CESifo Working Papers
www.cesifo.org/wp

# Why We Should Use High Values for the Smoothing Parameter of the Hodrick-Prescott Filter

Gebhard Flaig

CESIFO WORKING PAPER NO. 3816
CATEGORY 12: EMPIRICAL AND THEORETICAL METHODS
MAY 2012

CESifo *Working Paper No. 3816*

# Why We Should Use High Values for the Smoothing Parameter of the Hodrick-Prescott Filter

## Abstract

The HP filter is the most popular filter for extracting the trend and cycle components from an observed time series. Many researchers consider the smoothing parameter $\lambda = 1600$ as something like an universal constant. It is well known that the HP filter is an optimal filter under some restrictive assumptions, especially that the "cycle" is white noise. In this paper we show that one gets a good approximation of the optimal Wiener-Kolmogorov filter for autocorrelated cycle components by using the HP filter with a much higher smoothing parameter than commonly used. In addition, a new method - based on the properties of the differences of the estimated trend - is proposed for the selection of the smoothing parameter.

JEL-Code: C220, C520.

Keywords: Hodrick-Prescott filter, Wiener-Kolmogorov filter, smoothing parameter, trends, cycles.

*Gebhard Flaig*
*Department of Economics*
*University of Munich*
*Schackstraße 4*
*80539 Munich*
*Germany*
*gebhard.flaig@lrz.uni-muenchen.de*

May 2012
The author thanks Andreas Blöchl for careful reading of the paper and useful suggestions.

# 1 Introduction

The probably most popular filter for extracting a trend and a cycle component from an observed time series is the Hodrick-Prescott filter (Hodrick/Prescott 1997). The features of this filter were intensively studied in the literature (see, among many others, King/Rebelo 1993; Harvey/Jaeger 1993; Cogley/Nason 1994; Kaiser/Maravall 2001). The properties of the HP filter in the time and frequency domain depend mainly on a smoothing parameter $\lambda$ which governs the smoothness of the estimated trend and the shape of the estimated cycle. In most empirical applications a value of 1600 is used for $\lambda$ (quarterly data). Hodrick/Prescott motivate this value by the assumption that the ratio of the variance of the cyclical component to the variance of the second differences of the trend (the inverse signal-to-noise ratio) of US GDP is about 1600.

It is well known that the HP filter with a smoothing parameter $\lambda$ is optimal (in the sense that the mean square error of the estimated components is minimal) if the second differences of the trend follow a white noise process (the trend is integrated of order 2) and if the cyclical components is white noise as well. These assumptions are clearly not appropriate in many applications. For instance, when we specify the trend as a random walk (with constant drift), the second differences follow an MA(1), not a white noise process. Most economists would argue that also the cycle is not white noise but follows an autocorrelated stationary process. In all these cases, the suggestion of Hodrick/Prescott has no sound justification and the HP filter is a pure ad-hoc procedure with possibly dubious features.

An alternative to ad-hoc filters like the HP filter is the specification and estimation of unobserved components models (Harvey 1989). The Kalman filter and smoother deliver in this case the optimal filter weights which are identical to those of the classical Wiener-Kolmogorov approach (Gomez 1999). A second possibility is to estimate an ARIMA model and apply the Beveridge-Nelson decomposition (Beveridge/Nelson 1981) or the "canonical" decomposition (Box et al. 1978). The disadvantage of those procedures - at least from the standpoint of an applied economist who has to analyze many time series for a real time business cycle analysis - is that they are complex and time consuming. There is a demand for simple and easy-to-use filters.

In this paper we stick to the HP filter. The procedure is simple to understand, it is easy to write a computer program and efficient procedures are very fast (on a modern PC you can filter several thousands of time series with 200 observations in less than a second). The aim of the paper is to suggest some simple rules for choosing a reasonable value for the smoothing parameter $\lambda$. Firstly, assuming a doubly infinite time series we derive for different specifications of the trend and the cycle components the optimal Wiener-Kolmogorov filter and search for that $\lambda$ for which the gain of the Wiener-Kolmogorov filter and the gain of the HP filter have both a value of 0.5. This is an approximation to the goal of minimizing the difference between the two gain functions (Harvey/Trimbur 2008). The results imply that in most realistic settings we should use much higher values for the smoothing parameter than the true inverse signal-to-noise ratio. For instance, assuming for the cycle an AR(1) model with a parameter of 0.7, the optimal $\lambda$ is about five times higher than the inverse signal-to-noise ratio: So, we should not use $\lambda = 1600$ but rather a value of somewhat higher than 8000! These theoretical results are corroborrated by a simulation study for time series with 160 observations. It is shown that by choosing a high $\lambda$ one can achieve a remarkable efficiency gain (compared with the Wiener-Kolmogorov filter).

The obtained results are useful but not really applicable in practice as we don't know the true signal-to-noise ratio. In the second part of the paper we derive a simple rule for the determination of a reasonable value for $\lambda$. The basic idea is that the first and/or second differences of the extracted trend should not exhibit a cyclical behaviour. We propose to use the HP filter with different values of $\lambda$ and select the minimum value for which the first and second differences of the generated trend show no cyclical behaviour. This choice can be based on visual inspection or on a more formal analysis in the time or frequency domain.

The paper is organized as follows: Section 2 outlines the model, the optimal Wiener-Kolmogorov filter and the Hodrick-Prescott filter. In section 3 we derive the optimal smoothing parameter for autocorrelated cycles. In section 4 we discuss a new suggestion for selecting $\lambda$. Section 5 reports the results for an empirical application and section 6 concludes.

# 2 Theoretical framework

## 2.1 The model

We specify a time series $\{y_t\}$ as the sum of a non-stationary trend component $\{\mu_t\}$ and a stationary cycle $\{c_t\}$:

$$y_t = \mu_t + c_t \tag{1}$$

The trend component is modeled as a n-fold integrated variable

$$(1 - L)^n \mu_t = \eta_t \tag{2}$$

where $n$ is a positive integer and $\eta_t$ is white noise with $E\eta_t = 0$ and $Var\,\eta_t = \sigma_\eta^2$.
The cycle is specified as a stationary AR(2) process (with AR(1) as a special case)

$$c_t = \varphi_1 c_{t-1} + \varphi_2 c_{t-2} + \epsilon_t \tag{3}$$

or

$$\Phi(L)\, c_t \;=\; \epsilon_t \qquad (3a)$$

where $\Phi(L) = 1 - \varphi_1 L - \varphi_2 L^2$ and $\epsilon_t$ is white noise with $E\,\epsilon_t = 0$ and $Var\,\epsilon_t = \sigma_\epsilon^2$. It is further assumed that $\eta_t$ and $\epsilon_t$ are uncorrelated at all leads and lags.

In the following we derive some properties of the components and the time series in the frequency domain. The specification of the trend and cycle components implies the folowing model for the observed time series

$$y_t \;=\; (1-L)^{-n}\eta_t + \Phi(L)^{-1}\epsilon_t$$

The stationary form is given by

$$(1-L)^n y_t \;=\; \eta_t + \frac{(1-L)^n}{\Phi(L)}\epsilon_t$$

The spectral density of the cycle is given by (Sargent, 1987, p. 262):

$$g_c(\omega) \;=\; \Big[1+\varphi_1^2+\varphi_2^2-2\varphi_1(1-\varphi_2)\cos\omega - 2\varphi_2\cos 2\omega\Big]^{-1}\sigma_\epsilon^2/2\pi$$

$$= \Big[1+\varphi_1^2+\varphi_2^2-2\varphi_1(1-\varphi_2)\cos\omega - 2\varphi_2\cos 2\omega\Big]^{-1}\frac{(1+\varphi_2)[(1-\varphi_2)^2-\varphi_1^2]}{1-\varphi_2}\frac{\sigma_c^2}{2\pi} \qquad (4)$$

where $\sigma_c^2$ is the variance of $\{c_t\}$. $\omega$ is the angular frequency, measured in radians.

The pseudo spectrum of $y_t$ is given by

$$g_y(\omega) \;=\; [2(1-\cos\omega)]^{-n}\,\sigma_\eta^2/2\pi + g_c(\omega)$$

$$= (\sigma_\eta^2/2\pi)\{[2(1-\cos\omega)]^{-n} + \lambda^*\tilde{g}_c(\omega)\} \qquad (5)$$

where $\lambda^* = \sigma_c^2/\sigma_\eta^2$ is the true inverse signal-to-noise ratio and $\tilde{g}_c(\omega)$ is defined as $\tilde{g}_c(\omega) = 2\pi g_c(\omega)/\sigma_c^2$. The pseudo-spectrum $g_y(\omega)$ is infinite at $\omega = 0$ and can be derived along the arguments presented in Harvey (1989, chapter 2.4) or Kaiser/Maravall (2001, chapt. 2.5). The key element concerning the trend part is that $\{\eta_t\}$ has a flat spectrum with value $\sigma_\eta^2/2\pi$ and the filter $(1-L)^{-n}$ has the power transfer function $[(1-e^{-\omega i})(1-e^{\omega i})]^{-n} = [2(1-\cos\omega)]^{-n}$, where $i = \sqrt{-1}$ is the imaginary unit.

With the same technique the (pseudo-) spectrum of $(1-L)^d y_t$ (where $d$ is a positive integer) can be derived as

$$g_{\Delta dy}(\omega) = \frac{\sigma_\eta^2}{2\pi}\left\{[2(1-\cos\omega)]^{d-n} + \lambda^*[2(1-\cos\omega)]^d \tilde{g}_c(\omega)\right\} \tag{6}$$

## 2.2 The optimal filter

We use as the optimal filter the Wiener-Kolmogorov filter. It minimizes the mean square error of the estimated component

$$\mathrm{MSE}_{\hat{\mu}} \;=\; E(\hat{\mu}-\mu)^2$$

It is easy to show that the optimal estimator is given by the conditional expectation

$$\hat{\mu} \;=\; E(\mu|y)$$

Assuming that all shocks are normally distributed we can express the filter formula as the linear function

$$\hat{\mu}_t \;=\; \sum_{j=-\infty}^{\infty} m_j y_{t-j}$$

where the weight $m_j$ is given by the coefficient of $L^j$ in the polynomial

$$M(L) = \left(\sigma_\eta^2/|(1-L)^n|^2\right) / \left(\sigma_\eta^2/|(1-L^n)|^2 + \sigma_\epsilon^2/|\Phi(L)|^2\right) \tag{7}$$

where we follow the convention to denote $A(L)A(-L)$ as $|A(L)|^2$. The formula is a simple application of the general framework developed by Whittle (1983) and Bell (1984) and described by Harvey (1989) and Kaiser/Maravall (2001). The numerator is the autocovariance generating function of $\{\mu_t\}$, the denominator the autocovariance generating function of $\{y_t\}$.

The power transfer function of the low-pass filter M(L) is obtained by replacing the lag operator $L$ by $e^{-i\omega}$ in $|M(L)|^2 = M(L)M(-L)$. The gain function $|M(\omega)| = \sqrt{|M(\omega)|^2}$ is then given by

$$|M(\omega)| = \sigma_\eta^2[2(1-\cos\omega)]^{-n} / \left\{\sigma_\eta^2[2(1-\cos\omega)]^{-n} + \tilde{g}_c(\omega)\sigma_c^2\right\} \tag{8}$$

where $\tilde{g}_c(\omega) = g_c(\omega)\cdot 2\pi/\sigma_c^2$ (already defined after equation (5)).

Using the defintion $\lambda^* = \sigma_c^2/\sigma_\eta^2$ we can write

$$|M(\omega)| \;=\; 1/\left[1+\lambda^*[2(1-\cos\omega)]^n \tilde{g}_c(\omega)\right] \tag{8a}$$

Applying the same procedure to the cyclical component, we get

$$\hat{c}_t = \sum_{j=-\infty}^{\infty} (1 - m_j) y_{t-j} = \sum_j h_j y_{t-j}$$

The gain function of the high-pass filter $H(L) = 1 - M(L)$ is given by

$$|H(\omega)| = 1 - |M(\omega)|$$

Often the properties of a filter are assessed by exploring its gain function. But, as Kaiser/Maravall (2001) notes, "this function only tells part of the story". It is much more useful to consider the spectrum of a generated component. The spectrum is derived as the product of the squared gain (the power transfer function) and the spectrum of the observed time series (see, e.g., Harvey, 1993).

We can derive the spectra of $\hat{\mu}$ and $\hat{c}$ as

$$g_{\hat{\mu}}(\omega) = |M(\omega)|^2 g_y(\omega) \tag{9}$$

and

$$g_{\hat{c}}(\omega) = |H(\omega)|^2 g_y(\omega) \tag{10}$$

The spectrum of $(1 - L)^d \hat{\mu}_t$ is given by

$$g_{\Delta d \hat{\mu}} = [2(1 - \cos\omega)]^d \, g_{\hat{\mu}}(\omega) \tag{11}$$

where $d$ is a positive integer.

## 2.3 The HP filter

We use the HP filter for extracting the trend and cycle from a time series. Suppose a doubly infinite series, the cycle is estimated by the high-pass filter (King/Rebelo 1993)

$$\tilde{c}_t = \tilde{H}(L) y_t$$

where

$$\tilde{H}(L) = \frac{\lambda(1-L)^2(1-L^{-1})^2}{1+\lambda(1-L)^2(1-L^{-1})^2} = \frac{\lambda L^{-2}(1-L)^4}{1+\lambda L^{-2}(1-L)^4}$$

$\lambda$ denotes not longer the true inverse signal-to-noise ratio, but is a prespecified smoothing parameter. If we replace $L$ by $e^{-i\omega}$ we get the frequency response function $\tilde{H}(\omega)$.

The spectrum of $\tilde{c}_t$ is then given by

$$g_{\tilde{c}}(\omega) = |\tilde{H}(\omega)|^2 g_y(\omega) \tag{12}$$

where the transfer function $|\tilde{H}(\omega)|^2$ is given by $|\tilde{H}(\omega)|^2 = \tilde{H}(\omega)\,\tilde{H}(-\omega)$ and $g_y(\omega)$ is the pseudo-spectrum of $\{y_t\}$.

$|\tilde{H}(\omega)|^2$ can be expressed as

$$|\tilde{H}(\omega)|^2 \quad = \quad \left(\frac{4\lambda(1-\cos\omega)^2}{1+4\lambda(1-\cos\omega)^2}\right)^2.$$

The trend is estimated by the low-pass filter

$$\tilde{\mu}_t = \tilde{M}(L)y_t = \left(1-\tilde{H}(L)\right)y_t = \left(1+\lambda(1-L)^2(1-L^{-1})^2\right)^{-1} y_t \tag{13}$$

The pseudo spectrum of $\tilde{\mu}_t$ is given by

$$g_{\tilde{\mu}}(\omega) = |\tilde{M}(\omega)|^2 g_y(\omega) \tag{14}$$

$|\tilde{M}(\omega)|^2$ can be expressed as

$$|\tilde{M}(\omega)|^2 = \left(1+4\lambda(1-\cos\omega)^2\right)^{-2} \tag{15}$$

We can also derive the spectrum of $(1-L)^d\tilde{\mu}_t$ (where $d$ is a positive integer) as

$$g_{\Delta d\tilde{\mu}}(\omega) \quad = \quad [2(1-\cos\omega)]^d\, g_{\tilde{\mu}}(\omega)$$

As already mentioned the choice of 1600 for the smoothing parameter $\lambda$ seems to be the "industry standard". The HP filter is the optimal filter if the trend follows an integrated random walk, the cycle is white noise (and not correlated with the trend shocks) and $\lambda$ is set to the inverse signal-to-noise ratio $\sigma_c^2/\sigma_n^2$(Kaiser/Maravall 2001). Even if the value of 1600 for $\lambda$ is optimal for US GDP it may be not optimal for GDP data of other countries or for other time series like investment (Harvey/Trimbur 2008). The possibly more important and interesting question is whether it is optimal to set $\lambda$ equal to the inverse signal-to-noise ratio in cases when the cycle is not white noise. We will deal with this problem in the next section.

# 3 The optimal value for the HP smoothing parameter

## 3.1 The general procedure

In this section we derive the optimal value for the HP smoothing parameter $\lambda$ in cases where the cyclical component is not white noise but rather follows a stationary autocorrelated process. We tackle this task in the following way: Firstly, we specify an AR process for the cycle $\{c_t\}$, derive the spectrum $g_c(\omega)$ and use equation (8a) for calculating the gain function $|M(\omega)|$ for the optimal Wiener-Kolmogorov filter. Secondly, from $|M(\omega)|$ we determine numerically the frequency $\omega_0$, where the gain has the value 0.5: $|M(\omega_0)| = 0.5$. Third, we use the relation (Gomez 2001) $\lambda_{HP} = [2\sin(\omega_0/2)]^{-4}$ for calculating that value of $\lambda$ for which the gain of the HP filter has a value of 0.5 at frequency $\omega_0$: At frequency $\omega_0$, the gain functions of the optimal Wiener-Kolmogorov and of the HP filter intersect. As Harvey/Trimbur (2008) note, this criterion is an approximation to minimizing the distance between the two gain functions.

## 3.2 Numerical calculations

In the following we use different specifications of the trend and cycle components for calculating the "optimal" $\lambda_{HP}^{opt}$ with the outlined procedure. For the trend component, we use alternatively a random walk (RW(1)) and an integrated random walk (RW(2)). For the cycle component we specify AR(1) and AR(2) models.

### 3.2.1 Trend RW(2) and cycle AR(1)

The model is given by

$$\begin{aligned}(1-L)^2\mu_t &= \eta_t \\ c_t &= \varphi_1 c_{t-1} + \epsilon_t\end{aligned}$$

$\eta_t$ and $\epsilon_t$ are white noise. The calculations are carried out for different values of the inverse signal-to-noise ratio. Table 1 shows the "optimal" values for the HP smoothing parameter.

Except cases with rather high values of the autoregressive parameter $\varphi_1$, the ratio $\lambda_{HP}^{opt}/(\sigma_c^2/\sigma_\eta^2)$ does not depend much on the true inverse signal-to-noise ratio $(\sigma_c^2/\sigma_\eta^2)$. However, the optimal $\lambda_{HP}^{opt}$ increases strongly with the autoregressive parameter $\varphi_1$. For $\varphi_1 = 0.5$, a relatively modest degree of autocorrelation, the optimal $\lambda_{HP}^{opt}$ is about three times higher than the inverse signal-to-noise ratio. For $\lambda = 0.9$, $\lambda_{HP}^{opt}$ is about ten times higher than $(\sigma_c^2/\sigma_\eta^2)$. This implies that the standard value of $\lambda_{HP} = 1600$ is much too low, even in cases where the assumption of Hodrick/Prescott $(\sigma_c^2/\sigma_\eta^2 = 1600)$ is valid.

Figure 1 shows the filter weights of the optimal Wiener-Kolmogorov (WK) filter (continuous line)

Table 1: Optimal $\lambda_{HP}^{opt}$ for different values of $\varphi_1$ and $\sigma_c^2/\sigma_\eta^2$ (Trend: RW(2), Cycle: AR(1))

| $\varphi_1$ $\sigma_c^2/\sigma_\eta^2$ | 800 | 1600 | 6400 |
|---|---|---|---|
| 0.0 | 800(1.0) | 1600(1.0) | 6400(1.0) |
| 0.1 | 975(1.2) | 1950(1.2) | 7811(1.2) |
| 0.3 | 1462(1.8) | 2938(1.8) | 11821(1.8) |
| 0.5 | 2304(2.9) | 4664(2.9) | 18926(3.0) |
| 0.7 | 4039(5.0) | 8359(5.2) | 34820(5.4) |
| 0.9 | 7439(9.3) | 18248(11.4) | 94043(14.7) |

Note: The numbers in parentheses denote the ratio $\lambda_{HP}^{opt}/(\sigma_c^2/\sigma_\eta^2)$

for the AR(1) model with $\varphi_1 = 0.7$ and the true inverse signal-to-noise ratio $(\sigma_c^2/\sigma_\eta^2) = 1600$ (equation (7)), of the HP(1600) filter (dashed line) and of the HP(8356) filter (dotted line) (equation (13)). The HP(8356) filter is the "optimal" HP filter (in the sense explained above) for the model. The weights for the WK and the HP(8356) filter are almost identical. Only for the central observation and the first lag and lead the weights for the HP(8356) filter are slightly lower than the weights for the WK filter. The weights for the standard HP(1600) filter, however, are far away from the optimal weights.

The pattern of filter weights carries over to the gain function of the filters (equations (8a), (15)). Figure 2 shows the gain for the three low-pass filters. The gains of the WK and the HP(8356) filter, respectively, are contiguous, whereas the gain of the HP(1600) filter is moved to the right (to higher frequencies). Especially for frequencies between 0.1 and 0.3 (this corresponds to periods of about 60 und 20 quarters) the gain of the HP(1600) filter is much higher than the gain of the optimal filter. Consequently, the trend extracting HP(1600) filter is too responsive to fluctuations which are commonly counted as business cycles. Similar results are obtained when we use different values for the autoregressive parameter and the true inverse signal-to-noise ratio. In all cases the message remains the same: The best approximation of the optimal Wiener-Kolmogorov filter is achieved by choosing a $\lambda$ higher than $\sigma_c^2/\sigma_\eta^2$.

Figure 1: Filter weights for WK, HP(1600) and HP(8356) filters (RW(2); AR(1); $\varphi_1 = 0.7$)

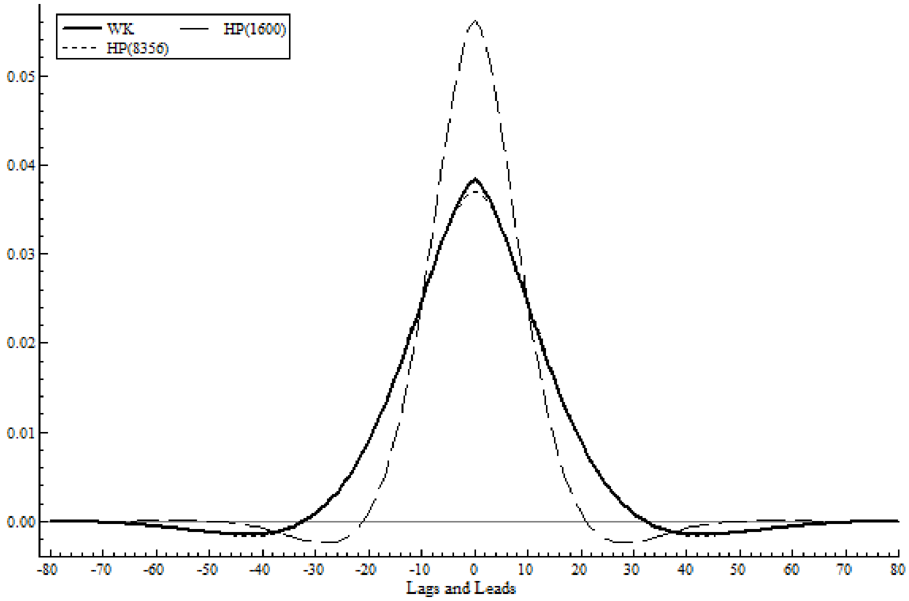



Figure 2: Gain functions of WK, HP(1600) and HP(8356) filters (RW(2), AR(1), $\varphi_1 = 0.7$)

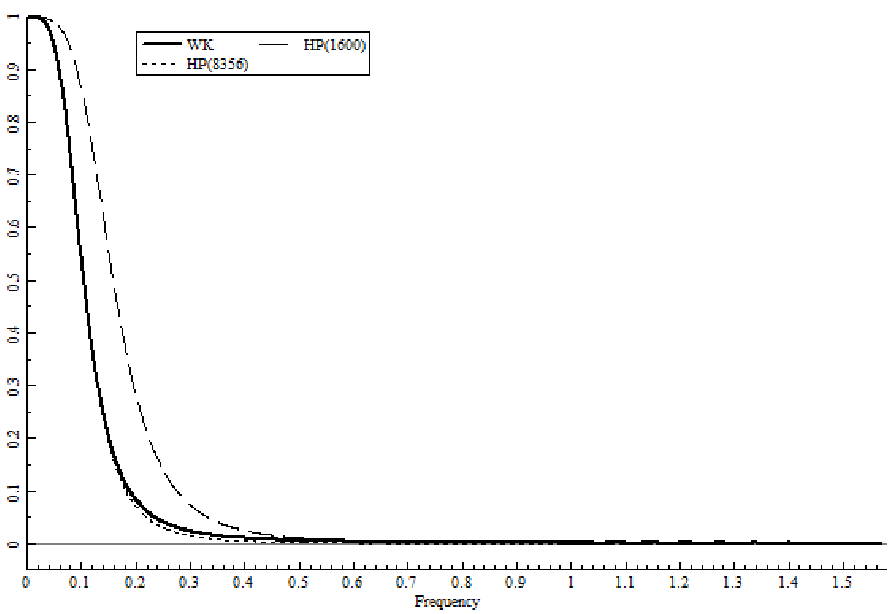
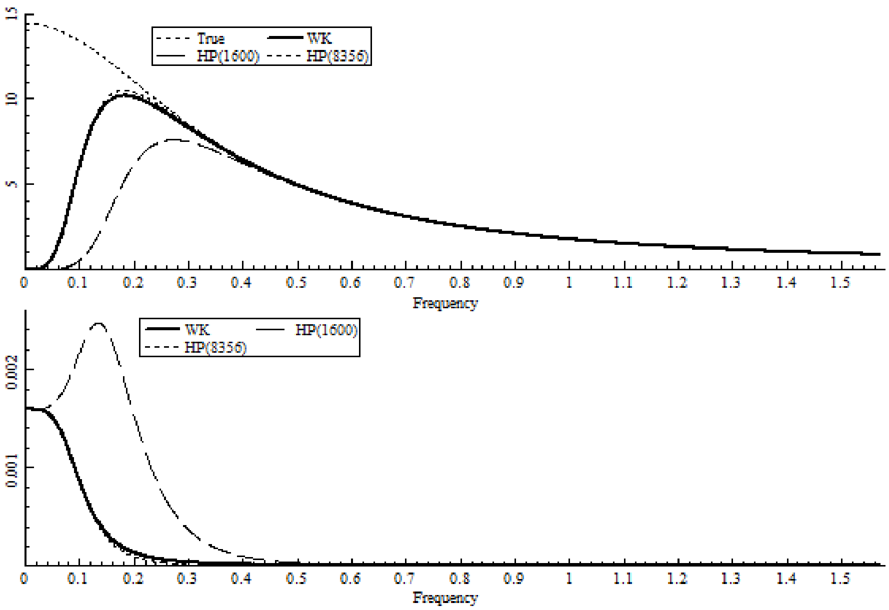


Figure 3: Spectra for the generated cycles (upper part) and for the second differences of generated trends (lower part) (RW(2); AR(1), $\varphi_1 = 0.7$)




In Figure 3 we evaluate the implied spectra of the generated cycle $\hat{c}$ and of the second differences the generated trend $\hat{\mu}$ (equations (10), (11), (12) and (16)). The upper part shows the spectra of the generated cycles (together with the spectrum of the true cycle). The spectrum of the cycle generated with the HP(1600) filter is again markedly different from that of the cycles generated with optimal filters. As the variance of a stationary time series is proportional to the integral over the spectrum, it is clear that the variance of the cycle generated by HP(1600) is much lower than the variance of the cycles generated by the WK and the HP(8536) filters. In addition, the peak of the spectrum of the cycle generated by HP(1600) is at a higher frequency. Consequently, the HP(1600) produces shorter and smaller cycles than the optimal filters.

An analagous pattern applies for the spectra of the second differences of the generated trends. Most important is that the HP(1600) filter produces a spectrum of $(1-L)^2\hat{\mu}_t$ with a pronounced peak in the region of business cycle frequencies: A value for $\lambda$ "too low" produces cycles in the second differences of the estimated trend. We will argue below that this feature can be used in practical applications for determining a reasonable value for the smoothing parameter.

#### 3.2.2 Trend RW(2) and cycle AR(2)

The model is given by

$$\begin{aligned} (1-L)^2\mu_t &= \eta_t \\ c_t &= \varphi_1 c_{t-1} + \varphi_2 c_{t-2} + \epsilon_t \end{aligned}$$

Table 2: Optimal $\lambda_{HP}^{opt}$ for different values of $\varphi_1$, $\varphi_2$ and $\sigma_c^2/\sigma_\eta^2$ (Trend: RW(2), Cycle: AR(1))

| $\varphi_1$, $\varphi_2$ \|\| $\sigma_c^2/\sigma_\eta^2$ | 800 | 1600 | 6400 |
|---|---|---|---|
| 1.109, -0.36 | 3627(4.5) | 7297(4.6) | 29385(4.6) |
| 1.663, -0.81 | 2476(3.1) | 4657(2.9) | 17233(2.7) |
| 1.177, -0.36 | 4871(6.1) | 9944(6.2) | 40753(6.4) |
| 1.765, -0.81 | 8379(10.5) | 15887(9.9) | 58905(9.2) |

There are many combinations of $\varphi_1$ and $\varphi_2$ compatible with the stationarity assumption. In the following we restrict the analysis to four combinations with complex roots in the AR polynomial: 1.) $\varphi_1 = 1.109$, $\varphi_2 = -0.36$; 2.) $\varphi_1 = 1.663$, $\varphi_2 = -0.81$; 3.) $\varphi_1 = 1.177$, $\varphi_2 = -0.36$; 4.) $\varphi_1 = 1.765$, $\varphi_2 = -0.81$. The parameters are chosen accordingly to the AR part of structural time series models (Harvey 1993, chapt. 6.5). We set $\varphi_1 = 2\rho\cos\omega_c$ and $\varphi_2 = -\rho^2$, where $|\rho| < 1$ is a damping factor and $\omega_c$ is the frequency of a cyclical function. The roots of the AR polynomial are a pair of complex conjugates with modulus $1/\rho$. Model 1 has a damping factor $\rho$ of 0.6 and a frequency $\omega_c$ of 0.393 (16 quarters), model 2 a damping factor of 0.9 and a frequency of 0.393, model 3 a damping factor of 0.6 and a frequency of 0.196 (32 quarters) and model 4 a damping factor of 0.9 and a frequency of 0.196. We have two models with a short and two models with a long cycle, combined with two different damping factors, 0.6 and 0.9, respectively.

Table 2 shows the optimal values for the HP smoothing parameter for the different models and three different values of the true inverse signal-to-noise ratio (800, 1600 and 6400). In all cases the optimal value for $\lambda$ is much higher than $\sigma_c^2/\sigma_\eta^2$. For parameter combination 1.) is is about 4.5 times higher, for combination 2.) about 3 times higher, for combination 3.) about 6 times higher and for combination 4.) about 10 times higher. For instance, if we assume a pronounced cycle with a period of 8 years and a true inverse signal-to-noise ratio of 1600, the optimal value for the HP smoothing parameter is 15887!

Figure 4 shows the filter weights of the optimal Wiener-Kolmogorov (WK) filter (continuous line), for the AR(2) model with $\varphi_1 = 1.765$, $\varphi_2 = -0.81$ and $\sigma_c^2/\sigma_\eta^2 = 1600$, of the standard HP(1600) filter (dashed line) and of the HP(15887) filter (dotted line). Contrary to the AR(1) case, the weighting pattern of the WK filter can not be fully replicated by a HP filter with a suitably chosen $\lambda$. The reason is that the weight function of the HP filter is always relatively smooth, whereas the weight function of the WK filter has in the model under consideration a sharp discontinuity for the central observation. However the difference between the weights of the WK and of the HP(15887) filter is very small for lags and leads higher than 4. In contrast, the shape of the weights of the HP(1600) filter is very different (the near coincidence with the weight of the WK filter for the central observation is accidental).

Figure 4: Filter weights for the WK, HP(1600) and HP(15887) filters. (RW(2), AR(2), $\varphi_1 = 1.765$, $\varphi_2 = -0.81$)

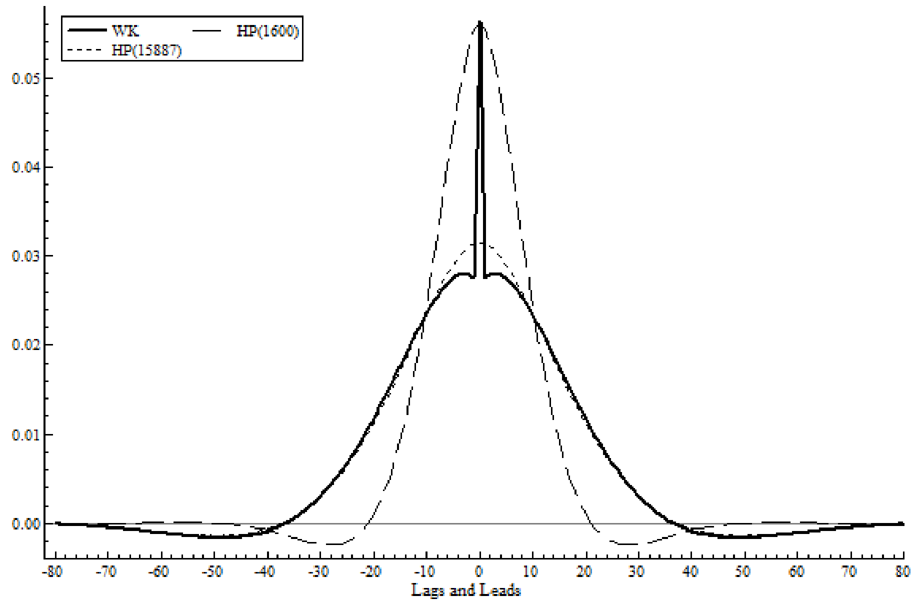



Figure 5 shows the gain for the three low-pass filters. For frequencies lower than about 0.25 (a period of about 6 years) the gain function of the WK and the HP(15887) filters are almost identical. For higher frequencies, the gain of the HP(15887) filter converges to zero, whereas the gain of the WK filter has small positive values. Similarily to the AR(1) model, the gain function of the traditional HP(1600) filter is very different from the gain of the WK filter.

Figure 5: Gain functions of WK, HP(1600) and HP(15887) filters (RW(2); AR(2), $\varphi_1 = 1.765$, $\varphi_2 = -0.81$)

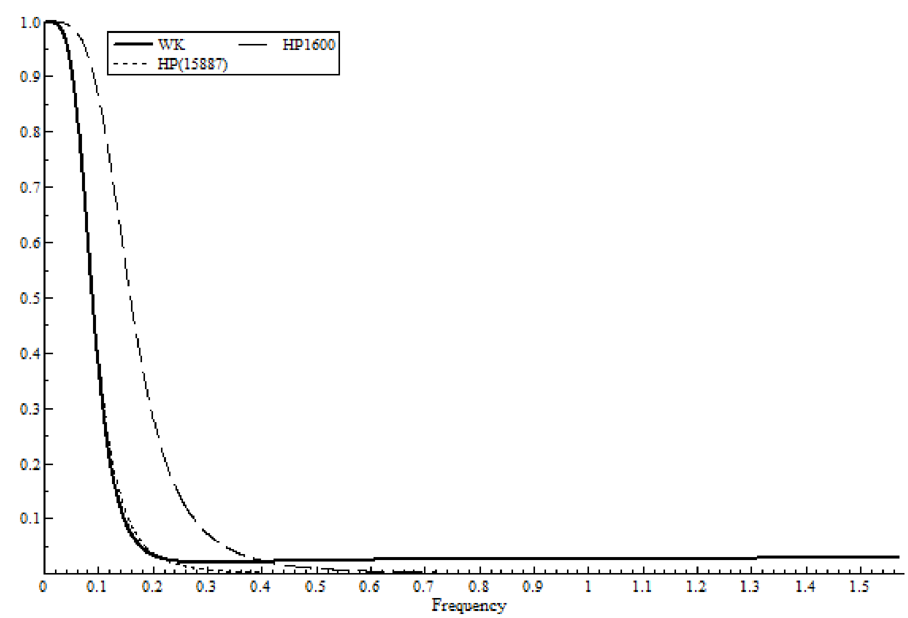


Figure 6: Spectra for the generated cycles (upper part) and for the second differences of generated trends (lower part). (RW(2); AR(2), $\varphi_1 = 1.765$, $\varphi_2 = -0.81$)

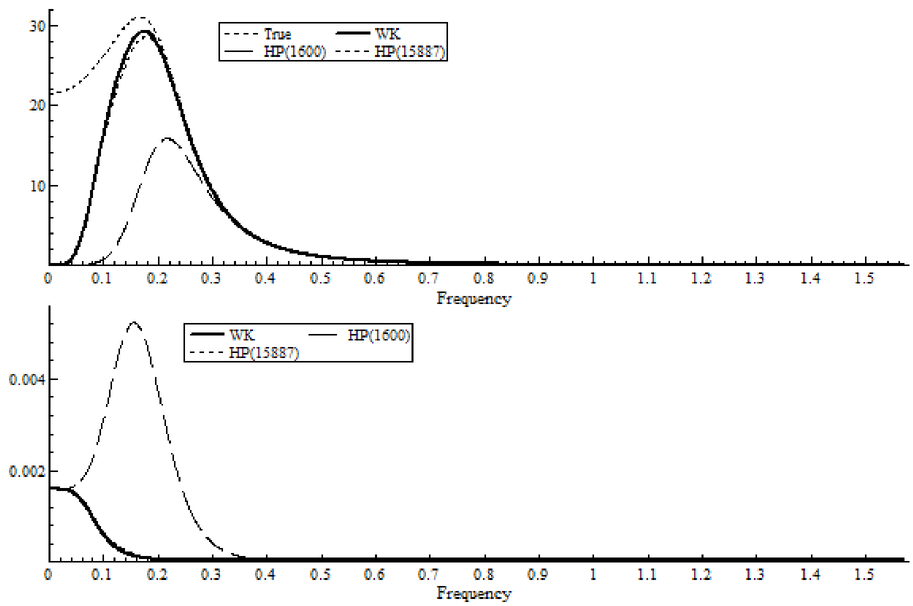



Figure 6 shows the spectra of the generated cycles (upper part) and of the second differences of the generated trend (lower part). Again, the spectra are very similar for the components generated by the WK and the HP(15887) filters, whereas the spectra generated by using the traditional HP(1600) filter are very different. The cycle is clearly underestimated by HP(1600) and the spectrum of the second differences of the estimated trend shows a very pronounced peak in the region of business cycle frequencies.

### 3.2.3 Trend RW(1) and cycle AR(1)

In this section we repeat the calculations for the case where the first differences of the trend are white noise (the trend is a random walk). Now, the inverse signal-to-noise ratio is much lower than in the case where the second differences of the trend are white noise. For instance, the unobserved components model for US GDP estimated by Watson (1985) implies a ratio $\sigma_c^2/\sigma_\eta^2$ of about 30. We calculated the optimal values for the HP filter for different models of the stationary process and three different values of the true inverse signal-to-noise ratios (10, 30 and 60).

Table 3 presents the optimal value for the HP smoothing parameter for different values of the AR(1) parameter and the true data-generating value of the inverse signal-to-noise ratio. In case of a RW(1)-trend the optimal values depend both an $\varphi_1$ and $\sigma_c^2/\sigma_\eta^2$. In all combinations (even for $\varphi_1 = 0$, i.e., the cycle is white noise) the optimal $\lambda$ is much higher than the true inverse signal-to-noise ratio.

Figure 7 shows the weights for the WK, the HP(1600) and the HP(26313) filter. The last fil-

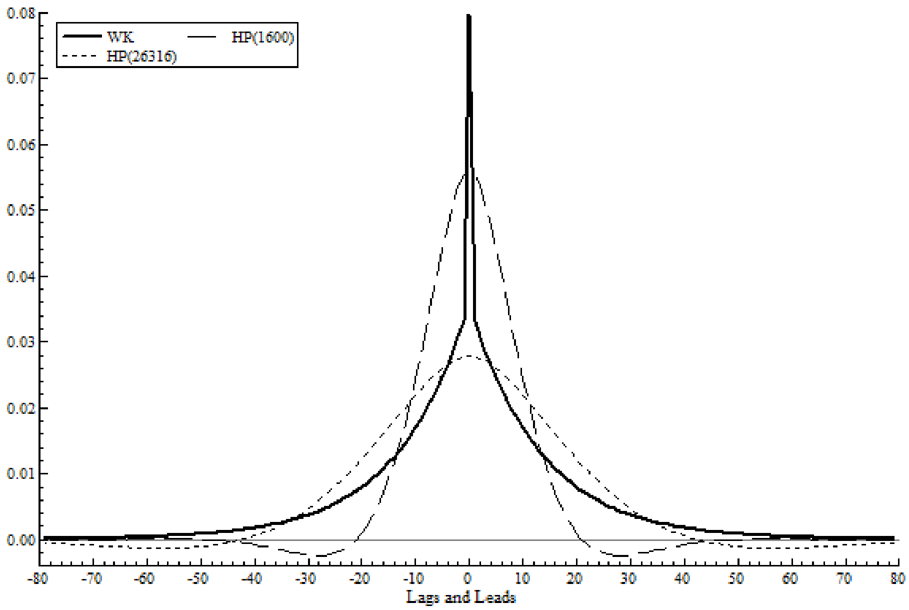
Table 3: Optimal $\lambda_{HP}$ for different values of $\varphi_1$ and $\sigma_c^2/\sigma_\eta^2$ (Trend: RW(1), Cycle: AR(1))

| $\varphi_1 \| \sigma_c^2/\sigma_\eta^2$ | 10 | 30 | 60 |
|---|---|---|---|
| 0.0 | 100(10.0) | 900(30.0) | 3600(60.0) |
| 0.1 | 146(14.6) | 1335(44.5) | 5360(89.4) |
| 0.3 | 322(32.2) | 3036(101.2) | 12282(204.7) |
| 0.5 | 784(78.4) | 7744(258.1) | 31692(528.2) |
| 0.7 | 2390(239.0) | 26316(877.2) | 110411(1840.2) |
| 0.9 | 10002(1000.2) | 230451(7681.7) | 1103807(18396.8) |

Note: The number in parentheses denote the ratio $\lambda_{HP}/(\sigma_c^2/\sigma_\eta^2)$.

ter is the "optimal" HP filter for $\varphi_1 = 0.7$ and $\sigma_c^2/\sigma_\eta^2 = 30$. The weights for both HP filters do not follow closely the pattern for the WK filter. This is not really surprising as it is well known that the WK filter for a random walk trend is the exponential smoothing filter (King/Rebelo 1993; Proietti 2007; Harvey/Delle Monache 2009).

Figure 7: Filter weights for WK, HP(1600) and HP(26316) (RW(1), AR(1), $\varphi_1 = 0.7$)




The very poor performance of the HP filter is confirmed in Figure 8 where the gain functions are shown. In the region of business cycle frequencies (say, about 0.2) the gains of the HP filter are far away from the gain of the WK filter: The HP(1600) filter transfers too much from business cycle fluctuations to the trend, the HP(26316) too little.

Figure 8: Gain functions of WK, HP(1600) and HP(26316) filters (RW(1); AR(1), $\varphi_1 = 0.7$)

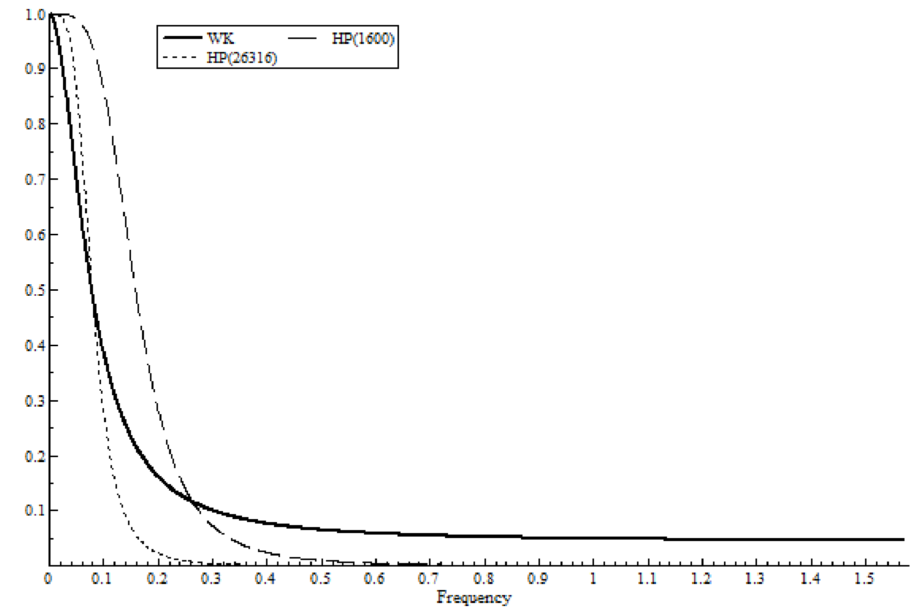
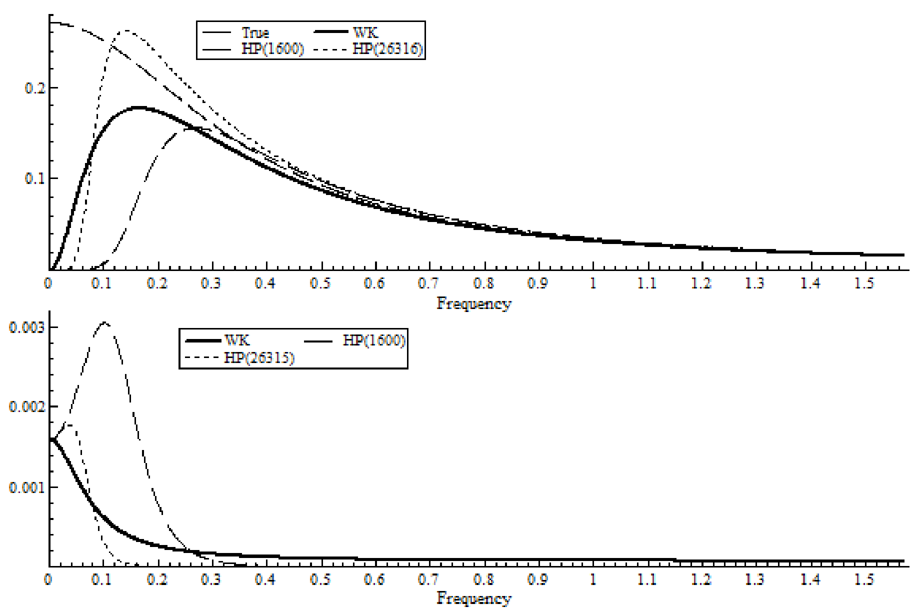



Figure 9: Spectra for the generated cycles (upper part) and for the second differences of generated trends (lower part). (RW(1); AR(1), $\varphi_1 = 0.7$)




The distortionary effects can also be seen in the spectra for the cycle and the first differences of the trend (Figure 9). The HP(1600) filter underestimates the cycle and leads to a cyclical movement in the first differences, the HP(26316) filter overestimates the cycle. The conclusion from these exercises is that the HP filter does not work satisfactorily when the trend follows a random walk. In this case exponential smoothing may be a much better choice.

### 3.3 A simulation study

The results in the previous sections are derived under the assumption of a double infinite time series. In finite time series we have at the start and the end of the sample asymmetric filters. In order to check the ability of the HP filter with a relatively high smoothing parameter to replicate the main properties of the WK filter for autocorrelated cycle processes we carry out a simulation study for a finite time series with 160 observations.

The model is given by

$$\begin{aligned} y_t &= \mu_t + c_t \\ (1-L)^2\mu_t &= \eta_t \\ c_t &= \varphi_1 c_{t-1} + \varphi_2 c_{t-2} + \epsilon_t \end{aligned}$$

$\eta_t$ and $\epsilon_t$ are mutually uncorrelated white noise processes. The inverse signal-to-noise ratio $\sigma_c^2/\sigma_\eta^2$ is set to the three alternative values 800, 1600 and 6400.

We generate series for $\{\mu_t\}$, $\{c_t\}$ and $\{y_t\}$, $t = 1, ..., 160$ and filter the "observed" time series $\{y_t\}$ with the optimal WK filter and the HP filter, using for the latter different values of the smoothing parameter $\lambda$.

The estimated trend values $\hat{\mu}$ are generated by using the matrix formula (McElroy 2008; Flaig 2012):

$$\hat{\mu} = (C_c^{-1} + \lambda D'D)^{-1} C_c^{-1} y$$

$C_c$ is the $T \times T$ correlation matrix of $\{c_t\}$, $\lambda$ is the true inverse signal-to-noise ratio $\sigma_c^2/\sigma_\eta^2$ and $D$ is the $(T-2) \times (T-2)$ differencing matrix, given by

$$D = \begin{pmatrix} 1 & -2 & 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & & & \\ 0 & 0 & 0 & 0 & \cdots & 1 & -2 & 1 \end{pmatrix}$$

For estimating $\mu$ with the HP filter we set $C_c = I$ and $\lambda$ to a prespecified value.

The simulation study consists of 1000 replications of the described procedure. For each replication we calculated the mean absolute error $MAE = \left(\sum_t |\mu_t - \hat{\mu}_t|\right)/T$ and the mean squared error $MSE = \left(\sum_t (\mu_t - \hat{\mu}_t)^2\right)/T$ for the different filters.

Table 4: Relative efficiency of different HP filters compared to WK filter (Trend: RW(2), Cycle: AR(1)

| $\varphi_1 \| \sigma_c^2/\sigma_\eta^2$ | 800 | | 1600 | | 6400 | |
|---|---|---|---|---|---|---|
| | HP(800) | HP(opt) | HP(1600) | HP(opt) | HP(6400) | HP(opt) |
| 0.0 | 1.000 | 1.000(1) | 1.000 | 1.000(1) | 1.000 | 1.000(1) |
| 0.1 | 0.996 | 0.996(1) | 0.994 | 0.994(1) | 0.996 | 0.996(1) |
| 0.3 | 0.964 | 0.996(2) | 0.970 | 0.995(2) | 0.972 | 0.992(2) |
| 0.5 | 0.912 | 0.990(3) | 0.914 | 0.992(3) | 0.912 | 0.992(3) |
| 0.7 | 0.829 | 0.984(5) | 0.828 | 0.984(5) | 0.814 | 0.984(5) |
| 0.9 | 0.796 | 0.964(10) | 0.740 | 0.953(12) | 0.682 | 0.943(14) |

Note: The numbers in parentheses denote the ratio of the "optimal" HP smoothing parameter to the true inverse signal-to-noise ratio.

Following Harvey/Delle Monache (2009) we assess the efficiency of a filter by the ratio $MSE_{WK}/MSE_{HP}$, where $MSE_{WK}$ and $MSE_{HP}$ are the mean squared error of the Wiener-Kolmogorov filter and the HP filter, respectively. Table 4 shows the relative efficiency of different HP filters compared to the WK filter for different values of $\varphi_1$ for an AR(1) model of the cycle and for different values of the true inverse signal-to-noise ratio $\sigma_c^2/\sigma_\eta^2$. The efficiency in the columns labeled as HP(opt) are obtained in the following way:

For each model (characterized by $\varphi_1$ and $\sigma_c^2/\sigma_\eta^2$) we generate 1000 series of 160 observations for trend $\mu$, cycle $c$ and the "observed" time series $y$ (trend + cycle). For each time series we calculate the mean square error $MSE = \left(\sum_t(\mu_t - \hat{\mu}_t)^2\right)/T$ for the WK filter and for 15 HP filters with smoothing parameter $\lambda_j = j\sigma_c^2/\sigma_\eta^2$, $j = 1, 2, ..., 15$. An entry in Table 4 shows the mean of 1000 values for $MSE_{WK}/MSE_{HP}$. For each true inverse signal-to-noise ratio (800, 1600 and 6400) there are two columns of results. The first column shows the relative efficiency of the HP($\sigma_c^2/\sigma_\eta^2$) filter, the second the relative efficiency when we choose the "optimal" $\lambda$. The numbers in parentheses denote the ratio of the "optimal" smoothing parameter to $\sigma_c^2/\sigma_\eta^2$.

The results indicate that in case the cycle follows an AR(1) process one can get an impressive efficiency gain by choosing an HP smoothing parameter higher than the true inverse signal-to-noise ratio. Take, for example, a model with $\sigma_c^2/\sigma_\eta^2 = 1600$ and $\varphi_1 = 0.7$. Compared with the WK filter, the HP(1600) filter has a relative efficiency of 0.83, whereas the HP(8000) filter has a relative efficiency of 0.98.

Table 5 reports the results for four AR(2) processes. The results confirm the conclusions for the AR(1) case. By choosing a reasonably high value of the HP smoothing parameter one can achieve a

Table 5: Relative efficiency of different HP filters compared to WK filter (Trend: RW(2), Cycle: AR(2))

| $\varphi_1$, $\varphi_2$ \|\| $\sigma_c^2/\sigma_\eta^2$ | 800 | | 1600 | | 6400 | |
|---|---|---|---|---|---|---|
| | HP(800) | HP(opt) | HP(1600) | HP(opt) | HP(6400) | HP(opt) |
| 1.109, -0.36 | 0.835 | 0.980(5) | 0.837 | 0.980(5) | 0.819 | 0.974(5) |
| 1.663, -0.81 | 0.810 | 0.935(3) | 0.828 | 0.949(3) | 0.841 | 0.943(3) |
| 1.177, -0.36 | 0.794 | 0.972(6) | 0.779 | 0.968(6) | 0.778 | 0.962(6) |
| 1.765, -0.81 | 0.629 | 0.927(11) | 0.626 | 0.912(10) | 0.640 | 0.919(10) |

Note: The numbers in parentheses denote the ratio of the "optimal" HP smoothing parameter to the true inverse signal-to-noise ratio.

remarkable efficiency gain compared with the HP filter where $\lambda$ is set to the true inverse signal-to-noise ratio. Compared with the AR(1) model for the cycle the maximal efficiency is now somewhat lower. For instance, for case 4 ($\varphi_1 = 1.765$, $\varphi_2 = -0.81$) and a true inverse signal-to-noise ratio of 1600, the relative efficiency (compared to the WK filter) is 91 %. However, the HP(1600) filter has only a relative efficiency of 63 %. The reward of using a high smoothing parameter is still high.

The results generated by the simulation study using a finite length of the time series confirm the conclusions of the theoretical analysis for doubly infinite series: When the stationary component of a time series is autocorrelated, the optional value for the HP smoothing parameter is several times higher than the inverse signal-to-noise ratio.

# 4 Choosing $\lambda$ in practical applications: A new proposal

It is common knowledge that the HP filter may induce spurious cycles. An often used example is the case of a random walk as the input series (Kaiser/Maravall 2001). In this case, the HP filter typically produces cycles with periods between 8 and 10 years ($\lambda = 1600$). The main argument here concentrates on the contrary danger. The basic assumption is that the cyclical component is not white noise but follows an autocorrelated process. In this case it is necessary to choose a value for the smoothing parameter $\lambda$ that is much higher than the true or assumed inverse signal-to-noise ratio. The problem for the practitioner is that we do not know the parameters of the data-generating process (at least in situations where it is too difficult or too costly to estimate the parameter of structural models).

In this section we propose the following (partial) solution to this problem. We start with the assumption that the generated trend component should not exhibit any cyclical features. Since the trend is not stationary, we concentrate on the first and/or second differences of the trend. We identify a possible cycle in the differences of the generated trend using the spectrum of

$(1-L)^d \tilde{\mu}_t$, $d = 1, 2$, where $\tilde{\mu}_t$ is the HP-generated trend component.

Using the results in section 2.3, we can write the spectrum $g_{\Delta d\tilde{\mu}}$ of $(1-L)^d \tilde{\mu}_t$ as

$$\begin{aligned} g_{\Delta d\tilde{\mu}} &= [2(1-\cos\omega)]^d \, |\tilde{M}(\omega)|^2 g_y(\omega) \\ &= (\sigma_n^2/2\pi) \left[1 + 4\lambda(1-\cos\omega)^2\right]^{-2} \left\{[2(1-\cos\omega)]^{d-n} + [2(1-\cos\omega)]^d \, \lambda^* \tilde{g}_c(\omega)\right\} \end{aligned}$$

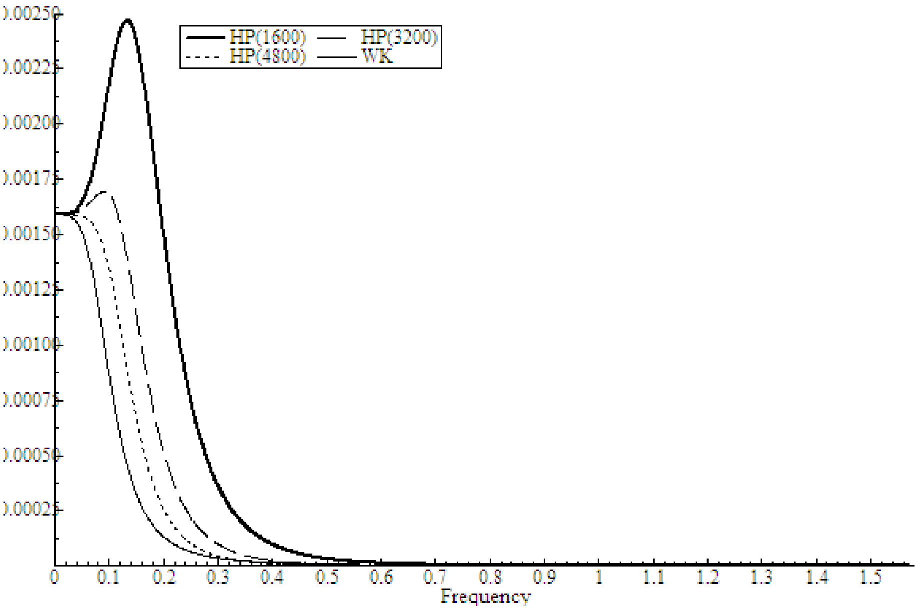
where $\lambda$ is the HP smoothing parameter and $\lambda^*$ is the true inverse signal-to-noise ratio. If $d < n$, $(1-L)^d \tilde{\mu}_t$ is not stationary.

Given the parameters of the true data-generating process, the shape of $g_{\Delta d\tilde{\mu}}$ is determined by the smoothing parameter $\lambda$. In the following we concentrate on the case $d = n$. If $\lambda = 0$, $g_{\Delta d\tilde{\mu}}$ is an increasing function of $\omega$, if $\lambda$ is very high, $g_{\Delta d\tilde{\mu}}$ is a decreasing function of $\omega$. For values in between, it is possible that $g_{\Delta d\tilde{\mu}}$ has a peak for $0 < \omega < \pi$. If this occurs, we have cycles in the differences of the generated trend.

Figure 10 shows the spectra of $g_{\Delta 2\tilde{\mu}}$ for a model where the trend is an integrated random walk ($n = 2$) and the cycle follows an AR(1) process with $\varphi_1 = 0.7$. The thick continuous line shows the spectrum for the HP(1600) filter. It has a pronounced peak at frequency 0.133 (which corresponds to a period of 47 quarters). The dashed line shows the spectrum for the HP(3200) filter. We have a peak at frequency 0.091, which is less pronounced than the peak for the HP(1600).

The spectrum of the HP(4800) filter (dotted line) has no peak, but is nevertheless shifted to the right compared to the WK filter (thin continuous line). The trend generated by the HP(4800) is still to responsive to fluctuations with business cycle frequencies. We know that the “optimal” value for $\lambda$ in this case is 8359 (see section 3.2.1). We conclude that the lowest value for $\lambda$ that does not generate a peak in the spectrum of the second differences of the extracted trend is about half as high as the “optimal” value. This is roughly confirmed by calculations for other models of the cycle.

Figure 10: Spectra of second differences of generated trends for different values of $\lambda$ (RW(2), AR(1), $\varphi_1 = 0.7$)

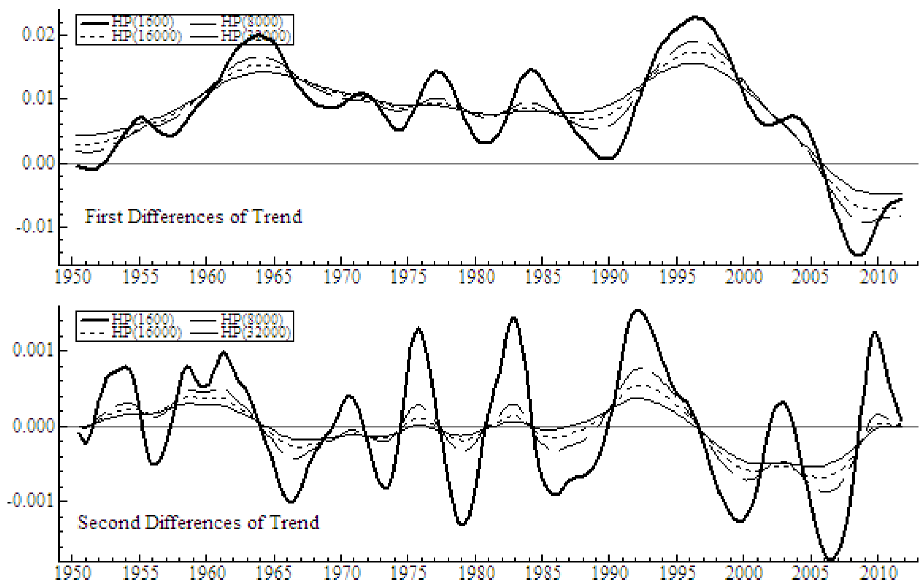



# 5 An empirical example: US Private Investment

In this section we study the effects of different values of the HP smoothing parameter on the properties of the estimated trend of US Real Gross Private Domestic Investment (logarithmic values; 1950:1 - 2011:4). We estimate the trend component with the HP filter with values for $\lambda$ of 1600, 8000, 16000 and 32000. Figure 11 shows in the upper part the growth rates for the generated trends, in the lower parts changes in the growth rates (second differences). The thick continuous line shows the results for the HP(1600) filter. Both first and second differences display clearly cycles with a period of about 8 years. Oscillations with this period are counted by many economists as business cycles. If one accepts that business cycles can appear in the growth rates of the trend, it is fine. But the general definition of the trend does not allow for cyclical elements. In this interpretation, the generated "trend" is a mixture of trend and cycle and, consequently, an artefact of the filter.

To a lesser degree, the first and second differences of trends generated by the HP(8000) (dashed line) and HP(16000) (dotted line) show a similar picture. When we use HP(32000), the differences display no cyclical element.

Figure 11: First and second differences of HP generated trends (US Private Investment)




From the perspective of the criterion that the trend and its differences should not exhibit any form of a cycle it is clear from the previous discussion that for real investment a smoothing parameter of 1600 is not appropriate. The minimum reasonable values for $\lambda$ is in the region of 20000 to 40000. Harvey/Trimbur (2008) suggest a value of 32000 (based on a somewhat different line of arguments). Using the result of section 4 one can argue that the "optimal" value may be even higher, say about 60000.

The choice of the smoothing parameter has far-reaching consequences for the size and the dynamic properties of the HP-generated cycle component. In the example of US Private Investment, the standard deviation of the cycle is 0.077 for $\lambda = 1600$ and 0.094 for $\lambda = 32000$. And the autocorrelation function decays at a much slower rate for higher $\lambda$- values. For instance, the autocorrelation coefficient at lag 1 (4) is 0.80 (0.03) for $\lambda = 1600$ and 0.87 (0.29) for $\lambda = 32000$. It is left for future research to analyze the implication of different smoothing parameters for other variables (GDP, consumption, employment, etc.) and to explore the consequences for the construction of "stylized facts" of the business cycle.

# 6 Summary and conclusions

When we interpret the Hodrick-Prescott filter as a model-baed filter it is well known that it is the optimal Wiener-Kolmogorov filter if the trend follows an integrated random walk, the cycle is white noise and the smoothing parameter $\lambda$ is set to the inverse signal-to-noise ratio. In the traditional trend-cycle decomposition these assumptions are in many cases clearly implausible and the HP filter lacks a sound theoretical foundation.

In this paper we concentrate on the situation where the cycle follows an AR(1) or AR(2) process and ask the question whether it is possible to reach a reasonable approximation of the optimal Wiener-Kolmogorov filter by the HP filter with an appropriate chosen value for the smoothing parameter $\lambda$. The analysis is done in the following way: First, we calculate from the gain function of the optimal Wiener-Kolmogorov filter the frequency where the gain is 0.5. Secondly, we determine the value of $\lambda$ for which the gain function of the HP filter has also the value of 0.5 at the same frequency. In the last step we compare the mean square error of both filters (the WK filter and the HP filter with the "optimized" value of $\lambda$). These calculations are carried out for different specifications of the AR-parameters of the cycle, different values of the inverse signal-to-noise ratio and different specifications of the trend component.

The general result is that in case the trend follows an integrated random walk one gets a relatively good approximation of the weights and the gain function of the optimal Wiener-Kolmogorov filter by choosing a value for the HP smoothing parameter much higher than the inverse signal-to-noise ratio. For example, when the cycle follows an AR(1) process with a parameter $\varphi_1 = 0.9$, the "optimal" value of $\lambda$ is more than 10 imes higher than the true inverse signal-to-noise ratio.

Smoothing parameters "too low" have a twofold distortionary effect. They produce trends with first and/or second differences which exhibit cyclical features. The trend is too responsive to business cycle fluctuations. This implies that the variance and the period of the generated cycle are too low. The relevance of the cyclical component is underestimated.

When the trend is a random walk the approximation is not really satisfactory. It is not possible to replicate the shape of the weight and gain function of the optimal Wiener-Kolmogorov filter by the HP filter. However, the general result remains that we should select higher values for $\lambda$ than usually chosen (e.g., $\lambda = 1600$).

These findings are useful, but not really applicable in practice as we do not know the true inverse signal-to-noise ratio. The problem is how to choose an appropriate value for the HP smoothing parameter. The suggestion proposed in this paper is to rely on the properties of the first and/or second differences of the extracted trend. The proposal is based on the assumption that the differences of the extracted trend should not show any cyclical behaviour (in the sense that the spectrum of the differences has a peak in the region of business cycle frequencies). It is shown that by choosing a high enough smoothing parameter a peak can always be avoided. In practical applications we could estimate the trend by applying the HP filter with different values for $\lambda$ and search for the lowest among them which does not produce cycles in the differences of the generated trend.

In the last section the proposal procedure is applied to US private real investment. We find that the lowest value for $\lambda$ that does not produce cycles in the differences of the trend component is about 32000. With some caution, we can conclude that the “optimal” value of the smoothing parameter may be approximately 60000!

The general conclusion of this paper is that the “industry standard” of $\lambda = 1600$ may be much too low for many macroeconomic time series. In order to produce reasonable and reliable trend-cycle decomposition much higher values are necessary. It is left for future research to explore the implications for the “stylized facts” of business cycles.

## References


Bell, William (1984), Signal Extraction for nonstationary time series. Annals of Statistics, 12, 646 - 664

Beveridge, Stephen and Charles R. Nelson (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, 151 - 174.

Box, George E. P., Steven C. Hillmer and George C. Tiao (1978), Analysis and Modeling of Seasonal Time Series. In: Arnold Zellner (ed.), Seasonal Analysis of Economic Time Series. U.S. Department of Commerce.

Cogley, Timothy and James Nason (1995), Effects of the Hodrick-Prescott filter on trend and difference stationary time series. Implications for business cycle research. Journal of Economic Dynamic and Control, 19, 253 - 278

Flaig, Gebhard (2012), Penalized Least Squares and Some Simple Extensions of the HP filter. Mimeo

Gomez, Victor (1999), Three Equivalent Methods for Filtering Finite Nonstationary Time Series. Journal of Business & Economic Statistics 17, 109-116

Harvey, Andrew (1989), Forecasting, Structural Time Series Models and the Kahman Filter. Cambridge University Press.

Harvey, Andrew (1993), Time Series Models. MIT Press

Harvey, Andrew and Albert Jaeger (1993), Detrending, Stylized Facts and the Business Cycle. Journal of Applied Econometrics, 8, 231 - 247

Harvey, Andrew and Thomas Trimbur (2008), Trend Estimation and the Hodrick-Prescott Filter. Journal of the Japan Statistical Society, 38, 41-49

Harvey, Andrew and Davide Delle Monache (2009), Computing the Mean Square Error of Unobserved Components Extracted by Misspecified Time Series Models. Journal of Economic Dynamics & Control, 33, 283-295

Hodrick, Robert and Edward Prescott (1997), Postwar U.S. Business Cycles. An Empirical Investigation. Journal of Money, Credit, and Banking, 29, 1-16.

Kaiser, Regina and Agustin Maravall (2001), Measuring Business Cycles in Economic Time Series. Springer Verlag

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

McElroy, Tucker (2008), Matrix Formulas for Nonstationary ARIMA Signal Extraction. Econometric Theory 24, 988-1009

Proietti, Tommaso (2007), Signal Extraction and Filtering by Linear Semiparametric Methods. Computational Statistics & Data Analysis 52, 935-958

Sargent, Thomas (1987), Macroeconomic Theory. Academic Press.

Watson, Mark (1986), Univariate Detrending Methods with Stochastic Trends. Journal of Monetary Economics, 18, 49 - 75

Whittle, P. (1983), Prediction and Regulation. Blackwell