^{-i\omega}) d\omega, \end{aligned}$$

since the object in square brackets is unity. The expression, $R(0)$, can be evaluated using (A.4).

$$\begin{aligned} R(f) - R(f+1) &= \int_a^b \left( e^{-i\omega f} + e^{i\omega f} \right) g(e^{-i\omega}) d\omega \\ &= \frac{1}{2} \int_{-\pi}^{\pi} B(e^{-i\omega}) g(e^{-i\omega}) \left( e^{-i\omega f} + e^{i\omega f} \right) d\omega \\ &= \int_{-\pi}^{\pi} B(e^{-i\omega}) g(e^{-i\omega}) e^{-i\omega f} d\omega. \end{aligned}$$

The last integral in this expression can be evaluated using (A.4). This gives us a recursive way to compute $R(0), R(1), ...$, using closed-form formulas.

## B. Second Moments of Aggregate Variables

We compare the second moment properties of aggregate time series, filtered using the various procedures discussed in this paper. Results are presented for various frequencies: the high frequencies with period of oscillation between 2 and 6 quarters; business cycle frequencies with period of oscillation between 6 and 32 quarters; low frequencies with period of oscillation between 32 and 80 quarters; and very low frequencies with period of oscillation between 80 and 160 quarters. We also present results based on combining the high and business cycle frequencies, so that we can do a direct comparison with results based on the HP filter.

Our findings are presented in Tables B.1-B.5. The tables are differentiated according to the frequency band isolated by the filter method used. The band pass filter approximations considered are Random Walk, Random Walk Fixed, Random Walk Symmetric, the Baxter-King method (BK), and Trigonometric Regression. All but the last two of these methods are defined in Table 2. BK is defined in (2.14) and Trigonometric Regression is defined in (3.1). The HP filter used in each table is always the same, and sets $\lambda = 1600$. The data cover the period 1948 to 1998. In most cases, they are logged prior to filtering. Exceptions are cases where the variable can potentially be negative. Thus, '(Ex-Im)/GDP' represents net exports deflated by GDP, and this is not logged. The same is true for 'Chg Invntry/GDP' which is the change in the stock of inventories (i.e., inventory investment), scaled by GDP. In addition, interest rates have not been logged.

Our basic finding is that results are very similar across the different filtering methods, except at the very lowest frequencies. There, they differ sharply. It is our impression that nothing of substantive turns on which filter is used to isolate the higher frequency components of the data.

Three observations about the sensitivity of the Random Walk results to the choice of frequency band are worth noting. First, consistent with findings in Hornstein (1998), the correlation between inventory investment and sales is negative in the high frequencies, and positive in the lower frequencies. Perhaps this reflects the predominance of different types of shocks in these frequency bands. Second, consumption is about 25 percent more volatile than output in the very lowest frequencies. This stands in striking contrast to its behavior in the higher frequencies, where consumption is substantially less volatile than output. Third, the correlation between the price level and output is more negative in the low frequencies than in the higher frequencies. On the other hand, the correlation between output and inflation increases as one goes from the high frequencies to the low frequencies. The switch in the sign of the correlation between inflation and output and the price level and output has been noted before (see Ball and Mankiw (1994), Christiano (1991), and Cooley and Ohanian (1991)), and has generated

considerable discussion. That the phenomenon is more pronounced in the lower frequencies has not been noted before.

## References


[1] Ball,Laurence, and N. Gregory Mankiw, 1994, 'A Sticky-Price Manifesto,' Carnegie-Rochester Conference Series on Public Policy, 41 (0), December, pages 127-151.

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

[3] Barro, Robert J., 1987, Macroeconomics, John Wiley & Sons, second edition.

[4] Barro, Robert J., and David B. Gordon, 1983, 'A Positive Theory of Monetary Policy in a Natural Rate Model,' *Journal of Political Economy* 91 (August): 589-610.

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

[6] Baxter, Marianne, Robert G. King, 1999, 'Measuring Business Cycles Approximate Band-Pass Filters for Economic Time Series,' *International Economic Review*, forthcoming.

[7] Chari, V.V., Lawrence J. Christiano and Martin Eichenbaum, 1998, 'Expectation Traps and Discretion,' *Journal of Economic Theory*.

[8] Christiano, Lawrence J., 1991, 'Modeling the Liquidity Effect of a Monetary Shock,' Federal Reserve Bank of Minneapolis *Quarterly Review*.

[9] Christiano,Lawrence J. and Martin Eichenbaum, 1990, 'Unit Roots in Real GNP: Do We Know, and Do We Care?' Carnegie-Rochester Conference Series on Public Policy, 32 (0), Spring, pages 7-61.

[10] 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.

[11] Christiano, Lawrence J., and Terry J. Fitzgerald, 1998, 'The Business Cycle: It's Still a Puzzle,' *Economic Perspectives*, Federal Reserve Bank of Chicago.

[12] Cooley, Thomas F., and Lee Ohanian, 1991, 'The Cyclical Behavior of Prices,' *Journal of Monetary Economics*, 28 (1), August, pages 25-60.

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

[14] Friedman, Milton, 1968, 'The Role of Monetary Policy,' American Economic Review, 58 (March), 1-17.

[15] Granger, Clive W. J., 1966, 'The Typical Spectral Shape of an Economic Variable,' *Econometrica*, vol. 34, no. 1, pp. 150-61.

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

[17] 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.

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

[19] Ireland, Peter, 1997, 'Sustainable Monetary Policies,' *Journal of Economic Dynamics and Control*, vol. 22, November, pp.87-108.

[20] Ireland, Peter N., 1998, 'Does the Time-Consistency Problem Explain the Behavior of Inflation in the United States?', manuscript, Department of Economics, Boston College.

[21] Kilian, Lutz , 1998, 'Small - Sample Confidence Intervals for Impulse Response Functions', May, *The Review of Economics and Statistics*.

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

[23] 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.

[24] 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.

[25] Koopmans**,** Leonid, 1974, The Spectral Analysis of Time Series, New York, New York, Academic Press.

[26] Kydland, Finn, and Edward C. Prescott, 1977, 'Rules Rather than Discretion: the Inconsistency of Optimal Plans,' *Journal of Political Economy*, June, vol. 85, no. 3, pp. 473-491.

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

[28] Lucas, Robert E., 1980, 'Two Illustrations of the Quantity Theory of Money,' *American Economic Review*, December.

[29] Phelps, Edmund, 1968, 'Money-Wage Dynamics and Labor-Market Equilibrium,' *Journal of Political Economy*, August, part 2, pp. 678-711.

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

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

[32] Phillips, A. W., 1958, 'The Relationship Between Unemployment and the Rate of Change of Money Wages in the United Kingdom, 1861-1957,' *Economica* 25, (November).

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

[34] 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.

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

[36] Sargent, Thomas J., 1999, *The Conquest of American Inflation*, Princeton University Press.

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

[38] Sims, Christopher, 1988, 'Projecting Policy Effects with Statistical Tools,' *Revista de Analisis Economico*, vol. 3, pp. 3-20.

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

[40] 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.

[41] Stock, James, and Mark Watson, 1998, 'Business Cycle Fluctuations in US Macroeconomic Time Series,' NBER Working Paper 6528, forthcoming in Taylor and Woodford, *Handbook of Macroeconomics*.

[42] Taylor, John B., 1999, *Monetary Policy Rules*, University of Chicago Press.

[43] Whiteman, Charles, 1984, 'Lucas on the Quantity Theory: Hypothesis Testing Without Theory,' *American Economic Review*, vol. 74, no. 4, pp. 742-749.