The international platform of Ludwig-Maximilians University's Center for Economic Studies and the Ifo Institute

CESifo Working Papers

www.cesifo.org/wp

The Hodrick-Prescott Filter with a Time-Varying Penalization Parameter. An Application for the Trend Estimation of Global Temperature.

Andreas Blöchl
Gebhard Flaig

CESIFO WORKING PAPER NO. 4577
CATEGORY 12: EMPIRICAL AND THEORETICAL METHODS
JANUARY 2014

CESifo *Working Paper No. 4577*

The Hodrick-Prescott Filter with a Time-Varying Penalization Parameter. An Application for the Trend Estimation of Global Temperature.

Abstract

In this paper we use the Hodrick-Prescott filter for analysing global temperature data. We are especially concerned with a reliable estimation of the trend component at the end of the data sample. To this end we employ time-varying values for the penalization parameter. The optimal values are derived by a comparison with an ideal filter. The method is applied to temperature data for the northern hemisphere from 1850 to 2012. The main result is that for the optimal specification of the flexible penalization the trend component of temperature is still increasing, possibly with a somewhat lower pace.

JEL-Code: Q540, C220.

Keywords: climate change, global warming, trend, Hodrick-Prescott filter, flexible penalization.

Andreas Blöchl
Department of Economics
University of Munich
Schackstrasse 4
Germany - 80539 Munich
andreas.bloechl@lrz.uni-muenchen.de

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

1 Introduction

It is a generally accepted fact that global temperature increased noticeably all through the last three decades of the previous century. However, during the last ten to 15 years (say from 2000 to 2012) the temperature data show fluctuations more around a constant value or even a slight decrease. This observation triggered a sometimes heated discussion concerning the appropriate interpretation and possible consequences for climate policies. Some argue that the data indicate only a "pause" in the long-run increase in temperature that is caused by cyclical fluctuations around an unchanged trend. This behaviour might be traced back to a declining insolation due to the eleven years solar cycle as well as to the rise of sulfur emissions during the last 15 years (Kaufmann et al., 2011), or to a heat uptake by the oceans (Meehl et al., 2011). However, others claim that we should interpret the recent temperature data as a sign of a trend reversal.

This situation is a typical example for the difficulties and problems arising when we try to get a reliable estimate of the trend at the current end of a data sample. Here we have only current and past data and one has to use an asymmetric filter. The situation is different from that in the middle of a time series where we have past and future observations which allow the use of a symmetric filter (for surveys of different filter techniques see Mills, 2006, 2009). Symmetric filters deliver much more reliable estimates of an unobserved component like the trend of a time series. This means that asymmetric filters have some special problems which are addressed in the following.

In this paper we compare two filters which are widely used in climatology with a filter which is well-known in economics as the Hodrick-Prescott (Hodrick/Prescott, 1997). The first two filters are the binomial filter and the Gaussian filter. Both are two-sided symmetric filters. For the estimation of the trend value in a given period they use past, present and future observations. The only possibility to estimate with a symmetric filter the trend at the end of the sample is to extend the time series by forecasts. This procedure is not satisfactory, since the quality of the trend estimation depends heavily on the quality of the forecasts. For a reliable estimate of the current trend one needs at least 10 years of forecasts of the time series. In many cases there may not exist a consensus concerning the appropriate time series model used for forecasting. Consequently, the generated trend depends on arbitrary assumptions.

The Hodrick-Prescott filter, that is based on the ideas of Whittaker (1923), Henderson (1924) and Leser (1961), is the result of a well defined optimization problem. In the middle of the time series (all observations which are far enough away from the end of the sample) the filter is symmetric, at the end it gets more and more asymmetric. The properties of the filter depend on a penalization parameter that has to be specified by the researcher. We discuss a method how one can find a reasonable value for this parameter using spectral analysis. A novel contribution to the literature is that we allow time-varying parameter values. This enables us to improve the filter quality especially at the end of the time series and to get more reliable estimates of the current trend.

The paper is organized as follows. The first two sections summarize the binomial filter, the Gaussian filter and the Hodrick-Prescott filter. Afterwards it is explained how the penalization parameter of the Hodrick-Prescott filter is selected using spectral analysis and why the Hodrick-Prescott filter is superior to the Gaussian filter in the frequency domain. Afterwards the paper deals with the excess variability of the Hodrick-Prescott filter at the margins of a time series and describes how it can be reduced by a flexible penalization. Finally, empirical examples for the HadCRUT4 data for the northern hemisphere annual average temperature are presented.

2 The binomial and Gaussian filter

Two widely used filters for modeling the trend of temperature data are the binomial filter and the Gaussian filter. The binomial filter is for instance used by the Met Office Hadley Centre for the construction of smoothed annual average temperature values. Gaussian filters are for instance used in HISTALP project (Auer/Böhm/Schöner, 2001). Both filters are low-pass filters which extract from an observed time series the trend, i.e. the fluctuations with "low" frequencies. Low-pass filters let pass fluctuations with low frequencies with no or only minor modifications and suppress fluctuations with high frequencies. In principle, the filters are a sort of a finite symmetric weighted moving average. For the estimation of the trend value in a given period they use N observations (N is an odd integer): The current value, $(N-1)/2$ past and $(N-1)/2$ future values. If y_t denotes the observed time series, the estimated trend for period t, $\hat{\mu}_t$, is calculated as

$$\hat{\mu}_t = \sum_{j=-n}^{n} c_j y_{t-j}, \tag{1}$$

where c_j are the filter weights and $n = (N-1)/2$.

2.1 The binomial filter

The filter weights c_j of the binomial filter are given by the normalized binomial coefficients

$$c_j = \binom{N-1}{n+j} / 2^{N-1}, \quad j = -n, ..., 0, ...n. \tag{2}$$

The factor 2^{N-1} ensures that the sum of the filter weights is one. This is a necessary condition for all low-pass filters used for smoothing and for the extraction of the trend component of a time series (Osborn, 1995). The formulation implies that the filter function is symmetric, i.e. $c_j = c_{-j}$.

Alternatively, the filter weights can be calculated by the $(N-2)$ fold convolution of the sequence (0.5, 0.5) with itself (for the concept of convolutions see Gourieroux/Monfort, 1997, Appendix 7.4).

In the following we analyse the properties of filters in the frequency domain. The frequency response function is the Fourier transform of the sequence of the filter weights and is defined as (Mills, 2003)

$$a(\omega, N) = \sum_{j=-n}^{n} e^{-i\omega j} c_j, \tag{3}$$

where ω is the angular frequency and i is the imaginary unit ($i^2 = -1$). Inserting formula (2) for c_j we get

$$a(\omega, N) = 2^{-(N-1)} \sum_{j=-n}^{n} \binom{N-1}{n+j} e^{-i\omega j} = 2^{-(N-1)} \sum_{j=0}^{N-1} \binom{N-1}{j} e^{-i\omega(j-n)} =$$

$$= 2^{-(N-1)} e^{i\omega n} \sum_{j=0}^{N-1} \binom{N-1}{j} e^{-i\omega j}.$$

Using the Binomial theorem we could write the sum as $(1 + e^{-i\omega})^{N-1}$.

From this we could derive (using Euler's formula)

$$a(\omega, N) = 2^{-(N-1)} e^{i\omega(N-1)/2} \left[e^{-i\omega/2}(e^{-i\omega/2} + e^{i\omega/2}) \right]^{N-1}$$

$$= 2^{-(N-1)} \left(\frac{\cos(\omega/2)}{2} \right)^{N-1} = \cos^{N-1}(\omega/2). \tag{4}$$

Additional information about the properties of a filter yields the gain function. This is defined as

$$g(\omega, N) = \sqrt{a(\omega, N) a(-\omega, N)}. \tag{5}$$

The gain is interpreted as the factor by which the amplitude of an oscillation with frequency ω is damped or amplified. Considering the gain for a bandwidth of frequencies, usually the interval $[0, \pi]$, gives insight about the qualities of a filter as an instrument for trend estimation. For a symmetric filter the gain is identical to the frequency response function.

For an application of the binomial filter the number of used observations, N, has to be determined. One possibility is to choose N in order that for a prespecified frequency ω_0 the gain function has a specific value c_0, say 0.5. From the above formula for the frequency response function we could calculate N as $N = \ln(c_0)/\ln(\cos(\omega_0/2)) + 1$. As an example, suppose you want to construct a filter for which the gain function is less than 0.5 for all fluctuations with a periodicity less than 5 (that is with a frequency greater than $2\pi/5$), then $N = 5.54$. To ensure that N is an odd integer, we could round this number to 5 periods.

2.2 The Gaussian filter

For large values of N the binomial filter converges to a Gaussian low-pass filter (Panofsky/Brier, 1958; Mitchell et al., 1966). The Gaussian filter is a bi-infinite filter. The filter weights $\{c_j\}$ of the Gaussian low-pass filter are given by

$$c_j = (2\pi\sigma^2)^{-\frac{1}{2}} \exp\left(\frac{-j^2}{2\sigma^2}\right), \quad j = 0, \pm 1, \pm 2, \ldots . \tag{6}$$

The normalization factor in front of the exponential function ensures that the filter weights sum up to one. The only free parameter to be chosen by the researcher is σ^2. The frequency response function is given by the Fourier transform of $\{c_j\}$ as

$$a(\omega, \sigma^2) = \exp\left(\frac{-\omega^2\sigma^2}{2}\right), \tag{7}$$

where ω denotes the angular frequency.

An appropriate value for the free parameter σ^2 can be obtained by a similar reasoning as in the case of the binomial filter. One can show that the frequency response function is about 0.5 for fluctuations with a period of 5.3 times the standard deviation. Suppose you want to construct a Gaussian filter for which the gain function has a value of 0.5 for a period of 5. Then σ has a value of about $5/5.3 \approx 0.94$.

In principle, the Gaussian filter requires a bi-infinite time series. However, since the filter weights converge monotonically to zero, it is possible to use a finite filter as an approximation. One possibility is to use only the observations for which the filter weights are greater than a prespecified percentage value of the weight for the central observation (say, 5 %) and to rescale the used weights in order to get a sum of one.

3 The Hodrick-Prescott filter

3.1 Technical framework

The basic idea of the Hodrick-Prescott filter (Hodrick/Prescott, 1997, henceforth denoted as HP-filter) is to decompose a time series $\{y_t\}_{t=1}^T$ in two components $y_t = \mu_t + c_t$. μ_t is the trend component and c_t defines the rest, usually the sum of the cyclical component and irregular noise. The trend μ_t is estimated by solving the following minimization problem:

$$\min_{\mu_t} \sum_{t=1}^{T} (y_t - \mu_t)^2 + \lambda \sum_{t=2}^{T-1} [(\mu_{t+1} - \mu_t) - (\mu_t - \mu_{t-1})]^2. \tag{8}$$

In general the minimization problem consists of two parts. The first one aims to minimize the squared difference between μ_t and y_t, which results in a close fit of the trend to the original series. The second part minimizes the squared second differences of μ_t. These can be seen as a measure for the volatility of the trend. Clearly there is a trade off between both parts of the minimization problem. While the first part would generate a very flexible trend function, the second part penalizes the volatility of the trend. This trade off is solved by the so called penalization parameter λ, that puts weight on the second part.

Selecting a high value for λ emphasizes the second part of the minimization problem, which results in a smooth trend estimation. On the other hand, for $\lambda \to 0$, the trend is equal to the original series $\{y_t\}_{t=1}^{T}$. Thus, the basic feature of the HP-filter is the selection of the penalization parameter λ, as it completely determines the shape of the estimated trend. Figure 1 shows trend estimations for a HP-filter with different values of λ, that was applied to the HadCRUT4-data for the northern hemisphere yearly temperature averages from 1850-2012 (Morice et al., 2012):

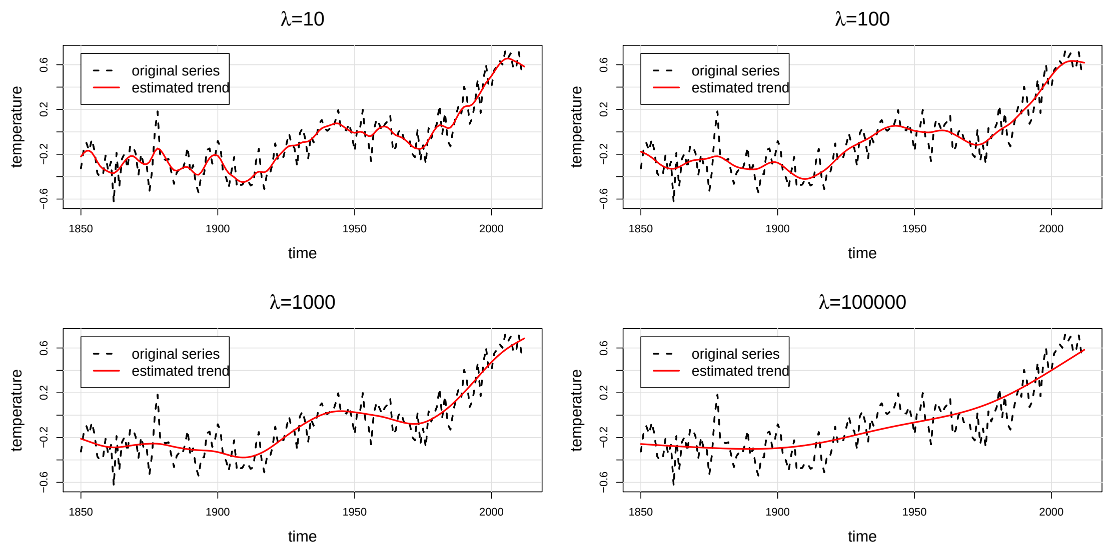

Figure 1: Trend estimations of HP-filter with different values of λ

Figure 1 shows that the volatility of the trend decreases as λ becomes higher. While it is very wiggly for $\lambda = 10$, it is smooth for $\lambda = 100000$. In order to calculate the trend by the HP-filter, it is more practical to write the filter in matrix notation. The solution of the minimization problem in (8) can be written as (Mc Elroy, 2008):

$$\hat{\boldsymbol{\mu}} = (\mathbf{I} - \lambda \boldsymbol{\Delta}'\boldsymbol{\Delta})^{-1}\mathbf{y}, \tag{9}$$

with $\hat{\boldsymbol{\mu}} = (\hat{\mu}_1, ..., \hat{\mu}_T)'$, $\mathbf{y} = (y_1, ..., y_T)'$ and $\boldsymbol{\Delta}$ as a $(T-2) \times T$ differencing matrix.

$$\boldsymbol{\Delta} = \begin{pmatrix} 1 & -2 & 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 & -2 & 1 \end{pmatrix}.$$

The product of $\boldsymbol{\Delta}$ and $\mathbf{y}$ yields the second differences of $\mathbf{y}$.

3.2 Selecting a value for λ

As pointed out in chapter 3.1, the decisive feature of the HP-filter is the selection of a value for the penalization parameter λ. There are no general rules how to choose λ, however it is possible to use spectral analysis for this selection. The basic idea of spectral analysis is to decompose a time series into oscillations with different frequencies, which is called the representation of the series in the frequency domain. Based on this decomposition, the trend can be defined by those oscillations with high periodicities, while the rest consists of oscillations with medium and low periodicities. This way it is possible to derive subjective but nevertheless precise definitions of trend and cycle.

The HP-filter as a tool for trend estimation suppresses and eliminates high frequencies, while it leaves low frequencies almost unchanged. This impact of the HP-filter on a time series in the frequency domain can be described by the gain function, which is calculated by the filter weights of the HP-filter. Given formula (9) the filter weights are contained in the hat matrix $\boldsymbol{H}(\lambda) = (\mathbf{I} - \lambda\boldsymbol{\Delta}'\boldsymbol{\Delta})^{-1} \in \mathbb{R}^{T\times T}$, so that $\hat{\boldsymbol{\mu}} = \boldsymbol{H}(\lambda)\boldsymbol{y}$. Let h_{ij} denote the j^{th} element in the i^{th} row of $\boldsymbol{H}(\lambda)$, then $\hat{\mu}_t$ is calculated as:

$$\hat{\mu}_t = \sum_{j=1}^{T} h_{tj} y_j. \tag{10}$$

Given the filter weights h_{ij} the frequency response function for an estimation $\hat{\mu}_t$ and a frequency ω can be calculated for a certain λ as (e.g. Mills, 2003):

$$a_t(\omega, \lambda) = \sum_{j=1-t}^{T-t} h_{t,j+t} e^{-ij\omega}, \tag{11}$$

from which the corresponding gain can be obtained as

$$g_t(\omega, \lambda) = \left[\left(\sum_{j=1-t}^{T-t} h_{t,j+t} \cos(\omega j) \right)^2 + \left(\sum_{j=1-t}^{T-t} h_{t,j+t} \sin(\omega j) \right)^2 \right]^{\frac{1}{2}}. \tag{12}$$

Considering the gain for all frequencies within a certain bandwidth, usually $[0, \pi]$, yields the gain function. This gain function can be used as a measure in order to select an appropriate value for λ. To this regard, first of all a definition of the trend component in the frequency domain has to be made. The definition that the trend consists of frequencies lower than a certain cut off frequency ω^{cf} can be characterized by an ideal gain function. Such an ideal gain function is shown in the left plot of Figure 2:

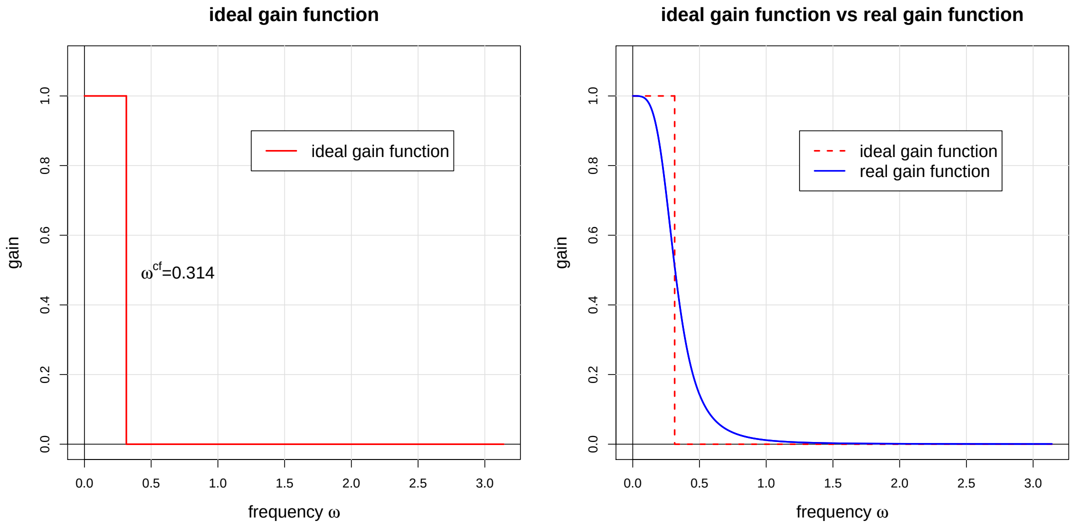

Figure 2: Ideal gain function and gain function of a HP-filter

The left plot of Figure 2 shows an ideal gain function for a cut off frequency of $\omega^{cf} = 0.314$. This cut off frequency would imply that the trend is defined by fluctuations with periods of above 20 years. Such an ideal gain function can never be achieved by a real filter, as this would require an infinite number of filter weights (Openheim/Schafer 1989). However an ideal gain function can be used as a tool in order to find an appropriate selection for the penalization of the HP-filter. To this regard the right plot of Figure 2 shows the gain function of an HP-filter with $\lambda = 100$ that is applied to a series with 163 observation (what is the same length as the HadCRUT4 northern hemisphere data set). Its gain function is compared to the ideal gain function with $\omega^{cf} = 0.314$. Clearly the real gain function cannot fully imitate the ideal one. However one can try to find an appropriate penalization by selecting λ such that the real gain function is as close as possible to the ideal one.

To this point a so called *loss* is defined, that is a measure for the deviation between the gain functions. Let $g^*(\omega)$ denote the ideal gain for frequency ω, then the *loss* for estimation $\hat{\mu}_t$ is calculated as:

$$l_t(\lambda) = \int_0^{\pi} [g^*(\omega) - g_t(\omega, \lambda)]^2 d\omega. \tag{13}$$

Thus, the *loss* is the squared deviation between the ideal and the real gain function within the interval $[0, \pi]$. Now λ can be selected such, that $l_t(\lambda)$ is minimized, i.e. the HP-filter yields the best approximation of the ideal gain function. Note that similar approaches have been made by Baxter/King (1999) for the so called Baxter-King filter and by Tödter (2002) for the HP-filter. Minimizing (13) is numerical complicated for continuous frequencies, however it can be easily approximated by a sufficient high number of discrete frequencies. Let $\boldsymbol{\omega} \in \mathbb{R}^{k\times 1}$ be a vector of discrete frequencies, e.g. $\boldsymbol{\omega} = (0, 0.001, 0.002, ..., \pi)'$, then (13) can be approximated by:

$$l_t(\lambda) = \sum_{j=1}^{k} [g^*(\omega_j) - g_t(\omega_j, \lambda)]^2 \cdot \delta. \tag{14}$$

k is the number of elements in $\boldsymbol{\omega}$, and δ defines the distance between the elements of $\boldsymbol{\omega}$, i.e. $\delta = \omega_j - \omega_{j-1}$. The minimization of (14) with regard to λ can be done by methods like fisher scoring or Newton-Raphson. As $l_t(\lambda)$ is minimized with respect to one parameter only, also a grid search is suitable.

Formula (14) is used to derive the optimal penalization parameter for series with 163 observations. Hereby the optimal λ's for a set of cut off periodicities between ten and 50 years are calculated, where the *loss* is minimized for the medium (82^{th}) estimation. The resulting values for λ are shown in Table 1. Moreover, in order to show the performance of the HP-filter in relation to the Gaussian filter, Table 1 compares the *loss* of the HP-filter with those of the Gaussian filter for different values of n.
For the Gaussian filter σ is selected such, that its gain has a value of 0.5 for the given cut off periodicities. This makes the Gaussian filter directly comparable to the HP-filters that minimize $l_t(\lambda)$. That is because two gain functions are approximately as equal as possible, when they have a gain of 0.5 for the same frequency (Harvey/Trimbur 2008). Thus, for a given n, the Gaussian filters considered here yield gain functions that can be seen as approximations of ideal gain functions with a certain cut off periodicity.

Table 1: optimal values of λ and *loss* for different cut off periodicities

periodicity	$\boldsymbol{\lambda}$	***loss* HP-filter**	***loss* Gauss ($n = 10$)**	***loss* Gauss ($n = 20$)**
10	9	0.0635	0.0828	0.0828
20	127	0.0307	0.0408	0.0414
30	637	0.0204	0.0303	0.0276
40	1984	0.0153	0.0364	0.0203
50	4756	0.0122	0.0475	0.0164

As Table 1 shows, λ increases with a higher cut off periodicity. While for a cut off periodicity of ten years $\lambda = 9$ is sufficient, for a cut off periodicity of 50 years λ has to be set to 4756. Comparing the *loss* of the HP-filter and the Gaussian filter, the HP-filter is clearly superior with regard to approximating an ideal gain function. Even if a relative high loss of 20 observations at each margin is accepted, the *loss* of the Gaussian filter is still around 25 percent higher than the one of the HP-filter. In order to describe the differences between the two filters, Figure 3 plots the gain functions of the HP-filter and the Gaussian filter with $n = 20$ for the cut off periodicities in Table 1.

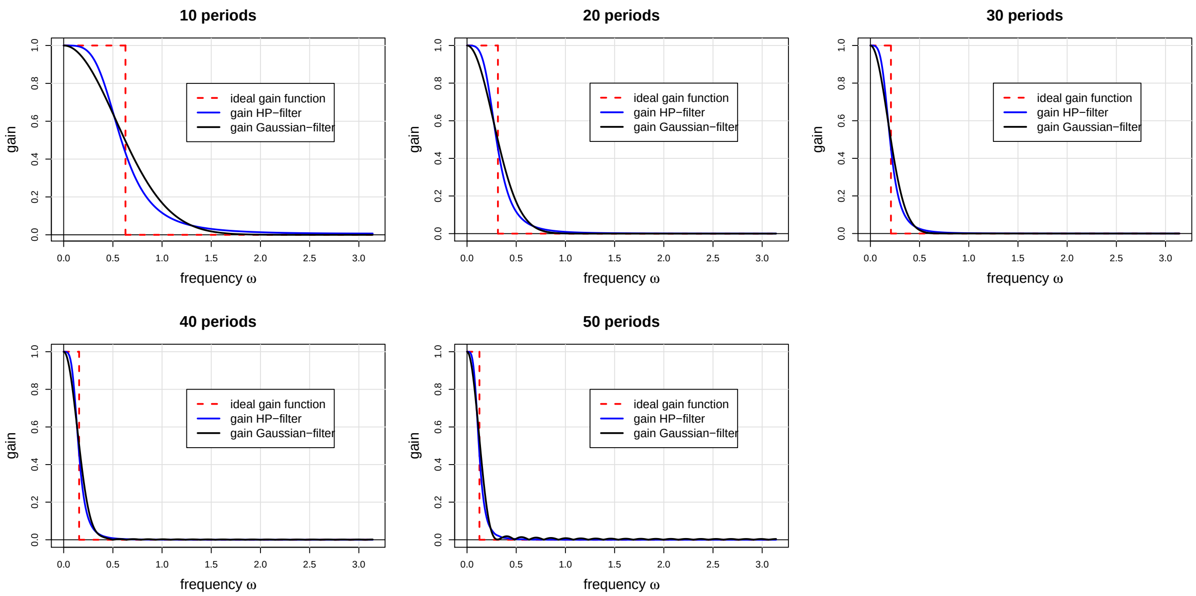

Figure 3: Gain functions of HP-filter and Gaussian filter for different cut off periods

Clearly the gain function of the HP-filter stays longer at a value close to one for low frequencies but then decreases more rapidly to zero than the gain function of the Gaussian filter. As a consequence the HP-filter is more suitable for approximating an ideal gain function than the Gaussian filter. Moreover, at a cut off periodicity of 50, the gain function of the Gaussian filter oscillates for high frequencies. Thus, the HP-filter is preferable not only as it is able to generate estimations at the margins of the series but also as it offers a much more accurate gain function. It should be taken into consideration that Table 1 displays the optimal values for the estimation in the middle of the series. For other estimations this might not be the optimal penalization, as the filter weights are not the same for every estimation. Compared to the middle of the time series the weight structure especially changes at the margins. How to account for a changing filter weight structure is explained in the next sections.

3.3 Changing filter weight structures

A clear advantage of the HP-filter compared to the Baxter-King filter (Baxter/King, 1999) or to the Gauss-filter is that it can render trend estimations for the most recent periods. However, this feature is paid by an increasing asymmetry of the filter weights at the margins, inducing an rising excess variability for these estimations. In this regard it is useful to consider the HP-filter in the frequency domain, as spectral analysis allows to describe and quantify the excess variability. As mentioned, the filter weights strongly change their structure to the margins of the series, while they have a similar, almost symmetric shape around the data middle. As an example this is shown in Figure 4, which displays the filter weights of a HP-filter with $\lambda = 1984$, that is the optimal value for a cut off periodicity of 40 years and 163 observations (see Table 1).

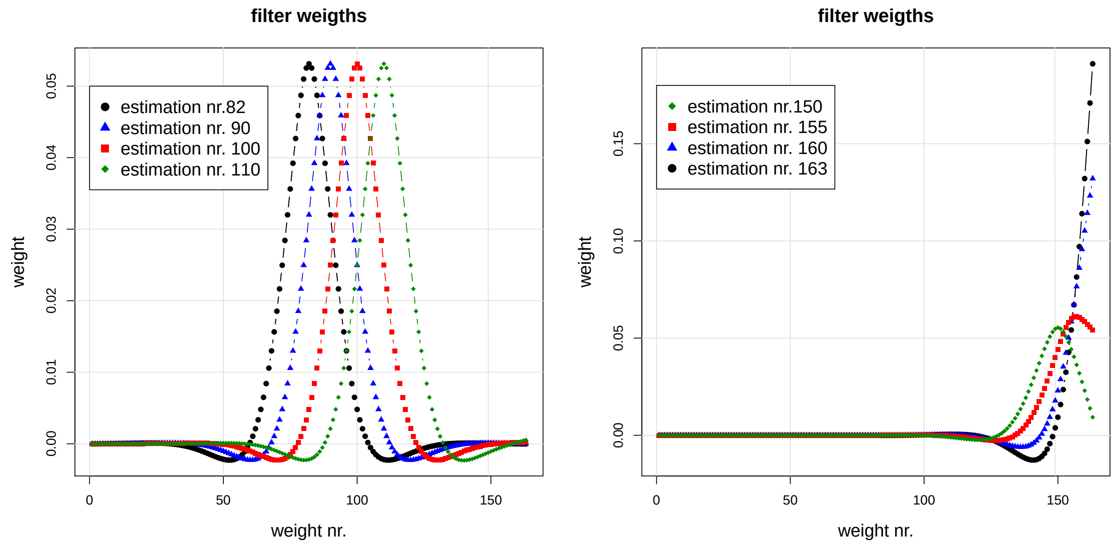

Figure 4: Filter weights of HP-filter for different time periods (163 observations)

The left plot of Figure 4 shows the filter weights for estimations near the middle of the data. Clearly they have an almost equal structure. On the other hand the right plot displays the filter weights for estimations close the end of the series. With an decreasing distance to the margin the structure changes and the symmetry completely disappears. This change of the filter weight structure affects the gain function of the HP-filter. To demonstrate this, Figure 5 shows the gain function for different periods.

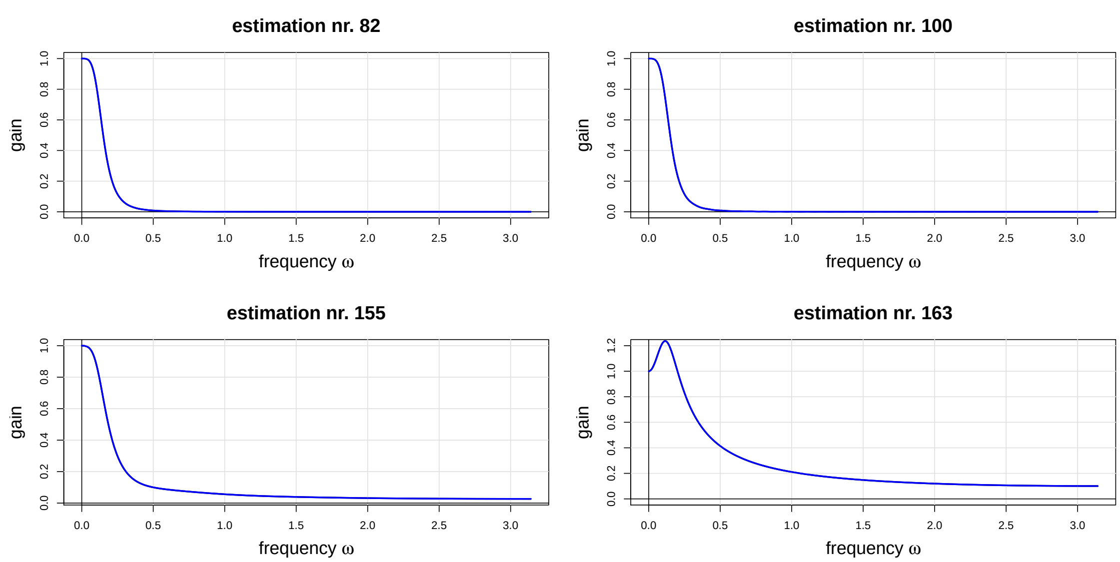

Figure 5: Gain functions of HP-filter for different time periods (163 observations)

For the 82^{th} and 100^{th} estimation, the gain functions are almost identical. For both time periods, high frequencies are completely suppressed. This is different for the 155^{th} and

163^{th} estimation. Especially for the last estimation high frequencies are not completely eliminated any more, which results in increased volatility of the trend function. However, on purpose to describe the excess variability at the margins, it is not practicable to use the gain function as one can hardly consider it for all estimations. A much more appropriate measure is the $loss\ function$. The value of the $loss$ can easily be depicted for all estimations in one graph, which allows to evaluate the rising volatility at the margins. To this point Figure 6 shows the $loss\ functions$ for a HP-filter with different cut off periodicities that is applied to a series with 163 observations.

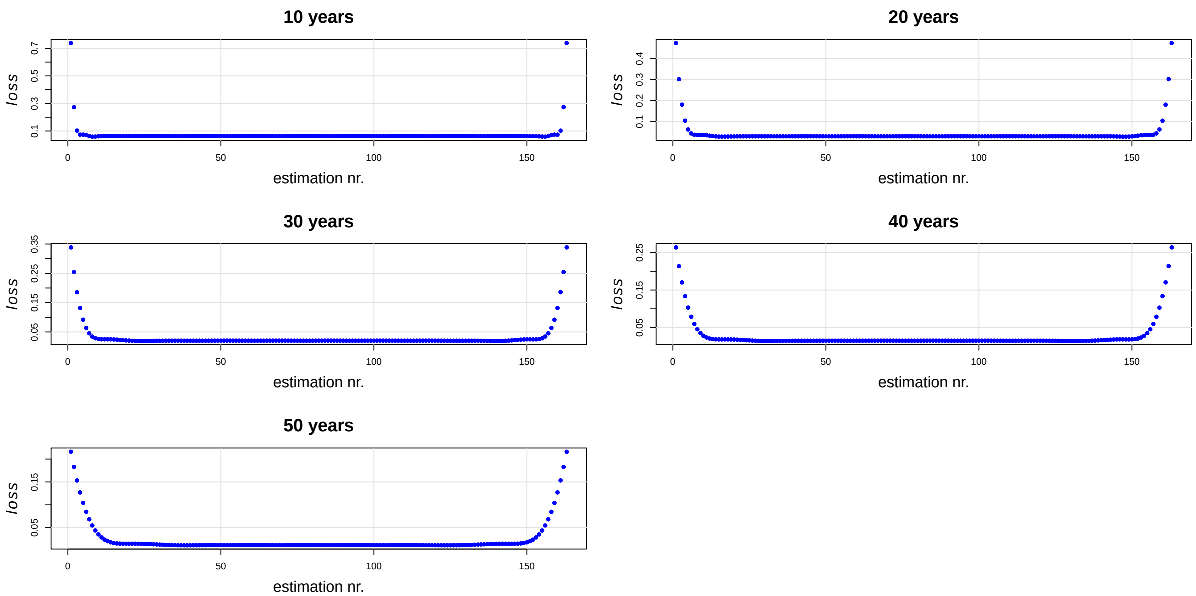

Figure 6: $loss\ functions$ for different cut off periodicities

Obviously, the $loss$ is rather low and almost equal for the estimations around the data middle. As one approaches the margins of the series, the $loss$ strongly increases which describes the increasing excess variability for those estimations. Furthermore, as figure 6 shows, the number of estimations that is affected by the excess variability, depends on the value of λ. It increases with higher values of λ. For $\lambda = 9$ only about the first and last five estimations are affected, for $\lambda = 4756$ the $loss$ is increased for about more than the first and last 25 estimations. While the selection of λ affects the excess variability, this is not true for the data length. Studies show that the number of estimations that show an increased $loss$ is independent of the length of the time series (e.g. Blöchl, 2013).

3.4 Introducing a flexible penalization

Section 3.3 showed that the $loss\ function$ can be used as a measure to describe the excess variability at the margins. In order to reduce the excess variability now a flexible penalization of the HP-filter is introduced (a flexible penalization was already suggested by Razzak/Richard (1995) in order to account for breaks in the data). As Figure 6 showed the variability increases to the margins. This variability can be compensated by a flexible penalization that sets higher values for λ at the margins. Such a changing penalization can be introduced by changing the model framework slightly. Given formula (9), a flexible penalization can be implemented by replacing the scalar λ by a vector $\boldsymbol{\lambda} \in \mathbb{R}^{T-2\times 1}$, where $\boldsymbol{\lambda} = (\lambda_1, ..., \lambda_{T-2})'$ and constructing the matrix $\boldsymbol{K} = \text{diag}(\boldsymbol{\lambda})$. Then the HP-filter can be written to

$$\hat{\boldsymbol{\mu}} = (\mathbf{I} - \boldsymbol{\Delta}'\boldsymbol{K}\boldsymbol{\Delta})^{-1}\mathbf{y}. \tag{15}$$

The HP-filter can be interpreted as a continuous connection of lines that change their slope at the points in time $t = 2, 3, ..., T-1$. The parameters $\lambda_1, ..., \lambda_{T-2}$ regulate to what degree the slope of the trend can change at these points in time, i.e. λ_t determines how strongly the trend can change its slope at the point in time $t+1$. Setting the first and last few λ's higher thus results in a smoother trend at the margins, which reduces the excess variability.

An important question is how fast the value of λ shall rise to the margins. Figure 3 suggests, that a linear increase might be suitable. Furthermore a criterion is needed in order to determine the pace of the rise of the penalization. This criterion should secure, that the flexible penalization induces a decrease of the *loss* at the margins without affecting the one around the data middle. To this regard it turns out that a appropriate criterion is to minimize the cumulative *loss* for all estimations $\sum_{t=1}^{T} l_t(\boldsymbol{\lambda})$. This sum is minimized with respect to the parameter vector $\boldsymbol{\lambda}$, where the first and last few values of the penalization rise linearly. Thus the last m values of the penalization can be described by a linear function:

$$\lambda_{T-2-m+j} = \alpha + \beta j, \quad \text{where } j = 1, ..., m. \tag{16}$$

The first m values of the penalization are defined analogously, i.e. $\lambda_1 = \lambda_{T-2}, \lambda_2 = \lambda_{T-3}, ..., \lambda_m = \lambda_{T-1-m}$. The intercept α can be seen as given with regard to the cut off periodicity, as the penalization shall not rise for estimations around the middle. Consequently the penalization is completely determined by β. The cumulative *loss* is then given to:

$$L(\boldsymbol{\lambda}(\beta)) = \sum_{t=1}^{T} l_t(\boldsymbol{\lambda}(\beta)), \tag{17}$$

$$\text{where } \; l_t(\boldsymbol{\lambda}(\beta)) = \sum_{j=1}^{k} [g^*(\omega_j) - g_t(\omega_j, \boldsymbol{\lambda}(\beta))]^2 \cdot \delta. \tag{18}$$

For a given m, $L(\boldsymbol{\lambda}(\beta))$ can be minimized with respect to β using minimization algorithms like fisher scoring or Newton-Raphson. m cannot be included into the minimization as it is an integer. Thus the minimization has to be done for different values of m, where this value is chosen, that leads to the lowest cumulative *loss*. Performing this minimization for the different cut off periodicities yielded the following results for m and β for series with the same number of observations as the HadCRUT4 data set.

Table 2: optimal values for m and β

periodicity	$\boldsymbol{\alpha}$	$\boldsymbol{\beta}$	$\boldsymbol{m}$
10	9	14.49	6
20	127	137.22	13
30	637	490.81	20
40	1984	1180.79	27
50	4756	2283.44	34

The slope β clearly rises, when higher cut off periodicities and thus higher values of λ are chosen. Furthermore m increases for higher cut off periodicities. This is in line with Figure 3 which shows, that the number of estimations affected by the excess variability rises, when λ is increased. To show the effect of this flexible penalization, Table 3 summarizes the $loss$ in the data middle, at the margin and the cumulative $loss$ for the fixed and the flexible penalization.

Table 3: $loss$ for fixed and flexible penalization

periodicity	penalization	$l(82, \boldsymbol{\lambda})$	$l(163, \boldsymbol{\lambda})$	$L(\boldsymbol{\lambda})$
10	fixed	0.0635	0.7381	12.2269
	flexible	0.0635	0.3775	11.6428
20	fixed	0.0307	0.4731	7.0226
	flexible	0.0307	0.2184	6.3586
30	fixed	0.0204	0.3385	5.3499
	flexible	0.0204	0.1524	4.6803
40	fixed	0.0153	0.2635	4.5286
	flexible	0.0153	0.1170	3.8562
50	fixed	0.0122	0.2160	4.0401
	flexible	0.0125	0.0951	3.3664

As Table 3 shows, the flexible penalization could reduce the $loss$ for the last estimation, while it hardly affected the estimation in the middle of the series (82^{th} estimation). Moreover the cumulative $loss$ could be reduced in every of the five cases. The decrease of $L(\boldsymbol{\lambda})$ was between five and 17 percent, for the $loss$ of the last estimation even around 50 percent. To get a complete overview of how the flexible penalization affects the results, Figure 7 compares the $loss$ of the fixed and flexible penalization for all estimations and cut off periodicities.

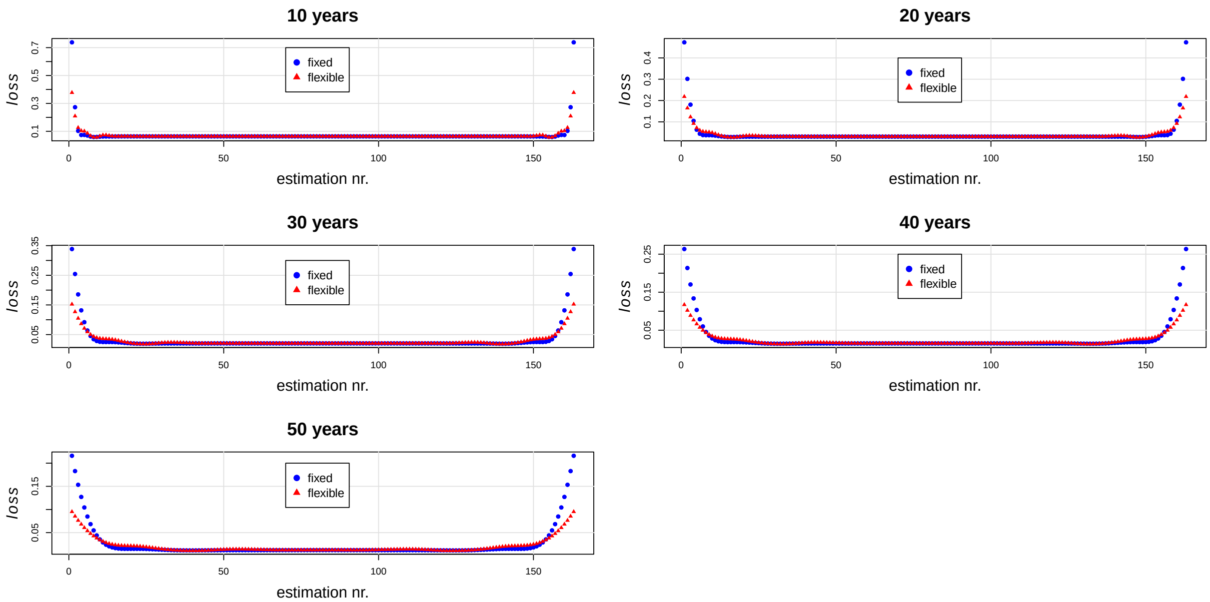

Figure 7: $loss\ functions$ for fixed and flexible penalization

Clearly the $loss$ was reduced at the margins of the time series in every case. Furthermore, for most estimations around the data middle, the results were hardly affected by the flexible penalization. Only for few estimations the $loss$ was increased slightly. Considering Figure 7 and the fact that $L(\boldsymbol{\lambda})$ declined for every cut off periodicity, it turns out that the criterion of minimizing $L(\boldsymbol{\lambda})$ by a linear increase of the penalization to the margins is appropriate to reduce the excess variability without worsening the results for estimations closer to the middle of the series (a simpler approach to get a handle on the excess variability by flexible penalization was already made by Bruchez (2003)).

4 Empirical application

The HP-filter with the flexible penalization is now used to estimate the trend of the HadCRUT4 data (Morice et al., 2012) for the northern hemisphere annual average temperature. The trend component is calculated for each of the cut off periodicities of ten to 50 years, where the penalizations derived in section 3.4 are used. The results are compared to the "standard case" of a fixed penalization in order to point out the different implications of the flexible penalization. First of all the trend component with a ten year cut off period is estimated. The results are shown in Figure 8.

The left plot of Figure 8 shows the trend with a ten years cut off period for fixed and flexible penalization. There are clear differences between both at the beginning of the series, while the deviation is rather small at the end (A detailed graph is added in the appendix). However, if one considers the right plot that shows the first differences of both trends, also deviations at the end of the data can be observed. Both trend functions decrease since 2006, but the trend according to the flexible penalization decreases slower since 2010.

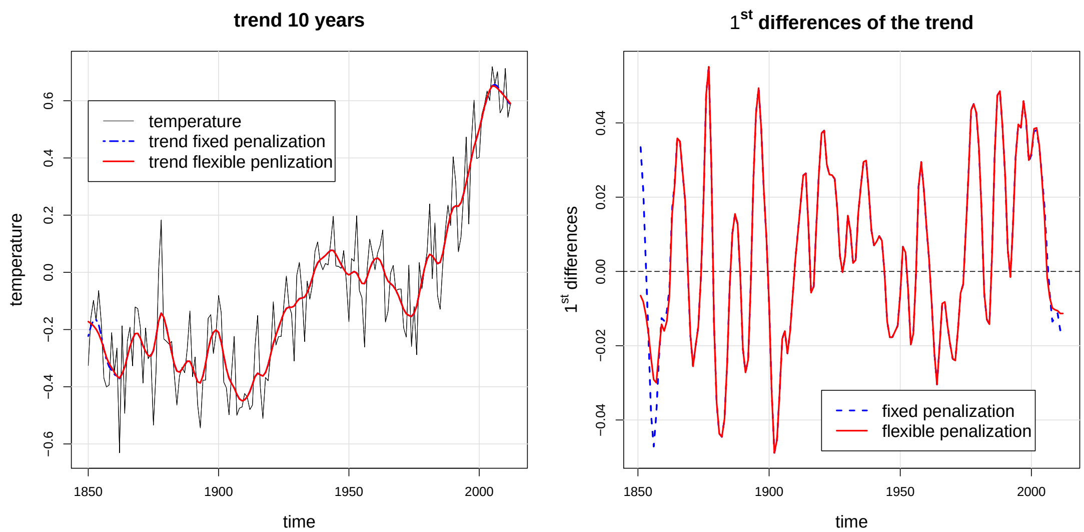

Figure 8: Trend for ten years cut off period and first differences for fixed and flexible penalization

The right plot shows that, irrespective of fixed and flexible penalization the smoothed values do not represent a "trend". The generated differences exhibit pronounced cyclical features and do not represent the long-run development. In many cases we observe decreases of the first differences, even in the period 1970 to 2000 for which a trend increase is generally accepted.

Next, the trend with a cut off period of 20 years is estimated. Figure 9 shows the trend estimations for the fixed and flexible penalization as well as their first differences.

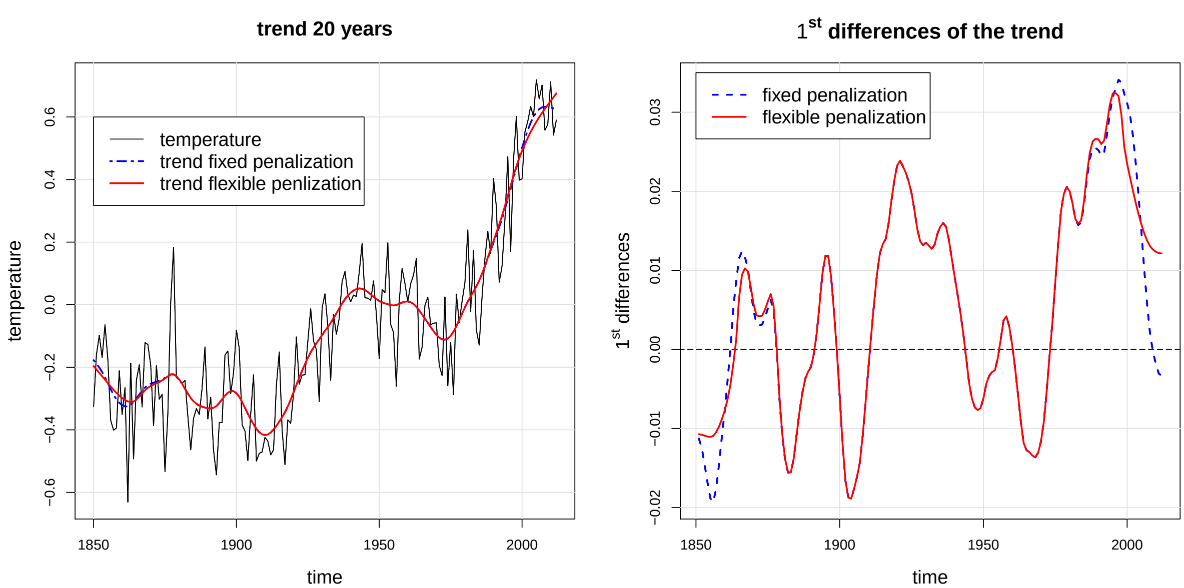

Figure 9: Trend for 20 years cut off period and first differences for fixed and flexible penalization

For the trend with a cut off period of 20 years significant differences between both penalizations at the margins can be observed. The fixed penalization approach yields a trend that starts to decrease from the year 2006, while the trend for the flexible penalization also rises from 2006 onwards. Thus the flexible penalization corrected the trend estimation, inducing a completely changed development of the long run temperature compared to the case of fixed penalization. However, its growth has also slowed down since 1996, which can be seen at the first differences on the right plot of Figure 9.

The trend with a cut off period of 30 years and the first differences are shown in Figure 10:

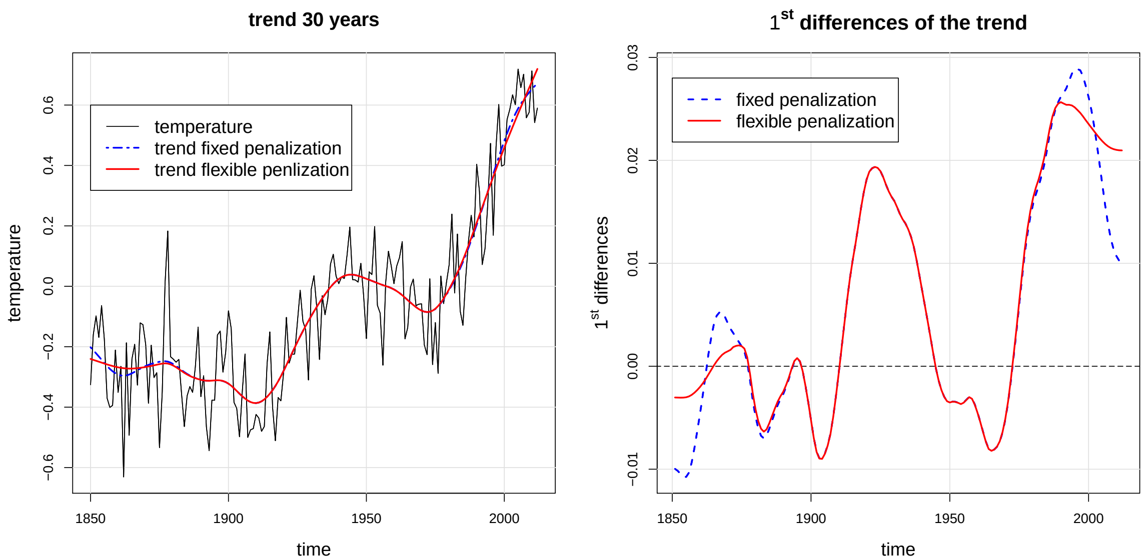

Figure 10: Trend for 30 years cut off period and first differences for fixed and flexible penalization

Also for the trend with a cut off period of 30 years, there are obvious differences between fixed and flexible penalization. Both trends show an increasing behaviour for the most recent years. But the trend growth rate according to fixed penalization strongly decreased since 1996, while there was only a slight decrease for the trend using the flexible penalization.

Next, the trend with a cut off period of 40 years is considered. The different trend estimations and the first differences are shown in Figure 11. For the trend with a cut off period of 40 years the differences between fixed and flexible penalization at the end of the series are not as obvious as for those with cut off periods of 20 and 30 years. However, one can see that the trend according to the flexible penalization increased more rapidly during the last three or four years. This becomes obvious by considering the first differences. The trend growth started to decrease during the last ten to 15 years. Hereby the trend growth of the fixed penalization approach decreased strongly since 1996, while there is just a slight decrease for the flexible penalization approach.

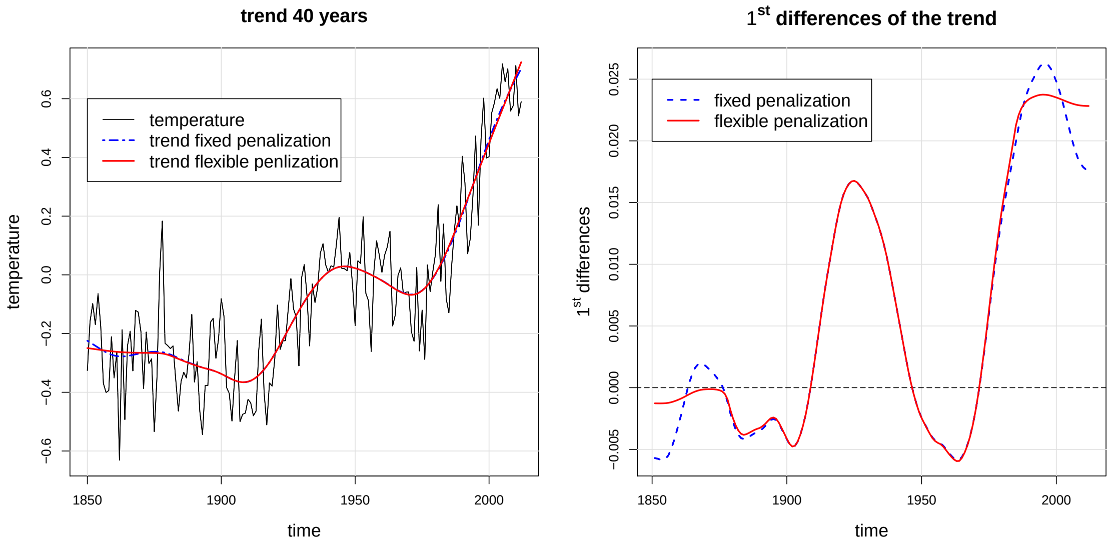

Figure 11: Trend for 40 years cut off period and first differences for fixed and flexible penalization

Finally the trend with a cut off period of 50 years is estimated, that is shown in Figure 12.

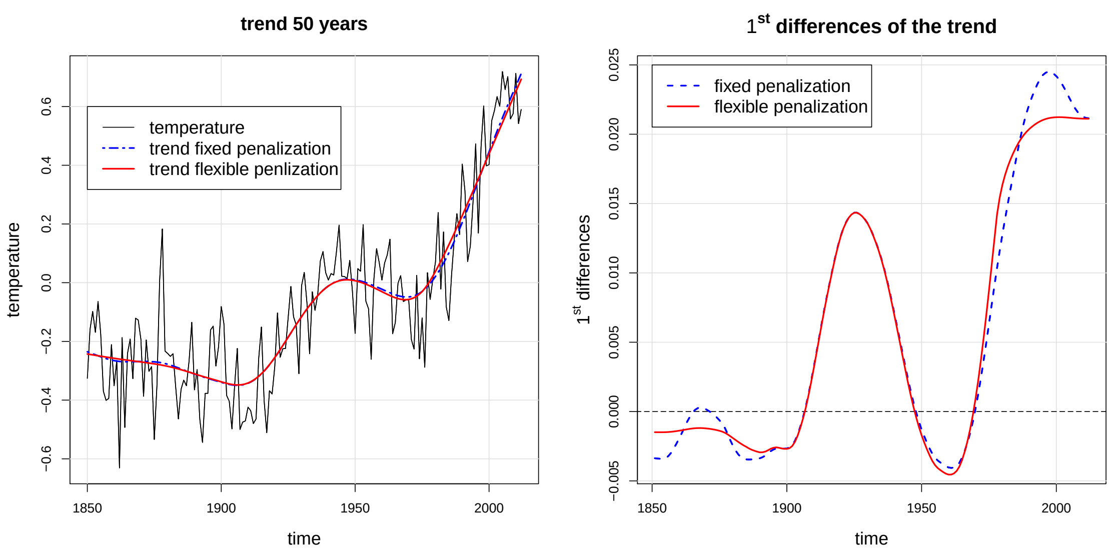

Figure 12: Trend for 50 years cut off period and first differences for fixed and flexible penalization

There seem to be no significant differences between the estimations. Both trend functions show an almost linear increase since about 1970, while the trend according to the fixed penalization even grows slightly faster during the most recent periods. Differences become obvious, when the right plot of Figure 12 is considered. From the year 1998 onwards the trend growth for the fixed penalization approach decreased strongly. On the other hand, the growth of the trend according to the flexible penalization remained almost unchanged during the most recent periods.

Considering the estimations for the trend functions with cut off periods between 20 and 50 years, it can be seen that the "standard approach" of the fixed penalization would imply, that the growth of the trend temperature has strongly decreased during the past ten to 15 years, or even that the trend temperature is already declining. The picture is completely different, when the excess variability at the margins is reduced by flexible penalization. This correction changes the implications of the trend estimations significantly. According to the results of the flexible penalization, the decrease of the trend growth rate is much lower. For cut off periodicities of 30 years or above, it seems to have only stabilized on a high level. Thus, the correction of the excess variability contradicts the thesis of a turning point in the northern hemisphere temperature. At least it shows, that more time is needed to clearly identify the current situation as the beginning of a downturn of the trend temperature.

4.1 Summary and conclusion

It is a great challenge to get reasonable and reliable estimates of the long-run trend component of a time series at the end of the data sample. An illustrative example is the current fierce discussion concerning the correct interpretation of the global temperature data for the last ten years. For periods far enough away from the end of the observed time series one uses past, present and future values for the estimation of the trend (symmetric filter). However, in carrying out a real-time estimation at the end of the sample one can only use asymmetric filters. Asymmetric filters induce phase shifts and other distorted features (in comparison to symmetric filters). For instance, for asymmetric low pass filters the maximum of the gain function is generally not located at frequency zero but at a positive frequency what might induce an amplification of cycles with periods not considered to be part of the trend (see as an example the effects for the HP filter in Figure 5). In addition, fluctuations with high frequencies (noise) are not totally removed.

In this paper we propose the use of a filter known in economics as the Hodrick-Prescott filter for the estimation of the trend component of global temperature data. This filter is based on a well defined optimization problem. It requires the pre-specification of a penalization parameter for the second differences of the trend component which governs the degree of smoothness of the generated trend. We derive suitable values for the penalization parameter by interpreting the Hodrick-Prescott filter as an approximation to an ideal filter and compare the results with the binomial and the Gaussian filter. An ideal low-pass filter lets pass all fluctuations with frequencies below a given cut-off value without any modifications and removes all fluctuations with frequencies greater than the cut-off value. We consider cut-off frequencies for trends which comprise waves with periods greater than ten, 20, 30, 40 and 50 years, respectively. In all cases the Hodrick-Prescott filter is closer to an ideal filter than the binomial or the Gaussian filter.

In the middle of the time series the Hodrick-Prescott filter is symmetric, for periods more at the border it gets more and more asymmetric. In order to mitigate the mentioned negative effects of asymmetric filters we allow for time varying values of the penalization parameter.

To be more precise, we minimize a loss function based on the squared differences between the gain functions of an ideal and the Hodrick-Prescott filter, respectively. As the result we get increasing values for the penalization parameter towards the end of the time series.

In the second part of the paper we apply the proposed method for the estimation of the trend component of yearly temperature data for the northern hemisphere from 1850 to 2012. For cut-off periods of 10 and 20 years we get estimated trends for which the first differences exhibit clearly strong cyclical features: The generated empirical trend component is not what we would expect for a "true" trend. For cut-off periods of 30, 40 and 50 years the estimated trends and their first differences show no cyclical behaviour and look similar.

For the following interpretation we concentrate on the trend with a cut-off period of 30 years (Figure 10). In case of the fixed penalization the value for the penalization parameter λ is 637. In case of the flexible penalization only the observations in the middle of the time series get the same weight. The first and the last 20 observations get higher values (for example, for the first and the last observation the value of the penalization parameter is about 10 500).

For the northern hemisphere temperature data the trend component decreased slightly between 1850 and 1910 and then increased until the end of the forties of the last century. After a slight decrease in the following 20 years we estimate a strong increase since about the year 1970. The maximum of the increase in temperature (not in the level) occurred in the second half of the nineties with a rise of 0.03 degrees per year for the fixed penalization and 0.025 degrees for the flexible penalization. During about the last 15 years for both penalization methods the pace of warming decreased somewhat but remained positive. For our preferred flexible penalization method the estimated increase in the year 2012 is 0.02 degrees per year. When we apply the flexible penalization method to trends with cut-off periods of 40 and 50 years we estimate for the last 15 year a roughly constant rise in temperature of somewhat more than 0.02 degrees per year. In summary: Using data until 2012 there is no indication of a change in the sign of the trend direction. The trend component is still increasing, only the pace of the increase is somewhat lower but is still positive.

A Appendix

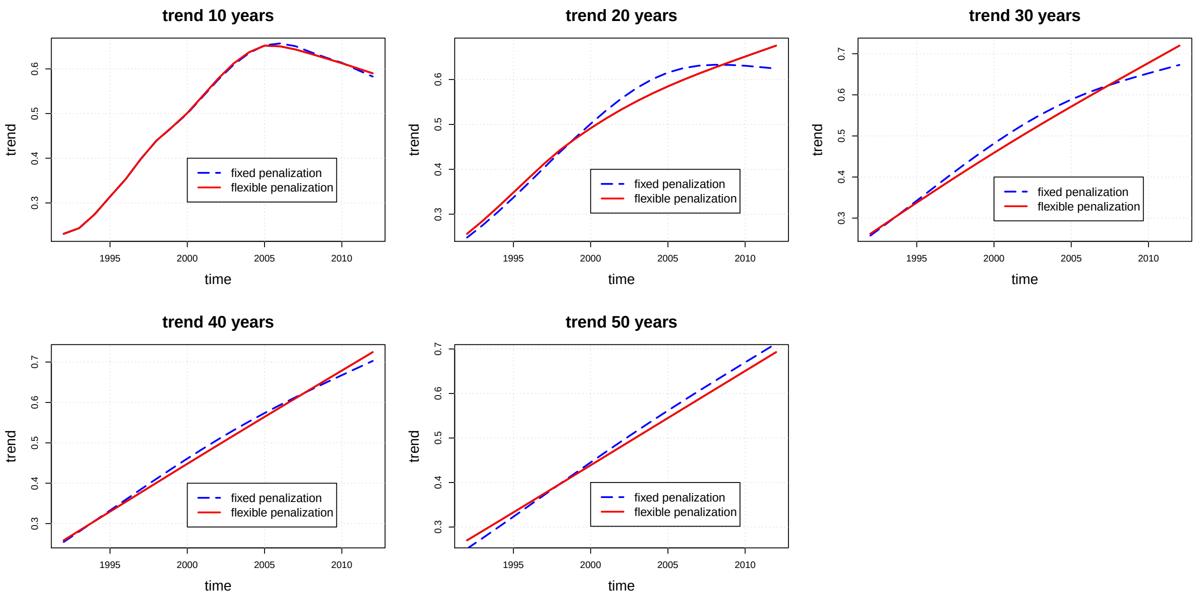

Figure 13: Estimated trend for fixed and flexible penalization from 1992-2012

B References

Baxter, M., King, R. (1999): *"Measuring Business Cycles: Approximate Band-Pass Filters for Economic Time Series,"* The Review of Economics and Statistics 81, 575-593.

Blöchl, A. (2013): *"Reducing the Excess Variability of the Hodrick-Prescott Filter by Flexible Penalization,"* Working Paper.

Bruchez, P. A., (2003): *"A Modification of the HP-filter - Aiming at Reducing the End-Point Bias,"* Swiss Federal Finance Administration, Working Paper.

Gourieroux, C., Monfort, A. (1997): *"Time Series and Dynamic Models,"* Cambridge University Press.

Harvey, A., Trimbur, T. (2008): *"Trend Estimation and the Hodrick-Prescott Filter,"* Journal of the Japan Statistical Society, 38, 41-49.

Henderson, R. (1924): *"On a new method of graduation,"* Transactions of the Actuarial Society of America, 25, 29-40.

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

Kaufmann, R. K., Kauppi, H., Mann, M. L., Stock, J. H. (2011): *"Reconciling anthropogenic climate change with observed temperature 1998-2008,"* Proceedings of the National Academy of Sciences, USA 108(29), 11 790-11 793.

Kupce, E. (1996): *"Binomial Filters,"* in: D.N. Rutledge (ed.), Signal Treatment and Signal Analysis in NMR. Elsevier.

Leser, C. E. V. (1961): *"A simple method of trend construction,"* Journal of the Royal Statistical Society, Series (B), 23, 91-107.

McElroy, T. (2008): *"Matrix Formulas for Nonstationary ARIMA Signal Extraction,"* Econometric Theory 24, 988-1009.

Meehl, G. A., Arblaster, J. M., Fasullo, J. T., Hu, A., Trenberth, K. E. (2011): *"Model-based evidence of deep-ocean heat uptake during surface-temperature hiatus periods,"* Nature Climate Change, (1)7, 360-364.

Mills, T. C. (2003): *"Modelling Trends and Cycles in Economic Time Series,"* Palgrave Macmillan. Houndmills/Basingstoke/Hampshire/New York.

Mills, T. C. (2006): *"Modelling Current Trends in Northern Hemisphere Temperatures,"* International Journal of Climatology, 26, 867-884.

Mills, T. C. (2009): *"Modelling Current Temperature Trends,"* Journal of Data Science, 7, 89-97.

Mitchell, J., M. et al. (1966): *"Climate Change,"* Technical Note No. 79, World Meteorological Organisation.

Morice, C. P., Kennedy, J. J., Rayner, N. A., Jones, P. D. (20012), *"Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 dataset,"* Journal of Geophysical Research, 117, D08101, doi:10.1029/2011JD017187.

Oppenheim, A. V., Schafer, R. W. (1989): *"Discrete-Time Signal Processing,"* Prentice-Hall, Englewood Cliffs, New Jersey.

Osborn, D. (1995): *"Moving Average Detrending and the Analysis of Business Cycles,"* Oxford Bulletin of Economics and Statistics, 57, 547-558.

Panofsky, H., A., Brier, G., W. (1958): *"Some applications of statistics to meteorology,"* University Park, Pa. : Pennsylvania State University Press.

Razzak, W., Richard, D., (1995): *"Estimates of New Zealand's Output Gap using the Hodrick Prescott filter with an non-constant Smoothing Parameter,"* Reserve Bank of New Zealand, Discussion Paper G95/8.

Tödter, K. H. (2002): *"Exponential Smoothing as an Alternative to the Hodrick-Prescott Filter,"* In: Klein, I., Mittnik, S. (eds.) Contributions to Modern Econometrics-From Data Analysis to Economic Policy. Boston: Kluwer Academic Publishers, S. 223-237.

Whittaker, E. T. (1923): *"On a new method of graduation,"* Proceedings of the Edinburgh Mathematical Society, 41, 63-75.