Penalized Splines as Time Series Filters in Economics

Theoretical and Practical Aspects in the Frequency and Time Domain

Inaugural-Dissertation
zur Erlangung des Grades Doctor oeconomiae publicae (Dr. oec. publ.) an der
Ludwig-Maximilians-Universität München

2014

vorgelegt von

Andreas Blöchl

Referent: Prof. Dr. Gebhard Flaig
Koreferent: Prof. Dr. Gerhard Illing
Promotionsabschlussberatung: 5. November 2014

Datum der mündlichen Prüfung: 21. Oktober 2014

Prüfer: Prof. Dr. Gebhard Flaig, Prof. Dr. Gerhard Illing, Prof. Dr. Timo Wollmershäuser

Acknowledgements

I am deeply grateful to my supervisor Prof. Dr. Gebhard Flaig for his permanent support and his insightful comments. I thank him for the very friendly environment that let me focus on my thesis and especially for always having the opportunity to ask him for a discussion or his advice. I also thank Prof. Dr. Gerhard Illing and Prof. Dr. Timo Wollmershäußer for their efforts.

A special thank goes to Martina Grass and Ursula Baumann for the valuable backup and to Florian Buck for his efforts and helpful comments on my thesis. Finally, I thank Renate Meitner and Ulrike Gasser for their support and Jakob Eberl for being the best roommate during the past three years.

Contents

List of Tables

List of Figures

Guide to notation

This section provides a brief overview of the notation in the thesis.

The field of real numbers is denoted as $\mathbb{R}$ and its $n \times m$ dimensional space as $\mathbb{R}^{n \times m}$.
Vectors are denoted by lowercase bold letters and always refer to column vectors:

$$\boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}.$$

Matrices are denoted by uppercase bold characters. The first index of the elements refers to the number of the row, the second index to the number of the column:

$$\boldsymbol{X} = \begin{pmatrix} x_{11} & \dots & x_{1m} \\ \vdots & \ddots & \vdots \\ x_{n1} & \dots & x_{nm} \end{pmatrix}.$$

$\boldsymbol{X} \in \mathbb{R}^{n \times m}$ denotes a matrix with n rows and m columns and entries from the field of the real numbers. For partitioned matrices the submatrices are denoted by bold uppercase characters:

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_{11} & \boldsymbol{X}_{12} \\ \boldsymbol{X}_{21} & \boldsymbol{X}_{22} \end{pmatrix}.$$

The transpose of a vector $\boldsymbol{y}$ or a matrix $\boldsymbol{X}$ is denoted by $\boldsymbol{y'}$ and $\boldsymbol{X'}$.
The inverse of a quadratic, invertible matrix $\boldsymbol{X}$ is denoted by $\boldsymbol{X}^{-1}$.
For a quadratic matrix $\boldsymbol{X}$ with n rows and columns the trace is denoted by

$$\operatorname{tr}(\boldsymbol{X}) = \sum_{i=1}^{n} x_{ii}.$$

For a vector $\boldsymbol{x}$ with n elements the $n \times n$ diagonal matrix is defined as

$$\operatorname{diag}(\boldsymbol{x}) = \begin{pmatrix} x_1 & & \boldsymbol{0} \\ & \ddots & \\ \boldsymbol{0} & & x_n \end{pmatrix}.$$

The Euclidean norm of a vector is denoted as

$$||\boldsymbol{y}|| = \sqrt{\boldsymbol{y'y}}.$$

The angular frequency is denoted as ω. The equivalent periodicity to a frequency ω is given by the relation

$$P = \frac{2\pi}{\omega}.$$

Autoregressive processes of order p are denoted as AR(p).
Autoregressive moving average processes of orders p and q are denoted as ARMA(p,q).
Integrated autoregressive moving average processes of orders p and q and integration order d are denoted as ARIMA(p, d, q).

$\{y_t\}_{t=1}^{T}$ denotes a time series with T elements that start in $t = 1$ and end in T.

L denotes the backshift operator for time series such that

$$\mathrm{L}^k y_t = y_{t-k}.$$

1 Preface

A fundamental conception in economic theory is that time series consist of different, independent components, predominantly trend and cycle. This conception is not only basis of a wide range of scientific fields, but it is furthermore important in many practical disciplines like economic policy, that essentially depends on a precise evaluation of the current status of trend and cycle. Business cycle research has a long tradition in economics, where its origins trace back into the second part of the nineteenth century (Mills 2003 p.1). In this regard Gabsich/Lorenz (1989 p.7) describe business cycle research by the explanation of observed up- and downturns in macroeconomic variables (for example the GDP, the inflation rate or the unemployment rate) as well as the analysis of conditions for periodic fluctuations in a model economy. Although the estimation of trend and cycle has been an important field in economics for almost 150 years, it is still a controversial topic today. Meanwhile, the literature knows several different definitions of trend and cycle, however, there is still no generally accepted definition (Stamfort 2005 p.7). From a very general point of view, the trend represents the long run development of a time series, while the cycle contains the economic activity characterized by booms and recessions and is supposed to exhibit a kind of sine pattern.

The literature provides three basic technical definitions of the cyclical component, which are the *classical cycle*, the *growth cycle* and the *growth rate cycle* (Anas/Ferrara 2002). The *classical cycle* might be the most widespread conception in economics (Anas/Ferrara 2004 p.90). It characterizes downturns by phases of absolute decline of the time series and upturns by phases of absolute increase. The *growth cycle* is described as the deviation of the observed series from the trend component, which is known as the output gap. Here, in phases of economic expansions the observed series shows a higher growth rate than the trend, while it is lower during downturns. Given the definition of the *growth cycle*, the trend component describes the output for a normal degree of capacity utilization, or under full employment in a Keynesian context (cf. Okun 1962, Bundesbank 2014 p.14). Thus, the *growth cycle* is a major element in business cycle theory, and is often regarded as the most important indicator of the cyclical state (Flaig/Plötscher 2001 p.221). A crucial issue of the growth cycle is the estimation of the trend component, which is the central topic of this thesis. Finally, the *growth rate cycle* defines upturns by periods, in which the growth rate of the observed series with regard to the previous period is positive, while downturns are phases that exhibit a negative growth rate.

In analogy to the definition of trend and cycle, the literature provides several conceptions about the reasons and duration of business cycles. Based on Schumpeter (1939) the most relevant theories about business cycles are summarized in Gabisch/Lorenz (1989). The *Kondratieff Cycle* describes cyclical movements due to technological progress and structural change like the beginning of the industrial electrification or the comprehensive building of railroads. Its duration is supposed to be between 40 and 60 years.

The *Juglar Cycle* explains phases of economic expansion and decline by the life cycle of investment goods and exhibits a duration between five to eleven years. The investigation of up- and downturns in the economy usually refers to the *Juglar Cycle* (Gabisch/Lorenz 1989 p.9).
Another widespread type of business cycle is the *Kitchin Cycle*. This cycle is supposed to be generated by random exogenous shocks, which temporarily lead to deviations of economic variables from their equilibrium level. The duration of the *Kitchin Cycle* is assumed to be between two and four years.

Gabisch/Lorenz (1989 p.9) note that it is questionable if these cycles really exist. Furthermore, even if the existence of these cycles is accepted, it is difficult to observe the cyclical components, as they overlap and generate in sum an ambiguous behavior (Schumpeter 1939). This fact complicates the isolation of certain cyclical components. As an example, Figure 1.1 plots the three types of business cycles as well as the series resulting from the sum of the cyclical components.

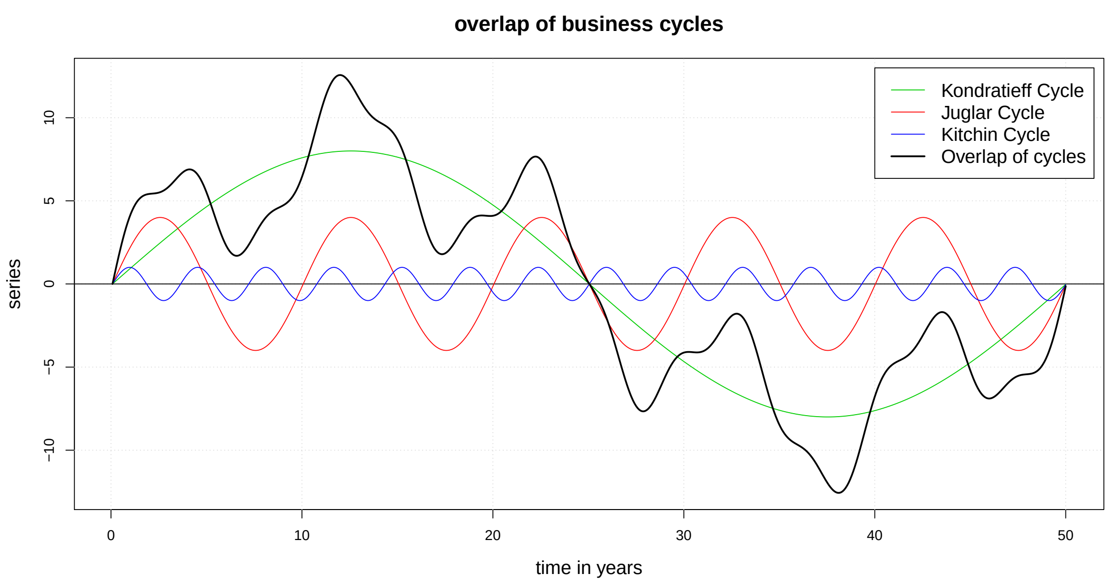

Fig. 1.1: Overlap of different business cycles
c.f. Schumpeter (1939) and Stamfort (2005 p.6)

Figure 1.1 illustrates one of the problems that might arise on the isolation of business cycles. For example, the resulting time series exhibits phases of upswings, when the *Kondratieff Cycle* is in a downturn, or a boom, when the *Juglar Cycle* is already falling.

To estimate the cyclical component and the trend in the sense of the *growth cycle*, meanwhile a wide range of instruments is available. One class of methods aims at smoothing the time series, thereby eliminating cyclical fluctuations and isolating the trend. The methods in this class can be summarized as univariate time series filters. Moreover, trend and cycle can be derived by econometric models like production functions, which are employed by most international organisations (Bundesbank 2014 p.14). To this point production functions model the aggregate output by the factors labour and capital as well as the use of technology. Nevertheless, production functions require the application of univariate filters to estimate the input factors (Bundesbank 2014 p.15).

This thesis deals with the challenges in the use of univariate time series filters. There is a great variety of univariate time series filters. In economics probably the Hodrick-Prescott filter (Hodrick/Prescott 1997) and the Baxter-King filter (Baxter/King 1999) are the most prevalent members of this class. More recently, also penalized splines (O'Sullivan 1986) have become popular tools in economics. These filters smooth an observed time series in order to extract the trend component. As pointed out, there is no general definition of trend and cycle in economics. Consequently, there is an ongoing discussion about the adequate degree of smoothing, when the trend is estimated by univariate linear filters.

The focus of this thesis lies on penalized splines as well as the Hodrick-Prescott filter, which are closely related methods. The Hodrick-Prescott filter derives the trend by the solution of a minimization problem, while penalized splines estimate the trend as a function of time $f(t)$. Both are so called penalized methods, as the outcome predominantly depends on the choice of a single penalization parameter. This parameter regulates the smoothness of the estimated trend. High values of the penalization parameter induce a smooth trend, while low values enable the trend to become very flexible. For the Hodrick-Prescott filter as the probably most widespread instrument for trend extraction in economics there is a vast amount of literature about the selection of the penalization parameter. Here, basically two different point of views can be distinguished that arise from structural time series models and from spectral analysis.

Structural time series models assume that a time series can be decomposed into unobserved components, where these components are regarded as independent, stochastic processes (Flaig 2003 p.258). Assuming that the series can be decomposed into a trend and a cyclical component, and that both are defined by an ARIMA-process, leads to the framework of the Wiener-Kolmogorov filter (Whittle 1983, Bell 1984). In business cycle theory the trend is usually described as an integrated process, while the cycle exhibits stationary, autocorrelated patterns. Given this conception of trend and cycle as ARIMA-processes, the Wiener-Kolmogorov filter yields the minimum mean squared error linear estimation for known parameters of the processes for trend and cycle. Hodrick/Prescott (1997) show that the Hodrick-Prescott filter is equal to the setting of the Wiener-Kolmogorov filter for a second fold integrated random walk as the trend, a white noise process as the cycle and when the penalization parameter is equal to the inverse signal to noise ratio. In this case, the inverse signal to noise ratio is defined as the ratio between the variance of the cycle and the variance of the second differences of the trend. For the US GDP Hodrick/Prescott (1997) decide to set the penalization parameter to 1600 for quarterly data, which meanwhile has become an 'industry standard' in economics (Flaig 2012 p.23). Given the value of 1600, Ravn/Uhlig (2002) recommend a penalization of 6.25 for yearly data and 129600 for monthly data as equivalent values.

Even though a penalization of 1600 is selected in most applications, this choice is seen as controversial in the literature. Danthine/Giradin (1989) criticize the suggestion of Hodrick/Prescott as dubious and subjective. Their critique is confirmed by the fact, that the assumptions about trend and cycle are not necessarily given in real time series. Especially a white noise business cycle appears to be a misspecification for most economic time series. Moreover, McCallum (2000) and also Flaig (2012) argue that a penalization of 1600 for quarterly data might be too low. For this value McCallum claims that economic crisis might misleadingly be attributed to the trend component. Flaig shows that for most series with an autocorrelated cycle a much higher penalization is required, if the Hodrick-Prescott filter shall yield comparable results to the Wiener-Kolmogorov filter. As the model parameters required for the Wiener-Kolmogorov filter are usually unknown, he suggests to set the penalization so high that the trend and/or its second differences show no cyclical behaivor anymore. Since Hodrick/Prescott derive the penalization considering only US variables, the value of 1600 is also often criticized as not data driven (e.g. Schlicht 2005, Kauermann et al. 2011). To this point Schlicht (2005) shows that the Hodrick-Prescott filter can be incorporated into a linear mixed model in order to receive an estimation for the penalization parameter. This model framework, however, is limited to a white noise error term. Since the error term represents the cyclical component, this is a restrictive assumption for most economic time series.

In contrast to the Hodrick-Prescott filter, penalized splines allow a data driven derivation of the penalization for the case of an autocorrelated residual structure, which represents the business cycle. Thus, the usually autocorrelated pattern of the cycle can be incorporated into the model. In this regard, generalized cross validation (Hastie/Tibshirani 1990) and the mixed model approach (e.g. Brumback et al. 1999) appear to be the most relevant methods to estimate the penalization parameter. Generalized cross validation aims at facing the tradeoff between minimizing the bias as well as the variance of the estimation, which results in minimizing the mean squared error between the true trend process and the estimation for $f(t)$. However, the selection of the penalization parameter using generalized cross validation is often not appropriate for the extraction of the trend component, since it yields estimations that fit very closely to the observed data and are thus too wiggly for most economic time series that contain autocorrelated error terms. This is already explored by Diggle/Hutchinson (1989), Altman (1990) and Hart (1991). To overcome this drawback, Kohn et al. (1992) and Wang (1998) extend this method to account for autocorrelated residual structures. Nevertheless, Opsomer et al. (2001), Proietti (2005) and also Dagum/Giannerini (2006) demonstrate that this technique is very sensitive to the assumption about the autocorrelation structure. With regard to the mixed model approach Krivobokova/Kauermann (2007) show that the results are robust even if the assumed autocorrelation deviates from the true (but unknown) one. This is a clear advantage to generalized cross validation with autocorrelated errors. Moreover, the mixed model framework yields the best linear unbiased predictors, which allows deriving unbiased confidence intervals for trend and cycle (e.g. Ruppert at al. 2003). More recently, it is demonstrated that the Hodrick-Prescott filter is a special case of a penalized spline (Paige, 2010). It follows that the mixed model framework of the Hodrick-Prescott filter can be extended to account for autocorrelated error terms. However, the resulting filter is not the same as the original Hodrick-Prescott filter.

Another possibility to motivate a reasonable selection of the penalization parameter arises from spectral analysis. Spectral analysis allows transferring a time series into the superposition of oscillations with different periodicities. Given the conception that trend and cycle can be distinguished more or less precisely by their spectral properties, it is possible to define these components by bandwidths of periodicities. In this sense the trend as the long run development of the series is attributed to high periodicities, while the cyclical component is characterized by rather medium and low periodicities. A widespread definition of the cycle traces back to Burns/Mitchell (1946), who define the cycle by fluctuations between six and 32 quarters. With regard to business cycle theories like the *Kondratieff Cycle*, the *Juglar Cycle* and the *Kitchin Cycle*, which all describe the cycle by aspects of duration, it appears to be straightforward to employ spectral analysis for the derivation of the penalization parameter. The penalization of the filter can be selected such that the filter extracts mainly the desired bandwidth of periodicities from a time series, while all other periodicities are eliminated. Following this argumentation Tödter (2002) recommends values for the penalization parameter of the Hodrick-Prescott filter between 453 and 3417 for quarterly data, which refer to assumed maximal cyclical durations between six and ten years. The properties of the Hodrick-Prescott filter in the frequency domain are well examined, amongst others by King/Rebelo (1993), Harvey/Jaeger (1993) or Cogley/Nason (1995). King/Rebelo and Harvey/Jaeger demonstrate that the Hodrick-Precott filter might induce phase shifts as well as spurious cycles, when it is employed for the extraction of the business cycle. Thus, the filtered series might exhibit cyclical behaviour, even if the observed series contains no cyclical component, and the filtering process might furthermore change the temporal relation to other series. A detailed overview about the characteristics of the Hodrick-Prescott filter is also provided by Stamfort (2005).

The estimation of trend and cycle still faces unsolved problems. One challenge arises, if the trend component shall be extracted for series that contain structural breaks. Structural breaks are problematic, since techniques like the Hodrick-Prescott filter or penalized splines are not able to react in time to massive changes in the observed series. Thus, such breaks distort the estimation of trend and cycle. Amongst others, this problem affects the estimation of trend and cycle for the German economy due to the reunification in 1990. This can be seen in Figure 2.2 (page 17), which shows a rapid change of the German GDP from 1990 to 1991. This break does not permit a reliable estimation of trend and cycle for the whole series. In order to avoid distortions by structural breaks Razzak/Richard (1995) and also Pollock (2009) suggest to use a time-varying penalization for the Hodrick-Prescott filter or penalized splines. They propose to set a very low penalization at the time of the break, which allows a fast adaptation of the estimated trend to the observed data at this point in time. Moreover, Schlicht (2008) shows how to include dummy variables into the

Hodrick-Prescott filter to account for breaks. However, these methods are limited to a priori selected values for the penalization parameter, and do not allow a data driven choice of the penalization. To overcome the problem of structural breaks when the penalization parameter shall be estimated, this thesis extends the mixed model framework for penalized splines. In chapter 2 it is demonstrated how a time-varying penalization can be implemented and estimated within a mixed model to avoid distortions by structural breaks. This model is afterwards employed to estimate trend and cycle of the German GDP.

An even more general problem of time series filters is the estimation of trend and cycle at the margins of a series. Completely symmetric filters like the Baxter-King filter cannot derive estimations for the first and last periods. Partially symmetric filters like penalized splines are principally able to yield estimations for the margins of the time series. However, the reliability of the estimated trend and cycle strongly decreases to the margins. This decreasing reliability is generated by an asymmetric filter weight structure for estimations at the beginning and the end. For estimations around the middle the filter weights of a penalized spline are (almost) symmetric, so that the estimations arise from a weighted average of past present and future observations. At the end of the series, however, the filter weights become strongly asymmetric due to a lack of future observations. The increasing asymmetry of the filter weight structure induces a too high variability of the estimations at the margins, which is called excess variability. This excess variability is a problem, since researchers and policy makers are often especially interested in trend and cycle of the most recent periods. An approach to avoid this excess variability is to attach forecasts to the end of the series. This allows a more symmetric filter weight structure for the estimations at the margins. Nevertheless, in most cases a high number of forecasts is required to reduce the excess variability. As the forecasts exhibit failures that increase with a rising forecasting horizon, this method appears to be of limited practicability. Instead, chapters 4 and 5 of this thesis offer another approach to deal with the problem of the excess variability, when trend and cycle are derived by frequency domain considerations. They show how to employ a time-varying penalization to reduce the excess variability. As the variability of the estimation undesirably increases to the margins, the penalization is allowed to rise to the ends. It is demonstrated that this method is capable of strongly reducing the excess variability at the margins, whereas estimations closer to the middle are hardly affected.

The general outline of this thesis is as follows. The mixed model framework of penalized splines for series with structural breaks is described in the second chapter of this thesis. The third chapter shows that penalized splines as mixed models can be integrated into the framework of the Wiener-Kolmogorov filter. In detail it is demonstrated that under certain settings of the model parameters, the spline within a mixed model is equal to the Wiener-Kolmogorov filter, when the trend is a second-fold integrated random walk and the cycle follows a stationary ARMA-process. This model is employed to estimate trend and cycle of the German GDP as well as the index of the German industrial production, which is a proxy for the GDP. The basic result is that there is a very smooth and almost linear development of the German economy on the long run. The reduction of the excess variability by a time-varying penalization is explained in chapters 4 and 5. Finally, in chapter 6 penalized splines are compared to the Baxter-King filter. The Baxter-King filter is a widespread instrument for the extraction of trend and cycle in economics. It is motivated by frequency domain aspects and aims at a precise extraction of frequency bands. Moreover, Baxter/King (1999) postulate features for an ideal time series filter. Beside a precise extraction of frequency bands the filter should not induce phase shifts and thus not alter the original relation between the filtered to other series. Additionally, the estimated cycle should be stationary, even if the observed series is integrated up to order two or contains a quadratic time trend. The sixth chapter examines in how far penalized splines meet these requirements. It is shown that penalized splines approximately do not induce phase shifts and render stationary cycles, where these features depend on the number of observations and the selected penalization. Furthermore, the ability of penalized splines to extract frequency bands is compared to the one of the Baxter-King filter. The basic result is, that penalized splines are superior when used for trend estimation, but not necessarily for the extraction of the cyclical component.

In summary, this thesis sheds light on penalized splines as instruments for the estimation of trend and cycle in economics. It is shown how penalized splines can be modified to get a handle on problems that arise, when trend and cycle are estimated. The thesis examines their characteristics with regard to features that are often postulated for time series filters in economics, compares them to predominantly used methods and integrates them into economic theory.

2 Trend Estimation with Penalized Splines as Mixed Models for Series with Structural Breaks[1]

Summary

Penalized splines are popular tools for trend estimation, as they offer a data driven derivation of the penalization parameter by the incorporation into a mixed model framework. However, this approach might fail to estimate trend and cycle, when time series contain structural breaks. This chapter shows how the model framework can be extended to get a reasonable estimation for the penalization parameter, even if the data exhibit break points. In this regard, the variances of the random effects and thus the penalization, is allowed to vary at the time of the break. Moreover, it is discussed, which parameter settings are preferable for this purpose. Finally, the technique is employed to estimate trend and cycle of the German GDP, which contains a structural break due to the reunification in 1990. The basic outcome is that there is a linear long run development of the German economy.

[1] This chapter refers to Blöchl (2014c).

2.1 Introduction

An important field of economic research is the decomposition of a time series into a trend and a cyclical component. The trend represents the long run development of the data, whereas the cycle contains the economic activity characterized by phases of upswings and downturns. On purpose to estimate trend and cycle, several different methods have been developed. The probably most popular is the Hodrick-Prescott filter, which traces back to Whittaker (1923), Henderson (1924) and Leser (1961). The Hodrick-Prescott filter (Hodrick/Prescott 1997) is a linear filter, predominantly depending on the choice of a single penalization parameter λ that controls the smoothness of the estimated trend. The choice of λ has been an issue of research, especially since there is no general definition of trend or cycle (Stamfort 2005 p.7). To this point Hodrick/Prescott (1997) suggest a value of $\lambda = 1600$ for quarterly data. Ravn/Uhlig (2002) commend values of 6.25 and 129600 for yearly and monthly data respective.

Nevertheless, Hodrick/Prescott's choice of the penalization parameter is often criticized as arbitrary (Danthine/Giradin 1989), because the derivation of this value can be seen as dubious concerning economic and statistical aspects. The derivation of $\lambda = 1600$ arises from the equivalence of the Hodrick-Prescott filter to the Wiener-Kolmogorov filter (Whittle 1983, Bell 1984), when the trend is a twofold integrated random walk and the cycle is white noise as well as subjective assumptions about the variance of trend and cycle (Hodrick/Prescott 1997). These circumstances are clearly not necessarily given in real time series. Schlicht (2005) shows how the Hodrick-Prescott filter can be written as a linear mixed model in order to estimate the penalization. However, this approach is limited to a white noise error term which is restrictive, as business cycles are usually supposed to be autocorrelated.

In contrast to the Hodrick-Prescott filter penalized splines (O'Sullivan 1986) within a mixed model framework allow estimating the penalization parameter for an autocorrelated residual structure. Thus, the usually autocorrelated pattern of business cycles can be incorporated into the model. A further advantage of this method is that the results are robust with regard to a misspecification of the residual autocorrelation structure as long as the misspecification is not too large (Krivobokova/Kauermann 2007). This is a positive feature, since the true autocorrelation structure is unknown in most cases. Paige (2010) shows that the Hodrick-Prescott filter is a special case of a penalized spline. Consequently, the mixed model approach with an autocorrelated residual structure is available for this filter. However, the resulting filter is not equivalent to the original Hodrick-Prescott filter any more.

Even though the mixed model framework of splines helps to overcome the subjective choice of the penalization, it fails to estimate trend and cycle when series contain structural breaks. Such breaks can be due to different reasons, for example the contribution of East Germany to the German GDP from 1991 onwards, or the abrupt rise in the German population after the reunification. Such break points lead to undesirable distortions when trend and cycle are estimated with time series filters like the Hodrick-Prescott filter or penalized splines. Thus, techniques that allow for a consideration of break points are required. Schlicht (2008) suggests extending the Hodrick-Prescott filter by dummy variables to account for structural breaks. Alternatively Razzak/Richard (1995) and also Pollock (2009) propose not to use single but different penalization parameters within the Hodrick-Prescott filter. In detail, they propose to change the penalization parameter at the time of the break point to reduce the generated distortions. Even if this method allows accounting for breaks, it still suffers from the shortcoming of the disputed choice of the penalization.

In this chapter a related approach to deal with structural breaks is developed. It extends the mixed model framework for penalized splines to account for breaks by a time-varying penalization. This allows deriving a data driven estimation of trend and cycle, even if the time series contains a structural break. The chapter is structured as follows. The first section shortly summarizes penalized splines with a truncated polynomial basis and how they can be incorporated into a linear mixed model (Brumback et al. 1999). Afterwards it is shown how the model framework can be extended by a flexible penalization in order to account for breaks points. Finally, empirical examples are provided, where the focus lies on estimating the trend of the German GDP.

2.2 The model framework

2.2.1 tp-splines as linear mixed models

This chapter deals with tp-splines. Here, tp-splines denote splines with a truncated polynomial basis that trace back to Brumback et al. (1999). This kind of spline is advantageous because of its easy implementation and its obvious connection to linear mixed models. Estimating the trend component with a tp-spline means the trend function is modelled in dependence of time t. To this point the time variable t, $t = 1, ..., T$ is divided into $m-1$ intervals by setting m knots $1 = \kappa_1 < \kappa_2 < ... < \kappa_m = T$. The distance between the knots generally can vary, but usually equidistant knots are selected. In this case the knots are defined as (e.g. Fahrmeir et al. 2009 p.301-302)

$$\kappa_j = 1 + (j-1)h, \quad \text{where } h = \frac{T-1}{m-1}.$$

After setting the knots, a tp-spline of degree l for a time series $\{y_t\}_{t=1}^T$ can be written as

$$y_t = f(t) + \varepsilon_t = \delta_1 + \delta_2 t + \delta_3 t^2 + ... + \delta_{l+1} t^l + \delta_{l+2}(t-\kappa_2)_+^l + ... + \delta_d (t-\kappa_{m-1})_+^l + \varepsilon_t, \quad (2.1)$$

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

where ε_t denotes the error term that represents the business cycle and $d = m + l - 1$. Writing the model in matrix notation yields:

$$\boldsymbol{y} = \boldsymbol{Z\delta} + \boldsymbol{\varepsilon}, \quad (2.2)$$

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

Here, $\boldsymbol{\delta} = (\delta_1, ..., \delta_d)'$, $\boldsymbol{\varepsilon} = (\varepsilon_1, ..., \varepsilon_T)'$ and $\boldsymbol{y} = (y_1, ..., y_T)'$. The truncated polynomials allow a great flexibility of $f(t)$, as they induce that the coefficient of the highest polynomial changes at every knot. The degree of the change and thus the flexibility of $f(t)$ is determined by the values of the coefficients $\delta_{l+2}, ..., \delta_d$. Consequently, the flexibility of the trend function can be regulated by controlling the coefficients of the truncated polynomials. This is done by the penalized least squares criterion (e.g. Fahrmeir et al. 2009 p.308):

$$PLS(\lambda) = \sum_{t=1}^{T} [y_t - f(t)]^2 + \lambda \sum_{j=l+2}^{d} \delta_j^2. \tag{2.3}$$

The second part of $PLS(\lambda)$ controls the smoothness of the estimated trend function. It is weighted by a factor λ that is called the penalization parameter. Increasing the value of λ induces a smoother trend, as the coefficients of the truncated polynomials become absolutely smaller. Finally, the solution of (2.3) in matrix notation is given to (e.g. Fahrmeir et al. 2009 p.313)

$$\hat{\boldsymbol{\delta}} = (\boldsymbol{Z}'\boldsymbol{Z} + \lambda\boldsymbol{K})^{-1}\boldsymbol{Z}'\boldsymbol{y}, \quad \text{where} \quad \boldsymbol{K} = \operatorname{diag}(\underbrace{0, ..., 0}_{l+1}, \underbrace{1, ..., 1}_{m-2}), \tag{2.4}$$

$$\text{so that} \quad \hat{\boldsymbol{y}} = \boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z} + \lambda\boldsymbol{K})^{-1}\boldsymbol{Z}'\boldsymbol{y}. \tag{2.5}$$

To receive an estimator for λ a penalized tp-spline can be interpreted as a linear mixed model as already shown by Brumback et al. (1999) (for a detailed discussion of mixed models see Searle et al. 1992, Vonesh/Chinchilli 1997, Pinheiro/Bates 2000 or McCulloch/Searle 2001). A tp-spline within a mixed model has the form (e.g. Ruppert et al. 2003 p.108)

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{U}\boldsymbol{\gamma} + \boldsymbol{\varepsilon} = \boldsymbol{Z}\boldsymbol{\theta} + \boldsymbol{\varepsilon}, \tag{2.6}$$

$$\text{where } \boldsymbol{X} = \begin{pmatrix} 1 & 1 & \dots & 1^l \\ 1 & 2 & \dots & 2^l \\ \vdots & \vdots & \ddots & \vdots \\ 1 & T & \dots & T^l \end{pmatrix} \text{ and } \boldsymbol{U} = \begin{pmatrix} (1-\kappa_2)_+^l & \dots & (1-\kappa_{m-1})_+^l \\ \vdots & \ddots & \vdots \\ (T-\kappa_2)_+^l & \dots & (T-\kappa_{m-1})_+^l \end{pmatrix}.$$

Consequently, $\boldsymbol{Z} = [\boldsymbol{X}, \boldsymbol{U}]$, $\boldsymbol{\theta}' = [\boldsymbol{\beta}', \boldsymbol{\gamma}']$, $\boldsymbol{\beta} \in \mathbb{R}^{(l+1)\times 1}$ and $\boldsymbol{\gamma} \in \mathbb{R}^{(m-2)\times 1}$. Additionally, it is assumed that $\boldsymbol{\varepsilon} \sim \mathrm{N}(0, \boldsymbol{R})$ and $\boldsymbol{\gamma} \sim \mathrm{N}(0, \boldsymbol{G})$. This means an autocorrelated residual structure can be allowed. A special feature of the representation of the spline as a mixed model is its interpretation as a hierarchical model (e.g. Fahrmeir et al. 2009 p.261):

$$\boldsymbol{y}|\boldsymbol{\gamma} \sim \mathrm{N}(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{U}\boldsymbol{\gamma}, \boldsymbol{R}) \text{ and the marginal distribution } \boldsymbol{\gamma} \sim \mathrm{N}(0, \boldsymbol{G}).$$

Moreover, the spline can be seen as a marginal model:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}^* \text{ with } \boldsymbol{\varepsilon}^* = \boldsymbol{U\gamma} + \boldsymbol{\varepsilon}.$$

The distribution of $\boldsymbol{y}$ is then given by $\boldsymbol{y} \sim \mathrm{N}(\boldsymbol{X\beta}, \boldsymbol{V})$, where $\boldsymbol{V} = \boldsymbol{R} + \boldsymbol{UGU'}$. Given the hierarchical model of $\boldsymbol{y}$ the parameter vector $\boldsymbol{\theta}$ can be estimated by maximizing the resulting log-likelihood function with respect to $\boldsymbol{\theta}$ (e.g. Ruppert et al. 2003 p.100):

$$\log L(\boldsymbol{\theta}) = -\frac{1}{2}(\boldsymbol{y} - \boldsymbol{Z\theta})'\boldsymbol{R}^{-1}(\boldsymbol{y} - \boldsymbol{Z\theta}) - \frac{1}{2}\boldsymbol{\gamma}'\boldsymbol{G}^{-1}\boldsymbol{\gamma}. \tag{2.7}$$

If it is further specified that $\boldsymbol{R} = \sigma^2\boldsymbol{\Omega}$ and $\boldsymbol{G} = \mathrm{diag}(\tau^2)$ (e.g. Kauermann et al. 2011), i.e. the error term is autocorrelated with a constant variance σ^2 and the autocorrelation matrix $\boldsymbol{\Omega}$ and the parameters in $\boldsymbol{\gamma}$ are uncorrelated and have the constant variance τ^2, then this is equivalent to minimizing (e.g. Fahrmeir et al. 2009, Kauermann et al. 2011)

$$\min_{\boldsymbol{\theta}} \; (\boldsymbol{y} - \mathbf{Z}\boldsymbol{\theta})'\boldsymbol{\Omega}^{-1}(\boldsymbol{y} - \boldsymbol{Z\theta}) + \lambda\boldsymbol{\gamma}'\boldsymbol{\gamma}, \tag{2.8}$$

where $\lambda = \frac{\sigma^2}{\tau^2}$. As $\boldsymbol{G} = \mathrm{diag}(\tau^2)$, the solution of the maximization of formula (2.7) is (Robinson 1991, Hayes/Haslett 1999, also Ruppert et al. 2003 p.100)

$$\tilde{\boldsymbol{\theta}} = \begin{bmatrix} \tilde{\boldsymbol{\beta}} \\ \tilde{\boldsymbol{\gamma}} \end{bmatrix} = (\boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{Z} + \frac{1}{\tau^2}\boldsymbol{K})^{-1}\boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{y}, \tag{2.9}$$

where the estimators $\tilde{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{\gamma}}$ are the best, linear unbiased predictors (BLUPs) of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$. A remaining problem is that usually the matrices $\boldsymbol{R}$ and $\boldsymbol{G}$ are unknown and have to be estimated. To this regard both matrices are written in dependence of a vector of parameters $\boldsymbol{\vartheta}$, where $\boldsymbol{\vartheta}$ depends on the assumed correlation structures of $\boldsymbol{\varepsilon}$ and $\boldsymbol{\gamma}$. Then the log-likelihood is derived by the interpretation as a marginal model which is given except of a constant term to (e.g. Ruppert et al. 2003 p.100-101):

$$l(\boldsymbol{\beta}, \boldsymbol{\vartheta}) = -\frac{1}{2}\left[\log(|\boldsymbol{V}(\boldsymbol{\vartheta})|) + (\boldsymbol{y} - \boldsymbol{X\beta})'\boldsymbol{V}(\boldsymbol{\vartheta})^{-1}(\boldsymbol{y} - \boldsymbol{X\beta})\right]. \tag{2.10}$$

Differentiating and solving with respect to $\boldsymbol{\beta}$ yields

$$\tilde{\boldsymbol{\beta}}(\boldsymbol{\vartheta}) = \left[\boldsymbol{X}'\boldsymbol{V}(\boldsymbol{\vartheta})^{-1}\boldsymbol{X}\right]^{-1}\boldsymbol{X}'\boldsymbol{V}(\boldsymbol{\vartheta})^{-1}\boldsymbol{y}. \tag{2.11}$$

Reinserting into (2.10) finally yields the profile log-likelihood:

$$l_p(\boldsymbol{\vartheta}) = -\frac{1}{2}\left(\log|\boldsymbol{V}(\boldsymbol{\vartheta})| + [\boldsymbol{y} - \boldsymbol{X}\tilde{\boldsymbol{\beta}}(\boldsymbol{\vartheta})]'\boldsymbol{V}(\boldsymbol{\vartheta})^{-1}[\boldsymbol{y} - \boldsymbol{X}\tilde{\boldsymbol{\beta}}(\boldsymbol{\vartheta})]\right). \tag{2.12}$$

Instead of the profile log-likelihood in most applications the restricted log-likelihood (Searle et al. 1992) is maximized:

$$l_r(\boldsymbol{\vartheta}) = l_p(\boldsymbol{\vartheta}) - \frac{1}{2}\log|\boldsymbol{X}'\boldsymbol{V}(\boldsymbol{\vartheta})^{-1}\boldsymbol{X}|. \tag{2.13}$$

The restricted log-likelihood is more accurate in small samples, since it takes into account the degrees of freedom for of the fixed effects (Ruppert et al. 2003 p.101). By maximizing $l_r(\boldsymbol{\vartheta})$ with respect to $\boldsymbol{\vartheta}$ one receives $\hat{\boldsymbol{\vartheta}}$ and thus also the estimators $\hat{\boldsymbol{R}}$ and $\hat{\boldsymbol{G}}$. As also σ^2 and τ^2 are contained in $\boldsymbol{\vartheta}$, immediately $\hat{\lambda} = \frac{\hat{\sigma}^2}{\hat{\tau}^2}$ can be calculated. Furthermore, $\hat{\boldsymbol{R}}$ and $\hat{\tau}^2$ can be inserted into (2.9), which yields the estimators $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\gamma}}$ that are called the estimated BLUPs or EBLUPs of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$.

A further issue is the assumption of the residual autocorrelation structure, as the true autocorrelation is usually unknown. To this regard Krivobokova/Kauermann (2007) show that the results are robust, even if the assumed autocorrelation structure is not equal to the real (but unknown) autocorrelation of the cycle. However, the assumed autocorrelation may not be too different from the true one. A further advantage of the interpretation of splines as linear mixed models is that the estimations are almost independent of the number of selected knots as long as it is not too low. Without the interpretation as a mixed model the flexibility of a penalized tp-spline increases with the number of knots (e.g. Fahrmeir et al. 2009 p.301). Kauermann/Opsomer (2011) show that the estimation within a mixed model framework also yields a higher penalization parameter, when the number of knots is increased. This increase of the penalization just compensates the higher number of knots.

2.2.2 Time-varying penalization parameters

The coefficient of the highest polynomial of a penalized tp-spline can change at every knot. To this point the penalization parameter controls the extent of the change and thus the smoothness of the trend. Low values of λ allow rather large changes of the coefficient for the highest polynomial, which results in a close fit of the trend to the data. However, if the trend is supposed to be smooth, but there is an abrupt structural break in the series, then it is not sufficient to choose one single value for the penalization parameter. To achieve a smooth trend function a high penalization has to be selected, while the adaptation to the break point would require a low value of λ. This problem can be solved by selecting different values for the penalization parameter. The product of λ and the i^{th} diagonal element of the penalty matrix $\boldsymbol{K}$ yields the penalization of the coefficient δ_i, where the coefficients $\delta_{l+2}, ..., \delta_d$ regulate the change of the trend function at the knots $\kappa_2, ..., \kappa_{m-1}$. It follows that if the trend shall be smooth in general, but change abruptly at a certain knot, then the scalar penalization parameter λ can be replaced by a vector of penalization parameters

$$\boldsymbol{\lambda} = (\underbrace{0, ..., 0}_{l+1}, \underbrace{\lambda_1, ..., \lambda_{m-2}}_{m-2})' \in \mathbb{R}^{d\times 1}. \tag{2.14}$$

The coefficients $\delta_1, ..., \delta_{l+1}$ do not need to be penalized so that the first $l+1$ elements of $\boldsymbol{\lambda}$ are set to zero. From this setting the penalization can be selected separately at each knot, where λ_{j-1} controls the penalization at the knot κ_j.

This concept can be transferred to the mixed model framework. This allows estimating the penalization even when the series contains structural breaks. So far, it is assumed that the variance of the random effects and thus the penalization is the same at every point in time. Now, to estimate varying penalization parameters, the variance of the fixed effects is allowed to change over time. A time-varying variance is already suggested by Crainiceanu et al. (2005), who model the variance as a function that smoothly changes over time. But to avoid distortions by structural breaks, it is not useful to model a smoothly varying penalization parameter. Instead, the penalization parameter shall change abruptly at the break points. This can be done as follows:

Assume that there is a structural break at time t^* so that the time series abruptly changes from period t^*-1 to t^*. If equidistant knots are chosen, there will likely be no knots lying exactly on the points in time t^*-1 and t^*. This should be the case in order to account exactly for the break in the data and to generate a trend function that adapts accurately to the break. Consequently, knots should be inserted at t^*-1 and t^* in addition to the equidistant knots. Moreover, it has to be checked if there is a knot lying between t^*-1 and t^*. In this case either this knot has to be removed or the number of knots m has to be changed. If m equidistant knots are chosen, the number of knots extends to $m+2$ so that the sequence of knots is now given to $\kappa_1, \kappa_2, ..., t^*-1, t^*, ...\kappa_m$. One exception arises for $m = T$ equidistant knots, as then there is a knot exactly at the points in time $t = 1, ..., T$. In this case no additional knots have to be inserted.

To achieve a penalization that differs at the break from the one for the rest of the series, the variance of the random effects is allowed to change at the knots t^*-1 and t^*. Thus, the variance of the random effects at the break is now given to υ^2, where the variance is still τ^2 at all other knots. The covariance matrix of the random effects is then given to

$$\boldsymbol{G} = \text{diag}(\underbrace{\tau^2, ..., \tau^2}_{\text{s}-1}, \upsilon^2, \upsilon^2, \underbrace{\tau^2, ..., \tau^2}_{\text{m}-\text{s}-1}). \tag{2.15}$$

s is the number of knots before t^*-1. Since there are two different variances of the random effects, two penalization parameters are obtained

$$\lambda_1 = \frac{\sigma^2}{\tau^2} \text{ and } \lambda_2 = \frac{\sigma^2}{\upsilon^2}.$$

λ_2 regulates the penalization at the knots t^* and t^*-1, while λ_1 determines how the trend function can change at all other knots. This model allows estimating a penalization at the time of the break that is different from the penalization for the rest of the series. Thus, the model is able to account for the structural break. Section 2.3 gives empirical examples and shows that this model indeed yields a penalization that adapts to the break.

To simplify the maximization of the restricted log-likelihood the variance υ^2 can also be a priori set to very high values, which equivalently yields a low penalization at the break. The maximization of the restricted log-likelihood yields the estimators $\hat{\boldsymbol{R}}$ and $\hat{\boldsymbol{G}}$, where $\hat{\boldsymbol{G}} = \text{diag}(\hat{\tau}^2, ..., \hat{\tau}^2, \hat{\upsilon}^2, \hat{\upsilon}^2, \hat{\tau}^2 ..., \hat{\tau}^2)$. Then the matrix

$$\hat{\boldsymbol{B}} = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \hat{\boldsymbol{G}}^{-1} \end{pmatrix}$$

is defined. Afterwards the parameters of the penalized tp-spline with the time-varying penalization can be estimated according to (2.9) by

$$\hat{\boldsymbol{\theta}} = \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\gamma}} \end{bmatrix} = (\boldsymbol{Z}'\hat{\boldsymbol{R}}^{-1}\boldsymbol{Z} + \hat{\boldsymbol{B}})^{-1}\boldsymbol{Z}'\hat{\boldsymbol{R}}^{-1}\boldsymbol{y}. \tag{2.16}$$

A special characteristic of the time-varying penalization is that the estimated trend can change its slope abruptly after the break. If a very low penalization is estimated at the knots $t^* - 1$ and t^*, then the trend can change abruptly from period $t^* - 1$ to t^* and from period t^* to $t^* + 1$. Thus, the growth of the trend in the period after the break can strongly differ from the one in the period before the break. This is a positive feature, since structural breaks might influence the long run development of a time series. Moreover, if it is assumed that a break heavily changes the structural conditions that underlie a time series, then this approach can also be extended by assuming a specific variance of the random effects after the break. Consequently, the general structure of the trend would be allowed to change after the structural break.

2.2.3 The optimal degree of the tp-spline

One remaining question concerns the optimal degree of the spline. The aim is to construct a trend function, whose estimated slope can change abruptly at a single point in time. The best way to achieve this feature is to use a tp-spline of degree one. A tp-spline of degree l is a continuous function that changes the coefficient of the l^{th} polynomial at each knot. With regard to a tp-spline of degree one this means the estimated trend is just a line that changes its slope at every knot. The change of the slope is regulated by the penalization parameter λ. High values of λ induce that the slope can only change slightly, while low values of λ allow large changes of the slope. If the penalization adopts very low values at the break, then the slope of the tp-spline of degree one can change abruptly and adapts to the break. The same is not true for tp-splines of degrees $l \geq 2$. As only the coefficient of the highest polynomial is able to change at each knot, these splines contain coefficients of lower polynomials that do not change, which hinders an abrupt change of the trend. This is shown in Figure 2.1, which displays trend estimations for a series that contains a break. The trend is estimated with tp-splines of degrees one to four using a time-varying penalization to account for the break point. The tp-spline of degree one changes its slope accurately at the break, while all others start rising their slopes too soon and are distorted.

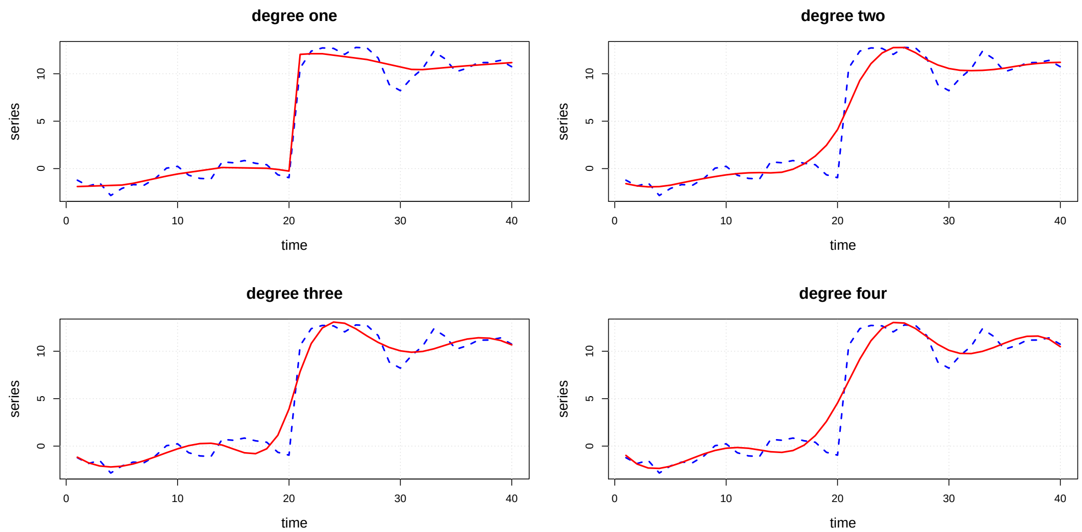

Fig. 2.1: Trend estimation with splines of different degrees for series with a structural break

2.3 Empirical application

2.3.1 German GDP

In this section the described methods are applied to real time series. The German reunification offers good examples for structural breaks, since most economic variables change abruptly at the beginning of 1991. To this regard data of the real German GDP from the first quarter 1970 to the second quarter 2013 are considered[2]. The data are adjusted for season and calendar effects.

Fig. 2.2: Seasonally adjusted real German GDP from 1970-2013

[2]The data are from the German Federal Statistical Office.

The analysis of this series is problematic, since it refers only to West Germany before 1991 and contains both East and West Germany afterwards. The German GDP is displayed in Figure 2.2. Obviously, there is an abrupt fall from the fourth quarter 1990 to the first quarter 1991. This is due to a change of the base period used for the calculation as well as the reunification and may not be interpreted as a decline of the German GDP.

However, this fall causes distortions, when the trend of this series is estimated by a fixed penalization. This is shown in the left plot of Figure 2.3, which displays the estimated trend of the GDP. The right plot shows the results, when a flexible penalization is used. In both cases the trend is estimated with a tp-spline of degree one, where the knots are set at every point in time. Note that this specification is equal to the Hodrick-Prescott filter (Proietti/Luati 2007, Paige 2010). If this special spline is estimated within a mixed model assuming that the error term is autocorrelated, then it can be seen as an extended version of the HP-filter taking into account an autocorrelated cyclical component. In this case the error term, i.e. the business cycle is assumed to follow an AR(1)-process.

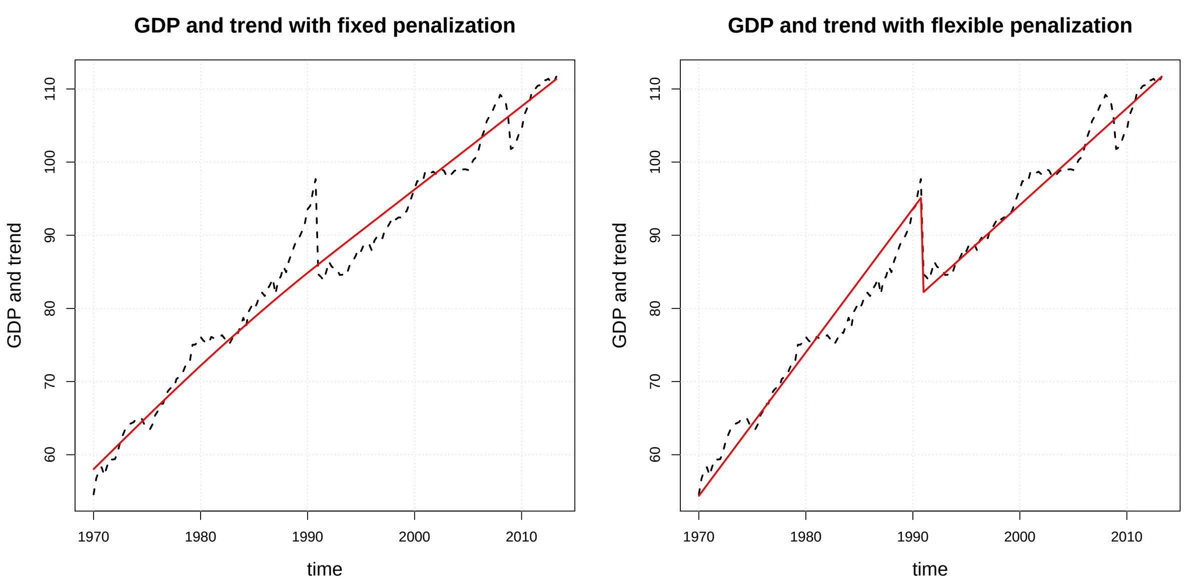

Fig. 2.3: Trend estimation for the real German GDP with fixed and flexible penalization

Clearly, the spline with the fixed penalization cannot account for the break and is distorted. The spline with the flexible penalization is able to adapt to the break and yields a much more reliable estimation for the trend of the German GDP. In both cases the penalization parameter λ is estimated infinitely high, which results in the (almost) linear shape of the trend. Moreover, the flexible penalization at the break allows the estimated trend to abruptly change its slope after the reunification. This is reasonable, as also East Germany is included in the GDP after the reunification. The slope of the estimated trend changes after the reunification. While it is 0.49 per quarter before 1991 it is only 0.33 afterwards. Thus, this model indicates that the trend growth has declined after the reunification.

To show the effects of another autocorrelation structure of the error term, the trend of the GDP is again estimated assuming that the business cycle follows an AR(2)-process. The results are shown in Figure 2.4:

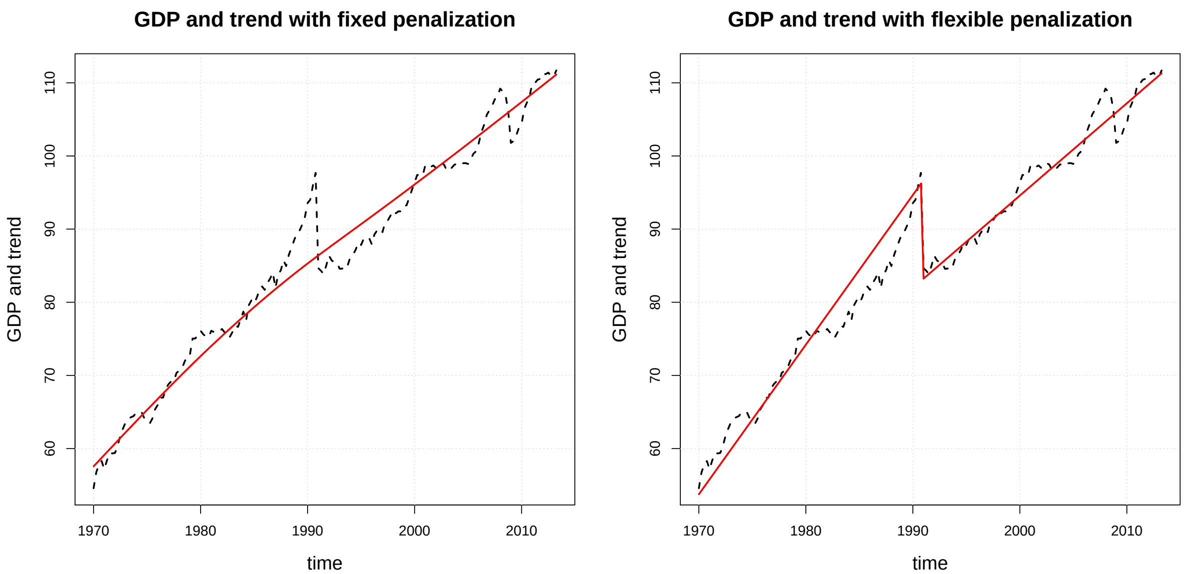

Fig. 2.4: Trend estimation for the real German GDP with fixed and flexible penalization

There are only slight differences to the case of the AR(1) error term. There might be a bit more curvature in the trend of the fixed penalization, while for the flexible penalization the slope before the reunification is with 0.51 per quarter a bit higher and with 0.32 after the reunification a bit lower than before. Beside the trend component, also the estimated business cycle is affected by the structural break. This is shown in Figure 2.5, that plots the estimated business cycles for fixed and flexible penalization in the case of an AR(1) error term.

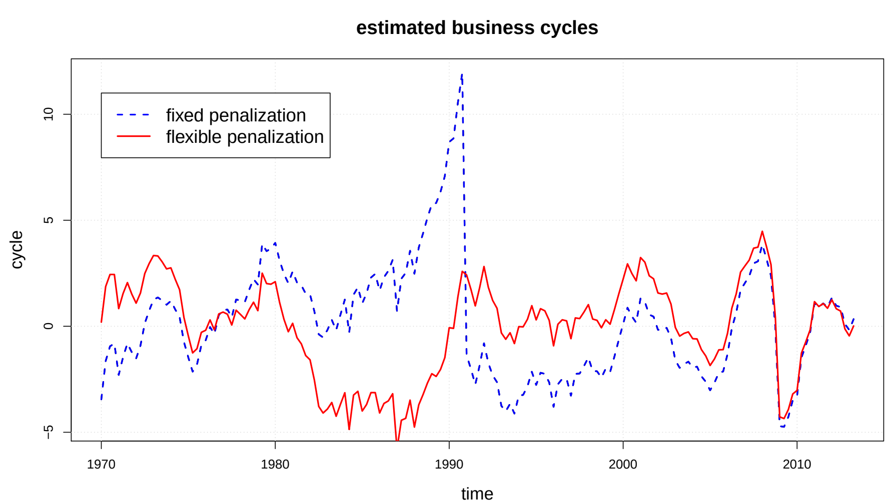

Fig. 2.5: Estimated cyclical component for fixed and flexible penalization

With the exception of the years 2009-2013 the estimated business cycles deviate. The largest difference is around the years 1990 and 1991, as the cycle according to the flexible penalization excludes the break after the reunification. Moreover, the cyclical component according to the flexible penalization exhibits far higher values than the one of the fixed penalization during the years 1970-1978 and 1991-2008, while it is lower during the period 1978-1991. This example shows, that it is possible to derive a data driven estimation for trend and cycle of the German GDP even for time series that start before 1991, without having distortions by the reunification. To check for the robustness of the assumed autocorrelation structure, finally the autocorrelation- and partial autocorrelation function of the business cycle according to the flexible penalization of the estimation with the AR(1) residual structure are considered:

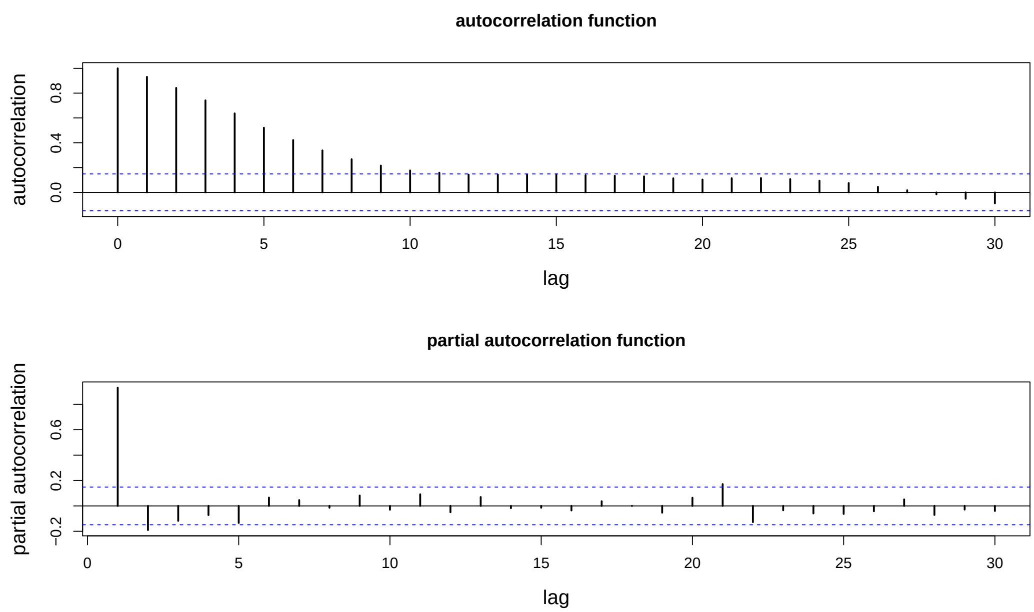

Fig. 2.6: Autocorrelation- and partial autocorrelation function for the estimated residuals

For the estimation within the mixed model framework the cyclical component is assumed to follow an AR(1)-process. The autocorrelation function of the resulting cycle exhibits a damped behavior, while the partial autocorrelation function is only highly significant for the first lag and there is almost no significance after the second lag. Thus, the assumption of a cycle that follows an AR(1)-process seems not to be very far away from the resulting autocorrelation structure of the estimated cycle.

2.3.2 German employable population

Another example is the employable population[3] of Germany from 1959 to 2010.[4] The data exhibit an abrupt rise from the year 1990 to 1991. This is caused by the German reunification, which raised the employable population abruptly from about 43 million to more than 55 million. Again the trend is estimated using a penalized tp-spline of order one and as many knots as observations, where the error term is assumed to follow an AR(1)-process. The estimation is done with a fixed and a flexible penalization. The resulting trend of the fixed penalization is shown on the left side of Figure 2.7. Because of the abrupt rise, the estimated trend is just a straight line. This is obviously far away from the real trend, since the data suggest an increasing trend before, and a decreasing trend after the reunification.

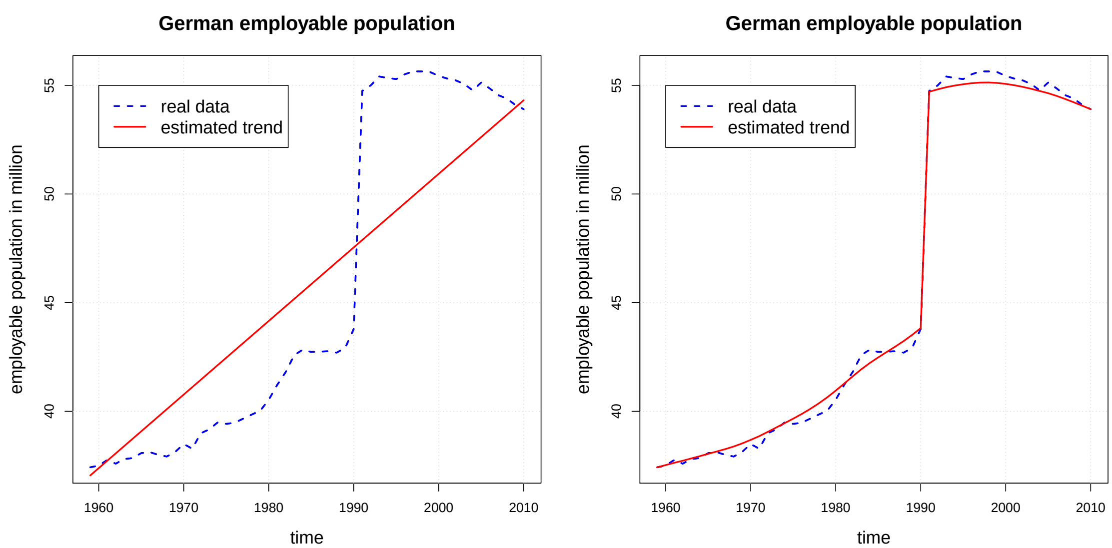

Fig. 2.7: Estimated trend of the employable population with fixed and flexible penalization

The plot on the right side of Figure 2.7 shows the estimated trend for the flexible penalization, where the penalization is allowed to be flexible at the knots 1990 and 1991. Now, the estimated trend is not a straight line any more, but it adapts to the break after the reunification. Moreover, the trend increases before the year 1990 and decreases afterwards, which is in line with the behavior of the expected trend component now.

2.4 Conclusion

This chapter shortly summarizes penalized splines with a truncated polynomial basis and how they can be incorporated into a mixed model framework. The interpretation as a linear mixed model allows a data driven derivation of the penalization parameter and helps

[3]The employable population is the population of age between 15 and 65.
[4]The data are from the German Federal Statistical Office.

to overcome the drawback of the subjective choice of the penalization. This approach enables to use an autocorrelated residual structure, which is reasonable, since the cyclical component is seldom expected to be white noise in economic time series. Beside the penalization parameter, penalized splines include more model parameters like the degree of the basis functions, the number of knots and the residual autocorrelation structure. However, in most cases these parameters only slightly influence the resulting trend estimation as shown in Ruppert (2002), Krivobokova/Kauermann (2007), Claeskens et al. (2009) and Kauermann/Opsomer (2011).

Nevertheless, penalized tp-splines within mixed models can fail to estimate trend and cycle, when time series exhibit structural breaks. It is shown in section 2.3 that such breaks can cause distortions in the estimation. In this regard, this chapter extends the mixed model framework of penalized splines to account for structural breaks. The variance of the random effects is allowed to change at the break and in the period before. This yields a time-varying penalization that enables the trend function to adapt to the break, where the penalization can still be estimated by maximum likelihood. Moreover, it is shown that a tp-spline of degree one is most suitable for estimating the trend of time series containing structural breaks. tp-splines of higher degrees are not able to change abruptly enough.

The algorithm described in this chapter is demonstrated on empirical examples. The trend of the real German GDP from 1970-2013 is estimated, where the data exhibit a break in the year 1991 due to the reunification. While the setting with the fixed penalization is distorted downwards, the approach with the flexible penalization can account for the break. Given the assumptions of an error term that follows an AR(1)-, or an AR(2)-process, the estimation with the flexible penalization yields an infinitely high penalization parameter and a linear trend. Moreover, the estimated trend growth rate declines after the reunification.

Thus, based on the ideas of Crainiceanu et al. (2005), who suggest to let the penalization parameter smoothly change over time, and Razzak/Richard (1995), who use a flexible penalization within the Hodrick-Prescott filter to account for breaks, this chapter offers a useful tool to achieve a data driven estimation of trend and cycle for series that exhibit structural breaks like most German data due to the reunification.

3 Penalized Splines as Wiener-Kolmogorov Filters Investigating the long run development of the German Economy

Summary

For the Hodrick-Prescott filter it is known that it is equal to an optimal Wiener-Kolmogorov filter, when the trend is a second fold integrated random walk and the cycle is just white noise. However, the Hodrick-Prescott filter can be extended, when it is written as a penalized spline and transferred into a linear mixed model framework. This chapter shows that in this case the optimal Wiener-Kolmogorov filter for a second fold integrated random walk as the trend and a stationary autocorrelated cycle arises. The mixed model framework allows the estimation of the parameters by maximum likelihood. This method is employed to examine the long run development of the German GDP as well as the industrial production index. The outcome is that there is a very smooth long run development in the German economy, which is in line with existing literature like Flaig (2005).

3.1 Introduction

On purpose to estimate trend and cycle of economic time series, the predominantly used technique is the Hodrick-Prescott filter (Hodrick/Prescott 1997), which traces back to the ideas of Whittaker (1923), Henderson (1924) and Leser (1961). The output of this filter depends on the choice of a penalization parameter λ that controls the smoothness of the estimated series. There are no general rules how to select a suitable value for λ, especially as there is no general definition of trend and cycle (Stamfort 2005 p.7). The most commonly used value traces back to the suggestion of Hodrick/Prescott (1997), who propose to use $\lambda = 1600$ for quarterly data. Their suggestion is based on the equivalence of the Hodrick/Prescott filter to an optimal Wiener-Kolmogorov filter (Whittle 1983, Bell 1984) for a second fold integrated random walk as the trend and a white noise process as the cyclical component, as well as subjective assumptions about the variance of the growth rates of trend and cycle. As these assumptions are rather restrictive, the suggestion of Hodrick/Prescott is criticized as dubious (Danthine/Giradin 1989). Additionally, the choice of $\lambda = 1600$ might be too low for most economic time series (Mc Callum 2000, Flaig 2012). Furthermore, the suggestion is criticized as not data driven (Schlicht 2005, Kauermann et al. 2011).

To this point Schlicht (2005) shows how to write the Hodrick-Prescott filter as a linear mixed model in order to estimate the penalization parameter. However, this approach is limited to a white noise residual structure and thus a white noise business cycle, what is seldom true for economic time series. Nevertheless, this shortcoming can be corrected, as the Hodrick-Prescott filter belongs to the class of penalized splines (Paige 2010, for penalized splines see O'Sullivan 1986, Eilers/Marx 1996, Ruppert et al. 2003). Penalized splines offer the advantage to estimate trend and cycle data driven, where the most common methods are generalized cross validation (Hastie/Tibshirani 1990, also Ruppert 2002), or the incorporation into a mixed model framework (e.g. Brumback et al. 1999). Both methods allow for an autocorrelated residual structure (see Kohn et al. (1992) or Wang (1998) for the generalized cross validation with correlated errors). But while generalized cross validation is very sensitive about the specification of the autocorrelation structure (Opsomer et al. 2001, Proietti 2005), the mixed model approach appears to be relatively robust, as long as the deviation between assumed and true (but unknown) autocorrelation structure is not too large (Krivobokova/Kauermann 2007). Thus, it immediately follows that the Hodrick-Prescott filter can also be written as a linear mixed model with an autocorrelated error term. This chapter shows that in this case the optimal Wiener-Kolmogorov filter for a second fold integrated random walk as the trend and an autocorrelated cycle arises. The mixed model framework allows the estimation of the model parameters by maximum likelihood.

A further issue that is summarized in this chapter is the calculation of confidence intervals for the estimated trend. As penalized splines and the Hodrick-Prescott filter are cases of the so called ridge regression, they do not yield unbiased estimates (e.g. Ruppert et al. 2003 p.133 et seq.). However, this shortcoming can be avoided when the mixed model interpretation of splines and the Hodrick-Prescott filter is employed, which yields a further advantage of the trend estimation within a mixed model framework.

The first sections of this chapter briefly summarize the Hodrick-Prescott filter and penalized splines with a truncated polynomial basis (Brumback et al. 1999) and describe the link between both methods. Moreover, it is explained how penalized splines can be interpreted as linear mixed models. Furthermore, the basic aspects of the calculation of confidence intervals for splines are discussed. Afterwards it is shown how these methods fit into the framework of the Wiener-Kolmogorov filter. Finally, the mixed model framework of the Hodrick-Prescott filter is used to derive an estimation for the trend component of the German GDP as well as of the industrial production index.

3.2 Model framework

3.2.1 The Hodrick-Prescott filter

The Hodrick-Prescott filter (henceforth denoted as HP-filter) decomposes a time series $\{y_t\}_{t=1}^T$ into two components, i.e. $y_t = \mu_t + c_t$. μ_t is regarded as the trend while c_t represents the rest, usually the sum of cycle and irregular effects. The trend μ_t is estimated by solving the following minimization problem:

$$\min_{\mu_t} \sum_{t=1}^{T} (y_t - \mu_t)^2 + \lambda \sum_{t=2}^{T-1} \left[(\mu_{t+1} - \mu_t) - (\mu_t - \mu_{t-1})\right]^2 . \tag{3.1}$$

The first part of the minimization problem causes a close fit of the estimation to the original data, while the second part penalizes the volatility of the trend. The parameter λ controls the smoothness of the trend function. Increasing λ makes the trend component become less flexible. For $\lambda = 0$ the trend is equal to the original series, whereas for $\lambda \to \infty$ it is just a straight line (Stamfort 2005 p.25). The solution to the minimization in (3.1) can be expressed in matrix notation (McElroy 2008, Faig 2012 p.16), which allows a fast and easy calculation:

$$\hat{\boldsymbol{\mu}} = (\boldsymbol{I} + \lambda \boldsymbol{\Delta}'\boldsymbol{\Delta})^{-1}\boldsymbol{y}, \tag{3.2}$$

where $\boldsymbol{\Delta} \in \mathbb{R}^{(T-2)\times T}$ is a differencing matrix, such that the product of $\boldsymbol{\Delta}$ and $\boldsymbol{y}$ yields the second differences of $\boldsymbol{y}$. $\boldsymbol{\Delta}$ is described by formula (3.31) in section 3.3.

Furthermore, for $T \to \infty$ the HP-filter approximatively renders a stationary cyclical component, even if the observed series $\{y_t\}_{t=1}^T$ is integrated up to order four. Disadvantages of the HP-filer are, that it might induce phase shifts (King/Rebelo 1993) as well as spurious cycles (Harvey/Jaeger 1993, Cogley/Nason 1995). This means the filter might change the original temporal relation of the filtered to other series and the component c_t might exhibit cyclical behavior, even if the observed series contains no cycle.

3.2.2 Penalized tp-splines

This section describes briefly penalized splines. For a detailed discussion see e.g. Fahrmeir et al. (2009) or Ruppert et al. (2003). Although there are many different types of splines, it is focused on truncated polynomial splines (Brumback et al. 1999, henceforth denoted as tp-splines). Even if tp-splines tend to be numerical instable, they are advantageous due to their relative easy implementation and interpretation as well as their straight link to linear mixed models and the HP-filter. Estimating the trend component with a tp-spline means the trend function is modelled in dependence of time t. After dividing the variable t, $t = 1, ..., T$, into $m-1$ intervals by setting m knots $1 = \kappa_1 < \kappa_2 < ... < \kappa_m = T$ a tp-spline of degree l for a time series $\{y_t\}_{t=1}^T$ can be written to:

$$y_t = f(t) + \varepsilon_t = \delta_1 + \delta_2 t + ... + \delta_{l+1} t^l + \delta_{l+2}(t-\kappa_2)_+^l + ... + \delta_d (t-\kappa_{m-1})_+^l + \varepsilon_t, \tag{3.3}$$

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

where ε_t denotes the error term that represents the business cycle and $d = m + l - 1$. In matrix notation the model is defined to:

$$\boldsymbol{y} = \boldsymbol{Z\delta} + \boldsymbol{\varepsilon}, \tag{3.4}$$

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

Here, $\boldsymbol{\delta} = (\delta_1, ..., \delta_d)'$, $\boldsymbol{\varepsilon} = (\varepsilon_1, ..., \varepsilon_T)'$ and $\boldsymbol{y} = (y_1, ..., y_T)'$. tp-splines can be interpreted as a continuous function of piecewise defined polynomials of degree l. The coefficient of the highest polynomial changes at every knot because of the truncated polynomials, which allows a high flexibility of $f(t)$. Consequently, the smoothness of the trend can be determined by controlling the parameters of the truncated polynomials $\delta_{l+2}, ..., \delta_d$, since they regulate the change of the coefficient of the highest polynomial (e.g. Fahrmeir et al. 2009 p.308). This can be done by estimating the vector of coefficients by minimizing the penalized least squares criterion.

$$PLS(\lambda) = \sum_{t=1}^{T} [y_t - f(t)]^2 + \lambda \sum_{j=l+2}^{d} \delta_j^2. \tag{3.5}$$

The first part of $PLS(\lambda)$ aims at a close fit of the trend to the observed series, while the second part penalizes a too high volatility. This tradeoff is solved by the parameter λ that puts weight on the second part. Increasing the value for λ reduces the volatility of the trend. The solution of (3.5) is given to (e.g. Fahrmeir et al. 2009 p.313)

$$\hat{\boldsymbol{\delta}} = (\boldsymbol{Z}'\boldsymbol{Z} + \lambda\boldsymbol{K})^{-1}\boldsymbol{Z}'\boldsymbol{y}, \text{ where } \boldsymbol{K} = \text{diag}(\underbrace{0, ..., 0}_{l+1}, \underbrace{1, ..., 1}_{m-2}). \tag{3.6}$$

$$\text{so that } \hat{\boldsymbol{y}} = \boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z} + \lambda\boldsymbol{K})^{-1}\boldsymbol{Z}'\boldsymbol{y}. \tag{3.7}$$

Given formula (3.7) it can be shown that for a certain selection for the parameters m and l the penalized tp-spline is equal to the HP-filter (Paige 2010). Setting $l = 1$ and knots at every point in time t, $t = 1, 2, ..., T$, implies that the design matrix $\boldsymbol{Z}$ is quadratic and invertible (c.f. Paige 2010 p.870). In this case it is generally defined as

$$\boldsymbol{Z} = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & \dots & 0 & 0 \\ 1 & 2 & 0 & 0 & 0 & \dots & 0 & 0 \\ 1 & 3 & 1 & 0 & 0 & \dots & 0 & 0 \\ 1 & 4 & 2 & 1 & 0 & \dots & 0 & 0 \\ 1 & 5 & 3 & 2 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 1 & T-1 & T-3 & T-4 & T-5 & \dots & 1 & 0 \\ 1 & T & T-2 & T-3 & T-4 & \dots & 2 & 1 \end{pmatrix} \in \mathbb{R}^{T \times T}.$$

Given this special structure of $\boldsymbol{Z}$, formula (3.7) can also be expressed as (a derivation is provided in appendix 3.A):

$$\hat{\boldsymbol{y}} = (\boldsymbol{I} + \lambda\boldsymbol{Z}'^{-1}\boldsymbol{K}\boldsymbol{Z}^{-1})^{-1}\boldsymbol{y}. \tag{3.8}$$

Taking into account that $\hat{\boldsymbol{y}} = \hat{\boldsymbol{\mu}}$ and that $\boldsymbol{Z}'^{-1}\boldsymbol{K}\boldsymbol{Z}^{-1} = \boldsymbol{\Delta}'\boldsymbol{\Delta}$ (a proof is given in appendix 3.B), it becomes obvious that the HP-filter is identical to a penalized tp-spline of order one and with knots at every observed point in time $t = 1, ..., T$. Note that this formulation is also equivalent to the so called local linear model (Proietti 2007). The equivalence between tp-splines and the HP-filter allows a useful interpretation of the HP-filter. The trend generated by the HP-filter is a continuous connection of lines that can change their slope at the points in time $t = 2, 3, .., T-1$. The penalization parameter λ controls to what extend the slope of the lines can change at these points in time. Choosing high values for λ means the slope can change only slightly, which results in a smooth estimated trend. Furthermore, the equivalence of the HP-filter and a penalized spline implies that the mixed model framework with an autocorrelated residual structure can also be applied to the HP-filter to derive a data driven estimation for its penalization parameter.

3.2.3 Splines within a mixed model framework

Penalized tp-splines can be interpreted as a linear mixed model in order to derive a data driven estimation of λ (for a detailed discussion of mixed models see Searle et al. 1992, Vonesh/Chinchilli 1997, Pinheiro/Bates 2000 or McCulloch/Searle 2001). A tp-spline within a mixed model framework is defined to (e.g. Ruppert et al. 2003 p.108)

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{U}\boldsymbol{\gamma} + \boldsymbol{\varepsilon} = \boldsymbol{Z}\boldsymbol{\theta} + \boldsymbol{\varepsilon}, \tag{3.9}$$

$$\text{where } \boldsymbol{X} = \begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & 2 & \dots & 2^l \\ \vdots & \vdots & \ddots & \vdots \\ 1 & T & \dots & T^l \end{pmatrix} \text{ and } \boldsymbol{U} = \begin{pmatrix} (1-\kappa_2)_+^l & \dots & (1-\kappa_{m-1})_+^l \\ \vdots & \ddots & \vdots \\ (T-\kappa_2)_+^l & \dots & (T-\kappa_{m-1})_+^l \end{pmatrix}.$$

Consequently, $\boldsymbol{Z} = [\boldsymbol{X}, \boldsymbol{U}]$, $\boldsymbol{\theta}' = [\boldsymbol{\beta}', \boldsymbol{\gamma}']$, $\boldsymbol{\beta} \in \mathbb{R}^{(l+1)\times 1}$ and $\boldsymbol{\gamma} \in \mathbb{R}^{(m-2)\times 1}$. Moreover, it is assumed that $\boldsymbol{\varepsilon} \sim \mathrm{N}(0, \boldsymbol{R})$ and $\boldsymbol{\gamma} \sim \mathrm{N}(0, \boldsymbol{G})$, so that in general an autocorrelated residual structure can be allowed. When the tp-spline is written as a mixed model, it can equivalently be interpreted as a hierarchical model (e.g. Fahrmeir et al. 2009 p.261):

$$\boldsymbol{y}|\gamma \sim \mathrm{N}(\boldsymbol{X\beta} + \boldsymbol{U\gamma}, \boldsymbol{R}) \text{ and the marginal distribution } \boldsymbol{\gamma} \sim \mathrm{N}(0, \boldsymbol{G}).$$

Furthermore, the spline can be described by a marginal model:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}^* \text{ with } \boldsymbol{\varepsilon}^* = \boldsymbol{U\gamma} + \boldsymbol{\varepsilon}.$$

The distribution of $\boldsymbol{y}$ is then defined by $\boldsymbol{y} \sim \mathrm{N}(\boldsymbol{X\beta}, \boldsymbol{V})$, where $\boldsymbol{V} = \boldsymbol{R} + \boldsymbol{UGU'}$. Furthermore, it is assumed that $\boldsymbol{R} = \sigma^2\boldsymbol{\Omega}$ and $\boldsymbol{G} = \mathrm{diag}(\tau^2)$ (e.g. Kauermann et al. 2011), i.e. the error term is autocorrelated with a constant variance σ^2 and the autocorrelation matrix $\boldsymbol{\Omega}$ and the parameters $\boldsymbol{\gamma}$ are uncorrelated and have the constant variance τ^2. Given the hierarchical model of $\boldsymbol{y}$ the parameter vector $\boldsymbol{\theta}$ can be estimated by maximizing the resulting log-likelihood function with respect to $\boldsymbol{\theta}$ (e.g. Ruppert et al. 2003 p.100):

$$\log L(\boldsymbol{\theta}) = -\frac{1}{2}(\boldsymbol{y} - \boldsymbol{Z\theta})'\boldsymbol{R}^{-1}(\boldsymbol{y} - \boldsymbol{Z\theta}) - \frac{1}{2}\gamma'\boldsymbol{G}^{-1}\gamma. \tag{3.10}$$

This is equivalent to minimizing (e.g. Fahrmeir et al. 2009, Kauermann et al. 2011)

$$\min_{\boldsymbol{\theta}}(\boldsymbol{\theta}) = (\boldsymbol{y} - \mathbf{Z}\boldsymbol{\theta})'\boldsymbol{\Omega}^{-1}(\boldsymbol{y} - \boldsymbol{Z\theta}) + \lambda\gamma'\gamma, \tag{3.11}$$

$$\text{where } \lambda = \frac{\sigma^2}{\tau^2}.$$

As $\boldsymbol{G} = \mathrm{diag}(\tau^2)$, the solution of the maximization of formula (3.10) is (e.g. Robinson 1991, Hayes/Haslett 1999, also Ruppert et al. 2003 p.100)

$$\tilde{\boldsymbol{\theta}} = \begin{bmatrix} \tilde{\boldsymbol{\beta}} \\ \tilde{\gamma} \end{bmatrix} = (\boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{Z} + \frac{1}{\tau^2}\boldsymbol{K})^{-1}\boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{y}, \tag{3.12}$$

where the estimators $\tilde{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{\gamma}}$ are the best linear unbiased predictors (BLUPs) of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$. BLUP means for any vectors $\boldsymbol{s}, \boldsymbol{v} \in \mathbb{R}^{T\times 1}$ the predictors $\tilde{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{\gamma}}$ minimize (e.g. Ruppert et al. 2003 p.99):

$$\mathbb{E}\left[(\boldsymbol{s}'\boldsymbol{X}\tilde{\boldsymbol{\beta}} + \boldsymbol{v}'\boldsymbol{U}\tilde{\boldsymbol{\gamma}}) - (\boldsymbol{s}'\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{v}'\boldsymbol{U}\boldsymbol{\gamma})\right]^2$$

subject to

$$\mathbb{E}\left[\boldsymbol{s}'\boldsymbol{X}\tilde{\boldsymbol{\beta}} + \boldsymbol{v}'\boldsymbol{U}\tilde{\boldsymbol{\gamma}}\right] = \mathbb{E}\left[\boldsymbol{s}'\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{v}'\boldsymbol{U}\boldsymbol{\gamma}\right].$$

In most situations the covariance matrices $\boldsymbol{R}$ and $\boldsymbol{G}$ are unknown and have to be estimated. To this regard both matrices are written in dependence of a vector of parameters $\boldsymbol{\vartheta}$, where $\boldsymbol{\vartheta}$ depends on the assumed autocorrelation structure of $\boldsymbol{\varepsilon}$. From this the log-likelihood is derived by the interpretation as a marginal model. It is given except of a constant term to (e.g. Ruppert et al. 2003 p.100-101):

$$l(\boldsymbol{\beta}, \boldsymbol{\vartheta}) = -\frac{1}{2}\left[\log(|\boldsymbol{V}(\boldsymbol{\vartheta})|) + (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'\boldsymbol{V}(\boldsymbol{\vartheta})^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})\right]. \tag{3.13}$$

Differentiating and solving with respect to $\boldsymbol{\beta}$ yields

$$\tilde{\boldsymbol{\beta}}(\boldsymbol{\vartheta}) = [\boldsymbol{X}'\boldsymbol{V}(\boldsymbol{\vartheta})^{-1}\boldsymbol{X}]^{-1}\boldsymbol{X}'\boldsymbol{V}(\boldsymbol{\vartheta})^{-1}\boldsymbol{y}. \tag{3.14}$$

Reinserting into (3.13) yields the profile log-likelihood:

$$l_p(\boldsymbol{\vartheta}) = -\frac{1}{2}\left(\log|\boldsymbol{V}(\boldsymbol{\vartheta})| + [\boldsymbol{y} - \boldsymbol{X}\tilde{\boldsymbol{\beta}}(\boldsymbol{\vartheta})]'\boldsymbol{V}(\boldsymbol{\vartheta})^{-1}[\boldsymbol{y} - \boldsymbol{X}\tilde{\boldsymbol{\beta}}(\boldsymbol{\vartheta})]\right). \tag{3.15}$$

Instead of the profile log-likelihood in most applications the restricted log-likelihood (e.g. Searle et al. 1992) is maximized:

$$l_r(\boldsymbol{\vartheta}) = l_p(\boldsymbol{\vartheta}) - \frac{1}{2}\log|\boldsymbol{X}'\boldsymbol{V}(\boldsymbol{\vartheta})^{-1}\boldsymbol{X}|. \tag{3.16}$$

The restricted log-likelihood is more accurate in small samples, because it takes into account the degrees of freedom for of the fixed effects (Ruppert et al. 2003 p.101). The maximization of $l_r(\boldsymbol{\vartheta})$ with respect to $\boldsymbol{\vartheta}$ yields $\hat{\boldsymbol{\vartheta}}$ and thus the estimators $\hat{\boldsymbol{R}}$ and $\hat{\boldsymbol{G}}$. Because σ^2 and τ^2 are contained in $\boldsymbol{\vartheta}$ also $\hat{\lambda} = \dfrac{\hat{\sigma}^2}{\hat{\tau}^2}$ can be received. The estimators for the autocovariance matrix $\hat{\boldsymbol{R}}$ and $\hat{\tau}^2$ can be inserted into (3.12) which yields the estimators $\hat{\boldsymbol{\beta}}$ and $\hat{\gamma}$ that are called the estimated BLUPs or EBLUPs of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$.

An advantage of the estimation of tp-splines within the mixed model framework is that the results hardly depend on the number of knots. In general the flexibility of the spline increases when a higher number of knots is selected (e.g. Fahrmeir et al. 2009 p.301). However, Kauermann/Opsomer (2011) demonstrate that the mixed model framework compensates the higher number of knots by an increase of the penalization. As a consequence, the number of knots has no significant effect on the results, as long as it is not too low. Furthermore, Krivobokova/Kauermann (2007) demonstrate that the results are robust with regard to a misspecification of the residual correlation structure. However, the assumed correlation structure may not be too different from the true (but unknown) one.

3.2.4 Confidence intervals

The mixed model framework for splines offers a further advantage, when confidence intervals are calculated. To this point, first of all it is explained why it is not useful to derive confidence intervals from the normal spline model of section 3.2.2. For a vector $\boldsymbol{t} = (1, 2, ..., T)'$ the estimated tp-spline can be expressed as $\hat{\boldsymbol{y}} = \hat{f}(\boldsymbol{t}) = \boldsymbol{H}\boldsymbol{y}$, where $\boldsymbol{H} = \boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z} + \lambda\boldsymbol{K})^{-1}\boldsymbol{Z}'$. Without loss of generality it is assumed that $\boldsymbol{\varepsilon} \sim \mathrm{N}(\boldsymbol{0}, \sigma^2\boldsymbol{I})$ so that $\mathrm{Cov}(\boldsymbol{H}\boldsymbol{y}) = \sigma^2\boldsymbol{H}\boldsymbol{H}'$. Let $\boldsymbol{h_t} \in \mathbb{R}^{T\times 1}$ denote a vector containing the t^{th} row of $\boldsymbol{H}$, then it follows that $\hat{f}(t) = \boldsymbol{h_t}'\boldsymbol{y}$ and $\hat{f}(t) \sim \mathrm{N}\left(\mathbb{E}\left[\hat{f}(t)\right], \sigma^2||\boldsymbol{h_t}||^2\right)$. From this a variable with a standard normal distribution can be derived by (e.g. Ruppert at al. 2009 p.136)

$$\frac{\hat{f}(t) - \mathbb{E}\left[\hat{f}(t)\right]}{\sigma||\boldsymbol{h_t}||} \sim \mathrm{N}(0, 1). \tag{3.17}$$

For large datasets σ can be replaced by its estimation $\hat{\sigma}$. The confidence intervals are calculated by

$$\hat{f}(t) \pm z\left(1 - \frac{\alpha}{2}\right)\hat{\sigma}||\boldsymbol{h_t}||. \tag{3.18}$$

$z\left(1 - \frac{\alpha}{2}\right)$ is the corresponding quantile of the standard normal distribution. However, this confidence interval refers to $\mathbb{E}[\hat{f}(t)]$ and not to $f(t)$ (e.g. Ruppert et al. 2003 p.136). $\hat{f}(t)$ is estimated by penalized least squares, which is a special case of a ridge regression, that is biased (e.g. Fahrmeir et al. 2009 p.172). This can be demonstrated easily. While the ordinary least squares estimator of $\boldsymbol{\delta}$ is unbiased, the penalized least squares estimator is biased because

$$\mathbb{E}(\hat{\boldsymbol{\delta}}) = \mathbb{E}\left[(\boldsymbol{Z}'\boldsymbol{Z} + \lambda\boldsymbol{K})^{-1}\boldsymbol{Z}'\boldsymbol{y}\right] = (\boldsymbol{Z}'\boldsymbol{Z} + \lambda\boldsymbol{K})^{-1}\boldsymbol{Z}'\mathbb{E}(\boldsymbol{y}) = (\boldsymbol{Z}'\boldsymbol{Z} + \lambda\boldsymbol{K})^{-1}\boldsymbol{Z}'\boldsymbol{Z}\boldsymbol{\delta} \neq \boldsymbol{\delta}.$$

Thus, also the confidence interval in (3.18) is biased. However, this bias can be corrected by employing the mixed model interpretation of the penalized spline. As seen in section 3.2.3, the estimators derived from the maximum likelihood estimation are unbiased, so that it is better to construct confidence intervals from the mixed model framework. In the mixed model framework the parameter vector $\boldsymbol{\beta}$ is fixed, while $\boldsymbol{\gamma}$ is random. To get information about the precision of the parameter estimators one calculates the covariance

$$\mathrm{Cov}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}) = \mathrm{Cov}\left[\begin{pmatrix} \tilde{\boldsymbol{\beta}} - \boldsymbol{\beta} \\ \tilde{\gamma} - \gamma \end{pmatrix}\right] = \mathrm{Cov}\left[\begin{pmatrix} \tilde{\boldsymbol{\beta}} \\ \tilde{\gamma} - \gamma \end{pmatrix}\right]. \tag{3.19}$$

From formula (3.12) for the BLUPs of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ (3.19) can be calculated by (e.g. Ruppert et al. 2003 p.103)

$$\mathrm{Cov}\left[\begin{pmatrix} \tilde{\boldsymbol{\beta}} \\ \tilde{\gamma} - \gamma \end{pmatrix}\right] = (\boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{Z} + \frac{1}{\tau^2}\boldsymbol{K})^{-1}. \tag{3.20}$$

This relation can be motivated by the posteriori distribution of the Bayesian mixed model (see for example Fahrmeir et al. (2009) for a more detailed discussion). Inserting the estimated covariance matrices $\hat{\boldsymbol{R}}$ and $\hat{\boldsymbol{G}}$ yields the estimated covariance

$$\text{Cov}\left[\begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma} \end{pmatrix}\right] = (\boldsymbol{Z}'\hat{\boldsymbol{R}}^{-1}\boldsymbol{Z} + \frac{1}{\hat{\tau}^2}\boldsymbol{K})^{-1}. \tag{3.21}$$

Finally, to get inference about the precision of $\hat{f}(\boldsymbol{t})$ one calculates

$$\text{Cov}(\boldsymbol{Z}\hat{\boldsymbol{\theta}} - \boldsymbol{Z}\boldsymbol{\theta}) = \boldsymbol{Z}(\boldsymbol{Z}'\hat{\boldsymbol{R}}^{-1}\boldsymbol{Z} + \frac{1}{\hat{\tau}^2}\boldsymbol{K})^{-1}\boldsymbol{Z}'. \tag{3.22}$$

From (3.22) the standard deviation of $\hat{f}(t) - f(t)$ can finally be written as (e.g. Ruppert et al. 2003 p.140)

$$\widehat{\text{st.dev}}\left[\hat{f}(t) - f(t)\right] = \hat{\sigma}\left[\boldsymbol{z_t}'\left(\boldsymbol{Z}'\hat{\boldsymbol{\Omega}}\boldsymbol{Z} + \frac{\hat{\sigma}^2}{\hat{\tau}^2}\boldsymbol{K}\right)^{-1}\boldsymbol{z_t}\right]^{\frac{1}{2}}. \tag{3.23}$$

$\hat{\boldsymbol{\Omega}}$ is the estimated autocorrelation matrix of $\boldsymbol{\varepsilon}$ and $\boldsymbol{z_t} \in \mathbb{R}^{d\times 1}$ is a vector containing the t^{th} row of $\boldsymbol{Z}$. It follows that

$$\frac{\hat{f}(t) - f(t)}{\widehat{\text{st.dev}}\left[f(\hat{t}) - f(t)\right]} \sim \text{N}(0,1), \tag{3.24}$$

so that for a sufficient large sample size, the unbiased $100(1-\alpha)\%$ confidence interval can be calculated as

$$\hat{f}(t) \pm z\left(1 - \frac{\alpha}{2}\right)\hat{\sigma}\widehat{\text{st.dev}}\left[\hat{f}(t) - f(t)\right]. \tag{3.25}$$

3.3 Penalized splines and the Wiener-Kolmogorov filter

It is known that the HP-filter can be integrated into the framework of the Wiener-Kolmogorov filter (Hodrick/Prescott 1997, for the Wiener-Kolmogorov filter see Whittle 1983, Bell 1984 also Harvey 1989 and Kaiser/Maravall 2001). As penalized splines are equal to the HP-filter under certain values for the parameters (Proietti/Luati 2007, Paige 2010) there is also a link between penalized splines and the Wiener-Kolmogorov filter. To understand the relationship between these filters, it is useful to derive their connection from a very general standpoint. Assume that a time series $\{y_t\}_{t=1}^T$ can be written as the sum of a trend μ_t and a cyclical component c_t

$$y_t = \mu_t + c_t. \tag{3.26}$$

This can equivalently be expressed in matrix notation:

$$\boldsymbol{y} = \boldsymbol{\mu} + \boldsymbol{c}, \tag{3.27}$$

where $\boldsymbol{y} = (y_1, ..., y_T)'$, $\boldsymbol{\mu} = (\mu_1, ..., \mu_T)'$ and $\boldsymbol{c} = (c_1, ..., c_T)'$. If trend and cycle are defined as ARIMA(p,d,q) models and L denotes the backshift operator such that $\text{L}^k y_t = y_{t-k}$, then the components can be expressed as:

$$\Phi^{\mu}(\mathrm{L})\mu_t = \Theta^{\mu}(\mathrm{L})\varepsilon_t \quad \text{and} \quad \Phi^{c}(\mathrm{L})c_t = \Theta^{c}(\mathrm{L})\eta_t, \tag{3.28}$$

$$\text{where } \Phi^{\mu}(\mathrm{L}) = 1 - \varphi_{\mu,1}\mathrm{L}^1 - ... - \varphi_{\mu,p}\mathrm{L}^p,$$

$$\text{and } \Theta^{\mu}(\mathrm{L}) = 1 + \theta_{\mu,1}\mathrm{L}^1 + ... + \theta_{\mu,q}\mathrm{L}^q.$$

$\Phi^c(\mathrm{L})$ and $\Theta^c(\mathrm{L})$ are defined analogously. Both ε_t and η_t are independent white noise variables. If μ_t and c_t are integrated of (arbitrary) order n and r, $\Phi^{\mu}(\mathrm{L})$ and $\Phi^c(\mathrm{L})$ have characteristic polynomials with n and r roots on the unit circle.

Consequently, $(1-L)^n\mu_t = u_t$ and $(1-L)^r c_t = v_t$ are stationary and the vectors $\boldsymbol{u} = (u_{n+1}, ..., u_T)'$ and $\boldsymbol{v} = (v_{r+1}, ..., v_T)'$ have the autocovariance matrices $\boldsymbol{\Sigma_u}$ and $\boldsymbol{\Sigma_v}$. By introducing differencing matrices $\boldsymbol{\Delta_\mu}$ and $\boldsymbol{\Delta_c}$ of dimension $(T-n)\times T$ and $(T-r)\times T$, where as an example the product of $\boldsymbol{\Delta_\mu}$ and $\boldsymbol{\mu}$ yields the n^{th} first differences of $\boldsymbol{\mu}$, this can be expressed as:

$$\boldsymbol{\Delta_\mu \mu} = \boldsymbol{u} \quad \text{and} \quad \boldsymbol{\Delta_c c} = \boldsymbol{v}.$$

On purpose to estimate the trend component $\boldsymbol{\mu}$ McElroy (2008) shows that the minimum mean squared error linear estimate of $\boldsymbol{\mu}$ is given by:

$$\hat{\boldsymbol{\mu}} = (\boldsymbol{\Delta_c}'\boldsymbol{\Sigma_v}^{-1}\boldsymbol{\Delta_c} + \boldsymbol{\Delta_\mu}'\boldsymbol{\Sigma_u}^{-1}\boldsymbol{\Delta_\mu})^{-1}\boldsymbol{\Delta_c}'\boldsymbol{\Sigma_v}^{-1}\boldsymbol{\Delta_c y}. \tag{3.29}$$

This is the matrix formulation of the Wiener-Kolmogorov filter. The autocovariance matrices of $\boldsymbol{u}$ and $\boldsymbol{v}$ can equivalently be written as $\boldsymbol{\Sigma_u} = \boldsymbol{\Gamma_u}\sigma_u^2$ and $\boldsymbol{\Sigma_v} = \boldsymbol{\Gamma_v}\sigma_v^2$, where $\boldsymbol{\Gamma_u}$ and $\boldsymbol{\Gamma_v}$ are the autocorrelation matrices of $\boldsymbol{u}$ and $\boldsymbol{v}$, while σ_u^2 and σ_v^2 are the respective variances. Consequently, $\hat{\boldsymbol{\mu}}$ can also be expressed by:

$$\hat{\boldsymbol{\mu}} = (\boldsymbol{\Delta_c}'\boldsymbol{\Gamma_v}^{-1}\boldsymbol{\Delta_c} + \frac{\sigma_v^2}{\sigma_u^2}\boldsymbol{\Delta_\mu}'\boldsymbol{\Gamma_u}^{-1}\boldsymbol{\Delta_\mu})^{-1}\boldsymbol{\Delta_c}'\boldsymbol{\Gamma_v}^{-1}\boldsymbol{\Delta_c y}. \tag{3.30}$$

Now, consider the special case, where the trend is a second order integrated random walk and the cyclical component is a white noise process. Then the model above simplifies to:

$$(1-L)^2\mu_t = \varepsilon_t,$$

$$c_t = \eta_t.$$

This implies that $\boldsymbol{\Gamma_u} = \boldsymbol{\Gamma_v} = \boldsymbol{I}$ as well as $\boldsymbol{\Delta_c} = \boldsymbol{I}$. The differencing matrix $\boldsymbol{\Delta_\mu} \in \mathbb{R}^{(T-2)\times T}$ is defined as:

$$\boldsymbol{\Delta_\mu} = \begin{pmatrix} 1 & -2 & 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 & -2 & 1 \end{pmatrix}. \tag{3.31}$$

Thus, the minimum mean squared error estimation of the trend component reduces to (Flaig 2012 p.16):

$$\hat{\boldsymbol{\mu}} = (\boldsymbol{I} + \frac{\sigma_v^2}{\sigma_u^2}\boldsymbol{\Delta}'_{\boldsymbol{\mu}}\boldsymbol{\Delta}_{\boldsymbol{\mu}})^{-1}\boldsymbol{y}, \tag{3.32}$$

what is the matrix formula for the Hodrick-Prescott filter, when λ is equal to the inverse signal to noise ratio $\frac{\sigma_v^2}{\sigma_u^2}$. This makes obvious that the HP-filter is minimizing the mean squared error, when the trend is a two-fold integrated random walk and the cycle is just white noise. From this Hodrick/Prescott (1997) suggest to set $\sigma_v = 5$ and $\sigma_u = 1/8$ for quarterly data, which results in a penalization of $\lambda = 1600$, that has meanwhile become an "industry standard" in economics (Flaig 2012 p.23).

When the HP-filter is written as a penalized spline and incorporated into a mixed model, then usually an autocorrelated residual structure $\boldsymbol{R}$ is assumed. It can be shown that the formulation of the HP-filter as a mixed model according to (3.12) is a special case of the Wiener-Kolmogorov filter in (3.29) or (3.30). Assume that the trend component is just like before a second order integrated random walk, but the cycle is now an autocorrelated stationary process with the autocorrelation structure $\boldsymbol{\Gamma_v}$ and variance σ_v^2. Then it follows that $\boldsymbol{\Gamma_u} = \boldsymbol{I}$ and $\boldsymbol{\Delta_c} = \boldsymbol{I}$, while $\boldsymbol{\Delta_\mu}$ is still defined like in (3.31). For these assumptions the Wiener-Kolmogorov filter is defined as

$$\hat{\boldsymbol{\mu}} = (\boldsymbol{\Gamma_v}^{-1} + \frac{\sigma_v^2}{\sigma_u^2}\boldsymbol{\Delta_\mu}'\boldsymbol{\Delta_\mu})^{-1}\boldsymbol{\Gamma_v}^{-1}\boldsymbol{y}. \tag{3.33}$$

When the setting of the tp-spline is such that it is equal to the HP-filter, i.e. T equidistant knots are chosen and $l = 1$ is selected, then $\boldsymbol{Z}$ is generally defined like in section 3.2.2. Using this, the mixed model representation of the HP-filter from (3.12) can be written as (the derivation is equivalent to appendix 3.A).

$$\hat{\boldsymbol{y}} = \boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{Z} + \frac{1}{\tau^2}\boldsymbol{K})^{-1}\boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{y} = \boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{\Omega}^{-1}\boldsymbol{Z} + \frac{\sigma^2}{\tau^2}\boldsymbol{K})^{-1}\boldsymbol{Z}'\boldsymbol{\Omega}^{-1}\boldsymbol{y} =$$

$$= (\boldsymbol{\Omega}^{-1} + \frac{\sigma^2}{\tau^2}\boldsymbol{Z}'^{-1}\boldsymbol{K}\boldsymbol{Z}^{-1})^{-1}\boldsymbol{\Omega}^{-1}\boldsymbol{y}. \tag{3.34}$$

Noting that $\boldsymbol{Z}'^{-1}\boldsymbol{K}\boldsymbol{Z}^{-1} = \boldsymbol{\Delta_\mu}'\boldsymbol{\Delta_\mu}$ then it immediately follows that

$$\hat{\boldsymbol{y}} = (\boldsymbol{\Omega}^{-1} + \frac{\sigma^2}{\tau^2}\boldsymbol{\Delta_\mu}'\boldsymbol{\Delta}_\mu)^{-1}\boldsymbol{\Omega}^{-1}\boldsymbol{y}. \tag{3.35}$$

As $\boldsymbol{\Omega} = \boldsymbol{\Gamma_v}$ and $\lambda = \dfrac{\sigma^2}{\tau^2} = \dfrac{\sigma_v^2}{\sigma_u^2}$ it becomes obvious that the HP-filter within the mixed model framework is equal to the Wiener-Kolmogorov filter when the trend is a two-fold integrated random walk and the cycle is a stationary autocorrelated process. It follows for this setting that the Wiener-Kolmogorov filter can be estimated by maximum likelihood estimation. Moreover, in this setting a penalized tp-spline within a mixed model is the mean squared error minimizing filter for T equidistant knots and $l = 1$.

3.4 Empirical application

In this section the HP-filter is used to estimate the trend component of the quarterly real GDP of Germany, which is adjusted for seasonal and calendar effects. Moreover, the long run development of the German industrial production index is investigated. The HP-filter is written as a penalized tp-spline within a mixed model framework in order to derive an estimator for the penalization parameter λ. Although this setting is special, the results can be seen as general for penalized splines and the mixed model interpretation. Kauermann/Opsomer (2011) show that the results are almost independent of the number of knots m, as long as m is not too low. Furthermore, Claeskens et al. (2009) show that the estimations are hardly influenced by the selected spline basis, which is also demonstrated by Ruppert (2002).

An important question concerns the assumed autocorrelation structure of the error term. According to Krivobokova/Kauermann (2007) the results of the maximum likelihood estimation are robust to a misspecification of the autocorrelation structure, as long as the assumed autocorrelation doesn't deviate too much from the true autocorrelation structure. However, it is worth to consider that different additive time series models require different assumptions about the error term. If it is assumed that the series can be decomposed simply in

$$y = \text{trend} + \text{cycle},$$

then it might be sufficient to choose an AR(p)-process for the residual structure. Here, the trend is represented by the estimated spline and the cycle by an AR(p)-process. However, especially the industrial production index contains short term fluctuations, i.e. a noise component (Flaig 2005 p.419). Thus, in addition the (more likely) decomposition

$$y = \text{trend} + \text{cycle} + \text{noise},$$

should be considered. In this case it is not sufficient to choose an AR(p)-process for the residual correlation structure. This is because an AR(p)-process cannot capture the noise in the residual component. For such a decomposition of the time series an ARMA(p,q)-process has to be used for the autocorrelation structure, because it allows including the irregular noise in the error term. If the cycle is assumed to follow an AR(p)-process and the noise is expressed by a white noise variable, then the sum of cycle and noise is given by an ARMA(p,p)-process (e.g. Schlittgen/Streitenberg 2001 p.133). Consequently, an ARMA(p,p)-process is necessary in this situation. To check for both cases of time series models, the estimation is done for AR(1) and AR(4) as well as ARMA(1,1) and ARMA(4,4) residual structures.

As a further robustness check the trend of the real German industrial production[5] is estimated. This offers a higher robustness of the estimation, as the index of the industrial

[5]The data of the GDP and the industrial production are from the German Federal Statistical Office.

production is published monthly, so that compared to quarterly data a three times larger data basis is available. Although the industrial production covers only about one fourth of the German GDP (Flaig 2005 p.419), it is a good proxy for the economic activity, since it is highly correlated with the GDP and a main factor that drives the economic development (Langmantel 1999, Abberger/Nierhaus 2008).

First of all the trend estimation for the German GDP is considered.[6][7] The results are shown in Figure 3.1.

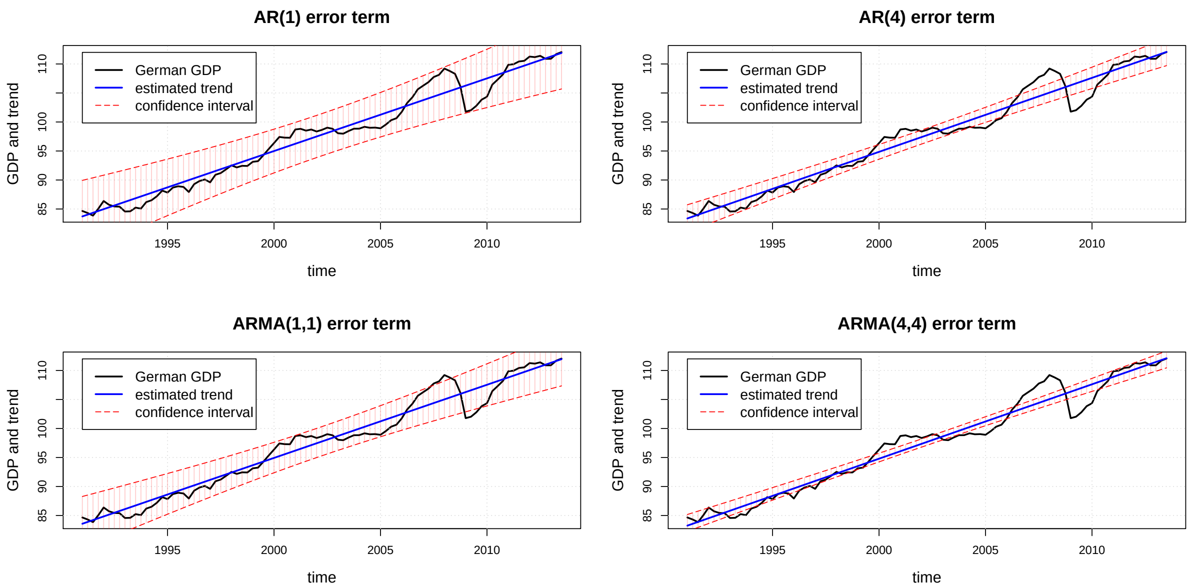

Fig. 3.1: Estimated trend of the real German GDP for different residual autocorrelation structures

Independent of the selected autocorrelation structure for the error term an infinitely high value for λ is estimated. This implies a linear trend in every of the four situations. Nevertheless, the different assumptions about the residual autocorrelation structure yield different estimations for σ^2, which results in varying confidence intervals. The estimations are almost similar. There are only very slight differences with regard to the intercept and the slope. For the case of the AR(1) error term the slope is 1.25 per year, while it is 1.28 per year in the case of the ARMA(4,4) error term. Furthermore, the confidence intervals become smaller for higher autocorrelation structures of the error term, since the estimated variance $\hat{\sigma}^2$ strongly decreases.

Figure 3.2 shows the estimated business cycle resulting from the ARMA(4,4) error term. It is smoothed with a HP-filter in the mixed model framework, where a white noise structure for the error term is assumed. To this point a penalization parameter of $\hat{\lambda} = 0.31$ is estimated.

[6] For the optimization the R routine 'nlminb' was used.
[7] The starting values for the optimization were selected according to appendix 3.C.

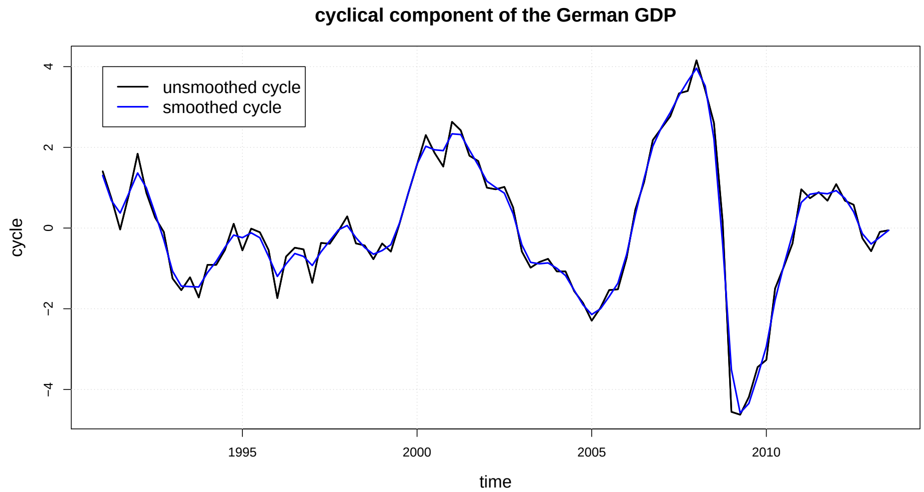

Fig. 3.2: Estimated cyclical component of the real German GDP

To check for the robustness of the assumptions about the error term, the autocorrelation and partial autocorrelation function of the estimated residuals of the estimation with the ARMA(4,4) autocorrelation structure are calculated. These are shown in Figure 3.3:

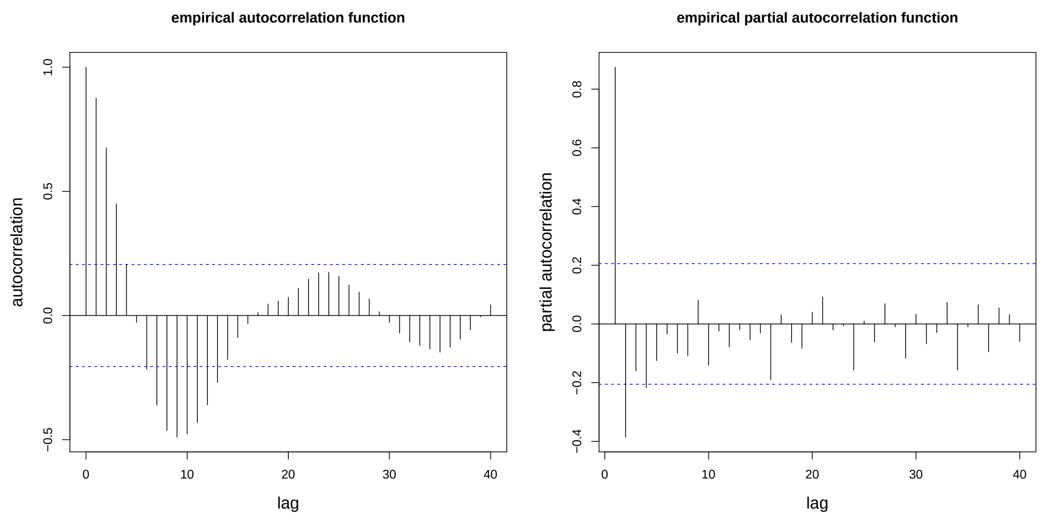

Fig. 3.3: Empirical autocorrelation- and partial autocorrelation functions

While the autocorrelation function is damped and slowly converges to zero, the partial autocorrelation function is only significant to the first, second and fourth lag. This implies that the assumed ARMA(4,4) autocorrelation structure seems to be no massive misspecification. However, even if an AR(2)-process would be used for the estimation, the shape of the trend would deviate only slightly. Just the confidence intervals would differ from the ARMA(4,4) case, because of different estimated parameters for the autocorrelation structure.

The results of the estimation for the industrial production index are shown in Figure 3.4.

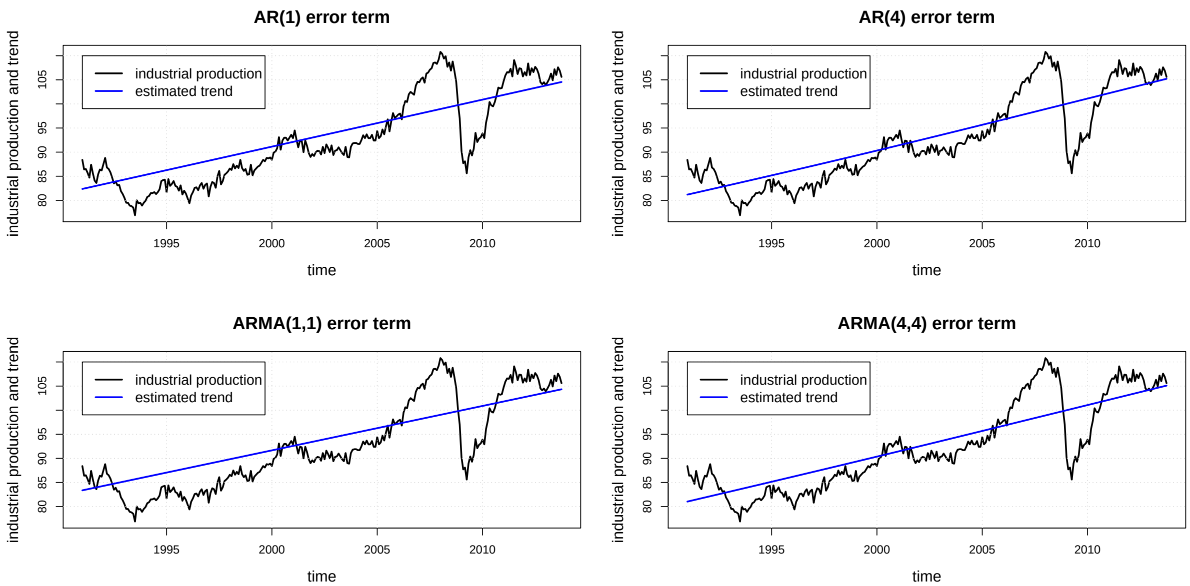

Fig. 3.4: Trend of the German industrial production index for different residual autocorrelation structures

Also for the industrial production index the penalization is estimated (almost) infinitely high, which results in the linear shape of the trend functions. Figure 3.4 shows that there are differences between the trend estimations, although they are rather small. While the trend is linear for all residual autocorrelation structures there are differences in the intercept and the slope, especially when the results of the AR(1) and ARMA(1,1) error terms are compared to those of the AR(4) and ARMA(4,4) error terms. In detail the slope for the AR(1) case is 0.975 per year, whereas it is 1.025 in the case of an ARMA(4,4) error term. To check for the robustness of the assumed autocorrelation structure Figure 3.5 displays the empirical autocorrelation- and partial autocorrelation functions for the residuals of the AR(1) and ARMA(4,4) estimation:

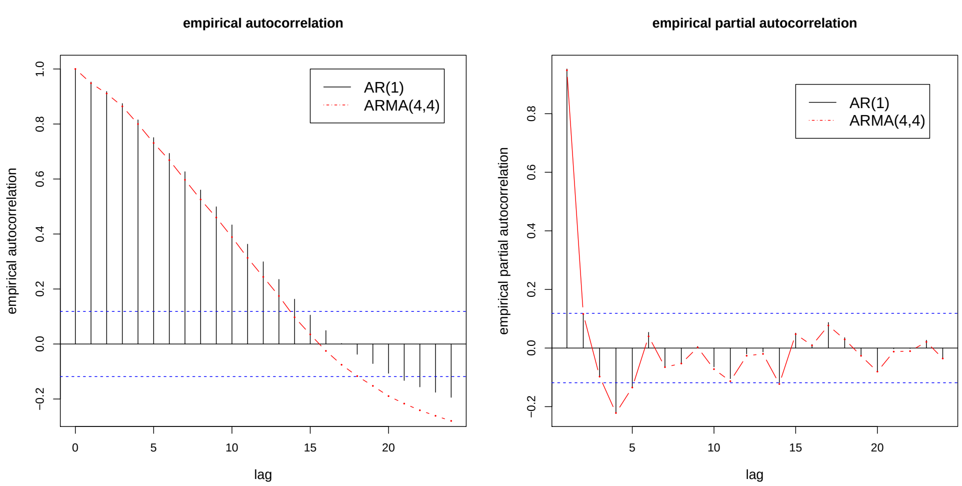

Fig. 3.5: Empirical autocorrelation and partial autocorrelation functions

The empirical autocorrelation and partial autocorrelation functions are almost equal for both cases. The empirical autocorrelation is damped and converges slowly to zero. The empirical partial autocorrelation adopts higher and clearly significant values for the first and fourth lag. Thus, higher autocorrelation structures rather than an AR(1)-process might be adequate, so that the results of the AR(4) or ARMA(4,4) cases appear to be preferable.

The analysis of the trend of the German economy by the HP-filter within a mixed model framework suggests that there is a very smooth and constant development on the long run. The estimation for the real GDP turns out to be very robust with regard to the assumed autocorrelation structure of the error term. In every case an infinitely high penalization is estimated, which is far above the commonly used penalization of $\lambda = 1600$ for quarterly data. The estimation yields a linear trend function with a slope of 1.28 per year. Moreover, the empirical autocorrelation and partial autocorrelation functions indicate, that the assumed autocorrelation structure for the error term is no massive misspecification.
The trend estimation of the real industrial production as an alternative measure for the economic development shows a very similar picture. For every assumed residual autocorrelation structure a linear trend is estimated. These results are quite similar to those of Flaig (2005), who uses unobserved components models to examine the long run development of the German industrial production index, where the trend from 1991 to 2005 also shows a very smooth shape.

3.5 Conclusion

This chapter reviews basic theory about penalized splines with a truncated polynomial basis as well as their link to linear mixed models. This link is advantageous, as it can be employed to derive an estimation for the penalization parameter λ. Moreover, it is focused on the calculation of confidence intervals, which is an interesting topic, since penalized splines as special cases of the ridge regression are biased, which is a drawback when confidence intervals are derived. To this regard the mixed model interpretation of penalized splines offers a further advantage for calculating confidence intervals, as they yield the best linear unbiased predictors.

Afterwards it is focused on the link between the HP-filter, penalized tp-splines and the Wiener-Kolmogorov filter. As the HP-filter is equal to a penalized tp-spline under certain settings of the parameters it can be incorporated into a mixed model framework. It is shown that for an autocorrelated error term the resulting model is equivalent to the Wiener-Kolmogorov filter where the trend is a second fold integrated random walk and the cycle follows a stationary autocorrelated process. The estimation within the mixed model framework by maximum likelihood furthermore offers the advantage that it yields unbiased estimators and thus also correct confidence intervals.

The possibility to incorporate the Hodrick-Prescott filter into a mixed model framework is used to estimate the trend component of the real German GDP. As a robustness check, also the trend component of the industrial production is estimated, as it is highly correlated with the real GDP and offers a wider data basis, because it is published monthly. Although studies indicate that the exact assumption about the autocorrelation structure of the error term is of subordinate importance as long as the assumed structure doesn't deviate too strong from the true one, the estimation is done for different residual autocorrelation structures. It turns out that the results are hardly affected by the assumed autocorrelation. In every case a very high penalization is estimated inducing for both series a linear trend. Moreover, the empirical residual autocorrelation structures do not deviate strongly from the assumed one, so that there is unlikely a massive misspecification of the error term.

As a result this chapter employs the Hodrick-Prescott filter within a mixed model framework in order to derive the trend component of the real German GDP. This analysis indicates that there is a very smooth a linear long run development of the German economy which is in line with existing literature like Flaig (2005).

Appendix

3.A Derivation of formula 3.8

First, recall that for two invertible matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ (e.g. Fahrmeir et al. 2009 p.450)

$$(\boldsymbol{AB})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}.$$

This is used to derive

$$\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z}+\lambda\boldsymbol{K})^{-1}\boldsymbol{Z}'\boldsymbol{y} = \left[(\boldsymbol{Z}'\boldsymbol{Z}+\lambda\boldsymbol{K})\boldsymbol{Z}^{-1}\right]^{-1}\boldsymbol{Z}'\boldsymbol{y} =$$

$$= (\boldsymbol{Z}'\boldsymbol{Z}\boldsymbol{Z}^{-1}+\lambda\boldsymbol{K}\boldsymbol{Z}^{-1})^{-1}\boldsymbol{Z}'\boldsymbol{y} = (\boldsymbol{Z}'+\lambda\boldsymbol{K}\boldsymbol{Z}^{-1})^{-1}\boldsymbol{Z}'\boldsymbol{y} =$$

$$= \left[\boldsymbol{Z}'^{-1}(\boldsymbol{Z}'+\lambda\boldsymbol{K}\boldsymbol{Z}^{-1})\right]^{-1}\boldsymbol{y} = (\boldsymbol{Z}'^{-1}\boldsymbol{Z}'+\lambda\boldsymbol{Z}'^{-1}\boldsymbol{K}\boldsymbol{Z}^{-1})^{-1}\boldsymbol{y} =$$

$$= (\boldsymbol{I}+\lambda\boldsymbol{Z}'^{-1}\boldsymbol{K}\boldsymbol{Z}^{-1})^{-1}\boldsymbol{y}.$$

3.B Proof that $(\boldsymbol{Z}')^{-1}\boldsymbol{K}\boldsymbol{Z}^{-1} = \boldsymbol{\Delta_\mu}'\boldsymbol{\Delta_\mu}$

Given the special form of $\boldsymbol{Z}$ for $l=1$ and T equidistant knots, the inverse of $\boldsymbol{Z}$ is in general given to (see also Paige 2010 p.870):

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

Then the product of the transpose of $\boldsymbol{Z}^{-1}$ and $\boldsymbol{K}$ yields:

$$(\boldsymbol{Z}^{-1})'\boldsymbol{K} = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & -2 & 1 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 1 & -2 & 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -2 & 1 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & 1 & -2 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 1 & -2 \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 1 \end{pmatrix},$$

where the product of the two matrices has the dimension $T \times T$. This product can be decomposed in two matrices $\boldsymbol{E_1} = \mathbf{0} \in \mathbb{R}^{T\times 2}$ and $\boldsymbol{E_2} \in \mathbb{R}^{T\times T-2}$ so that $(\boldsymbol{Z^{-1}})'\boldsymbol{K} = (\boldsymbol{E_1}, \boldsymbol{E_2})$. Having a closer look on $\boldsymbol{E_2}$ it becomes clear that $\boldsymbol{E_2} = \boldsymbol{\Delta_\mu}'$. Consequently, the product $(\boldsymbol{Z}^{-1})'\boldsymbol{K} = (\mathbf{0}, \boldsymbol{\Delta_\mu}')$. Moreover, $\boldsymbol{Z}^{-1}$ is partitioned into $(\boldsymbol{Z_1}', \boldsymbol{Z_2}')'$, with $\boldsymbol{Z_1} \in \mathbb{R}^{2\times T}$ and $\boldsymbol{Z_2} \in \mathbb{R}^{T-2\times T}$. Taking into consideration that $\boldsymbol{Z_2} = \boldsymbol{\Delta_\mu}$ it can finally be written:

$$(\boldsymbol{Z}^{-1})'\boldsymbol{K}\boldsymbol{Z}^{-1} = \left(\mathbf{0}, \boldsymbol{\Delta}'_{\boldsymbol{\mu}}\right)\begin{pmatrix}\boldsymbol{Z_1}\\ \boldsymbol{\Delta_\mu}\end{pmatrix} = \mathbf{0}\boldsymbol{Z_1} + \boldsymbol{\Delta_\mu}'\boldsymbol{\Delta_\mu} = \boldsymbol{\Delta_\mu}'\boldsymbol{\Delta_\mu}$$

3.C Selection of the starting values for the optimization of the restricted log-likelihood

As mentioned in section 3.4 for the optimization of the restricted log-likelihood function the R routine 'nlminb' is used, which requires an initial set of starting values of the model parameters. Thus, it shall be described shortly, how starting values were found. To this point the decomposition

$$y_t = \mu_t + c_t$$

is chosen, where y_t is the original time series, μ_t is the trend and c_t is the cycle. In a first step, a Hodrick-Prescott filter with penalization λ is applied, in order to get an initial guess of trend and cycle. Thus:

$$\hat{\mu}_t = HP_\lambda(y_t),$$

$$\hat{c}_t = y_t - \hat{\mu}_t.$$

The cycle is supposed to follow an (stationary) ARMA(p,q)-process. Consequently, the starting values for the parameters of the residual correlation structure are gained by estimating c_t as the selected ARMA(p,q)-process. According to the theory of the ideal Wiener-Kolmogorov filter, the trend is supposed to be an integrated process of order n.

$$(1-L)^n\mu_t = \eta_t,$$

where η_t is stationary. As the penalization parameter of the ideal Wiener-Kolmogorov filter (Whittle 1983, Bell 1984, Kaiser/Maravall 2001) is given by the ratio of the variances $\frac{\sigma_c^2}{\sigma_\eta^2}$, the starting value for the variance of the error term is simply the empirical variance of $\hat{c}_t$. To get a starting value for the variance of the random effects, the first differences of $\hat{\mu}_t$ are calculated so many times, until they render a stationary process, which yields $\hat{\eta}_t$. Then the empirical variance of $\hat{\eta}_t$ is used as starting value. Using this procedure, different sets of starting values can be obtained by selecting different values for the penalization of the Hodrick-Prescott filter in the first step.

4 Penalized Splines as Frequency Selective Filters Reducing the Excess Variability at the Margins[8]

Summary

The outcome of penalized splines as instruments for trend estimation predominantly depends on the selection of the penalization parameter. This chapter derives the penalization by frequency domain aspects and points out their link to rational square wave filters. As a novel contribution it focuses on the so called excess variability at the margins. This excess variability describes the insufficient suppression of high frequencies at the ends of the series, which induces an undesired increase of the variability for estimations at the margins. It is shown that the too high volatility at the margins can be reduced considerably by a time-varying penalization, increasing the precision of the estimations for the most recent periods.

[8]This chapter refers to Blöchl (2014b).

4.1 Introduction

A fundamental challenge in economics, especially in business cycle research, is to decompose a time series into trend and cycle. There exists a wide range of instruments to estimate these components, 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 derive this value by interpreting the filter as an optimal Wiener-Kolmogorov filter (Whittle 1983, Bell 1984). Since 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 for most economic time series (McCallum 2000, Flaig 2012) and not data driven (e.g. Schlicht 2005, Kauermann et al. 2011). 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 autocorrelated errors (Diggle/Hutchinson 1989, Altman 1990, Hart 1991), but it can be extended to account for an autocorrelated residual structure (Kohn et al. 1992, Wang 1998). Nevertheless, Opsomer et al. (2001), Proietti (2005) and also Dagum/Giannerini (2006) demonstrate that this technique is very sensitive to the assumptions about the residual autocorrelation structure. To this point the mixed model approach is advantageous, as it is relatively robust with regard to a misspecification of the residual autocorrelation 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 is demonstrated by Tödter (2002) for the Hodrick-Prescott filter. This chapter shows 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 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 chapter, a high number of forecasts is required so that this approach is of limited practicability. Instead, this chapter 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 is already suggested by Razzak/Richard (1995) and Pollock (2009) in order to account for structural breaks and also Crainiceanu et al. (2005) introduce a time-varying penalization for splines within a mixed model framework. Moreover, Bruchez (2003) offers an ad-hoc approach to reduce the excess variability by a time-varying penalization.

This chapter is structured as follows. In the first section penalized splines with a truncated polynomial basis are discussed briefly. Then it is shown how to choose the penalization to extract certain frequency bands. Afterwards it is explained how spectral analysis and a time-varying penalization can be employed 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 4.4 provides an empirical example.

4.2 Penalized splines

A 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 chapter always equidistant knots are used. If ε_t denotes the error term, a tp-spline of degree l, henceforth denoted as $tp(l)$, is defined as:

$$y_t = f(t) + \varepsilon_t = \beta_1 + \beta_2 t + ... + \beta_{l+1} t^l + \beta_{l+2}(t-\kappa_2)_+^l + ... + \beta_d (t-\kappa_{m-1})_+^l + \varepsilon_t, \quad (4.1)$$

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

where $d = m + l - 1$. In this sense $f(t)$ represents the trend and ε_t the cycle. 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:

$$\boldsymbol{y} = \boldsymbol{Z}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{4.2}$$

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

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 (e.g. Fahrmeir et al. 2009 p.308).

$$\min_{\boldsymbol{\beta}} PLS(\lambda) = \sum_{t=1}^{T} [y_t - f(t)]^2 + \lambda \sum_{j=l+2}^{d} \beta_j^2. \tag{4.3}$$

The solution of this minimization is (e.g. Fahrmeir et al. 2009 p.313):

$$\hat{\boldsymbol{\beta}}(\lambda) = (\boldsymbol{Z}'\boldsymbol{Z} + \lambda\boldsymbol{K})^{-1}\boldsymbol{Z}'\boldsymbol{y} \text{ where } \boldsymbol{K}_{d\times d} = \text{diag}(\underbrace{0, ..., 0}_{l+1}, \underbrace{1, ..., 1}_{m-2}). \tag{4.4}$$

The fitted values for a given λ are defined by:

$$\hat{\boldsymbol{y}}(\lambda) = \underbrace{\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z} + \lambda\boldsymbol{K})^{-1}\boldsymbol{Z}'}_{\boldsymbol{H}(\lambda)}\boldsymbol{y}. \tag{4.5}$$

$\boldsymbol{H}(\lambda)$ 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 observed 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 connected to square wave filters that are related to 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.

4.3 The optimal penalization

4.3.1 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 in the medium and short run. Consequently, it is defined by a bandwidth of 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 p.189). The gain function of a spline for estimation $\hat{y}_t$ and a frequency ω is given to (e.g. Mills 2003 p.80):

$$g_t(\omega, \lambda) = \sqrt{\left[\sum_{j=1-t}^{T-t} h_{t,j+t} \cos(\omega j)\right]^2 + \left[\sum_{j=1-t}^{T-t} h_{t,j+t} \sin(\omega j)\right]^2}. \tag{4.6}$$

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) for the Baxter-King filter or Tödter (2002) for the Hodrick-Prescott filter, who aim to minimize the deviation between the gain function of the filter and the ideal gain function.

As an example Figure 4.1 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.

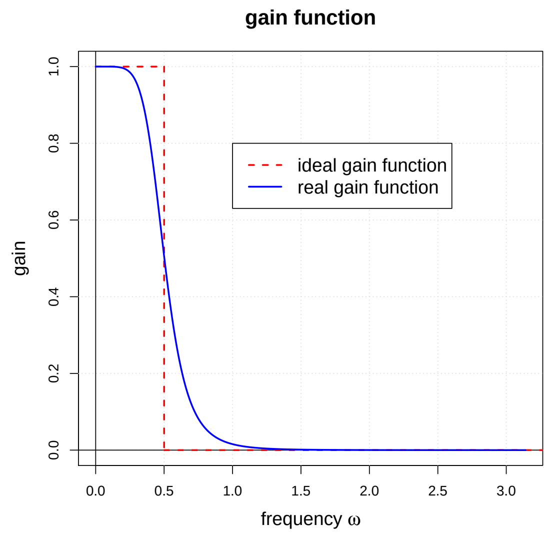

Fig. 4.1: Examples for an ideal gain function and a real gain function

The ideal gain function is opposed to a real gain function of a $tp(2)$ with $\lambda = 250$ and 140 equidistant knots that is applied to a series with 140 observations. The gain function of the $tp(2)$ refers to the 70^{th} estimation. Now, the approach in this chapter aims to minimize the squared deviation between the real 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:

$$l_t(\lambda) = \int_0^\pi [g^*(\omega) - g_t(\omega, \lambda)]^2 d\omega. \tag{4.7}$$

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:

$$l_t(\lambda) = \sum_{i=1}^{n} [g^*(\omega_i) - g_t(\omega_i, \lambda)]^2 \cdot \delta, \tag{4.8}$$

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).

4.3.2 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 4.2 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 4.2 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.

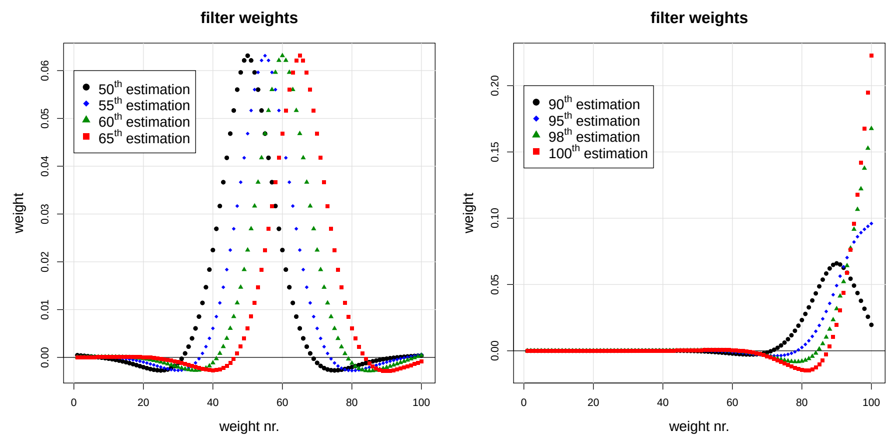

Fig. 4.2: Filter weights for different estimations

This change of the filter weight structure affects the gain. Figure 4.3 shows the gain functions of estimations for different periods, which refer to the $tp(1)$ of the example above:

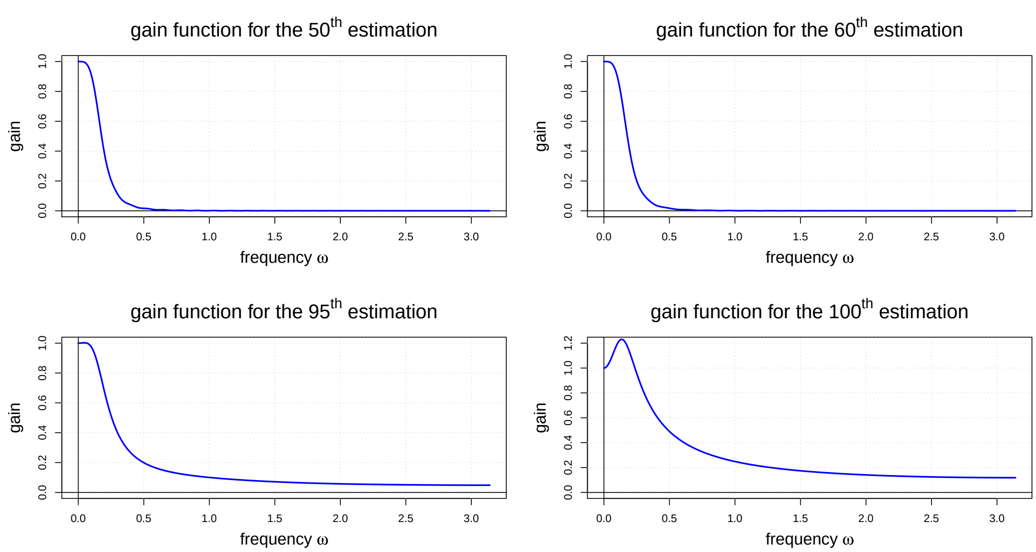

Fig. 4.3: Gain functions of the $tp(1)$ for estimations around the middle and at the margin

The gain functions 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 than 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 4.2 and 4.3 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 representing the excess variability by one single graph. This is denoted as the $loss\ function$. As an example the left plot of Figure 4.4 shows the $loss\ function$ for the $tp(1)$ with $\lambda = 1000$ and $T = m = 100$. For the ideal gain function a cut-off frequency of $\omega^{cf} = 0.196$ is chosen, which implies a periodicity of eight years in the case of quarterly data. Furthermore, the right plot of Figure 4.4 displays for every $\hat{y}_t$, $t = 1, ...T$, that value of λ that minimizes the $loss$ for this specific estimation.

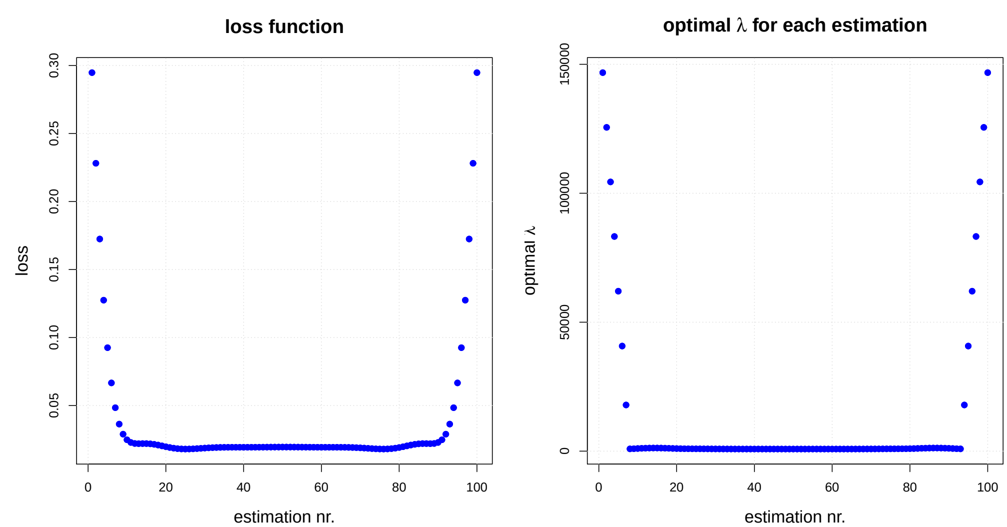

Fig. 4.4: $loss\ function$ and optimal values for λ

The left plot of Figure 4.4 shows that the $loss$ is rather low and similar for most estimations around the middle. 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 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 is 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 4.5 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.

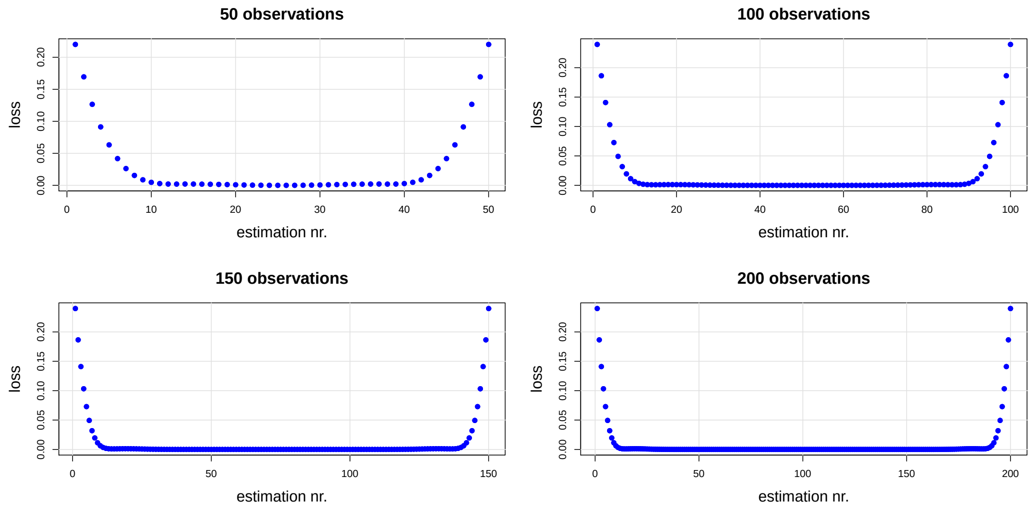

Fig. 4.5: *loss functions* of a $tp(1)$ for series of different length

Figure 4.5 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 4.6 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.

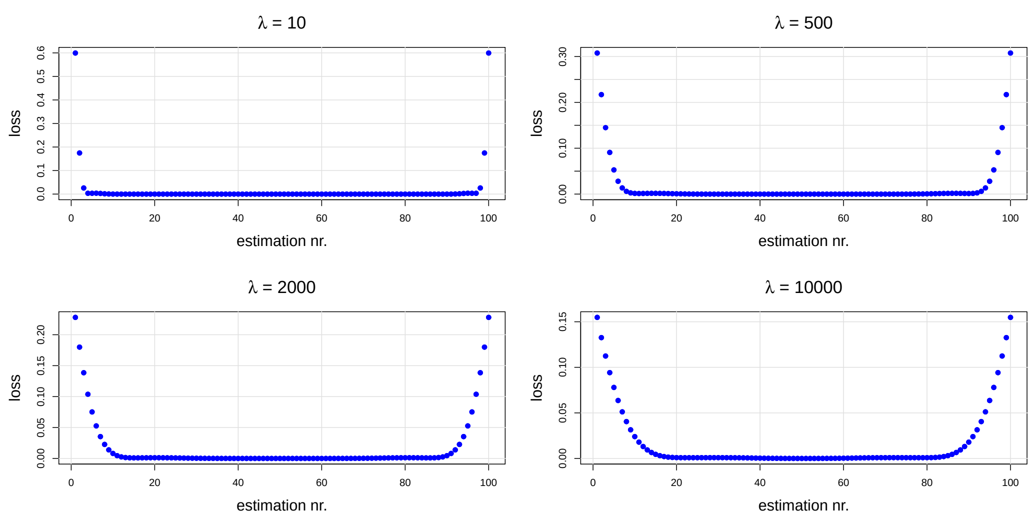

Fig. 4.6: *loss function* of a $tp(1)$ for different values of λ

Clearly, the number of estimations that is affected by the excess variability depends on the value of λ. For $\lambda = 10$ only 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 4.5 and 4.6 that the excess variability depends on the penalization, but not on the number of observations.

4.3.3 The time-varying penalization and the number of knots

The previous section describes 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 4.2, 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 the penalty matrix is defined as $\tilde{\boldsymbol{K}} = \text{diag}(\boldsymbol{\lambda})$. The penalized spline with a time-varying penalization can then be written in matrix notation as

$$\hat{\boldsymbol{y}}(\boldsymbol{\lambda}) = \boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z} + \tilde{\boldsymbol{K}})^{-1}\boldsymbol{Z}'\boldsymbol{y}. \tag{4.9}$$

The change of the coefficient of the highest polynomial at each knot is determined by the penalization parameter, which now is able to vary over time. Given the vector $\boldsymbol{\lambda}$, λ_i determines the change at the knot κ_{i+1}, so that a specific penalization can be set at each knot. Figure 4.4 shows that the penalization needs to increase to the ends, 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 4.4 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

$$\lambda_{m-2-j+i} = \alpha_0 + \alpha_1 \cdot i, \qquad i = 1, ..., j. \tag{4.10}$$

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

$$\lambda_1 = \lambda_{m-2}, \lambda_2 = \lambda_{m-3}, ..., \lambda_j = \lambda_{m-1-j}. \tag{4.11}$$

α_0 is the value for the λ's closer to the middle which do not need to rise. As seen in Figure 4.4 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 (4.10) and (4.11), 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 2014a)

$$L(\boldsymbol{\lambda}) = \sum_{t=1}^{T} l_t(\boldsymbol{\lambda}), \tag{4.12}$$

by the time-varying penalization. As it turns 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 the focus usually not lies on a single estimation, but on the trend of the whole time series. Given the condition of a linear increase, the cumulative *loss* $L(\boldsymbol{\lambda})$ is minimized subject to (4.10) and (4.11). 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. It has to be considered 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 selected that yields to the lowest cumulative *loss*.

Beside the flexible penalization secondly, a value for the number of knots m has to be chosen. 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 select a high number of knots, as it allows an accurate determination of the flexible penalization at the margins. As seen in Figure 4.4, 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 chapter 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 numerical instable, when the number of knots is too high (e.g. Fahrmeir et al. 2009 p.303). Hence, for splines of higher degrees the number of knots should be set to a high value that still allows a numerical stable estimation.

4.3.4 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) or Baxter/King (1999). The simulated series contains 140 observations, which might be not unrealistic for most quarterly economic time series. Table 4.1 shows the resulting parameters for tp-splines of degrees one, two and three.

spline	$\boldsymbol{m}$	$\boldsymbol{\alpha_0}$	$\boldsymbol{\alpha_1}$	$\boldsymbol{j}$
$tp(1)$	140	821	654	21
$tp(2)$	140	79678	112500	28
$tp(3)$	140	$18.7 \cdot 10^6$	$40.6 \cdot 10^6$	35

Tab. 4.1: Optimal parameters for $\omega^{cf} = 0.196$

For example, according to Table 4.1, the last j values of $\boldsymbol{\lambda}$ for the $tp(1)$ are defined by $\lambda_{m-2-j+i} = 821 + 654 \cdot i$, $i = 1, ..., 21$. The first 21 values of the penalization $\lambda_1, ..., \lambda_{21}$ are defined analogously to (4.11). The remaining parameters $\lambda_{22}, ..., \lambda_{117}$ do not require an increased penalization and are set to 821. Table 4.1 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 4.2 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 (4.5):

spline	**penalization**	**70$^{\text{th}}$ estimation**	**140$^{\text{th}}$ estimation**	$\boldsymbol{L(\lambda)}$
$tp(1)$	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

Tab. 4.2: *loss* for a cut-off periodicity of $\omega^{cf} = 0.196$

Consider 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 4.7 that displays the gain functions for the splines in the middle (70$^{\text{th}}$ estimation) and at the margin (140$^{\text{th}}$ 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.

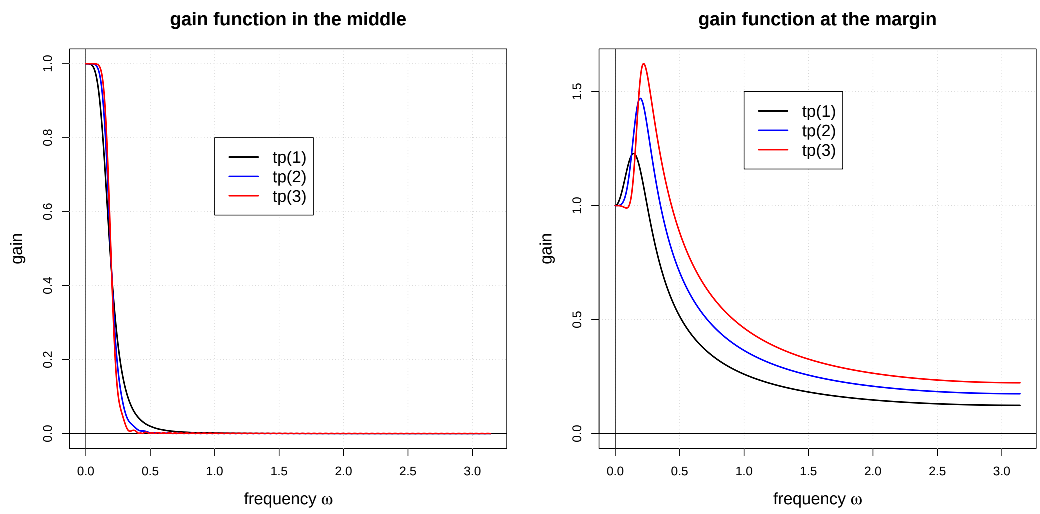

Fig. 4.7: Gain functions in the middle and at the margin

This also shows the link of penalized tp-splines and rational square wave filters that is 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.

The results for the flexible penalization in Table 4.2 show that the *loss* at the margins is reduced strongly in every case by around 38-55 percent, while the *loss* in the middle is hardly affected. Moreover, the cumulative *loss* is decreased for every spline. To demonstrate the effects of the flexible penalization for the whole time series, Figure 4.8 displays the *loss functions* for both the fixed and the flexible penalization.

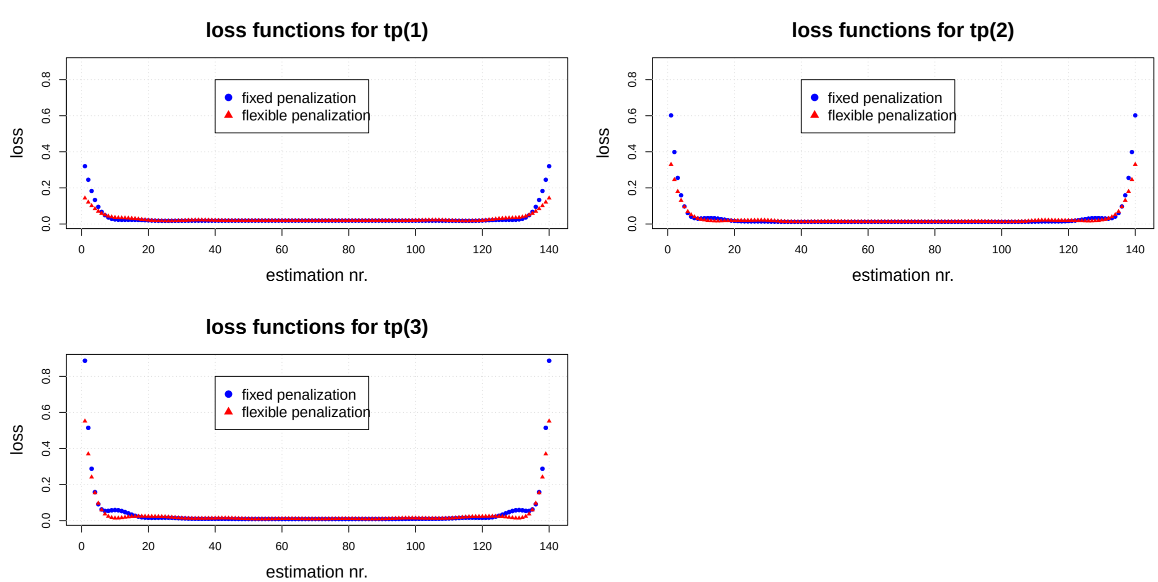

Fig. 4.8: *loss functions* for fixed and flexible penalization

Obviously, 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 4.8 is that in every case the $loss$ at the margins can be reduced considerably by the flexible penalization, while it is increased only slightly for some estimations closer to the middle. As the cumulative $loss$ decreases 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 induces hardly any notable increases of the $loss$ and clearly improves the results at the margins. Consequently, Figure 4.8 shows that the flexible penalization is able to reduce the excess variability at the margins and to yield more precise 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 4.8 shows, in this case at least forecasts for the next 15 periods are required. For quarterly data this implies that data for the next four years have to be predicted. 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 4.9 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).

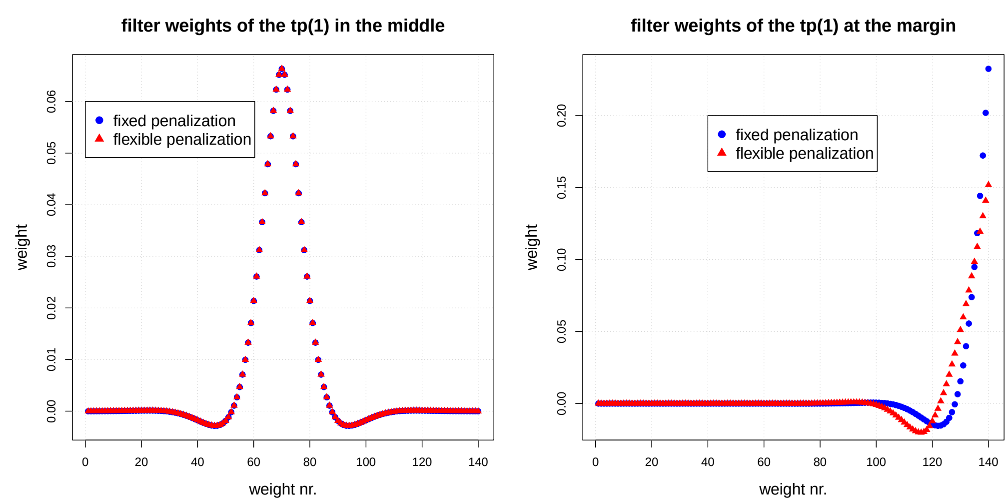

Fig. 4.9: Filter weights for fixed and flexible penalization

The left plot of Figure 4.9 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 4.2, 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.

4.4 Empirical application

To demonstrate the effect of the time-varying penalization on real time series, the trend component of the seasonally adjusted quarterly real GDP of Switzerland is estimated.[9] 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 $j = 21$, where $m = 135$.

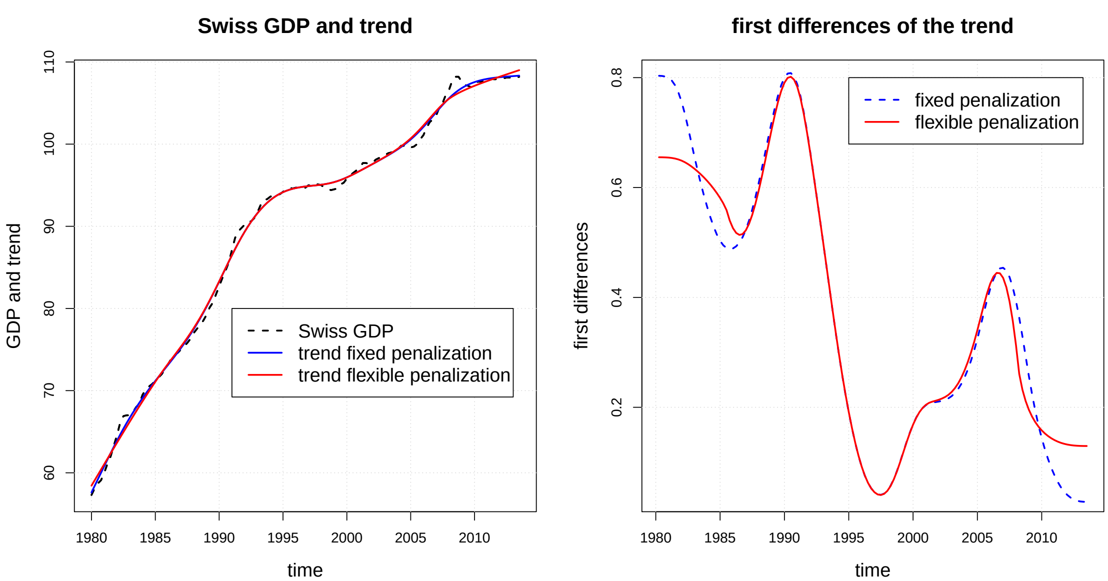

Fig. 4.10: Trend estimation of the real Swiss GDP and first differences of the trend

The left plot of Figure 4.10 shows the Swiss GDP as well as the estimated trend resulting from the fixed and the flexible penalization. Especially at the end of the series, there are clear differences between the estimations. In both cases the trend growth rate declines

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

after 2008, but the decline is larger for the fixed penalization. The trend according to the flexible penalization exhibits much larger growth rates and lies above the one of the fixed penalization for the most recent periods. The right plot of Figure 4.10 shows the first differences of both trend estimations. Also the first differences only deviate at the margins of the series. In both cases the first differences have decreased 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 4.11.

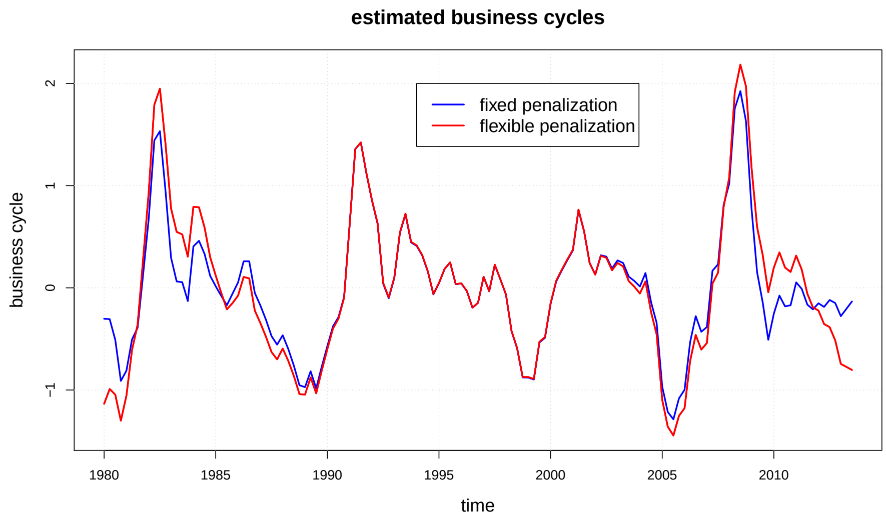

Fig. 4.11: Estimated business cycle for fixed and flexible penalization

The same pattern can be observed: The flexible penalization affects the margins of the series, while it has no effect in the middle. Predominantly the output gap at the end of the series is much larger in the case of the time-varying penalization. As seen in Figures 4.2 and 4.9, 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.

4.5 Conclusion

One of the unsolved problems of trend estimation is to get precise 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 of this chapter uses penalized splines to estimate the trend component. The penalization is 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 is demonstrated that the deviation of the ideal gain function increases strongly for estimations at the margins. This behavior is 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 is tackled by a time-varying penalization. In detail, the penalization is increased linearly to the margins, where the increase is such that the cumulative *loss* is minimized. It is 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 chapter shows that the degree of the spline strongly influences its properties in the frequency domain and points 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 a tradeoff between a precise gain function of the estimations for most periods and a low excess variability at the margins.

Finally, this chapter demonstrates the effects of this time-varying penalization for a real time series. To this point the trend of the Swiss GDP is estimated. There are 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 shows 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 chapter cannot completely eliminate the undesired volatility at the margins. Nevertheless, it can be shown that it is possible to improve the precision 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.

5 Reducing the Excess Variability of the Hodrick-Prescott Filter by Flexible Penalization[10]

Summary

In most applications of the Hodrick-Prescott filter for quarterly data a penalization of 1600 is selected. Although this value is often criticized, it can be justified by frequency domain considerations, as it describes the trend by periodicities above nine to ten years. Given this penalization, this chapter aims at reducing the excess variability at the margins by a time-varying penalization. The previous chapter focuses on approximating an ideal gain function, so that it is assumed that trend and cycle can be strictly separated by a certain cut-off frequency. However, as there is no general, precise definition of the duration of business cycles, it can be argued that the abrupt transition of an ideal gain function from one to zero is not necessary (Pollock 2000 p.318). As a consequence, this chapter does not regard an ideal gain function as optimal, but the gain function of the Hodrick-Prescott filter in the middle of the series, in particular as it exhibits no abrupt transition from one to zero. Thus, the major difference to chapter 4 is the reference to the gain function in the middle of the series instead of an fictive, ideal gain function.

[10]This chapter refers to Blöchl (2014a).

5.1 Introduction

The Hodrick-Prescott filter (Hodrick/Prescott 1997) is one of the most popular tools for trend estimation in economics (Flaig/Wollmershäuser 2007 p.17). Its advantages are clearly an easy and numerical fast and stable implementation, while the shape of the estimated trend completely depends on the choice of a single penalization parameter λ. In most applications λ is set to 1600 for quarterly data according to the suggestion of Hodrick/Prescott (1997), which can be seen as an "industry standard" (Flaig 2012 p.23) for economic trend estimation. However, this choice is often criticized in the literature as dubious (Danthine/Giradin 1989), not data driven (e.g. Schlicht 2005, Kauermann et al. 2011) and too low for most economic time series (Mc Callum 2000, Flaig 2012). As an alternative Schlicht (2005) describes how the Hodrick-Prescott filter can be incorporated into a mixed model framework to derive a data driven estimation of λ. This approach is based on the assumption of a white noise business cycle, but it was shown that it can be extended to account for autocorrelated residuals (e.g. Proietti 2007, Paige 2010). Another possibility to choose λ is suggested by Flaig (2012), who proposes to select λ such high, that the trend component doesn't feature any cyclical behavior any more. Nevertheless, the choice of $\lambda = 1600$ can be justified by frequency domain considerations. Using quarterly data this selection implies that the trend approximately consists of oscillations with periodicities above nine to ten years (Tödter 2002, Maravall/del Rio 2001), which is reasonable according to economic conceptions of trend and cycle (e.g. Burns/Mitchell 1946, Baxter/King 1999).

The estimation of trend and cycle with the Hodrick-Prescott filter by frequency domain aspects exhibits unsolved problems. One massive problem is the increasing variability of the trend estimation to the margins, called excess variability. A rising excess variability means the filter cannot suppress high frequencies for the first and last couple of estimations any more. This induces an increasing volatility of the estimated trend at the margins compared to the rest of the series, making the trend estimations for the first and last periods more volatile than desired, which heavily reduces their reliability. This is a drawback, as researchers and politicians are mainly interested in the trend of the most recent periods that is deterred by the excess variability.

An often applied approach to solve the problem of the excess variability is to use ARIMA models. ARIMA models are employed to derive forecasts that are attached to the time series. This way the original margin of the series moves closer to the middle of the data and is thus less affected by the excess variability. However, this method exhibits the drawback that the forecasts feature failures that rise with an increasing forecast horizon. Consequently, also the estimated trend is subject to this uncertainty.

This chapter tackles the problem of the excess variability using a time-varying penalization and spectral analysis. A time-varying penalization is introduced for the HP-filter by Razzak/Richard (1995) to account for structural breaks and by Crainiceanu et al. (2005) for penalized splines (O'Sullivan 1986) within a mixed model framework. It is shown in this chapter that the volatility at the margins can be reduced by letting λ increase to the ends of the series. Given that the estimation of trend and cycle is motivated by frequency domain considerations, the gain function of the Hodrick-Prescott filter in the middle of the series is regarded as optimal. The flexible penalization is selected such that the gain function for the estimations at the margins is adjusted to the one in the middle.

This chapter does not discuss the selection of λ in general. Instead, it sticks to the standard choice of $\lambda = 1600$ and shows how the excess variability can be reduced for this selection. The first section of this chapter briefly reviews the Hodrick-Prescott filter. Afterwards the features of the filter in the frequency domain are examined, especially its characteristics at the margins. Then it is shown how the gain function can be used as a measure for the excess variability and how flexible penalization can reduce the increasing volatility at the margins. Finally, this chapter gives some empirical examples and points out the different implications between the results of the flexible penalization and the standard approach.

5.2 The Hodrick-Prescott filter

5.2.1 General framework

The Hodrick-Prescott filter (HP-filter) decomposes a time series $\{y_t\}_{t=1}^{T}$ into two components

$$y_t = \mu_t + c_t, \tag{5.1}$$

where μ_t is the trend and c_t is the rest, usually the sum of cycle and irregular effects. μ_t is estimated by solving the following minimization problem:

$$\min_{\mu_t} \sum_{t=1}^{T} (y_t - \mu_t)^2 + \lambda \sum_{t=2}^{T-1} [(\mu_{t+1} - \mu_t) - (\mu_t - \mu_{t-1})]^2. \tag{5.2}$$

The minimization problem consists of two parts. The first one is the squared deviation of the trend from the original series. Minimizing the first part yields a trend that is identical to $\{y_t\}_{t=1}^{T}$. The second part consists of the squared second differences and is a measure for the volatility of the trend. Minimizing the second part yields a linear trend. Obviously, there is a tradeoff between both parts. This tradeoff is solved by the penalization parameter λ that puts weight on the second part. Thus, the smoothness of the estimated trend can be completely regulated by the selection of λ. A low value of λ generates a flexible trend, whereas high values induce a smooth trend.

Solving the minimization problem in (5.2) yields the filter in matrix notation (Mc Elroy 2008, Flaig 2012 p.16):

$$\hat{\boldsymbol{\mu}} = (\boldsymbol{I} - \lambda \boldsymbol{\Delta}' \boldsymbol{\Delta})^{-1} \boldsymbol{y}, \tag{5.3}$$

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

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

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

5.2.2 The Hodrick-Prescott filter in the frequency domain

To examine the characteristics of the HP-filter in the frequency domain, it is reasonable to consider the structure of its filter weights at first. Given formula (5.3) the filter weights are contained in the matrix $(\boldsymbol{I} - \lambda\boldsymbol{\Delta}'\boldsymbol{\Delta})^{-1} \hat{=} \boldsymbol{H}(\lambda) \in \mathbb{R}^{T \times T}$, where the t^{th} row of $\boldsymbol{H}(\lambda)$ contains the weights for the estimation $\hat{\mu}_t$, i.e.

$$\hat{\mu}_t = \sum_{j=1}^{T} h_{tj} y_j. \tag{5.4}$$

h_{tj} is the j^{th} element of the t^{th} row of $\boldsymbol{H}(\lambda)$. An important feature of this weight matrix is that the filter weights have a very similar, (almost) symmetric structure for estimations in the middle of the data, while this structure changes to the margins. This can be seen in Figure 5.1 that plots the weights for different estimations for a HP-filter with $\lambda = 1600$ that is applied to a series with 100 observations.

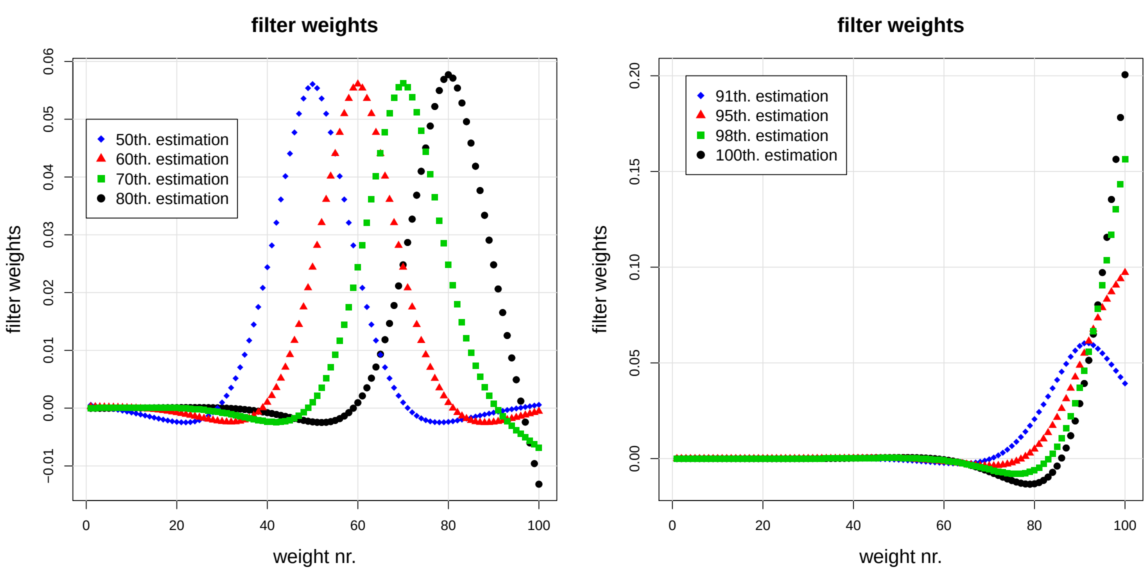

Fig. 5.1: Filter weights in the middle and at the margin

The filter weights in the middle look similar and converge symmetrically to zero. To the margins the filter weights become more and more asymmetric. Furthermore, the highest weight of the estimations at the margins is far above of those around the middle. This means the estimations at the margins are much more affected by single observations than those closer to the middle of the data. This behavior of the filter weight structure causes the excess variability at the margins, as a single observation can heavily influence the estimated trend.

The change in the weight structure to the margins also becomes obvious, when the HP-filter is considered in the frequency domain. In the frequency domain a time series is interpreted as the superposition of oscillations with different frequencies (for a detailed discussion see Harvey (1993), Hamilton (1994) or Mills (2003)), where the trend as the long run development of a series is supposed to consist of those oscillations with a high periodicity. The HP-filter as a tool for trend estimation extracts oscillations with high periodicities and eliminates those with lower periodicities. This behavior can be described by the gain function. Given the filter weights h_{tj} that arise for a certain value of λ, the gain for estimation $\hat{\mu}_t$ and frequency ω can be calculated as (e.g. Mills 2003 p.80)

$$g_t(\omega, \lambda) = \sqrt{\left(\sum_{j=1-t}^{T-t} h_{t,j+t}\cos(\omega j)\right)^2 + \left(\sum_{j=1-t}^{T-t} h_{t,j+t}\sin(\omega j)\right)^2}\,. \tag{5.5}$$

The gain is interpreted as the factor by which the amplitude of an oscillation with a certain frequency is damped or amplified by a filter. To show the effects of the changing filter weight structure, Figure 5.2 displays the gain functions for different estimations.

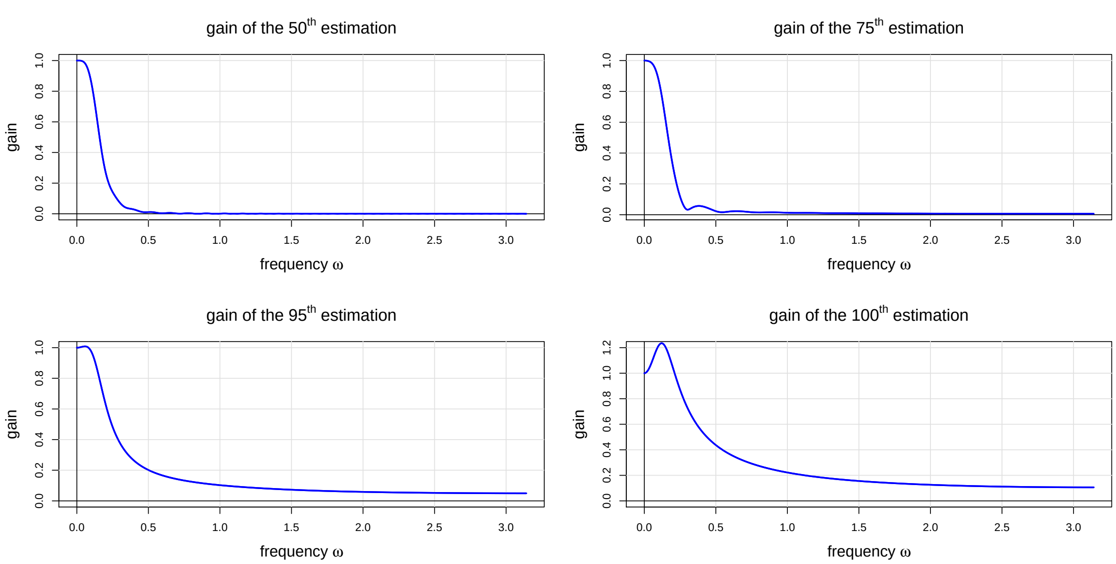

Fig. 5.2: Gain functions for different estimations

For the 50^{th} and 75^{th} estimation the gain functions are very similar, since they are induced by an almost equal weight structure. However, with a decreasing distance to the margins the gain function starts to change like for the 95$^{\text{th}}$ and 100$^{\text{th}}$ estimation. For the estimations at the margins high frequencies cannot be completely eliminated any more, which causes an increasing volatility of the resulting trend function. This increase of the volatility is known as the excess variability. To examine the excess variability it is not practicable to consider the gain functions for all estimations of the time series. It is much more practicable to consider the deviation of the gain function of a certain estimation $\hat{\mu}_t$ from the one of the estimation in the middle of the series. Let $g_m(\omega, \lambda)$ denote the gain for frequency ω for the middle estimation $\hat{\mu}_m$, where $m = T/2$, or the next integer if T is odd, and $g_t(\omega, \lambda)$ the one for estimation $\hat{\mu}_t$, then a *loss* can be defined:

$$l(t, \lambda) = \int_0^\pi [g_m(\omega, \lambda) - g_t(\omega, \lambda)]^2 d\omega. \tag{5.6}$$

$l(t, \lambda)$ is the squared deviation of the gain for estimation $\hat{\mu}_t$ from the one for the middle estimation in the interval $[0, \pi]$. The calculation of (5.6) for continuous frequencies is difficult, however it can be easily approximated by a sufficient high number of discrete frequencies, e.g. for $\boldsymbol{\omega} = (0, 0.001, 0.002, ..., \pi)' \in \mathbb{R}^{n \times 1}$:

$$l(t, \lambda) = \sum_{i=1}^{n} [g_m(\omega_i, \lambda) - g_t(\omega_i, \lambda)]^2 \cdot \delta. \tag{5.7}$$

n is the number of elements in $\boldsymbol{\omega}$ and δ is the distance between the elements, i.e. $\delta = \omega_j - \omega_{j-1}$. Calculating the *loss* for all $t = 1, ..., T$ gives an overview of which estimations are affected by the increase of the variability. An important question is which factors influence the excess variability. To shed light on this issue first of all a HP-filter with $\lambda = 1600$ is applied to time series of length 50, 100, 150 and 200. Then the *loss* is calculated for each element of these series, which is henceforth denoted as the *loss function*.

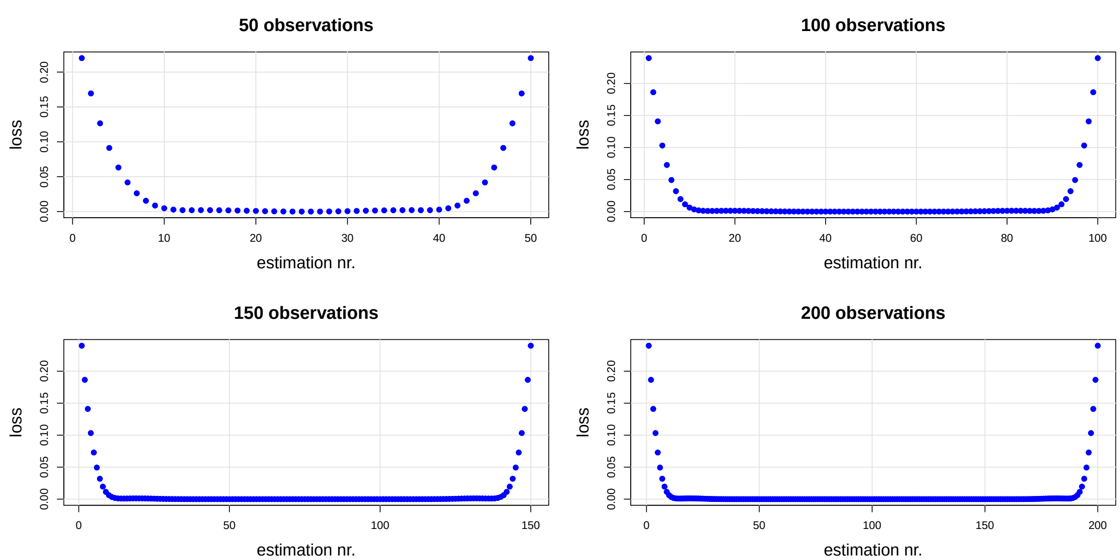

Fig. 5.3: *loss function* for different numbers of observations

Figure 5.3 shows that the $loss$ is very similar and almost zero for the estimations around the middle of the data and starts to increase abruptly at the margins, which is in line with Figures 5.1 and 5.2 that indicate that the gain of the HP-filter only changes for the estimations close to the ends of the series. Independent of the number of observations the first and last ten or eleven estimations exhibit a rise in the $loss$. Consequently, the excess variability does not depend on the number of observations.

However, it can be shown that the number of affected estimations depends on the value of λ. To this regard Figure 5.4 displays the $loss\ functions$ for different values of λ and a series with 100 observations:

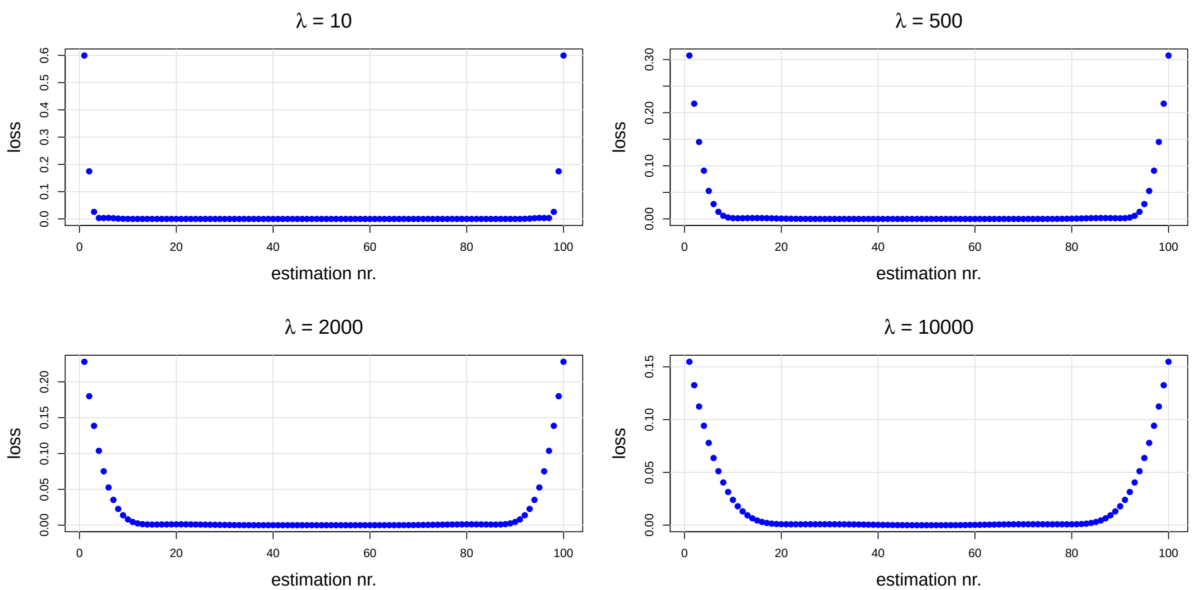

Fig. 5.4: $loss\ function$ for different values of λ

In contrast to the length of the series, the value of λ affects the number of estimations that show an increased excess variability. This number rises with the value of λ. For $\lambda = 10$ about the first and last three estimations are affected, for $\lambda = 500$ about the first and last eight estimations, and for $\lambda = 10000$ more than the first and last 20 estimations show an increased $loss$. But while the number of affected estimations rises with increasing values of λ, the degree of the excess variability is worse for lower values of λ. For $\lambda = 10000$ the $loss$ of the last estimation is about 0.15, while for $\lambda = 10$ it is about 0.6. However, the fact that the excess variability depends on λ is of subordinate importance here, as this chapter only focuses on the standard case of $\lambda = 1600$.

5.3 Reducing the excess variability

5.3.1 Introducing a flexible penalization

Given the *loss function* it is possible to quantify and describe the excess variability. The next step is to find techniques to reduce this variability. To this point a time-varying penalization for the HP-filter is introduced that allows selecting different values of λ at different points in time. According to Figure 5.3, the first and last ten or eleven estimations show an increased *loss*, when a HP-filter with $\lambda = 1600$ is applied. Consequently, the penalization should rise for the estimations at the margins of the series. This can be done easily by changing the model framework of the HP-filter slightly. Given formula (5.3) the scalar λ has to be replaced by a vector $\boldsymbol{\lambda} \in \mathbb{R}^{(T-2)\times 1}$, where $\boldsymbol{\lambda}' = (\lambda_1, ..., \lambda_{T-2})$ and the matrix $\boldsymbol{K} = \text{diag}(\boldsymbol{\lambda})$ is constructed. Then the HP-filter with a flexible penalization can be written as

$$\hat{\boldsymbol{\mu}} = (\boldsymbol{I} - \boldsymbol{\Delta}'\boldsymbol{K}\boldsymbol{\Delta})^{-1}\boldsymbol{y}. \tag{5.8}$$

As the HP-filter is equal to a penalized spline of order one with a truncated polynomial basis and knots at every point in time $t = 1, ..., T$ (Paige 2010), it can be interpreted as a continuous connection of lines, where the slope of the lines changes at the points in time $t = 2, 3, ..., T-1$. To this regard the penalization regulates to what extend the slope can change at these points in time. High values of λ let the slope only change slightly, which results in a smooth trend estimation, whereas low values for the penalization allow large changes of the slope, inducing a flexible trend function. λ_1 regulates the degree to what the slope of the trend function can change at $t = 2$ and λ_{T-2} determines the change of the slope at the point in time $T-1$. In general λ_t regulates the change of the slope at the point in time $t+1$. In order to reduce the excess variability the values of $\boldsymbol{\lambda}$ are increased to the margins. This is explained in detail in the next sections.

5.3.2 Direct method

A question that arises is how the penalization should increase to the margins. The general aim is to reduce the *loss* at the margins without increasing it in the middle of the data. To this point a criterion is defined, which will turn out to be suitable for this purpose. The penalization at the margins is increased such that the cumulative *loss* of all estimations is minimized. This is reasonable as one is usually not just interested in the trend in the middle, but in the whole series. Moreover, this section shows that this criterion leads to a reduced excess variability at the margins without strongly affecting the estimations in the middle of the series. Defining the cumulative *loss* in dependence of $\boldsymbol{\lambda}$ as $L(\boldsymbol{\lambda})$ yields:

$$L(\boldsymbol{\lambda}) = \sum_{t=1}^{T} l(t, \boldsymbol{\lambda}), \tag{5.9}$$

$$\text{where } \; l(t, \boldsymbol{\lambda}) = \sum_{i=1}^{n} [g_m(\omega_i, \lambda) - g_t(\omega_i, \boldsymbol{\lambda})]^2 \cdot \delta. \tag{5.10}$$

Here, $g_t(\omega_i, \boldsymbol{\lambda})$ is the gain of the t^{th} estimation for frequency ω_i using the flexible penalization and $g_m(\omega_i, \lambda)$ is the gain for the estimation in the middle of the series, when a single, fixed penalization parameter is used. This criterion is subject to the conditions that the penalization is 1600 in the middle and that it rises to the margins. Considering the *loss function* over all estimations suggests that a linear increase of the penalization to the margins might be appropriate. As a consequence the last k values of $\boldsymbol{\lambda}$ rise by

$$\lambda_{T-2-k+j} = 1600 + \alpha j, \;\; j = 1, ..., k. \tag{5.11}$$

The intercept can be seen as given, since λ is set to 1600 for the estimations in the middle. As the penalization needs to increase to both margins the first k values of the penalization are defined as

$$\lambda_1 = \lambda_{T-2}, \; \lambda_2 = \lambda_{T-3}, ..., \; \lambda_k = \lambda_{T-1-k}. \tag{5.12}$$

Consequently, the first and last k values of $\boldsymbol{\lambda}$ are defined by (5.11) and (5.12), while all others are still set to 1600. Beside a linear increase of the penalization also other functions like a quadratic or cubic increase or a polynomial of degree two could be assumed. However, the results would hardly change, but especially the polynomial function would make the calculation much more expensive. Minimizing (5.9) with respect to (5.11) and (5.12) can be done by algorithms like fisher scoring or Newton-Raphson. However, the minimization by these algorithms can only be done for α as k is an integer. Thus, $L(\boldsymbol{\lambda})$ is minimized for different values of k and that value is selected that yields the lowest value of $L(\boldsymbol{\lambda})$.

Applying this algorithm to a simulated time series with 100 observations yields $k = 27$ and $\alpha = 1294.72$. The penalization parameter for the last 27 λ's then rises by $\lambda_{71+j} = 1600 + 1294.72 \cdot j, \quad j = 1, ..., 27$, while $\lambda_1, ..., \lambda_{27}$ are defined according to (5.12) (This chapter focuses on the case of $\lambda = 1600$, but a table with corresponding values for k and α for other values of λ is provided in appendix 5.A). Table 5.1 shows the *loss* for the fixed and the flexible penalization for the middle of the data and the margin (50^{th} and 100^{th} estimation) as well as $L(\boldsymbol{\lambda})$.

penalization	**50^{th}estimation**	**100^{th}estimation**	**$L(\boldsymbol{\lambda})$**
fixed	0	0.23956	1.76382
flexible	0.00015	0.09078	1.16872

Tab. 5.1: *loss* and cumulative *loss* for fixed and flexible penalization

Table 5.1 shows that the flexible penalization reduces the *loss* for the 100^{th} estimation as well as $L(\boldsymbol{\lambda})$, while it increases the one for the 50^{th} estimation only slightly. $L(\boldsymbol{\lambda})$ can be reduced by around 34 percent and for the 100^{th} estimation the *loss* even declines by 62 percent. This indicates that the excess variability can be reduced without affecting the estimations in the middle of the data. To get a complete picture of how the flexible penalization changes the gain functions of all estimations Figure 5.5 displays the *loss* for the whole series for the fixed and the flexible penalization.

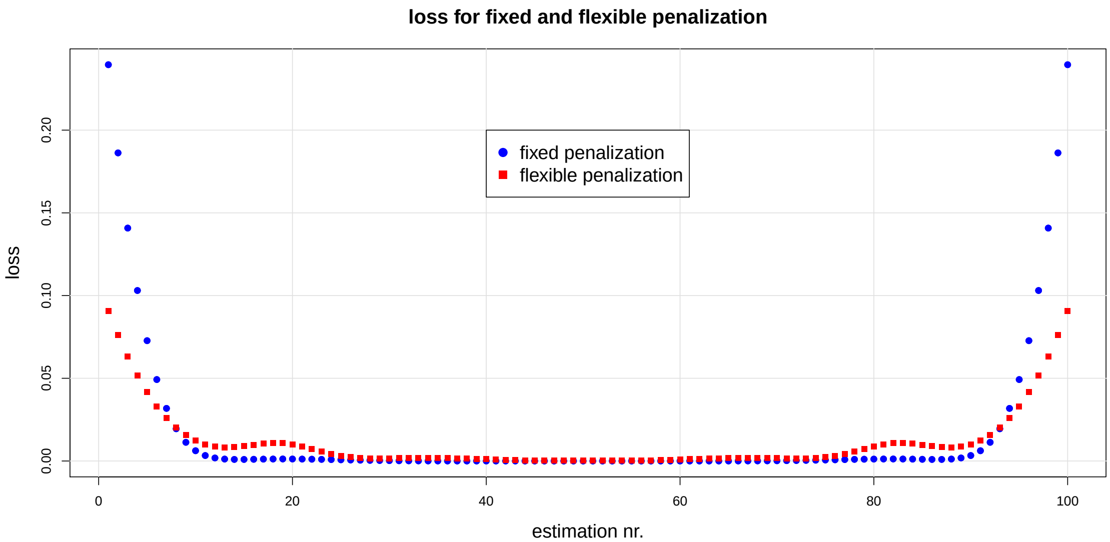

Fig. 5.5: *loss function* for a fixed and a flexible penalization

Clearly, the *loss* strongly decreases for about the first and last six estimations. While the *loss* is strongly reduced at the margins, it is slightly increased by the flexible penalization for about the estimations 10-20 and 80-90. However, the decrease at the margins outweighs this slight increase, which results in the strong decrease of the cumulative *loss*. Thus, the time-varying penalization offers a tool to reduce the excess variability without strongly affecting the trend in the middle of the data (note that Bruchez (2003) suggests an ad-hoc approach to reduce the excess variability of the HP-filter). Section 5.4 applies this method to real time series and shortly discusses the implications that arise, when the time-varying penalization instead of the fixed one is used.

5.3.3 Indirect method

The direct method focuses directly on the gain function. As this approach requires the calculation of a high number of gain functions within the optimization, it can become numerical expensive, especially for long time series. A numerical much cheaper way of implementing the flexible penalization is the indirect method that focuses on the filter weights instead of the gain function. An often used measure for the degree of fit of linear smoothers is $df(\lambda) = \mathrm{tr}\,[\boldsymbol{H}(\lambda)]$ (e.g. Ruppert et al. 2003 p.81). A low value of df indicates a high degree of smoothing. This can be explained by the behavior of the filter weights with respect to the value of λ. For linear smoothers like the HP-filter the filter weights h_{tj} for an estimation $\hat{\mu}_t$ symmetrically decrease to zero as $|t-j| \to \infty$, where the rate of the convergence depends on the value of λ (Dagum/Luati 2004). For high values of λ the rate is rather slow, which induces a smooth trend estimation, while for low values of λ the rate of the convergence is fast. This can be seen in Figure 5.6, which plots the filter weights of a HP-filter with different values of λ that is applied to a series with 100 observations.

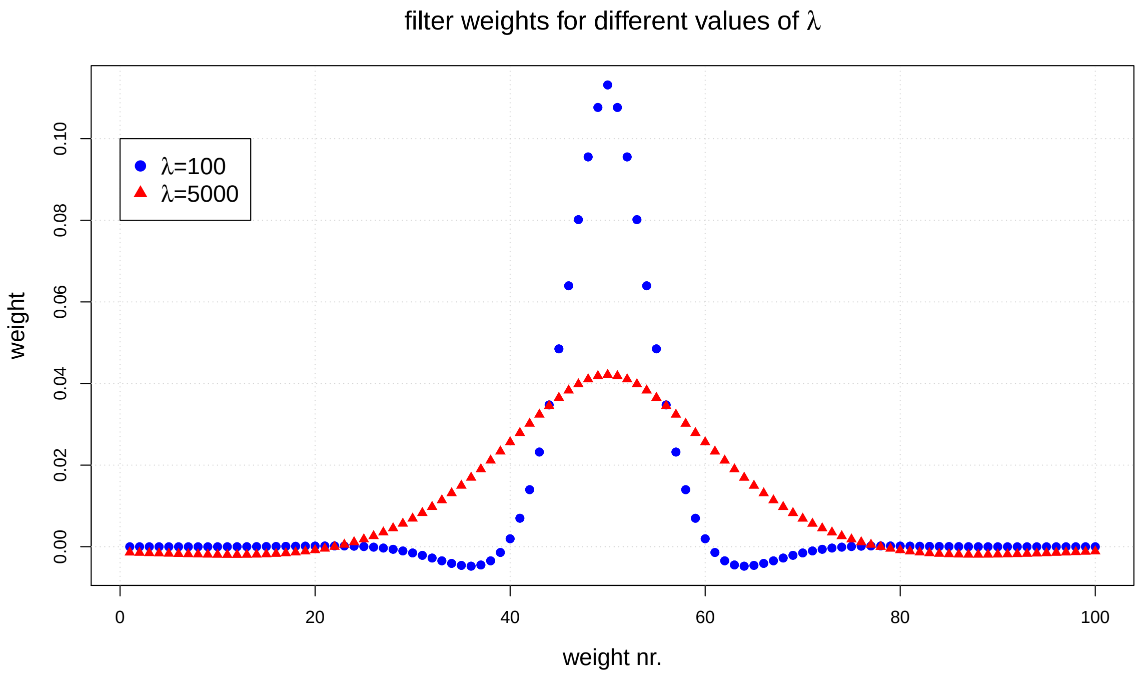

Fig. 5.6: Filter weights of a HP-filter for different values of λ

The filter weights in Figure 5.6 refer to the 50$^{\text{th}}$ estimation. For $\lambda = 100$ the filter weights decrease much faster to zero than for $\lambda = 5000$. Since the filter weights of the HP-filter furthermore sum up to one (Eubank 1983) the middle weights h_{tt} are rather high for low values of λ and rather low for high values of λ. Consequently the filter weight h_{tt} can be seen as a measure for the degree of smoothing of estimation $\hat{\mu}_t$ and thus as an indicator for its gain function.

The indirect approach uses this fact for a numerical faster implementation. In the direct method the gain function of every estimation $\hat{\mu}_t$ is compared to the one of the estimation in the middle $\hat{\mu}_m$ and the cumulative squared deviation is calculated. Now, a numerical much cheaper approach can be implemented. According to the notation in this chapter h_{mm} is the middle weight of the estimation $\hat{\mu}_m$. In the ideal case where there is no excess variability at the margins all weights h_{tt}, $t = 1, ..., T$ should be equal to h_{mm}. This implies that in the ideal case $df^*(\lambda) = T \cdot h_{mm}(\lambda)$.

Now, the indirect approach approximates this ideal measure to the real degree of smoothing df. In this sense the increase of the flexible penalization is such that the deviation between df^* and $df(\boldsymbol{\lambda}) = \mathrm{tr}\left[\boldsymbol{H}(\boldsymbol{\lambda})\right]$ is minimized. Here, $h_{mm}(\lambda)$ refers to the HP-filter with a fixed value of λ, while $\boldsymbol{H}(\boldsymbol{\lambda})$ is the hat matrix of the HP-filter with a flexible penalization. To put higher weight on larger deviations the squared diagonal elements of the hat matrix can be used for the calculation, i.e. $df_{sq}(\boldsymbol{\lambda}) = \mathrm{tr}\left[\boldsymbol{H}(\boldsymbol{\lambda})^2\right]$ and $df^*_{sq}(\lambda) = T \cdot h_{mm}(\lambda)^2$. The minimization problem of the indirect approach is finally given to:

$$\min_{\boldsymbol{\lambda}} |df^*_{sq}(\lambda) - df_{sq}(\boldsymbol{\lambda})|, \tag{5.13}$$

$$\text{subject to } \lambda_{T-2-k+j} = 1600 + \alpha j, \quad j = 1, ..., k,$$

$$\text{and } \lambda_1 = \lambda_{T-2}, \lambda_2 = \lambda_{T-3}, ..., \lambda_k = \lambda_{T-1-k}.$$

The deviation of the cumulative *loss* between the direct and the indirect method is in most cases lower than 0.1 percent. However, the time for the optimization is reduced immensely by the indirect method, which is helpful as the optimization has to be done for different values of k in order to find the ideal increase of the filter weights.

5.4 Empirical application

In this section the flexible penalization is applied to empirical time series to point out how the results might change, when a flexible instead of a fixed penalization is employed. First of all the seasonally adjusted quarterly real GDP of Switzerland is considered[11]. The data start in the first quarter 1980 and end in the third quarter 2013 so that there are 135 observations. The minimization of the cumulative *loss* $L(\boldsymbol{\lambda})$ with respect to α and k yields $\alpha = 1304.22$ and $k = 27$. The trend is also estimated by a HP-filter with a fixed penalization of $\lambda = 1600$. The results are shown in Figure 5.7:

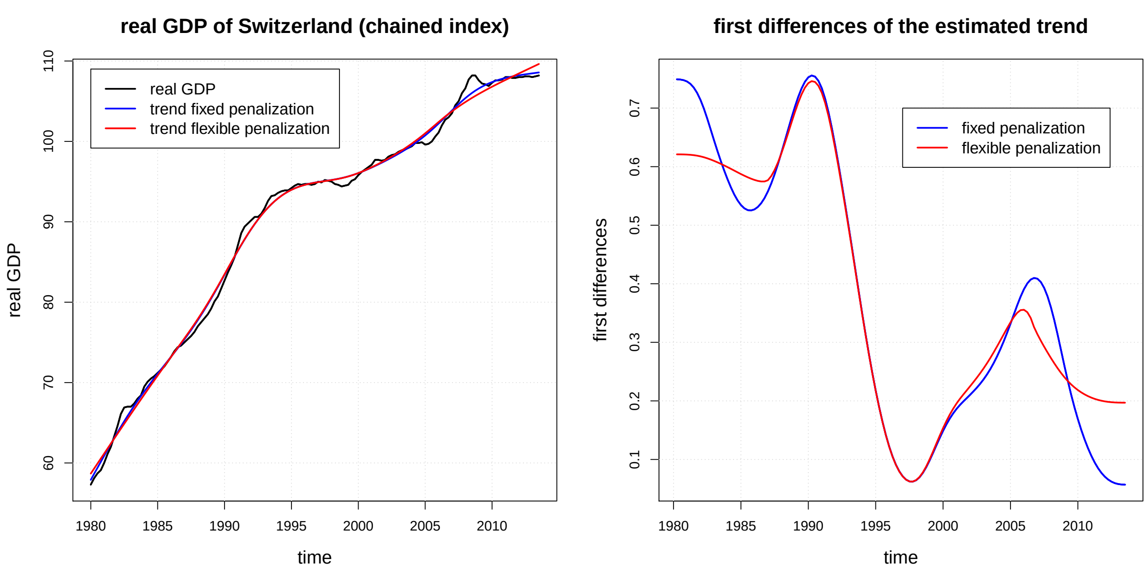

Fig. 5.7: Estimated trend for the Swiss real GDP and first differences of the trend

The growth of the trend according to the fixed penalization has strongly decreased since about 2007. The trend of the flexible penalization exhibits only a slight decline of its growth

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

since 2007 and its growth rate is clearly above the case of the fixed penalization for the last three years. This can also be seen, when the first differences on the right plot are considered. Furthermore, the trend of the flexible penalization is clearly above the one of the fixed penalization for the most recent years. A further interesting feature of the flexible penalization becomes obvious, when the resulting estimates of the business cycle are considered.

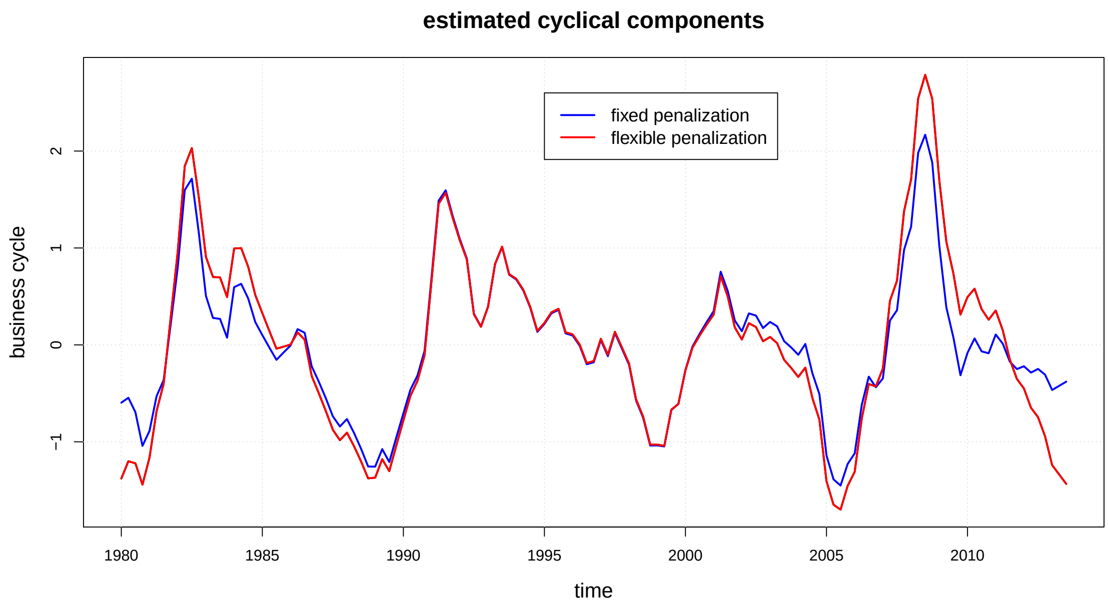

Fig. 5.8: Estimated business cycles for fixed and flexible penalization

Figure 5.8 shows that the output gap at the end of the series is much smaller for the fixed penalization. This smaller output gap is a direct consequence of the excess variability. As Figure 5.1 shows, the filter weights for estimations at the margins are very high for the last observation. Thus, the last observation is the most influential for these estimations, which induces that the estimations at the margin are deterred to the value of the last observation. As a result in this example the excess variability in trend leads to a rather too low output gap at the margin of the series. The cyclical component of the flexible penalization instead shows a far larger output gap at the end of the series due to the reduced excess variability.

Next, the seasonally adjusted quarterly real GDP of Denmark is considered[12]. The data start in the first quarter 1991 and end in the third quarter 2013 so that there are 91 observations in this series. Minimizing the cumulative *loss* $L(\boldsymbol{\lambda})$ yields $\alpha = 1242.48$ and $k = 27$. The trend of the real GDP is estimated using both the flexible and a fixed penalization. The results are shown in the left plot of Figure 5.9.

[12]The data are from the national census bureau of Denmark, http://www.statbank.dk.

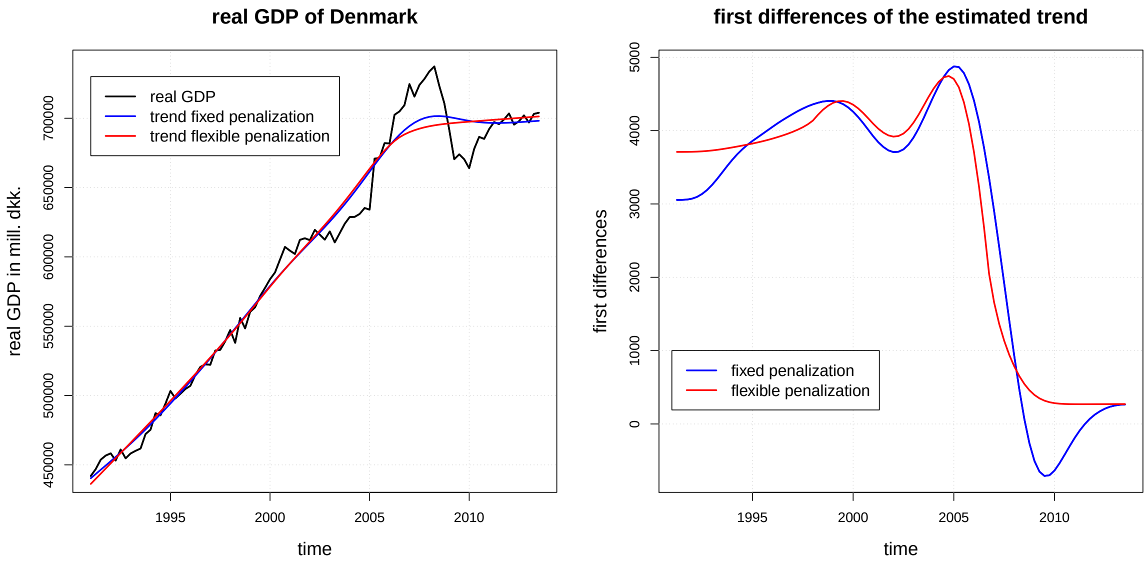

Fig. 5.9: Estimated trend for the Danish real GDP and first differences of the trend

There are clear differences between the estimations. In both cases the growth rate of the trend decreases from about 2005, but the decline is much larger for the fixed penalization. The right plot of Figure 5.9 shows the first differences of the trend estimations. Here, the differences become even more obvious. According to the fixed penalization the growth of the trend even turns negative for the years 2008-2012. In contrast the trend growth rate according to the flexible penalization declines from the year 2005, but seems to have stabilized on a lower level during the last three years.

5.5 Conclusion

The Hodrick-Prescott filter is the probably most widespread instrument for trend estimation in economics. Compared to the competing Baxter-King filter (Baxter/King 1999) it offers the advantage of yielding estimates for the most recent periods. However, the fact that the filter weights strongly change at the margins leads to an increased excess variability for these periods. This means compared to the estimation in the middle of the data the trend at the margins is too volatile. Especially as researchers are predominantly interested in the most recent periods, the excess variability turns out to be a serious problem of the Hodrick-Prescott filter. An existing method to overcome this problem is to use ARIMA models in order to prolong the time series by adding forecasts at the margins. Nevertheless, the predictions exhibit failures that increase with a rising forecast horizon. As for $\lambda = 1600$ more than ten periods have to be predicted, this method seems to be of limited practicability.

This chapter combines spectral analysis with a flexible penalization in order to reduce the excess variability at the margins. To this point the *loss*, i.e. the squared deviation of the gain function from the one in the middle of the series is employed as a measure to describe and quantify the excess variability. To reduce the increased volatility of the estimations at the ends of the series the penalization is allowed to increase linearly to the margins. The exact rise of the penalization is determined such that the cumulative *loss* is minimized. In this regard it is shown that this criterion not only leads to a lower cumulative *loss*, but that it reduces the excess variability at the margins without strongly affecting the estimations closer to the middle of the series.

To show the empirical implications that can arise when the flexible penalization is used instead of a fixed one the HP-filter with a flexible penalization is applied to estimate the trend of the real GDP of Switzerland and Denmark. It is shown that in both cases the estimated trend according to the flexible penalization considerably differs from the trend estimation with the fixed penalization. Especially the estimation for Switzerland using the flexible penalization shows that the current data of the Swiss real GDP do not allow concluding that the trend growth rate has decreased immensely since 2007, like the HP-filter with the fixed penalization suggests.

This chapter offers an approach to improve the precision of the estimations at the margins. Although the excess variability cannot be completely eliminated it is reduced by more than 62 percent for the last estimation. As the empirical examples show, the results of the flexible penalization can strongly differ from the standard approach with a fixed λ of 1600. Given these results, the flexible penalization might be an interesting tool for researchers, since it increases the precision of the estimations for the most recent periods.

Appendix

5.A Parameters of the time-varying penalization for different values of λ

The following table provides the corresponding values of k and α for different values of λ. Please note that the values can vary slightly for series of different length. The table also shows the length of the series to which the values refer, which is denoted as T_{ref}. However, in most cases the results only change slightly, when these values are used for series with a length that deviates from the reference length in this table as long as the series are not too short.

λ	k	α	T_{ref}	λ	k	α	T_{ref}
5	5	19	100	1000	24	880	125
10	6	44	100	1250	25	1165	150
20	8	47	100	1500	26	1414	150
30	9	66	100	1750	27	1617	150
40	10	72	100	2000	27	2279	150
50	10	126	100	2500	28	3198	150
60	10	210	100	3000	29	4073	150
70	12	99	100	3500	31	3713	150
80	12	134	100	4000	32	4207	150
90	12	179	100	4500	33	4606	150
100	13	144	100	5000	34	4904	150
200	15	325	100	6000	38	3617	200
300	17	403	100	8000	41	4452	200
400	19	412	100	10000	43	5611	200
500	20	513	100	12000	46	5673	200
600	21	590	125	14000	47	7133	250
700	21	853	125	16000	50	6660	250
800	22	881	125	18000	51	7752	250
900	23	887	125	20000	52	8796	250

Tab. 5.2: Values for k and α for different values of λ

6 Penalized Splines in the Light of Baxter/King

Summary

This chapter opposes penalized splines with a truncated polynomial basis to the so called Baxter-King filter, which is a widespread instrument for the estimation of trend and cycle in economics. It is investigated, whether penalized splines meet the requirements for an ideal time series filter formulated by Baxter/King (1999). This chapter especially focuses on the ability of penalized splines to extract frequency bands and points out their link to rational square wave filters. In this regard it is shown that penalized splines can be regarded as superior to the Baxter-King filter, when they are used as lowpass filters, but that the Baxter-King filter exhibits advantageous properties as a bandpass filter.

6.1 Introduction

The decomposition of time series into trend and cycle plays an important role in macroeconomics. Here, the conception that trend and cycle exhibit characteristic spectral properties often motivates a reasonable derivation of these components. Given this conception, spectral analysis has become an important tool in economics as it enables to transfer a time series into the frequency domain and to represent it by the superposition of oscillations with different periodicities. It allows describing the business cycle, which is supposed to show a kind of sine pattern, by fluctuations within a certain bandwidth of periodicities. For example Burns/Mitchell (1946) find evidence that the cycle of the U.S. economy exhibits a duration between six and 32 quarters. The trend component as the long run development of the series is usually characterized by fluctuations with high periodicities.

One of the most widespread tools in economics for the extraction of trend and cycle by frequency domain aspects is the so called Baxter-King filter (Baxter/King 1999). To this point Baxter/King (1999) formulate ideal properties for a time series filter in the frequency domain, which are the accurate extraction of frequency bands, the elimination of unit roots and the absence of phase shifts. They develop the Baxter-King filter as a tool that meets these requirements. Nevertheless, this filter still suffers from diverse shortcomings. A massive one is that the filter is not able to render estimates for the most recent periods, which are often the most important for researchers.

In this regard penalized splines (O'Sullivan 1986, Eilers/Marx 1996) as asymmetric filters are generally able to yield estimations for the most recent periods. They are closely related to the Hodrick-Prescott filter (Hodrick/Prescott 1997, Paige 2010), where the outcome predominantly depends on the selection of a single penalization parameter λ. Penalized splines have become popular tools for trend estimation in economics, since they allow a data-driven estimation of trend and cycle by generalized cross validation (Hasti/Tibshirani 1990, also Kohn et al. 1992, Wang 1998), or the incorporation into a mixed model framework (Brumback et al. 1999, also Ruppert et al. 2003). This chapter focuses on penalized splines in the frequency domain and investigates, whether they meet the requirements of Baxter/King. To this point splines with a truncated polynomial basis (Brumback et al. 1999) are used due to their fast and easy implementation as well as their link to rational square wave filters (Pollock 2000, 2003). It is shown that these splines asymptotically meet the requirements of Baxter/King and exhibit advantageous features as time series filters for trend estimation.

The chapter is structured as follows. The first section summarizes theory about time series filters in the frequency domain. Afterwards the Baxter-King filter as well as penalized splines with a truncated polynomial basis are explained briefly. The next sections discuss in how far penalized splines meet the requirements for ideal filters formulated by Baxter/King (1999). Finally, the ability of splines to extract frequency bands is compared to the one of the Baxter-King filter.

6.2 Ideal filters in the frequency domain

Given the conception that trend and cycle can be distinguished by their spectral characteristics, spectral analysis is a helpful device for the construction of time series filters. For a detailed discussion of spectral analysis see Granger/Hatanaka (1964), Harvey (1993), Hamilton (1994) or Mills (2003). The basic idea of spectral analysis is to transfer a time series $\{y_t\}_{t=1}^{T}$ into the superposition of oscillation with different frequencies. From this it is possible to evaluate the effects of an instrument for the estimation of trend and cycle on the time series in the frequency domain. There are basically two effects characterized by the gain and the phase. The gain can be interpreted as the factor by which the amplitude of an oscillation is weighted, when an instrument for the extraction of trend or cycle is applied. Given the notation of such an instrument as a linear filter (e.g. Harvey 1993 p.189)

$$x_t = \sum_{j=-q}^{s} h_j y_{t-j}, \tag{6.1}$$

it is calculated for given filter weights h_j and a frequency ω as (e.g. Mills 2003 p.80)

$$g(\omega) = \sqrt{\left[\sum_{j=-q}^{s} h_j \cos(\omega j)\right]^2 + \left[\sum_{j=-q}^{s} h_j \sin(\omega j)\right]^2}. \tag{6.2}$$

A gain of zero induces that a frequency ω is completely suppressed and a gain of one implies that it is not affected by the filter. Beside the change of the amplitude of certain oscillations linear filters can induce phase shifts. This means the peaks and bottoms of an oscillation are shifted along the time axis. The phase is given as (e.g. Mills 2003 p.81)

$$\theta(\omega) = \tan^{-1} \frac{\sum h_j \sin(\omega j)}{\sum h_j \cos(\omega j)}. \tag{6.3}$$

Figure 6.1 exemplary shows the effects of a filter on a time series in the frequency domain:

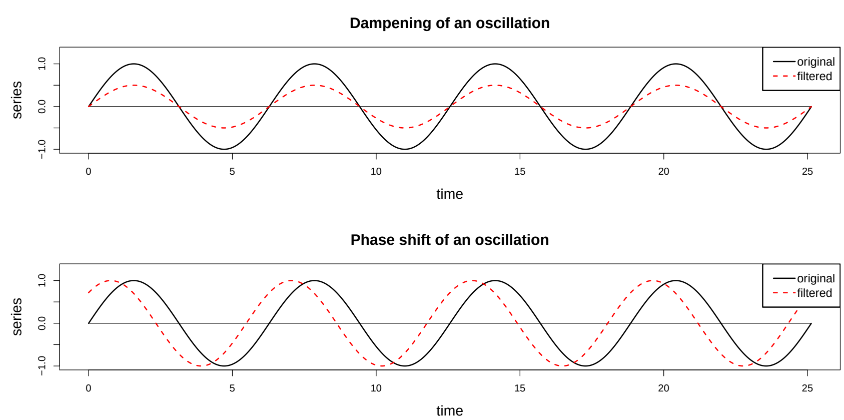

Fig. 6.1: Effects of linear filters in the frequency domain
(c.f. King/Rebelo 1993 p.215)

Considering the gain function, i.e. the gain over a bandwidth of frequencies, usually $[0, \pi]$, gives information about what frequencies are extracted by the filter. An ideal instrument for trend estimation suppresses oscillations with high frequencies, while it does not affect those with lower frequencies. Because it leaves low frequencies unchanged it is called lowpass filter. Two further ideal filter types are the highpass filter and the bandpass filter. The highpass filter completely suppresses low frequencies and leaves all others unchanged so that it eliminates the trend. The bandpass filter completely suppresses high and low frequencies and is applied to extract the cyclical component. Figure 6.2 shows examples for the gain functions of the ideal filter types (c.f. Schlittgen/Streitenberg 2001 p.173):

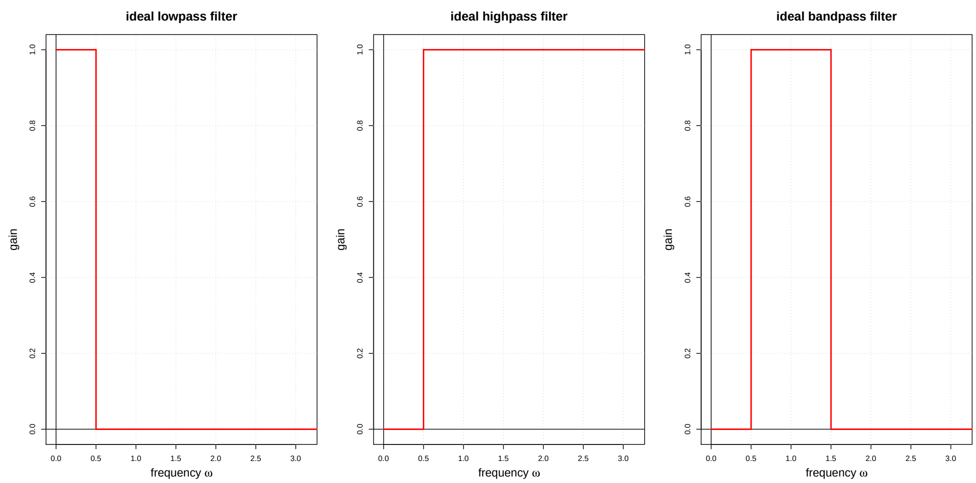

Fig. 6.2: Gain functions of ideal filters

Given the effects of linear filters on series in the frequency domain, Baxter/King (1999) postulate features for time series filters. The first one is that time series filters should exhibit a gain function similar to an ideal gain function. It is not possible to completely realize an ideal gain function for finite time series (Oppenheim/Schafer 1989). In this regard Baxter/King urge that a lowpass filter should at least exhibit a gain of one for $\omega = 0$. This condition holds for the lowpass filter, if the filter weights sum up to one so that $\sum_{j=q}^{s} h_j = 1$ (Baxter/King 1999). The highpass filter should at least fulfil $g(0) = 0$, which is the case, if the weights sum up to zero, i.e. $\sum_{j=q}^{s} h_j = 0$.

A further condition for a time series filter is that it should not induce phase shifts, as this might change the original relation between the filtered and other series. To this point a linear filter does not induce phase shifts, if its filter weights are symmetric (e.g. Mills 2003 p.81). Thus, the filter must fulfil $q = s$ as well as $h_j = h_{-j} \; \forall \; j = 1, 2, 3, ...$

Finally, the series has to be stationary after the trend component is removed, even if the original series is not stationary, i.e. includes a deterministic time trend or unit roots. It can be shown that linear filters eliminate quadratic time trends as well as stochastic trends of order two, if they are symmetric and their filter weights sum up to zero (Osborn 1995, Baxter/King 1999).

Based on these conditions Baxter/King (1999) develop the widespread Baxter/King-filter, which is summarized in the next section. Afterwards it is examined in how far penalized splines meet these requirements for time series filters.

6.3 Lowpass filters

6.3.1 The Baxter-King filter

One of the most popular filters in economics is the so called Baxter-King filter (Baxter/King 1999, henceforth denoted as BK-filter). The basic idea is to construct a linear filter

$$x_t = \sum_{j=-\infty}^{\infty} b_j y_{t-j},, \tag{6.4}$$

that has a gain function with minimal deviation to an ideal gain function of a lowpass filter with a certain cut-off frequency ω^{cf}. The filter shall further be symmetric in order to avoid phase shifts. Let $g^*(\omega)$ be the ideal gain for frequency ω and $g(\omega)$ the one of the filter, then the BK-filter shall minimize

$$\int_0^{\pi} [g^*(\omega) - g(\omega)]^2 \, d\omega, \tag{6.5}$$

subject to

$$b_j = b_{-j} \;\; \forall \, j = 1, 2, \dots \,.$$

Thus, the BK-filter minimizes the squared deviation between ideal and real gain function. In the case of an ideal infinite filter the filter weights are derived by the inverse Fourier transformation of the ideal gain function (c.f. Baxter/King 1999 p.577 and 592-593).

$$b_0 = \frac{\omega^{cf}}{\pi} \quad \text{and} \quad b_j = \frac{\sin(j\omega^{cf})}{j\pi}, \;\; \forall \, j = 1, 2, \dots \,.$$

Of course an infinite linear filter is not practicable so that the length of the filter weights has to be limited to $N = 2n + 1$. It can be shown that for a real finite filter the weights a_j are equal to the ideal weights b_j (Baxter/King 1999, Sargent 1987). Baxter/King furthermore demand that the gain of the filter is one for $\omega = 0$. This implies that $\sum\limits_{j=-n}^{n} a_j = 1$, which finally results in

$$a_j^* = a_j + \eta, \quad \text{where} \quad \eta = \frac{1 - \sum\limits_{j=-n}^{n} a_j}{2n + 1}. \tag{6.6}$$

The decisive feature of the BK-filter is the selection of n. A high value of n results in a good approximation of the ideal gain function. However, the BK-filter cannot yield estimations for the first and last n periods of the series. This is a drawback, as researchers often are predominantly interested in the most recent periods. Thus, there is a tradeoff between a good approximation of the ideal gain function and receiving estimates for the most recent periods. To this regard Baxter/King (1999) suggest to use $n = 12$ for quarterly data, as this value would yield a sufficient gain function without losing too much periods.

6.3.2 Penalized splines

This chapter describes briefly the most important facts about penalized splines. It is focused on tp-splines that denote splines with a truncated polynomial basis, which trace back to Brumback et al. (1999). For a detailed discussed see also Ruppert et al. (2003). Even if tp-splines tend to be numerical instable (Fahrmeir et al. 2009 p.303), they are advantageous due to their relative easy implementation and straight interpretation. Estimating the trend component for a time series $\{y_t\}_{t=1}^T$ with a tp-spline means the trend function is modelled in dependence of time t, $t = 1, ..., T$. After dividing the variable t into $m-1$ intervals by setting m knots $1 = \kappa_1 < \kappa_2 < ... < \kappa_m = T$, a tp-spline of degree l, henceforth denoted as $tp(l)$, is defined to:

$$y_t = f(t) + \varepsilon_t = \beta_1 + \beta_2 t + ... + \beta_{l+1} t^l + \beta_{l+2}(t-\kappa_2)_+^l + ... + \beta_d (t-\kappa_{m-1})_+^l + \varepsilon_t, \quad (6.7)$$

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

where ε_t is the error term and $d = m + l - 1$. Writing $f(t)$ in matrix notation yields:

$$\boldsymbol{y} = \boldsymbol{Z}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad (6.8)$$

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

Here, $\boldsymbol{\beta} = (\beta_1, ..., \beta_d)'$, $\boldsymbol{\varepsilon} = (\varepsilon_1, ..., \varepsilon_T)'$ and $\boldsymbol{y} = (y_1, ..., y_T)'$. tp-splines can be interpreted as a continuous function of piecewise defined polynomials of degree l. Because of the truncated polynomials the coefficient of the highest polynomial changes at every knot, which enables $f(t)$ to become very flexible. The parameters of the truncated polynomials determine to what degree the coefficient of the highest polynomial can change at the knots. Absolute high values of $\beta_{l+2}, ..., \beta_d$ allow a great flexibility of the trend function. To regulate the smoothness of the trend, the absolute values of these parameters have to be controlled. This is done by estimating the vector of coefficients by minimizing the penalized least squares criterion

$$\min_{\boldsymbol{\beta}} PLS(\lambda) = \sum_{t=1}^{T} [y_t - f(t)]^2 + \lambda \sum_{j=l+2}^{d} \beta_j^2. \quad (6.9)$$

The second part of $PLS(\lambda)$ controls the smoothness of the estimated trend function. It is weighted by the penalization parameter λ. Increasing the value of λ induces a smoother trend, as the estimated coefficients of the truncated polynomials become absolutely smaller.

The solution of (6.9) in matrix notation is given by:

$$\hat{\boldsymbol{y}}(\lambda) = \boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z} + \lambda\boldsymbol{K})^{-1}\boldsymbol{Z}'\boldsymbol{y}, \tag{6.10}$$

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

6.4 Penalized splines and ideal filters

6.4.1 The gain function

In this section it is investigated in how far penalized tp-splines can approximate ideal gain functions. A good lowpass filter at least should have a gain of one for oscillations with zero frequency. This is fulfilled if the filter weights sum up to one. From formula (6.10) a penalized tp-spline can be easily expressed as a linear filter. Defining $\boldsymbol{H}(\lambda) = \boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z} + \lambda\boldsymbol{K})^{-1}\boldsymbol{Z}'$ allows to write the tp-spline as

$$\hat{\boldsymbol{y}}(\lambda) = \boldsymbol{H}(\lambda)\boldsymbol{y}. \tag{6.11}$$

The t^{th} row of $\boldsymbol{H}(\lambda)$ contains the filter weights for estimation $\hat{y}_t$. Let h_{ij} define the j^{th} element in the i^{th} row of $\boldsymbol{H}(\lambda)$, then $\hat{y}_t$ is defined as

$$\hat{y}_t = \sum_{j=1}^{T} h_{tj} y_j. \tag{6.12}$$

For this notation the gain of the tp-spline for an estimation $\hat{y}_t$ and a certain frequency ω can be calculated as:

$$g_t(\omega, \lambda) = \sqrt{\left[\sum_{j=1-t}^{T-t} h_{t,j+t}\cos(\omega j)\right]^2 + \left[\sum_{j=1-t}^{T-t} h_{t,j+t}\sin(\omega j)\right]^2}. \tag{6.13}$$

A gain of one for frequency $\omega = 0$ is given for every estimation $\hat{y}_t$, $t = 1, ..., T$, if the filter weights sum up to one for every row of $\boldsymbol{H}(\lambda)$, i.e.

$$\boldsymbol{H}(\lambda)\boldsymbol{1} = \boldsymbol{1}.$$

This condition holds for any penalized spline in general as long as an intercept is included (Eubank 1983).[13] Thus, penalized tp-splines have a gain of one for $\omega = 0$. The spline can easily be formulated as a highpass filter:

$$\hat{\boldsymbol{\varepsilon}} = \boldsymbol{y} - \boldsymbol{H}(\lambda)\boldsymbol{y} = [\boldsymbol{I} - \boldsymbol{H}(\lambda)]\,\boldsymbol{y}.$$

[13] A specific proof for penalized tp-splines is provided in appendix 6.A.

It follows immediately that $g(0) = 0$ holds for all estimations of the highpass filter, as $[\boldsymbol{I} - \boldsymbol{H}(\lambda)]\mathbf{1} = \mathbf{0}$. Consequently, the basic requirements for low- and highpass filters are fulfilled. Moreover, Figure 6.3 shows the gain functions for splines of different degrees that aim to approximate an ideal gain function with a cut-off frequency of $\omega^{cf} = 0.785$ in the case of a lowpass filter. The gain functions refer to the 51^{th} estimation of a series with 101 observations.

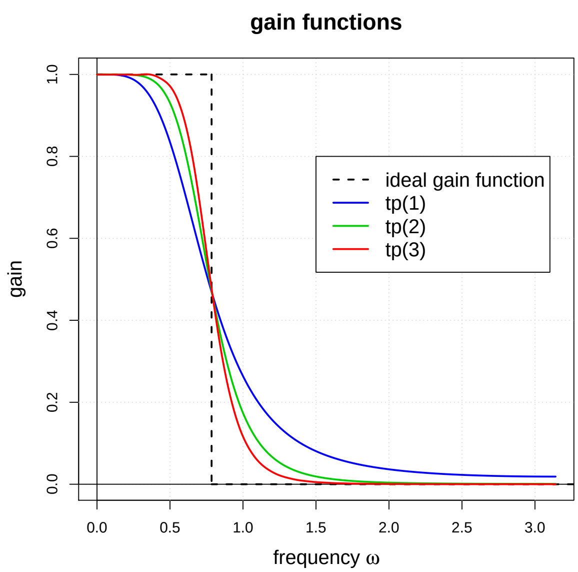

Fig. 6.3: Gain functions for splines of different degrees

All splines approximate the ideal gain function. The approximation becomes more accurate for higher degrees of the splines. This shows the connection of penalized tp-splines and rational square wave filters (Pollock 2000, 2003). For knots at every point in time Proietti (2007) describes a $tp(l)$ as a trend plus noise model, where the trend is a $l+1$-fold integrated random walk, which is a special case of the Wiener-Kolmogorov filter (Whittle 1983, Bell 1984, also Harvey 1989, Kaiser/Maravall 2001). Since rational square wave filters are closely related to the framework of the Wiener-Kolmogorov filter penalized tp-splines can be linked to the class of square wave filters. In detail the degree of the spline regulates the bandwidth for the transition of the gain function from one to zero, and the penalization parameters controls the approximate cut-off frequency.

One advantage of penalized splines to the BK-filter is that they yield a trend estimation for the whole time series, while the BK-filter cannot estimate trend or cycle for the first and last n periods. However, one has to be aware that the estimation of splines heavily loses its reliability to the margins of the series. This is caused by the problem that the filter weight structure cannot be symmetric for the first and last estimations. This asymmetry induces far too volatile estimations for periods at the margins. To this point Figure 6.4 shows the gain functions of the $tp(3)$ of the example above for different periods:

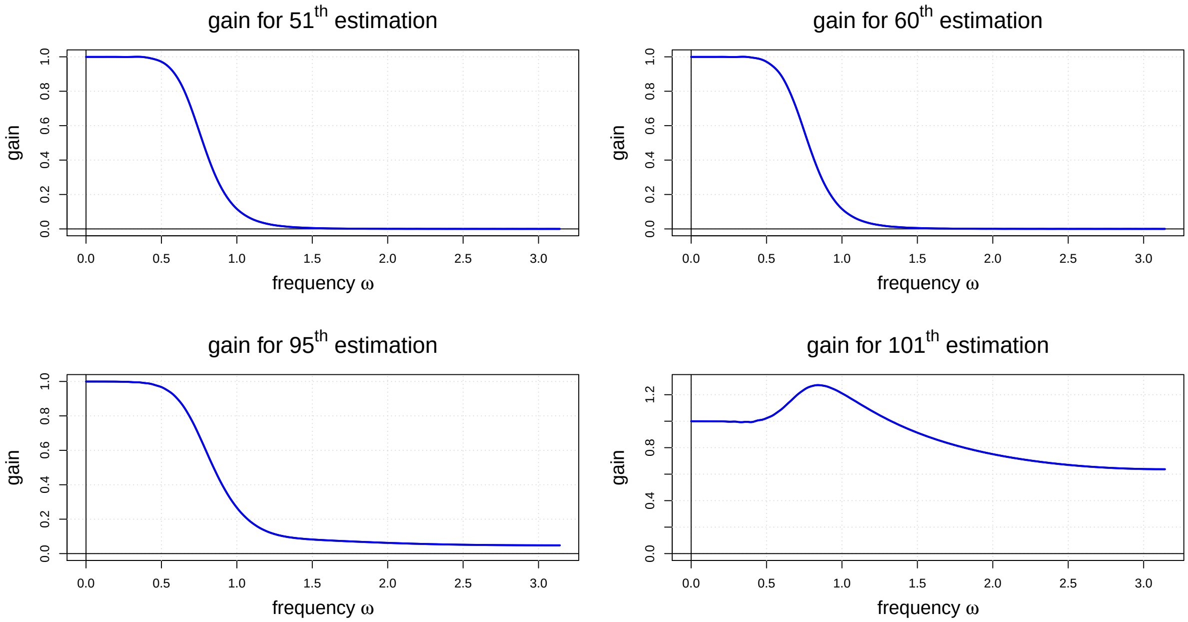

Fig. 6.4: Gain functions for different periods

The gain function of the 51^{th} and 60^{th} estimation are good approximations of an ideal gain function. However, for estimations closer to the margins the spline is not able to suppress higher frequencies any more, which is shown by the gain functions for the 95^{th} and 101^{th} estimation. The insufficient suppression of high frequencies heavily affects the reliability at the margins, which is called excess variability. For a detailed discussion of the excess variability of splines to the margins see Blöchl (2014b). The number of elements that show an increased variability at the margins depends on the value of the penalization and the degree of the spline. A higher penalization as well as a higher degree of the spline ceteris paribus increases the number of estimations that are affected by the excess variability.

6.4.2 Phase shifts and the elimination of unit roots

It is reasonable to examine the induction of phase shifts and the elimination of unit roots together, as both are based on similar conditions. Linear filters do not induce phase shifts when they are symmetric and they are able to turn time series that are integrated of order one or two into stationary series, if furthermore their filter weights sum up to zero (Osborn 1995, Baxter/King 1999).

It is already shown in section 6.4.1 that the filter weights of a penalized spline sum up to zero, when it is used as a highpass filter. The condition of symmetry would be fulfilled if the entries of the hat matrix $\boldsymbol{H}(\lambda)$ are symmetric around the weights h_{ii}. Of course, for a finite time series this condition can never hold for all estimations, especially at the margins. As an example Figure 6.5 displays the filter weights for a penalized tp-spline of degree two, 25 knots and λ=100 that is applied to a series containing 101 observations:

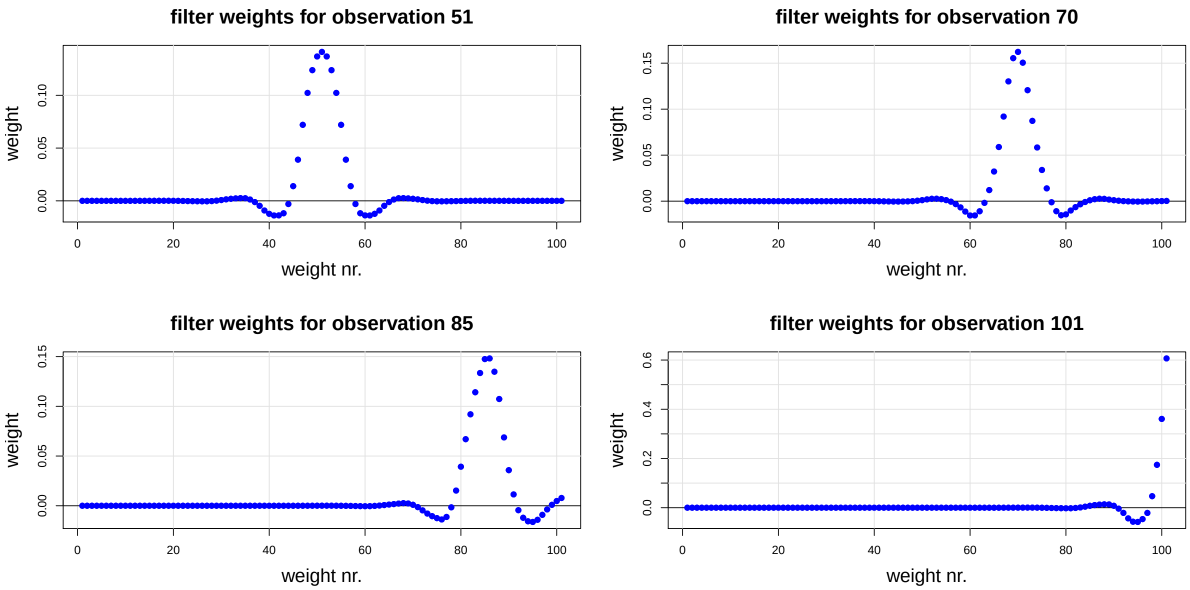

Fig. 6.5: Filter weights of a tp-spline for different estimations

Although the filter weights refer to the tp-spline as a lowpass filter, they are also representative for the case of a highpass filter, as the weights of the corresponding highpass filter just arise from $\boldsymbol{I} - \boldsymbol{H}(\lambda)$. Observation 51 is the middle of the data. Obviously, the corresponding weights are symmetric. With an increasing distance to the middle the weights become more asymmetric, which is a drawback to completely symmetric filters like the Baxter-King filter.

To this point Dagum/Luati (2004) note that $\boldsymbol{H}(\lambda)$ is centroysmmetric, which means $h_{ij} = h_{T+1-i,T+1-j}$. This implies that for an odd number of observations the filter weights for the estimation in the middle of the data are symmetric. Furthermore, linear smoothers generate filter weights that symmetrically decrease to zero with a growing distance to the middle weight (Dagum/Luati 2004). This means $h_{ij} \to 0$ as $|i - j| \to \infty$. Consequently, for an infinite series the filter weights would be symmetric for every estimation. As a result penalized splines asymptotically do not induce phase shifts and are able to eliminate unit roots, when $T \to \infty$.

Moreover, the decrease of the filter weights for growing $|i - j|$ depends on the selected penalization. For low values of λ the filter weights converge very fast to zero. This results in a rather wiggly trend, as only the nearest observations are influential for the estimation. A high value of λ induces a slow decline of the weights. Thus, for a given estimation also observations far away play a significant role. This leads to a rather smooth trend estimation. Figure 6.6 shows this fact for a tp-spline with $l = 1$ and T equidistant knots, that is applied to a series with 101 observations. The weight structures refer to the middle of the series.

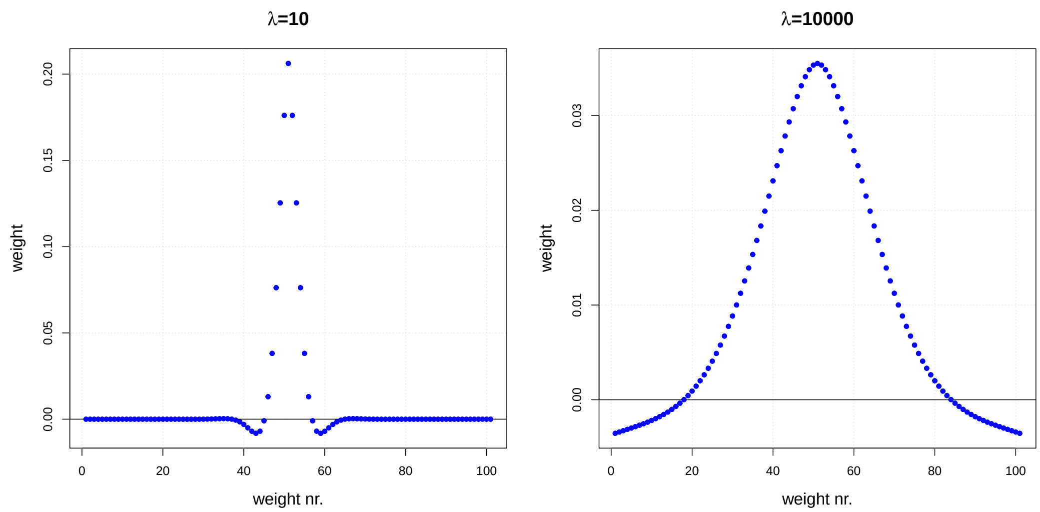

Fig. 6.6: Filter weights for a $tp(1)$ with low and high penalization

The left plot shows the filter weights for the case of a low penalization. Here, only the nearest twelve observations have a notable influence on the estimation. This is different for the case of $\lambda = 10000$, where almost all observations have a weight that is different from zero and where the convergence to zero would require much more than the length of the time series. As a consequence, the induction of phase shifts and the approximative ability of splines to eliminate unit roots is rather given in cases of a low penalization. The effect of the number of observations and the penalization on the ability of splines to eliminate unit roots can easily be demonstrated. To this point a penalized tp-spline with varying values of λ is applied to simulated time series with different numbers of observations. The series are simulated as the sum of a r-fold integrated random walk as the trend component and a stationary AR(1)-process as the cyclical component, i.e. $y_t = \mu_t + c_t$ where

$$(1 - L)^r \mu_t = \eta_t, \tag{6.14}$$

$$c_t = \varphi c_{t-1} + u_t. \tag{6.15}$$

η_t and u_t are independent white noise variables with variances σ_η^2 and σ_u^2. For the simulation the parameters are set to $r = 1, \varphi = 0.5$ and $\sigma_\eta^2 = \sigma_u^2 = 1$. The simulated time series contain 50, 100 and 250 observations, while for each case 500 series are generated. The trend component of the series is estimated by a tp-spline of order one and knots at every point in time t, $t = 1, 2, ..., T$. Note that this setting is equal to the Hodrick-Prescott filter (Paige 2010). Finally, a Dickey-Fuller test (Dickey/Fuller 1979) is used to test, if the resulting residual components are stationary, where a constant term and a linear trend are included into the test equation. Table 6.1 shows the resulting average p-values of the tests:

length	$\boldsymbol{\lambda = 10}$	$\boldsymbol{\lambda = 10^2}$	$\boldsymbol{\lambda = 10^3}$	$\boldsymbol{\lambda = 10^4}$	$\boldsymbol{\lambda = 10^5}$	$\boldsymbol{\lambda = 10^6}$
$\boldsymbol{T = 50}$	0.011	0.035	0.159	0.334	0.438	0.457
$\boldsymbol{T = 100}$	$\leq$0.01	$\leq$0.01	0.018	0.101	0.278	0.429
$\boldsymbol{T = 250}$	$\leq$0.01	$\leq$0.01	$\leq$0.01	0.010	0.034	0.178

Tab. 6.1: Average p-values of the Dickey-Fuller test

Low p-values indicate that on average the unit root has been eliminated. The p-values increase for higher values of λ, while they decrease, when the number of observations rises. Furthermore, note that the results would be similar for splines of higher degrees, when the penalization is set to an equivalent level. The results also do not basically change, if a higher signal to noise ration is selected, or if the value for φ is changed.
To show the practical relevance of this issue a probably more realistic example is constructed. Consider data for the quarterly German GDP that are at the moment consistently available form the first quarter 1991 to the third quarter 2013, which makes 91 observations. To this regard 1000 time series with 91 observations are simulated by the same process as before, but this time the parameters are set to $r = 2, \varphi = 0.9, \sigma_\eta = 0.25$ and $\sigma_u = 10$ so that the inverse signal to noise ratio is $\frac{\sigma_u^2}{\sigma_\eta^2}$ is 1600. This might be not unrealistic for quarterly data (Hodrick/Prescott 1997). Just like before the trend is estimated by a $tp(1)$ with knots at every point in time, which is equal to the Hodrick-Prescott filter (Paige 2010). Defining the trend component by fluctuations with periodicities above six, eight or ten years implies values for λ of 453, 1410 and 3417, if the deviation between real and ideal gain function is minimized (Tödter 2002). Again the Dickey-Fuller test is applied to each residual component. Figure 6.7 shows the empirical cumulative density functions of the p-values for each value of λ:

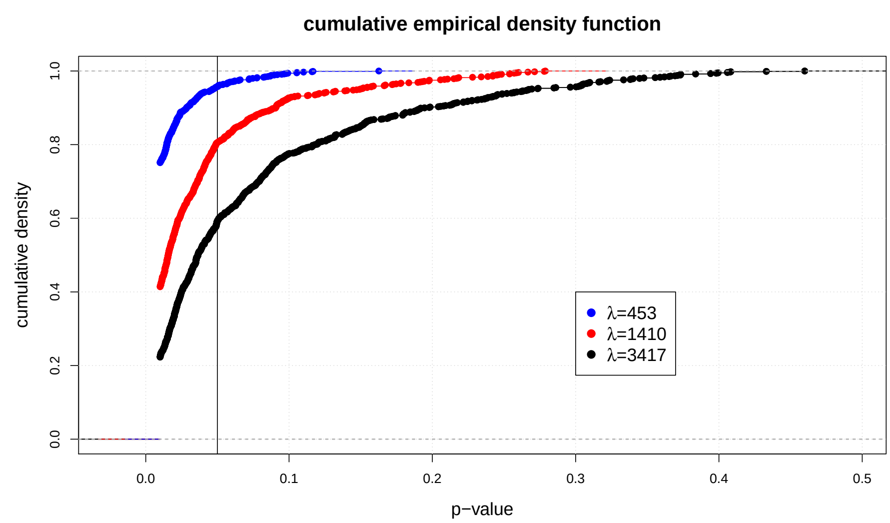

Fig. 6.7: p-values of the Dickey-Fuller test for different values of λ

The empirical cumulative density function is shifted to the right as λ increases. While for $\lambda = 453$ about 95 percent of the p-values are smaller than 0.05, this is only the case for about 60 percent of the p-values for $\lambda = 3417$. The average p-value for $\lambda = 3417$ is 0.073, while it is 0.037 for $\lambda = 1410$ and 0.016 for $\lambda = 453$. This demonstrates that the ability of splines to eliminate unit roots decreases for higher values of λ.

6.5 Comparison to the Baxter-King filter

6.5.1 Lowpass filters

In this section the ability of penalized tp-splines to extract certain frequency bands is compared to the one of the BK-filter. To this point the penalization of splines is selected such that the gain function of the spline shows a minimal squared deviation to an ideal gain function with a certain cut-off frequency ω^{cf}, which is identical to the criterion of the BK-filter. Let $g^*(\omega)$ define the ideal gain for frequency ω, and $g_t(\omega, \lambda)$ the gain of the tp-spline for estimation $\hat{y}_t$, frequency ω and a certain value of λ, then the penalization is selected such that it minimizes:

$$l(t, \lambda) = \int_0^{\pi} \left[g^*(\omega) - g_t(\omega, \lambda)\right]^2 d\omega. \tag{6.16}$$

$l(t, \lambda)$ is the squared deviation between real and ideal gain function in the interval $[0, \pi]$ and henceforth denoted as *loss*. The calculation can easily be simplified by a sufficient large number of discrete frequencies.

$l(t, \lambda)$ depends on the period of the estimation, for which the *loss* is minimized. As seen in Figure 6.4, the gain functions are not equal for estimations for different periods. However, as it is also shown later, the gain functions of the splines are almost equal for the majority of estimations except of the margins. Thus, the penalization is selected such that it minimizes $l(t, \lambda)$ for the estimation in the middle of the series, i.e. for $t = T/2$, or its next integer if T is odd. Most other estimations require almost the same degree of penalization in order to minimize the *loss* (Blöchl 2014b). Furthermore, the number of knots has to be selected. As a low number of knots might be restrictive and hindering to achieve a precise gain function, it is always set as high as possible.

The BK-filter is opposed to penalized tp-splines for the case of a lowpass filter. To this regard two different cut-off frequencies are considered that are set to $\omega_h^{cf} = 0.785$ and $\omega_l^{cf} = 0.196$. Both imply a cut-off periodicity of eight years for the cases of yearly data (ω_h^{cf}) and quarterly data (ω_l^{cf}). As the degree of the tp-spline affects its properties in the frequency domain (Blöchl 2014b), the BK-filter is opposed to a $tp(1)$ as well as a $tp(3)$. For quarterly data Baxter/King (1999) suggest to set $n = 12$, as this would yield a sufficient tradeoff between a good adaptation to an ideal gain function and a low number of dropped observations at the margins. Consequently, n is set to 12 for the case of ω_l^{cf} so that for quarterly data the first and last three years are lost. For the case of ω_h^{cf} a value of $n = 3$ is selected, as this implies the loss of three years at the ends for yearly data. Table 6.2 shows the results for the BK-filter and splines for the case of a time series with 130 observations. The *loss* refers to the estimation in the middle (65$^{\text{th}}$ estimation). Furthermore, Table 6.2 shows the minimal value of n that is necessary to achieve better results than penalized tp-splines.

	$\boldsymbol{\omega_l^{cf}}$		$\boldsymbol{\omega_h^{cf}}$	
BK-filter	n=12	0.0228	n=3	0.0890
	n=36	0.0095	n=9	0.0357
spline	$tp(1)$	0.0191	$tp(1)$	0.0814
	$tp(3)$	0.0095	$tp(3)$	0.0379

Tab. 6.2: *loss* for the BK-filter and tp-splines

For both $n = 3$ and $n = 12$ the BK-filter exhibits a worse adaptation to the ideal gain function than penalized tp-splines. In the case of ω_l^{cf} the *loss* of the BK-filter is about 20 percent higher than the one of the $tp(1)$ and even 2.4 times higher than the *loss* of the $tp(3)$. For ω_h^{cf} the deviation between the *loss* of the BK-filter and the tp-splines is about the same degree. Moreover, to achieve comparable results to the $tp(3)$, in both cases at least a three times higher value for n has to be selected. This would imply that nine years at each margin are lost. To show the differences between the BK-filter and the tp-splines, Figure 6.8 plots the gain functions of the $tp(3)$ and the BK-filter for both cut-off frequencies. The gain function of the $tp(3)$ refers to the 65$^{\text{th}}$ estimation.

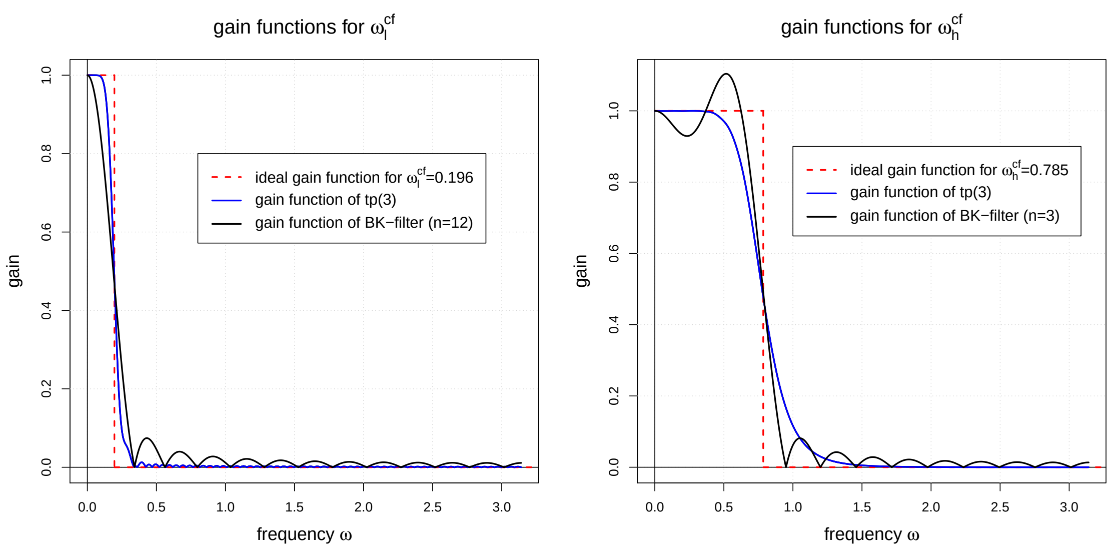

Fig. 6.8: Gain functions of the BK-filter and the $tp(3)$

In both cases the gain function of the $tp(3)$ clearly exhibits preferable characteristics. The BK-filter is not able to completely suppress high frequencies, which is already explored by Goldrian (2005) for low cut-off frequencies. For ω_h^{cf} a certain bandwidth of frequencies is even amplified. However, so far the analysis is incomplete as the gain functions of the splines might not be equal for estimations for different periods. The gain functions of the splines change especially to the margins so that the *loss* for the estimations of all periods, henceforth denoted as the *loss function*, should be compared. Figures 6.9 and 6.10 display the *loss functions* for ω_h^{cf} for the $tp(1)$ and the $tp(3)$ and oppose them to the BK-filter.

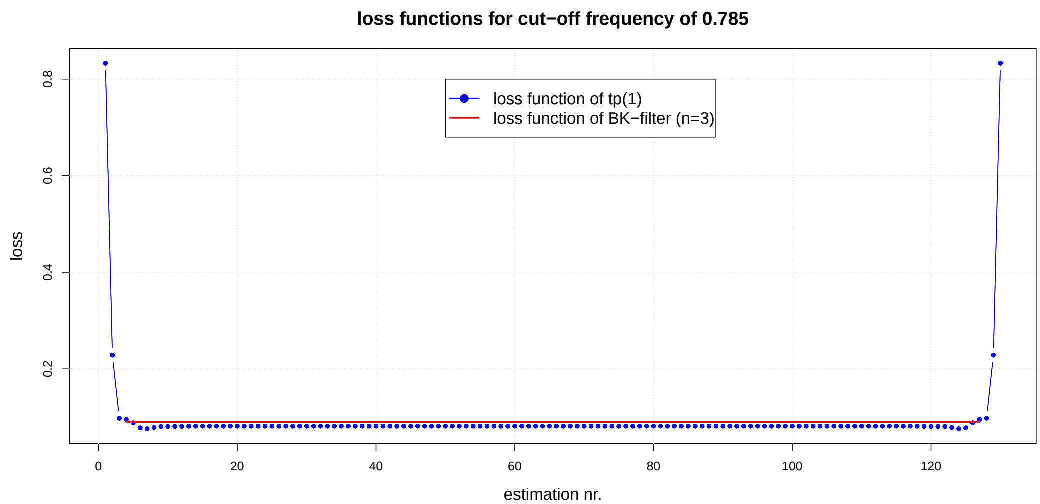

Fig. 6.9: *loss functions* of the BK-filter and the $tp(1)$ for ω_h^{cf}

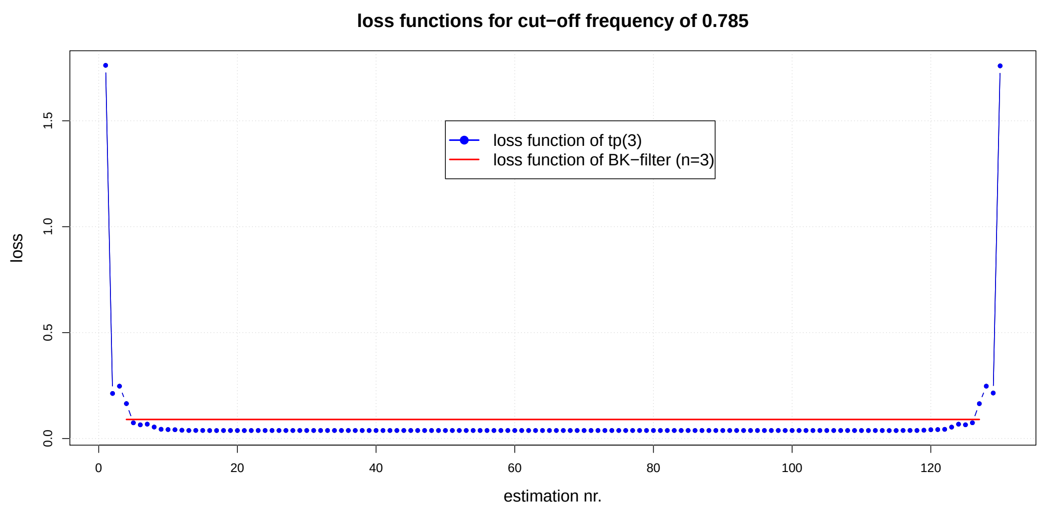

Fig. 6.10: *loss functions* of the BK-filter and the $tp(3)$ for ω_h^{cf}

Figures 6.9 and 6.10 show that the *loss* of the splines is almost equal for the majority of estimations and increases only for few estimations at the margins of the series. The *loss* of the BK-filter is equal for the estimations of every period, since the filter weights are the same for all periods. This, however, is paid by the loss of n estimations at the margins. The *loss* of the $tp(1)$ as well as the $tp(3)$ is clearly lower than the one of the BK-filter for almost all estimations. Just one estimation at each margin of the series exhibits a *loss* above the one of the BK-filter. Figures 6.11 and 6.12 display the *loss functions* for the case of ω_l^{cf}:

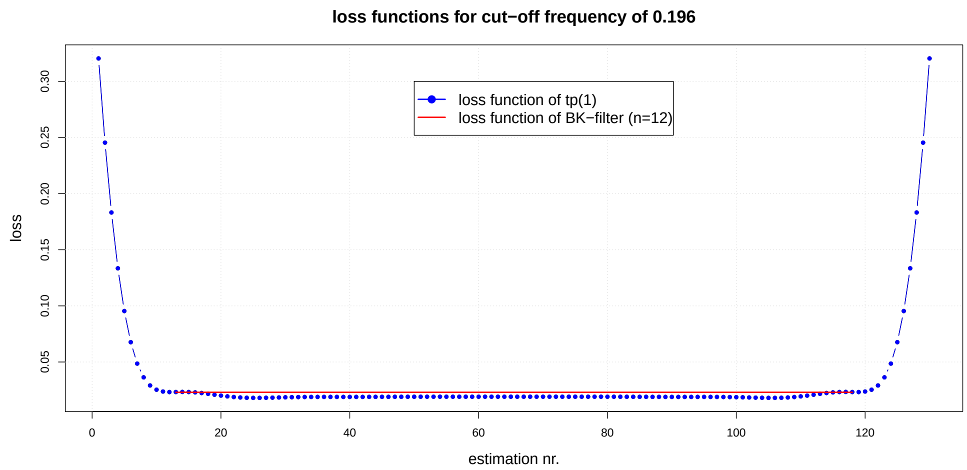

Fig. 6.11: $loss\ functions$ of the BK-filter and the $tp(1)$ for ω_l^{cf}

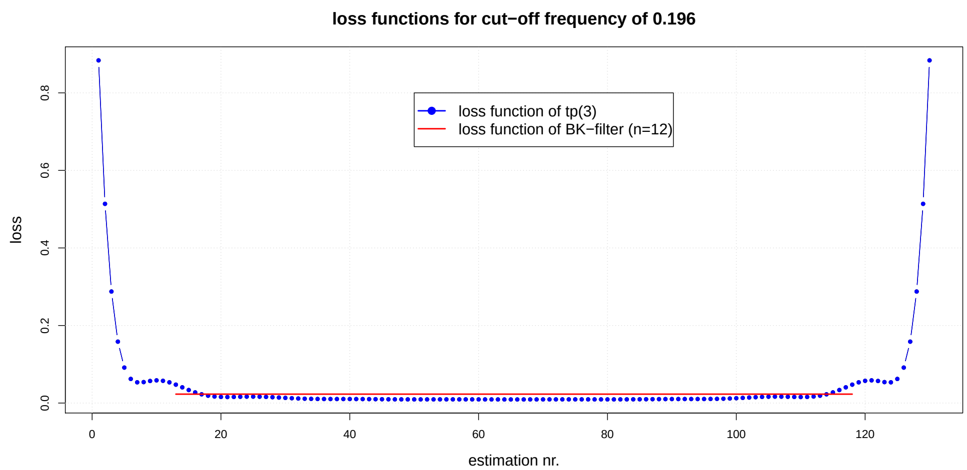

Fig. 6.12: $loss\ functions$ of the BK-filter and the $tp(3)$ for ω_l^{cf}

The situation of ω_l^{cf} is similar to the case of ω_h^{cf}. For the $tp(1)$ the estimations for almost all periods exhibit a lower $loss$ than the BK-filter. Only for few estimations at the margins there is no notable difference between the filters. The $tp(3)$ suffers from a far higher variability at the ends than the $tp(1)$. There are four estimations at each margin that notably show a higher $loss$ than the BK-filter. However, for the majority of estimations the $tp(3)$ clearly yields better results.

The analysis shows that, compared to the BK-filter, penalized tp-splines have advantageous features as lowpass filters. The accuracy of the gain function can be controlled by the degree of the spline, while there is a tradeoff between a good approximation of an ideal gain function for the majority of estimations and a low increase of the variability at the margins. Although tp-splines suffer from an increasing variability at the margins they yield for (almost) all estimations better results than the BK-filter for the standard selection of n=12 for quarterly, and n=3 for yearly data. To achieve comparable results to splines a much higher loss of estimations at the margins would have to be accepted for the BK-filter.

Moreover, the excess variability of penalized splines at the margins can be reduced notably by a time-varying penalization (Blöchl 2014b).

6.5.2 Bandpass filters

The previous comparison demonstrates that penalized tp-splines appear to be the preferable choice as lowpass filters. This is also true for the case of highpass filters, as highpass filters directly arise from the lowpass filter. However, it is worth comparing penalized tp-splines and the BK-filter as bandpass filters. The bandpass filter is constructed from two lowpass filters with a high and a low cut-off frequency ω_h^{cf} and ω_l^{cf}. Let $\boldsymbol{H_h}$ and $\boldsymbol{H_l}$ denote the hat matrices that contain the weights of the lowpass filters, then the weight matrix of the bandpass filter is defined to $\boldsymbol{H_h} - \boldsymbol{H_l}$ (c.f. Baxter/King 1999 p.578). For this comparison the cut-off frequencies are set to $\omega_l^{cf} = 0.196$ and $\omega_h^{cf} = 1.048$. This implies for quarterly data that the cycle is defined by periodicities between six and 32 quarters. Table 6.3 shows the basic results for a series with 130 observations. For the BK-filter n was set to twelve.

filter	***loss*** **(65^{th} estimation)**
$tp(1)$	0.136
$tp(3)$	0.060
BK-filter (n=12)	0.046

Tab. 6.3: *loss* for the BK-filter and tp-splines

The values in Table 6.3 refer to the estimation in the middle of the series. In the case of a bandpass filter the BK-filter exhibits a lower *loss* than the $tp(1)$ and the $tp(3)$. To this regard Figure 6.13 displays the corresponding gain functions:

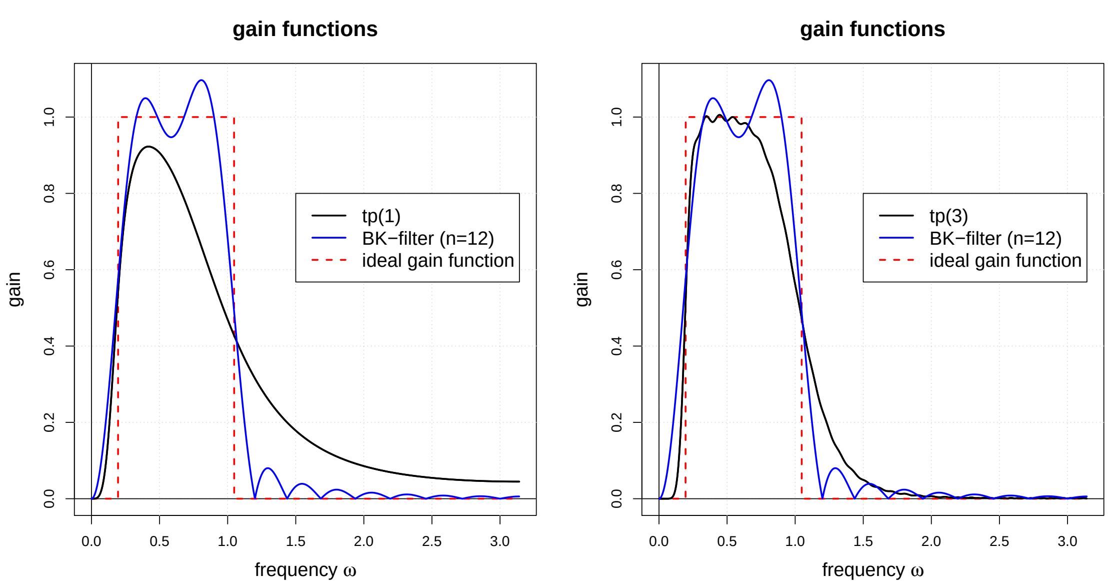

Fig. 6.13: Gain functions of the BK-filter and penalized splines

Figure 6.13 shows the reasons for the better results of the BK-filter. While both splines exhibit a slightly better adaptation to the ideal gain function at the low cut-off frequency, the BK-filter yields a much better adaptation at the high cut-off frequency. To get a complete overview about the performances of the filters, Figures 6.14 and 6.15 display the *loss functions*:

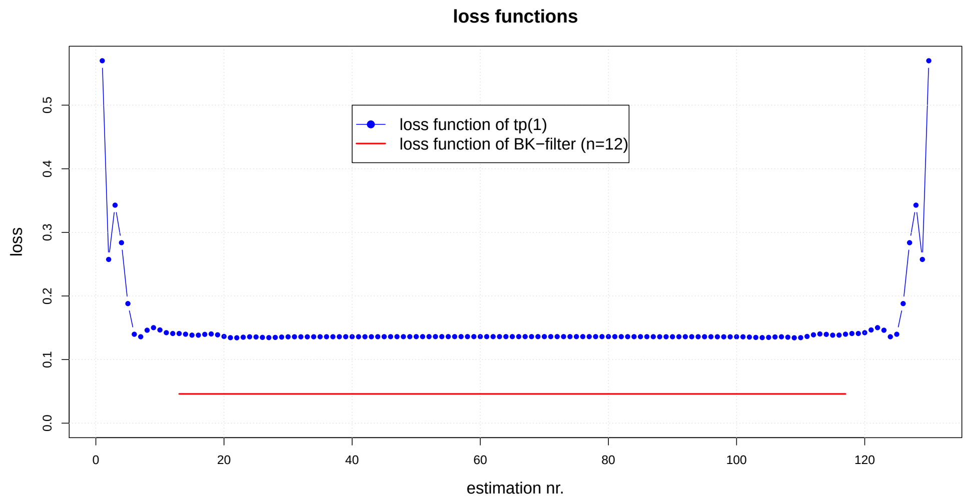

Fig. 6.14: *loss functions* of the $tp(1)$ and the BK-filter

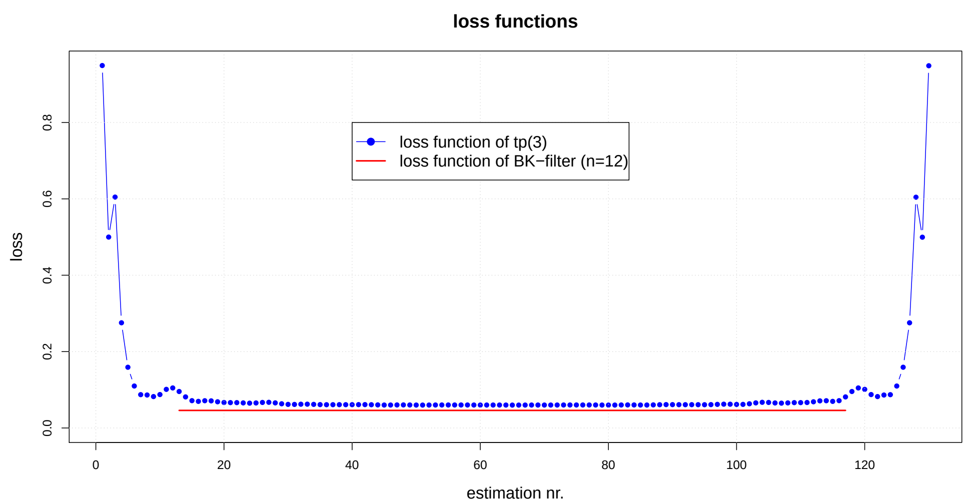

Fig. 6.15: *loss functions* of the $tp(3)$ and the BK-filter

Figures 6.14 and 6.15 show that the BK-filter exhibits a lower *loss* for all estimations. In this setting it cannot yield estimations for the first and last twelve periods. However, for the $tp(3)$ about an equivalent number of estimations is strongly affected by the excess variability. The $tp(1)$ exhibits only five estimations that are strongly affected by the excess variability at each margin. Nevertheless, the BK-filter would yield better results than the $tp(1)$, even if n would be set to five.

6.6 Conclusion

The characterization of trend and cycle by their spectral properties is a widespread and helpful conception for a reasonable separation of these components in economics. In this regard the BK-filter is one of the most popular tools to derive trend and cycle by frequency domain aspects. It arises from the postulations of Baxter/King for an ideal time series filter. The aim of this chapter is to shed light on in how far penalized splines meet these requirements. It summarizes theory about time series filters in the frequency domain, penalized splines with a truncated polynomial basis as well as the BK-filter, and analyzes the properties of penalized splines with regard to the extraction of frequency bands, the elimination of unit roots and the induction of phase shifts. Furthermore, the ability of penalized splines to extract frequency bands is compared to the one of the BK-filter.

This chapter shows that the filter weights of penalized tp-splines sum up to one for every estimation so that one basic condition of Baxter/King is fulfilled. It ensures that the gain of a penalized tp-spline is one for a frequency of zero. Moreover, this chapter demonstrates that the gain function of a penalized spline strongly depends on its degree. A higher degree allows a more precise extraction of frequency bands as the transition of the gain function from one to zero becomes faster. This behavior points out the link of penalized tp-splines and rational square wave filters. The degree of the spline controls the transition of the gain function, while the penalization parameter determines the approximate cut-off frequency.

A basic advantage of the BK-filter is that it renders symmetric filter weights for every estimation. This filter weight structure does not induce phase shifts and yields an identical gain function for every estimation. However, the condition of completely symmetric filter weights does not permit to estimate the trend component for the first and last n periods, which is a major drawback of this filter. Penalized tp-splines as asymmetric filters generally are able to derive estimations for all periods. This property is paid by an change in the filter weights structure that becomes increasingly asymmetric to the margins of the series.

The change of the filter weight structure induces phase shifts and affects the gain functions of penalized tp-splines. The gain functions are good approximations of an ideal gain function for most estimations. However, penalized tp-splines increasingly lose the ability to suppress high frequencies to the margins of the series, which results in a too volatile trend estimation at the margins. The filter weights of a penalized tp-spline are centrosymmetric. In detail, they decrease symmetrically to zero from the weight h_{ii} of the hat matrix $\boldsymbol{H}(\lambda)$, which means $h_{ij} \to 0$ as $|i-j| \to \infty$. The speed of the convergence depends on the penalization parameter λ, which becomes slower for high values of λ. Thus, the ability of penalized splines to eliminate unit roots or quadratic time trends as well as to avoid phase shifts depends on the value of the penalization and the number of observations, which is demonstrated in section 6.4.2.

In the last section the ability of both filters to extract frequency bands is compared. This is done for the case of lowpass filters for the fictive situations of quarterly and yearly data at first. To this point a cut-off periodicity of eight years is assumed. It is shown that penalized tp-splines clearly yield more precise gain functions than the BK-filter, when the standard setting of $n = 12$ for quarterly and $n = 3$ for yearly data is considered. To achieve comparable results to penalized splines at least three times higher values of n have to be selected. This comparison also accounts for the excess variability of penalized tp-splines to the margins. It is shown that penalized splines yield better results for the estimation of all periods except of the margins. Nevertheless, the number of estimations at the margins that are lost by the BK-filter is above the number of estimations that are affected by the excess variability.

Afterwards the filters are compared for the case of bandpass filters, where the cut-off periodicities are set to six and 32 quarters. Here, the situation turns out to be different. It is shown that the BK-filter with $n = 12$ yields better results than penalized tp-splines. Moreover, the number of estimations that are strongly affected by the excess variability, when the cycle is estimated by penalized tp-splines, is about as high as the number of estimations lost at each margin, when the BK-filter is applied. Consequently, for quarterly data penalized splines appear to be superior as lowpass filters, but the BK-filter seems to have advantageous properties, when it is employed as a bandpass filter.

Appendix

6.A Proof that filter weights of a penalized tp-spline sum up to one

The hat matrix of a penalized tp-spline is defined as:

$$\mathbf{H} = \boldsymbol{Z}(\boldsymbol{Z'}\boldsymbol{Z} + \lambda\boldsymbol{K})^{-1}\boldsymbol{Z'}.$$

As an intercept is included, the design matrix $\boldsymbol{Z} \in \mathbb{R}^{T\times d}$ has only ones as the first column. $\boldsymbol{Z}$ is divided into

$$\boldsymbol{Z} = (\boldsymbol{Z_1}, \boldsymbol{Z_2}),$$

where $\boldsymbol{Z_1} = \mathbf{1} \in \mathbb{R}^{T\times 1}$ and $\boldsymbol{Z_2} \in \mathbb{R}^{T\times(d-1)}$. Furthermore, $\lambda\boldsymbol{K}$ can be portioned into:

$$\boldsymbol{K} = \begin{pmatrix} \boldsymbol{K_1} & \boldsymbol{K_2} \\ \boldsymbol{K_3} & \boldsymbol{K_4} \end{pmatrix},$$

with the portioned matrices $\boldsymbol{K_1} = 0 \in \mathbb{R}^{1\times 1}$, $\boldsymbol{K_2} = \mathbf{0} \in \mathbb{R}^{1\times(d-1)}$, $\boldsymbol{K_3} = \mathbf{0} \in \mathbb{R}^{(T-1)\times 1}$ and $\boldsymbol{K_4} \in \mathbb{R}^{(T-1)\times(d-1)}$. This yields:

$$(\boldsymbol{Z'}\boldsymbol{Z}+\lambda\boldsymbol{K})^{-1} = \left(\begin{pmatrix} \boldsymbol{Z_1'} \\ \boldsymbol{Z_2'} \end{pmatrix}(\boldsymbol{Z_1}\ \boldsymbol{Z_2}) + \lambda\boldsymbol{K}\right)^{-1} = \left(\begin{pmatrix} \boldsymbol{Z_1'Z_1} & \boldsymbol{Z_1'Z_2} \\ \boldsymbol{Z_2'Z_1} & \boldsymbol{Z_2'Z_2} \end{pmatrix} + \lambda\begin{pmatrix} \boldsymbol{K_1} & \boldsymbol{K_2} \\ \boldsymbol{K_3} & \boldsymbol{K_4} \end{pmatrix}\right)^{-1} =$$

$$\left(\begin{pmatrix} \boldsymbol{Z_1'Z_1} & \boldsymbol{Z_1'Z_2} \\ \boldsymbol{Z_2'Z_1} & \boldsymbol{Z_2'Z_2} \end{pmatrix} + \lambda\begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{K_4} \end{pmatrix}\right)^{-1} = \left(\begin{pmatrix} \boldsymbol{Z_1'Z_1} & \boldsymbol{Z_1'Z_2} \\ \boldsymbol{Z_2'Z_1} & \boldsymbol{Z_2'Z_2} + \lambda\boldsymbol{K_4} \end{pmatrix}\right)^{-1} = \begin{pmatrix} \boldsymbol{J_1} & \boldsymbol{J_2} \\ \boldsymbol{J_3} & \boldsymbol{J_4} \end{pmatrix}$$

Partial inversion general (e.g. Fahrmeir et al. 2009 p.451):

$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{E} & \boldsymbol{F} \\ \boldsymbol{G} & \boldsymbol{L} \end{pmatrix}$$

$$\boldsymbol{A^{-1}} = \begin{pmatrix} \boldsymbol{E^{-1}}(\boldsymbol{I} + \boldsymbol{F}\boldsymbol{D^{-1}}\boldsymbol{G}\boldsymbol{E^{-1}}) & -\boldsymbol{E^{-1}}\boldsymbol{F}\boldsymbol{D^{-1}} \\ -\boldsymbol{D^{-1}}\boldsymbol{G}\boldsymbol{E^{-1}} & \boldsymbol{D^{-1}} \end{pmatrix},$$

where $\boldsymbol{D} = \boldsymbol{L} - \boldsymbol{G}\boldsymbol{E^{-1}}\boldsymbol{F}$. This implies that:

$$\boldsymbol{D} = \boldsymbol{Z_2'Z_2} + \lambda\boldsymbol{K_4} - \boldsymbol{Z_2'Z_1}(\boldsymbol{Z_1'Z_1})^{-1}\boldsymbol{Z_1'Z_2} = \lambda\boldsymbol{K_4} + \boldsymbol{Z_2'}(\boldsymbol{I} - \boldsymbol{H_1})\boldsymbol{Z_2},$$

$$\text{with } \boldsymbol{H_1} = \boldsymbol{Z_1}(\boldsymbol{Z_1'Z_1})^{-1}\boldsymbol{Z_1'}.$$

Consequently $\boldsymbol{J_1}, \boldsymbol{J_2}, \boldsymbol{J_3}$ and $\boldsymbol{J_4}$ can be written to:

$$\boldsymbol{J_1} = (\boldsymbol{Z_1'Z_1})^{-1}(\boldsymbol{I} + \boldsymbol{Z_1'Z_2}(\lambda\boldsymbol{K_4} + \boldsymbol{Z_2'}(\boldsymbol{I} - \boldsymbol{H_1})\boldsymbol{Z_2})^{-1}\boldsymbol{Z_2'Z_1}(\boldsymbol{Z_1'Z_1})^{-1}),$$

$$\boldsymbol{J_2} = -(\boldsymbol{Z_1'Z_1})^{-1}\boldsymbol{Z_1'Z_2}(\lambda\boldsymbol{K_4} + \boldsymbol{Z_2'}(\boldsymbol{I} - \boldsymbol{H_1})\boldsymbol{Z_2})^{-1},$$

$$\boldsymbol{J_3} = -(\lambda\boldsymbol{K_4} + \boldsymbol{Z_2'}(\boldsymbol{I} - \boldsymbol{H_1})\boldsymbol{Z_2})^{-1}\boldsymbol{Z_2'Z_1}(\boldsymbol{Z_1'Z_1})^{-1},$$

$$\boldsymbol{J_4} = (\lambda\boldsymbol{K_4} + \boldsymbol{Z_2'}(\boldsymbol{I} - \boldsymbol{H_1})\boldsymbol{Z_2})^{-1}.$$

Calculating the hat matrix of the penalized tp-spline yields:

$$\boldsymbol{H} = \boldsymbol{Z}(\boldsymbol{Z'Z})^{-1}\boldsymbol{Z'} = (\boldsymbol{Z_1Z_2})\begin{pmatrix}\boldsymbol{J_1} & \boldsymbol{J_2}\\ \boldsymbol{J_3} & \boldsymbol{J_4}\end{pmatrix}\begin{pmatrix}\boldsymbol{Z_1'}\\ \boldsymbol{Z_2'}\end{pmatrix} =$$

$$\begin{aligned}
&\boldsymbol{Z_1}(\boldsymbol{Z_1'Z_1})^{-1}\boldsymbol{Z_1'} + \boldsymbol{Z_1}(\boldsymbol{Z_1'Z_1})^{-1}\boldsymbol{Z_1'Z_2}(\lambda\boldsymbol{K_4} + \boldsymbol{Z_2'}(\boldsymbol{I} - \boldsymbol{H_1})\boldsymbol{Z_2})^{-1}\boldsymbol{Z_2'Z_1}(\boldsymbol{Z_1'Z_1})^{-1}\boldsymbol{Z_1'} -\\
&-\boldsymbol{Z_1}(\boldsymbol{Z_1'Z_1})^{-1}\boldsymbol{Z_1'Z_2}(\lambda\boldsymbol{K_4} + \boldsymbol{Z_2'}(\boldsymbol{I} - \boldsymbol{H_1})\boldsymbol{Z_2})^{-1}\boldsymbol{Z_2'} -\\
&-\boldsymbol{Z_2}(\lambda\boldsymbol{K_4} + \boldsymbol{Z_2'}(\boldsymbol{I} - \boldsymbol{H_1})\boldsymbol{Z_2})^{-1}\boldsymbol{Z_2'Z_1}(\boldsymbol{Z_1'Z_1})^{-1}\boldsymbol{Z_1'} +\\
&+\boldsymbol{Z_2}(\lambda\boldsymbol{K_4} + \boldsymbol{Z_2'}(\boldsymbol{I} - \boldsymbol{H_1})\boldsymbol{Z_2})^{-1}\boldsymbol{Z_2'} =\\
&= \boldsymbol{H_1} + \boldsymbol{H_1Z_2}(\lambda\boldsymbol{K_4} + \boldsymbol{Z_2'}(\boldsymbol{I} - \boldsymbol{H_1})\boldsymbol{Z_2})^{-1}\boldsymbol{Z_2'H_1} -\\
&-\boldsymbol{H_1Z_2}(\lambda\boldsymbol{K_4} + \boldsymbol{Z_2'}(\boldsymbol{I} - \boldsymbol{H_1})\boldsymbol{Z_2})^{-1}\boldsymbol{Z_2'} -\\
&-\boldsymbol{Z_2}(\lambda\boldsymbol{K_4} + \boldsymbol{Z_2'}(\boldsymbol{I} - \boldsymbol{H_1})\boldsymbol{Z_2})^{-1}\boldsymbol{Z_2'H_1} +\\
&+\boldsymbol{Z_2}(\lambda\boldsymbol{K_4} + \boldsymbol{Z_2'}(\boldsymbol{I} - \boldsymbol{H_1})\boldsymbol{Z_2})^{-1}\boldsymbol{Z_2'}.
\end{aligned}$$

Rearranging terms finally yields:

$$\boldsymbol{H} = \boldsymbol{H_1} + (\boldsymbol{I} - \boldsymbol{H_1})\boldsymbol{Z_2}(\lambda\boldsymbol{K_4} + \boldsymbol{Z_2'}(\boldsymbol{I} - \boldsymbol{H_1})\boldsymbol{Z_2})^{-1}\boldsymbol{Z_2'}(\boldsymbol{I} - \boldsymbol{H_1}).$$

Since $\boldsymbol{H_1} = \boldsymbol{Z_1}(\boldsymbol{Z_1'Z_1})^{-1}\boldsymbol{Z_1'}$ and $\boldsymbol{Z_1} = \boldsymbol{1}$ this is equivalent to:

$$\boldsymbol{H} = \frac{\boldsymbol{11'}}{T} + (\boldsymbol{I} - \boldsymbol{H_1})\boldsymbol{Z_2}(\lambda\boldsymbol{K_4} + \boldsymbol{Z_2'}(\boldsymbol{I} - \boldsymbol{H_1})\boldsymbol{Z_2})^{-1}\boldsymbol{Z_2'}(\boldsymbol{I} - \boldsymbol{H_1}).$$

Let $p = d - 1$ then furthermore $(\boldsymbol{I} - \boldsymbol{H_1})\boldsymbol{Z_2}$ can be written to:

$$(\boldsymbol{I} - \boldsymbol{H_1})\boldsymbol{Z_2} = \left(\boldsymbol{I} - \frac{\boldsymbol{11'}}{T}\right)\boldsymbol{Z_2} = \boldsymbol{Z_2} - \frac{\boldsymbol{11'}}{T}\boldsymbol{Z_2} = \boldsymbol{Z_2} - \boldsymbol{1}\left(\frac{1}{T}\boldsymbol{1'Z_2}\right) =$$

$$= \boldsymbol{Z_2} - \left(\boldsymbol{1}\frac{\sum\limits_{i=1}^{T} z_{1i}}{T}, ..., \boldsymbol{1}\frac{\sum\limits_{i=1}^{T} z_{pi}}{T}\right) =$$

$$= \boldsymbol{Z_2} - (\boldsymbol{1}\bar{z}_1, ..., \boldsymbol{1}\bar{z}_p) =$$

$$= (\boldsymbol{z_1} - \bar{\boldsymbol{z}}_{\boldsymbol{1}}, ..., \boldsymbol{z_p} - \bar{\boldsymbol{z}}_{\boldsymbol{p}}) = \tilde{\boldsymbol{Z}}.$$

$\boldsymbol{z_i}$ is the i^{th} column of the matrix $\boldsymbol{Z_2}$ and $\boldsymbol{\bar{z}_i}$ is a $T \times 1$ vector containing the mean of $\boldsymbol{z_i}$. Thus $\tilde{\boldsymbol{Z}}$ is the matrix of the deviations from the mean for every variable z_i, $i = 1, ..., p$. As a consequence $\tilde{\boldsymbol{Z}}'\mathbf{1} = 0$ since:

$$\tilde{\boldsymbol{Z}}'\mathbf{1} = \begin{pmatrix} \boldsymbol{z_1} - \boldsymbol{\bar{z}_1} \\ \dots \\ \boldsymbol{z_p} - \boldsymbol{\bar{z}_p} \end{pmatrix} \mathbf{1} = \begin{pmatrix} \sum_{i=1}^{T} (z_{1i} - \bar{z}_1) \\ \dots \\ \sum_{i=1}^{T} (z_{pi} - \bar{z}_p) \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{T} z_{1i} - T\bar{z}_1 \\ \dots \\ \sum_{i=1}^{T} z_{pi} - T\bar{z}_p \end{pmatrix} = \begin{pmatrix} T \sum_{i=1}^{T} \frac{1}{T} z_{1i} - T\bar{z}_1 \\ \dots \\ T \sum_{i=1}^{T} \frac{1}{T} z_{pi} - T\bar{z}_p \end{pmatrix} =$$

$$= \begin{pmatrix} T\bar{z}_1 - T\bar{z}_1 \\ \dots \\ T\bar{z}_p - T\bar{z}_p \end{pmatrix} = 0$$

Considering that $(\boldsymbol{I} - \boldsymbol{H_1})$ is symmetric, it can bee seen quickly that $\boldsymbol{H}\mathbf{1} = 1$, which means that the rows of $\boldsymbol{H}$ sum up to one:

$$\boldsymbol{H}\mathbf{1} = \frac{\mathbf{1}\mathbf{1}'}{T}\mathbf{1} + (\boldsymbol{I} - \boldsymbol{H_1})\boldsymbol{Z_2}(\lambda \boldsymbol{K_4} + \boldsymbol{Z_2'}(\boldsymbol{I} - \boldsymbol{H_1})\boldsymbol{Z_2})^{-1}\boldsymbol{Z_2'}(\boldsymbol{I} - \boldsymbol{H_1})\mathbf{1} =$$

$$= \frac{\mathbf{1}\mathbf{1}'}{T}\mathbf{1} + (\boldsymbol{I} - \boldsymbol{H_1})\boldsymbol{Z_2}(\lambda \boldsymbol{K_4} + \boldsymbol{Z_2'}(\boldsymbol{I} - \boldsymbol{H_1})\boldsymbol{Z_2})^{-1}((\boldsymbol{I} - \boldsymbol{H_1})\boldsymbol{Z_2})'\mathbf{1} =$$

$$\frac{\mathbf{1}\mathbf{1}'}{T}\mathbf{1} + (\boldsymbol{I} - \boldsymbol{H_1})\boldsymbol{Z_2}(\lambda \boldsymbol{K_4} + \boldsymbol{Z_2'}(\boldsymbol{I} - \boldsymbol{H_1})\boldsymbol{Z_2})^{-1}\tilde{\boldsymbol{Z}}'\mathbf{1} =$$

$$= \frac{\mathbf{1}\mathbf{1}'}{T}\mathbf{1} = \frac{\mathbf{1}}{T}\mathbf{1}'\mathbf{1} = \frac{\mathbf{1}}{T}T = \mathbf{1}.$$

References

Abberger, K., Nierhaus, W. (2008): *"Die ifo-Kapazitätsauslastung - ein gleichlaufender Indikator der deutschen Industriekonjunktur,"* ifo Schnelldienst, 61(16), 15-23.

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

Anas, J., Ferrara, L. (2002): *"Detecting cyclical turning points; the ABCD approach and two probabilistic indicators,"* paper presented at the 26$^{\text{th}}$ Ciret Conference in Taiwan, October 2002 (www.coe.ccip.fr).

Anas, J., Ferrara, L. (2004): *"A Comparative Assessment of Parametric and Non-Parametric Turning Points Detection Methods: The Case of the Euro Zone Economy,"* Monographs of official statistics: Papers and proceedings of the third Eurostat colloquium on modern tools for business cycle analysis, Luxembourg: Office for Official Publications of the European Communities, 2004.

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. (2014a): *"Reducing the Excess Variability of the Hodrick-Prescott Filter by Flexible Penalization,"* Munich Discussion Papers in Economics No. 17940.

Blöchl, A. (2014b): *"Penalized Splines as Frequency Selective Filters - Reducing the Excess Variability at the Margins,"* Munich Discussion Papers in Economics No. 20687.

Blöchl, A. (2014c): *"Trend Estimation with Penalized Splines as Mixed Models for Series with Structural Breaks,"* Munich Discussion Papers in Economics No. 18446.

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

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.

Bundesbank (2014): *"Monatsbericht April 2014,"* Monatsbericht der deutschen Bundesbank, Jahrgang 66, Nr. 4.

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

Claeskens, G., Krivobokova, T., Opsomer, J. (2009): *"Asymptotic properties of penalized spline estimators,"* Biometrika, 96, 529-544.

Colgley, T., Nason, J. (1995): *"Effects of the Hodrick-Prescott filter on trend and difference stationary time series: Implications for business cycle research,"* Journal of Economic Dynamics and Control, 19, 253-278.

Crainiceanu, C., Ruppert, D., Carroll, R. (2005): *"Spatially adaptive Bayesian P-splines with heteroscedastic errors."* Johns Hopkins University Dept. of Biostatistics Working Paper Series No. 1061, Berkeley Electronic Press.

Dagum, E. B., Luati, A. (2004): *"A linear transformation and its properties with special applications in time series filtering,"* Linear Algebra and its Applications, 338, 107-117.

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.

Dickey, D. A., Fuller, W. A. (1979): *"Distributions of the Estimators for Autoregressive Time Series with a Unit Root,"* Journal of the American Statistical Association, 74, 427-431.

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.

Eubank, R. L. (1983): *"The Hat Matrix of Smoothing Splines,"* Technical Report No. SMU/DS/TR-179.

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

Flaig, G., Plötscher, C. (2001): *"Estimating the Output Gap Using Business Survey Data. A Bivariate Structural Time Series Model for the German Economy,"* ifo Studien, 221-232.

Flaig, G. (2003): *"Strukturelle Zeitreihenmodelle,"* Wirtschaftswissenschaftliches Studium, 5, 258-263.

Flaig, G. (2005): *"Time Series Properties of the German Production Index,"* Allgemeines Statistisches Archiv, 89, 419-434.

Flaig, G., Wollmershäuser, T. (2007): *"Does the Euro-zone Diverge? A Stress Indicator for Analyzing Trends and Cycles in Real GDP and Inflation,"* W. Franz, H.J. Ramser, M. Stadler (ed.), Dynamik internationaler Märkte. Tübingen: Mohr Siebeck.

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

Gabisch, G., Lorenz, H. W. (1989): *"Business Cycle Theory,"* Springer-Verlag, Berlin.

Goldrian, G. (2005): *"Weaknesses of the Baxter-King-filter: Is a pattern-based filter an alternative?,"* Jahrbücher für Nationalökonomie und Statistik, 225 (4), 386-393.

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, Cambridge.

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

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

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

Hayes, K., Haslett, J. (1999): *"Simplifying general least squares,"* American Statistician, 53, 376-381.

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

Hodrick, R. J., Prescott, E. C. (1997); *"Post-War 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, New York.

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

Kauermann, G., Opsomer, J. (2011): :*"Data-driven selection of the spline dimension in penalized spline regression,"* Biometrika, 98(1), 225-230.

King, R., Rebelo, S. (1993): *"Low frequency filtering and real business cycles,"* Journal of Economic Dynamics and Control, 17, 207-231.

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.

Langmantel, E. (1999): *"Das ifo-Geschäftsklima als Indikator für die Prognose des Bruttoinlandsprodukts,"* ifo Schnelldienst, 52(16-17), 16-21.

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

Maravall, A., del Rio, A. (2001): *"Time Aggregation and the Hodrick-Prescott-Filter,"* Banco de Espana Servicio de Estudios, Documento de Trabajo n. 0108.

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.

McCulloch, C. E., Searle, S. R. (2001): *"Generalized, Linear, and Mixed Models,"* Wiley, New York.

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.

Okun, A. (1962): *"Potential GNP: Its Measurement and Significance,"* American Statistical Association, Proceedings of the Business and Economics Statistics Section, 98-103.

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.

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

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

Pinheiro, J. C., Bates, D. M. (2000): *"Mixed-Effects Models in S and S-PLUS,"* Springer-Verlag, New York.

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): *"Investigating economic trends and cycles,"* Boumans, M. (Ed.), Palgrave Handbook of Econometrics, Volume 2: Applied Econometrics, Palgrave Macmillan, 243-307.

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.

Proietti, T., Luati, A. (2007): *"Least Squares Regression: Graduation and Filters,"* Boumans, M. (Ed.), Measurement in Economics: A Handbook, ch. 16, Academic Press.

Ravn, M. O., Uhlig, H. (2002): *"On adjusting the Hodrick-Prescott filter for the frequency of observations,"* The Review of Economics and Statistics, 84, 371-380.

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.

Robinson, G. K. (1991): *"That BLUP s a good thing: The estimation of random effects,"* Statistical Science, 6, 15-51.

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

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

Sargent, T. J. (1987): *"Macroeconomic Theory,"* Academic Press, Boston.

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.

Schlicht, E. (2008): *"Trend Extraction from Time Series with Structural Breaks and Missing Observations,"* Journal of the Japanese Statistical Society, Vol. 38 No. 2, 205-292.

Schlittgen, R., Streitenberg, B. (2001): *"Zeitreihenanalyse,"* Oldenburg Wissenschaftsverlag, München, Wien.

Schumpeter, J. A. (1939): *"Business Cycles. Volume I,"* McGraw-Hill, New York, London.

Searle, S. R., Casella, G., McCulloch, C. E. (1992): *"Variance Components,"* Wiley, New York.

Stamfort, S. (2005): *"Berechnung trendbereinigter Indikatoren für Deutschland mit Hilfe von Filterverfahren,"* Deutsche Bundesbank, Discussion Paper Nr. 19.

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

Vonesh, E. F., Chinchilli, V. M. (1997): *"Linear and Nonlinear Models for the Analysis of Repeated Measures,"* Marcel Dekker, New York.

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

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

Whittle, P. (1983): *"Prediction and Regulation by Linear Least-Square Methods,"* University of Minnesota Press, Minneapolis.

Eidesstattliche Versicherung

Ich versichere hiermit eidesstattlich, dass ich die vorliegende Arbeit selbständig und ohne fremde Hilfe verfasst habe. Die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sowie mir gegebenen Anregungen sind als solche kenntlich gemacht. Die Arbeit wurde bisher keiner anderen Prüfungsbehörde vorgelegt und auch noch nicht veröffentlicht. Sofern ein Teil der Arbeit aus bereits veröffentlichten Papers besteht, habe ich dies ausdrücklich angegeben.

München, 25. Juni 2014

Andreas Blöchl