LMU
LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

VOLKSWIRTSCHAFTLICHE FAKULTÄT

Andreas Blöchl:

Penalized Splines as Frequency Selective Filters - Reducing the Excess Variability at the Margins

Munich Discussion Paper No. 2014-3

Department of Economics
University of Munich

Volkswirtschaftliche Fakultät
Ludwig-Maximilians-Universität München

Penalized Splines as Frequency Selective Filters - Reducing the Excess Variability at the Margins

Andreas Blöchl

April 2014

Abstract

Penalized splines have become a popular tool to model the trend component in economic time series. The outcome of the spline predominantly depends on the choice of a penalization parameter that controls the smoothness of the trend. This paper derives the penalization of splines by frequency domain aspects and points out their link to rational square wave filters. As a novel contribution this paper focuses on the so called excess variability at the margins that describes the undesired increasing variability of the trend estimation to the ends of the series. It will be shown that the too high volatility at the margins can be reduced considerably by a time varying penalization, which yields more reliable estimations for the most recent periods.

Keywords: excess variability, penalized splines, spectral analysis, time varying penalization, trends

Andreas Blöchl
Universitiy of Munich
Schackstraße 4
80539 Munich
Andreas.Bloechl@lrz.uni-muenchen.de

1 Introduction

An important and fundamental challenge in economics, especially in business cycle research, is a reasonable decomposition of time series into trend and cycle. Meanwhile a wide range of instruments in order to estimate these components exists, where penalized splines (O'Sullivan, 1986; Eilers/Marx, 1996) are among the most popular tools. There are strong similarities between penalized splines and the Hodrick-Prescott filter (Hodrick/Prescott, 1997), which might be the most widespread instrument for trend estimation in economics and Paige (2010) shows that the Hodrick-Prescott filter indeed is a special case of a penalized spline. The decisive feature of these instruments is that the estimated trend predominantly depends on the choice of a single penalization parameter λ that determines the smoothness of the trend.

In most applications of the Hodrick-Prescott filter λ is set to 1600 for quarterly data. This traces back to Hodrick/Prescott (1997), who derived this value by interpreting the filter as an optimal Wiener-Kolmogorov filter (Whittle, 1983; Bell, 1984). As this derivation is based on rather unrealistic assumptions the choice of $\lambda = 1600$ is often criticized as dubious (Danthine/Giradin, 1989). Furthermore, it is criticized as too low (McCallum, 2000; Flaig, 2012) and not data driven (e.g. Schlicht, 2005; Kauermann et al., 2011). To this point data driven methods like generalized cross validation (Hastie/Tibshirani, 1990) and the incorporation of the Hodrick-Prescott filter or penalized splines into a linear mixed model help to overcome this problem. Generalized cross validation induces a too wiggly trend estimation for most series with correlated errors (Diggle/Hutchinson, 1989; Altman, 1990; Hart, 1991), but it can be extended to account for a correlated residual structure (Kohn et al., 1992; Wang, 1998). Nevertheless, Oppsomer et al. (2001), Proietti (2005) and also Dagum/Giannerini (2006) demonstrated that this technique is very sensitive to the assumptions about the residual correlation structure. To this point the mixed model approach is advantageous, as it is relatively robust with regard to a misspecification of the residual correlation structure (Krivobokova/Kauermann, 2007).

In economics the separation of trend and cycle is often motivated by the conception that these components are characterized by distinguishable spectral properties. In this sense the trend as the long run development of the series is described by fluctuations with high periodicities and the cyclical component by medium and low periodicities. This allows defining trend and cycle by bandwidths of frequencies. A very common definition traces back to Burns/Mitchell (1946), who describe the cycle by periodicities between six and 32 quarters. From this point of view the penalization can be selected such that the filters mainly extract the desired frequencies. Such a method was demonstrated by Tödter (2002) for the Hodrick-Prescott filter. This paper will show a related approach for penalized splines, where splines based on a truncated polynomial basis are considered. This type of splines is interesting, as it is closely related to the Hodrick-Prescott filter. Moreover, Proietti (2007) describes the link between these splines and square wave filters (Pollock, 2000, 2003).

The extraction of frequency bands by linear filters like penalized splines exhibits unsolved problems. One massive problem is that linear filters loose the ability to suppress high frequencies for estimations at the ends of the series. This is due to the increasing asymmetry of the filter weights for estimations at the margins and leads to an undesirable increase of the volatility of the estimations for the first and last periods. This increasing volatility at the margins is often called excess variability. Especially as researchers are predominantly interested in the trend of the most recent periods this excess variability turns out to be a serious problem, since it heavily affects the reliability of the estimations at the ends of the series. An existing method to overcome this problem is to attach forecasts to the end of the series. However, as it is also shown in this paper, a high number of forecasts is required so that this approach is of limited practicability. This paper describes another approach to get a handle on the excess variability. It is shown that the excess variability can be reduced considerably by a time varying penalization, where the penalization is allowed to increase to the margins. A time varying penalization was already suggested by Razzak/Richard (1995) and Pollock (2009) in order to account for structural breaks and also Crainiceanu et al. (2005) introduced a time varying penalization for splines within a mixed model framework.

The paper is structured as follows: In the first section penalized splines based on a truncated polynomial basis are discussed briefly. Then it is shown how to choose the penalization in order to extract certain frequency bands. Afterwards it is explained how spectral analysis and a time varying penalization can be used to tackle the problem of the excess variability at the margins. The next section describes the effects of the time varying penalization and also compares the properties of different splines in the frequency domain. Finally section 4 gives an empirical example.

2 Penalized splines

A meanwhile popular instrument to estimate the trend component of a series $\{y_t\}_{t=1}^{T}$ are penalized splines with a truncated polynomial basis (Brumback et al., 1999; also Ruppert et al., 2003), following denoted as tp-splines. In a first step the explanatory variable time t, $t = 1, ..., T$, is divided into $m-1$ intervals by setting m knots $1 = \kappa_1 < \kappa_2 < ... < \kappa_{m-1} < \kappa_m = T$. The distance between the knots generally can vary, but in this paper always equidistant knots will be used. After the knots are set a tp-spline of degree l, henceforth denoted as $tp(l)$, is defined as:

$$y_t = f(t) + \varepsilon_t = \beta_1 + \beta_2 t + ... + \beta_{l+1} t^l + \beta_{l+2}(t-\kappa_2)_+^l + ... + \beta_d (t-\kappa_{m-1})_+^l + \varepsilon_t, \quad (1)$$

$$\text{with } (t-\kappa_j)_+^l = \begin{cases} (t-\kappa_j)^l & , t \geq \kappa_j \\ 0 & , \text{else} \end{cases},$$

where $d = m+l-1$. The first part is a polynomial of degree l, while the second part consists of truncated polynomials that enable $f(t)$ to become very flexible. In matrix notation the tp-spline is defined to:

$$\boldsymbol{y} = \boldsymbol{Z}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{2}$$

$$\text{with } \boldsymbol{Z} = \begin{pmatrix} 1 & 1 & \dots & 1^l & (1-\kappa_2)_+^l & \dots & (1-\kappa_{m-1})_+^l \\ \vdots & \vdots & \dots & \vdots & \vdots & \dots & \vdots \\ 1 & T & \dots & T^l & (T-\kappa_2)_+^l & \dots & (T-\kappa_{m-1})_+^l \end{pmatrix},$$

where $\boldsymbol{\beta} = (\beta_1, ..., \beta_d)'$, $\boldsymbol{\varepsilon} = (\varepsilon_1, ..., \varepsilon_T)'$ and $\boldsymbol{y} = (y_1, ..., y_T)'$. tp-splines can be interpreted as a continuous function of piecewise defined polynomials of degree l. Due to the truncated polynomials the coefficient of the highest polynomial changes at every knot, which allows the spline to become very flexible. To regulate the flexibility of tp-splines and to receive a smoother function the concept of penalization is used. As the coefficients $\beta_{l+2}, ..., \beta_d$ drive the flexibility of the tp-spline, the volatility of the estimated function can be determined by controlling the absolute values of these coefficients. For a given parameter λ the vector $\boldsymbol{\beta}(\lambda)$ is estimated by minimizing the penalized least squares criterion.

$$\min_{\boldsymbol{\beta}} PLS(\lambda) = \sum_{t=1}^{T} [y_t - f(t)]^2 + \lambda \sum_{j=l+2}^{d} \beta_j^2. \tag{3}$$

The solution of this minimization is:

$$\hat{\boldsymbol{\beta}}(\lambda) = (\boldsymbol{Z}'\boldsymbol{Z} + \lambda \boldsymbol{K})^{-1} \boldsymbol{Z}'\boldsymbol{y}, \tag{4}$$

$$\text{where } \boldsymbol{K}_{d \times d} = \text{diag}(\underbrace{0, ..., 0}_{l+1}, \underbrace{1, ..., 1}_{m-2}).$$

The fitted values for a given λ are defined by:

$$\hat{\boldsymbol{y}}(\lambda) = \underbrace{\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z} + \lambda \boldsymbol{K})^{-1} \boldsymbol{Z}'}_{\boldsymbol{H}(\lambda)} \boldsymbol{y}. \tag{5}$$

$\boldsymbol{H}(\lambda)$ is the hat matrix of the spline that contains the filter weights. The penalized least squares criterion describes a tradeoff between a close fit of the trend to the original data and a smooth trend function. The smoothness of the trend can be regulated by the penalization parameter λ, where high values of λ induce a smooth trend. An interesting feature of tp-splines is their link to the Wiener-Kolmogorov filter (see also Harvey, 1989; Kaiser/Maravall, 2001; McElroy, 2008) and square wave filters (Pollock, 2000, 2003). For T equidistant knots Proietti (2007) describes a tp-spline of degree l as a time series model where the trend is a $l+1$-fold integrated random walk and the cycle is white noise. Thus tp-splines are closely related to square wave filters that arise from the model framework of the Wiener-Kolmogorov filter. Moreover, Paige (2010) shows that for $l = 1$ and knots at every point in time $t = 1, 2, ..., T$, the spline is equal to the Hodrick-Prescott filter.

3 The optimal penalization

3.1 The penalization by frequency domain aspects

The conception that trend and cycle are characterized by their spectral properties can be employed to derive the penalization of splines. For a detailed discussion of spectral analysis see Granger/Hatanaka (1964), Harvey (1993), Hamilton (1994) or Mills (2003). In general spectral analysis allows decomposing a series $\{y_t\}_{t=1}^{T}$ into oscillations of different frequencies. This is utilized to define trend and cycle by certain bandwidths of frequencies. The trend represents the long run development of the time series and is supposed to be smooth so that it is described by oscillations with low frequencies, i.e. high periodicities. The cycle contains economic activity characterized by booms and recessions and is more volatile over time, since it reflects the development on the medium and short run. Consequently it is defined by medium and high frequencies. Extracting the trend by frequency domain aspects implies that oscillations with higher frequencies are suppressed, while those with lower frequencies are left unchanged.

To this point the gain function provides information about the impact of an instrument for trend estimation on the original series in the frequency domain. Using the matrix notation of a penalized tp-spline $\hat{\boldsymbol{y}}(\lambda) = \boldsymbol{H}(\lambda)\boldsymbol{y}$, where h_{ij} is the j^{th} element in the i^{th} row of $\boldsymbol{H}(\lambda)$, it is obvious that the spline defines a linear filter $\hat{y}_t = \sum_{j=1}^{T} h_{tj} y_j$ (e.g. Harvey, 1993). The gain function of a spline for estimation $\hat{y}_t$ and a frequency ω is given to (e.g. Mills, 2003):

$$g_t(\omega, \lambda) = \sqrt{\left[\sum_{j=1-t}^{T-t} h_{t,j+t} \cos(\omega j)\right]^2 + \left[\sum_{j=1-t}^{T-t} h_{t,j+t} \sin(\omega j)\right]^2}. \tag{6}$$

The gain can be interpreted as the factor by which an oscillation of frequency ω is damped or amplified, when a linear filter is applied. For an ideal instrument in order to extract the trend component the gain function should be one for low frequencies up to a certain cut-off frequency ω^{cf}, and zero for higher frequencies. This would imply that low frequencies are not affected by the filter while higher frequencies are completely eliminated. Such an instrument is called lowpass filter. The gain function of an ideal lowpass filter can be used to define trend and cycle in the frequency domain and to construct a selection criterion for the penalization parameter of splines. Such approaches were made by Baxter/King (1999) or Tödter (2002), who aimed to minimize the deviation between the gain function of the filter and the ideal gain function.

As an example Figure 1 shows an ideal gain function with a cut-off frequency of $\omega^{cf} = 0.5$. It adopts a value of one for frequencies in the interval $[0, 0.5]$. For all higher frequencies the ideal gain function has a value of zero.

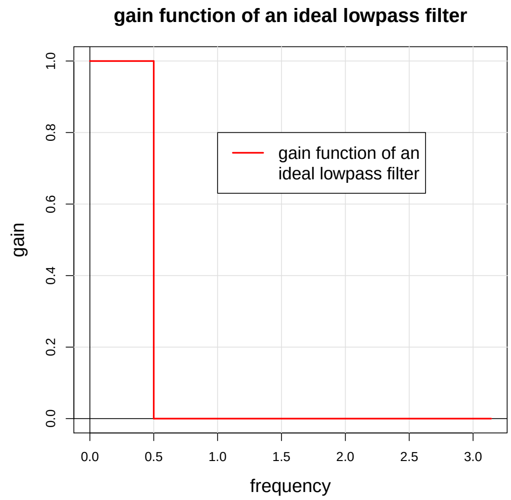

Figure 1: Example for an ideal gain function

The approach in this paper aims to minimize the square deviation between the gain function of a penalized spline form an ideal gain function by the selection of the penalization parameter λ. Note that it is not possible to completely realize an ideal gain function, since this would require an infinite number of filter weights (Oppenheim/Schafer, 1989). Let the gain of the spline for estimation $\hat{y}_t$, parameter λ and frequency ω be denoted as $g_t(\omega, \lambda)$ and the ideal gain as $g^*(\omega)$, then a so called *loss* $l_t(\lambda)$ can be defined:

$$l_t(\lambda) = \int_0^\pi [g^*(\omega) - g_t(\omega, \lambda)]^2 d\omega. \tag{7}$$

The *loss* is the squared deviation of the real gain function from the ideal one in the interval $[0, \pi]$. Now λ is selected such that $l_t(\lambda)$ is minimized. Since the minimization of $l_t(\lambda)$ is numerically complicated for continuous values of ω, it is approximated by a sufficient high number of discrete frequencies. If $\boldsymbol{\omega} \in \mathbb{R}^{n\times 1}$ denotes a vector of frequencies from zero to π in very small steps, e.g. $\boldsymbol{\omega} = (0, 0.001, 0.002, ..., \pi)'$, then $l_t(\lambda)$ can be written for discrete values:

$$l_t(\lambda) = \sum_{i=1}^{n} [g^*(\omega_i) - g_t(\omega_i, \lambda)]^2 \cdot \delta, \tag{8}$$

where n is the number of elements in $\boldsymbol{\omega}$ and δ is the distance between the elements in $\boldsymbol{\omega}$, i.e. $\delta = \omega_j - \omega_{j-1}$. The minimization can be done by algorithms like Newton-Raphson, fisher scoring or a grid search. If a grid search is used then a fast and stable implementation of penalized splines is required, which is for example described in the appendix of Ruppert et al. (2003).

3.2 Accounting for a time varying gain function

Before the penalization is selected by defining an ideal gain function and minimizing the *loss*, it has to be focused on the problem that the gain functions of estimations for different periods might not be equal. This is due to a changing structure of the filter weights, especially at the margins of the series. As an example Figure 2 displays the filter weights for a $tp(1)$ that is applied to a series with 100 observations with $\lambda = 1000$ and 100 equidistant knots. The left plot of Figure 2 shows the filter weights for estimations near the middle of the data. Clearly they have a very similar and almost symmetric structure. In contrast the right plot displays the weights for estimations close to the margin. The weight structure increasingly changes for estimations closer to the end of the series.

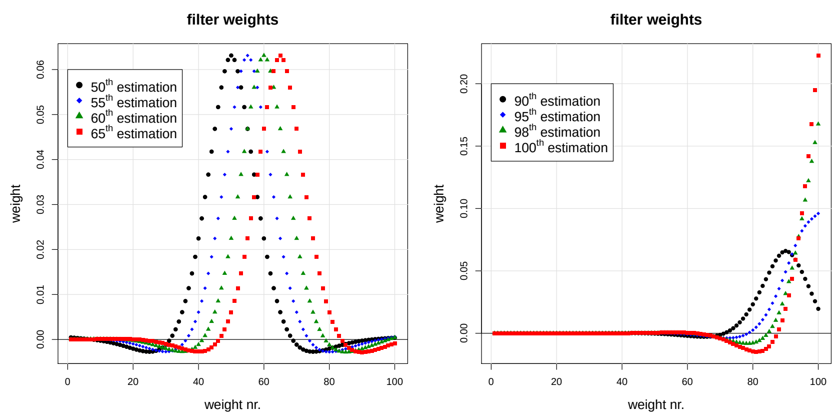

Figure 2: Filter weights for different estimations

This change of the filter weight structure affects the gain. Figure 3 shows the gain functions of estimations for different periods, which refer to the $tp(1)$ of the example above:

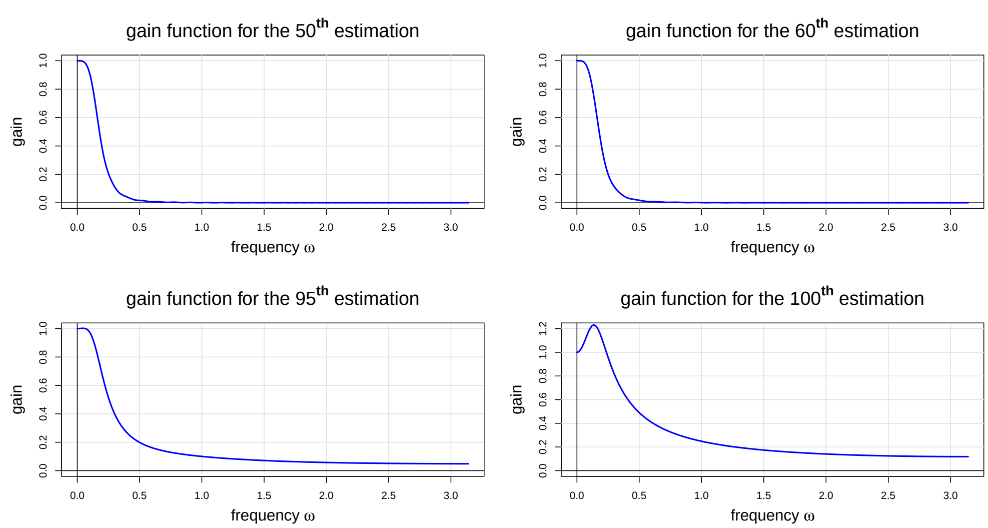

Figure 3: Gain functions of the $tp(1)$ for estimations around the middle and at the margin

The gain function for the 50^{th} and 60^{th} estimation look very similar. Both are good approximations of an ideal gain function and effectively eliminate high frequencies. This is different for estimations for periods at the margins. The gain functions for the 95^{th} and 100^{th} estimation adopt values greater that one for certain frequency bands and are not able to eliminate high frequencies. This insufficient suppression of high frequencies induces a too volatile trend estimation at the ends of the series.

Figures 2 and 3 motivate the reasons for the excess variability at the margins. However, the gain function or the filter weigh structure are not appropriate in order to describe this excess variability, as it is at least costly to consider the gain function or the filter weights for the estimations of all periods. It is much more practicable to regard the $loss$ over the estimations for all periods, as this allows to represent the excess variability with one single graph. This is following denoted as the $loss\ function$. As an example the left plot of Figure 4 shows the $loss\ function$ for the $tp(1)$ with $\lambda = 1000$ and $T = m = 100$. For the ideal gain function a cut-off periodicity of $\omega^{cf} = 0.196$ was chosen, which implies a periodicity of eight years in the case of quarterly data. Furthermore the right plot of Figure 4 displays for every $\hat{y}_t$, $t = 1, ...T$, that value of λ that minimizes the $loss$ for this specific estimation.

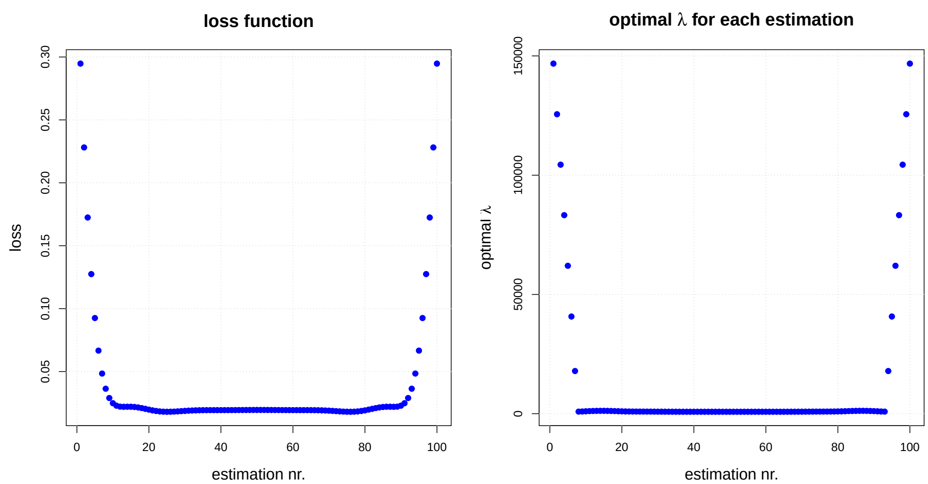

Figure 4: $loss\ function$ and optimal values for λ

The left plot of Figure 4 shows that the $loss$ is rather low and similar for most estimations in the middle of the data. However, for the estimations for about the first and last ten periods the $loss$ starts to increase. This illustrates that the excess variability mainly affects the margins of the series. Moreover, the right plot shows that except of the margins all estimations would require almost the same penalization in order to minimize the $loss$. At the margins the required penalization heavily increases and is up to 180 times higher than in the middle.

In order to develop methods to reduce the volatility at the margins it is useful to be aware of the factors that determine the excess variability. Beside the degree of the spline, which will be examined in detail later, two remaining potential factors are the value of the penalization parameter and the length of the series. To examine the influence of the length of the time series, Figure 5 shows the *loss functions* of a $tp(1)$ with $\lambda = 1600$ that is applied to series with varying length. To avoid any influence of the number of knots m was set equal to the number of observations T in every case.

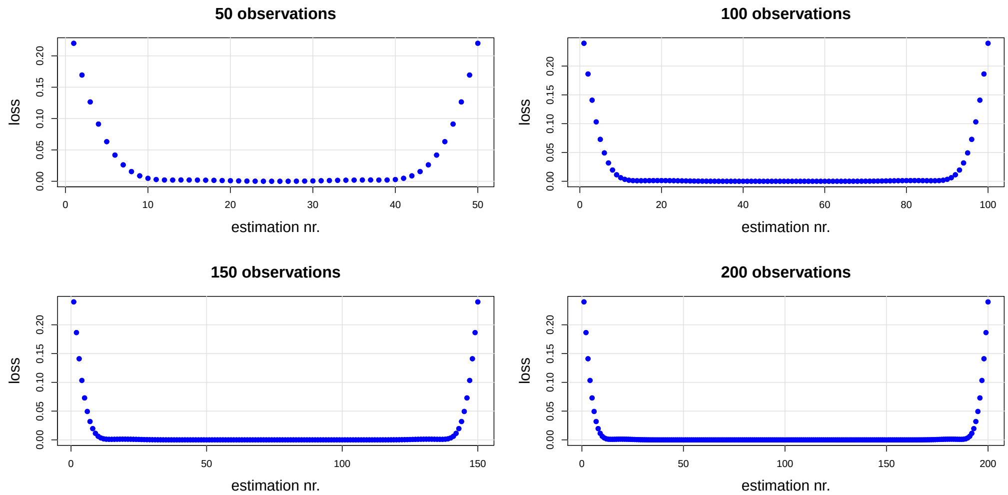

Figure 5: *loss functions* of a $tp(1)$ for series of different length

Figure 5 shows that independent of the number of observations about the first and last ten estimations are affected by the excess variability. Thus the length of the series seems to have no effect on the excess variability. This is different for the value of λ. Figure 6 displays the *loss functions* of a $tp(1)$ with different values of λ that is applied to a series with 100 observations. In every case m was set equal to T.

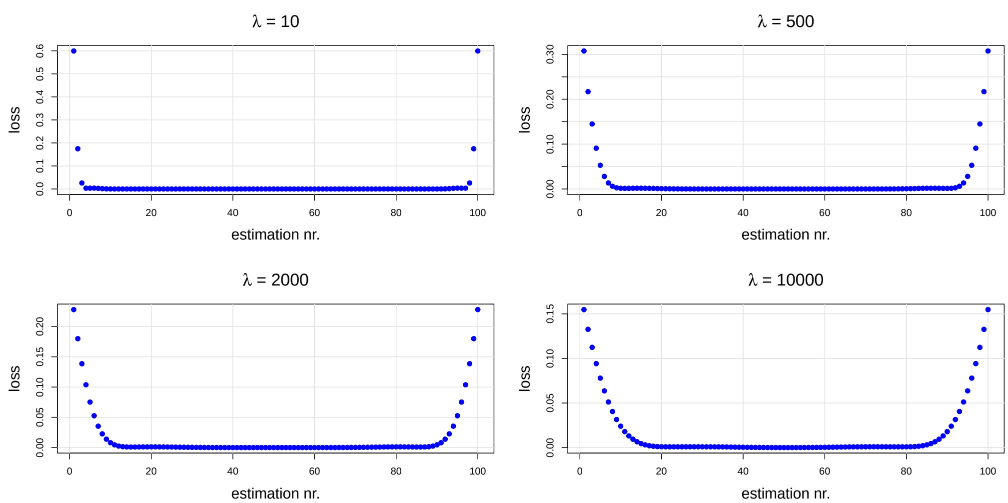

Figure 6: *loss function* of a $tp(1)$ for different values of λ

Clearly the number of estimations that is affected by the excess variability depends on the value of λ. For $\lambda = 10$ only about the first and last three estimations show an increased *loss*. However, for $\lambda = 10000$ about the first and last 20 estimations are affected by the excess variability. The number of estimations that exhibit an increased *loss* to the margins rises with higher values of the penalization. It follows from Figures 5 and 6 that the excess variability depends on the penalization but not on the number of observations.

3.3 The time varying penalization and the number of knots

The previous section described the undesired increase of the volatility to the margins. This and the next section show how the excess variability can be reduced by a time varying penalization that is allowed to increase to the margins. Recall the matrix formula of a penalized tp-spline from section 2, where the penalization was defined by the product of λ and the penalty matrix $\boldsymbol{K} = \text{diag}(0, ..., 0, 1, ..., 1)$. The product of λ and the i^{th} diagonal element of $\boldsymbol{K}$ gives the degree of penalization for the coefficient β_i. To achieve a flexible penalization the scalar λ has to be replaced by a vector $\boldsymbol{\lambda} = (0, ..., 0, \lambda_1, \lambda_2, ..., \lambda_{m-2})' \in \mathbb{R}^{d\times 1}$ and one defines the penalty matrix as $\tilde{\boldsymbol{K}} = \text{diag}(\boldsymbol{\lambda})$. The penalized spline with a time varying penalization can then be written in matrix notation as

$$\hat{\boldsymbol{y}}(\boldsymbol{\lambda}) = \boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z} + \tilde{\boldsymbol{K}})^{-1}\boldsymbol{Z}'\boldsymbol{y}. \tag{10}$$

As each coefficient of the truncated polynomials controls the change of the trend function at a certain knot, the penalization is not fixed any more but can vary over time. Figure 4 showed the penalization needs to increase to the ends of the series, while all other estimations require about the same degree of penalization. Consequently it seems appropriate to let the values of λ rise to the ends of the series to reduce the excess variability.

The basic purpose of the time varying penalization is to set a higher penalization at the margins of the series in order to reduce the excess variability. To this point it has to be considered how the penalization shall rise to the margins. Figure 4 suggests that a linear increase of the penalization might be appropriate so that the first and last j penalization parameters increase to the margins by a linear function. Then the last j values of $\boldsymbol{\lambda}$ can be expressed as

$$\lambda_{m-2-j+i} = \alpha_0 + \alpha_1 \cdot i, \quad i = 1, ..., j. \tag{11}$$

The first j λ's are defined just conversely, i.e.

$$\lambda_1 = \lambda_{m-2}, \lambda_2 = \lambda_{m-3}, ..., \lambda_j = \lambda_{m-1-j}. \tag{12}$$

α_0 is the value for the λ's closer to the middle which do not need to rise. As seen in Figure 4 the majority of estimations around the middle require about the same penalization. Thus it is sufficient to choose for α_0 that value of λ that minimizes the *loss* for the estimation in the middle of the series. As a consequence the first and last j λ's are defined according to (11) and (12) while all other λ's are set to α_0.

A further condition for the time varying penalization is that it shall reduce the excess variability at the margins without increasing the *loss* of estimations for periods closer to the middle. A criterion that is able to fulfil this condition is to minimize the cumulative *loss* of the estimations for all periods (see also Blöchl, 2014)

$$L(\boldsymbol{\lambda}) = \sum_{t=1}^{T} l_t(\boldsymbol{\lambda}), \tag{13}$$

by the time varying penalization. As it will turn out in the next section this criterion is suitable to reduce the variability at the margins without strongly affecting all other estimations. The minimization of the cumulative *loss* is reasonable, as one is usually not interested in a single estimation, but in the trend of the whole time series. Given the condition of a linear increase, the cumulative *loss* $L(\boldsymbol{\lambda})$ is minimized subject to (11) and (12). This can be done by algorithms like Newton Raphson, fisher scoring or a grid search. Because α_0 is fixed, the two remaining parameters are α_1 and j. One has to consider that j is an integer, so $L(\boldsymbol{\lambda})$ is minimized over α_1 for different, fixed values of j. Finally this value for j is chosen that yields to the lowest cumulative *loss*.

Besides the flexible penalization also a value for the number of knots m has to be selected. In general a higher number of knots allows a greater flexibility for the trend function. As Ruppert (2002) shows, there is a minimum number of knots that is necessary to achieve a reasonable fit of the trend function, but there are hardly changes if the number is further increased. Moreover, a higher number of knots can slightly increase the mean squared error (Ruppert, 2002). However, in order to reduce the excess variability at the margins it is preferable to choose a high number of knots, as it allows an accurate determination of the flexible penalization at the margins. As seen in Figure 4, the estimations for the first and last periods require different values of the penalization parameter. If the number of knots is too low then it might be not possible to set appropriate values of the penalization for all estimations at the margins. Thus the number of knots should generally be set as high as possible. In this paper for the $tp(1)$ m is always set equal to T. As equidistant knots are chosen this setting is identical to the Hodrick-Precott filter. Splines of higher degrees can become numeric instable, when the number of knots is too high (e.g. Fahrmeir et al., 2009). Hence for splines of higher degrees the number of knots should be set to a high value that still allows a numeric stable estimation.

3.4 Effects of the time varying penalization

The previous sections showed methods how to select the penalization of splines by frequency domain aspects and how to describe the increasing variability to the margins. It was argued that a time varying penalization can help to reduce this undesired increase of the variability, where the penalization shall rise to the margins. In order to show the effects of this time varying penalization it is applied to a simulated time series. To this point a cut-off frequency of $\omega^{cf} = 0.196$ is selected. Assuming that the exemplary data are quarterly this implies a cut-off periodicity of eight years, which is in line with Burns/Mitchell (1946). The simulated series contains 140 observations which might be not unrealistic for most quarterly economic time series. Table 1 shows the resulting parameters for tp-splines of degrees one, two and three.

Table 1: Optimal parameters for $\omega^{cf} = 0.196$

spline	m	α_0	α_1	j
$tp(1)$	140	821	654	21
$tp(2)$	140	79678	112500	28
$tp(3)$	140	$18.7 \cdot 10^6$	$40.6 \cdot 10^6$	35

For example, according to Table 1, the last j values of $\boldsymbol{\lambda}$ for the $tp(1)$ are defined by $\lambda_{m-2-j+i} = 821 + 654 \cdot i$, $i = 1, ..., j$, where $j = 21$. The first 21 values of the penalization $\lambda_1, ..., \lambda_{21}$ are defined analogously to (12). The remaining parameters $\lambda_{22}, ..., \lambda_{117}$ do not require an increased penalization and are set to 821. Table 1 shows that tp-splines of higher degrees require a larger increase of the penalization, where the increase also has to start closer to the middle. For the $tp(1)$ it is sufficient to increase the penalization for the first and last 21 values while for the $tp(3)$ the first and last 35 penalization parameters need to increase. Table 2 shows the basic results of the different types of splines for fixed and flexible penalization. For the fixed penalization the penalization parameter λ is set to the value of α_0 and the spline is calculated according to formula (5):

Table 2: *loss* for a cut-off periodicity of $\omega^{cf} = 0.196$

spline	penalization	70^{th} estimation	140^{th} estimation	$L(\boldsymbol{\lambda})$
$tp(1)$	fixed	0.019	0.320	4.706
	flexible	0.019	0.144	4.035
$tp(2)$	fixed	0.013	0.602	5.259
	flexible	0.013	0.330	4.264
$tp(3)$	fixed	0.009	0.886	6.232
	flexible	0.010	0.552	4.911

Let's just focus on the results for the fixed penalization at first. Splines of higher degrees yield a lower *loss* in the middle but a much higher *loss* at the margins. The *loss* of the $tp(1)$ in the middle is with 0.019 almost two times higher than the one of the $tp(3)$. However, the *loss* of the $tp(3)$ at the margin is almost three times higher than the one of the $tp(1)$. The different results of the splines also become obvious when their gain functions are considered. These are shown in Figure 7 that displays the gain functions for the splines in the middle (70^{th} estimation) and at the margin (140^{th} estimation). In the middle clearly tp-splines with a higher degree can extract frequency bands more precisely. The frequency band for the transition of the gain function from one to zero gets smaller when the degree of the spline is increased.

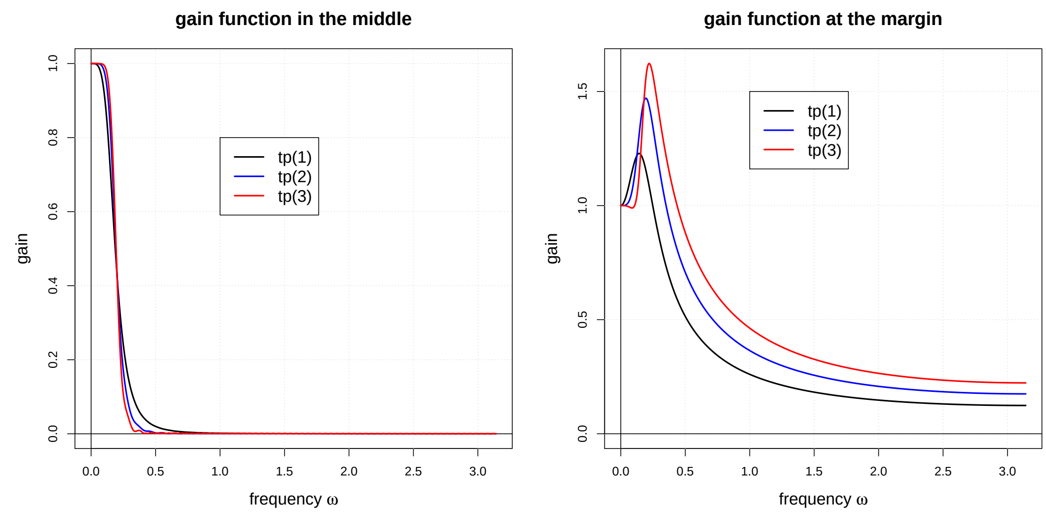

Figure 7: Gain functions in the middle and at the margin

This also shows the link of penalized tp-splines and rational square wave filters that is already described in Proietti (2007). The degree of the spline controls the transition of the gain function from one to zero while the penalization parameter determines the approximate cut-off frequency. At the margin splines of higher degrees increasingly loose the ability to suppress high frequencies inducing a much higher excess variability.

If the results for the flexible penalization in Table 2 are considered then one can see that the *loss* at the margins was reduced strongly in every case by around 38-55 percent, while the *loss* in the middle was hardly affected. Moreover the cumulative *loss* was decreased for every spline. To show the effects of the flexible penalization for the whole time series, Figure 8 displays the *loss functions* for both the fixed and the flexible penalization.

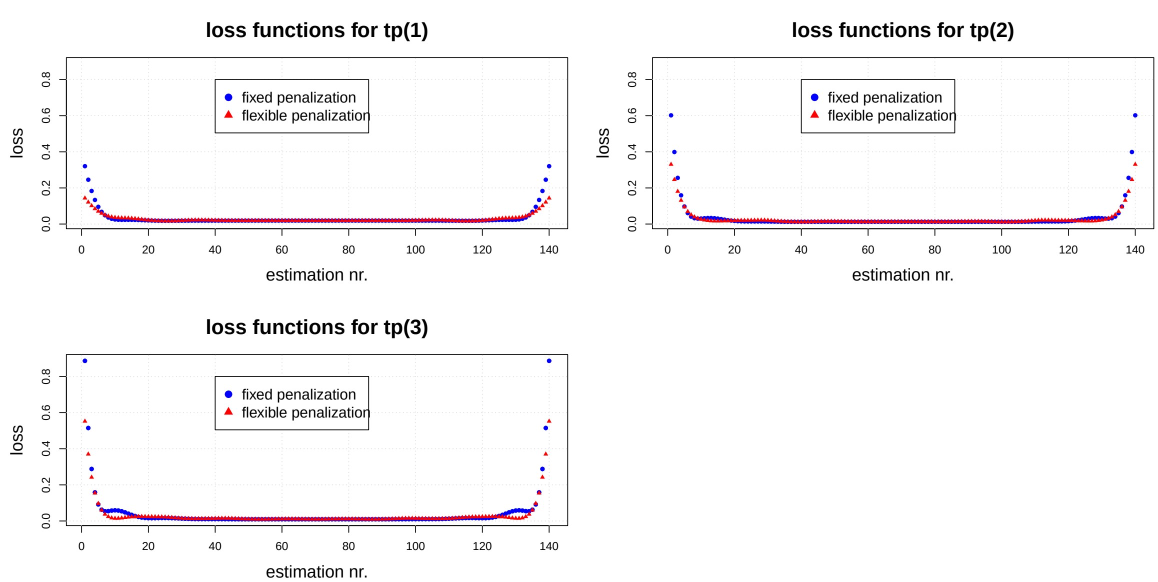

Figure 8: *loss functions* for fixed and flexible penalization

Clearly the $tp(1)$ exhibits the lowest excess variability at the margins. Moreover, for the $tp(2)$ and $tp(3)$ more estimations are affected by the excess variability than for the $tp(1)$. The most important result of Figure 8 is that in every case the *loss* at the margins could be reduced considerably by the flexible penalization while it was increased only slightly for some estimations closer to the middle. As the cumulative *loss* was decreased for every spline the reduction of the *loss* at the margins clearly outweighs these slight increases. Especially for the $tp(2)$ and $tp(3)$ the flexible penalization induced hardly any notable increases of the *loss* and clearly improves the results at the margins. Consequently Figure 8 shows that the flexible penalization is able to reduce the excess variability at the margins and to yield more reliable estimations for the most recent periods. Another important result is that there is clearly a tradeoff between a good approximation of an ideal gain function for most estimations around the middle and a low excess variability at the margins. Splines of higher degrees yield a better adaptation to an ideal gain function for the majority of estimations but also exhibit a far higher excess variability at the margins.

An existing method to overcome the problem of the excess variability at the margins is to attach forecasts to the end of the series. As Figure 8 shows in this case at least forecasts for the next 15 periods are required. For quarterly data this implies that one hast to predict data for the next four years. As the prediction errors are likely to be large for such a distance, the approach to add forecasts seems to be of limited practicability.

Finally, to get a better understanding of the effects of this time varying penalization it is worth considering the filter weights. Figure 9 plots the weights of the $tp(1)$ with the fixed and the flexible penalization of the example above in the middle (70^{th} estimation) and at the margin (140^{th} estimation).

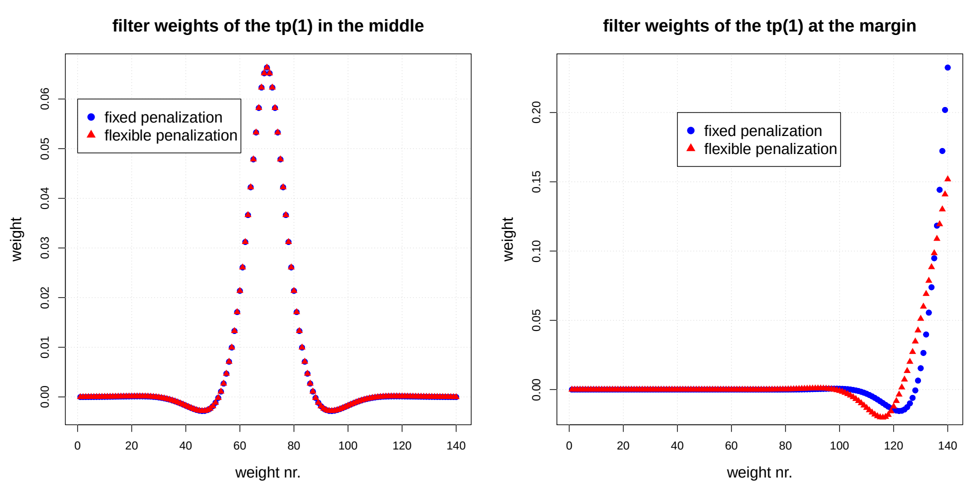

Figure 9: Filter weights for fixed and flexible penalization

The left plot of Figure 9 shows the weights for the estimation in the middle. There are almost no differences between fixed and flexible penalization. The weights are symmetric where the highest weight is about 0.07. This is different for the estimation at the margin.

As seen in Figure 2 the weight structure at the margin is not symmetric and the weights for the last few observations are very high. Consequently the estimation for the last period is predominantly influenced by few observations at the end of the series, especially by the last one. This causes the excess variability and deters the estimations for the periods at the margin to the value of the last observation. The time varying penalization dampens this behavior of the weights at the margin. The right plot shows that the time varying penalization reduced the weights that are attached to the last five observations, while others closer to the middle are increased in absolute values. Due to this declined influence of the last few observations the distortion of the trend estimation for periods at the margin to the value of the last observations is reduced.

4 Empirical application

In order to demonstrate the effect of the flexible penalization on real time series, the trend component of the seasonally adjusted quarterly real GDP of Switzerland is estimated[1]. The data start in the first quarter 1980 and end in the third quarter 2013 so that there are 135 observations. The trend shall be defined by a cut-off periodicity of eight years and is estimated with a $tp(1)$. The resulting optimal values for the flexible penalization are α_0=821, $\alpha_1 = 845$ and $j = 21$.

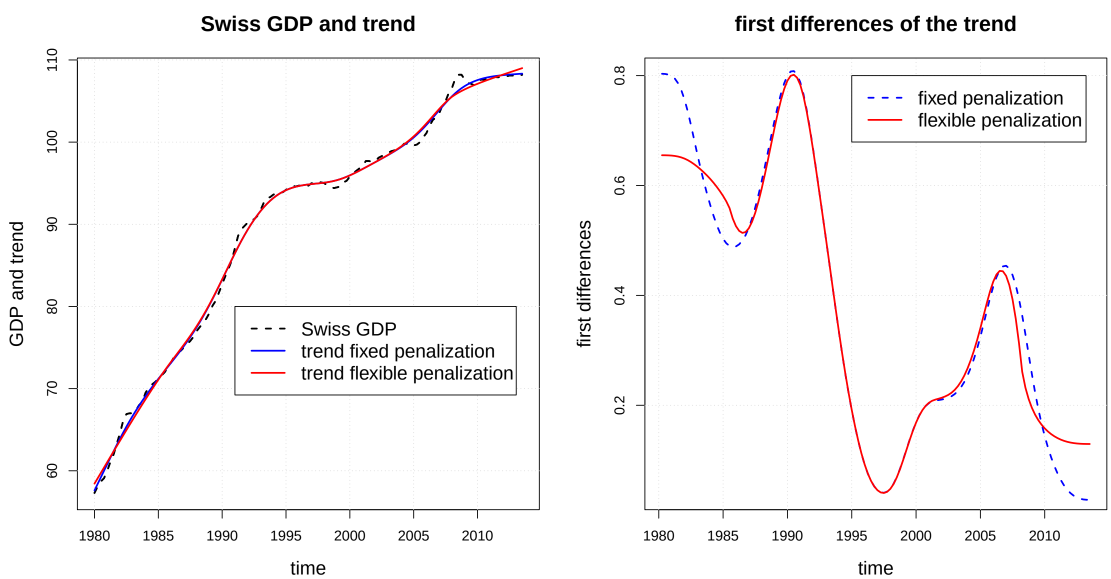

Figure 10: Trend estimation of Swiss GDP and first differences for fixed and flexible penalization

The left plot of Figure 10 shows the Swiss GDP as well as the estimated trend resulting from the fixed and flexible penalization. Especially at the end of the series are clear differences between the estimations. In both cases the trend growth rate declines after 2008, but

[1]The data are from the Swiss Secretariat of Economic Affairs, http://www.seco.admin.ch.

the decline is much larger for the fixed penalization. The trend according to the flexible penalization exhibits much larger growth rates and lies clearly above the one of the fixed penalization for the most recent periods. The right plot of Figure 10 shows the first differences of the trend for the estimations. Also the first differences only deviate at the margins of the series. In both cases the first differences decrease since about 2008, but the growth rate of the trend according to the time varying penalization has stabilized on a higher level.

It is also interesting to consider the effects of the time varying penalization on the business cycle. This is shown in Figure 11 for both cases.

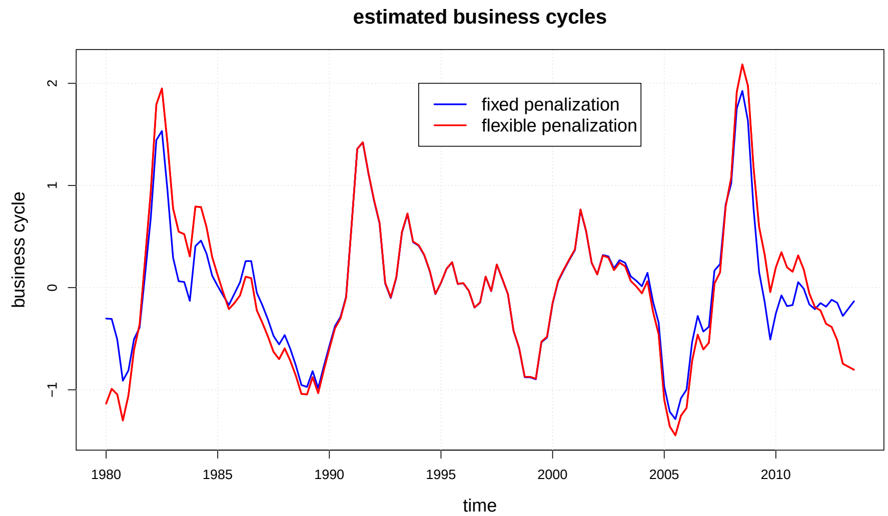

Figure 11: Estimated business cycle for fixed and flexible penalization

Also here the flexible penalization affected the margins of the series while it had no effect in the middle. It attributed the recession in the years 2008 and 2009 to a larger degree to the business cycle. Furthermore the output gap at the end of the series is much larger for the flexible penalization. As seen in Figures 2 and 9 for the estimations close to the end of the series the last observation is the most influential, which causes the excess variability. Thus in most cases the output gap is distorted to zero at the end of the series, as the last estimations tend to the value of the last observation. The flexible penalization reduces this distortion, which results in a much larger output gap at the margin in this example.

5 Conclusion

One of the unsolved problems of trend estimation is to get reliable results for the most recent periods. Due to the increasing asymmetry of the filter weights the volatility of estimations at the margins undesirably increases, which is known as the excess variability. The approach in this paper used penalized splines in order to estimate the trend component. The penalization was selected such that the gain function of the spline shows a minimal deviation from an ideal gain function. On the basis of this approach it was demonstrated that the deviation of the ideal gain function increases strongly for estimations at the margins. This behavior was described and visualized by the *loss function* that shows the deviation between real and ideal gain function over all periods.

The increasing variability of estimations at the margins was tackled by a time varying penalization. In detail the penalization was increased linearly to the margins, where the increase was such that the cumulative *loss* was minimized. It was shown that this criterion is capable of reducing the excess variability without strongly affecting other estimations closer to the middle of the series.

Moreover this paper showed that the degree of the spline strongly influences its properties in the frequency domain and pointed out the link between penalized tp-splines and rational square wave filters. The degree of the spline controls the transition of the gain function from one to zero while the penalization parameter determines the approximate cut-off frequency. Splines of higher degrees exhibit a more rapid transition so that they are better approximations of an ideal gain function. However, splines of higher degrees also suffer from a far higher excess variability at the margins. Thus there is clearly a tradeoff between a precise gain function of the estimations for most periods and a low excess variability at the margins.

Finally the paper demonstrated the effects of this time varying penalization for a real time series. To this point the trend of the Swiss GDP was estimated. There were clear differences between the trend according to the time varying penalization and the trend of the "standard approach" with a fixed penalization. In detail the time varying penalization showed higher trend growth rates over the last five years as well as a far higher output gap for the most recent periods.

The approach of this paper to get a handle on the excess variability could not completely eliminate the undesired volatility at the margins. Nevertheless, it could be shown that it is capable of improving the reliability of the estimation to the ends of the series. Thus this time varying penalization might be a useful instrument for researchers, especially when the focus lies on the most recent periods.

6 References

Altman, N. S. (1990): *"Kernel smoothing of data with correlated errors,"* Journal of the American Statistical Association 85, 749-759.

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.

Bell, W. (1984): *"Signal extraction for nonstationary time series,"* Annals of Statistics, 12, 646-664.

Blöchl, A. (2014): *"Reducing the Excess Variability of the Hodrick-Prescott Filter by Flexible Penalization,"* Munich Discussion Paper No. 2014-1.

Brumback, B. A., Ruppert, D., Wand, M. P. (1999): Comment on *"Variable selection and function estimation in additive nonparametric regression using a data-based prior,"* Journal of the American Statistical Association, 94, 794-797.

Burns, A. M., Mitchell, W. C. (1946): *"Measuring Business Cycles,"* New York: National Bureau of Economic Research.

Crainiceanu, C., Ruppert, D., Carroll, R. (2005): *"Spatially adaptive Bayesian P-splines with heteroscedastic errors,"* working paper.

Dagum, E. B., Giannerini, S. (2006): *"A critical investigation on detrending procedures for nonlinear processes,"* Journal of Macroeconomics, 28 (1), 175-191.

Danthine, J., Girardin, M. (1989): *"Business Cycles in Switzerland. A Comparative Study,"* European Economic Review, 33(1), 31-50.

Diggle, P. J., Hutchinson, M. F. (1989): *"On spline smoothing with autocorrelated errors,"* Australian Journal of Statistics 31, 166-182.

Eilers, P., Marx, B. (1996): *"Flexible smoothing with B-splines and penalties,"* Statistical Science, 11, 89-121.

Fahrmeir, L., Kneib, T., Lang, S. (2009): *"Regression-Modelle, Methoden und Anwendungen,"* Springer Verlag, Berlin Heidelberg.

Flaig, G. (2012): *"Why we should use high values for the smoothing parameter of the Hodrick-Prescott filter,"* CESifo Working Paper Series No. 3816.

Granger, C. W., Hatanaka, M. (1964): *"Spectral Analysis of Economic Time Series,"* Princeton Universtity Press, Princeton/New Jersey.

Hamilton, J. D. (1994): *"Time Series Analysis,"* Princeton University Press. Princeton/New Jersey.

Hart, J. D. (1991): *"Kernel regression estimation with time series errors,"* Journal of the Royal Statistical Society B 53, 173-187.

Harvey, A., C. (1989): *"Forecasting, Structural Time Series Models and the Kallman Filter,"* Cambridge University Press.

Harvey, A. C. (1993): *"Time Series Models,"* Harvester Wheatsheaf, Hertfordshire.

Hastie, T., Tibshirani, R. (1990): *"Generalized Additive Models,"* London: Chapman and Hall.

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

Kaiser, R., Maravall, A. (2001): *"Measuring Business Cycles in Economic Time Series,"* Springer Verlag.

Kauermann, G., Krivobokova, T., Semmler, W. (2011): *"Filtering Time Series with Penalized Splines,"* Studies in Nonlinear Dynamics and Econometrics 15(2), Article 2.

Kohn, R., Ansley, C., Wong, C. (1992): *"Nonparametric spline regression with autoregressive moving average errors,"* Biometrika, 79, 335-346.

Krivobokova, T., Kauermann, G. (2007): *"A note on penalized spline smoothing with correlated errors,"* Journal of the American Statistical Association, 102, 1328-1337.

McCallum, B. T. (2000): *"Alternative monetary policy rules: A comparison with historical settings for the United States, the United Kingdom and Japan,"* Economic Quarterly of the Federal Reserve Bank of Richmond, 86, 49-79.

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

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

O'Sullivan, F. (1986): *"A statistical perspective on ill-posed inverse problems (c/r: P519-527),"* Statistical Science, 1, 502-518.

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

Opsomer, J., Wang, Y., Yang, Y. (2001): *"Nonparametric regression with correlated errors,"* Statistical Science, 16, 134-153.

Paige, R. L. (2010): *"The Hodrick-Prescott filter: A special case of penalized spline smoothing,"* Electronic Journal of Statistics, 4, 856-874.

Pollock, D. S. G. (2000): *"Trend estimation and de-trending via rational square-wave filters,"* Journal of Econometrics 99, 317-334.

Pollock, D. S. G. (2003): *"Improved frequency selective filters,"* Computational Statistics & Data Analysis 42, 279-297.

Pollock, D. S. G. (2009): *"Improved frequency selective filters,"* Computational Statistics & Data Analysis 42, 279-297.

Proietti, T. (2005): *"Forecasting and signal extraction with misspecified models,"* Journal of Forecasting, 24, 539-556.

Proietti, T. (2007): *"Signal Extraction and Filtering by Linear Semiparametric Methods,"* Computational Statistics and Data Analysis 52, 935-958.

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.

Ruppert, D. (2002): "Selecting the number of knots for penalized splines," Journal of Computational and Graphical Statistics, 11, 735-757.

Ruppert, D., Wand, M., Carroll, R. (2003): *"Semiparametric Regression,"* Cambridge University Press.

Schlicht, E. (2005): *"Estimating the Smoothing Parameter in the so-called Hodrick-Prescott Filter,"* Journal of the Japanese Statistical Society, Vol. 35 No. 1, 99-119.

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

Wang, Y. (1998): *"Mixed effects smoothing spline analysis of variance,"* Journal of the Royal Statistical Society, Series B, 60, 159-174.

Whittle, P. (1983): *"Prediction and Regulation,"* Blackwell.