UNIVERSIDADE TÉCNICA DE LISBOA

INSTITUTO SUPERIOR DE ECONOMIA E GESTÃO

MESTRADO EM: Econometria Aplicada e Previsão

Multivariate Filtering with Common Factors

Ana Regina Nunes Pereira

ORIENTAÇÃO: Prof. Doutor João Valle e Azevedo

JÚRI:

Presidente: Prof. Doutor Artur Carlos Barros da Silva Lopes

Vogais:

Prof. Doutor Paulo Manuel Marques Rodrigues

Prof. Doutor João Carlos Henriques da Costa Nicolau

Prof. Doutor João Valle e Azevedo

June/2009

Abstract

This study discusses four commonly used optimal approximations to the infinite order moving average filter that ideally extracts from a time series fluctuations within a specified range of periodicities. Based on our findings, we use two of those approximations in the estimation of two macroeconomic signals: business cycle fluctuations and medium to long run component of output growth rate. This study distinguishes itself from related literature by showing how to successfully incorporate in the multivariate band-pass approximations factors estimated from a large panel of time series.

As illustration, we apply these approximations to U.S. data. We evaluate the real-time performance of the indicators and provide forecasting comparisons. The results suggest that the multivariate indicator outperforms the competing univariate indicator across all different settings considered. Moreover, multivariate methods that target smooth growth are useful to forecast quarterly GDP growth rate at short-term and to forecast yearly GDP growth.

Keywords: dynamic factor models, band-pass filter, business cycle fluctuations, smooth component, coincident indicator, macroeconomic fluctuations.

Resumo

Este estudo discute quatro aproximações óptimas ao filtro de médias móveis infinitas que idealmente isola de uma série temporal flutuações compreendidas num determinado intervalo de periodicidades. De acordo com as nossas conclusões, utilizamos duas dessas aproximações na estimação de dois sinais macroeconómicos: flutuações de ciclo económico no produto e a componente de médio e longo prazo da taxa de crescimento do produto. Este estudo distingue-se da literatura corrente ao mostrar como integrar nas aproximações do filtro banda multivariado factores estimados a partir de um largo painel de séries temporais.

Como ilustração, aplicamos estas aproximações a dados dos E.U.A.. Avaliamos o desempenho dos indicadores em tempo real e apresentamos comparações em termos de previsão. Os resultados sugerem que o indicador multivariado tem um desempenho claramente superior ao do indicador univariado em todos os cenários considerados. Adicionalmente, os métodos multivariados que aproximam o crescimento alisado são úteis na previsão da taxa de crescimento trimestral do PIB a curto prazo e para previsão do crescimento anual do PIB.

Palavras Chave: modelos dinâmicos de factores, filtro de banda, flutuações de ciclo económico, componente alisada, indicador coincidente, flutuações macroeconómicas.

Contents

List of Figures

List of Tables

1 Introduction

A main concern of macroeconomic analysts, policy makers and others lies in tracking, in real time, the state of the economy. So, they focus attention on signals that aim at summarizing information embedded in the time series movements of major macroeconomic aggregates.

The idea that an individual time series can be seen as the sum of multiple components driven by different kinds of shocks follows from Persons' (1919) work. Ever since several methods have been proposed in the literature to measure those different components; in particular the trend and cyclical component. Methods can be divided into two groups: economic-based models and statistical-based models. The first group uses economic theory to explain the mechanism of fluctuations while the second group uses purely statistical assumptions to identify the components. A brief survey of the literature indicates as references for a discussion of economic-based models Singleton (1988), King et al. (1988, 1991), Blanchard and Quah (1989) and Cochrane (1994). Alternatively, statistical-based models include deterministic detrending, first differences, Beveridge and Nelson's (1981) decomposition, unobserved components models, filtering techniques, Stock and Watson's (1998) index models or Altissimo et al.'s (2008) projection problem.

In this study the main objective will be to construct a real time indicator for two specific macroeconomic signals: business cycle fluctuations of aggregate output and smooth component of output growth (henceforth smooth growth). We will use a filtering approach to directly obtain such signals, which amounts to apply to the series of interest a filter specifically designed to isolate only the fluctuations within a pre-specified range of periodicities. Unlike most methods, filtering techniques clearly permit to achieve an explicit separation between components and additionally provide a simple way to accomplish such task. The downside follows from the statistical nature of this approach that compromises the interpretation of the extracted components from an economic point of view.

The first step is to specify the threshold values of the cyclical periods that distinguish the different components of a time series. We overcome this problem adopting the definition of business cycle fluctuations as those fluctuations with period between [6, 32] quarters in the pseudo-spectrum of Gross Domestic Product (GDP) series, as advocated by Baxter and King (1999), and smooth growth as output growth excluding the fluctuations with cyclical period lower than 4 quarters, exactly as in Altissimo et al. (2008). These signals undoubtedly contain important information to assess the direction of the economy. Moreover, we view forecasts of the smooth growth indicator as being useful to forecast GDP growth itself because the possibly unpredictable short-run oscillations, approximated by conventional models, have been eliminated. Our empirical application provides important insights regarding this matter.

Both signals can be obtained by applying an infinite order moving average filter to the series of interest. However, this procedure implies infinite data and thus an approximate filter is required for empirical purposes. According to an optimization criterion, good approximations to the ideal infinite sample filter are proposed by Hodrick and Prescott (1997), Baxter and King (1999) and Christiano and Fitzgerald (2003) in an univariate context and by Valle e Azevedo (2007) in a multivariate context. In the first part of this study we analyze in detail the properties of each of these approximations and discuss which ones are suitable for real time analysis. We conclude, given our objective, that Hodrick and Prescott's filter as well as Baxter and King's filter detain undesirable features. Therefore, only the two remaining approximations from those mentioned earlier can be used to construct a real-time indicator.

The second part of this work provides an empirical application of the optimal filter suggested by Christiano and Fitzgerald and of the optimal filter suggested by Valle e Azevedo to U.S. quarterly GDP or GDP growth rate, depending on the signal. The univariate filter is used as a benchmark filter for comparisons and the multivariate approximation is adopted as our proposed indicator. The multivariate optimal filter does not only exploit the information in a single time series but also that in a large panel of monthly economic variables. To compress this additional information, the panel is assumed to be described by a dynamic factor model, as originally developed by Geweke (1977) and Sargent and Sims (1977). To be specific, each variable of the panel is assumed to be described by a few number of common factors plus an idiosyncratic error. These factors are unobservable variables and therefore estimated using either principal components as in Stock and Watson (2002a, 2002b) or generalized principal components as in Forni et al. (2005). Accordingly, we compare the two methods of approximating the factor space. The extracted factors are then incorporated as covariates in the multivariate approximation of the signals of interest.

In view of the above, our main contribution to the current literature is the integration of recent developments in the analysis of dynamic factor models in the approximation of band-pass filters. Specifically, we implement an approximation to the signal of interest using a multivariate approach that combines information derived from a large panel of time series (reduced by estimation of common factors). Moreover, our multivariate approach can be used in any similar signal extraction problem since it can easily be adapted to optimally approximate any other signal of interest or equivalently to extract any other range of periodicities.

To simulate a true real time exercise both signals are defined on the GDP of the current quarter and the release delays of all the variables involved is taken into account. This implies that our real activity indicators will be timely and that our method is flexible enough to easily produce real time estimates. Furthermore, the elimination of fluctuations with low periodicities implies that the indicators will display little short-run oscillations, thereby giving a clear picture of current cyclical and growth prospects. Finally, due to the monthly frequency of the variables of the panel we are able to obtain an update of the multivariate indicator each month and not just at a quarterly frequency as for the univariate filter. All these features stress the advantages inherent to our method.

Our findings reveal that the multivariate indicator outperforms the competing univariate indicator across all settings considered in all months of the year. These settings include variations in the estimation procedure of factors and second order moments. In detail, the best performing indicator for both signals is the multivariate filter using two monthly factors and moments derived from a parametric estimator. We conclude that exploiting information in other variables other than real GDP is helpful in mitigating the approximations errors arising from missing data.

In addition, as a by-product we use the best performing multivariate filter for forecasting quarterly and yearly output growth rates. This exercise gives important insights on whether it is more relevant, for forecasting purposes, to target a smooth version of a time series or instead the original time series containing the irregular oscillations. We found that multivariate methods that target smooth growth are useful for short-term forecast of quarterly growth rates and that at long horizons all methods seem useless for forecasting purposes. These results support the findings of Runstler et al. (2008) and Reichlin et al. (2008). In terms of forecasting the yearly growth rate we report a surprising accuracy in the multivariate forecasts that are derived from approximations to smooth growth, even at the end of the first quarter of the year, where the task is rather demanding.

This work is organized into 5 sections. In section 2 we define precisely the characteristics of the signals that we aim at approximating and make clear how this task can be accomplished.

After this we discuss individually each approximate filter and exam their properties in order to rule out those unfeasible for real-time analysis. Section 3 discusses topics related with the practical implementation of the approximations. In detail, we survey methods to estimate second order moments and introduce factor models, focusing on factor space approximation and on how to determine the number of factors. Section 4 presents the sample to which we apply the approximate filters and evaluates the performance of various real-time indicators. Moreover, some forecasting results are also discussed. Finally, section 5 concludes with a summary of our findings and further research topics.

2 Setting the problem

In this section we define precisely the signals (or components) that we aim to isolate from a time series and review some of the methods available to optimally approximate them.

Our objective is to obtain two distinct macroeconomic signals: business cycle fluctuations and the medium to long run component of the output growth rate. To this effect, we will center our attention on real Gross Domestic Product (GDP) since it is the best available proxy of the aggregate economic activity. The business cycle fluctuations will be extracted from the logarithm of GDP, denoted by y_t, while the medium to long run component will be extracted from the GDP growth rate, denoted by $\Delta y_t = (1 - L)y_t$ where L is the lag operator.

In a frequency domain perspective[1], business cycle fluctuations can be identified as those *fluctuations with a specified range of periodicities* (see Baxter and King, 1999) in the (pseudo-) spectrum of y_t[2]. Following Burns and Mitchell (1946), the range limits should be set to 6 and 32 quarters, meaning that the cyclical component of GDP is composed of all cycles with period no less than 6 quarters and no more than 32 quarters. This definition of business cycle fluctuations is completely arbitrary in the sense that it is always possible to extract fluctuations in any other range of periodicities that one might understand as cyclical movements. Furthermore, we argue that the definition of the different components of a time series, in particular that of business cycle fluctuations, is model-dependent. For now we adopt this widely used definition but we will contrast it with others, in particular with the Hodrick and Prescott filter (see below). The medium to long run component of GDP growth rate (henceforth smooth growth) is defined as the output growth cleaned of fluctuations with period less than one year (4 quarters), as in Altissimo et al. (2008)[3]. Note that this signal can be seen as the trend-cycle component of the GDP growth rate, which excludes variability of short duration. Accordingly, it will vary smoothly over time and it is precisely this characteristic that makes this signal an interesting predictor of GDP growth. This last idea will be explored in section 4. Moreover, both signals exclude irregular and seasonal variation by eliminating all fluctuations with period less than 6 quarters in the case of business cycle fluctuations and 4 quarters in the case of smooth growth. This reduces the

[1] For more details on the frequency domain approach see Appendix A.

[2] The spectrum is only well defined if y_t is a stationary process. Given that GDP may contain a unit root we define instead the pseudo-spectrum, which is not well defined at frequency zero.

[3] Altissimo et al. (2008) define this signal to construct the new EuroCoin indicator, which is a coincident indicator for euro area growth.

need to account for GDP data revisions if these have most of its power concentrated in the high frequencies. Nevertheless, we still have to account for other types of revisions.

We now turn to the particular question of how to obtain these signals. It is well-known that any arbitrary signal can be extracted by applying a two-sided infinite order moving average to the series of interest (z_t). As presented in more detail in Appendix B the application of such a linear filter to a covariance stationary time series z_t results in

$$x_t = \sum_{j=-\infty}^{\infty} B_j z_{t-j} = B(L) z_t$$

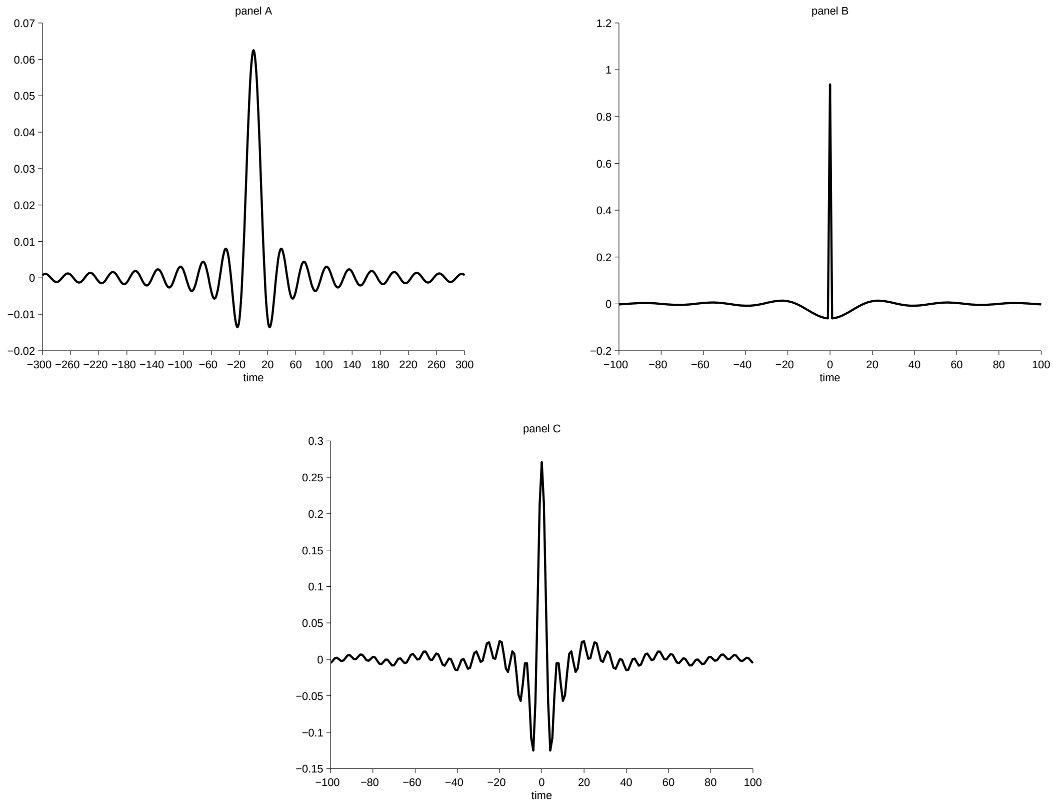
where x_t denotes the component to be extracted from z_t, $B(L) = \sum_{j=-\infty}^{\infty} B_j L^j$ where L is the lag operator and $\{B_j : j = 0, \pm 1, \pm 2, \ldots\}$ are weights verifying $\sum_{j=-\infty}^{\infty} |B_j| < \infty$ (absolutely summable).

In order to isolate only specific movements, we will have to chose a particular design for the weight sequence, which is better handled via frequency domain analysis. In this perspective, the focus is on the spectral density function (or spectrum) which decomposes the time series fluctuations into orthogonal frequencies. Each component of a time series can be distinguished from others by the different amount of time it requires to complete a whole cycle, or equivalently, we may say that each component is connected with specific periodicities. In turn, these latter periodicities (p) relate to frequencies (ω) through $p = \frac{2\pi}{\omega}$, suggesting that each component can be indistinctly identified either by their range of periodicities or range of frequencies. Under the assumption that z_t is a covariance stationary process with absolutely summable autocovariance function $\gamma_z(k), k = 0, \pm 1, \pm 2, \ldots$, the spectrum of the filtered series, $S_x(\omega)$, is given by

$$S_x(\omega) = \left|B\left(e^{-i\omega}\right)\right|^2 S_z(\omega), \quad -\pi \leq \omega \leq \pi \tag{1}$$

where $B\left(e^{-i\omega}\right) = \sum_{j=-\infty}^{\infty} B_j e^{-i\omega j}$ is the frequency response function (or the Fourier transform) of the linear filter $B(L)$ and $S_z(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma_z(h) e^{-i\omega h}$ is the spectrum of z_t. The function $\left|B\left(e^{-i\omega}\right)\right|^2$ in equation (1) is known as transfer function (or square gain function) and works as a weighting sequence of the spectrum of the series to be filtered. Given a particular frequency $\bar{\omega}$, if $\left|B\left(e^{-i\bar{\omega}}\right)\right|^2 > 1$ (or < 1) then the fluctuations of z_t with periodicity $\frac{2\pi}{\bar{\omega}}$ will pass to x_t with their properties emphasized (attenuated) while if $\left|B\left(e^{-i\bar{\omega}}\right)\right|^2 = 0$ (or $= 1$) then the fluctuations of z_t with periodicity $\frac{2\pi}{\bar{\omega}}$ will be completely removed (exactly preserved) from x_t. Given the role of the gain function in the spectrum of the filtered series and the correspondence of frequencies to periodicities it is straightforward to construct a linear filter that exactly preserves fluctuations within specific periodicities and that completely eliminates all the other undesirable fluctuations. This type of linear filters are referred to as ideal filters and a discussion of their characteristics is provided in Appendix B. Figure 1 shows the ideal behavior of the infinite moving average weights (in the time domain) when the goal is to extract a component related to low frequencies (panel A), high frequencies (panel B) or intermediate frequencies (panel C).

Figure 1: Time domain representation of the weights of an ideal low pass filter (panel A), an ideal high pass filter (panel B) and an ideal band-pass filter (panel C).



Reviewing, to extract a particular signal we have to first specify its characteristics in terms of frequencies or periodicities and then apply the infinite order moving average filter with the proper weights to the series of interest. Obviously, in empirical studies some finite version of those filters is used instead, but this will be subject for discussion in the next subsection. For now we focus on translating the previous ideas to the specified signals.

Define the following decomposition of y_t and Δy_t:

$$y_t = BC(L)y_t + (1 - BC(L))y_t \tag{2}$$

$$\Delta y_t = SG(L)\Delta y_t + (1 - SG(L))\Delta y_t \tag{3}$$

where the logarithm of GDP and the GDP growth rate are regarded as a sum of two components; one with power only at the frequencies of interest, namely $BC(L)y_t$ and $SG(L)\Delta y_t$, and a second with power outside those frequencies. Business cycle fluctuations and smooth growth rates are obtained by applying to y_t and Δy_t, respectively, a infinite, symmetric and two-sided filter as follows

$$\begin{aligned} BC(L)y_t &= \sum_{j=-\infty}^{\infty} BC_j L^j y_t \\ SG(L)\Delta y_t &= \sum_{j=-\infty}^{\infty} SG_j L^j \Delta y_t \end{aligned}$$

with

$$BC_0 = \tfrac{\omega_u - \omega_l}{\pi}, \quad BC_j = \tfrac{\sin(\omega_u j) - \sin(\omega_l j)}{\pi j}, \ \ |j| \geq 1 \text{ with } \omega_l = \tfrac{2\pi}{32} \text{ and } \omega_u = \tfrac{2\pi}{6} \tag{4}$$

$$SG_0 = \frac{\omega_u}{\pi}, \qquad SG_j = \frac{\sin(\omega_u j)}{\pi j}, \ \ |j| \geq 1 \text{ with } \omega_u = \frac{2\pi}{4}$$

where ω_l is the lowest frequency and ω_u is the highest frequency of the frequency band of interest. Accordingly the spectra of the desired components are given by

$$S_{BC}(\omega) = \left|BC\left(e^{-i\omega}\right)\right|^2 S_y(\omega), \quad -\pi \leq \omega \leq \pi \quad \omega \neq 0 \tag{5}$$

$$S_{SG}(\omega) = \left|SG\left(e^{-i\omega}\right)\right|^2 S_{\Delta y}(\omega), \quad -\pi \leq \omega \leq \pi$$

where ω is a frequency defined in radians, $BC\left(e^{-i\omega}\right) = \sum_{j=-\infty}^{\infty} BC_j e^{-i\omega j}$ is the frequency response function of the linear filter $BC(L)$, $SG\left(e^{-i\omega}\right) = \sum_{j=-\infty}^{\infty} SG_j e^{-i\omega j}$ is the frequency response function of the linear filter $SG(L)$, $S_y(\omega)$ is the pseudo-spectrum of real GDP and $S_{\Delta y}(\omega)$ the spectrum of GDP growth rate. Remembering the equivalence between periodicities and frequencies as well as the signals' exact definitions it is evident what will be the designs of the squared gain functions. Extracting business cycle fluctuations amounts to isolate the interval of frequencies $[2\pi/32; 2\pi/6]$ from the pseudo-spectrum of y_t whereas extracting the smooth growth amounts to isolate all frequencies lower than $2\pi/4$ from the spectrum of Δy_t. Thus, in the first case the filter must retain without distortion fluctuations in a time series between a lower and upper bound frequency and remove completely all variations outside this range of frequencies. In the case of the smooth growth component the filter must eliminate the high frequency movements and preserve the frequencies connected to long and medium term fluctuations.

As a result, the filter $BC(L)$ will be an ideal band-pass filter and so its gain function is

$$\left|BC\left(e^{-i\omega}\right)\right|^2 = \begin{cases} 1, & \omega_l \leq |\omega| \leq \omega_u \\ 0, & \text{elsewhere} \end{cases} \tag{6}$$

with $\omega_l = \frac{2\pi}{32}$ and $\omega_u = \frac{2\pi}{6}$ denoting the lower and upper frequencies. The linear filter $BC(L)$ removes unit roots because $BC(1) = 0$. So, the signal $x_t = BC(L)y_t$ is stationary, even when y_t is integrated of order 1, and contains only fluctuations with frequencies in the specified range. In the presence of a unit root process, like real GDP usually is, one must define the spectrum of an integrated series in order to interpret the relation in (5) as usual. The pseudo-spectrum of an integrated series is routinely defined as the limit of the spectrum of a stationary process when the smallest autoregressive roots converge to 1 (see Harvey, 1993, Den Haan and Sumner, 2004, Young et al., 1999). Accordingly, if y_t follows and ARIMA(p,1,q) process its pseudo-spectrum is

$$S_y(\omega) = \frac{\sigma_\epsilon^2}{2\pi} \frac{\psi(e^{-i\omega})\psi(e^{i\omega})}{(1 - e^{-i\omega})(1 - e^{i\omega})}$$

where y_t satisfies $(1 - L)y_t = \psi(L)\epsilon_t$ and ϵ_t is a white noise sequence with variance σ_ϵ^2. This function is well defined for all frequencies except at $\omega = 0$. In any case, we conclude that equation (5) holds by definition if $S_y(\omega)$ is defined as the pseudo-spectrum. When $\omega \in [\omega_l; \omega_h] \subseteq]0; \pi]$ the pseudo-spectrum is well defined and $BC(e^{-i\omega}) = 1$ so $S_x(\omega) = S_y(\omega)$. If $\omega \in]0; \pi]/[\omega_l; \omega_h]$ then

$BC(e^{-i\omega}) = 0$ by definition and $S_x(\omega) = 0$. Finally, if the pseudo-spectrum is not well defined (or $\omega = 0$) we conclude that $S_x(0) = 0$. This follows from two observations: first $BC(L)$ can be factored as $BC(L) = (1-L)^2 BC_{**}(L) = (1-L)BC_*(L)$ because the ideal band-pass filter removes unit roots and this implies that $BC_*(1) = 0$ and second $x_t = BC(L)y_t = BC_*(L)(1-L)y_t = BC_*(L)z_t$ but z_t is stationary and its spectrum is well defined for all frequencies. Some researchers however argue that (5) does not hold because unit root processes theoretically do not have a spectral density function. Follows that it is a meaningless discussion if the focus is on the signal because, regardless of the interpretation given to $S_y(\omega)$, the ideal band-pass filter will still isolate a component with fluctuations within the specified range of periodicities. However, the spectrum of x_t often will exhibit a sharp peak at the lowest frequency of the band of interest and some researchers wrongly associate that pattern with the presence of spurious cycles. Clearly that is not a consequence of filtering a unit root process as argued by Cogley and Nason (1995) and others because a similar behavior arises when filtering highly persistent but stationary processes (see Pedersen 2001).

For the smooth growth case we set up an ideal low pass filter and so the gain function of $SG(L)$ follows

$$\left|SG\left(e^{-i\omega}\right)\right|^2 = \begin{cases} 1, & |\omega| \leq \omega_u \\ 0, & \text{elsewhere} \end{cases}$$

with $\omega_u = \frac{2\pi}{4}$ denoting the threshold frequency.

At this point it is clear that the extraction of the signals entails a infinite number of weights. In such case an approximation is required. The usual approach is to obtain an optimal approximation, in other words, the minimum mean square error (MSE) estimate of the signal. In detail, this involves specifying the mean square error as loss function

$$L = E[(x_t - \hat{x}_t)^2]$$

where $x_t = B(L)y_t$ is an arbitrary signal of interest with $B(L) = \sum_{j=-\infty}^{\infty} B_j L^j$ and $\hat{x}_t = \hat{B}(L)y_t$ is an approximation to the signal x_t with $\hat{B}(L)$ in general given by $\hat{B}(L) = \sum_{j=-f}^{p} \hat{B}_j L^j$ where p and f denote the number of past and future observations, respectively, considered in the approximation. In frequency domain representation we have[4]

$$\tilde{L}(\omega) = \int_{-\pi}^{\pi} \left| B\left(e^{-i\omega}\right) - \hat{B}\left(e^{-i\omega}\right) \right|^2 S_y(\omega) d\omega$$

where $B(z)$ denotes the frequency response function of the ideal filter, $\hat{B}(z)$ denotes the frequency response function of the approximate filter and $S_y(\omega)$ the spectrum (or pseudo-spectrum) of y_t. To obtain an approximate filter with optimal weights the following optimization problem is solved

$$\underset{\{\hat{B}_j\}_{j=-f}^{p}}{Min} \; \tilde{L}(\omega) = \underset{\{\hat{B}_j\}_{j=-f}^{p}}{Min} \int_{-\pi}^{\pi} \left| B\left(e^{-i\omega}\right) - \hat{B}\left(e^{-i\omega}\right) \right|^2 S_y(\omega) d\omega. \tag{7}$$

The objective function equals the square modulus of the difference between the frequency domain representation of the ideal filter and the frequency domain representation of the approximate filter, weighted at each frequency by the spectrum of the time series to be filtered. The optimal weights will depend on the ideal weights and on the properties of the series to be filtered due to the presence of the spectrum in the loss function. However, as the true times series representation is unknown the spectrum is in practice replaced by its empirical counterpart (see below for a discussion of this point).

2.1 Optimal approximations

This section analyzes in some detail four suggested approximations to business cycle fluctuations, i.e., solutions to the minimization problem in equation (7) with $B\left(e^{-i\omega}\right)$ given in equation (6). An approximation to smooth growth is obtained by adapting the following approximate filters to a different frequency band and to the case of filtering a stationary time series.

[4] To obtain the frequency domain representation of the loss function define $x_t = \Gamma(L) y_t$ with $\Gamma(L) = B(L) - \hat{B}(L)$. In Appendix A we show that the spectrum of x_t is

$$S_x(\omega) = \left|\Gamma(e^{-i\omega})\right|^2 S_y(w), \quad -\pi \leq \omega \leq w.$$

Through the Fourier transform of $S_x(\omega)$ we get

$$\gamma_x(k) = \int_{-\pi}^{\pi} \left|\Gamma(e^{-i\omega})\right|^2 S_y(\omega) e^{i\omega k} d\omega, \quad k = 0, \pm 1, \pm 2, ...$$

but given that the loss function is the variance of x_t it follows that

$$L = \gamma_x(0) = \int_{-\pi}^{\pi} \left|\Gamma(e^{-i\omega})\right|^2 S_y(\omega) d\omega$$

To abbreviate notation throughout this section $B(L)$ denotes the time domain representation of the ideal filter that extracts a business cycle component.

2.1.1 The Baxter and King approximation

Baxter and King (1999) propose an approximate filter to extract business cycle fluctuations from a time series based on six criteria that restrain the filter design; see table 1.

Table 1: Baxter and King's (1999) objectives to be met by their optimal approximation.

	Baxter and King's objectives
1.	*"... the filter should extract a specified range of periodicities and otherwise leave the properties of this extracted component unaffected"*
2.	*"... the ideal band-pass filter should not introduce phase shifts, i. e., that it not alter the timing relationships between the time series at any frequency;"*
3.	*"... method be an optimal approximation to the ideal band-pass filter; we specify a particular quadratic loss function for discrepancies between the exact and approximate filter"*
4.	*"... the application of an approximate band-pass must result in a stationary time series even when applied to trending data."*
5.	*"... the method yield business-cycle components that are unrelated to the length of the sample period."*
6.	*"... method be operational"*

The first entry in table 1 means that the ideal filter is a band-pass filter with a gain function of one over the $[-\omega_u; -\omega_l] \cup [\omega_l; \omega_u]$ frequency range and zero at all other frequencies, exactly as in equation (6).

To avoid phase shifts (second objective in table 1) it is necessary to set the phase of the filter equal to zero. The complex nature of the frequency response function, $B\left(e^{-i\omega}\right)$, implies a polar representation as follows

$$B\left(e^{-i\omega}\right) = \left|B\left(e^{-i\omega}\right)\right| e^{-i\theta(\omega)}$$

where $\left|B\left(e^{-i\omega}\right)\right|$ denotes the gain and $\theta(\omega)$ the phase of the filter. Therefore, the second objective corresponds to setting $\theta(\omega) = 0$ in the previous equation, which implies:

$$B\left(e^{-i\omega}\right) = \left|B\left(e^{-i\omega}\right)\right|.$$

The fourth objective is attained under two restrictions: first the optimal weights must be symmetric and second must sum to zero. Mathematically,

$$bk_j = bk_{-j} \quad \forall j \ \Rightarrow BK(L) = bk_o + \sum_{j=1}^{K} bk_j \left(L^j + L^{-j}\right) \tag{8}$$

$$BK(1) = \sum_{j=-K}^{K} bk_j = bk_o + 2\sum_{j=1}^{K} bk_j = 0 \tag{9}$$

where $BK(L) = \sum_{j=-K}^{K} bk_j L^j$ is the time domain representation of Baxter and King's approximate filter (BK filter) and K the maximum lag length[5]. In the frequency domain we are constraining the response of the filter to be zero at the zero frequency, which is intuitive since low frequencies are related to the trend component. Nevertheless, these assumptions do not restrict the trend to be stochastic. In fact, Baxter and King (1999) prove that these two assumptions remove either stochastic trends up to second order or up to quadratic deterministic trends.

The last two objectives in table 1 state that the weights should not be time-dependent and that the method should be of easy implementation. Finally, the third objective describes how the weights of the approximate filter are obtained in order to have an optimal approximation to the ideal symmetric band-pass filter.

Adding the constraint in equation (9), to ensure a trend elimination property in the approximate filter, the weights of the BK filter solve the following optimization problem:

$$\underset{\{bk_j\}_{j=-K,\ldots,K}}{Min} \quad \mathrm{E}\left[\left(\sum_{j=-\infty}^{\infty} b_j y_{t-j} - \sum_{j=-K}^{K} bk_j y_{t-j}\right)^2\right]$$

$$s.t. \ \ BK(1) = \sum_{j=-K}^{K} bk_j = bk_o + 2\sum_{j=1}^{K} bk_j = 0$$

where $\left\{b_j : j = 0, \pm 1, \ldots\right\}$ denotes the weights of the ideal band-pass filter and $\left\{bk_j : j = 0, \pm 1, ..., \pm K\right\}$

[5]In the approximate filter of Baxter and King $p = f = K$.

the weights of the optimal filter. In the frequency domain we have

$$\underset{\{bk_j\}_{j=-K,\ldots,K}}{Min} \int_{-\pi}^{\pi} \left|B\left(e^{-i\omega}\right) - BK\left(e^{-i\omega}\right)\right|^2 d\omega$$
$$s.t. \;\; BK(1) = \sum_{j=-K}^{K} bk_j = bk_o + 2\sum_{j=1}^{K} bk_j = 0$$

where $B(z) = \sum_{j=-\infty}^{\infty} b_j z^j$ and $BK(z) = \sum_{j=-K}^{K} bk_j z^j$ with z taken to be a complex scalar. The objective function written in frequency domain reveals that Baxter and King implicitly assume an independent and identically distributed (i.i.d.) process for the time series y_t since no spectrum appears in the objective function. As a result, the square error terms are all equally weighted and the optimal weights will not directly depend on the true data generating process (DGP).

The $2K+1$ weights of the approximate filter are obtained from the first order conditions (FOC) derived from the implied Lagrange function:

$$\mathfrak{L}\left(bk_{-K}, \ldots, bk_K, \lambda\right) = \int_{-\pi}^{\pi} \left|B\left(e^{-i\omega}\right) - BK\left(e^{-i\omega}\right)\right|^2 d\omega - \lambda \sum_{j=-K}^{K} bk_j$$

$$\begin{cases} \frac{\partial \mathfrak{L}(bk_{-K},\ldots,bk_K,\lambda)}{\partial bk_j} & = -\int_{-\pi}^{\pi}\left(e^{-i\omega j} + e^{i\omega j}\right)\left[B\left(e^{-i\omega}\right) - BK\left(e^{-i\omega}\right)\right] d\omega - \lambda = 0, \quad j = -K, ..., K \\ \frac{\partial \mathfrak{L}(bk_{-K},\ldots,bk_K,\lambda)}{\partial \lambda} & = -BK\left(1\right) = 0 \end{cases}$$

After some algebraic manipulations[6] we obtain

$$\begin{cases} bk_j & = b_j + \frac{\lambda}{4\pi} = b_j - \frac{b_o + 2\sum_{j=1}^{K} b_j}{1+2K}, \quad j = -K, ..., K \\ \lambda & = -\frac{4\pi}{2K+1}\left(b_o + 2\sum_{j=1}^{K} b_j\right) \end{cases}$$

where $\left\{b_j : j = 0, \pm 1, ...\right\}$ represent the weights of the ideal band-pass filter. The BK optimal weights have a few interesting features. First, they only depend on the weights of the ideal band-pass filter and on the nonnegative integer K that truncates the moving average. This indicates that the BK approximate filter also satisfies the sixth criterion, i.e., easy implementation. Second, the BK weights just differ from the ideal weights by a standardization factor $\frac{\lambda}{4\pi}$ that corrects the latter weights to ensure trend removal from the original series.

[6] For details see for example Everts (2006).

2.1.2 The Christiano and Fitzgerald approximation

Baxter and King's approximation to the filter that ideally isolates business cycle fluctuations has two important flaws; first it is unsuitable for real-time analysis and second does not take into account the true DGP. The approximate filter proposed by Christiano and Fitzgerald (2003) surmounts both limitations by defining the solution as a function of all available data points and using the spectrum of the series to be filtered as a weighting function of the approximation errors. However, their approach is not a complete novelty in the literature. Geweke (1978) and Pierce (1980) have presented the time domain solution to the same problem analyzed in Christiano and Fitzgerald (2003) but in the context of seasonal adjustment. It is shown that the best approximation to the filter is equivalent to apply the ideal filter to the series of interest, but with the particularity that this series is extended with optimal backcasts and forecast when data points are not available. The solution proposed by Geweke (1978) allows for the inclusion of multivariate information but does not deal with unit root processes whereas Pierce (1980) deals with unit roots but only in a univariate context. So, Christiano and Fiztgerald's (2003) contribution is the derivation of the solution to the problem of extracting business cycle fluctuations in real-time in a frequency domain perspective.

They start by assuming a time series decomposition as in equation (2) and set their approximate filter as the solution to T (sample length) projection problems defined as follows

$$\hat{x}_t = P\left(x_t / \Im_T\right), \quad t = 1, 2, ..., T$$

where $\Im_T$ denotes the available information set. The solution is a linear combination of the available data

$$\hat{x}_t = \sum_{j=-f}^{p} cf_{tj} y_{t-j} = CF_t(L) y_t, \quad t = 1, 2, ..., T$$

where $f = T - t$, $p = t - 1$, $\{cf_{tj} : j = -f, ..., p\}$ are the $p + f + 1$ weights of Christiano and Fitzgerald's optimal filter (CF filter) and $CF_t(L) = \sum_{j=-f}^{p} cf_{tj} L^j$. Note that at time $t = T$, $f = 0$ and $\hat{x}_T = \sum_{j=0}^{p} cf_j y_{T-j}$, which means that we have a one-sided filter at the end of the sample and thus feasible to be used in real time analysis. Moreover, since the researchers omit the symmetry assumption of the weights, the CF optimal filter will not have trend elimination properties and so before its application all trends must be removed from the series to be filtered.

For a given t the optimal weights solve the following optimization problem written in the frequency domain:

$$\underset{\{cf_j\}_{j=-f,\dots,p}}{Min} \int_{-\pi}^{\pi} \left| B\left(e^{-i\omega}\right) - CF\left(e^{-i\omega}\right) \right|^2 S_y(\omega) d\omega$$

where $B(z) = \sum_{j=-\infty}^{\infty} b_j z^j$ is the frequency response function of the ideal filter, $CF(z) = \sum_{j=-f}^{p} cf_j^{p,f} z^j$ is the frequency response function of the optimal filter and $S_y(w)$ is the spectrum of y_t. In contrast with the BK objective function, the errors between the ideal filter and the optimal filter are penalized by the spectrum of y_t which implies that the loss function will explicitly depend on the choice of the time series representation. Furthermore, because the weights vary over time the optimization problem must be solved for each sample observation in order to get the T sets of weights needed to extract the desired component.

Christiano and Fitzgerald's MSE filter is derived and analyzed in detail in the working paper version of Christiano and Fitzgerald (2003) for two types of processes. Following their notation let

$$y_t = y_{t-1} + \theta(L)\varepsilon_t$$

where ε_t is a zero mean white noise process with $E(\varepsilon_t^2) = 1$ and $\theta(L) = \theta_0 + \theta_1 L + \theta_2 L^2 + ... + \theta_q L^q$, a finite q order lag polynomial. When $\theta(1) = 0$, y_t is a zero mean covariance stationary process modeled as

$$y_t = y_{t-1} + \theta(L)\varepsilon_t \Leftrightarrow y_t = \frac{\theta(L)}{1-L}\varepsilon_t \Leftrightarrow y_t = \tilde{\theta}(L)\varepsilon_t$$

with $\tilde{\theta}(L)$ a $(q-1)$ lag polynomial. When $\theta(1) \neq 0$, y_t is a unit root process, i.e., difference stationary. Finally, if $\theta(1) \neq 0$ and $\theta(L) = 1$, y_t is a random walk process

$$y_t = y_{t-1} + \varepsilon_t.$$

In general the spectral density function is

$$\begin{aligned} S_y(\omega) &= \frac{1}{2\pi}\left\{\gamma_y(0) + 2\sum_{k=1}^{+\infty}\gamma_y(k)\cos(\omega k)\right\} = \\ &= \frac{1}{2\pi}\frac{\theta(e^{-i\omega})\theta(e^{-i\omega})}{(1-e^{-i\omega})(1-e^{i\omega})}, \qquad -\pi \leq \omega \leq \pi \end{aligned}$$

where $g(\omega) = \theta(e^{-i\omega})\theta(e^{i\omega}) = c_0 + c_1\left(e^{-i\omega} + e^{i\omega}\right) + ... + c_q\left(e^{-i\omega q} + e^{i\omega q}\right)$. Thus $c_{-\tau} = c_\tau$,

$\forall\tau$ and $c_\tau = 0$ for $\tau > q$, reflecting that $\begin{bmatrix} c_0 & c_1 & ... & c_q \end{bmatrix}$ are constants that follow from the covariance function of $\theta(L)\varepsilon_t$.

In the non-stationary case we have to ensure that the filtered series is covariance stationary since no symmetry is forced upon the weights. Accordingly, the optimal filter weights must sum to zero

$$CF(1) = \sum_{j=-f}^{p} cf_j = 0$$

following that[7]

$$CF^*(L) = \frac{CF(L)}{(1-L)}$$

where $CF^*(L) = \sum_{j=-f}^{p-1} cf_j^* L^j = cf_{p-1}^* L^{p-1} + ... + cf_{-f}^* L^{-f}$. Through the method of indeterminate coefficients we obtain a relationship between the weights of both filters

$$cf_j^* = -\sum_{k=j+1}^{p} cf_k, \quad j = p-1, ..., -f$$

or in matrix form

$$cf^* = Q \cdot cf \Leftrightarrow \underbrace{\begin{bmatrix} cf_{p-1}^* \\ cf_{p-2}^* \\ \vdots \\ cf_{-f}^* \end{bmatrix}}_{(p+f)\times 1} = \underbrace{\begin{bmatrix} -1 & 0 & \cdots & 0 & 0 \\ -1 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ -1 & -1 & \cdots & -1 & 0 \end{bmatrix}}_{(p+f)\times(p+f+1)} \cdot \underbrace{\begin{bmatrix} cf_p \\ cf_{p-1} \\ \vdots \\ cf_{-f} \end{bmatrix}}_{(p+f+1)\times 1}$$

[7] If $L = 1$ is a root of the lag polynomial $CF(L)$ then the latter can be factorize as

$$CF(L) = (1-L)CF^*(L) \Leftrightarrow CF^*(L) = \frac{CF(L)}{(1-L)}$$

with $CF^*(L) = \sum_{j=-f}^{p-1} cf_j^* L^j = cf_{p-1}^* L^{p-1} + ... + cf_{-f}^* L^{-f}$.

Adding the constraint and replacing $S_y(w)$ in the objective function follows[8]

$$\underset{\{cf_j\}_{j=-f,\dots,p}}{Min} \int_{-\pi}^{\pi} \left| \widetilde{B}\left(e^{-i\omega}\right) - CF^{*}(e^{-iw}) \right|^2 g(e^{-i\omega}) d\omega$$

$$s.t.\ \ CF(1) = \sum_{j=-f}^{p} cf_j = 0$$

where $\widetilde{B}(z) = \frac{B(z)}{(1-z)}$ with z taken to be a complex number. The FOC's of this minimization problem are

$$\begin{cases} \int_{-\pi}^{\pi} \widetilde{B}\left(e^{-i\omega}\right) g(e^{-i\omega}) e^{i\omega j} d\omega = \int_{-\pi}^{\pi} CF^{*}(e^{-i\omega}) g(e^{-i\omega}) e^{i\omega j} d\omega, & j = p-1, ..., -f \\ CF(1) = cf_p + ... + cf_{-f} = 0 \end{cases}$$

Christiano and Fitzgerald solve this problem by replacing the first $p+f$ conditions by a system of linear equations in cf_j, $j = p, \dots, -f$ such as

$$R(j) - R(j-1) = S(j) - S(j-1), \quad j = p-1, \dots, -f+1 \tag{10}$$

$$R(-f) = S(-f) \tag{11}$$

with $R(j) = \int_{-\pi}^{\pi} \widetilde{B}\left(e^{-iw\omega}\right) g(e^{-i\omega}) e^{i\omega j} d\omega$ and $S(j) = \int_{-\pi}^{\pi} CF^{*}(e^{-i\omega}) g(e^{-i\omega}) e^{i\omega j} d\omega$ for $j = p-1, \dots, -f$. The left side of equation (10) equals

$$\begin{aligned} R(j) - R(j-1) &= \int_{-\pi}^{\pi} B\left(e^{-i\omega}\right) g(e^{-i\omega}) e^{i\omega j} d\omega = \\ &= \begin{cases} 2\pi b_j \theta_0^2 & \text{, if } q = 0 \\ 2\pi \left(b_j c_0 + \sum_{i=1}^{q} \left[b_{|j|+i} + b_{|j|-i}\right] c_i\right) & \text{, if } q > 0 \end{cases} \end{aligned}$$

[8]Note that,

$$\begin{aligned} Q &= \int_{-\pi}^{\pi} \left| B\left(e^{-i\omega}\right) - CF\left(e^{-i\omega}\right) \right|^2 S_y(\omega) d\omega = \\ &= \int_{-\pi}^{\pi} \left| B\left(e^{-i\omega}\right) - CF\left(e^{-i\omega}\right) \right|^2 \frac{g(e^{-i\omega})}{\left(1-e^{-i\omega}\right)\left(1-e^{i\omega}\right)} d\omega = \\ &= \int_{-\pi}^{\pi} \left| \frac{B\left(e^{-i\omega}\right)}{\left(1-e^{-i\omega}\right)} - \frac{CF\left(e^{-i\omega}\right)}{\left(1-e^{-i\omega}\right)} \right|^2 g(e^{-i\omega}) d\omega = \\ &= \int_{-\pi}^{\pi} \left| \widetilde{B}\left(e^{-i\omega}\right) - CF^{*}(e^{-i\omega}) \right|^2 g(e^{-i\omega}) d\omega \end{aligned}$$

for $j = p-1, \ldots, -f+1$ and where $\{b_j\}$ are the ideal weights as in equation (4). While, the right side of equation (10) equals

$$S(j) - S(j-1) = \int_{-\pi}^{\pi} CF(e^{-i\omega}) g(e^{-i\omega}) e^{i\omega j} d\omega = \\ = \begin{cases} 2\pi \cdot cf_j \cdot \theta_0^2 & \text{, if } q = 0 \\ 2\pi \cdot A_j \cdot cf & \text{, if } q > 0 \end{cases}$$

for $j = p-1, \ldots, -f+1$ and where $\{cf_j\}$ are the optimal weights, cf is a $(p+f+1)$ column vector $cf = \begin{bmatrix} cf_p & \cdots & cf_0 & \cdots & cf_{-f} \end{bmatrix}'$ and A_j is a $(p+f+1)$ row vector that involves the constants c_τ, $\tau = 0, ..., q$, and that varies according to the values of j[9]. Finally, equation (11) translates to

$$\int_{\omega_l}^{\omega_u} \left[\frac{e^{-i\omega f}}{1 - e^{-i\omega}} + \frac{e^{i\omega f}}{1 - e^{i\omega}} \right] g(e^{-i\omega}) d\omega = 2\pi \cdot F \cdot Q \cdot cf = 2\pi \cdot A_{-f} \cdot cf$$

where F is a $(p+f)$ row vector $F = \begin{bmatrix} 0 & \cdots & 0 & c_q & c_{q-1} & \cdots & c_0 \end{bmatrix}$ while Q and cf mantain the same definitions as above.

At last, the $(p+f+1)$ optimal weights for the non-stationary case are obtained by solving the following linear system

$$d = 2\pi \cdot A \cdot cf \Leftrightarrow \underbrace{\begin{bmatrix} \int_{-\pi}^{\pi} B\left(e^{-i\omega}\right) g(e^{-i\omega}) e^{i\omega(p-1)} d\omega \\ \vdots \\ \int_{-\pi}^{\pi} B\left(e^{-i\omega}\right) g(e^{-i\omega}) e^{i\omega(-f+1)} d\omega \\ \int_{\omega_l}^{\omega_u} \left[\frac{e^{-i\omega f}}{1-e^{-i\omega}} + \frac{e^{i\omega f}}{1-e^{i\omega}} \right] g(e^{-i\omega}) d\omega \\ 0 \end{bmatrix}}_{(p+f+1)\times 1} = 2\pi \cdot \underbrace{\begin{bmatrix} A_{p-1} \\ \vdots \\ A_{-f+1} \\ A_{-f} \\ \iota \end{bmatrix}}_{(p+f+1)\times(p+f+1)} \cdot \underbrace{\begin{bmatrix} cf_p \\ \vdots \\ cf_{-f+2} \\ cf_{-f+1} \\ cf_{-f} \end{bmatrix}}_{(p+f+1)\times 1}$$

For the stationary case we drop the constraint imposed over the zero frequency of the frequency response function. So, only small adjustments are done to adapt the linear system to the stationary case. In detail, the last two rows of A are replaced by A_p and A_{-f}, respectively, and the last two rows of d are replaced by $\int_{-\pi}^{\pi} B\left(e^{-i\omega}\right) g(e^{-i\omega}) e^{i\omega l} d\omega$ for $l = p$ and $-f$.

Finally, assuming that y_t is a random walk process means solving the optimization problem

[9] For further details on this row vectors see the working paper version of Christiano and Fitzgerald (2003).

for the non-stationary case setting $q = 0$ and $\theta_0 = 1$. These assumptions imply

$$
\begin{aligned}
y_t &= y_{t-1} + \varepsilon_t \quad \text{with } E(\varepsilon_t^2) = 1 \\
S_y(\omega) &= \frac{1}{2\pi} \frac{1}{(1 - e^{-i\omega})(1 - e^{i\omega})}, \qquad -\pi \leq \omega \leq \pi
\end{aligned}
$$

and

$$
\underset{\{cf_j\}_{j=-f,\dots,p}}{Min} \int_{-\pi}^{\pi} \left| \widetilde{B}\left(e^{-i\omega}\right) - CF^*(e^{-i\omega}) \right|^2 d\omega
$$

$$
s.t. \;\; CF(1) = \sum_{j=-f}^{p} cf_j = 0
$$

From this minimization problem we get a closed form expression for the optimal weights which simplifies the filtering procedure:

$$
cf_j = \begin{cases} \frac{1}{2} b_0 - \sum_{k=0}^{j-1} b_k & , \;\; j = p \\ b_j & , \;\; j = p-1, ..., -f+1 \;, \\ \frac{1}{2} b_0 - \sum_{k=j+1}^{0} b_k & , \;\; j = -f \end{cases} \qquad t = 1, ..., T
$$

where $p = t - 1$, $f = T - 1$ and $\{b_j\}$ stands for the weights of the ideal band-pass filter given in equation (4). It is worth noting that most optimal weights are equal to the ideal weights and that only the highest order terms are adjusted to ensure the stationarity of the filtered series. This particular filter is known in the literature as the optimal random walk filter. The random walk filter is a symmetric filter when $p = f$. First note that $cf_j = b_j$ for $j = p - 1, ..., 1 - p$ and that the ideal weights are symmetric. Secondly, it is easily proven that $cf_p = cf_{-p}$ replacing j by p and $-p$ in the expressions above. For $p = f$ and p fixed we have

$$
cf_j^{p,p} = b_j - \frac{1}{1+2p}\left(b_0 + 2\sum_{j=1}^{p} b_j\right), \quad j = 0, \pm 1, ..., \pm p
$$

which amounts to the Baxter and King optimal weights. Based on the filter weights of the random walk case, solutions to time series representations close to random walk are easily handled. For instance, in an i.i.d. case we have $S_y(w) = 1$ which means that we are in the BK case without constraints (assuming $p = f$ and p fix) and so the optimal weights are just a truncated version of the ideal weights. A near i.i.d. case corresponds to the BK case when we impose the stationarity condition of the filtered series. In this case the optimal weights are the ideal weights truncated and adjusted by a constant in order to verify $BK(1) = 0$ and $bk_j = bk_{-j} \;\; \forall j$.

Although the CF filter is made optimal according to the DGP, Christiano and Fitzgerald concluded that the pure random walk filter is nearly optimal even if the random walk process is not the true time series representation. A reason that can justified this result is the fact that many of the U.S. time series exhibit what is called a typical Granger spectral shape, in other words most series exhibit a low dominating spectrum. Nevertheless, gains are always achieved by using the true time series model in the optimization problem but they can be rather small when compared with the effort needed to run a model choice procedure or to implement the optimal procedure with the true DGP. Finally, they also stress that important efficiency gains are attained using all information available at each time t.

2.1.3 The Multivariate approximation

For the same frequency extraction problem of Christiano and Fitzgerald (2003), Valle e Azevedo (2007) has developed a multivariate approximation. Basically, he expands the CF univariate solution adding an arbitrary number of covariates to the solution. The idea is that those additional covariates will help to improve the estimate of the desired frequency component, in particular at the endpoints where the estimates are frequently poor.

Again, assume a decomposition for y_t as in equation (2) and therefore the filter that perfectly isolates the components of y_t in the $[2\pi/w_u, 2\pi/w_l]$ interval of periodicities will be the, already discussed, ideal band-pass filter. The problem amounts to obtaining an approximation to such filter since it requires an infinite amount of data. Similarly to the solution of Christiano and Fitzgerald, Valle e Azevedo defines an approximation as a weighted sum of past and future values of the series of interest but adds a weighted sum of past and future values of n covariance stationary variables, $z_1, ..., z_n$. This approach allows to exploit information from other variables that may help to predict the signal of interest. In detail, the multivariate approximation is expressed as

$$\hat{x}_t = \sum_{j=-f}^{p} \hat{B}_j y_{t-j} + \sum_{s=1}^{n} \sum_{j=-f}^{p} \hat{R}_{s,j} z_{s,t-j} = \hat{B}(L) y_t + \sum_{s=1}^{n} \hat{R}_s(L) z_{st} \tag{12}$$

where p denotes the number of past observations, f the number of future observations used from the available information set at each time t and $\left\{\hat{B}_j; \hat{R}_{s,j} : j = -f, \ldots, p; s = 1, \ldots, n\right\}$ are the $(n+1) \times (p+f+1)$ weights of the optimal multivariate filter. Such weights are chosen in such a way that the minimum MSE estimate of the ideally filtered series at time t is obtained, i.e.,

$$\underset{\left\{\hat{B}_j; \hat{R}_{1,j}; \ldots; \hat{R}_{n,j}\right\}_{j=-f,\ldots,p}}{Min} E[(x_t - \widehat{x}_t)^2].$$

The frequency domain formulation of the problem is now slightly different since it involves a vector process[10]

$$\underset{\left\{\hat{B}_j; \hat{R}_{1,j}; \ldots; \hat{R}_{n,j}\right\}_{j=-f,\ldots,p}}{Min} \int_{-\pi}^{\pi} \beta(\omega) S_{y,z_1,\ldots,z_n}(\omega) \beta'^{*}(\omega)$$

where $\beta(\omega) = \left[B\left(e^{-i\omega}\right) - \hat{B}\left(e^{-i\omega}\right) \quad -\hat{R}_1\left(e^{-i\omega}\right) \quad \cdots \quad -\hat{R}_n\left(e^{-i\omega}\right)\right]$, $B(z) = \sum_{j=-\infty}^{\infty} B_j z^j$, $\hat{B}(z) = \sum_{j=-f}^{p} \hat{B}_j z^j$, $\hat{R}_s(z) = \sum_{j=-f}^{p} \hat{R}_j z^j$ for $s = 1, \ldots, n$ and $S_{y,z_1,\ldots,z_n}(w)$ denotes the $(n+1) \times (n+1)$ spectral matrix of the vector process $\left[y \quad z_1 \quad \cdots \quad z_n\right]'$.

The solution to this multivariate optimization problem is derived in Valle e Azevedo (2007) assuming a M finite order moving average (MA(M)) representation for the vector process, as in

[10] Considering $\mathbf{W_t}$ a covariance stationary vector then its spectrum is

$$S_{\mathbf{W}}(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \mathbf{\Gamma}(k) e^{-i\omega k}, \quad -\pi \leq \omega \leq \pi$$

where $\mathbf{\Gamma}(k)$ denotes the covariance matrix, which can be recovered from the spectrum by

$$\mathbf{\Gamma}(k) = \int_{-\pi}^{\pi} S_{\mathbf{W}}(\omega) e^{i\omega k} d\omega, \qquad k = 0, \pm 1, \pm 2, \ldots$$

Hence, the spectrum of the filtered series $M_t = \sum_{j=-\infty}^{\infty} H_j \mathbf{W}_{t-j}$ is

$$S_M(\omega) = H(e^{-i\omega}) S_{\mathbf{W}}(\omega) H'(e^{i\omega}), \quad -\pi \leq \omega \leq \pi$$

Note that it generalizes the univariate case. Suppose that W_t is a univariate covariance stationary process then

$$\begin{aligned} S_M(\omega) &= H(e^{-i\omega}) S_W(\omega) H'(e^{i\omega}) = \\ &= H(e^{-i\omega}) H(e^{i\omega}) S_W(\omega) = \\ &= \left|H(e^{-i\omega})\right|^2 S_W(\omega), \quad -\pi \leq \omega \leq \pi \end{aligned}$$

Christiano and Fitzgerald (2003). Nonetheless, the case of non-stationarity of the series to be filtered is also handled. If y_t is a unit-root process then to ensure stationarity of the filtered series the previous problem is constrained with

$$\hat{B}(1) = 0 \Rightarrow \hat{B}(z) = (1-z)b(z) \Leftrightarrow b(z) = \hat{B}(z)/(1-z)$$

where $b(z) = \sum_{j=-f}^{p-1} b_j L^j = b_{p-1}z^{p-1} + ... + b_0 + b_{-f}z^{-f}$ and so the objective function is modified to

$$\underset{\{\hat{B}_j;\hat{R}_{1,j};...;\hat{R}_{n,j}\}_{j=-f,...,p}}{Min} \int_{-\pi}^{\pi} \alpha(\omega) S_{\Delta y,z_1,...,z_n}(\omega)\alpha(-\omega)'$$

with $\alpha(\omega) = \begin{bmatrix} \bar{B}\left(e^{-i\omega}\right) - b\left(e^{-i\omega}\right) & -\hat{R}_1\left(e^{-i\omega}\right) & \cdots & -\hat{R}_n\left(e^{-i\omega}\right) \end{bmatrix}$, $\bar{B}(z) = B(z)/(1-z)$ and $S_{\Delta y,z_1,...,z_n}(\omega)$ is the spectrum of the vector process $\begin{bmatrix} \Delta y & z_1 & \cdots & z_n \end{bmatrix}'$.

Define W_j as a $(n+1)$ column vector $\begin{bmatrix} b_j & \hat{R}_{1,j} & \cdots & \hat{R}_{n,j} \end{bmatrix}'$, $j = p-1, ..., -f$, then the FOC's are

$$\begin{cases} \frac{\partial \mathfrak{L}}{\partial W_j} = 0 \Leftrightarrow \int_{-\pi}^{\pi} e^{-i\omega j} \begin{bmatrix} S_{\Delta y}(\omega) \\ S_{z_1,\Delta y}(\omega) \\ \vdots \\ S_{z_n,\Delta y}(\omega) \end{bmatrix} \bar{B}\left(e^{-i\omega}\right) d\omega = \int_{-\pi}^{\pi} e^{-i\omega j} S_{\Delta y,z_1,...,z_n}(\omega) \begin{bmatrix} b\left(e^{i\omega}\right) \\ \hat{R}_1\left(e^{i\omega}\right) \\ \vdots \\ \hat{R}_n\left(e^{i\omega}\right) \end{bmatrix} dw \\ \frac{\partial \mathfrak{L}}{\partial \lambda} = 0 \Leftrightarrow \hat{B}(1) = 0 \end{cases}$$

for $j = p-1, \ldots, -f$.

Following the strategy of Christiano and Fitzgerald the $(n+1)\times(p+f)$ first equations of the FOC's are replaced by other equations. Define the left hand side as

$$S_j = \int_{-\pi}^{\pi} e^{-i\omega j} \bar{S}_{\Delta y,z_1,...,z_n}(\omega) \bar{B}\left(e^{-i\omega}\right) d\omega, \qquad j = p-1, ..., -f$$

then

$$S_j - S_{j-1} = \begin{bmatrix} B_j \gamma_{\Delta y}(0) + \sum_{i=1}^{M} \left(B_{j+i} + B_{j-i}\right) \gamma_{\Delta y}(i) \\ \vdots \\ B_j \gamma_{z_n,\Delta y}(0) + \sum_{i=1}^{M} \left(B_{j-i}\gamma_{z_n,\Delta y}(i) + B_{j+i}\gamma_{z_n,\Delta y}(i)\right) \end{bmatrix} = V_j$$

for $j = p-1, ..., -f+1$. The S_j are then obtained recursively by

$$S_j = S_{j-1} + V_j, \qquad j = p-1, ..., -f+1.$$

For $-f$ we evaluate the following integral numerically

$$S_{-f} = \int_{-\omega_u}^{-\omega_l} \frac{e^{i\omega f}}{1 - e^{i\omega}} \bar{S}_{\Delta y, z_1, ..., z_n}(\omega) d\omega + \int_{\omega_l}^{\omega_u} \frac{e^{i\omega f}}{1 - e^{i\omega}} \bar{S}_{\Delta y, z_1, ..., z_n}(\omega) d\omega.$$

The right hand side of the FOC's is replaced by

$$R_j = \int_{-\pi}^{\pi} e^{-i\omega j} S_{\Delta y, z_1, ..., z_n}(\omega) \begin{bmatrix} b\left(e^{i\omega}\right) \\ \hat{R}_1\left(e^{i\omega}\right) \\ \vdots \\ \hat{R}_n\left(e^{i\omega}\right) \end{bmatrix} d\omega = Q_j \hat{W}, \qquad j = p-1, ..., -f$$

with $Q_j = \begin{bmatrix} Q_{\Delta y,j} & Q_{\Delta y z_1,j} & \cdots & Q_{\Delta y z_n,j} \\ \vdots & \vdots & \ddots & \vdots \\ Q_{z_n \Delta y,j} & Q_{z_n z_1,j} & \cdots & Q_{z_n,j} \end{bmatrix}$ [11] and $\hat{W} = \begin{bmatrix} \hat{B} & R_1 & \cdots & R_n \end{bmatrix}'$ a column vector with all weights stacked.

Finally, only $n+1$ conditions that correspond to the case $j = p$ are missing. The weights $\left\{\hat{R}_{1,p}; ...; \hat{R}_{n,p}\right\}$ are obtained from $\tilde{S}_p = \tilde{Q}_p \hat{W}$, where $\tilde{S}_p$ equals S_p but with the first row deleted and $\tilde{Q}_p$ equals Q_p with the first row deleted. The final equation is given by the restriction impose at the beginning, $\hat{B}(1) = 0$, which gives $\hat{B}_p$.

So, the optimal weights $\hat{W}$ solve a linear system with $(n+1) \times (p+f+1)$ equations

$$V = Q\hat{W} \Leftrightarrow \hat{W} = Q^{-1}V \tag{13}$$

[11] For further details on the definition of the Q_j matrix see Valle e Azevedo (2007) Appendix A.

with $V = \begin{bmatrix} S_{-f} & \cdots & S_{p-1} & \tilde{S}_p & 0 \end{bmatrix}'_{[(n+1)\times(p+f+1)]\times 1}$,
$Q = \begin{bmatrix} Q_{-f} & \cdots & Q_{p-1} & \tilde{Q}_p & U \end{bmatrix}'_{[(n+1)\times(p+f+1)]\times[(n+1)\times(p+f+1)]}$,
$\hat{W} = \begin{bmatrix} \hat{B} & R_1 & \cdots & R_n \end{bmatrix}'_{[(n+1)\times(p+f+1)]\times 1}$ and $U = \begin{bmatrix} 1 & \cdots & 1 & 0 & \cdots & 0 \end{bmatrix}$.

To accommodate the case in which all data points are used, one just sets $p = t-1$ and $f = T-t$ and solves the linear system in equation (13) T times, like in Christiano and Fitzgerald. In the stationary case no restriction is imposed and some adjustments are done to the V and Q matrices

$$V = Q\hat{W} \Leftrightarrow \hat{W} = Q^{-1}V$$

with $V = \begin{bmatrix} S_{-f} & \cdots & S_{p-1} & S_p \end{bmatrix}'$, $Q = \begin{bmatrix} Q_{-f} & \cdots & Q_{p-1} & Q_p \end{bmatrix}'$ and the spectrum is from $S_{y,z_1,\ldots,z_n}(w)$ because now it is implied that y_t is a covariance stationary process. The optimal weights, i.e., the solution of the optimization problem depends (a) on the second order moments of $\begin{bmatrix} \Delta y, z_1, \ldots, z_n \end{bmatrix}$ or $\begin{bmatrix} y, z_1, ..., z_n \end{bmatrix}$, due to the presence of the spectrum in the objective function and (b) on the ideal weights as given in equation (4). Moreover, this filter does not remove stochastic or deterministic trends due to the asymmetrical weights.

As pointed out in the context of Christiano and Fitzgerard's filter, this type of solution cannot be seen as completely new in the literature given the studies of Geweke (1978) and Pierce (1980). But this particular approach has an important contribution since it gives the extension to the multivariate case of the solution obtained when filtering a unit root process.

2.1.4 The Hodrick and Prescott filter

In order to analyze the cyclical component of some macroeconomic variables, Hodrick and Prescott (1980, 1997) developed a detrending method based on the idea that a time series is the sum of two unobservable and uncorrelated components:

$$y_t = g_t + c_t, \quad t = 1, ..., T \tag{14}$$

where T is the sample size, g_t denotes a growth component, related to the long run movements of the series y_t, and c_t denotes a cyclical plus irregular component, defined as the deviations around the former component. So, given an estimate of the trend the so called cyclical fluctuation is simply the residual series. The estimate of the trend is obtained by minimizing the following cost function

$$\underset{\{g_t\}_{t=1,\ldots,T}}{Min} \sum_{t=1}^{T} (y_t - g_t)^2 + \lambda \sum_{t=2}^{T-1} [(g_{t+1} - g_t) - (g_t - g_{t-1})]^2 . \tag{15}$$

Since the average of the deviations from the trend is expected to be near zero for long time periods, the first term of the objective function sums the squares of those deviations. The second term sums the variation in the second difference of the growth component, $\left[(1-L)^2 g_t\right]$ where L denotes the lag operator, which gives the variability of the growth component (trend growth rate). Hodrick and Prescott state that this measures the smoothness of the long run component, expected to evolve smoothly over time, and thus penalize this term by a positive number λ, known as the smoothness parameter. The larger λ is the smoother is the growth component. If $\lambda = 0$ then the first term of the objective function in equation (15) has to be zero in order to minimize the function. As a result, the growth component equals the original series and the cyclical movements do not exist. In the limit, when $\lambda = \infty$, we have perfect smoothing, the growth component is simply a linear trend, and we obtain the maximum cyclical fluctuations. The smoothing parameter is not determined by optimization but derived in Hodrick and Prescott (1980, 1997) from a probability model. They show that if the cycle component and the second difference of the trend are zero mean i.i.d. normally distributed variables then the parameter equals the variance of the business cycle component divided by the variance of the acceleration in the trend component, $\lambda^{\frac{1}{2}} = \sigma_c / \sigma_{\Delta^2 g}$. From this formula results the standard value $\lambda = 1600$ for quarterly data. For other data frequencies some correction has to be done due to alterations in the variability of the series. Adjustment rules have been suggested by Backus and Kehoe (1992), Correia et al. (1992), Cooley and Ohanian (1991) and Ravn and Uhlig (2001). Alternatively, in the working paper version of Canova (1998) the smoothing parameter is understood as a signal extraction coefficient and is estimated by maximum likelihood. Also, Harvey and Jaeger (1993) attempt to estimate the parameter using maximum likelihood methods.

The FOC's from equation (15) are

$$\begin{aligned} -2(y_t - g_t) + 2\lambda [(g_t - g_{t-1}) - (g_{t-1} - g_{t-2})] - 4\lambda [(g_{t+1} - g_t) - (g_t - g_{t-1})] \\ + 2\lambda [(g_{t+2} - g_{t+1}) - (g_{t+1} - g_t)] = 0, \quad t = 1, \ldots, T \end{aligned} \tag{16}$$

Since, in equation (16) there are forward and backward differences, we can rewrite the FOC's using the backshift and forwardshift operator, L^j and L^{-j}, respectively, i.e.,

$$\begin{aligned} y_t &= \left[\lambda(1-L)^2(1-L^{-1})^2+1\right] g_t \Leftrightarrow \\ &\Leftrightarrow \quad g_t = \frac{1}{1+\lambda(1-L)^2(1-L^{-1})^2} y_t \Leftrightarrow \\ &\Leftrightarrow \quad g_t = G(L)y_t, \quad t=1,\dots,T \end{aligned}$$

Thus, the growth component is obtained from the original series by applying a linear filter to the raw data. The filter $G(L)$ is known as the HP growth filter and essentially consists in a low pass filter since it aims to extract only the long run movements associated to low frequencies. Straightforwardly, the HP cyclical filter is obtained as,

$$c_t = y_t - g_t = [1-G(L)]\, y_t = C(L)y_t$$

where $C(L) = \left[\lambda(1-L)^2(1-L^{-1})^2\right] / \left[1+\lambda(1-L)^2(1-L^{-1})^2\right]$. The $C(L)$ filter is a high pass filter that renders stationary any integrated process up to fourth order and that induces no phase shift due to symmetry.

Unlike the discussed optimal filters, the growth or cyclical HP filter is not presented, in the original framework, as an optimal solution to a frequency extraction problem. But, King and Rebelo (1993), section 4, show that for an infinite sample and a particular class of models the HP cyclical filter is the optimal approximation, in the MSE sense, to an ideal high pass filter. Due to this interpretation and the broadly use of the HP method for detrending time series, it is common practice to use this filtering technique as a benchmark when evaluating the performance of alternative procedures (see Baxter and King, 1999, Christiano and Fitzgerald, 2003 or Valle e Azevedo, 2007).

Assuming that the weights are square summable there is a valid Fourier transform of the cyclical HP filter. King and Rebelo (1993) showed that the transfer function[12] is given by

$$C(w) = \frac{4\lambda\left[1-\cos(w)\right]^2}{1+4\lambda\left[1-\cos(w)\right]^2}, \quad -\pi \le w \le \pi.$$

Hence, the frequency domain approach stresses the trend elimination properties of the filter because zero weight is placed on the zero frequency, $C(0) = 0$, and the fact that it does not

[12]Since de cyclical HP filter is symmetric the transfer function equals the frequency response function.

reweight the high frequencies since almost unit weight is placed on those frequencies.

This tool gained popularity given its use in the context of Real Business Cycle (RBC) models. In this context, the HP filter is used in a preliminary step to eliminate the trend component of several variables. The choice of the HP filter as a detrending method follows from its flexibility to incorporate (through the smoothness parameter) the researchers' preferences regarding the path of the trend.

2.2 Comparison

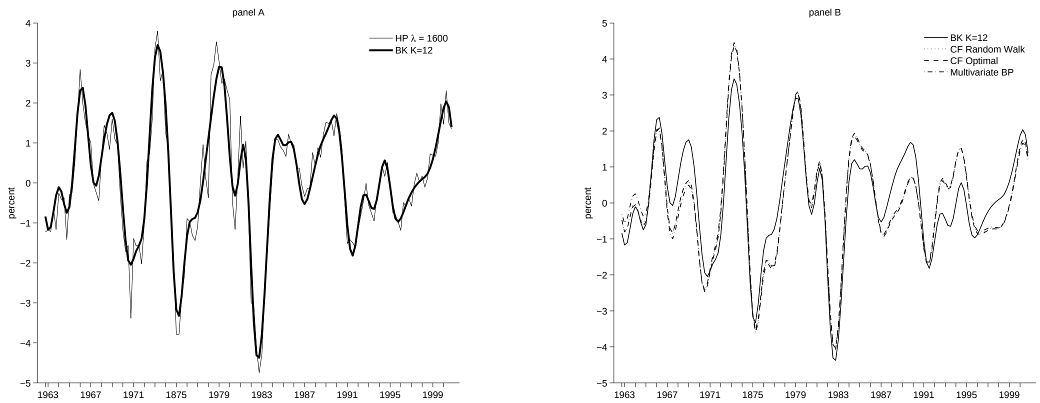
In this subsection we examine the limitations of the discussed approximate filters and simultaneously document their differences and similarities. This exercise is important to establish which filters may be used to construct a real time indicator (our main objective). We will focus our analysis on the filters that aim to approximate business cycle fluctuations, simply because these were the kind of filters that we discussed in more detail in the previous subsections. However, most of the conclusions are also valid for smooth growth approximate filters.

The HP and BK approximations are optimal in MSE sense under a particular set of conditions, hardly met in our particular application to GDP series. In the case of the HP filter such conditions imply assuming that the trend of a time series is integrated of order two and that the cyclical fluctuations follow a white noise process. In the case of the BK filter the implicit assumption considers that the time series to be filtered follows an i.i.d. process.

Harvey and Jaeger (1993), Cogley and Nason (1995) and Guay and St-Amant (1997) state that these filters, or to be rigorous the HP filter, will perform adequately when the spectrum of the original series has a peak at business cycle frequencies and will perform poorly when the spectrum is dominated by low frequencies (typical Granger shape). The idea follows from the fact that when filtering integrated or highly persistent (but stationary) processes it is common to detect a peak in the spectrum of the filtered series at the lowest frequency of the band of interest. These researchers take this as evidence of spurious cycles and alert to the danger of obtaining misleading conclusions when applying a band-pass filter to such processes. However, that is not true as argued by Pedersen (2001) and Valle e Azevedo (2007). They show that we can interpret the effects of filtering integrated time series in the same way as we do for stationary processes. So, the peak arises just because we cut a (pseudo-) spectrum that has most of its power around the zero frequency. The variance of the cycle component will consequently be concentrated around the lowest frequency and no spurious cycles are in fact induced by applying these filters to integrated time series. Intuitively, if the spectrum of a time series is dominated by low frequencies then it is hard to identify the business cycle fluctuations in the original series. Automatically the peaks in the spectrum of the filtered series are interpreted as distorted cycles when in fact they represent fluctuations that are definitely present in the time series.

Secondly, the HP filter is mainly a detrending method and less orientated to the measurement of business cycle fluctuations given the adopted definition. The so-called cyclical component of Hodrick and Prescott is obtained from the difference between the observed time series and the estimated growth component. Thus, unlike all the other business cycle approximation filters, discussed in the previous subsections, the noise component connected to very high frequencies is not removed but absorbed by the cyclical component. This might explain why in figure 2 panel A the HP approximation to business cycle fluctuations of real GDP is more ragged than the alternative approximations.

Figure 2: Approximations to the business cycle fluctuations of U.S. real GDP using alternative filtering techniques.



In fact the extracted components are not directly comparable because we are extracting different types of information from the data. This reveals that the definition of business cycle fluctuations is arbitrary and can be model-dependent. Moreover, if the HP cyclical filter retains both medium and high frequencies from a time series then it is expected to better approximate an ideal high pass filter than an ideal band-pass filter. Figures 3 (panel B) and 4 show that it is in fact the case. For the standard value of $\lambda = 1600$ the HP cyclical filter contains minor problems of leakage and compression and so is considered as a good approximation to the high pass filter with cutoff frequency $\omega = \frac{2\pi}{32}$. Instead, the trend filter will retain only the long periodicities, suggesting a good approximation to an ideal low pass filter as revealed by figure 3 panel A.

Figure 3: Panel A. Gain function of the ideal low pass filter that retains cycles of length lower than 32 quarters against the gain function of the Hodrick and Prescott growth filter with $\lambda = 1600$; Panel B. Gain function of the ideal high pass filter that retains cycles of length higher than 32 quarters against the gain function of the Hodrick and Prescott cyclical filter with $\lambda = 1600$.

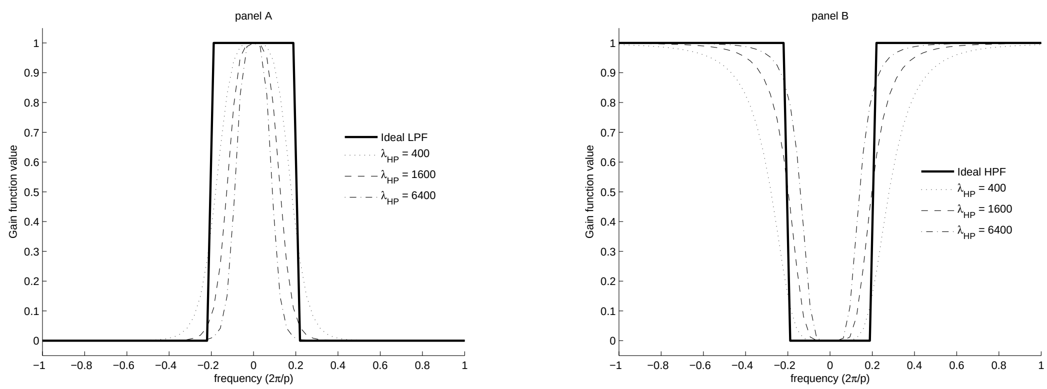
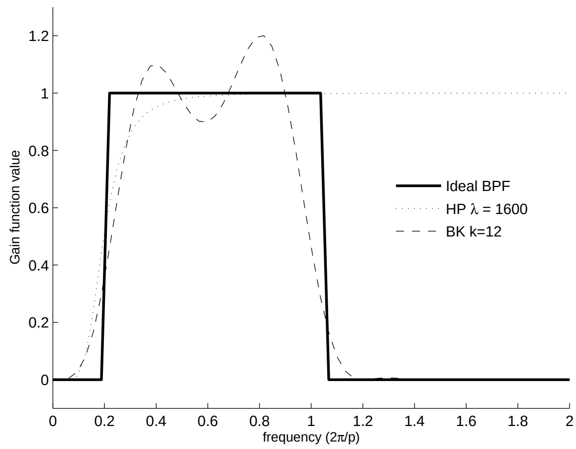
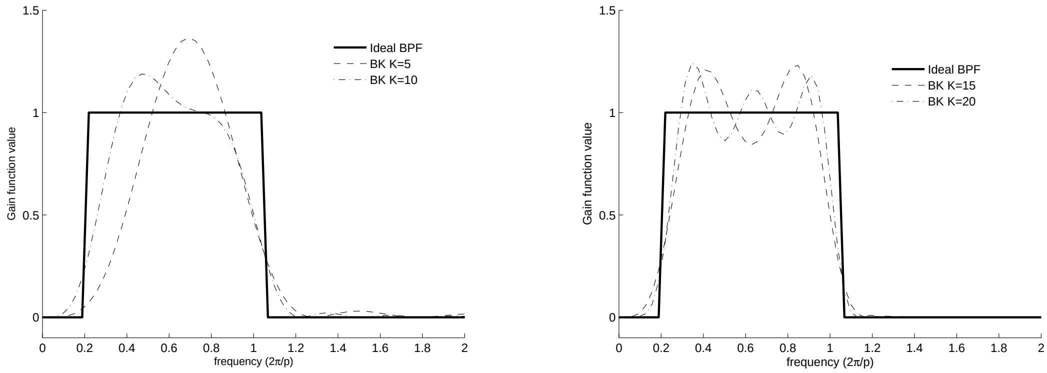


Figure 4: Gain function of the ideal band-pass filter passing frequencies in the range $\frac{2\pi}{32} \leq |\omega| \leq \frac{2\pi}{6}$ against the gain function of the Hodrick and Prescott cyclical filter with $\lambda = 1600$ and the Baxter and King filter truncated at lag and lead 12.



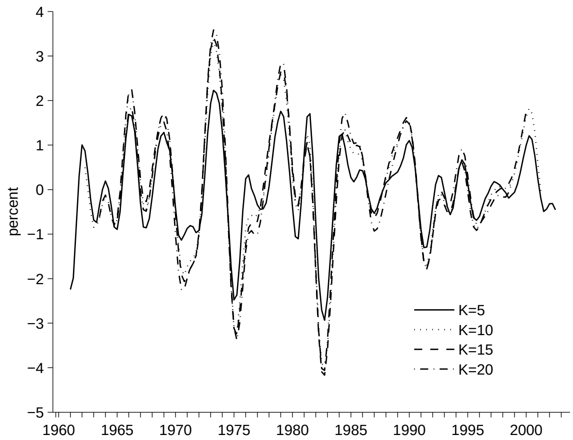
The band-pass approximation suggested by Baxter and King mainly depends on the value K (the maximum lag length of the MA filter). Figures 5 and 6 illustrate the effect of changes in the truncation of the infinite order MA filter on the gain function of the optimal filter and on the filtered series, respectively. The panels of the first figure show that the problems of leakage, outside the specified range of periodicities, and of both exacerbation and compression, within the specified range, tend to become less severe as the maximum lag length gets larger. In fact, the oscillations around the ideal value of the gain function, between the cutoff frequencies, seem to shrink as K increases. So, the BK filter is a reasonable approximation to the band-pass filter when the value of K is sufficiently large. The definition of sufficiently large depends on the sampling frequency and on the data properties. Furthermore, figure 6 shows that the shape of the approximations is not significantly affected by the different values of K while the size of the deviations from the trend seems to increase, until a threshold, as K gets larger.

Figure 5: Gain function of the ideal band-pass filter passing frequencies in the range $\frac{2\pi}{32} \leq |\omega| \leq \frac{2\pi}{6}$ against the gain function of the Baxter and King filter for $K = 5, 10, 15, 20$.



For a final comparison between the HP and BK filters' gains we analyse figure 4 that displays the gain function of the ideal band-pass filter with $\omega_l = \frac{2\pi}{32}$ and $\omega_h = \frac{2\pi}{6}$ as cutoff frequencies against the gain functions of the HP cyclical filter with $\lambda = 1600$ and of the BK filter with $K = 12$. Given the previous discussion it is not surprising that the BK gain function resembles more the ideal gain function than the HP gain function outside the specified frequency band. For periodicities of more than 6 quarters the HP gain substantially deviates from the ideal gain, but for periodicities lower than 32 quarters the problems of leakage are of the same size as those of the BK filter. Nevertheless, any weightning of undesirable frequencies, even if minor as seems to be the case, may lead to distortions of the true cycle. Within the specified range of periodicities, the HP gain exhibits a small problem of compression near the inferior cutoff frequency while the BK gain varies considerable, above and below the ideal value of 1. This investigation of the HP and BK gain functions values to comprehend their sensibility to the values of λ and K, respectively.

Figure 6: Approximation to the business cycle fluctuations of U.S. real GDP using the Baxter and King approximate filter with lag lengths of 5, 10, 15 and 20 quarters.



Thirdly, several studies document that the smoothing parameter present in the HP filter should fit the data in terms of frequency and intrinsic properties, but theory provides little guidance as to what this value should be. Thus, in applied work the degree of smoothness is simply a matter of choice subject therefore to individual judgements. An inadequate choice of λ may lead to attribute variability to the cyclical component that in fact is part of the trend component. A similar caveat arises in the context of the BK optimal filter given the requirement of a prior choice of the K value (and also of the frequency band of interest). Evidently that increasing K enhances the approximation to the ideal filter but also implies loosing more observations at both endpoints of the sample. As before, there is no good rule for setting K but the advice is obvious: weight the trade off from the opposing factors when setting the truncation point of the infinite order MA filter. From simulation results, Baxter and King (1999) suggest, as guideline, *that researchers use moving averages based on three years of past data and three years of future data, as well as the current observation*, for quarterly and annual U.S. data.

Finally, the BK filtering procedure implies, as already mentioned, loosing a total of $2K$ observations, where K is the maximum length of the finite MA filter. As a result, the BK filter cannot be used for real time analysis. Even so, some ideas appear in the literature to overcome this particular limitation. Van Norden (2004) suggested without much success to use BK filter for the full sample length replacing by zero the missing observations. While Stock and Watson (1998) suggested to expand the sample with K forecasts and K backcasts generated from a suitable time series model. Moreover, this feature of the BK approximations may suggest that the HP approximations might be preferred over the first because the HP filter generates estimates for both components at the beginning and end of the sample. The finite version of the HP filter in time domain representation, derived in Hodrick and Prescott (1997), shows that the cyclical component at each moment t can be obtained through a linear filter with time-varying weights involving all available data points. The problem, is that the first and last filtered observations have very poor properties given the one-sided nature of the applied filter, as documented in Baxter and King (1999) and others. But this difficulty is shared with all methods that aim at obtaining estimates at the endpoints. In the analysis of HP filtered series typically some observations are disregarded but in that case we stay in the same situation of Baxter and King, no available real-time estimates.

Alternatives to the approaches discussed so far are the CF filter and the multivariate filter. Unlike HP and BK filters, these filters are made optimal according to the DGP. This implies different spectra in the objective function for each particular case. Nevertheless, the BK optimization problem is a special case of the CF optimization problem, and in turn the last is a special case of the multivariate filter. To obtain the CF filter from the multivariate problem the second term in the right side of equation (12) is dropped. So, the main difference between the two filters is that the multivariate method exploits additional information from a wide set of variables in order to get a more accurate estimate of the desired component, in particular at the endpoints. Theoretically, as Valle e Azevedo (2007) states better estimates will be obtained by applying the multivariate filter if the true second order moments were known. Since they have to be replaced by estimates it is not clear what happens.

Secondly, the CF and multivariate approximations are flexible enough to easily handle the use of all available data points at each time t by setting $p = t - 1$ and $f = T - t$. In that case the optimal weights will vary over time allowing assessment of the characteristics of the estimated signal in real time. But on the other hand, this introduces instability in the estimates of the endpoints, which will vary every time new data becomes available just like the HP estimates. Moreover, a different amount of output data results from the application of the BK method and the CF or multivariate filter. The first only computes approximations at the middle of the sample while the latter computes approximations for all data points.

At last, the asymmetry of the CF and of the multivariate weights implies that both methods do not have trend elimination properties and may induce phase shifts. The removal of the trend from the original series before applying the filter might be seen as a another drawback if it is unclear whether a stochastic or determinist trend is present in the time series. Nevertheless, we argue that it is a minor problem when compared with the possible time displacement of the components. Christiano and Fitzgerald (2003) argue however that "*the degree of asymmetry and non-stationarity in the optimally filtered data is quantitatively small*".

The multivariate approach, in particular, has also a few more pitfalls. First, it requires data other than the series of interest. Such task combined with the need to run a variable choice procedure can be cumbersome in most applications, given the extra work. Secondly, the addition of a new variable to the covariate set implies the estimation of more $p+f+1$ parameters in each iteration and more than doubles the number of necessary estimates of second order moments.

All methods have their limitations but whether they are relevant in a specific case is an empirical question. Nevertheless, these filters have been applied to real data. The HP filter is commonly used as detrending method in real business cycle models while the BK filter has being applied mainly to confirm stylized facts. Important references include Baxter (1994), King, Stock and Watson (1995) and Cecchetti and Kashyap (1995) which study the relation among variables at different periodicities. Christiano and Fitzgerald (2003) provided an application of their approximate band-pass filter to investigate the correlation between M2 money growth and CPI inflation for U.S. data at different frequency bands. Finally, the multivariate filter is a very recent method and therefore little as been done regarding real data applications. In fact this will be one of the contributions of this work. Nevertheless, Valle e Azevedo (2007) illustrated the use of this approach to extract the business cycle component of GDP time series based on a set of other variables.

Regardless of the extensive application of the HP and BK filters in the measurement of trends and cycles we will only analyze the performance of the CF optimal filter and of the multivariate filter as real-time indicators. This choice is supported by three main conclusions; first the BK filter is useless for real-time analysis, second the HP filter may not extract from the data exactly the same kind of information that we desire and third empirical evidence shows that the CF filter and the multivariate filter perform better than the alternatives (see Christiano and Fitzgerald, 2003 and Valle e Azevedo, 2007).

3 Estimation of moments and factor models

Before turning to the empirical application we discuss a few technical details related to the implementation of the filters. Specifically, we discuss spectrum estimators and factor models.

3.1 Estimation of second order moments

The optimal solutions of Christiano and Fitzgerald and Valle e Azevedo exploit the second order properties (or the spectrum) of Δy_t in the first case and of the vector $(\Delta y_t, z_{1t}, \ldots, z_{nt})'$ in the second case[13]. In practice this and other population moments are unknown and need to be estimated. For this purpose, two different estimation methods[14] are suggested.

The non-parametric spectrum estimator of a covariance stationary process is well-known and given by:

$$\widehat{S}_{\Delta y}(\omega) = \frac{1}{2\pi}\left[\hat{\gamma}(0) + 2\sum_{k=1}^{M(T)} \kappa(k,T)\hat{\gamma}(k)\cos(\omega k)\right] \quad \text{(univariate)}$$

$$\widehat{S}_{\Delta y, z_1, \ldots, z_n}(\omega) = \frac{1}{2\pi}\left[\hat{\Gamma}(0) + \sum_{k=1}^{M(T)} \kappa(k,T)\left(\hat{\Gamma}(k)e^{i\omega k} + \hat{\Gamma}(k)'e^{-i\omega k}\right)\right] \quad \text{(multivariate)}$$

where we choose $\kappa(k,T)$ to be the Bartlett lag window for which

$$\kappa(k,T) = \begin{cases} 1 - \frac{|k|}{M(T)+1} & \text{if } |k| \leq M(T) \\ 0 & \text{if } |k| > M(T) \end{cases} \qquad \text{with } M(T) < T$$

$\hat{\gamma}(k)$ and $\hat{\Gamma}(k)$ are the sample autocovariance function and the sample covariance matrix at lag k, respectively:

$$\widehat{\gamma}(k) = \frac{1}{T}\sum_{t=k+1}^{T} \Delta y_t \Delta y_{t-k} \qquad k = 0, 1, \ldots, M(T)$$

$$\widehat{\Gamma}(k) = \frac{1}{T}\sum_{t=k+1}^{T} (\Delta y_t, z_{1t}, \ldots, z_{nt})'(\Delta y_{t-k}, z_{1t-k}, \ldots, z_{nt-k}) \qquad k = 0, 1, \ldots, M(T)$$

assuming that the series are demeaned and $M(T)$ is a truncation point that determines the

[13] We use Δy_t instead of y_t to make clear that we define the spectrum of a covariance stationary process.
[14] A brief overview on spectrum estimators and related topics is provided in Appendix A.

point at which the sample covariance function is ignored. This function of the sample size ensures the consistency of the estimator if $M(T)$ tends to infinity slower than T does, so that $\frac{M(T)}{T} \to 0$ as $T \to \infty$. Details on the interpretation and how to choose this value are provided in Appendix A. To obtain the optimal weights we just have to replace in the corresponding linear system (univariate or multivariate) the parameter M by $M(T)$ and the population moments by the weighted sample covariance estimates.

The necessary moments can also be obtained parametrically. To this effect we follow Priestley (1981) and, in particular, the work of Den Haan and Levin (1996, 2000), but with a different objective. We use their method to obtain a parametric estimate of the true covariance function at various lags while they aim to obtain a heterocedastic autocorrelation consistent estimate of the variance-covariance matrix. In the case of the univariate filter we start by fitting an autoregressive (AR) model to Δy_t and in the case of the multivariate filter by fitting a seemingly unrelated regression-vector autoregressive (SUR-VAR) model to $(\Delta y_t, z_{1t}, \ldots, z_{nt})'$. Furthermore, we allow each equation of the SUR-VAR model to have a different lag structure across groups of variables, in detail, we set that the dependent variable of an equation will have h_{s1} lags while the remaining group of n variables will have h_{s2} lags with $s = 1, 2, \ldots, n+1$. The procedure to choose h_{s1} and h_{s2} is as follows: for every combination of lag orders (h_{s1}, h_{s2}), with $h_{s1}, h_{s2} = 1, 2, \ldots, H$ and H the maximum lag length, we estimate each equation by ordinary least squares (OLS) using only the observations from $H+1$ to T of the sample to ensure the comparability of the results; next for each pair of lags we compute the values of the AIC and BIC criterion and select the optimal combination of lag orders (h^*_{s1}, h^*_{s2}) for each equation minimizing the information criteria. Given the optimal lag structures, we then use a SUR method to gain efficiency in parameter estimates, as suggested by Den Haan and Levin (2000). With the lag polynomial of the AR model (a(L)) or the lag polynomial of the SUR-VAR model (A(L)) estimated we compute the residuals ($\hat{\epsilon}$) and estimate their sample covariance function as follows:

$$\widehat{\gamma}_{\hat{\epsilon}}(k) = \frac{1}{T} \sum_{t=k+1}^{T} \hat{\epsilon}_t \hat{\epsilon}_{t-k} \qquad \text{for } k = 0, \pm 1, \ldots, \pm T-1$$

$$\widehat{\Gamma}_{\hat{\boldsymbol{\epsilon}}}(k) = \begin{cases} \frac{1}{T} \sum_{t=k+1}^{T} \hat{\boldsymbol{\epsilon}}'_t \hat{\boldsymbol{\epsilon}}_{t-k} & \text{for } k = 0, 1, \ldots, T-1 \\ \widehat{\Gamma}_{\hat{\boldsymbol{\epsilon}}}(-k)' & \text{for } k = -1, -2, \ldots, -(T-1) \end{cases}$$

Thereafter, we use this result to derive an estimate of the covariance function of the original process by means of a finite approximation to the theoretical autocovariance generating function expressed as

$$G(z) = a(z)^{-1} G_{\epsilon}(z) a(z^{-1})^{-1}$$
$$G(z) = A(z)^{-1} G_{\boldsymbol{\epsilon}}(z) A(z^{-1})^{-1\prime}$$

with z taken to be a complex scalar and $G_{\epsilon}(z) = \sum_{j=-\infty}^{\infty} \gamma_{\epsilon}(j) z^j$ or $G_{\epsilon}(z) = \sum_{j=-\infty}^{\infty} \mathbf{\Gamma}_{\boldsymbol{\epsilon}}(j) z^j$ the autocovariance generating function of the process $\epsilon_{\mathbf{t}}$. The consistency of the estimator is guaranted if the sample covariance function implied by the chosen parametric model resembles the true covariance function. As before, to solve the linear system and obtain the optimal weights we replace the theoretical moments by the parametric estimates.

This kind of framework to obtain spectrum estimates is also regarded as a prewhitening technique in the sense that the fitted parametric model acts as a filter whose output is expected to be nearly white noise. It is widely documented that spectra with very sharp peaks, like those from highly persistent processes, are not easy to estimate while flat spectra, as the ones from the white noise processes, are easier to estimate. So, instead of directly estimate the spectrum, Press and Tukey (1956) suggest to adjust an AR or VAR model to the data, with the objective of flattening the spectrum, and use the estimated spectrum of the residuals model, which by assumption are nearly white, to obtain an estimate for the spectrum of the original process. This is exactly our approach which means we are prewhitening the process Δy_t in the univariate case and the vector process $(\Delta y_t, z_{1t}, \ldots, z_{nt})'$ in the multivariate case.

Each estimation method has its advantages but it is relevant to notice that if we consider a large number of covariates in the multivariate approximation it will be complicated to estimate a VAR model due to the huge number of unknown parameters to be estimated with a finite sample. Thus, the consideration of the kernel based estimator for comparison.

3.2 Multivariate information

3.2.1 Factor model

Several recent studies give special attention to factor models and this sudden interest is connected with the need to exploit information concentrated in large panels of data. These models describe observable variables as a linear combination of unobservable variables, known as factors, plus an idiosyncratic error. Their ability to combine the variables variation into a few number of factors reduces the dimensionality problem. Exploiting this advantage we will use factor estimates as covariates in the multivariate approximation.

Assume that $\mathbf{X} = [x_{it}]_{i=1,...,N;\ t=1,...,T}$ denotes a matrix of T observations of N different demeaned covariance stationary time series. In the classic or strict model each element of $\mathbf{X}$ can be modeled as

$$\begin{aligned} x_{it} &= \lambda_{i1}F_{1t} + \lambda_{i2}F_{2t} + ... + \lambda_{r1}F_{rt} + \eta_{it} \qquad (17) \\ &= \boldsymbol{\lambda}_i\mathbf{F}_t + \eta_{it} = \chi_{it} + \eta_{it} \end{aligned}$$

or in vector notation as

$$\mathbf{x}_t = \boldsymbol{\Lambda}\mathbf{F}_t + \boldsymbol{\eta}_t \qquad t = 1, ..., T$$

where $\boldsymbol{\lambda}_i = \begin{bmatrix}\lambda_{i1} & \lambda_{i2} & \cdots & \lambda_{ir}\end{bmatrix}$ is a row vector of factor loadings with λ_{ij} the loading of the jth factor for the ith variable, $\boldsymbol{\Lambda}$ is an $(N \times r)$ matrix of loadings with rows equal to $\boldsymbol{\lambda}_i$, $\mathbf{F}_t = \begin{bmatrix}F_{1t} & F_{2t} & \cdots & F_{rt}\end{bmatrix}'$ is a column vector of r serially uncorrelated factors and $\boldsymbol{\eta}_t = \begin{bmatrix}\eta_{1t} & \eta_{2t} & \cdots & \eta_{Nt}\end{bmatrix}'$ a column vector of N unobservable variable-specific errors assumed to be serially and cross-sectionally uncorrelated. Moreover, $E\left[\mathbf{F}_t\boldsymbol{\eta}_t'\right] = 0$ which means that factors and idiosyncratic errors are mutually orthogonal.

The model in equation (17) decomposes the variables variation into two components. The first term is referred to as common component (χ_{it}) and is driven by a small number, r, of factors while the second term is referred to as idiosyncratic component (η_{it}). Furthermore, the classic factor model is a static model because the association between factors and variables is only contemporaneous.

In an economic context some of the assumptions made in the classic model are unrealistic, namely those over the idiosyncratic errors. Hence, Chamberlain (1983) and Chamberlain and

Rothschild (1983) introduced a static factor model which allows for some degree of serial correlation of the idiosyncratic errors. Several years later, Stock and Watson (2002a, 2002b) specified a factor model with serial and cross correlated idiosyncratic errors. Both models are known as approximate factor models due to the assumptions over the idiosyncratic disturbances. Precisely,

$$\mathbf{x}_t = \mathbf{\Lambda F}_t + \mathbf{e}_t \quad t = 1, ..., T$$

with $E\left[\mathbf{F}_t \mathbf{F}_t'\right] = \mathbf{\Sigma}_F$ a diagonal matrix, $E\left[\mathbf{e}_t' \mathbf{e}_{t-k}\right] = \gamma_e(k)$, $E\left[e_{it} e_{jt}\right] = \tau_{ij,t}$ and $E\left[\mathbf{F}_t \mathbf{e}_t'\right] = \mathbf{0}$.

Sargent and Sims (1977) and Geweke (1977) reinvented the classic factor model in another perspective (exact factor model). In their version of the model the factors are loaded using a lag structure as follows

$$\begin{aligned} x_{it} &= b_{i1}(L) f_{1t} + b_{i2}(L) f_{2t} + ... + b_{iq}(L) f_{qt} + \upsilon_{it} \\ &= \mathbf{b}_i(L) \mathbf{f}_t + \upsilon_{it} \end{aligned}$$

or in vector notation

$$\mathbf{x}_t = \mathbf{B}(L)\mathbf{f}_t + \boldsymbol{\upsilon}_t \quad t = 1, ..., T$$

where $\mathbf{b}_i(L) = \begin{bmatrix} b_{i1}(L) & b_{i2}(L) & \cdots & b_{iq}(L) \end{bmatrix}$ is a row vector of q lag polynomials of the form $b_{ij}(L) = \sum_{k=0}^{\infty} b_{ij,k} L^k$ for $j = 1, ..., q$, $\mathbf{f}_t = \begin{bmatrix} f_{1t} & f_{2t} & \cdots & f_{qt} \end{bmatrix}'$ is a column vector of q common shocks and $\boldsymbol{\upsilon}_t = \begin{bmatrix} \upsilon_{1t} & \upsilon_{2t} & \cdots & \upsilon_{Nt} \end{bmatrix}'$ is a vector of N idiosyncratic terms cross and serial uncorrelated. Moreover, $\mathbf{B}(L) = \left[b_{ij}(L)\right]_{i=1,...,N; j=1...,q}$ is a $(N \times q)$ matrix of factor loadings and as before the common component is assumed to be orthogonal to the idiosyncratic component.

Finally, Forni and Lippi (2001) and Forni et al. (2000, 2004) suggested a model (approximate dynamic factor model) that conveys all model versions discussed so far. Suggestively, they call it generalized dynamic factor model which in vector notation is given by

$$\mathbf{x}_t = \mathbf{B}(L)\mathbf{f}_t + \boldsymbol{\xi}_t \quad t = 1, ..., T$$

where $\mathbf{x}_t = \begin{bmatrix} x_{1t} & x_{2t} & \cdots & x_{Nt} \end{bmatrix}'$ is a column of N variables, $\mathbf{B}(L) = \left[b_{ij}(L)\right]_{i=1,...,N; j=1...,q}$ is a $(N \times q)$ matrix of one sided filters as $b_{ij}(L) = \sum_{k=0}^{\infty} b_{ij,k} L^k$ for $j = 1, ..., q$ with square summable coefficients, $\mathbf{f}_t = \begin{bmatrix} f_{1t} & f_{2t} & \cdots & f_{qt} \end{bmatrix}'$ is a white noise vector of q common orthogonal shocks with $E[\mathbf{f}_t] = 0$ and $E[\mathbf{f}_t'\mathbf{f}_t] = 1$ and $\boldsymbol{\xi}_t = \begin{bmatrix} \xi_{1t} & \xi_{2t} & \cdots & \xi_{Nt} \end{bmatrix}'$ is a column vector of nonorthogonal idiosyncratic terms. Further, $\xi_{it} \perp f_{jt-u}$ for any $j = 1, 2, ..., q$ and $u \in \mathbb{Z}$ meaning that the two components are mutually orthogonal at all leads and lags.

Following Stock and Watson (2002a, 2002b) and Forni et al. (2005) (restricted factor model), if the lag polynomials $b_{ij}(L)$ for $i = 1, ..., N$ and $j = 1, ..., q$ have a finite s order we have what is known in the literature as dynamic factor model. This model has the advantage of a static representation. Define

$$\mathbf{F}_t = \begin{bmatrix} F_{1t} \\ F_{2t} \\ \vdots \\ F_{rt} \end{bmatrix} = \begin{bmatrix} \mathbf{f}_t \\ \mathbf{f}_{t-1} \\ \vdots \\ \mathbf{f}_{t-s} \end{bmatrix}$$

with $\mathbf{f}_t = \begin{bmatrix} f_{1t} & f_{2t} & \cdots & f_{qt} \end{bmatrix}'$ and $r = q(s+1)$. Then,

$$\mathbf{x}_t = \mathbf{B}^*(L)\mathbf{f}_t + \boldsymbol{\theta}_t = \mathbf{C}\mathbf{F}_t + \boldsymbol{\theta}_t \qquad t = 1, ..., T$$

where $\mathbf{B}^*(L) = \left[b_{ij}^*(L)\right]_{i=1,...,N; j=1...,q}$ is a $(N \times q)$ matrix of factor loadings with $b_{ij}^*(L) = \sum_{k=0}^{s} b_{ij,k}^* L^k$, $C = [c_{ij}]_{i=1,2,...,N;\ j=1,2,...,r}$ is a $(N \times r)$ matrix of loadings and $\mathbf{F}_t$ and $\mathbf{f}_t$ are column vectors of factors. Thus, we obtain a static representation by appending the lagged factors as additional static factors. To distinguish the factors of the static representation from the factors of the dynamic representation we call the factors F_{jt} static factors and the factors f_{lt} dynamic factors.

3.2.2 Estimation of the factors

The static factors reflect unobservable variables and therefore have to be estimated from the panel data set. In the factor model literature several estimation procedures have been proposed, however we will only focus on two recent non-parametric approaches.

Stock and Watson (2002b) discuss factor estimation in an approximate static factor model with serial and cross correlated idiosyncratic errors and prove that the estimation of the static factors by principal components (PC) is a consistent estimation procedure if N and T both tend to infinity. Moreover, Bai (2003) establishes the asymptotic normality of the Stock and Watson estimator adding some other technical assumptions to the static approximate model. A principal component is a linear combination of the variables in the panel with coefficients given by the eigenvector of the sample variance-covariance matrix. Denote $\hat{\mathbf{F}}_t^{SW}$ as the Stock and Watson estimator of the static factors $\mathbf{F}_t$, then

$$\hat{\mathbf{F}}_t^{SW} = \begin{bmatrix} \hat{F}_{1t} & \hat{F}_{2t} & \cdots & \hat{F}_{rt} \end{bmatrix}' = \hat{\mathbf{S}}'\mathbf{x}_t = \begin{bmatrix} \hat{S}_1\mathbf{x}_t & \hat{S}_2\mathbf{x}_t & \cdots & \hat{S}_r\mathbf{x}_t \end{bmatrix}'$$

or in matrix notation

$$\hat{\mathbf{F}}^{SW} = \begin{bmatrix} \hat{\mathbf{F}}_1^{SW\prime} & \hat{\mathbf{F}}_2^{SW\prime} & \cdots & \hat{\mathbf{F}}_T^{SW\prime} \end{bmatrix}' = \frac{\mathbf{X}\hat{\mathbf{S}}}{N}$$

where $\hat{\mathbf{S}}$ is a $(N \times r)$ matrix, whose columns are the eigenvectors of the r largest eigenvalues of the sample variance-covariance matrix, $\hat{\mathbf{\Gamma}}_{\mathbf{x}}(0) = \sum_{t=1}^{T} \mathbf{x}_t\mathbf{x}_t'$. The matrix $\hat{\mathbf{S}}$ is referred to as the PC estimator of the factor loadings and $\hat{\mathbf{F}}^{SW}$ as the PC estimator of the static common factors. Notice that PC are organized by descending order of the eigenvalues meaning that the first estimated factor corresponds to the eigenvector with the maximal eigenvalue.

In more detail, each principal component is the optimal solution of the following broad problem:

$$\begin{aligned} \max_{\mathbf{a}_j \in \mathbb{R}^N} \quad & \mathbf{a}_j \hat{\mathbf{\Gamma}}_x(0)\mathbf{a}_j' \\ s.t. \quad & \mathbf{a}_j\mathbf{a}_j' = 1 \\ & \mathbf{a}_j \hat{\mathbf{\Gamma}}_x(0)\mathbf{a}_k' = 0 \qquad j < k; \quad j = 1, 2, \ldots, r \end{aligned}$$

where $\hat{\mathbf{\Gamma}}_x(0)$ is the sample variance-covariance matrix and $\mathbf{a_j}$ is an arbitrary vector of coefficients of a linear combination. Such linear combination will have maximum sample variance only if each $\mathbf{a_j}$ is set equal to an eigenvector of matrix $\hat{\mathbf{\Gamma}}_x(0)$. From the optimization problem we see that PC are uncorrelated linear combinations with the largest sample variances.

Alternatively, Forni et al. (2005) discuss factor estimation in a approximate dynamic factor model. Their approximation to the static factors, denoted by $\hat{\mathbf{F}}_t^{FHLR}$, follows from a frequency domain representation of the factor model and consists of the first r generalized principal components (GPC) of matrix $\mathbf{X}$. These GPC consist in linear combinations of variables, as the standard PC, but the coefficients equal the generalized eigenvectors of the pair of matrices $\left(\hat{\mathbf{\Gamma}}_\chi(0), \hat{\mathbf{\Gamma}}_\xi(0)\right)$. These matrices represent the estimated contemporaneous covariance matrix of the common and idiosyncratic component of the factor model, respectively. The Forni et al. estimator of the factors is given by

$$\hat{\mathbf{F}}_t^{FHLR} = \begin{bmatrix} \hat{F}_{1t} & \hat{F}_{2t} & \cdots & \hat{F}_{rt} \end{bmatrix}' = \hat{\mathbf{Z}}'\mathbf{x}_t = \begin{bmatrix} \hat{\mathbf{Z}}_1\mathbf{x}_t & \hat{\mathbf{Z}}_2\mathbf{x}_t & \cdots & \hat{\mathbf{Z}}_r\mathbf{x}_t \end{bmatrix}'$$

where $\hat{\mathbf{Z}}$ is a $(N \times r)$ matrix, whose columns are the generalized eigenvectors associated with the r largest generalized eigenvalues of the pair $\left(\hat{\mathbf{\Gamma}}_\chi(0), \hat{\mathbf{\Gamma}}_\xi(0)\right)$.

In this case each vector of coefficients is defined recursively by solving a slightly different problem

$$\begin{aligned} \max_{\mathbf{b}_j \in \mathbb{R}^N} \quad & \mathbf{b}_j \hat{\mathbf{\Gamma}}_\chi(0) \mathbf{b}_j' \\ s.t. \quad & \mathbf{b}_j \hat{\mathbf{\Gamma}}_\xi(0) \mathbf{b}_j' = 1 \\ & \mathbf{b}_j \hat{\mathbf{\Gamma}}_\xi(0) \mathbf{b}_k' = 0 \qquad j < k; \quad j = 1, 2, \ldots, r \end{aligned}$$

where $\mathbf{b}_j$ is an arbitrary vector of coefficients of a possible linear combination of the variables and $\hat{\mathbf{\Gamma}}_\chi(0)$ and $\hat{\mathbf{\Gamma}}_\xi(0)$ have the same definitions as above. Forni et al. (2005) show that the optimal solution to this constrained problem is to choose $\mathbf{b}_j$ equal to the generalized eigenvectors of the pair of matrices $\left(\hat{\mathbf{\Gamma}}_\chi(0), \hat{\mathbf{\Gamma}}_\xi(0)\right)$. This implies that the vector of coefficients is obtained from

$$\mathbf{b}_j \hat{\mathbf{\Gamma}}_\chi(0) = \hat{\nu}_j \mathbf{b}_j \hat{\mathbf{\Gamma}}_\xi(0) \qquad j = 1, 2, \ldots, r$$

where $\hat{\nu}_j$ denotes a generalized eigenvalue. Similarly to PC, the GPC are uncorrelated linear combinations which maximize the variance of the common component of the factor model.

A related estimation method is suggested in Altissimo et al. (2008). They argue that smoother factors can be constructed choosing as weights, in the r independent linear combinations, the generalized eigenvectors of a different pair of matrices. In this case they express the factor model as the sum of two components but assume that the common component can be further divided into a medium to long-run component and a short-run component. Since the goal is to obtain factors cleaned of short run oscillations and idiosyncratic errors, the coefficients are obtained by solving

$$\hat{\mathbf{Z}}_j \hat{\mathbf{\Gamma}}_\phi(0) = \hat{\alpha}_j \hat{\mathbf{Z}}_j (\hat{\mathbf{\Gamma}}_\chi(0) + \hat{\mathbf{\Gamma}}_\xi(0)) \qquad j = 1, 2, \ldots, r$$

where $\hat{\Gamma}_{\phi}(0)$ denotes the estimated variance-covariance matrix of the medium to long-run component of the common component and $\hat{\alpha}_j$ a generalized eigenvalue. The remaining objects maintain the definitions already given. The authors use these smooth factors to compute the new Euro-Coin indicator, which is an indicator of the medium to long run component of GDP growth rate for Euro Area aggregates.

Both estimators aim to approximate the same factor space but while the PC estimator only exploits the contemporaneous relations among variables the GPC estimator exploits entirely the dynamic covariance structure of the data when estimating $\hat{\mathbf{\Gamma}}_{\chi}(0)$ and $\hat{\mathbf{\Gamma}}_{\xi}(0)$[15]. Consequently, from the GPC estimator results linear combinations that are more efficient than those resulting from the PC estimator.

In the empirical application in section 4 we will compare both methods to approximate the factor space, i.e., we will estimate the common factors using either PC as in Stock and Watson (2002a, 2002b) or GPC as in Forni et al. (2005). However, we will not consider smooth factors because its irrelevant for filtering purposes.

3.2.3 Estimation of the number of factors

The number of factors necessary to account for most of the variables variation is in practice unknown and must be identified empirically. In this section we briefly give an overview of the existing methods to estimate the number of factors in a static or dynamic framework.

The recurrent use of factor models in economic applications is undoubtedly recent but there is already a large number of available criteria in the literature to deal with the estimation of these parameters. Firstly, it is noteworthy that in the factor model followed by Stock and Watson (2002a, 2002b) we must only identify the number r of static factors while in the factor model introduced by Forni and Lippi (2001) we must determine the number of static and dynamic factors, that is, r and q, respectively.

Regarding, the static factor model framework the first formal criterion for selecting the number of common factors was proposed by Bai and Ng (2002). They modified the classical information criteria (IC) to fit the case when both time and cross section variation is present in the data. As usual, the estimated number of factors, $\hat{r}$, is obtained from minimizing the IC in the

[15]This pair of matrices is obtained through an inverse Fourier transform of the estimated spectral density functions of the common and idiosyncratic components that use information contained in the whole covariance sequence $\hat{\mathbf{\Gamma}}_{\mathbf{x}}(k)$ for $k = 0, \pm 1, \pm 2, ...$For more details on the estimation see Forni et al (2005).

range $[1, r_{\max}]$, with $r_{\max}$ denoting a pre-specified upper bound for the true number of factors. The three suggested IC differ only in the penalty functions that in contrast with standard IC depend on both T and N. An alternative way suggested by Onatski (2005) for estimating r is based on the empirical distribution of the eigenvalues. The method involves counting the number of eigenvalues of the sample covariance matrix that are above a threshold. Moreover, such threshold is set as a slowly increasing function of N and T. Onatski (2005) proves the consistency of his method in the presence of either cross-sectional or serial correlation but never both and also provides a comparison of his method with the modified IC's of Bai and Ng (2002). Finally, Kapetanios (2004) proposed a sequential algorithm, named *Maximum Eigenvalue*, based on the widely accepted fact that when no factor structure exists the eigenvalues of the covariance matrix tend asymptotically to a constant whereas otherwise they tend to infinity.

Turning to the same question in the context of dynamic factors, Forni et al. (2000) proposed an informal method based on the relative size of the eigenvalues from the estimated spectral density matrix of the data. Other contributions to estimate q in a dynamic factor model with s finite include a modified version of Bai and Ng's IC by Amengual and Watson (2007) and a method that exploits the rank of the residuals of a VAR(p) in the static factors by Bai and Ng (2007). Recently, Hallin and Liška (2007) suggest a method for identifying the number of common factors in a generalized dynamic factor model. The criterion settles on the idea that the number of diverging eigenvalues of the spectral density matrix of the observations as N tends to infinity is equal to the number of primitive factors.

In our empirical application we will use one of these methods to establish an upper bound of the number of factors to be included in the multivariate approximations. Our goal is not to estimate a common component but to incorporate a reasonable number of factors in the multivariate filter that might produce an acceptable fit to the desired signal. Moreover, we justify this approach with the idea that only a few shocks drive economic fluctuations, consequently a large part of the data variation is explained by the first few factors (see Forni et al., 2005 or Gionnone, Reichlin and Sala, 2005).

4 Real Data Application

In this section we discuss the approximations of Christiano and Fitzgerald (2003) (henceforth also referred to as univariate approximation) and Valle e Azevedo (2007) (henceforth also referred to as multivariate approximation) to business cycle fluctuations and to smooth output growth using real data. Our targets will be precisely $BC(L)y_t$ and $SG(L)\Delta y_t$ as defined in section 2. We analyze two bands of periodicities: the $[6, 32]$ period band connected to business cycle fluctuations and the $[4, +\infty[$ period band connected to smooth growth.

4.1 Dataset

Our variable of interest will be the U.S. real GDP, either in logarithmic form or in growth rate. For such a time series we have quarterly observations from the first quarter of 1959 to the third quarter of 2003 ($T = 175$). Figure 7 displays the path followed by logarithm of real GDP and its growth rate (since the level is on logarithmic form) in the sample period. As expected, real quarterly GDP exhibits a clear upward trend in the logarithm, suggesting that it may have either stochastic or deterministic trends. The remaining fluctuations (cyclical and erratic) although present are not directly detectable. In contrast, the growth rate series has a stationary behavior that is mainly made of medium and short term movements. Moreover,

Figure 7: Logarithm of real GDP from 1959(q4)-2003(q3) and of the first difference of the logarithm of real GDP from 1960(q1)-2003(q4).

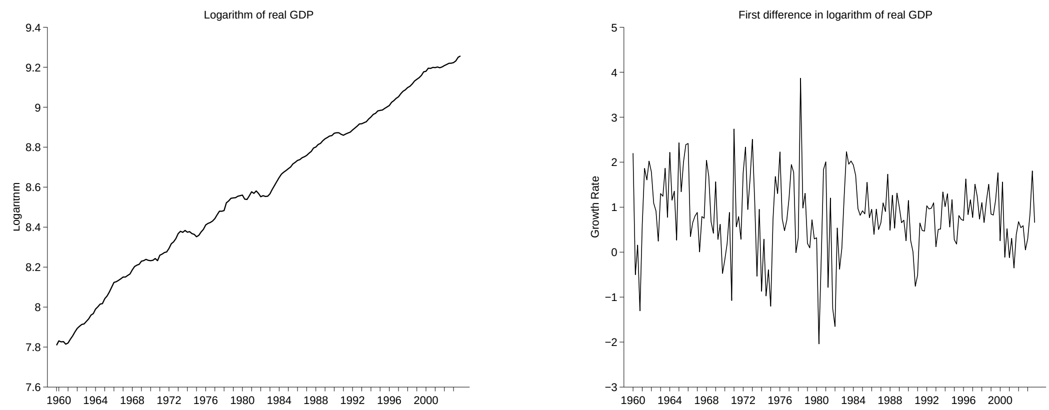
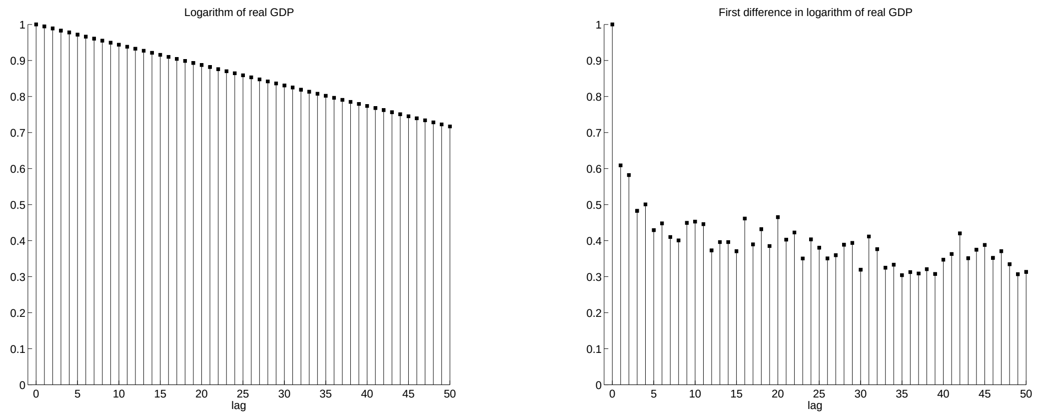


figure 8 shows the sample autocorrelation functions for the logarithm of real GDP and its growth rate, respectively. The autocorrelation in real GDP has a high level even for long lags, which reveals the highly persistent nature of the process connected with non-stationarity. On the other hand, the autocorrelation in the growth rate exhibits lower levels and a downward pattern as the lags increase. To decide on the type of non-stationarity present in real GDP we ran augmented

Figure 8: Sample autocorrelation function of the logarithm of real GDP and of the first difference of the logarithm of real GDP.



Dickey-Fuller (ADF) unit-root tests. For this purpose we estimated the following equation:

$$\Delta y_t = \alpha_0 + \alpha_1 t + \gamma y_{t-1} + \sum_{i=1}^{d} \beta_i \Delta y_{t-i} + \epsilon_t$$

where y_t is the logarithm of real GDP, Δy_t is the first difference in logarithm of real GDP, d is the maximum lag length and ϵ_t a error term. In the ADF tests the null hypothesis is that the time series contains an unit root and the alternative is that the time series was generated by a stationary process. If γ is not statistically different from zero this means that y_t may contain a unit root and if γ is statistically different from zero ($|\gamma| < 0$) then y_t is stationary.

Table 2 presents the results of the ADF test on real GDP with $d = 1$ (determined by BIC criterion). The null hypothesis is not rejected at 1, 5 and 10 percent level, indicating that real GDP has a unit root. This also implies that the first difference of real GDP is a stationary series.

Table 2: Results of ADF test on logarithm of real GDP.

ADF test	Test statistic	1% level	5% level	10% level
with trend	-2.479	$\circ^a$	$\circ$	$\circ$

[a] $\circ$: when the null hypothesis is not rejected and $\times$: when the null hypothesis is rejected.

Then the business cycle fluctuations are defined on a non-stationary series (log of real GDP) while the smooth growth is defined on a stationary variable (growth rate of real GDP). As a result, in the first case we must solve a constrained optimization problem to ensure the stationarity of the filtered series while in the second case we can ignore the restriction imposed on the frequency response function of the approximate filter.

For the multivariate indicator we need to have a group of covariates to estimate the factors, as discussed in section 3. For that purpose, we consider the same vintage panel of $N = 132$ U.S. monthly time series studied in Stock and Watson (2005). This large data set was downloaded directly at *http://www.princeton.edu/˜ mwatson* and includes several types of macroeconomic variables. Following Stock and Watson (2005) there are *14 categories: real output and income; employment and hours; real retail, manufacturing and trade sales; consumption; housing starts and sales; real inventories; orders; stock prices; exchange rates; interest rates and spreads; money and credit quantity aggregates; price indexes; average hourly earnings; and miscellaneous*. The complete data set covers the period from January of 1959 to December of 2003. Due to missing observations for some time series at the beginning of the sample and to the realignment of the variables to account for release delays we only considered a time period from February of 1960 to December of 2003 ($T^* = 527$). We use T^* instead of T to make clear that we have a monthly panel of time series. In order to apply a factor model framework all the time series in the monthly panel must be covariance stationary. So, all variables were transformed as suggested in Stock and Watson (2005), appendix A.

4.2 Covariates used in the multivariate approximation

The multivariate filter does not accommodate data with different frequencies[16], meaning that we cannot incorporate the estimated factors without any transformation. To deal with this problem we split each estimated monthly factor into three quarterly series as follows:

$$\begin{aligned}
&\widehat{F}^1_{l,t} = \widehat{F}_{l,t^*},\ \widehat{F}^1_{l,t-1} = \widehat{F}_{l,t^*-3},\ \widehat{F}^1_{l,t-2} = \widehat{F}_{l,t^*-6}, ... \\
&\widehat{F}^2_{l,t} = \widehat{F}_{l,t^*-1}, \widehat{F}^2_{l,t-1} = \widehat{F}_{l,t^*-4},\ \widehat{F}^2_{l,t-2} = \widehat{F}_{l,t^*-7}, ... \quad l = 1, ..., k \\
&\widehat{F}^3_{l,t} = \widehat{F}_{l,t^*-2},\ \widehat{F}^3_{l,t} = \widehat{F}_{l,t^*-5},\ \widehat{F}^3_{l,t} = \widehat{F}_{l,t^*-8}, ...
\end{aligned} \tag{18}$$

Accordingly, the covariate set in month t^* of quarter t considering k estimated monthly factors is

$$Z_{t^*}{=}(\widehat{F}^1_{1,t}, \widehat{F}^2_{1,t}, \widehat{F}^3_{1,t}..., \widehat{F}^1_{k,t}, \widehat{F}^2_{k,t}, \widehat{F}^3_{k,t})'.$$

The stationarity and the finite moving average representation of the extracted monthly factors, required to apply the multivariate filter, follows from the factor model assumptions. Furthermore, these properties are still valid regarding the vector of k split monthly factors. A discussion on the consistency of the extracted signal using the estimated factors can be found in Azevedo and Pereira (2008).

Both signals are define on a quarterly time series but in the case of the multivariate approximation we are able to compute in each month an estimate of those signals by updating the extracted factors with the new monthly information. Hopefully, this will improve considerable the approximation from one month to the other and from the univariate framework as well.

4.3 Number and estimation of the factors

As discussed in section 3, the monthly static factors will be estimated either by standard PC or by GPC. Both estimators set linear combinations of the variables in the panel but differ in the matrices exploited to obtain the coefficients.

An estimate of the true number of factors embedded in this specific panel is far from being consensual in the literature (see Stock and Watson, 2005, Bai and Ng, 2007 and Hallin and Liška,

[16] Real GDP or GDP growth rate both have quarterly frequency while the variables in the panel all have monthly frequency and therefore also the estimated factors.

2007). In particular, the number r of static factors. This value is generally estimated according to one of the methods discussed in section 3. As we argued, this is the appropriate path to follow if our purpose is to estimate the factor space and thereafter the common component. But our objective is the approximation of signals and in such a case it is wiser to check whether additional factors are helpful to improve the approximations. Otherwise we might obtain misleading conclusions. This approach is followed by Stock and Watson (2002a, 2002b) to obtain $\hat{k}$ in their factor augmented regression and by Altissimo et al. (2008) to determine the number of factors to include in the new EuroCoin indicator. So, we estimate k as follows: first we established as upper bound for the number of factors to be included $\hat{k}_{\max} = 6$ obtained by applying the CP1 criterion in Bai and Ng (2002) and then added split estimated factors up to the point where the increase in the squared empirical correlation between final estimates and real time estimates[17] in the third month of each quarter became negligible. In the case of business cycle fluctuations we found that the improvement in the fit became negligible with 5 static factors whereas in the case of the smooth growth 4 static factors seemed sufficient to stabilize the fit. We note that the values of $\hat{k}$ never reached the upper bond.

In another perspective several empirical applications, using U.S. data, point out that a small number of factors seems enough to explain most of the total variation in those series[18]. Therefore, as an alternative to the previous choice of $\hat{k}$ we also report performance results when using exactly the first 2 monthly static factors. It is only in this setting that we can follow a SUR-VAR approach to obtain second order moment estimates due to the "small" dimension of the covariate set.

The use of GPC estimator to extract the factors implies setting an estimate for the number of dynamic factors, $\hat{q}$. In line with the findings of Hallin and Liška (2007) we assumed $\hat{q} = 4$.

Below, we report the performance results for both choices of $\hat{k}$ and factor estimators.

4.4 Release delays

Since our main objective is to develop a real time indicator our targets must focus on GDP (or on GDP growth rate) of the current quarter and we have also to take into account the release delays of all the series involved in the filtering procedure.

The Bureau of Economic Analysis releases a final version of real GDP 6 months after the

[17]A clear definition of these objects is given below. For now it is only important to stress that both kinds of estimates considered estimated factors and second order moments obtained from the whole sample.

[18]See Stock and Watson (1999, 2002a, 2005); Giannone, Reichlin and Sala (2005) and Hallin or Liška (2007).

beginning of the quarter to which it refers. Meanwhile the same institute releases two earlier estimates that had potential to be used as covariates in the approximations. However, their inclusion in the covariate set in specific months did not produce substantial improvements and therefore we chosen not to consider such estimates. Nevertheless, table 3 reports the release delays of these three estimates.

Table 3: Release delays of U.S. GDP estimates by month of the quarter.

GDP estimate	1^{st} month	2^{nd} month	3^{rd} month
Final	2 quarters	2 quarters	1 quarter
Preliminary	2 quarters	1 quarters	1 quarter
Advanced	1 quarters	1 quarters	1 quarter

It is clear that if time t denotes the current quarter then the series of interest is not available. This fact implies to set f (number of future observations) negative in the solution so that only the information in fact available on GDP is taken into account. To be specific, in the first and second month of a given quarter we set $f = -2$ and in the third month of a given quarter we set $f = -1$. For the univariate filter, the estimates are not computed monthly but only in the third month of a quarter and thus $f = -1$. Of course we can also set $f = -2$ to have a univariate estimate in the first and second month of a quarter but no update occurred in quarterly information since the third month of the prior quarter that could enhance the approximation like in the multivariate case. In sum, if time t is the current quarter and $f = -1$ this means that only information on GDP up to period $t - 1$ is taken into account.

The variables included in the panel of predictors where also realigned to mimic a real-time data set. This task implied screening the release dates of the different variables and shift the time series when necessary[19]. At the end, the monthly panel is organized in such a way that at each filtering moment we use the data that would be in fact available at that month. For instance if t^* denotes the third month of quarter t then we assume $f = -1$ and the set of variables includes GDP series until period $t - 1$ and monthly predictors until period t^*.

[19]For instance, financial variables in general do not have release delays so no realignment was perform to these series.

4.5 Different types of estimates

Three different types of estimates were computed to evaluate the performance of the indicators: *final* estimates, *real-time* estimates and *today onwards* estimates. The *final* estimates are obtained by approximating the signals using the full sample at each filtering time; accordingly we set $p = t-1$ and $f = T-t$ in the solution. These estimates will be highly accurate, in particular in the middle of the sample given the use of future data points. The *real-time* approximations are computed using only the information available at each point in time. Specifically, in the multivariate case we gather the information until period t^* of quarter t and standardize it. Then, we estimate the static factors which we split into quarterly series. After the estimation of the necessary second order moments we obtain the filter weights by solving the linear system with $f = -1$ (in the case of a third month) or $f = -2$ (in the case of a first or second month) and $p = \bar{p}$, where $\bar{p}$ denotes a fixed integer. Finally, we apply the filter to this subset of data and obtain the *real-time* approximation of the signal for quarter t. The *today onwards* approximations use the whole sample to estimate the factors (only in the multivariate case) and the second order moments. These last estimates are subsequently used as inputs to solve the linear system from which we get the optimal weights. Such weights are then applied to the *real-time* subset of data and we obtain the *today onwards* approximation of the signal for quarter t. Note that when filtering we use only the available information in quarter t. Therefore we still set $f = -1$ (in the case of a third month) or $f = -2$ (in the case of a first or second month) and $p = \bar{p}$. With this exercise we hope to understand the revisions stemming from second order moments and factor space uncertainty.

To assess the performance of the indicators the *real-time* and *today onwards* estimates are compared with the *final* estimates obtained from using the univariate filter with moments derived from an AR model. Alternative multivariate filters and/or different methods to estimate second order moments deliver *final* estimates that are indistinguishable from those used.

4.6 Performance of the indicators

One important topic in filtering literature regards the revisions to which the estimates are subject to. These revisions arise from corrections of the past values of the time series and from the release of new data which has implications mainly in the estimates near the end of the sample. The effects of such revisions can be quite substantial, as documented in Orphanides and Van Norden (2002). In this particular application the estimates are not affected by data revisions due to the use of vintage data; but if not such revisions could not be a major problem because there is the idea that data revisions are mainly connected with high frequencies and these are completely eliminated with our filtering procedures. Nevertheless, we will try to mitigate as much as possible the effects of revisions due to the release of new data by disregarding the last 12 observations in the case of business cycle fluctuations (see Valle e Azevedo, 2007) and the last 4 observations in the case of output smooth growth (see Altissimo et al., 2008) when comparing the estimates.

The truncation point is a function of the sample size that guarantees the consistency of the spectrum estimator, as discussed in section 3, and consequently of the estimated signal. One of the assumptions made by Christiano and Fiztgerald (2003) and Valle e Azevedo (2007) to obtain the optimal solution referred to the finiteness of the MA representation of Δy_t and $(\Delta y_t, z_{1t}, \ldots, z_{nt})'$, respectively. So, it seems adequate to set the truncation point higher than the order of the MA representation. The problem is that the true order is in practice unknown. Hence, instead we followed a *window closing*[20] approach to set the truncation point. As a result, the projection problem for the business cycle case was solved considering 30 leads and 30 lags of the estimated covariance function and for the smooth growth case it was solved considering 40 leads and 40 lags of the estimated covariance function. But we note that in the range $20 < M < 40$ the results are very similar for both cases. M should be carefully chosen because considering too few covariances can leave out important information and considering too many covariances can introduce a very erratic behavior to the approximation as a result of the poorly estimated high order covariances.

The evaluation of the real time performance of the indicators makes use of the following criteria:

a) $Corr_t[x_t, \widehat{x}_t]$, where $\widehat{x}_t$ is the optimal approximation to the signal x_t. It is easy to show that if $\widehat{x}_t$ minimizes $E[(x_t - \widehat{x}_t)^2]$, then $E[(x_t - \widehat{x}_t)^2] = (1 - Corr_t[x_t, \widehat{x}_t]^2)Var[x_t]$. The dependence on t is eliminated if we fix p and f in the real-time approximations, as we do. $Corr_t[x_t, \widehat{x}_t]$ is therefore a good measure of the variance of the approximation error. We compute the sample counterpart of this statistic, using the estimated signal (say $\widehat{x}_t^*$) as $\widehat{x}_t$ and approximating x_t by the *final* estimates, denoted by x_t^F.

[20] For details on *window closing* see Appendix A.

b) Noise to Signal Ratio, computed as $\sum_t (\widehat{x}_t^* - x_t^F)^2 / \sum_t (x_t^F - \overline{x}^F)^2$.

c) The percentage of times $\widehat{x}_t^* - \widehat{x}_{t-1}^{*\ t}$ (where $\widehat{x}_{t-1}^{*\ t}$ is the approximation to the signals at $t-1$ using information up to time t) correctly signs the change $x_t^F - x_{t-1}^F$ (see Pesaran and Timmerman, 1992).

d) For the approximation to business cycle fluctuations, the percentage of times $\widehat{x}_t^*$ and x_t^F share the same sign (which gives an indication of whether $\widehat{x}_t^*$ indicates correctly if GDP is below or above the long-term trend).

Finally and before turning to the discussion of the results we clarify all the variations considered in the filters' settings:

- estimation of second order moments done by means of a non-parametric method using the Bartlett lag window as weight sequence or by means of estimating a SUR-VAR model for pre-whitening. Each estimation method was discussed in section 3 and most importantly the parametric framework implicitly limits the size of the covariate set;

- the factor space in the multivariate case is estimated by PC or by GPC. Both estimation methods were discussed in section 3;

- in the multivariate case, we report the statistics for two sets of covariates: one that only includes two monthly factors divided into six quarterly series and a second that includes split factors up to the point where the increase in the R^2 of the approximation is negligible. This topic was discussed early in this section;

- estimation of second order moments and factors (only in the multivariate case) takes into account only information available at each point (*real-time*) or, alternatively, the full sample (*today onwards*). These concepts were also introduced early in this section.

4.6.1 Business cycle fluctuations

To begin with we recall that we use the multivariate filter to construct a business cycle indicator for the U.S. economy and apply it to the same large panel of macroeconomic variables studied in Stock and Watson (2005). As a benchmark we considered naturally the univariate filter of

Christiano and Fitzgerald (2003) using either estimated moments derived from an AR model or from a non-parametric estimator.

In table 4 we report the performance results based on approximations done in a third month of the quarter. Each line corresponds to a different indicator whose description is given in the legend of the table.

Table 4: Evaluation statistics for the approximations to business cycle fluctuations in the U.S in the third month of the quarter. Evaluation period: 1978(q3) - 2000(q4).

Performance with respect to business cycle fluctuations (3rd month of the quarter)								
	Correlation		**Noise do Signal**		**Sign Concord.**		**% Correct Δ Sign**	
	Real Time	Today Onwards	Real Time	Today Onwards	Real Time	Today Onwards	Real Time	Today Onwards
Benchmark Filters[a]								
BPF AR	0.74	0.75	0.61	0.60	0.71	0.73	0.70	0.70
BPF KERNEL	0.72	0.78	0.62	0.57	0.69	0.70	0.67	0.69
with factors, $k = 5 < \widehat{r}, \widehat{q} = 4$								
MBPF PC KERNEL	0.75	0.96	0.59	0.26	0.69	0.92	0.76	0.87
MBPF GPC KERNEL	0.77	0.95	0.57	0.29	0.68	0.89	0.73	0.88
with factors, $k = 2 < \widehat{r}, \widehat{q} = 4$								
MBPF PC KERNEL	0.78	0.91	0.57	0.39	0.72	0.77	0.70	0.82
MBPF GPC KERNEL	0.77	0.90	0.58	0.41	0.68	0.77	0.71	0.82
MBPF PC VAR	0.84	0.86	0.49	0.46	0.72	0.76	0.76	0.78
MBPF GPC VAR	0.83	0.87	0.50	0.46	0.77	0.71	0.77	0.79

[a] BPF - univariate band-pass filter; AR - estimation of second moments by an AR model (BIC criterion for lag length); KERNEL - non-parametric estimation of second moments; MBPF - multivariate band-pass filter; PC - factor space estimated by principal components; GPC - factor space estimated by generalized principal components; VAR - estimation of second moments through pre-whitening with a SUR-VAR (BIC criterion for lag length in the various equations).

The multivariate indicators compare favourably with both univariate indicators either in full *real time* or in *today onwards*. This indicates that the variation of the variables in the panel provides valuable information to obtain a more exact signal of business conditions. In *real time* the gains are only modest when five monthly factors estimated by PC are used in the approximations. This behavior may stem from the poor estimation of moments and factors in the beginning of the evaluation period. Moreover, *real-time* approximations respond noticeably to the type of method used to obtain moment estimates. In fact using VARs to estimate second order moments leads to superior *real time* performance. When the full sample is used to estimate moments and factors, the multivariate approximations to the target are highly accurate, specifically when using moments estimated non-parametrically. This may indicate an over fitting behavior that only gets slightly more severe as we increase the number of monthly factors from two to five. But in that case the performance of the *real time* approximations deteriorates. Overall, the results show that it is better to use only a few static factors and a pre-whitening method to compute the indicator.

Focusing on the different factor space approximations, we conclude that the fit of the indicators with factors estimated by PC is quite similar to the fit of the indicators with factors estimated by GPC across all settings.

In line with the reported findings we select as the best performing filters the multivariate indicator using two monthly factors and parametric estimates of the moments and the multivariate indicator using five monthly factors and non-parametric estimates of the moments for the *real-time* and *today onwards* exercises, respectively.

Table 5 contains the evaluation results of the selected indicators sorted out by month of the quarter. Three things are worth stressing. Firstly, the clear monotonicity in the accuracy of the approximations across months; they tend to be more accurate in the third month, followed by the second month and then the first. Secondly, approximations using few monthly factors and a pre-whitening technique to derive the second order moments still seem to have a superior performance in *real-time* at the first and second month of the quarter than the competing settings. Recall that in the first two months of the quarter there are two quarters of information missing which make the approximation to the target a quite demanding task. Finally, the differences between the benchmark model and the multivariate frameworks in the first and second months of the quarter are still remarkable in the *today onwards* exercise.

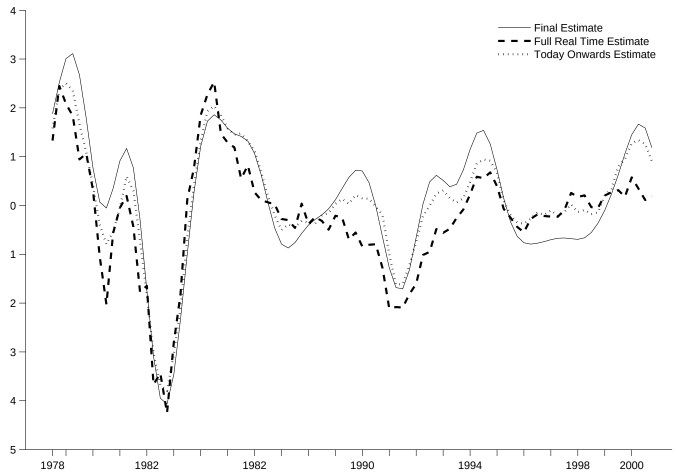
Table 5: Evaluation statistics for the approximation to business cycle fluctuations in the U.S. in every month of the quarter. Evaluation period: 1978(q3) - 2000(q4).

Performance with respect to business cycle fluctuations								
	Correlation		**Noise do Signal**		**Sign Concord.**		**% Correct Δ Sign**	
	Real Time	Today Onwards	Real Time	Today Onwards	Real Time	Today Onwards	Real Time	Today Onwards
BPF AR[a]								
1st/2nd month	0.72	0.73	0.62	0.62	0.71	0.73	0.57	0.59
3rd month	0.74	0.75	0.61	0.60	0.71	0.73	0.70	0.70
MBPF PC KERNEL								
$(k = 5 < \hat{r})$								
1st month	0.74	0.94	0.61	0.32	0.73	0.88	0.60	0.82
2nd month	0.75	0.95	0.60	0.29	0.74	0.90	0.59	0.89
3rd month	0.75	0.96	0.59	0.26	0.69	0.92	0.76	0.87
MBPF PC VAR								
$(k = 2 < \hat{r})$								
1st month	0.79	0.82	0.55	0.52	0.73	0.72	0.70	0.74
2nd month	0.81	0.83	0.52	0.50	0.73	0.72	0.71	0.77
3rd month	0.84	0.86	0.49	0.46	0.72	0.76	0.76	0.78

[a] BPF - univariate band-pass filter; AR - estimation of second moments by an AR model (BIC criterion for lag length); KERNEL - non-parametric estimation of second moments; MBPF - multivariate band-pass filter; PC - factor space estimated by principal components; VAR - estimation of second moments through pre-whitening with a SUR-VAR (BIC criterion for lag length in the various equations).

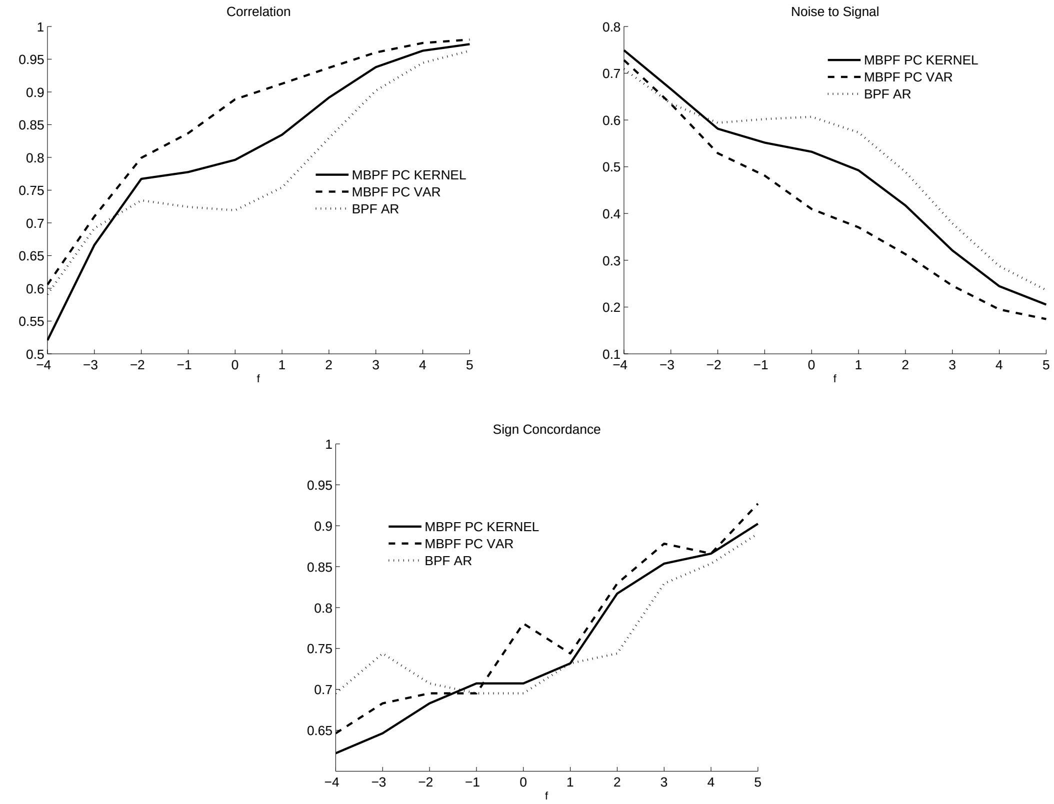
Figure 9 displays the best performing *real-time* and *today onwards* approximations in the most favorable situation, i.e., in the third month of the quarter, as well as the *final* estimates of business cycle fluctuations. Over most of the evaluation period there is an agreement in the behavior of the two real time estimates, the main distinction occurs in the less volatile part of the sample where the size of the approximations is different. Nonetheless, we visually confirm the quality of the approximations and argue that this particular multivariate method is the first to provide such accurate approximations to business cycle fluctuations in real-time.

Figure 9: U.S. GDP business cycle fluctuations: final estimates, *real-time* (MBPF PC VAR) and *today onwards* (MBPF PC KERNEL, $k = 5$ monthly factors) approximations. Evaluation period: 1978(q3)-2000(q4).



Moreover, we display in figure 10, the behavior of the best performing indicators in the third month of a given quarter, once 1,2,...,5 additional quarters of information (GDP and variables in the panel) are available as well as when 4,3,. . . ,0 quarters of information are missing. In the horizontal axis, -1 represents the *real-time* estimate, 1 represents the estimate obtained when one future data point is available and so forth. All the measures improve as more data becomes available and in all cases the multivariate filter has by far the best performance. Moreover, we note that even when 3 or 4 quarters of information are missing that the approximations provide some signal about the business conditions. The differences across methods tend to disappear only after 5 additional quarters of data are considered. This confirms that it is extremely hard to approximate business cycle fluctuations.

Figure 10: Evaluation of U.S. approximations to business cycle fluctuations: correlation with final estimates, noise to signal ratio and sign concordance when f future quarters of data are considered (MBPF PC VAR and MBPF PC KERNEL with $k = 5$ monthly factors). Evaluation period: 1978(q3)-2000(q4).



4.6.2 Smooth growth

We turn now to the real-time performance of smooth growth indicators. As before, we considered as benchmark the univariate filter using either estimated moments derived from an AR model or from a non-parametric estimator.

Table 6 reports the evaluation statistics for the approximations of smooth growth done in a third month of the quarter for the various dimensions under analysis.

Table 6: Evaluation statistics for the approximations to smooth growth in the U.S. in the third month of the quarter. Evaluation period: 1981(q3) - 2002(q4).

Performance with respect to smooth growth (3rd month of the quarter)						
	Correlation		**Noise do Signal**		**% Correct Δ Sign**	
	Real Time	Today Onwards	Real Time	Today Onwards	Real Time	Today Onwards
Benchmark Filters[a]						
BPF AR	0.72	0.76	0.59	0.55	0.72	0.74
BPF KERNEL	0.70	0.81	0.60	0.50	0.70	0.79
with factors, $k = 4 < \hat{r}, \hat{q} = 4$						
MBPF PC KERNEL	0.80	0.98	0.57	0.16	0.74	0.94
MBPF GPC KERNEL	0.80	0.98	0.57	0.16	0.76	0.93
with factors, $k = 2 < \hat{r}, \hat{q} = 4$						
MBPF PC KERNEL	0.83	0.96	0.52	0.22	0.77	0.86
MBPF GPC KERNEL	0.82	0.96	0.52	0.22	0.76	0.90
MBPF PC VAR	0.87	0.89	0.40	0.37	0.84	0.87
MBPF GPC VAR	0.85	0.89	0.42	0.38	0.81	0.83

[a] BPF - univariate band-pass filter; AR - estimation of second moments by an AR model (BIC criterion for lag length); KERNEL - non-parametric estimation of second moments; MBPF - multivariate band-pass filter; PC - factor space estimated by principal components; GPC - factor space estimated by generalized principal components; VAR - estimation of second moments through pre-whitening with a SUR-VAR (BIC criterion for lag length in the various equations).

Our findings are to a great extent similar to those of business cycle approximations since the filters' rank remains unchanged. The results show once more that our multivariate filters outperform the two benchmark filters in the various settings. Secondly, a distinct performance between the *real-time* and the *today-onwards* approximations, in particular for kernel based filters, is still observable. The *today onwards* approximations exhibit a nearly perfect linear association with the target, an extremely low noise to signal and a considerably high percentage of correctly detected sign changes. This high fit is most amplified if four monthly static factors are used, but on the other hand the performance of the *real-time* approximation deteriorates substantially. This does not occur if only two static factors are used and moments estimated parametrically. In fact the latter indicators are the best performing in *real-time*. Finally, the results are quantitatively the same for both factor space approximations. In general, we conclude that these findings alert again to the fact that fewer factors seem to produce a better out-of-sample fit to the target and justify why the approach followed by Altissimo et al. (2008)[21] might be misleading. Perhaps a larger time dimension would be needed to usefully incorporate a larger number of factors in the approximations.

In table 7 we report the evaluation criteria for the best performing filters in *real-time* and *today onwards* and for one benchmark filter by month of the quarter. The results show again that multivariate approximations are more accurate in the third month of the quarter, followed by the second month and then the first. Moreover, the results give the idea that with one less observation of GDP growth rate, as occurs in the first and second month of the quarter, that the performance of the univariate filter completely deteriorates while multivariate filters still perform considerably well, especially if only two monthly factors are included in the covariate set. Once more the findings suggest that few monthly factors are enough to compute remarkable good approximations to the signal in all three months of the quarter.

[21] Adding factors until the in-sample fit gains are negligible.

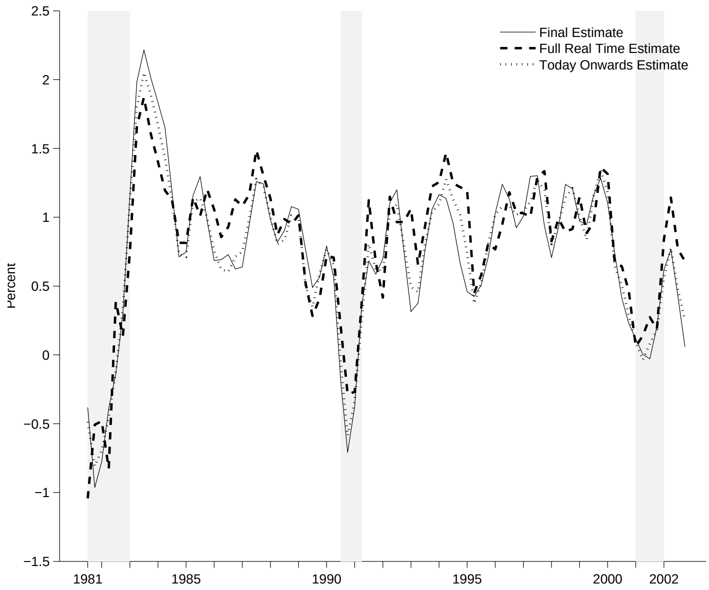
Table 7: Evaluation statistics for the approximations to smooth growth in the U.S. in every month of the quarter. Evaluation period: 1981(q3) - 2002(q4).

Performance with respect to smooth growth						
	Correlation		**Noise do Signal**		**% Correct Δ Sign**	
	Real Time	Today Onwards	Real Time	Today Onwards	Real Time	Today Onwards
BPF AR[a]						
1st/2nd month	0.31	0.40	0.80	0.75	0.72	0.72
3rd month	0.72	0.76	0.59	0.55	0.72	0.74
MBPF PC KERNEL						
($k = 4 < \hat{r}$)						
1st month	0.37	0.97	0.77	0.22	0.71	0.91
2nd month	0.56	0.98	0.73	0.19	0.74	0.93
3rd month	0.80	0.98	0.57	0.15	0.74	0.94
MBPF PC VAR						
($k = 2 < \hat{r}$)						
1st month	0.74	0.77	0.56	0.52	0.71	0.73
2nd month	0.81	0.85	0.48	0.43	0.77	0.76
3rd month	0.87	0.89	0.40	0.37	0.84	0.87

[a] BPF - univariate band-pass filter; AR - estimation of second moments by an AR model (BIC criterion for lag length); KERNEL - non-parametric estimation of second moments; MBPF - multivariate band-pass filter; PC - factor space estimated by principal components; VAR - estimation of second moments through pre-whitening with a SUR-VAR (BIC criterion for lag length in the various equations).

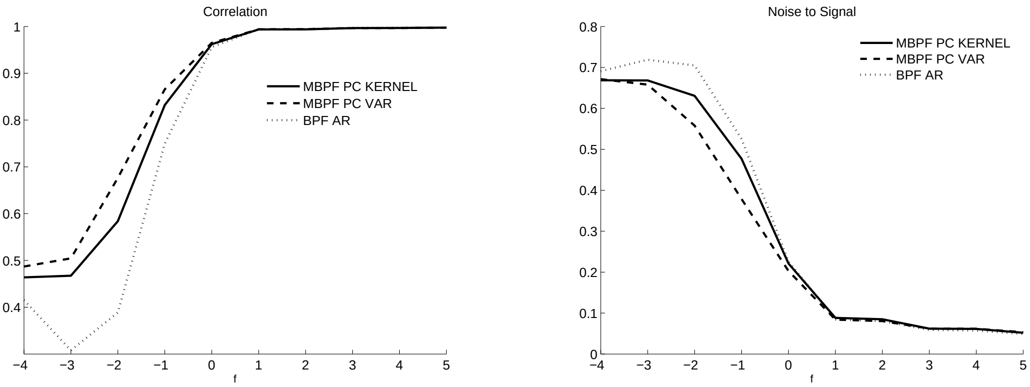
Figure 11 displays the *final* approximation against the best performing indicators for the smooth growth signal in the third month of the quarter. The shaded areas represent recessions as dated by the National Bureau of Economic Research. Although, *real-time* estimates are slightly more ragged than *today onwards* estimates we may conclude that both approximations are extremely accurate and timely track the upturns and downturns of output smooth growth.

Figure 11: Smooth growth of U.S. GDP: final estimates, *real-time* (MBPF PC VAR) and *today onwards* (MBPF PC KERNEL, $k = 4$ monthly factors) approximations. Evaluation period: 1981(q3)-2002(q4).



Finally, we look at the behavior of the estimates in a third month once $1, 2, \ldots, 5$ new observations are available and when $4, 3, \ldots, 0$ observations are missing either of GDP or of the panel variables (see figure 12). Again, in the horizontal axis -1 represents the *real-time* approximation, 1 represents the estimate obtained when one future data point is available and so forth. In short, both measures quickly converge to the expected target while the differences between frameworks vanish after 1 additional quarter of information is available. When some quarters of data are missing, the multivariate indicators still outperform the benchmark filter and more importantly still have an informative signal. The univariate approximation has a poor performance when more that one quarter of data is missing, confirming visually the results of table 7.

Figure 12: Evaluation of U.S. approximations to smooth growth: correlation with final estimates and noise to signal ratio when f future quarters of data are considered (MBPF PC VAR and MBPF PC KERNEL with $k = 4$ monthly factors). Evaluation period: 1981(q3)-2002(q4).



4.7 Forecast performance

4.7.1 Quarterly growth

As a by-product of the previous analysis we use the best multivariate indicator (MBPF PC VAR with $k = 2$ monthly factors) to forecast the quarterly GDP growth. We believe that this exercise can reveal important findings for two reasons. Firstly, the results in figure 12 indicate that the multivariate approximations are still informative even when 3/4 quarters of data are missing. Secondly, we rely on the unexplored idea that targeting a smooth version of a time series may be more useful than targeting the original series if the noisy fluctuations of a time series are indeed unpredictable. Most forecast models usually target the growth rate itself meaning that the short-run fluctuations are also being approximated regardless of the assumed restriction.

To assess the forecast performance of the best multivariate approximation to smooth growth signal we selected for comparison other six competing methods. Those methods are the following:

- the autoregressive model with lag length determined by the BIC criterion, denoted AR;
- the linear projection targeting GDP growth with second moments estimated by a pre-whitening technique, denoted VAR;
- the diffusion index model which amounts to a regression of GDP growth rate on its past

values, with lag length determined by the BIC criterion, and (a maximum of 2) split factors estimated by principal components as in Stock and Watson (2002a, 2002b), denoted DI-AR SW;

- the univariate approximation of the band-pass filter to smooth growth with non-parametric estimation of second order moments, denoted BPF KERNEL;
- the univariate approximation of the band-pass filter to smooth growth with second order moments derived from an AR model with lag length determined by the BIC criterion, denoted BPF AR;
- the multivariate approximation of the band-pass filter to smooth growth with second order moments estimated non-parametrically and using $k = 2$ split monthly factors estimated by principal components, denoted MBPF PC KERNEL.

For each method, we approximated the growth rate or instead the smooth growth rate at various forecast horizons (maximum of 3) and compare those estimates with the already available observations of quarterly GDP growth rate. All forecasts were made in *real-time*, meaning that data standardization, estimation of factors and moments, optimization and filtering are performed using only the data in fact available. The lag length selection and the number of factors identification were also done in real-time.

As evaluation statistic we have chosen, as is common practice, the relative MSE; setting as benchmark the AR model. In detail, we compute the ratio of the MSE of the forecast of each method to the MSE of the univariate regression forecast with lag length determined by the BIC criterion.

In table 8 we present the results for forecasts made in the third month of the quarter, from 1981(q3) to 2003(q4). At one-quarter horizon the DI AR - SW model outperforms all competing forecast methods, but relatively to the multivariate band-pass filters the improvement is quite small. Moreover, the parametric multivariate approximation to smooth growth improves over the comparable VAR model at one, two and three steps ahead. The latter model uses exactly the same estimated second order moments as MBPF PC VAR but instead targets the growth rate itself. Hence, this signals that it can be relevant for forecasting purposes to target only the predictable component of a series. In the case of 2 steps ahead forecasts, most methods perform as good as or slightly better than the autoregressive model, the exception being the DI AR - SW and VAR models. But note that RMSE of the benchmark forecasts oddly declines. For the highest horizon reported, we conclude that all methods perform poorly because the RMSE of the AR model is basically the standard deviation of GDP growth, meaning that we are better off with the mean growth rate as forecast. Although not reported, we computed forecasts at more than 3-steps ahead but the results confirmed the usefulness of all methods to produce accurate forecasts at long horizons.

Table 8: Ratio of the mean squared error of the forecasts of each method to the mean squared error of an univariate regression forecast (BIC for lag length). Evaluation period: 1981(q3) - 2003(q4).

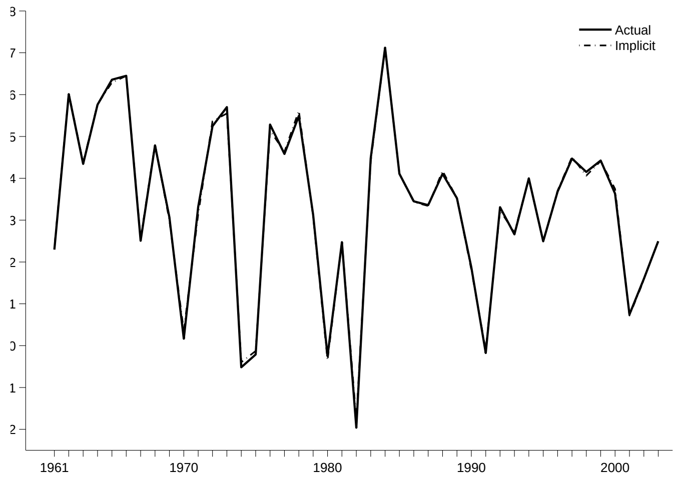
Simulated out-of-sample forecasting results: U.S. GDP growth rate			
		Relative MSE	
	One step ahead	2 steps ahead	3 steps ahead
Method	(current quarter)	(1 quarter ahead)	(3 quarters ahead)
VAR	0.84	1.10	1.30
BPF AR	1.00	1.00	0.98
BPF KERNEL	0.95	0.99	1.00
$k = 2 < \hat{r}$			
MBPF PC AR	0.76	0.95	0.98
MBPF PC KERNEL	0.74	0.99	1.14
DI AR - SW	0.69	1.19	1.49
RMSE, AR	0.00508	0.00496	0.00524

The main conclusions are that the multivariate approximations to smooth growth are clearly useful for short-term forecasting of the quarterly growth rate of GDP and that it is extremely difficult to forecast the growth rate at long horizons. In fact, there are some studies reporting that AR models forecast sometimes better than more sophisticated models at several steps ahead (see Runstler et al., 2008 and Angelini et al., 2008). Our findings suggest the same behavior.

4.7.2 Yearly growth

The previous results show that approximations to smooth growth are of limited use in forecasting quarterly growth but we argue that they can have a much more striking role in forecasting yearly GDP growth. Again, we invoke the fact that the multivariate approximations to smooth growth are still informative at longer horizons, as revealed in figure 12, while table 8 makes clear that all methods are uselessness in forecasting at more than 2 quarters ahead. In addition, the yearly growth implied by the quarterly smooth growth rates is indistinguishable from the yearly growth rate implied by the quarter on quarter GDP growth rates (see figure 13). Thus, by approximating accurately smooth growth at various horizons, we will approximate accurately yearly growth. If we use instead useless quarterly growth forecasts to forecast annual GDP growth, the results are unlikely to be promising.

Figure 13: U.S. yearly real GDP growth rate: actual and implicit in smooth growth of quarterly real GDP.



For the approximations that target the smooth growth, we forecast the yearly growth rate using the forecasts of the missing quarterly GDP growth rates and the most up-to-date approximations to smooth growth for the relevant past quarters. Therefore, in the filter-based approximations we do not use the already known quarterly growth rates, we use always the latest vintage of approximations to quarterly smooth growth rates. For the remaining methods, the yearly growth rate is derived from known quarterly growth rates as well as from the forecasts for the missing quarters. Then we compute the MSE between the actual yearly rate and that implied by the forecasts.

Given the results of table 8, when there are more than 2 quarters of data missing the 3 and 4-quarters ahead forecasts are replaced by real-time estimates of the mean growth rate of quarterly GDP. For example, at the end of the first quarter (March), no information on quarterly GDP of the current year is available, so forecasts for the four quarters are necessary. As explained earlier we assume the mean growth rate as the forecast for the third and fourth quarters and in practice only forecast the growth rate of the first and second quarters. Once in December we have information on GDP growth from the first three quarters and only need to forecast the last quarter of the current year. Then, since the multivariate approximations outperform the univariate methods in real-time we expect a superior forecast performance of the first methods at this point of the year.

Table 9 displays the results for forecasts made at the end of each quarter, this means in March, June, September and December. We report the MSE of each forecast approach to the MSE of an AR model (BIC for lag length) and as forecast methods we considered BPF AR, MBPF PC KERNEL, MBPF PC VAR and DI-AR SW, which denote the same methods introduced in the previous subsection.

Table 9: Ratio of the mean squared error of the forecasts of each method to the mean squared error of an univariate regression forecast (BIC for lag length). Evaluation period: 1985(q1) - 2003(q4).

Simulated out-of-sample forecasting results: U.S. GDP yearly growth rate

	Relative MSE of forecasts made at the end of:			
Methods	1st quarter	2nd quarter	3rd quarter	4th quarter
BPF AR	1.13	0.92	1.06	0.86
$k = 2 < \widehat{r}$				
MBPF PC KERNEL	1.02	0.71	0.71	0.64
MBPF PC VAR	1.12	0.63	0.54	0.55
DI AR - SW	0.98	0.75	0.55	0.49
RMSE, AR	0.0076	0.0057	0.0035	0.0014

The relative MSE shows that in March all methods perform similarly to the benchmark method, this is not surprising since there is no information on aggregate activity for the current year. In subsequent quarters, we report significant gains in forecasts made by multivariate approximations, in particular for our best performing *real-time* indicator of smooth growth and for diffusion indexes method. The latter is an important competitor as expected due to the results in table 8. The univariate method also improves as more data is becoming available but at a slower rate than all the other methods. This reveals that monthly information might have a role in forecasting yearly growth. As anticipated, in December (4th quarter) all methods perform very well since there is only one quarter of information missing.

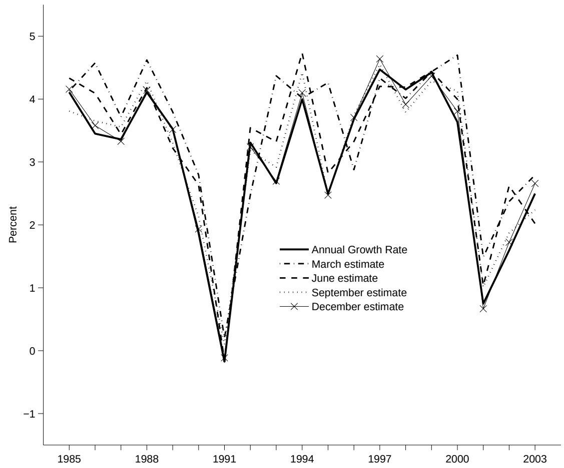
In addition, the RMSE for forecasts of the yearly growth made in the 12 months of the year using the multivariate approximation with moments derived from a SUR-VAR model and 2 monthly factors is reported in table 10. As we approach the end of the year the RMSE gets smaller and the significant jumps at the end of each quarter reveal that the forecasts improve mainly due to the release of GDP quarterly figures. Nevertheless, the new monthly information available within each quarter generally improves the forecasts.

Table 10: Root mean squared error of the forecasts of yearly growth made at the end of each month. Evaluation period: 1985(q1) - 2003(q4).

Simulated out-of-sample forecasting results: GDP yearly growth rate	
	RMSE of forecasts
Forecast moment	MBPF PC VAR
January	0.0090
February	0.0086
March	0.0081
April	0.0087
May	0.0085
June	0.0046
July	0.0044
August	0.0041
September	0.0026
October	0.0028
November	0.0027
December	0.0010

Finally, figure 14 displays the actual yearly growth rate against the forecasts made in March, June, September and December using the multivariate approximation. The forecasts correctly predict the clear downturns in 1991 and 2001 already in March. This gives an idea of the accuracy of the forecast performance of the multivariate approximation concerning the yearly growth rate.

Figure 14: Yearly GDP growth rate and real-time forecasts made at the end of March, June, September and December.



5 Conclusions

We have discussed the characteristics of four optimal approximations to the infinite order moving average filter that exactly extracts from a time series fluctuations within a specific range of periodicities. We argued that the Hodrick and Precostt and the Baxter and King filters are not useful to be used given our objectives. The first one because it extracts business cycle fluctuations that include short-term variability, which disagrees with our adopted definition. And the second because it does not produce estimates at the endpoints of the sample. Therefore, we have analyzed in detail approximations to business cycle fluctuations and to output smooth growth provided only by two optimal filters: the univariate filter suggested by Christiano and Fitzgerald (2003) and the multivariate filter suggested by Valle e Azevedo (2007). Both filters are designed to extract exactly the same fluctuations from a time series but use different information sets.

Our empirical application aimed at evaluating the real-time performance of these two indicators. So, the targets were defined on real GDP of the current quarter and the release delays of all the variables involved were taken into account. The use of a multivariate approach allowed to illustrate how common factors extracted from a panel of variables can be successfully incorporated in such an analysis. Overall, the results revealed that the multivariate indicator outperforms in various dimensions the univariate indicator. Furthermore, the approximations are reliable and accurate, even in the first two months of a quarter. From these findings we concluded that methods which explore various sources of information seem more suitable to track economic activity in real-time.

Finally, we used our best performing filter in real-time to forecast either quarterly or yearly GDP growth rate. What distinguishes our method from other forecasting methods is the fact that we target only the medium to long run component of output growth rate instead of the observable growth rate with the erratic fluctuations. We argued that if the short-run fluctuations are in fact unpredictable, which is not commonly assumed by most models, then targeting a smooth version of a time series may be more useful for forecasting purposes than targeting the original time series. From the comparison with other methods we concluded that moderate gains were obtained in short term forecasts of quarterly growth rate. But this approach has a more striking role in forecasting yearly growth rate.

Future research encompasses the approximation of any other important signal, such as core inflation, and the assessment of the advantages in forecasting exercises of targeting a smooth version of a time series.

APPENDIX

Appendix A: Spectral Analysis

Univariate

Time series analysis can be performed in two different perspectives: in the time domain or in the frequency domain. Typically it is carried out in a time domain perspective using, for example, the autocovariance and autocorrelation functions. However, in some specific contexts the frequency perspective has its advantages and so in the following paragraphs we present an overview of concepts associated to frequency domain theory.

The analysis of time series in the frequency domain is called spectral analysis and, essentially, it is based on the concepts of Fourier transform (henceforth FT) and inverse Fourier transform (henceforth IFT). In the time domain it is common practice to divide a time series into three components (trend, cycle and noise) according to their persistence in time. While from the frequency domain perspective a time series is considered as the sum of components with different frequencies of oscillation. As we will see, it is possible to establish a correspondence between the two approaches, which proves that they are theoretically equivalent.

The definitions of FT and IFT are deduce from the Fourier representation of a finite sequence of numbers that must be transformed to accommodate an infinite stochastic time series. To better comprehend this idea, let $\{z_t\}_{t=1}^{T}$ be a time series of length T. Consequently, the so called Fourier series of this sequence of T numbers is defined as

$$z_t = \sum_{k=1}^{\lfloor \frac{T}{2} \rfloor} \left[a_k \cos(t\omega_k) + b_k \sin(t\omega_k) \right], \qquad t = 1, 2, \ldots, T$$

where a_k and b_k are called Fourier coefficients and $\omega_k = \frac{2\pi k}{T}$ for $k = 1, ..., \lfloor \frac{T}{2} \rfloor$, with $\lfloor \cdot \rfloor$ denoting the largest integer smaller or equal than the operand, are known as the Fourier frequencies. Through this expression $\{z_t\}$ is described as a linear combination of orthogonal and periodic[22] functions. Note also that the summation entails $\lfloor \frac{T}{2} \rfloor$ processes with different frequencies of

[22] A periodic function is a function that repeats itself after some period and this property is called periodicity. So, a function f is periodic with period p if $f(x+p) = f(x)$. As an example, the cosine and sine functions are periodic functions of period 2π.

oscillation. Furthermore, using Euler's formula[23] the Fourier series of $\{z_t\}$ can be rewritten as

$$z_t = \begin{cases} \sum_{k=-\frac{(T-1)}{2}}^{\frac{(T-1)}{2}} c_k e^{i\omega_k t}, & T \text{ odd} \\ \sum_{k=-\frac{T}{2}+1}^{\frac{T}{2}} c_k e^{i\omega_k t}, & T \text{ even} \end{cases}$$

Christiano and Fitzgerald (1998) show, in the appendix of their paper, that with this representation, after the moment T, all data points are perfectly predicted. So, within the economic context this representation seems inadequate since we regard a particular time series $\{y_t\}$ as one of the many possible outcomes of an underlying stochastic process $\{y(w,t) : t = 0, \pm 1, \pm 2, ...\}$, which by definition is never predictable. Hence, we must adjust the Fourier series definition to our needs by assuming that the number of periodic functions is infinitely large or, equivalently, that the Fourier frequencies can be made as small as desired in the range $[-\pi; \pi]$. This is the idea behind the Spectral Representation Theorem which states that any real-valued, non-deterministic, absolutely summable[24] and covariance stationary process can be expressed as the weighted sum of orthogonal periodic components of different frequencies:

$$\begin{aligned} y_t &= \int_{-\pi}^{\pi} \left(a(\omega)\cos(\omega t) + b(\omega)\sin(\omega t)\right) d\omega \\ &= \int_{-\pi}^{\pi} f(\omega) e^{i\omega t} d\omega, \qquad \text{t} = 0, \pm 1, \pm 2, \ldots \end{aligned}$$

This last result defines the IFT of $f(\omega)$ which is also valid for sequences of finite duration and if $\{y_t\}$ is squared summable[25].

Finally, the FT of $\{y_t\}$ will be a function of frequency ω and is defined as

$$f(\omega) = \frac{1}{2\pi} \sum_{t=-\infty}^{\infty} y_t e^{-i\omega t}, \quad -\pi \leq \omega \leq \pi$$

with ω measured in radians. The FT and the IFT functions form a Fourier transform pair

[23] $e^{i\theta} = \cos(\theta) + i\sin(\theta)$

[24] $\{\gamma_k\}$ is absolutely summable if

$$\sum_{k=-\infty}^{\infty} |\gamma_k| < \infty$$

[25] $\{\gamma_k\}$ is square summable if

$$\sum_{k=-\infty}^{\infty} \gamma_k^2 < \infty$$

and can be applied to any time series that fulfill the requirements of the Spectral Representation Theorem. For more general time series, for example time series that are not absolutely summable, one must use the Fourier-Stieltjes integral in order to define a spectral representation[26].

In the time domain the joint distribution of a time series can be characterized by its autocovariance function while in the frequency domain the time series properties are exploited using the spectral density function. The analysis of this function, also known as spectrum of $\{y_t\}$ is called spectral analysis and the function itself corresponds to the FT of the absolutely summable autocovariance function of $\{y_t\}$. Note, that this fact stresses that the two approaches are equivalent because spectral analysis can be seen as the equivalent of the autocovariance function analysis performed in the time domain. The spectrum of $\{y_t\}$ is

$$S(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma_k e^{-i\omega k}, \quad -\pi \le \omega \le \pi$$

where γ_k =E$[(y_t - \mu)(y_{t-k} - \mu)]$, $k = 0, \pm 1, \pm 2, \ldots$, denotes the autocovariance function of the stationary process $\{y_t\}$ with mean μ for every t.

At each frequency ω the spectrum is a complex number and thus can be rewritten in polar form as

$$S(\omega) = c(\omega) + iq(\omega) = |S(\omega)| \, e^{i\phi(\omega)}, \quad -\pi \le \omega \le \pi$$

where $c(\omega)$ and $q(\omega)$ correspond to the real and imaginary parts, respectively, of the complex number, $|S(\omega)| = \sqrt{c(\omega)^2 + q(\omega)^2}$ is typically referred to as the amplitude, magnitude or gain and $\phi(\omega) = \arctan\left(\frac{q(\omega)}{c(\omega)}\right)$ the angle phase of the same complex number.

The spectrum can also be expressed in terms of the autocovariance generating function denoted by $g(z) = \sum_{k=-\infty}^{\infty} \gamma_k z^k$ for the sequence of autocovariances γ_k, $k = 0, \pm 1, \pm 2, \ldots$. The coefficient of z^0 is de variance of $\{y_t\}$ and the autocovariance of $\{y_t\}$ at lag k can be assess by the coefficient of z^k or z^{-k}. Thus, it is straightforward to get

$$S(\omega) = \frac{1}{2\pi} g(e^{-i\omega}). \tag{19}$$

Moreover, when y_t is a real-valued process with absolutely summable autocovariance function

[26]For further reading on this subject see Priestley(1981).

the spectrum can be written as

$$\begin{aligned}
S(w) &= \frac{1}{2\pi}\sum_{k=-\infty}^{\infty}\gamma_k e^{-iwk} \\
&= \frac{1}{2\pi}\left[\gamma_0 + \sum_{k=1}^{\infty}\gamma_k e^{-iwk} + \gamma_{-k}e^{iwk}\right] \\
&= \frac{1}{2\pi}\left[\gamma_0 + \sum_{k=1}^{\infty}\gamma_k \left(e^{-iwk} + e^{iwk}\right)\right] \\
&= \frac{1}{2\pi}\left[\gamma_0 + 2\sum_{k=1}^{\infty}\gamma_k \cos(wk)\right] \\
&= \frac{1}{2\pi}\gamma_0 + \frac{1}{\pi}\sum_{k=1}^{\infty}\gamma_k \cos(wk), \qquad -\pi \leq w \leq \pi
\end{aligned}$$

where in the second equality we use the symmetry property of the autocovariance function and in the third the fact that $e^{-i\theta} + e^{i\theta} = 2\cos(\theta)$. For this case the spectrum has some particular features:

- $S(\omega)$ is a continuous, real-valued and nonnegative function;
- $S(\omega)$ is a periodic function with period of 2π, i.e., $S(\omega) = S(\omega + 2\pi)$;
- $S(\omega)$ is a symmetric and even function, i.e., $S(\omega) = S(-\omega)$ because the cosine function is also even. The symmetry permits focusing only on frequency values in the $[0, \pi]$ interval.
- ω represents the frequency of a cycle expressed in radians and is related to the time domain by $\omega = \frac{2\pi}{p}$, where p is the periodicity, i.e., the amount of time required to complete a whole cycle[27]. Hence, high values of p are related to low frequencies whereas short periodicities are related to high frequencies.

The IFT of $S(\omega)$ is expressed as

$$\gamma_k = \int_{-\pi}^{\pi} S(\omega)e^{i\omega k}d\omega, \quad k = 0, \pm 1, \pm 2, \dots$$

[27] The function argument ωk is the period of oscillation and to express it in units of time we must quantify the amount by which ωk must increase so that the function repeats itself. Since we know that the function period is 2π we can write

$$k_2 > k_1 : k_2\omega - k_1\omega = 2\pi \Leftrightarrow k_2 - k_1 = \frac{2\pi}{\omega} \Leftrightarrow p = \frac{2\pi}{\omega}.$$

Then the period of oscillation of $\cos(\omega k)$, in units of time, is $\frac{2\pi}{\omega}$.

which permits to recover the autocovariance function from the spectrum. Further, when $k = 0$ we obtain the variance of the series,

$$\gamma_0 = var(y_t) = \int_{-\pi}^{\pi} S(\omega) d\omega$$

as the product of the different contributions of each frequency of the spectrum.

To give an idea of how the spectral density function may look we now present this function for some well known processes. The first example is the building block of time series models, i.e., the white noise process $\{\varepsilon_t\}$. It is a process with constant mean $\mathrm{E}(\varepsilon_t) = \mu_\varepsilon \ \forall t$, usually set to zero, constant variance $\mathrm{Var}(\varepsilon_t) = \sigma_\varepsilon^2 \ \forall t$ and uncorrelated at all leads and lags $\gamma_\varepsilon(k) = 0$ for $|k| \geq 1$. Hence, the spectral density function has the following structure

$$S_\varepsilon(\omega) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \gamma_\varepsilon(j) e^{-i\omega j} = \frac{\sigma_\varepsilon^2}{2\pi}, \qquad -\pi \leq \omega \leq \pi$$

which shows that all frequencies from $-\pi$ to π are equally powered by the same constant.

The second example is the general autoregressive moving average (ARMA) model with autoregressive order p and moving average order q, i.e.,

$$y_t = \alpha + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \\ + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q} \quad (20)$$

with $\{\varepsilon_t\}$ a zero mean white process with constant variance, $\mathrm{Var}(\varepsilon_t) = \sigma_\varepsilon^2 \ \forall t$. If the process is stationary and invertible then we can rewrite it as a moving average process of infinite order such that

$$y_t = c + \psi(L)\varepsilon_t$$

where $\psi(L) = \frac{1+\theta_1 L + \theta_2 L^2 + \ldots + \theta_q L^q}{1-\phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p}$. The autocovariance generating function of the former representation is well known and given by

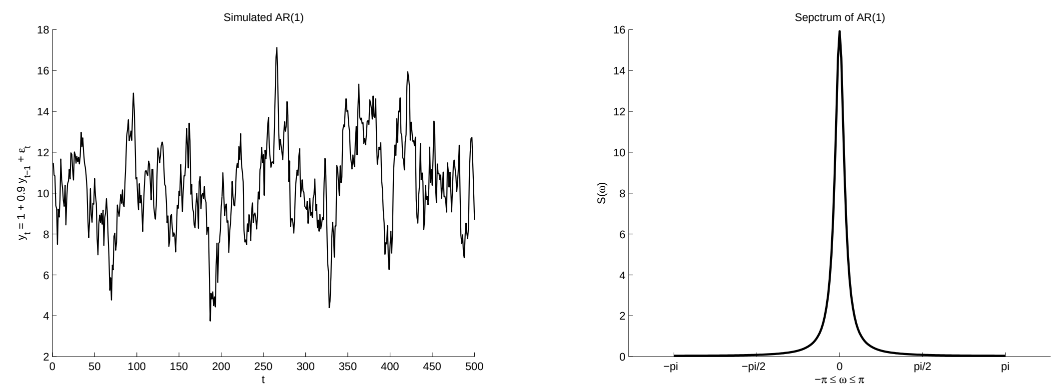
$$g_y(z) = \sigma_\varepsilon^2 \psi(z) \psi(z^{-1})$$

with z taken to be a complex scalar. It follows from equation (19) that the spectral density function of the ARMA(p,q) model is

$$S_y(\omega) = \frac{\sigma_\varepsilon^2}{2\pi}\psi(e^{-i\omega})\psi(e^{i\omega}) = \\ = \frac{\sigma_\varepsilon^2}{2\pi} \cdot \frac{1+\theta_1 e^{-i\omega}+\ldots+\theta_q e^{-iq\omega}}{1-\phi_1 e^{-i\omega}-\ldots-\phi_p e^{-ip\omega}} \cdot \frac{1+\theta_1 e^{i\omega}+\ldots+\theta_q e^{iq\omega}}{1-\phi_1 e^{i\omega}-\ldots-\phi_p e^{ip\omega}} \tag{21}$$

Figure 15 displays 500 values from a first order autoregressive model (henceforth AR(1))[28] process and the spectral density function for the same AR(1) process in the interval $[-\pi, \pi]$. It results from a simulation exercise setting $y_0 = 2$, $\phi = 0.9$, $\alpha = 1$ and $\{\varepsilon_t\}$ equal to a Gaussian white noise with zero mean and unit variance. Since $|\phi| < 1$ it is guarantee that the process is stationary. The horizontal axis of the right graph measures the frequency in radians and the vertical axis the spectrum. The closer ω is to zero the lower the frequency and the longer the correspondent cyclical period (periodicity). In this specific case the frequencies near zero are of greatest importance which means that long periodicities dominate this type of process. High frequencies near π are of little importance and therefore short duration movements do not influence much of the time series variation. Different values of ϕ will produce different spectrum shapes but the main point is that an autoregressive process always has a high spectrum power at the neighborhood of the zero frequency.

Figure 15: Left panel displays a simulated AR(1) model and right panel displays the population spectrum for an AR(1) model.



[28] An autoregressive model of order 1 is written as $y_t = \alpha + \phi y_{t-1} + \varepsilon_t$ and its spectrum equals $S(\omega) = \frac{\sigma_\varepsilon^2}{2\pi} \cdot \frac{1}{1+\phi^2-2\phi\cos(\omega)}$

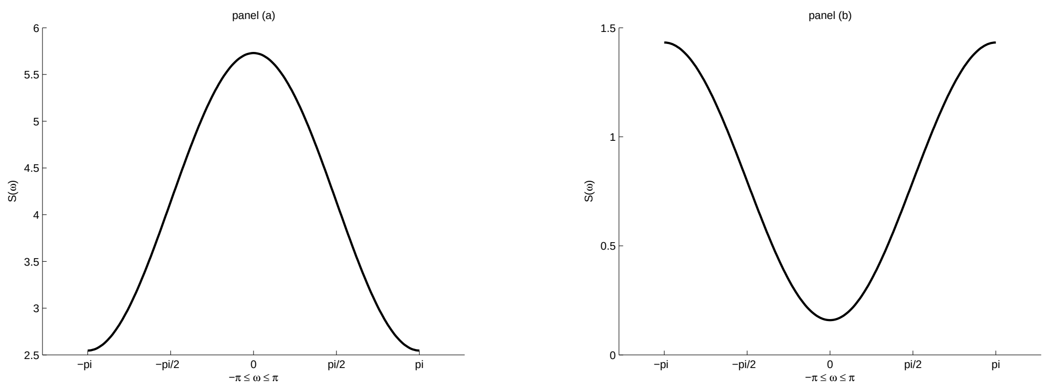
A random walk process can be seen as being an AR(1) process, i.e., $\phi = 1$. A parameter value of one originates a strong persistence process and so the random walk is not a covariance stationary process. As a consequence, its autocovariance function is not absolutely summable and the spectral density function is not well defined at the zero frequency. Even so, making $\phi \to 1$ in the AR(1) spectrum we get

$$\underset{\phi \to 1}{Lim} S(\omega) = \frac{\sigma_\varepsilon^2}{4\pi} \frac{1}{1 - \cos(\omega)}$$

This function is well behave except at the zero frequency where it becomes infinite.

Finally, the shape and pattern of the spectrum of a first order moving average process (henceforth MA(1))[29] depend on the value of θ. Panel A in figure 16 shows the spectrum in the interval $[-\pi, \pi]$ when $\theta > 0$ and panel B exhibits the case $\theta < 0$. The first panel shows that the low frequency components dominate the time series while the second shows the opposite. When $\theta < 0$ the time series contains mostly high frequency components.

Figure 16: Population spectrum for an MA(1) model. Panel A. $\theta = 5$; Panel B. $\theta = -2$.



In sum, we may conclude that low frequency dominating spectrum indicates a relatively smooth series, whereas a high frequency dominating spectrum implies a more ragged series. The population characteristics permit to infer that a sample spectrum with a strong peak near the zero frequency signals the possible presence of a trend (stochastic or deterministic) in the time

[29] The first order moving average model is given by $y_t = \varepsilon_t + \theta\varepsilon_{t-1}$ and its spectrum equals $S(\omega) = \frac{\sigma_\varepsilon^2}{2\pi}\left(1 + \theta^2 + 2\theta\cos(\omega)\right)$.

series and that we may have to consider the time series in first differences to obtain a covariance stationary process. Moreover, from the sample spectrum behavior we get an idea of the kind of cyclical movements that predominate in the time series.

Bivariate

In the previous section spectral analysis for the univariate case was presented but obviously all this frequency theory can be extended to the multivariate case. As an illustration we now turn to the discussion of spectral analysis in a bivariate context[30]. Let $y_t = [x_t \;\; z_t]'$ be a zero mean bivariate stochastic process with covariance matrix $\Gamma(k) = \mathrm{E}\left[y_t y'_{t-k}\right]$, for $k = 0, \pm 1, \pm 2, \ldots$. Following the fundamental idea of the univariate case, bivariate spectral analysis consists of decomposing the covariance between two time series into orthogonal frequency components. In order to define the spectral density function of $\{y_t\}$ we use the covariance matrix, instead of the autocovariance function. Under the assumption that $\Gamma(k)$ is absolutely summable it follows that

$$\begin{aligned} S_y(\omega) &= \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \Gamma(j) \mathrm{e}^{-i\omega j} = \\ &= \frac{1}{2\pi} \begin{bmatrix} \sum_{j=-\infty}^{\infty} \gamma_x(j) \mathrm{e}^{-i\omega j} & \sum_{j=-\infty}^{\infty} \mathrm{E}\left[x_t z_{t-j}\right] \mathrm{e}^{-i\omega j} \\ \sum_{j=-\infty}^{\infty} \mathrm{E}\left[z_t x_{t-j}\right] \mathrm{e}^{-i\omega j} & \sum_{j=-\infty}^{\infty} \gamma_z(j) \mathrm{e}^{-i\omega j} \end{bmatrix} = \\ &= \begin{bmatrix} S_{xx}(\omega) & S_{xz}(\omega) \\ S_{zx}(\omega) & S_{zz}(\omega) \end{bmatrix}, \qquad -\pi \leq \omega \leq \pi \end{aligned}$$

where $S_{xx}(\omega)$ and $S_{zz}(\omega)$ are the spectral density functions of $\{x_t\}$ and $\{z_t\}$, respectively, as defined in the univariate case and $S_{xz}(\omega)$ and $S_{zx}(\omega)$ are known as cross spectral densities or cross spectra. The two former functions split the covariance between $\{x_t\}$ and $\{z_t\}$ into uncorrelated components with different frequencies.

Previously, we have stated that the symmetry property around the zero lag of the autocovariance function implied that the univariate spectrum was a real and symmetric function. However, the bivariate spectrum may be complex even if the processes $\{x_t\}$ and $\{z_t\}$ are real valued because the symmetry property does not hold for the cross covariance function, i.e., $\mathrm{E}[x_t z_{t-k}] \neq \mathrm{E}[z_t x_{t-k}]$

[30]The extension to the more than two variables case is a straightforward adaptation of the bivariate case. For further details see Priestley(1981) or Hamilton(1994).

$\forall t$ and $k \neq 0$. This means that the cross spectrum is a complex number at each frequency and therefore has the following representation

$$S_{xz}(\omega) = c_{xz}(\omega) + iq_{xz}(\omega), \quad -\pi \leq \omega \leq \pi$$

where the real part, $c_{xz}(\omega)$, is called the cospectrum and the imaginary part, $q_{xz}(\omega)$, the quadrature spectrum. Again, any complex number may also be expressed in polar form as

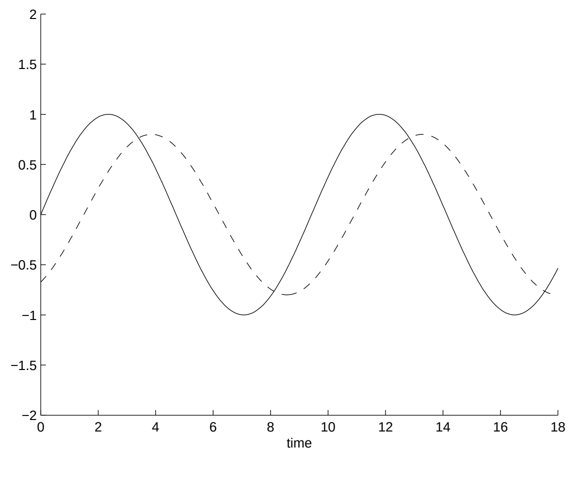
$$S_{xz}(\omega) = \alpha_{xz}(\omega)\mathrm{e}^{i\phi_{xz}(\omega)}, \quad -\pi \leq \omega \leq \pi$$

and the equivalence between representations is given by

$$\begin{aligned} \alpha_{xz}(\omega) &= |S_{xz}(\omega)| = \sqrt{c_{xz}^2(\omega) + q_{xz}^2(\omega)} \\ \phi_{xz}(\omega) &= \arctan\left(\frac{q_{xz}(\omega)}{c_{xz}(\omega)}\right) \end{aligned}$$

The function $\alpha_{xz}(\omega)$ is called the cross amplitude spectrum and measures the difference in amplitudes between $\{x_t\}$ and $\{z_t\}$ at each ω frequency (vertical difference). While the function $\phi_{xz}(\omega)$ is called the phase spectrum and gives the phase shift between frequency components of $\{x_t\}$ and $\{z_t\}$ (horizontal difference), i.e., measures the extent to which each frequency component of one time series leads/lags the other. So, if $\phi_{xz}(\omega) > 0$, $\forall\omega$ then we say that $\{x_t\}$ leads $\{z_t\}$ at all frequencies and when $\phi_{xz}(\omega) < 0$ $\forall\omega$, we say that $\{x_t\}$ lags $\{z_t\}$ at all frequencies. The phase shift expressed in time units equals $\frac{\phi_{xz}(\omega)}{\omega}$ and may be understood as the time delay between components of the two processes. Figure 17 illustrates the hypothetic effects of the gain and phase shift on the path of two time series. The horizontal difference between the lines reflects the phase shift between the time series. So we may say that the time series represented by the dashed line lags the other time series. The vertical difference measures the difference in the range of the time series. Overall, we see in this example that the gain effect shrinks the range of the periodic function while the phase shift delays the periodic function.

Figure 17: Gain and phase shift.



The covariance matrix of the bivariate process for all lags is recovered from the inverse process

$$\Gamma(k) = \int_{-\pi}^{\pi} S_y(\omega) e^{i\omega k} d\omega, \qquad k = 0, \pm 1, \pm 2, \dots$$

with the off-diagonal giving the IFT of the cross spectrum

$$\begin{aligned} \mathrm{E}\left[x_t z_{t-k}\right] &= \int_{-\pi}^{\pi} S_{xz}(\omega) e^{i\omega k} d\omega, \qquad k = 0, \pm 1, \pm 2, \dots \\ \mathrm{E}\left[z_t x_{t-k}\right] &= \int_{-\pi}^{\pi} S_{zx}(\omega) e^{i\omega k} d\omega, \qquad k = 0, \pm 1, \pm 2, \dots \end{aligned}$$

and the diagonal the IFT of the univariate spectrum

$$\begin{aligned} \gamma_x(k) &= \int_{-\pi}^{\pi} S_{xx}(\omega) e^{i\omega k} d\omega, \qquad k = 0, \pm 1, \pm 2, \dots \\ \gamma_z(k) &= \int_{-\pi}^{\pi} S_{zz}(\omega) e^{i\omega k} d\omega, \qquad k = 0, \pm 1, \pm 2, \dots \end{aligned}$$

Setting $k = 0$ we get $\mathrm{E}[x_t z_t] = \int_{-\pi}^{\pi} S_{xz}(\omega) d\omega$, which shows that the cross-spectral density decomposes the covariance of two series into the covariance of components at each ω frequency.

To finish, we display some typical cross spectral quantities that are frequently used, besides the phase spectrum, to judge the relationship between two time series, namely,

- The Coherency Spectrum

$$W_{xz}(\omega) = \frac{|S_{xz}(\omega)|}{\sqrt{S_{xx}(\omega) + S_{zz}(\omega)}} = \frac{\alpha_{xz}(w)}{\sqrt{S_{xx}(\omega) + S_{zz}(\omega)}}, \qquad -1 \leq W_{xz}(\omega) \leq 1$$

  It is evident that the coherency spectrum measures the correlation between the components of $\{x_t\}$ and the components of $\{z_t\}$ at frequency ω. It is the correlation coefficient between $\{x_t\}$ and $\{z_t\}$ in the frequency domain.

- The Gain

$$G_{xz}(\omega) = \frac{|S_{xz}(\omega)|}{S_{xx}(\omega)}$$

  This measures the extent to which the spectrum of $\{x_t\}$ has been modified to approximate the corresponding component of $\{z_t\}$. In other words, it is the regression coefficient from the least squares regression of $\{x_t\}$ on $\{z_t\}$ at frequency ω.

Estimation of the spectrum

All objects defined so far entail the knowledge of population moments or equivalently of infinite information. Therefore, their application to real data implies the definition of estimators. This section aims to present some of the most commonly used spectrum approximations.

Given a sample of T observations, the natural estimator of the population spectrum is the sample counterpart:

$$\begin{aligned}\widehat{S}(\omega) &= \frac{1}{2\pi} \sum_{k=-(T-1)}^{T-1} \hat{\gamma}(k) e^{-i\omega k} = \\ &= \frac{1}{2\pi} \left[\hat{\gamma}(0) + 2 \sum_{k=1}^{T-1} \cos(\omega k) \right] \qquad -\pi \leq \omega \leq \pi\end{aligned}$$

where $\hat{\gamma}(k)$ denotes the non-parametric estimator of the autocovariance function at lag k given by

$$\hat{\gamma}(k) = \begin{cases} \frac{1}{T}\sum_{t=k+1}^{T}\left(y_t - \overline{y}\right)\left(y_{t-k} - \overline{y}\right) & \text{for } k = 0, 1, \ldots, T-1 \\ \hat{\gamma}(-k) & \text{for } k = -1, , -2 \ldots, -(T-1) \\ 0 & \text{for } |k| > T-1 \end{cases}$$

where $\overline{y}$ is the sample mean[31].

Another spectrum estimator arises from the definition of finite Fourier transform and is referred to as periodogram:

$$\widehat{S}(\omega) = I(\omega) = \frac{1}{2\pi T}\left|\sum_{k=0}^{T} e^{-i\omega k} y_k\right|^2 \qquad -\pi \leq \omega \leq \pi$$

It is proven that both estimators produce the same estimates, which means that the periodogram is indistinguishable from the FT of the sample autocovariance.

As shown in Priestley (1981), the sample spectrum is an asymptotically unbiased estimator of $S(\omega)$ but its variance does not tend to zero as the sample size grows. This means that $I(\omega)$ is not a consistent estimator for $S(\omega)$. To construct an estimator of the spectrum with better properties we should first understand the source of the undesirable behavior. To this effect, Priestley (1981) points out that the periodogram *involves T sample autocovariances, and although the variance of each is O(1/T), the cumulative effect of the T terms produces a variance which is O(1)*, so *loosely speaking, it* (the sample spectrum) *contains "too many" sample autocovariances.* While Hamilton (1994) and Cochrane (2005) stress that we are using a sample of T observations to estimate T parameters and do not take into account this fact in the estimator. Whichever the point of view it seems that the problem draws from considering the poorly estimated high order autocovariances. Then, focusing on this idea a better estimator seems to be the truncated periodogram (or sample spectrum)

$$\widehat{S}_0(\omega) = \frac{1}{2\pi}\sum_{k=-M(T)}^{M(T)} \hat{\gamma}(k) e^{-i\omega k} \qquad -\pi \leq \omega \leq \pi$$

where $\hat{\gamma}(k)$ is the sample autocovariance at lag k and $M(T) < (T-1)$ the truncation point.

[31] Despite of being a biased estimator it produces positive definite autocovariance sequences. As a result, it guarantees that the spectrum estimates will be positive as required. This spectrum estimator is just the FT of the sample autocovariance.

By disregarding the sample autocovariances of order higher than $M(T)$ we get two conflicting effects; we reduce the variance of the estimates but on the other hand introduce some bias. However, the theoretical autocovariance function $\gamma(k)$ tends to zero as $|k| \to \infty$ and this ensures that omitting only the terms on the tail of the sample autocovariance will not seriously affect the estimator bias. As a result, the consistency of the above estimator depends on the properties of the truncation point. The truncated periodogram is a consistent estimator of the spectrum if $M(T)$ grows at a slower rate than the sample size grows so that $\frac{M(T)}{T} \to 0$ as $T \to \infty$.

The above discussion around the truncated sample spectrum was intended to introduce the broader class of spectrum estimators suggested by Grenander and Rosenblatt (1953)

$$\begin{aligned}\widehat{S}_*(\omega) &= \frac{1}{2\pi} \sum_{k=-M(T)}^{M(T)} \lambda(k,T)\hat{\gamma}(k)e^{-i\omega k} = \\ &= \frac{1}{2\pi}\left[\hat{\gamma}(0) + \sum_{k=1}^{M(T)} \lambda(k,T)\hat{\gamma}(k)\cos(\omega k)\right] \qquad -\pi \leq \omega \leq \pi\end{aligned}$$

where $M(T)$ and $\hat{\gamma}(k)$ are again the truncation point and the autocovariance estimates, respectively, and $\lambda(k,T)$ is referred to as lag window or kernel. In terms of the periodogram the estimator is

$$\widehat{S}_*(\omega) = \int_{-\pi}^{\pi} I(\theta)W(\omega-\theta)d\theta \qquad -\pi \leq \omega \leq \pi$$

where $W(\theta) = \frac{1}{2\pi}\sum_{s=-M(T)}^{M(T)} \lambda(s)e^{-i\theta s}$ is the spectral window, in other words the FT of the lag window. In the first expression the estimator is written as a weighted sum of the sample autocovariances whereas in the second it is written as a weighted integral of the periodogram. The lag window is a decreasing weight sequence that works to reduce the contribution of the high order sample autocovariances that affect the variance of the estimator. In the weighted covariance estimator the poorly estimated sample autocovariances are simultaneously eliminated through the truncation point and reduction of the contribution of the high order sample autocovariance considered by assuming a decreasing lag window. The spectral window works as a weight that smooths the periodogram in the neighborhood of a fixed ω frequency in order to attenuate the erratic behavior of the periodogram estimates and attempt to control the estimates of the variance. Ultimately, the weighted covariance (or smooth periodogram) estimator is consistent for $S(\omega)$ if we choose a suitable truncation point and an appropriate lag window (or spectral window).

The truncation point determines the point at which the sample autocovariance function is ignored. Such point also aims to balance the trade off between the bias, induced by disregarding the tail estimates of the autocovariance function, and the variance, diverging if too many poorly estimates are considered, to guarantee the consistency of the estimator. Hence, the important issue is that any proposed method to obtain a value for $M(T)$ must satisfy the convergence rate discussed above. Accordingly, the functions $M(T) = \sqrt{T}$ and $M(T) = T^{\alpha}$ with $0 < \alpha < 1$ are two possible choices. The problem is that they completely ignore the properties of the true underlying process and as Priestley (1981) pointed out is advisable to obtain the truncation point according to the sample characteristics. Therefore, we call attention for two more sophisticated methods surveyed in Priestley (1981). The first follows the well known property that the theoretical autocovariance function decays towards zero to suggest using the sample counterpart to determine the truncation point. The process is simple and consists of choosing $M(T)$ so that $\hat{\gamma}(k) \simeq 0, |k| > M(T)$. Naturally the sample autocovariance plot will not give us a unique value for the truncation point but it gives us a set of values for $M(T)$ which attempt to match the properties of the process. At the end which value to choose from this set should be a matter of evaluating the different estimates. In sum, this method only implies the computation of the sample autocovariance function, however because the latter will decay more slowly to zero than the theoretical function the selected $M(T)$ will be in general higher than the true $M(T)$. The second criterion is called *window closing*. We start with a small value of $M(T)$ and construct a set of possible truncation points by increasing $M(T)$. Then, for each value in the set we compute a spectrum estimate, obtaining a sequence of estimates. The first estimate will be the smoother one because it employs just a few sample estimates and as we increase the number of sample autocovariances used in the estimator the estimates will become more erratic. Finally, we determine the truncation point by detecting the point at which the smoothing has been relaxed *too far*. One of the criticisms made to this technique is that it relies on the researcher's judgement of *too smooth*.

The lag window weights the sample autocovariance function. The consistency of the estimator is established if this weight sequence decays at a proper rate. There are several proposals in the literature for the functional form of the lag window (or spectral window) and most of them are of the scale parameter form. This means that the lag window has a parameter that simply has the role of stretching or contracting the weights. Curiously, that parameter is the truncation point.

Table 11 presents in detail some of the most commonly used kernels in applied work. For the truncated periodogram there is no down weighting scheme because the weight sequence weights equally all the sample autocovariances up to lag $M(T)$. The problem with this lag window choice is that the spectrum estimates can be negative for some frequencies which is undesirable given the non-negative nature of the population spectrum. In contrast, all the other lag windows are decreasing weight functions and produce non-negative spectrum estimates for all frequencies. The General Tukey window is the general form of the original lag window suggested by Tukey (1949) where a was set to 0.23. In the general case a only has to satisfy $0 < a \leq \frac{1}{4}$ so that $\lambda(u) \geq 0$ for all u. Regarding the spectral windows, note that the Bartlett's window is the Dirichlet kernel of order M(T) and the General Tukey's window is just a linear combination of the former kernel. Parzen's spectral window has a difficult form and thus in applied work the approximation given in table 11 is typically used.

Table 11: Lag windows and corresponding spectral windows.

	Lag Window	Spectral Window
Truncated Periodogram	$\lambda(u)=\begin{cases} 1 & \text{if } \lvert u\rvert \leq M(T) \\ 0 & \text{if } \lvert u\rvert > M(T) \end{cases}$	$W^{\mathrm{TP}}(\theta)=\frac{1}{2\pi}\left\{\frac{sin\left[\left(M(T)+\frac{1}{2}\right)\theta\right]}{sen\left(\frac{\theta}{2}\right)}\right\}$
Bartlett	$\lambda(u)=\begin{cases} 1-\frac{\lvert u\rvert}{M(T)} & \text{if } \lvert u\rvert \leq M(T) \\ 0 & \text{if } \lvert u\rvert > M(T) \end{cases}$	$W^{\mathrm{B}}(\theta)=\frac{1}{2\pi M(T)}\left\{\frac{sin\left(\frac{M(T)\theta}{2}\right)}{sen\left(\frac{\theta}{2}\right)}\right\}^2$
General Tukey	$\lambda(u)=\begin{cases} 1-2a+2a\cos\left(\frac{\pi u}{M(T)}\right) & \text{if } \lvert u\rvert \leq M(T) \\ 0 & \text{if } \lvert u\rvert > M(T) \end{cases}$	$W^{\mathrm{GT}}(\theta)=aW^{\mathrm{B}}(\theta-\frac{\pi}{M(T)})+(1-2a)W^{\mathrm{B}}(\theta)+ +aW^{\mathrm{B}}(\theta+\frac{\pi}{M(T)})$
Parzen	$\lambda(u)=\begin{cases} 1-6\left(\frac{u}{M(T)}\right)^2+6\left(\frac{\lvert u\rvert}{M(T)}\right)^3 & \text{if } \lvert u\rvert \leq \frac{M(T)}{2} \\ 2\left(1-\frac{\lvert u\rvert}{M(T)}\right)^3 & \text{if } \frac{M(T)}{2} \leq \lvert u\rvert \leq M(T) \\ 0 & \text{if } \lvert u\rvert > M(T) \end{cases}$	$W^{\mathrm{GT}}(\theta)\simeq\frac{3}{8\pi M(T)^3}\left[\frac{\sin\left(\frac{M(T)\theta}{4}\right)}{\frac{1}{2}\sin\left(\frac{\theta}{2}\right)}\right]^4$

Overall, the main idea to retain is that with an appropriate choice of the truncation point and the lag window we obtain a consistent spectrum estimator valid for all types of stationary processes.

As an alternative to the non-parametric approach we can construct a spectrum estimator assuming a specific parametric model for the observed series. As suggested by Parzen (1974) and Akaike (1974c, 1976) first decide on the ARMA(p,q) specification (see equation 20) that best fits the data determining the orders p and q of the AR and MA polynomials, respectively, by the AIC or BIC criterion[32]. Then estimate the unknown model parameters $(\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q, \sigma_\epsilon^2)$

[32]The acronyms AIC and BIC denote Akaike's (1973) information criterion and Schwarz'(1978) bayesian information criterion, respectively.

using some proper method, like maximum likelihood. And, finally, replace such estimates in the theoretical spectral density function of an ARMA(p,q) model given in equation (21).

The sampling properties for this class of estimators are discussed in Hannah (1970), Parzen (1974) and Berk (1974) for the AR model and in An, Cgen and Hannah (1982), Hannan and Kavalieris (1983) and Den Haan and Levin (1998) for the ARMA model. Nonetheless, the idea is that if the model is correctly specified then the parametric spectrum estimator is consistent. Or in other words, it is sufficient that the autocovariance function of the estimated ARMA(p,q) process resembles to a great extent the autocovariance function of the true process to guarantee the consistency of the estimator.

Both types of estimators are widely used in applied work and there is no empirical evidence leading researchers to prefer one method over the other. Ultimately, the performance of each estimator depends on the true data generating process (unknown in practice) as shown in the simulation results provided by Beasmish (1977) or Beasmish and Priestley (1981).

Appendix B: Linear Filters

A linear filter in the time domain representation is defined as

$$x_t = \sum_{j=-\infty}^{\infty} b_j y_{t-j} = \sum_{j=-\infty}^{\infty} b_j L^j y_t = B(L) y_t$$

where $L^j y_t = y_{t-j}$ is the backshift operator, $\{b_j, j = 0, \pm 1, \pm 2, \ldots\}$ are the weights of the two-sided filter satisfying $\sum_{j=-\infty}^{\infty} |b_j| < \infty$ and $\{x_t\}$ is the filtered series. If $b_j = 0$ for $j < 0$ then the filter is said to be one-sided and if $b_j = b_{-j}$, $\forall j$ the filter is said to be symmetric.

To obtain the spectrum of the filtered series we first have to derive the autocovariance function. Assuming that $\{x_t\}$ is a zero mean process we have

$$\begin{aligned}
\gamma_x(k) &= E\left(x_t x_{t-k}\right) = \\
&= E\left[\left(\sum_{j=-\infty}^{\infty} b_j y_{t-j}\right)\left(\sum_{l=-\infty}^{\infty} b_l y_{t-k-l}\right)\right] = \\
&= \sum_{j=-\infty}^{\infty}\sum_{l=-\infty}^{\infty} b_j b_l E\left[y_{t-j} y_{t-k-l}\right] = \\
&= \sum_{j=-\infty}^{\infty}\sum_{l=-\infty}^{\infty} b_j b_l \gamma_y(k+l-j), \qquad k = 0, \pm 1, \pm 2, \ldots
\end{aligned}$$

Then the spectrum of $\{x_t\}$ is

$$\begin{aligned}
S_x(\omega) &= \frac{1}{2\pi}\sum_{k=-\infty}^{\infty} \gamma_x(k) e^{-i\omega k} = \\
&= \frac{1}{2\pi}\sum_{k=-\infty}^{\infty}\sum_{j=-\infty}^{\infty}\sum_{l=-\infty}^{\infty} b_j b_l \gamma_y(k+l-j) e^{-i\omega k} = \\
&= \frac{1}{2\pi}\sum_{h=-\infty}^{\infty}\sum_{j=-\infty}^{\infty}\sum_{l=-\infty}^{\infty} b_j b_l \gamma_y(h) e^{-i\omega(h-l+j)} = \\
&= \sum_{j=-\infty}^{\infty} b_j e^{-i\omega j} \sum_{l=-\infty}^{\infty} b_l e^{i\omega l} \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma_y(h) e^{-i\omega h} = \\
&= B(e^{-i\omega}) B(e^{i\omega}) S_y(\omega) = \\
&= \left|B(e^{-i\omega})\right|^2 S_y(\omega), \qquad -\pi \leq \omega \leq \pi
\end{aligned}$$

where $B(e^{-i\omega}) = \sum_{j=-\infty}^{\infty} b_j e^{-i\omega j}$ is called the frequency response function that can be understood as the FT of the filter weights, $\left|B(e^{-i\omega})\right|^2$ is named transfer function or squared gain and measures the relative importance of each frequency in the total variance since $\left|B(e^{-i\omega})\right|^2 = \frac{S_x(\omega)}{S_y(\omega)}$ and $S_y(\omega)$ denotes the spectral density function of the original time series. Finally, $\left|B(e^{-i\omega})\right|$ is known as gain of the filter because if $B(e^{-i\omega})$ can be a complex number then it has a polar form representation

$$B(e^{-i\omega}) = \left|B(e^{-i\omega})\right| e^{-i\phi(\omega)}$$

where $\left|B(e^{-i\omega})\right|$ is the gain and measures the difference in amplitude between the filtered series $\{x_t\}$ and the original series $\{y_t\}$ and $\phi(\omega)$ is the phase and measures the time displacement between the two time series.

In the spectrum of $\{x_t\}$ the transfer function works as a weighting function of the spectrum of $\{y_t\}$. Hence, it is easily seen that the filtered series will mainly include the frequency components of $\{y_t\}$ that are most weighted by the transfer function. If $\left|B(e^{-i\omega})\right| = 1$ for a particular frequency ω then the component of $\{y_t\}$ at the ω frequency is preserved without any distortion, whereas $\left|B(e^{-i\omega})\right| = 0$ for a particular frequency ω implies that the ω frequency component of $\{y_t\}$ is completely eliminated. Further, if $\left|B(e^{-i\omega})\right| < 1$ (or > 1) for a particular frequency ω then the ω frequency component of $\{y_t\}$ is retained but with some distortion.

The linear filters are classified as low pass, high pass or band-pass filters according to the type of frequencies that they preserve. For instance the difference operation is a linear filter frequently used to render a time series stationary. This means that the transfer function will assume higher values at medium and high frequencies than at low frequencies. So, that $\{x_t\}$ encompasses only the medium and short term components of $\{y_t\}$. For this reason the first difference filter is classified as a high pass filter. Explicitly, taking first differences of a time series consists in

$$x_t = y_t - y_{t-1} = (1 - L)y_t = F(L)y_t$$

Since $f_0 = 1$, $f_1 = -1$ and $f_j = 0$ for $|j| \geq 2$ the frequency response function is simply

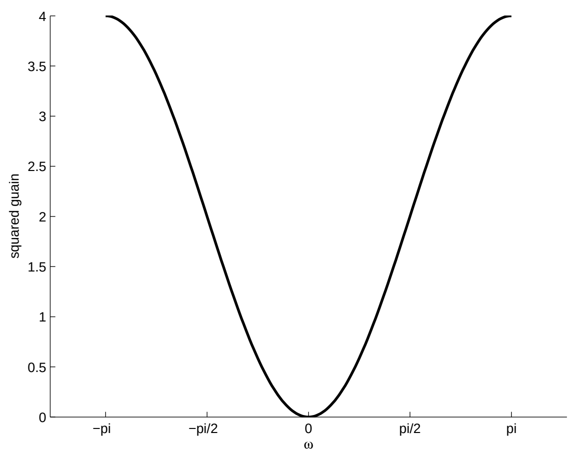
$$F(e^{-i\omega}) = 1 - e^{-i\omega}, \qquad \text{-}\pi \leq \omega \leq \pi$$

and the squared gain

$$\begin{aligned}\left|F(e^{-i\omega})\right|^2 &= F(e^{-i\omega})F(e^{i\omega}) = \\ &= \left(1-e^{-i\omega}\right)\left(1-e^{i\omega}\right) = \\ &= 2\left[1-\cos\left(\omega\right)\right], \qquad \text{-}\pi \leq \omega \leq \pi\end{aligned}$$

Figure 18 shows the squared gain of the first difference filter in the interval $[-\pi, \pi]$. At zero frequency the squared gain is zero and in its neighborhood is very low, which means that the low frequency components of $\{y_t\}$ will be eliminated or at least attenuated. Whereas the high and intermediate frequency components are preserve, although with distortion because $\left|B(e^{-i\omega})\right|^2 > 1$ for some frequencies.

Figure 18: Squared gain of the first difference filter.



A moving average filter is expressed as

$$x_t = \sum_{j=0}^{m-1} \frac{1}{m} y_{t-j} = \sum_{j=0}^{m-1} \frac{1}{m} L^j y_t = G(L) y_t, \qquad m \geq 2$$

where m stands for the moving average order. This filter is a very simple filter because its weights are constant for every j. The frequency response function of the filter is just

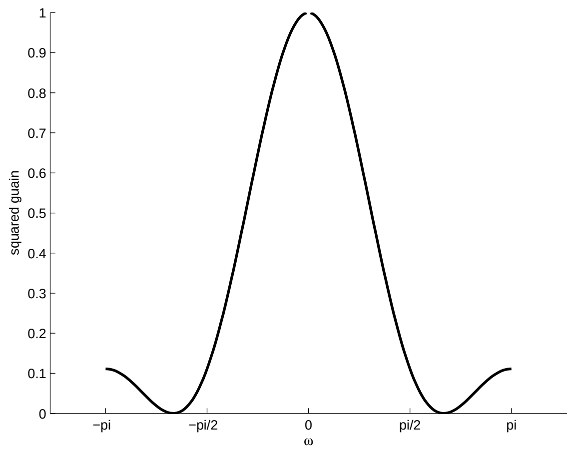
$$G(e^{-i\omega}) = \sum_{j=0}^{m-1} \frac{1}{m} e^{-i\omega j}, \qquad -\pi \leq \omega \leq \pi$$

and so the squared gain is

$$\begin{aligned} \left|G(e^{-i\omega})\right|^2 &= G(e^{-i\omega})G(e^{i\omega}) = \\ &= \frac{1}{m^2}\frac{1-\cos(mw)}{1-\cos(\omega)}. \end{aligned}$$

Figure 19 illustrates the path of the squared absolute value of the frequency response function from -π to π of a moving average filter of order 3. Clearly this filter passes low frequency components with a gain near one and attenuates high frequencies. So, $\{x_t\}$ will mainly contain the low frequency components of the original series and therefore this filter works as an example of a low pass filter. In a time domain perspective we say that $\{x_t\}$ will essentially contain the trend component because low frequencies are related with long periodicities.

Figure 19: Squared gain of the moving average filter of order 3.



Both examples exhibit an important feature of linear filters when used in practice, which is the fact that they usually do not leave the properties of the extracted components intact. In most cases the frequency components are preserved with some alteration but the researchers believe that those distortions are minor. Moreover, filtering techniques are several times employed with the objective to extract a particular component such as growth or cyclical fluctuations but as one can see there is always some weight given to frequencies that are not of interest. This phenomena is call leakage. The filters that exactly extract some desired range of frequencies are called ideal filters. Here we will survey three of the ideal linear filters since they will be mention in the main text.

(a) *Ideal low pass filter*

In this case we are looking for a filter that exactly retains the low frequency components of the $\{y_t\}$ series and that completely blocks the high and intermediate frequencies. For that purpose we must establish the following transfer function

$$\left|B^{LP}(e^{-i\omega})\right|^2 = \begin{cases} 1 & , \ |\omega| \leq \omega_l \\ 0 & , \ |\omega| > \omega_l \end{cases}$$

where ω_l sets the upper bound to what we consider low frequencies. The name ideal low pass filter obviously follows from the fact that it isolates exactly a range of low frequencies. The time domain representation of the ideal filter weights is obtained by working the IFT of $B^{LP}(e^{-i\omega})$. For $j \neq 0$

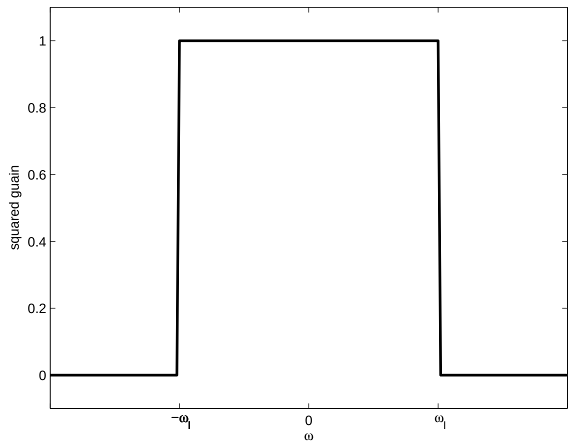
$$\begin{aligned} b_j^{LP} &= \frac{1}{2\pi}\int_{-\pi}^{\pi} B^{LP}(e^{-i\omega})e^{i\omega j}d\omega = \\ &= \frac{1}{2\pi}\int_{-\omega_l}^{\omega_l} e^{i\omega j}d\omega = \\ &= \frac{1}{2\pi}\left[\frac{e^{i\omega j}}{ij}\right]_{-\omega_l}^{\omega_l} = \\ &= \frac{sen(\omega_l j)}{\pi j}, \qquad |j| \geq 1 \end{aligned}$$

and for $j = 0$

$$\begin{aligned} b_0^{LP} &= \frac{1}{2\pi}\int_{-\pi}^{\pi} B^{LP}(e^{-i\omega})d\omega = \\ &= \frac{1}{2\pi}\int_{-\omega_l}^{\omega_l} 1 d\omega = \\ &= \frac{1}{2\pi}\left[\omega\right]_{-\omega_l}^{\omega_l} = \\ &= \frac{\omega_l}{\pi} \end{aligned}$$

From figure 20 we observe that this filter exactly preserves frequency components within $[-\omega_l, \omega_l]$ and that it completely eliminates the frequencies outside this interval.

Figure 20: Squared gain of the ideal low pass filter.



(b) *Ideal high pass filter*

In this case we want a filter that simultaneously preserves the high frequencies while eliminaties the low frequencies. Hence, we define the following

$$\left|B^{HP}(e^{-i\omega})\right|^2 = \begin{cases} 1 & , \ |\omega| \geq \omega_h \\ 0 & , \ |\omega| < \omega_h \end{cases}$$

The ideal weights, in time domain representation, for the high pass filter are easily obtain given that this filter can be constructed from a low pass filter considering $\omega_h = \omega_l$ and

$$B^{HP}(L) = 1 - B^{LP}(L)$$

where $B^{HP}(L)$ denotes the ideal high pass filter and $B^{LP}(L)$ the ideal low pass filter in time domain. Then,

$$b_j^{HP} = \begin{cases} 1 - b_0^{LP} & , \ j = 0 \\ -b_j^{LP} & , \ |j| \geq 1 \end{cases}$$

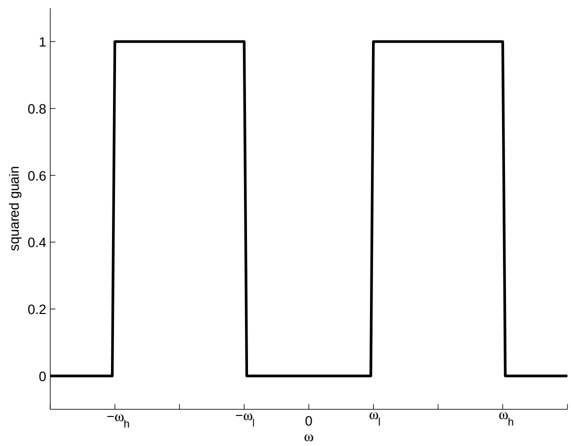
(c) *Ideal band-pass filter*

This last case is in the middle of the two previous filters. The objective is to eliminate all frequencies outside the interval $[-\omega_h, -\omega_l] \cup [\omega_l, \omega_h]$ and leave intact all frequencies within the interval. Hence, we have

$$\left|B^{BP}(e^{-i\omega})\right|^2 = \begin{cases} 1 & , \ |\omega| \in [\omega_l, \omega_h] \\ 0 & , \ |\omega| \notin [\omega_l, \omega_h] \end{cases}$$

Figure **??** reveals that we definitely remove components at higher frequencies than $|w_h|$ and at lower than $|w_l|$.

Figure 21: Squared gain of the ideal band-pass filter.



The filter weights in the time domain are given by

$$\begin{aligned} b_j^{BP} &= \frac{1}{2\pi}\int_{-\pi}^{\pi} B^{BP}(e^{-i\omega})e^{i\omega j}d\omega = \\ &= \frac{1}{2\pi}\int_{-\omega_h}^{-\omega_l} e^{i\omega j}d\omega + \int_{\omega_l}^{\omega_h} e^{i\omega j}d\omega = \\ &= \frac{1}{2\pi}\left[\frac{e^{i\omega j}}{ij}\right]_{-\omega_h}^{-\omega_l} + \left[\frac{e^{i\omega j}}{ij}\right]_{\omega_l}^{\omega_h} = \\ &= \frac{1}{2\pi}\frac{1}{ij}\left(e^{-i\omega_l j} - e^{-i\omega_h j} + e^{i\omega_h j} - e^{i\omega_l j}\right) = \\ &= \frac{sen(\omega_h j)}{\pi j} - \frac{sen(\omega_l j)}{\pi j}, \qquad |j| \geq 1 \end{aligned}$$

and for $j = 0$

$$\begin{aligned} b_0^{BP} &= \frac{1}{2\pi}\int_{-\pi}^{\pi} B^{BP}(e^{-i\omega})d\omega = \\ &= \frac{1}{2\pi}\int_{-\omega_h}^{-\omega_l} 1d\omega + \int_{\omega_l}^{\omega_h} 1d\omega = \\ &= \frac{\omega_h - \omega_l}{\pi}. \end{aligned}$$

Alternatively, the band-pass filter can be constructed from the subtraction of two ideal low pass filters, and the ideal weights are obtained as,

$$b_j^{BP} = b_j^{LP(\omega_h)} - b_j^{LP(\omega_l)}$$

where $b_j^{LP(\omega_h)}$ denotes the ideal weights of a low pass filter with cutoff at ω_h and $b_j^{LP(\omega_l)}$ the ideal weights of a low pass filter with cutoff at ω_l.

Finally, the IFT of the spectrum of $\{x_t\}$ allows recovering the autocovariance function of the filtered series

$$\gamma_x(k) = \int_{-\pi}^{\pi} S_x(\omega)e^{-i\omega k}dw, \qquad k = 0, \pm 1, \pm 2, \ldots$$

References

[1] Altissimo, F., Bassanetti, A., Cristadoro, R., Forni, M., Lippi, M., Reichlin, L. and Veronese, G. (2001), "A Real Time Coincident Indicator of the Euro Area Business Cycle," CEPR working paper No. 3108.

[2] Altissimo, F., Forni, M., Cristadoro, R., Lippi, M. and Veronese, G. (2008), "New Eurocoin: Tracking Economic Growth in Real Time," *Center for Economic Research (RECent)* 020, University of Modena and Reggio E., Department of Economics.

[3] Amengual, D. and Watson, M. (2007), "Consistent Estimation of the Number of Dynamic Factors in a Large N and T Panel," *Journal of Business and Economic Statistics*, American Statistical Association, 25, 91-96.

[4] Angelini, E., Camba-Mendez, G., Giannone, D., Reichlin, L. and Rünstler, G. (2008), "Short-term Forecasts of Euro Area GDP Growth," CEPR Discussion Papers 6746, C.E.P.R. Discussion Papers.

[5] Backus, D.K. and Kehoe, P.J. (1992), "International Evidence on the Historical Properties of Business Cycles," *American Economic Review*, 82(4), 864-888.

[6] Bai, J. (2003), "Inferential Theory for Factor Models of Large Dimensions," *Econometrica*, 71(1), 135-171.

[7] Bai, J., and Ng, S. (2002), "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, 70(1), 191-221.

[8] Bai, J., and Ng, S. (2007), "Determining the Number of Primitive Shocks in Factor Models," *Journal of Business and Economic Statistics*, 25(1), 52-60.

[9] Baxter, M. (1994), "Real Exchange Rates and Real Interest Differentials: Have We Missed the Business-Cycle Relationship?," *Journal of Monetary Economics*, 33(1), 5-37.

[10] Baxter, M. and King, R. (1999), "Measuring Business Cycles: Approximate Band-Pass Filters For Economic Time Series," *Review of Economics and Statistics*, 81(4), 575-93.

[11] Beveridge, S. and Nelson, C. R. (1981), "A New Approach to Decomposition of Economic Time Series Into Permanent and Transitory Components with Particular Attention to Measurement of the 'Business Cycle'," *Journal of Monetary Economics*, 7(2), 151-174.

[12] Blanchard, O. J. and Quah, D. (1989), "The Dynamic Effects of Aggregate Demand and Supply Disturbances," *American Economic Review*, 79(4), 655-673.

[13] Brockwell, P. J. and Davis, R. (1991), *Time Series: Theory and methods* (2nd ed.), Springer.

[14] Burns, A. F. and Mitchell, W. C. (1946), *Measuring Business Cycle*, NBER Book Series Studies in Business Cycles.

[15] Canova, F. (1998), "Detrending and Business Cycle Facts: A User's Guide," *Journal of Monetary Economics*, 41(3), 533-540.

[16] Cecchetti, S.G. and Kashyap, A.K. (1995), "International Cycles", NBER Working Papers 5310, *National Bureau of Economic Research*.

[17] Chamberlain, G. (1983), "Funds, Factors, and Diversification in Arbitrage Pricing Models," *Econometrica*, 51(5), 1305-1323.

[18] Chamberlain, G. and Rothschild, M. (1983), "Arbitrage, Factor Structure and Mean-Variance Analysis in Large Asset Markets," *Econometrica*, 51(5), 1281-1304.

[19] Christiano, L. and Fitzgerald, T. (1998), "The Business Cycle: It's Still a Puzzle," *Economic Perspectives*, Federal Reserve Bank of Chicago, issue Q IV, 56-83.

[20] Christiano, L. and Fitzgerald, T. (2003), "The Band-Pass Filter", *International Economic Review*, 44(2), 435-465.

[21] Cochrane, J. H. (1994), "Univariate vs. Multivariate Forecasts of GNP Growth and Stock Returns: Evidence and Implications for the Persistence of Shocks, Detrending Methods," NBER Working Papers 3427, National Bureau of Economic Research.

[22] Cochrane, J. H. (1997), "Time series for Macroeconomics and Finance," Lecture Notes, University of Chicago.

[23] Cogley, T. and Nason, J.M. (1995), "Effects of the Hodrick-Prescott Filter on Trend and Difference Stationary Time Series Implications for Business Cycle Research," *Journal of Economic Dynamics and Control*, 19(1-2), 253-278.

[24] Cooley, T. F. and Ohanian L.E. (1991), "The Cyclical Behavior of Prices," *Journal of Monetary Economics*, 28(1), 25-60.

[25] Correia, I.H., Neves, J.L., and Rebelo, S.T. (1992), "Business Cycles from 1850-1950 - New Facts about Old Data," *European Economic Review*, 36(2/3), 459-467.

[26] Croushore, D., and Stark, T. (2001), "A Real-Time Data Set for Macroeconomists," *Journal of Econometrics*, 105(1), 111-130.

[27] Croushore, D., and Stark, T. (2003), "A Real-Time Data Set for Macroeconomists: Does the Data Vintage Matter?," *Review of Economics and Statistics*, 85(3), 605-617.

[28] Den Haan, W.J. and Levin, A. (1996), "Inferences from Parametric and Non-Parametric Covariance Matrix Estimation Procedures," *NBER Tecnical Working Papers*, No. 0195, National Bureau of Economic Research.

[29] Den Haan, W.J. and Levin, A. (2000), "Robust Covariance Matrix Estimation with Data-Dependent VAR Prewhitening Order," *NBER Technical Working Papers*, No. 0255, National Bureau of Economic Research.

[30] Diebold, F. X. and Rudebusch, G. D. (1996), "Measuring Business Cycles: A Modern Perspective," *The Review of Economics and Statistics*, 78(1), 67-77.

[31] Estrella, A. (2007), "Extracting Business Cycle Fluctuations: What Do Time Series Filters Really Do?," *Federal Reserve Bank of New York*, Staff reports no. 289.

[32] Everts, M. (2006), "Band-Pass Filters," MPRA Paper 2049, University Library of Munich, Germany.

[33] Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2000), "The Generalized Dynamic Factor Model: Identification and Estimation," *The Review of Economics and Statistics*, 82(4), 540-554.

[34] Forni, M. and Lippi, M. (2001), "The Generalized Dynamic Factor Model: Representation Theory," *Econometric Theory*, 17(6), 1113-41.

[35] Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2004), "The Generalized Dynamic Factor Model: Consistency and Rates," *Journal of Econometrics*, 119(2), 231-255.

[36] Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2005), "The Generalized Dynamic Factor Model: One-Sided Estimation and Forecasting," *Journal of the American Statistical Association*, 100(471), 830-840.

[37] Geweke, J. (1977), "The Dynamic Factor Analysis of Economic Time Series," in Dennis J. Aigner and Arthur S. Goldberger (Eds.), *Latent Variables in Socio-Economic Models*, Amsterdam: North-Holland.

[38] Giannone, D., Reichlin, L. and Sala, L. (2005), "Monetary Policy in Real Time," in *NBER Macroeconomic Annual 2004*, eds. M. Gertler and K. Rogoff, Cambridge, MA: MIT Press, 161-200.

[39] Granger, C.W.J. (1966), "The typical Spectral Shape of an Economic Variable," *Econometrica*, 34(1), 150-161.

[40] Guay, A. and St-Amant, P. (1997), "Do the Hodrick-Prescott and Baxter-King Filters Provide a Good Approximation of Business Cycles?," Cahiers de recherche CREFE / CREFE Working Papers 53, CREFE, Université du Québec à Montréal.

[41] Guay, A. and St-Amant, P. (1996), "Do Mechanical Filters Provide a Good Approximation of Business Cycles?," Technical Reports 78, Bank of Canada.

[42] Hallin, M. and Liška, R. (2007), "Determining the Number of Factors in the General Dynamic Factor Model," *Journal of the American Statistical Association*, 102(478), 603-617.

[43] Hamilton, James D. (1994), *Time Series Analysis*, Princeton University Press.

[44] Harvey, A. C. and Trimbur, T. (2003), "Generalised Model-Based Filters For Extracting Trends and Cycles in Economic Time Series," *Review of Economics and Statistics*, 85(2), 244-255.

[45] Harvey, A.C. and Jaeger, A. (1993), "Detrending, Stylized Facts and the Business Cycle," *Journal of Applied Econometrics*, 8(3), 231-247.

[46] Higo, M. and Nakada, S.K. (1998), "How Can We Extract a Fundamental Trend from an Economic Time- Series?," Monetary and Economic Studies, Institute for Monetary and Economic Studies, *Bank of Japan*, 16(2), 61-111.

[47] Hodrick, R.J. and Prescott, E.C. (1997), "Postwar U.S. Business Cycles: an Empirical Investigation," *Journal of Money, Credit and Banking*, 29(1), 1-16.

[48] Kapetanios, G. (2004), "A New Method for Determining the Number of Factors in Factor Models with Large Datasets," Working Papers 525, Queen Mary, University of London, Department of Economics.

[49] King, R., Plosser, C. and Rebelo, S. (1988), "Production, Growth and Business Cycles: I. The Basic Neoclassical Model," *Journal of Monetary Economics*, 21(2-3), 195-232.

[50] King, R., Plosser, Stock, J. and Watson, M. (1991), "Stochastic Trend and Economic Fluctuations," *American Economic Review*, 81(4), 819-840.

[51] King, R. and Rebelo, S. (1993), "Low Frequency Filtering and Real Business Cycles," *Journal of Economic Dynamics and Control*, 17(1-2), 207-231.

[52] King, R. and Watson , M. (1994), "The Post-War U.S. Phillips Curve: a Revisionist Econometric History," Carnegie-Rochester Conference Series on Public Policy, 41(1), 157-219.

[53] King, R.G., Stock, J.H. and Watson, M.W., "Temporal Instability of the Unemployment-Inflation Relationship," Economic Perspectives of the Federal Reserve Bank of Chicago, May/June 1995, 19, 2-12.

[54] Nelson, C.R. and Plosser, C.I. (1982), "Trends and Random Walks in Macroeconomic Time series: Some Evidence and Implications," *Journal of Monetary Economics*, 10(2), 139-162.

[55] Onatski, A. (2005), "Determining the Number of Factors from Empirical Distribution of Eigenvalues," Discussion Papers 0405-19, Columbia University, Department of Economics.

[56] Orphanides, A. and van Norden, S. (2002), "The Unreliability of Output-Gap Estimates in Real Time," *Review of Economics and Statistics*, 84(4), 569-583.

[57] Osborn, D.R. (1995), "Moving Average Detrending and the Analysis of Business Cycles," Oxford Bulletin of Economics and Statistics, 57(4), 547-58.

[58] Pedersen, T. M. (2001), "The Hodrick-Prescott Filter, the Slutzky Effect, and the Distortionary Effect of Filters," *Journal of Economic Dynamics and Control*, 25(8), 1081-1101.

[59] Persons, W. M. (1919), "Indices of Business Conditions," *Review of Economics and Statistics*, 1, 5-107.

[60] Pesaran, M. and Timmermann , A. (1992), "A Simple Nonparametric Test of Predictive Performance," *Journal of Business and Economic Statistics*, 10(4), 461-465.

[61] Priestley, M. B. (1981). *Spectral Analysis and Time Series*, Academic Press, London

[62] Ravn, M. O. and Uhlig, H. (2001), "On Adjusting the HP-Filter for the Frequency of Observations," CEPR Discussion Papers 2858, C.E.P.R. Discussion Papers.

[63] Rünstler, G. (2004), "Modelling Phase Shifts Among Stochastic Cycles," *Econometrics Journal*, 7(1), 232-248.

[64] Rünstler, G., Barhoumi, K., Benk, S., Cristadoro, R., Den Reijer, A., Jakaitiene, A., Jelonek, P., Rua, A., Ruth, K. and Van Nieuwenhuyze, C. (2008), "Short-Term Forecasting of GDP Using Large Monthly Datasets - a Pseudo Real-Time Forecast Evaluation Exercise," Research series 200806-17, National Bank of Belgium.

[65] Sargent, T. J. and Sims, C. (1977), "Business Cycle Modelling Without Pretending to Have Too Much a Priori Economic Theory," in Cristopher A. Sims (ed.) New Methods in Business Research (Federal Reserve Bank of Minneapolis).

[66] Singleton, K. (1988), "Econometric Issues in the Analysis of Equilibrium Business Cycle Models," *Journal of Monetary Economics*, 21(2-3), 361-386.

[67] Stock, J. H. and Watson, M. W. (1998), "Diffusion Indexes," NBER Working Papers 6702, *National Bureau of Economic Research.*

[68] Stock, J. and Watson, M. W. (1999). Business Cycle Fluctuations in US Macroeconomic Time Series. In J. B. Taylor and M. Woodford (Eds.), *Handbook of Macroeconomics*, 3-64. Amsterdam:Elsevier Science Publishers.

[69] Stock, J.H. and Watson, M.W. (2002a), "Macroeconomic Forecasting Using Diffusion Indexes," *Journal of Business and Economic Statistics*, 20(2), 147-162.

[70] Stock, J.H. and Watson, M.W. (2002b), "Forecasting Using Principal Components from a Large Number of Predictors," *Journal of the American Statistical Association*, 97(460), 1167-79.

[71] Stock, J. H. and Watson, M. W. (2005), "Implications of Dynamic Factor Models for VAR Analysis," *NBER Working Papers*, 11467.

[72] Watson, M. W. (2007), "How Accurate are Real-Time Estimates of Output Trends and Gaps?," Economic Quarterly, *Federal Reserve of Richmond*, Spring:143-161.

[73] Valle e Azevedo, J. (2007), "Interpretation of the Effects of Filtering Integrated Time Series," *Banco de Portugal Working Paper 12/2007.*

[74] Valle e Azevedo, J. (2007), "A Multivariate Band-Pass filter," *Banco de Portugal Working Paper 17/2007.*

[75] Valle e Azevedo, J. and Pereira, A. (2008), "Approximating and Forecasting Macroeconomic Signals in Real-Time," *Banco de Portugal Working Paper 19/2008.*

[76] Valle e Azevedo, J., Koopman, S.J. and Rua, A. (2006), "Tracking the Business Cycle of the Euro Area: A Multivariate Model Based Band-Pass Filter," *Journal of Business and Economic Statistics*, 24(3), 278-290.

[77] van Norden, S. (2004), "Optimal Band-Pass Filtering and the Reliablity of Current Analysis," *EUROSTAT Working Papers and Studies.*

[78] Zarnowitz, V. and Ozyildirim, A. (2001), "Time Series Decomposition and Measurement of Business Cycles, Trends and Growth Cycles," Economics Program Working Papers 01-03, *The Conference Board*, Economics Program.