Using Penalized Spline, Generalized Additive Model and Mixed Model Regression Techniques to Examine Univariate and Multivariate Time Series and in particular Business Cycles

Timo Teuber

eingereicht im Juni 2012

Dissertation zur Erlangung des Grades eines Doktors der Wirtschaftswissenschaften

vorgelegt der Fakultät für Wirtschaftswissenschaften an der Universität Bielefeld

Dekan: Prof. Dr. Herbert Dawid

Gutachter: Prof. Dr. Peter Flaschel

Gutachter: Prof. Dr. Göran Kauermann

Gedruckt auf alterungsbeständigen Papier °° ISO 9706

Contents

List of Figures

List of Tables

1 Introduction

In this work, business cycles, which are short-term fluctuations in macroeconomic time series, will be analyzed. This work is focussing mainly on the work of Kauermann, Teuber, and Flaschel (2012), Teuber (2012b) and Teuber (2012a) in which the business cycles have been estimated simultaneously with a new, state-of-the-art estimation technique which revolutionizes the business cycle discussion. The approach takes the phrase business cycles more literally such that it will be assumed that the detrended time series follow a circular or elliptical trajectory over time. The underlying model uses a non-parametric approach to avoid specifying the structure of the unknown function. In more detail, a penalized spline approach will be used such that just the smoothness of the estimated function needs to be determined. It will be assumed that the time series can be decomposed into a long-term trend, which will be estimated in standard textbooks with the Hodrick-Prescott filter, a short-term component, which represents the business cycles with a length of more than one year and less than twelve years, and the residuals, which are white noise. Kauermann, Teuber, and Flaschel (2012) and Flaschel, Kauermann, and Teuber (2005) analyzed in more detail the long-term trends, which are known as Friedman and Goodwin cycles for the underlying time series. The short-term cycles have been the focus of the work of Kauermann, Teuber, and Flaschel (2012) and Proaño, Flaschel, Diallo, and Teuber (2008) and helped to provide evidence for several macroeconomic hypotheses due to the fact that the approach allows the time series to be decomposed in an objective way into a short-term and long-term component. Teuber (2012b) extended the work by Kauermann, Teuber, and Flaschel (2012) to estimate not only two time series simultaneously, but also any arbitrary number of time series. Furthermore, the topic has been investigated in more detail by Teuber (2012b), generating not only a methodology to date business cycles

but also to calculate instantaneously the duration of the business cycles and to smoothly identify any possible business cycle stage. Teuber (2012a) used a three-dimensional model instead of a two-dimensional model. Similar to the work of Teuber (2012b), the model was extended to estimate not only three time series but any arbitrary number of time series simultaneously.

It is remarkable that even though econometricians as well as statisticians have overcome the hurdles of parametric models for decades, most economical researchers still use parametric or awkward extensions to analyze and estimate time series. Thanks to the work of O'Sullivan (1988), Eubank (1989), Wahba (2003), Fan and Gijbels (1996), as well as Eilers and Marx (1996), the non-parametric estimation techniques and smoothers have become popular and are standard in any serious statistical tool. Ruppert, Wand, and Carroll (2003) give in their book numerous examples of the flexibility of the penalized spline approach, and how this model can be used in different research fields. In chapter 3, the statistical ground work will be presented: In chapter 3.1 the linear regression will be extended to a broader family of distributions. The non-parametric approach of penalized splines and B-splines will be shown in chapters 3.2 and 3.3, respectively. The mixed model, which consists of fixed and random effects, will be described in chapter 3.4. The Generalized Additive Model will be discussed in chapter 3.5 and chapter 3 will be concluded in chapter 3.6 with an example of how to use the previous approaches to model time-varying coefficients.

In chapter 4, the statistical models presented in chapter 3 will be applied to discuss several economic models. In chapter 4.1, the work of Ernst and Teuber (2006) for the European Socio-Economic Models of a Knowledge-based Society (ESEMK) will be presented in which the cyclicality of fiscal policies in OECD countries will be calculated using a time-varying coefficient

model.

The penalized spline approach was used by Flaschel, Tavani, Taylor, and Teuber (2008) to get a structuralist model of the wage-price spiral with non-linear demand pressure terms. The penalized spline approach is an elegant technique to estimate the first derivative of an estimated function, and given the nature of the approach a confidence and/or prediction interval can be calculated.

The research for the Organisation for Economic Co-operation and Development (OECD), which was published by Ernst and Teuber (2008) and Ernst and Teuber (2012), will be presented in chapter 4.3. The statistical tools presented in chapter 3 were useful for analyzing the Dutch tax benefit system and the life cycle employment. The research was very useful for calibrating an overlapping generation model for the Dutch economy.

In chapter 5, the statistical model of Kauermann, Teuber, and Flaschel (2012) will be presented in more detail. A multivariate approach which decomposes the time series into a long-term component and a short-term component, which are known as business cycles in standard textbooks, will be discussed. It will be assumed that the short-term component can be explained by a deterministic and a stochastic part, namely white noise. Furthermore, it will be assumed that the deterministic part describes a circular or elliptical structure over time. The radius and the angle of the unknown circular function will be left unspecified and a penalized spline approach will be used to estimate the function. The short-term and long-term time series will be estimated in a two-stage approach such that the ”best” long-term trend will be used to estimate the short-term fluctuations.

An extension of the two-dimensional model given by Teuber (2012b) will be shown in chapter 5.2. In this section, the two-dimensional circle will be reformulated: First, a two-dimensional ellipse will be defined, which allows the model to be more flexible, as it will be shown. Mathematically, the model can be extended in this form to fit not only two, but any numbers of time series, which be shown in chapter 6.3 and by Teuber (2012b).

Teuber (2012a) extended the two-dimensional model by Kauermann, Teuber, and Flaschel (2012) to use more than two time series. Instead of polar coordinates, the three-dimensional model will be preferred, such that spherical coordinates will be used. Given the more complex structure, the model encompasses a long-term cycle structure which might be wanted if a longer-term structure is reasonable. Teuber (2012a) extended the model, in a similar way as in Teuber (2012b) in the two-dimensional case, to be able to fit not just three time series but an arbitrary number of time series, as shown in section 5.3.

The empirical results of the statistical model in chapter 5 will be presented in chapter 6. Kauermann, Teuber, and Flaschel (2012) analyzed the bivariate trajectory of the employment rate and the inflation as well as the employment rate and the wage share for the US. The trajectory of the long-term trends show a circular form over a period of 50 years. The Goodwin and Friedman cycles are well known in standard textbooks and have been confirmed using the new statistical approach. The short-term fluctuations have been discussed in more detail by Proaño, Flaschel, Diallo, and Teuber (2008) and will be presented in chapter 6.1 such that the length of cycles, the causality as well as the leading and lagging characteristic will be shown for different political regimes. Furthermore, the chapter will discuss the findings of the work by Flaschel, Groh, Kauermann, and Teuber (2008).

The extended two-dimensional model formulated by Teuber (2012b) has been applied to the leading, coincident and lagging indicators provided by the Conference Board and will be shown in chapter 6.3. The three-dimensional model developed by Teuber (2012a) has been applied to the same data set and the results will be shown in chapter 6.4. It should be noted that the models by Kauermann, Teuber, and Flaschel (2012), Teuber (2012b), and Teuber (2012a) are not only state-of-the-art from a statistical point of view, but also provide new insights for the business cycle discussion.

2 Business Cycles

Business cycles refer to a fluctuation in the economy either locally or globally which occurs for several quarters and in general follows a certain pattern which is repeated constantly over time. Burns and Mitchell (1946) defined business cycles as a "type of fluctuations in the aggregate economic activity". Long and Plosser (1983) agreed with this definition and stated that "business cycles refers to the joint time-series behavior of a wide range of economic variables". Instead of observing these fluctuations directly in the time plot of a time series, one has to derive the "detrended" time series of the growth variables; see figure 2.1 for the GDP of the US from the year 1947. The "detrended" time series will reveal the fluctuations quite well.

Lucas (1977) redefines business cycles as "the deviation of the Gross National Product from a trend", which can differ from an exponential growth rate over time. Kydland and Prescott (1990) propose "a curve which students of business cycles and growth would draw in" and suggest using the Hodrick and Prescott (1997) filter. Although the trend is essential to define a good estimation of the fluctuations, there is at this time no "optimal" way to distinguish the trend from the fluctuations. Stock and Watson (1999) prefer in their extensive work the band pass filter developed by Baxter and King (1999).

Both methods have in common that the smoothing parameters need to be adjusted to find an "optimal" amount of smoothness of the trend function. Hodrick and Prescott (1997) and Kydland and Prescott (1990) used a fixed smoothing parameter ($\lambda = 1600$), which was subjectively proposed for one particular time series. Baxter and King (1999) and Stock and Watson (1999) choose the tuning parameter pair ($p = 6$ and $q = 32$) because they pointed

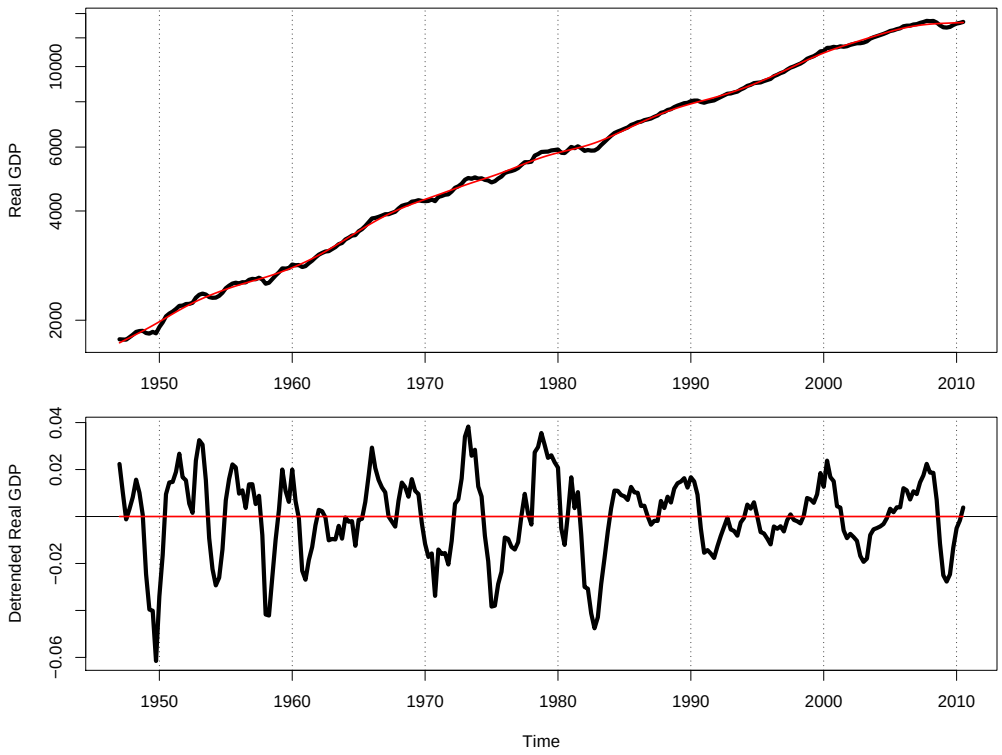

Figure 2.1: Real US Gross Domestic Product (GDP) over time: The upper plot shows the real US GDP (black line) and the long-term trend (red line) estimated by the Hodrick-Prescott filter($\lambda = 1600$). The lower plot shows the deviation of the observed real GDP from the long-term trend (black line); the long-term trend is equal to the horizontal axis at level zero (red line).

---

out that a cycle should last at least six and at most 32 quarters.

In the following sections, different kinds of fluctuations, and in particular business cycles, will be discussed in more depth. A classification of cyclicalities by the length of a "typical" cycle is a first starting point: Short-term cycles usually last one year or less, medium-term cycles (among them the

business cycles) with a length of more than a year and less than 12 years, and long-term cycles of more than 12 years. Furthermore, a brief overview of different economic models will be given to explain the cyclical behavior of the economy, which will be concluded with the Real business cycle theory. The chapter will be concluded with an empirical analysis of business cycles and an outlook on the following chapters.

2.1 Cyclicality in Economies

The definition of business cycles by Burns and Mitchell (1946) refers to "economic activities of nations that organize their work mainly in business enterprises". In this context, the fluctuations might be explained best from a microeconomic-based macroeconomy, which means that economic fluctuations can be observed as business cyclicalities/ fluctuations which will not be compensated by competitors and thus will be observed on an aggregated level. For example, business activities might fluctuate due to consumption habits: Restaurants are facing on an individual and aggregated level cyclicalities (measured by customers and/or revenues) either by demand and/or supply in the form of opening hours which are in the best case "optimized" subject to the cyclicality of demand over the day (with peaks around lunch and diner time) and/or week (with higher revenue either during the week or at weekends).

Most economic (aggregated) numbers however will not be calculated on an instant or daily basis, such that the short cyclicalities will in general not be discussed but have to be taken into account, if numbers of longer periods are to be calculated. For instance, the weekly reported new jobless claims has calendar effects if the number of weekdays differ due to holidays. Furthermore, it is obvious that monthly (raw) data for February might often differ

if the time series is not representing a level but a growth number. These calendar effects have to be accounted for as they might be the results of short-term fluctuations and/or economic activities.

More prolonged fluctuations are seasonal effects, which are most often cycles over the year. Numerous industries and sectors are face fluctuations over the year which are to some degree explainable due to the kind of business. For instance, the demand for heating oil, ice cream, sun lotion, fruit, and winter or summer clothing fluctuates over the year due to time-varying preferences and needs of customers as well as the supply. These might be influenced by the weather, such that the seasonal effect is not necessarily identical over the years in terms of timing, magnitude, and length.

It is worth noting that these seasonal effects which are characterized by a cycle length of a year might have an effect on more prolonged cycles. A bad harvest or a hurricane season might have an enormous effect on the following years either in terms of supply and/or a shift in demand.

Long-term fluctuations like the Kuznets (1934), Kuznets (1930), and Kondratieff (1984) waves are longer than the typical business cycles and might affect or be the underlying trend of the business cycles. The infrastructural investment waves by Kuznets are supposed to last 15-25 years and the technology cycles by Kondratieff 45-60 years. Although both long-term cycles might influence the business cycles directly via stronger/weaker long-term growth, which might be interpreted as a trend of the underlying time series, the explicit relationship will be neglected in general. It is worth noting that the interaction of medium-term cycles and long-term cycles is not one-sided, such that even business cycles events, for instance the oil price shock in the '70's or the financial crisis (2008-2009), might have an effect on the longer

cycles, which might determine the end of a new cycle. This new normality might be observed in different growth rates in production and/or new inflation/interest rate regimes.

Although the source of the medium-term (business) cycles and their length differ over time and between theories, the medium-term cycles face the same patterns as the short- and long-term cycles. Burns and Mitchell (1946) state that "a cycle consists of expansions occurring at about the same time in many economic activities, followed by similarly general recession, contractions, and revivals which merge into the expansion phase of the next cycle". The characteristic of the economic fluctuation can be stylized as cycles which however vary in duration from more than one year to ten or twelve years as defined by Burns and Mitchell (1946).

The National Bureau of Economic Research (NBER) provides for the US the dates of peaks and troughs of the business cycles, where a recession is "defined as a significant decline in economic activity spread across the economy, lasting more than a few months, normally visible in real GDP, real income, employment, industrial production, and wholesale-retail sales." The NBER does not define a recession in terms of two consecutive quarters of decline of real GDP; see http://www.nber.org/cycles.html.

2.2 Business Cycles in the Classical, Keynesian, Monetarism, Neo-Classical, and New-Keynesian View

Business cycles are easy to describe but hard to explain, not only because business cycle theory is a source of controversy. Mainstream theories are numerous and differ in the source of fluctuations and will be discussed here in a chronological order of appearance.

The classical school of thought started with Smith (1776) in which all economic agents are completely rational and markets clear instantaneously due to equilibrium forces. If the economy operates below natural output, nominal wages and prices will be cut, such that the production will reach the natural output. Following Say's Law (aggregated supply creates its own aggregated demand), the price will adjust such that demand will equal supply. Unemployment can be treated in this thought as a too slow reaction of wages on the labor supply, similar to the goods market where prices (here wages) should adjust immediately to reach natural output (here employment). The government has to ensure competition and any fiscal intervention is unwanted. The quantity theory of money suggests that inflation will not affect the optimal decisions, which results in no business cycle measured in real output, but in fluctuations in prices.

"The General Theory of Employment, Interest and Money" by Keynes (1936) was the start of the Keynesian revolution where no self-correction has been modelled due to the sticky prices. Wages will not be adjusted instantly due to minimum wages and/or contracts as well as the bargaining power of unions. In contrast to the Classical model, Keynes assumed that demand creates supply. The effective demand depends on the interest rates which are determined by money supply, such that money is no longer neutral. The government is able to intervene if markets do not clear via government expenditures, tax changes, money supply, and/or transfer payments. As a result, the Keynesian view is able to explain unemployment as well as fluctuations in aggregated demand.

The Monetarism view by Friedman (1957), Friedman (1968) and Friedman (2008) is based on the Classical view but has been extended by imperfect

information. The labor supply depends on the expected real wage based on expectations on the inflation. The presence of imperfect price information allows the output to deviate from the long-run natural output, which can be interpreted as business cycles. The labor market will clear at a natural rate of unemployment. The government should reduce restrictions (minimum wages, unions) to keep the natural rate of unemployment low. The markets are expected to clear by themselves such that the government should not intervene. Furthermore, the money supply should be transparent in a way to grow with output and avoid inflation via wage-price spiral due to anticipation of money supply shocks.

The Neo-Classical model is based on Friedman's model, but agents form rational expectations based on the available information, such that prediction errors can be treated as random and independent, which is a main difference in the assumptions of the Monetarism view. In accordance to the Classical model, the markets clear instantaneously by assumption, and the labor market will clear at the natural rate of unemployment. However, if money supply changes are announced, firms and households adapt their forecasts such that the effect vanishes; unannounced money supply changes will be adapted later via prices and wages such that, all in all, money is neutral. The economic output will deviate from natural output due to random deviations from forecasts and thus explains business cycles.

The Neo-Keynesian view is based on contract-based wages and price stickiness, such that markets do not clear instantaneously. Business cycles are caused by supply and demand shocks as well as monetary shocks which are hampered by sticky prices. However, the government can intervene similarly to the Keynesian model via stabilizers in the form of taxes, transfers, and expenditures. Furthermore, monetary policy in the form of money supply will

affect the interest rate, which will influence the effective demand via business investments.

2.3 Real Business Cycle Theory

In the real business cycle models, random fluctuations in productivity are the cause of economic fluctuations. The models by Kydland and Prescott (1982) and Long and Plosser (1983) are based on rational expectations given by the groundwork of Lucas (1977). Under the assumptions of complete price and wage flexibility and rational expectations, the theory explains economic fluctuations caused by technology shocks as well as natural shocks. Lucas (1980) notes that the equilibrium models of business cycles, which are based on Arrow and Debreu (1954), Arrow (1963) and Debreu (1959) contingent claims approach, describe an equilibrium path of fluctuations. In former models, the fluctuations were treated as deviations from equilibrium such that the economy has to converge back to the equilibrium over time.

Kydland and Prescott (1982) point out that neither factories nor ships are built in a day, such that a multiple-period construction is crucial to explain aggregated fluctuations. A competitive equilibrium model for one individual with a non-time separable utility function in one good economy has been modelled. The current utility thus depends on past work time choices, which admits greater intertemporal substitution of leisure (and working time). The equilibrium of the Robinson Crusoe economy with a noisy productivity parameter will be approximated to calibrate the model to measure the US economy. The selected parameters led to a good fit of the real output, consumption, investment, inventories, productivity, and real interest rates for the post-war period in terms of standard deviation and/or correlation with output.

Long and Plosser (1983) built a Robinson Crusoe economy with iid shocks which lead to serial- and auto-correlated output shocks in different sectors. The infinitely long-living individuals choose their preferred consumption-production plans over several commodities produced in different sectors. A shock in one sector will be "transferred" via the production technology into other sectors. The uncorrelated shocks lead to positive serial correlation and positive cross-sectional correlation. Long and Plosser (1983) point out that, similar to the Kydland and Prescott (1982) model, efforts to stabilize the economy can only serve to make consumers worse off.

Backus, Kehoe, and Kydland (1992) point out that "real business cycle theories have accounted for many of the features of postwar US business cycles". They extend the one-economy model to a two-country economy model such that an economy "can borrow and lend in international markets by running trade surpluses and deficits". The model was designed to explain the correlation between output fluctuations in different countries, which are assumed to derive from different technology shocks across countries but will be "exported" and "imported", respectively. Backus, Kehoe, and Kydland (1992) extent the Kydland and Prescott (1982) model and assumed in their open economy model that labor is immobile but capital mobile. The model has been calibrated to reflect the US and the European economic system: US output and consumption are modelled quite well, while investment seems to be too high in terms of variability. The model is not able to depict the trade balance in a numerical manner. Furthermore, it is not able to show the correct cyclicality of real output across countries, but positive correlations for consumption can however be verified.

All in all, even the simple business cycle models seem to depict the em-

pirical evidence of business cycles, but a simultaneous explanation seems to need a more sophisticated or at least elaborated model.

2.4 Empirical Analysis of Business Cycles

Given the theoretical background of the equilibrium models, more and more statistical approaches have been used to capture the behavior of the business cycles. Stock and Watson (1999) compare seventy macro-economic time series with the real output. Besides a graphical representation, the relationship was also captured by the autocorrelation statistic and the Granger causalities. Hamilton (1989) and Hamilton (2005) uses discrete Markov switching models to estimate univariate time series, and on the other hand the estimations produce as a by-product time series which can identify periods of recessions and the average duration of recessions as well as boom times. Stock (1987) shows that most of the statistical approaches can only be used in a univariate case. Furthermore, Stock (1987) distinguishes between the observed (linear) calendar time and the (non-linear) economic time, such that the business cycle pattern can be described by the economic time scale in a simpler form. Although this approach works well for the univariate case, it seems that the model will fail if it is extended to the multivariate case.

For the purpose of this work, the concept of business cycles will be discussed from a different perspective. The economic fluctuations stylized in figure 2.2 show that the trend is essential to identify the cycle of boom, recession, depression and expansion.

In general, the real GDP is used as an anchor or, better named, as a reference cycle which will be compared to the other time series. In figure 2.3, the (stylized) fluctuations of the real GDP (black curve) and the industry

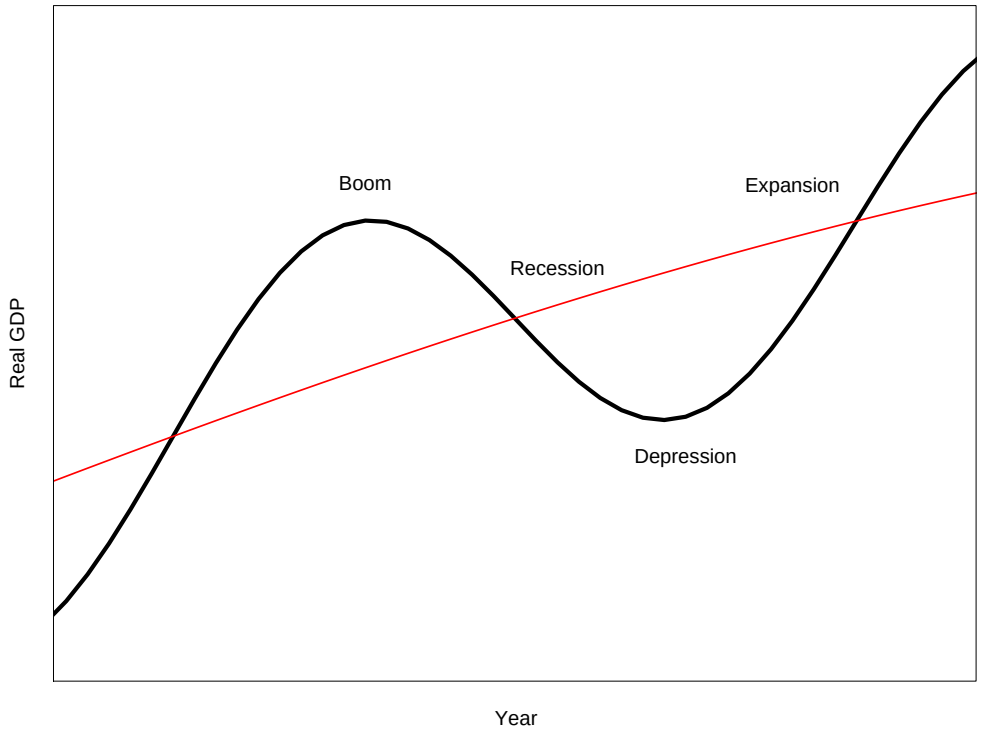

Figure 2.2: Stylized business cycle formation: The real GDP (black lines) oscillates around the long-term trend (red line), such that the peak/"boom" is followed by a decline (either in absolute or in relative terms) which is called "recession" until it bottoms out in "depression", which ends with the new "expansion" until a new "boom"/peak is reached.

---

production (green curve) are shown, which is an example for a pro-cyclical time series which means that industry production peaks roughly when the reference time series (real GDP) peaks. Furthermore, the industry production is neither leading nor lagging, such that the time series differ in the magnitude of the amplitudes at first glance.

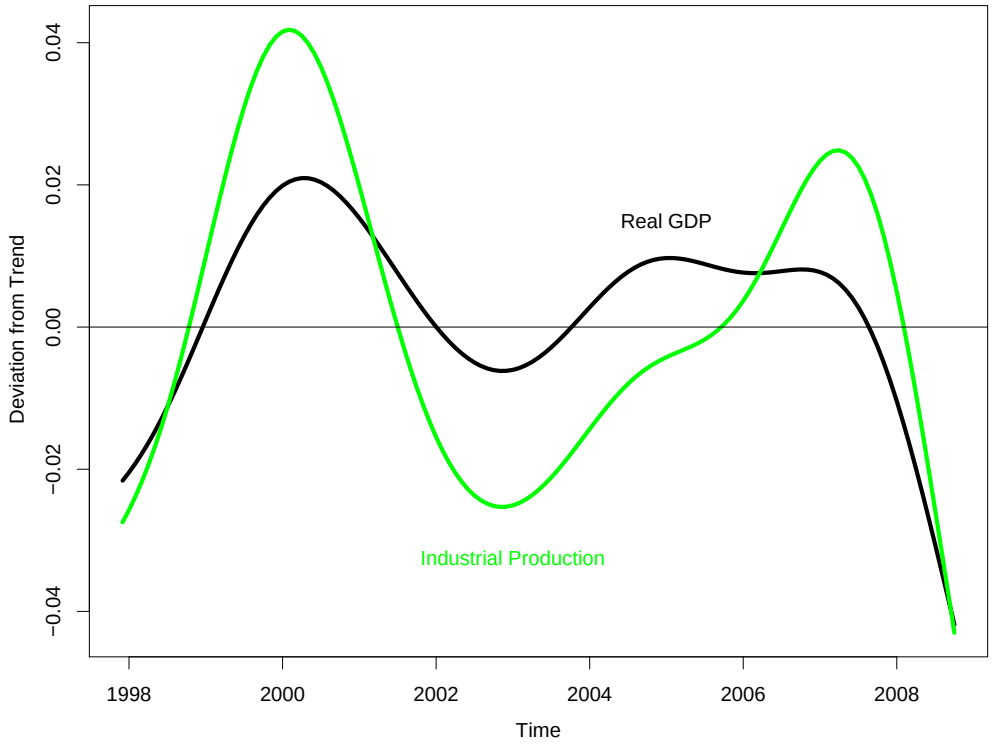

Figure 2.3: Real GDP fluctuation and a coincident time series: The real GDP (black line) and industrial production (green line) coincides.

The unemployment rate is an example of a counter-cyclical time series and is shown in figure 2.4. Whenever the GDP (black curve) is increasing, the unemployment rate (red curve) decreases, and vice versa. Obviously, the employment rate is pro-cyclical, such that the employment rate and real GDP coincides in a stylized view.

Furthermore, a distinction in terms of timing can be made, such that time series might lead, coincide, or lag the reference time series, namely real GDP. In figure 2.5, the leading indicator (green curves) peaks first, which is fol-

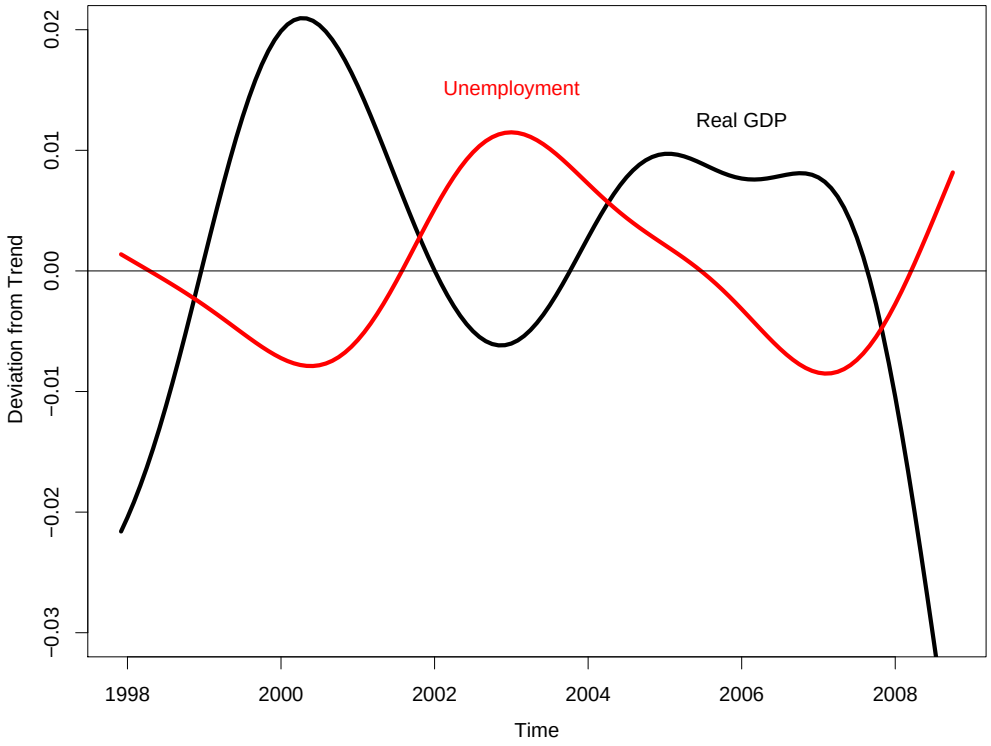

Figure 2.4: Real GDP fluctuation and a counter-cyclical time series: The real GDP (black line) and unemployment rate (red line) are counter-cyclical.

---

lowed by the coincident indicators (black curve) and the lagging indicators (red curve). The magnitude of the amplitudes is not of interest for this classification asthe time series can be scaled for the purpose of visualization.

The green curve in figure 2.5 is from data provided by the Conference Board which since January 1959 has provided the US Leading Index constructed out of ten economic indicators, namely average weekly hours (manufacturing), average weekly initial jobless claims, manufacturers' new orders (consumer), vendor performance (slower deliveries), manufacturers' new orders

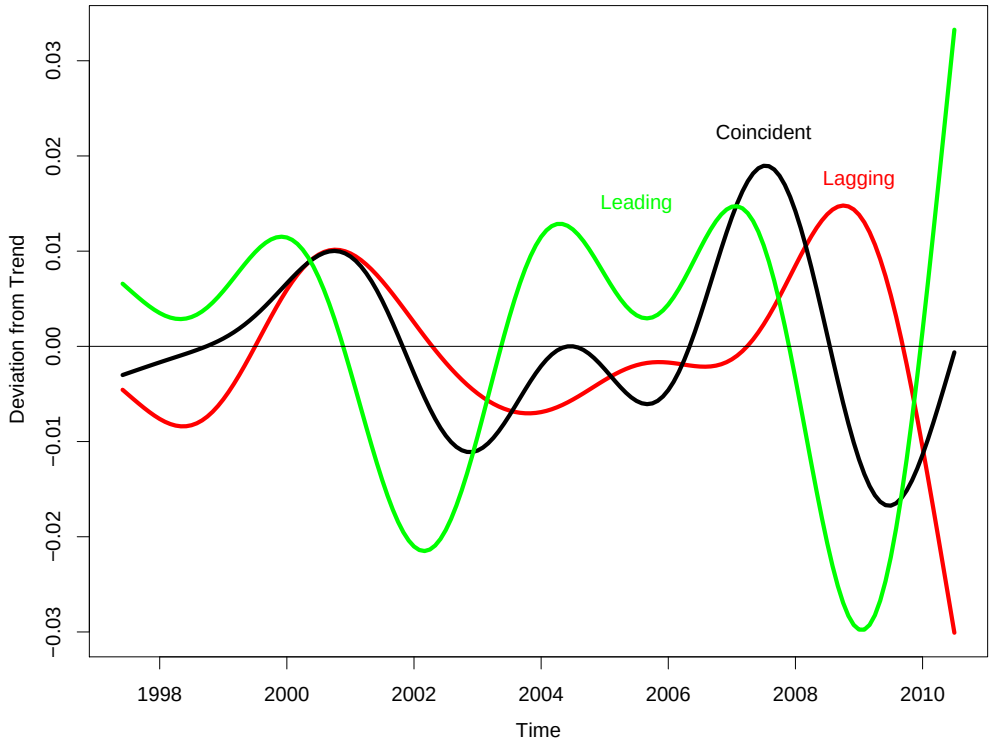

Figure 2.5: Conference Board leading, coincident and lagging indicators: Leading indicators (green line), coincident indicators (black line) and lagging indicators (red line).

(capital), building permits (new private housing units), stock prices (500 common stocks), M2 money supply, interest rate spread, and index of consumer expectations.

The black curve in figure 2.5 is given by the Conference Board which has provided the US Coincident Composite of 4 coincident indicators since January 1959. The four coincident indicators are unequally weighted combinations of the employees on non-agricultural payrolls, personal income less transfer

payments, industrial production and manufacturing, and trade sales. This means, referring back to figure 2.3, that the indicators coincide more or less with the real GDP.

Furthermore, the Conference Board provides the US Lagging Economic indicators, the red curve in figure 2.5, which consists of a total of seven lagging composite indicators, namely average duration of unemployment, inventories to sales ratio (manufacturing and trade), labor cost per unit of output (manufacturing), average prime rate, commercial and industrial loans, consumer installment credits to personal income ratio, and consumer price index for services.

In figure 2.6, the two-dimensional trajectories of the leading, coincident, and lagging indicators are shown. The trajectories of the two-dimensional plot of the pro-cyclical time series are shown in the three upper-left graphics turning clockwise. They have in common that the leading time series is measured on the y-axis and the lagging time series on the x-axis. The trajectory of the two pro-cyclical time series in the lower-right graphics are however turning counter-clockwise. Here, the leading time series is measured on the x-axis and the lagging time series on the y-axis. The time lag of the two time series will influence the trajectory in the way that the curve is more elliptical if the lag is small and the trajectory is more circular if the time lag is getting bigger. Figure 2.6 shows that no cycle is alike and that they vary in timing, amplitude, positioning, and length.

The graphical presentation shown in figure 2.6 is not new to economists, but it shows that the cyclicality should not be measured by the correlation due to the non-linear relation of the two time series. However, the cyclicality can be measured recalling the nature of cycles.

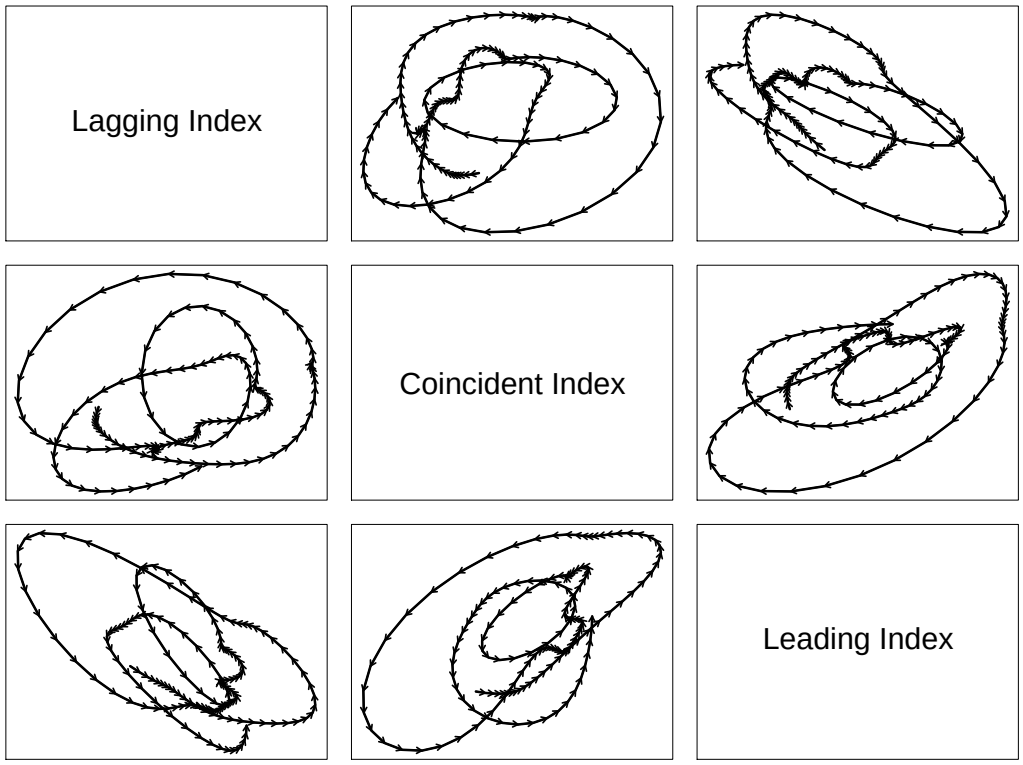

Figure 2.6: Smoothed two-dimensional trajectories of leading indicators, coincident indicators, and lagging indicators: First row: Middle plot (coincident vs lagging), right hand plot (leading vs lagging); second row: Left hand plot (lagging vs coincident), right hand plot (leading vs coincident); third row: Left hand plot (lagging vs leading), middle plot (coincident vs leading). The arrows indicate the trajectory path over time.

Under the assumption, that the classification of the Conference Board is valid, the trajectory of the price Phillips curve with the unemployment rate on the x-axis (counter-cyclical coincident time series) and the inflation on the the y-axis (lagging time series) is turning clockwise. The Goodwin cycles

with the employment rate (coincident time series) on the y-axis and the labor cost per unit of output (lagging time series) are turning clockwise. Both observations have been confirmed by theory and empirical research, and are well-known.

In chapter 5, a model will be presented to describe and estimate business cycles which can be treated as circles (cycles) similar to those shown in figure 2.6. The results of the estimated time series are presented in chapter 6. The statistical ground work for the circular presentation is given in the next two chapters and is needed to model the time-variation in the cyclicality of economic time series.

3 Generalized Linear Models, Penalized Splines, B-Splines, Generalized Linear Mixed Models, and Generalized Additive Models

In this chapter modern statistical tools will be discussed which are being used more and more these days due to their availability in most statistical packages. Classical statistical methods for regression and time series analysis are in most cases straightforward if the error terms are normal. The Ordinary Least Squares (OLS) or also called Linear Regression (LR) method is easy to use and to interpret. However, in most cases the assumption of normal errors is too stringent, such that a more generalized tool for non-Gaussian data must be used. In other cases, the relationship between the regressor and the other variable(s) is not known or not wanted to be specified in advance. Furthermore, other analysis requires the testing of a linear relationship for which the linear regression technique might not be the preferred choice to test the hypothesis.

In the following chapter 3.1 a generalization of the OLS method will be presented where the regression for a broader range of distributions, namely the family of exponential distributions, is possible and among them the Gaussian, binomial, multinomial, Poisson, exponential, negative binomial and gamma distributions are named. The most important advantage of the Generalized Linear Model (GLM) is that the general theoretical framework allows the use of a statistical tool (nowadays already standard in the main-stream statistical programs like S-Plus, R, Matlab, Stata, and SAS) for the regression, so only the underlying distribution needs to be specified.

Furthermore, the penalized spline regression technique, which will be ex-

plored in chapter 3.2, is a very powerful tool and can be used for scatterplot smoothing to get a first inspection of the (plotted) data. The advantage of penalized splines over the kernel estimation technique is one reason why the tool is a standard package in most statistical software programs. Due to the construction of penalized splines, the computing time can be dramatically reduced and the model is more flexible to use and to interpret if a parametric variable is included in the model. The selection of the smoothing parameter is fully worked out as will be shown in this work. In the following sections, a generalized version which can be used for the family of exponential distributions will be described so that the technique is useable for most applications. Afterwards, in chapter 3.3 a special case of penalized spline basis, namely B-splines, and the corresponding penalty matrix will be introduced.

In chapter 3.4, the (Generalized Linear) Mixed Model (GLMM) will be explained not only to introduce random effects, but also to show the linkage to the penalized splines approach such that the smoothing parameters can be selected in a more elegant way via the Restricted Maximum Likelihood (REML) technique. Nevertheless, random effects can be used if panel data for different groups (which consist of one or more individual observations) are collected and allow group-specific analysis of the parameters and/or test for these parameters.

In chapter 3.5, the (Generalized) Additive Model (GAM) will be introduced which allows the estimation of a model which should include more than one non-linear function to describe the relationship to the regressor.

The chapter will be concluded with an example in chapter 3.6 in which the previously mentioned frameworks will be used to estimate (time-) varying coefficients.

It is worth mentioning that the models can be combined and even parametric elements can be introduced such that a semi-parametric estimation can be used if preferred.

3.1 Generalized Linear Models

The extension of the linear regression model will be shown in these subsections so that non-normal response distributions and/or non-linear transformations can be modelled. McCullagh and Nelder (1989) introduced the generalized version of the linear models for which the response variable belongs to a specific family of distributions, namely the exponential family, which will be shown in the following sections.

3.1.1 Distribution of the Exponential Family

The variable Y belongs to the distribution of the exponential family if the probability density function can be written in the form

$$f(y) \quad = \quad \exp\left(\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\, \phi)\right) \tag{3.1}$$

with the location parameter θ, also called canonical parameter or natural parameter, the scale parameter ϕ, also called nuisance or dispersion parameter, and the known functions a, b, and c corresponding to the type of exponential family.

If the distribution of the random variable Y can be written in the form of equation (3.1) then the expected value is

$$\mathrm{E}\,(Y) \quad = \quad \mu = b'(\theta) \tag{3.2}$$

and the variance is

$$\mathrm{Var}\,(Y) \quad = \quad \sigma^2 = b''(\theta)\, a(\phi) \tag{3.3}$$

with b' and b'' as the first and second derivatives of the known function b.

Instead of modelling the mean μ as a linear function of the predictors $\mathbf{x}$ for the random variable Y, an one-to-one continuously differentiable, invertible transformation function g will be introduced, such that equation (3.2) can be written as

$$\mathrm{E}(Y) \;=\; \mu = g^{-1}(\eta) = g^{-1}(\mathbf{x}^T\boldsymbol{\beta}) \tag{3.4}$$

with g as link function, η as linear predictor, $\mathbf{x}$ as design vector of dimension p, and $\boldsymbol{\beta}$ as vector of the unknown parameters of dimension p which needs to be estimated. The inverse of the link function g will be defined as the response function h, i.e. $g^{-1} = h$.

It is worth mentioning that the model is assumed to be a linear model, i.e.

$$\eta \;=\; \mathbf{x}^T\boldsymbol{\beta} \tag{3.5}$$

and furthermore, one should remember that the expected value μ will be transformed (or linked) instead of the raw data y. This is important because, in general, a model where $g(y)$ is linear on x is not the same as a generalized linear model where $g(\mu)$ is linear on x. As shown above, a specific generalized linear model is fully characterized by the type of the exponential family and the link function.

3.1.2 Maximum Likelihood Estimation

An important feature of the generalized model is that the same methodology to fit the data, namely the maximum likelihood estimation technique, can be used. The likelihood function for the distribution in equation (3.1) is

$$L(\mathbf{y},\, \mathbf{x},\, \boldsymbol{\beta},\, \boldsymbol{\phi}) \;=\; \prod_{i=1}^{n} \exp\left(\frac{y_i\theta_i - b(\theta_i)}{a(\phi_i)} + c(y_i,\, \phi_i)\right) \tag{3.6}$$

with $\eta_i = \mathbf{x}_i^T \boldsymbol{\beta}$. In practice, the log-likelihood function

$$l(\mathbf{y}, \mathbf{x}, \boldsymbol{\beta}, \boldsymbol{\phi}) = \sum_{i=1}^{n} \left(\frac{y_i \theta_i - b(\theta_i)}{a(\phi_i)} + c(y_i, \phi_i) \right) \tag{3.7}$$

$$=: \sum_{i=1}^{n} l_i(\mathbf{y}, \mathbf{x}, \boldsymbol{\beta}, \boldsymbol{\phi}) \tag{3.8}$$

will be preferred, with l_i as individual likelihood function.

Remembering equation (3.4) $\theta_i = \mu^{-1}(h(\mathbf{x}_i^T \boldsymbol{\beta}))$, one can reformulate equation (3.3) as a function depending on $\boldsymbol{\beta}$

$$Var(Y_i) = \Sigma_i(\boldsymbol{\beta}) \tag{3.9}$$

as variance function of the i-th observation.

The individual score function is the first derivative of the individual log-likelihood function, i.e.

$$\mathbf{s}_i(\boldsymbol{\beta}) = \frac{\partial l_i}{\partial \boldsymbol{\beta}} \tag{3.10}$$

$$= \frac{\partial l_i}{\partial \theta_i} \frac{\partial \theta_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \boldsymbol{\beta}} \tag{3.11}$$

$$= \frac{y_i - \mu_i}{a(\phi_i)} \frac{1}{b''(\theta_i)} \frac{\partial \mu_i}{\partial \eta_i} \mathbf{x_i}^T \tag{3.12}$$

$$= (y_i - \mu_i) \Sigma_i(\boldsymbol{\beta})^{-1} \frac{\partial \mu_i}{\partial \eta_i} \mathbf{x_i}^T \tag{3.13}$$

using equations (3.9) and (3.2) to reformulate equations (3.12) to (3.13). Defining the Jacobian of the response function h as $D_i(\boldsymbol{\beta}) := \partial h(\eta_i)/\partial \eta$, the individual score function can be written as

$$\mathbf{s}_i(\boldsymbol{\beta}) = \mathbf{x_i}^T D_i(\boldsymbol{\beta}) \Sigma_i(\boldsymbol{\beta})^{-1} (y_i - \mu_i) \tag{3.14}$$

with the matrix

$$W_i(\boldsymbol{\beta}) = D_i(\boldsymbol{\beta}) \Sigma_i(\boldsymbol{\beta})^{-1} D_i(\boldsymbol{\beta})^T. \tag{3.15}$$

The Fisher information (matrix), also called expected information (matrix), is the expected value of the (negative) second derivative of the log-likelihood function. The calculation for the individual Fisher information of the i-th observation

$$\begin{aligned} I_i(\boldsymbol{\beta}) &= \mathrm{E}\left(\mathbf{s}_i(\boldsymbol{\beta})\,\mathbf{s}_i(\boldsymbol{\beta})^T\right) && (3.16)\\ &= \mathbf{x}_i^T D_i(\boldsymbol{\beta})\Sigma_i(\boldsymbol{\beta})^{-1} D_i(\boldsymbol{\beta})^T \mathbf{x}_i && (3.17)\\ &= \mathbf{x}_i^T W_i(\boldsymbol{\beta})\mathbf{x}_i && (3.18)\\ &= \mathrm{E}\left(-\frac{\partial^2 l_i(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T}\right) && (3.19) \end{aligned}$$

will be preferred for the sake of simplicity.

Defining the first derivative of μ with respect to $\boldsymbol{\beta}$ evaluated at η_i, i.e.

$$\begin{aligned} \mathbf{Z}_i(\boldsymbol{\beta}) &:= \partial\mu/\partial\boldsymbol{\beta} && (3.20)\\ &= \mathbf{x}_i^T \mathbf{D}_i(\boldsymbol{\beta}) && (3.21) \end{aligned}$$

the score function in equation (3.13) can then be written as

$$\mathbf{s}_i(\boldsymbol{\beta}) = \mathbf{Z}_i(\boldsymbol{\beta})\Sigma_i^{-1}(\boldsymbol{\beta})(y_i - \mu_i(\boldsymbol{\beta})) \quad (3.22)$$

and the Fisher information in equation (3.17) as

$$\mathbf{I}_i(\boldsymbol{\beta}) = \mathbf{Z}_i(\boldsymbol{\beta})\Sigma_i^{-1}(\boldsymbol{\beta})\mathbf{Z}_i(\boldsymbol{\beta})^T \quad (3.23)$$

which is valid for generalized linear models with the simplification in equation (3.21) as well as for the rather general case in equation (3.20) and even for the non-linear exponential family models which will be used later on.

Given the score function in equation (3.21)

$$\mathbf{s}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \mathbf{s}_i(\boldsymbol{\beta}) \tag{3.24}$$

and the Fisher information matrix in equation (3.23)

$$\mathbf{I}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \mathbf{I}_i(\boldsymbol{\beta}) \tag{3.25}$$

a Fisher scoring algorithm with the Fisher scoring iterations defined as

$$\hat{\boldsymbol{\beta}}^{(k+1)} = \hat{\boldsymbol{\beta}}^{(k)} + \mathbf{I}\left(\hat{\boldsymbol{\beta}}^{(k)}\right)^{-1} \mathbf{s}\left(\hat{\boldsymbol{\beta}}^{(k)}\right) \qquad \text{for } k = 0, 1, 2, \ldots \tag{3.26}$$

can be used, given an initial suitable parameter vector $\hat{\boldsymbol{\beta}}^{(0)}$.

3.1.3 Examples of the Exponential Family

In this subchapter several distributions which belong to the family of exponential distributions will be introduced to set the ground for the following chapters because of their suitable characteristics for describing macroeconomic data. Among the family of exponential distributions, the Gaussian (normal) one, which is for most economists the default distribution to estimate data, will be presented. In this case, the Generalized Linear Model converges to the well-known Linear Regression. In the most simple version the Ordinary Least Squares (OLS) approach is a special case of the GLM approach which will be shown below. Further examples of the Bernoulli, Poisson, and gamma distributions will be given in more detail.

For the Gaussian distribution the probability density function in equation (3.1) is characterized by $b(\theta) = 0.5\theta^2$, $a(\phi) = \phi^2$ and $c(y,\, \phi) = -0.5\log(2\pi\phi^2) - 0.5\frac{y^2}{\phi^2}$. It can be seen by applying equations (3.2) and (3.3) that the expected value is $\mu = b'(\theta) = \theta$ and the variance is $\sigma^2 = b''(\theta)\, a(\phi) = \phi^2$. For the

sake of simplicity the natural link function, namely the identity function, for the Gaussian distribution will be used such that $\mu = \eta = \mathbf{x}^T\boldsymbol{\beta}$ holds. The individual score function can be written as

$$\mathbf{s}_i(\boldsymbol{\beta}) = \frac{y_i - \mu_i}{\phi^2}\mathbf{x}_i^T \tag{3.27}$$

which can easily be seen using equation (3.14). Given equation (3.16), the Fisher information matrix can be calculated via equation (3.19) which is in this case given by

$$\mathbf{I}_i(\boldsymbol{\beta}) = \mathbf{x_i}^T\frac{1}{\phi^2}\mathbf{x}_i \qquad . \tag{3.28}$$

Assuming homoscedasticity and remembering equations (3.4) and (3.5), the Fisher scoring in equation (3.26) can be written as

$$\begin{aligned}
\hat{\boldsymbol{\beta}}^{(k+1)} &= \hat{\boldsymbol{\beta}}^{(k)} + (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T(\mathbf{y} - \boldsymbol{\mu}) && (3.29)\\
&= \hat{\boldsymbol{\beta}}^{(k)} + (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} - (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\mu} && (3.30)\\
&= \hat{\boldsymbol{\beta}}^{(k)} + (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} - (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}\hat{\boldsymbol{\beta}}^{(k)} && (3.31)\\
&= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} && (3.32)
\end{aligned}$$

which is the OLS estimator.

The probability function for the binomial distribution in the form in equation (3.1) is characterized by $\theta(\mu) = \log\left(\frac{\pi}{1-\pi}\right)$, $b(\theta) = \log(1 + \exp(\theta))$ and $c(y,\, \theta) = 0$ and $a(\phi) = 1$. The expected value is given by equation (3.2) as

$$\mathrm{E}\,(y) = \mu = b'(\theta) = \frac{\exp(\theta)}{1 + \exp(\theta)} = \pi \tag{3.33}$$

remembering that $\theta(\mu) = \log\left(\frac{\pi}{1-\pi}\right)$. The variance in equation (3.3) can be written as

$$\mathrm{Var}\,(y) = \sigma^2 = b''(\theta)a(\phi) = \frac{\exp(\theta)}{(1 + \exp(\theta))^2}\,. \tag{3.34}$$

Instead of the natural link function $\eta = \log\left(\frac{\pi}{1-\pi}\right) = g(\pi)$ or rewritten as $h(\eta) = \frac{\exp(\eta)}{1+\exp(\eta)} = \pi$, other link functions can be used. Instead of the logistic regression, often the probit model using the link function $\pi = \Phi(\eta) = h(\eta)$ with Φ as cumulative normal probability function will be used. Furthermore, the complementary log-log model using the link function $\pi = 1 - \exp(-\exp(\eta)) = h(\eta)$ which can be written as $g(\pi) = \log(-\log(1-\pi))$ can be used to model binomial data. In the case of the natural link function, i.e. the logistic regression, one gets $\mathbf{Z}_i(\boldsymbol{\beta}) = \mathbf{x}_i^T D_i(\boldsymbol{\beta})$ with $D_i(\boldsymbol{\beta}) = \frac{\partial h}{\partial \eta}(\eta_i) = \frac{\exp(\eta_i)}{(1+\exp(\eta_i))^2}$ such that the score function in equation (3.24) can be written as

$$\mathbf{s}_i(\boldsymbol{\beta}) = \mathbf{Z}_i(\boldsymbol{\beta})\boldsymbol{\Sigma}_i^{-1}(y_i - \mu_i) \tag{3.35}$$

and the Fisher information in equation (3.25)

$$\mathbf{I}_i(\boldsymbol{\beta}) = \mathbf{Z}_i(\boldsymbol{\beta})\boldsymbol{\Sigma}_i^{-1}\mathbf{Z}_i(\boldsymbol{\beta})^T\,, \tag{3.36}$$

with $\mu_i = \frac{\exp(\eta_i)}{1+\exp(\eta_i)}$. For the estimation of the unknown parameter(s) $\boldsymbol{\beta}$, the Fisher scoring method should be used with the initial parameter(s) $\boldsymbol{\beta}^{(0)}$.

The probability function for the Poisson distribution can be written in the form of equation (3.1) with $a(\phi) = 1$, $b(\theta) = \exp(\theta)$ and $c(y,\, \theta) = -\log(y!)$. Using the natural link function $\theta(\mu) = \log(\lambda)$, the expected value in equation (3.2) can be written as

$$\mathrm{E}\,(y) = \mu = b'(\theta) = \exp(\theta) = \lambda \tag{3.37}$$

and the variance in equation (3.3) can be written as

$$\mathrm{Var}\,(y) = \sigma^2 = b''(\theta)a(\phi) = \exp(\theta) = \lambda\,. \tag{3.38}$$

Instead of the natural (inverse) link function $h(\eta) = \exp(\eta)$ and $g(\mu) = \log(\mu)$ other link functions can be used. Recalling $D_i(\boldsymbol{\beta}) = \frac{\partial h}{\partial \eta}(\eta_i) = \exp(\eta_i)$ and $\mathbf{Z}_i(\boldsymbol{\beta}) = \mathbf{x}_i^T D_i(\boldsymbol{\beta})$, the score function can be written as

$$\mathbf{s}_i(\boldsymbol{\beta}) = \mathbf{Z}_i(\boldsymbol{\beta})\boldsymbol{\Sigma}_i^{-1}(y_i - \mu_i) \tag{3.39}$$

and the Fisher information as

$$\mathbf{I}_i(\boldsymbol{\beta}) = \mathbf{Z}_i(\boldsymbol{\beta})\boldsymbol{\Sigma}_i^{-1}\mathbf{Z}_i(\boldsymbol{\beta})^T\,, \tag{3.40}$$

with $\mu_i = \exp(\eta_i)$. For the Poisson regression an explicit solution is not available, such that the Fisher scoring algorithm should be applied with the initial parameter(s) $\boldsymbol{\beta}^{(0)}$.

The gamma distribution can be expressed in the form of equation (3.1) setting $\theta(\mu) = -\mu^{-1}$, $a(\phi) = \phi^{-1}$, $b(\theta) = -\log(\theta)$ and $c(y,\, \theta) = \phi^{-1}\log(\frac{y}{eta}) - \log(y\Gamma(\phi^{-1}))$. The natural (inverse) link function will be used, such that $h(\eta) = -\frac{1}{\eta}$ and $g(\mu) = -\frac{1}{\mu}$. The expected value given in equation (3.2) can be written as

$$\mathrm{E}\,(y) = \mu = b'(\theta) = -\frac{1}{\theta} = \mu \tag{3.41}$$

and the variance in equation (3.3) can be written as

$$\mathrm{Var}\,(y) = \sigma^2 = b''(\theta)a(\phi) = \frac{1}{\theta^2}\frac{1}{\phi} = \frac{\mu^2}{\phi}\,. \tag{3.42}$$

Setting $D_i(\boldsymbol{\beta}) = \frac{\partial h}{\partial \eta}(\eta_i) = \frac{1}{\eta_i^2}$ and $\mathbf{Z}_i(\boldsymbol{\beta}) = \mathbf{x}_i^T D_i(\boldsymbol{\beta})$, the score function can be written as

$$\mathbf{s}_i(\boldsymbol{\beta}) = \mathbf{Z}_i(\boldsymbol{\beta})\boldsymbol{\Sigma}_i^{-1}(y_i - \mu_i) \tag{3.43}$$

and the Fisher information as

$$\mathbf{I}_i(\boldsymbol{\beta}) = \mathbf{Z}_i(\boldsymbol{\beta})\boldsymbol{\Sigma}_i^{-1}\mathbf{Z}_i(\boldsymbol{\beta})^T\,, \tag{3.44}$$

with $\mu_i = \frac{1}{\eta_i}$. For the gamma distribution no explicit solution is available, such that the Fisher scoring algorithm is the preferred choice to estimate the unknown parameter(s) $\boldsymbol{\beta}$, starting with the initial parameter(s) $\boldsymbol{\beta}^{(0)}$.

Besides the four distributions introduced in this subchapter, more distributions belong to the family of exponential distributions which will not be discussed in this work explicitly. But, given the probability density function in equation (3.1), the maximum likelihood estimation can easily be applied given the above guidance.

3.2 Penalized Splines

The penalized spline regression technique is a regression without assuming a specific functional form linking the explanatory variable with the dependent variable(s). However, it will be assumed that the functional form is a "smooth" function. The penalized spline regression technique is often referred to as a non-parametric or non-linear regression, although it can be treated as an over-parameterized regression due to numerous parameters. Whether the solution of a penalized spline regression is linear or non-linear depends on the model setup, such that it results in a linear or a non-linear regression. For more details about non-parametric smoothing techniques and in particular penalized splines, see O'Sullivan (1988), Silverman (1985), Eubank (1989), Parker and Rice (1985), Hastie and Tibshirani (1990), Eilers and Marx (1996), Ruppert, Wand, and Carroll (2003), or Ruppert, Wand, and Carroll (2009) for an extensive introduction.

For the sake of simplicity, the univariate case will be discussed in this subsection, where the observation pairs $(x_i,\, y_i)$ for $i = 1, \ldots, n$ are given and it is assumed that one can write

$$\mathrm{E}\,(y|\, x) = f(x) \tag{3.45}$$

with f as an unknown, but smooth function. For the univariate, linear regression of first order, the function f is assumed to be

$$f(x) = \beta_0 + \beta_1 x \tag{3.46}$$

which would imply the corresponding basis

$$\mathbf{X} = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} . \tag{3.47}$$

For the spline regression a sequence of knots over the support of x are defined, namely $\kappa_1 < \ldots < \kappa_K$ where the number of knots is relatively small compared to the number of observations; in most cases $K = \max(0.2n, 40)$ is a suitable choice. The corresponding basis for a (truncated) linear spline basis is defined as

$$\mathbf{X} := \begin{pmatrix} 1 & x_1 & (x_1 - \kappa_1)_+ & \ldots & (x_1 - \kappa_K)_+ \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_n & (x_n - \kappa_1)_+ & \ldots & (x_n - \kappa_K)_+ \end{pmatrix} \tag{3.48}$$

with the operator $(\cdot)_+ = \max\{\cdot, 0\}$ as linear spline basis function. In this case the function f would be described by linear functions "knot" together at each knot point $\kappa_.$. More generally, the corresponding basis of a (truncated) linear spline basis of q-th order is defined as

$$\mathbf{X} := \begin{pmatrix} 1 & x_1 & \ldots & x_1^q & (x_1 - \kappa_1)_+^q & \ldots & (x_1 - \kappa_K)_+^q \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 1 & x_n & \ldots & x_n^q & (x_n - \kappa_1)_+^q & \ldots & (x_n - \kappa_K)_+^q \end{pmatrix} \tag{3.49}$$

for $q = 0, 1, ...$ and with the operator $(\cdot)_+^q := \max\{\cdot, 0\}^q$. It is assumed that the unknown function can be described given the spline basis, i.e.

$$f(x) = \mathbf{X}\boldsymbol{\beta} \tag{3.50}$$

such that the unknown parameters $\boldsymbol{\beta}$ need to be estimated. Using a simple, unconstrained estimation, the resulting estimated function $\hat{f}$ would be heavily overfitted, meaning that the estimated function is too flexible and

responds to small changes, which might appear to be random. For this reason, constraints on the estimation will be used to circumvent the problem such that instead of the unconstrained likelihood function a constrained (or latter called penalized) likelihood function will be maximized. The unknown function f will be estimated solving the problem

$$\min_{\boldsymbol{\beta}} -\sum_{i=1}^{n} l_i(y_i, \boldsymbol{\beta}) \tag{3.51}$$

under the constraint

$$\boldsymbol{\beta}^T \mathbf{D} \boldsymbol{\beta} \leq c \tag{3.52}$$

with a penalty matrix $\mathbf{D}$ and an arbitrary constant c. Or, rewritten as Lagrange function

$$-\sum_{i=1}^{n} l_i(y_i, \boldsymbol{\beta}) + \frac{1}{2}\lambda \left(\boldsymbol{\beta}^T \mathbf{D} \boldsymbol{\beta} - c\right) \tag{3.53}$$

with an arbitrary constant c and λ as Lagrange multiplier. Ruppert, Wand, and Carroll (2003) show that instead of solving the Lagrange function (3.53), the penalized likelihood function

$$\sum_{i=1}^{n} l_i(y_i, \boldsymbol{\beta}) - \frac{1}{2}\lambda \boldsymbol{\beta}^T \mathbf{D} \boldsymbol{\beta} \tag{3.54}$$

can be maximized with $\lambda \geq 0$. The Lagrange multiplier λ in equation (3.54) will be called a smoothing parameter, because the amount of smoothness will be determined by the parameter, similar to the constant c in the constraint shown in equation (3.52). For $\lambda = 0$, the estimation is unconstrained and leads to the most wiggly estimation. For $\lambda \to \infty$, the effect of the spline basis vanishes and depending on the design matrix $\mathbf{X}$ and the penalty matrix $\mathbf{D}$, a less complex solution will be achieved.

For the case of Gaussian response data, the solution of the problem in equation (3.54) is given by

$$\hat{\boldsymbol{\beta}}(\lambda) = \left(\mathbf{X}^T\mathbf{X} + \lambda\mathbf{D}\right)^{-1}\mathbf{X}^T\mathbf{y} \tag{3.55}$$

and the unknown function is estimated by

$$\hat{f}_\lambda = \mathbf{X}\hat{\boldsymbol{\beta}}(\lambda) = \mathbf{X}\left(\mathbf{X}^T\mathbf{X} + \lambda\mathbf{D}\right)^{-1}\mathbf{X}^T\mathbf{y}\,. \tag{3.56}$$

It is worth noting, that in this case the regression type is a linear regression due to the linear relationship of the estimated response and the observed response variable, i.e.

$$\hat{\mathbf{y}} = \mathbf{P}_\lambda\mathbf{y} \tag{3.57}$$

with

$$\mathbf{P}_\lambda = \mathbf{X}\left(\mathbf{X}^T\mathbf{X} + \lambda\mathbf{D}\right)^{-1}\mathbf{X}^T \tag{3.58}$$

as projection matrix, also sometimes called hat-matrix or smoother matrix.

For unconstrained regression techniques, the complexity of the used model can be measured by the degrees of freedom which are given by the trace of the projection matrix, and is in general the number of used parameters. The generalized version

$$\begin{aligned}\operatorname{tr}(\mathbf{P}_\lambda) &= \operatorname{tr}\left(\mathbf{X}\left(\mathbf{X}^T\mathbf{X} + \lambda\mathbf{D}\right)^{-1}\mathbf{X}^T\right)\\ &= \operatorname{tr}\left(\mathbf{X}^T\mathbf{X}\left(\mathbf{X}^T\mathbf{X} + \lambda\mathbf{D}\right)^{-1}\right)\\ &=: df(\lambda)\end{aligned} \tag{3.59}$$

will give the equivalent number of parameters, also called equivalent degrees of freedom or estimated degrees of freedom, which is a positive, finite real number.

For one of the various different suitable choices for a penalty matrix, here a (truncated) linear spline basis of q-th order will be defined in equation (3.49), the penalty matrix is given by

$$\mathbf{D} = \mathbf{0}_{q+1\times q+1} \oplus \mathbf{I}_k \tag{3.60}$$

with $\mathbf{0}_{n\times m}$ as $n \times m$ matrix which contains only zeros, $\mathbf{I}_d$ as $d \times d$ identity matrix, the operator defined as $\mathbf{A}_{n\times m} \oplus \mathbf{B}_{p\times q} := \begin{pmatrix} \mathbf{A} & \mathbf{0}_{n\times q} \\ \mathbf{0}_{p\times m} & \mathbf{B} \end{pmatrix}$. In this case, the penalizing term in equation (3.52) can be written as

$$\boldsymbol{\beta}^T \mathbf{D} \boldsymbol{\beta} = \beta_{q+1}^2 + \ldots + \beta_{q+K}^2 \,. \tag{3.61}$$

For this model, the degrees of freedom vary between $q+1$ for $\lambda \to \infty$ and $q+1+K$ for $\lambda = 0$. In the case of $\lambda \to \infty$, the model converges to

$$f(x) = \beta_0 + \beta_1 x + \ldots + \beta_q x^q \tag{3.62}$$

which is a parametric polynomial regression model of q-th order.

3.3 B-Splines

Eilers and Marx (1996) propose a so-called B-splines basis and a suitable penalty matrix where the order of the penalty can be set to suit the desire of the user. de Boor (1978) illustrates the construction of a (general) B-spline basis in more depth. Eilers and Marx (1996) show that a B-spline basis of q-th order consists of $q+1$ polynomial pieces each of degree q which join at q inner knots. At these joining points, up to order $q-1$, the derivatives are continuous. B-splines are positive on a domain spanned by $q+2$ knots and are zero everywhere else. Except at the boundaries, the B-spline overlaps with $2q$ polynomial pieces of its neighbors. At any given point x there are $q+1$ B-splines non-zero.

Given the knots $\kappa_1 < \ldots < \kappa_K$ over the support of the variable x, the B-splines of order $q = 0$ are defined as

$$B_{j,\,0}(x) = \begin{cases} 1 & \text{if } \kappa_j \leq x < \kappa_{j+1} \\ 0 & \text{else} \end{cases} \tag{3.63}$$

and for higher orders $q \in \mathbb{N}$ recursively as

$$B_{j,\,q}(x) = \frac{x - \kappa_j}{\kappa_{j+q} - \kappa_j} B_{j,\,q-1}(x) + \frac{\kappa_{j+q+1} - x}{\kappa_{j+q+1} - \kappa_{j+1}} B_{j+1,\,q-1}(x) \tag{3.64}$$

which are shown for the order $q = 0, 1, 2,$ and 3 in figure 3.1.

Given the same sequence of knots, the truncated spline basis $\mathbf{X}$ of q-th order defined in equation (3.49) can be constructed by the corresponding B-spline basis of q-th order which is set to $\mathbf{B} := (B_{j,\,q})_{j=1,\ldots,K}$. This means, that there exists a matrix $\mathbf{L}_{X,\,B}$ so that $\mathbf{X} = \mathbf{L}_{X,\,B}\mathbf{B}$ holds. Furthermore, there exists a matrix $\mathbf{L}_{B,\,X}$ so that $\mathbf{B} = \mathbf{L}_{B,\,X}\mathbf{X}$ holds.

Given the parameters $\boldsymbol{\theta} := (\theta_1, \ldots, \theta_K)^T \in \mathbb{R}^K$, the function f will be defined as weighted linear combination of a q-th order B-spline basis, i.e.

$$f(x) := \sum_{j=1}^{K} B_{j,\,q}(x)\theta_j \,. \tag{3.65}$$

Eilers and Marx (1996) propose a generalized p-th order penalty of the form

$$\sum_{j=p+1}^{K} (\Delta^p \theta_j)^2 \tag{3.66}$$

with the difference operator Δ^p of p-th order. Defining the matrix $\mathbf{L}_p$ as matrix representation of the difference operator then in matrix-vector writing the penalty in equation (3.67) can be written as

$$\boldsymbol{\theta}^T \mathbf{L}_p^T \mathbf{L}_p \boldsymbol{\theta} = \boldsymbol{\theta}^T \mathbf{D}_p \boldsymbol{\theta} \tag{3.67}$$

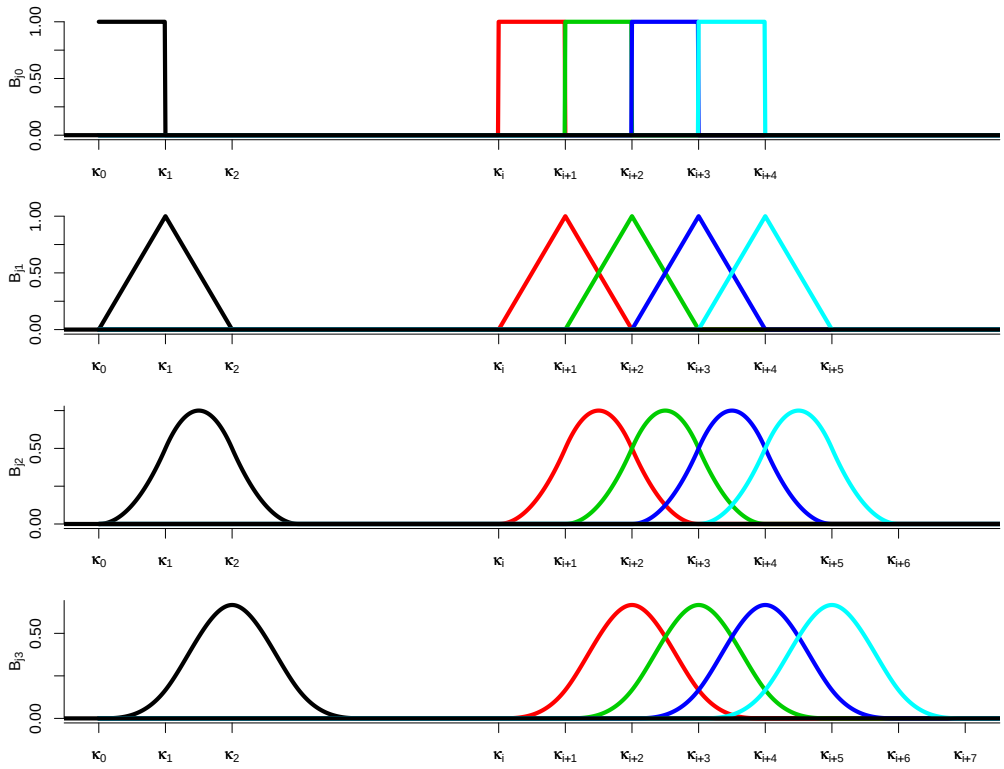

Figure 3.1: Isolated and overlapping B-splines for different degrees constructed with a given set of knots: First row: Degree zero, second row: Degree one, third row: Degree two, Fourth Row: Degree three. Each colored line (black, red, green, blue, and light blue) represents a different B-spline.

---

with penalty matrix $\mathbf{D}_p := \mathbf{L}_p^T \mathbf{L}_p$.

Maximizing the penalized likelihood similar to (3.54), the fitted function f converges to a polynomial of degree $p-1$ if the smoothing parameter is large, i.e. $\lambda \to \infty$, and if the B-spline basis is at least of order p, i.e. $q \geq p$.

For the selection of the optimal amount of smoothness, the smoothing pa-

rameter should minimize the Akaike Information Criterion (AIC) as proposed by Eilers and Marx (1996). The AIC originally proposed by Akaike (1974) trades off the "goodness of fit" against the complexity of a model measured by the degrees of freedom. In the general form, the AIC is defined as

$$\mathrm{AIC}(\lambda) := -2\sum_{i=1}^{n} l_i(y_i, \boldsymbol{\theta}, \lambda) + 2df(\lambda) \tag{3.68}$$

with $\boldsymbol{\theta}$ as parameters which maximize the (penalized) likelihood in equation (3.54) given the smoothing parameter λ and $df(\lambda)$ as corresponding degrees of freedom, defined in equation (3.59). For the case of Gaussian data, equation (3.68) can be written as

$$\mathrm{AIC}(\lambda) := \log(\mathrm{RSS}(\lambda)) + 2\frac{df(\lambda)}{n} \tag{3.69}$$

with n as number of observations and RSS as residual sum of squares, i.e. $RSS(\lambda) := \sum_{i=1}^{n} \left(y_i - \hat{f}_\lambda(x_i)\right)^2$.

But, the selection can also be based on cross-validation for which the i-th observation will be left out to estimate the response of the i-th observation, i.e.

$$\mathrm{CV}(\lambda) := \frac{1}{n}\sum_{i=1}^{n} \left(y_i - \hat{f}_\lambda^{-i}(x_i)\right)^2 \tag{3.70}$$

which can be written in a simpler form if the regression is linear; see equation (3.57). In the case of a linear relationship, i.e. $\hat{f}(x_i) = \sum_j p_{ij} y_j$, with the projection matrix $\mathbf{P} = (p_{ij})_{i,j=1,\ldots,n}$, one may simplify

$$\hat{f}_\lambda^{-i}(x_i) = \sum_{j\neq i} \frac{p_{ij}}{1-p_{ii}} y_j \tag{3.71}$$

$$= \frac{1}{1-p_{ii}}\hat{f}_\lambda(x_i) - \frac{p_{ii}}{1-p_{ii}} y_i \tag{3.72}$$

such that the cross-validation criterion in equation (3.70) can be written as

$$\mathrm{CV}(\lambda) := \frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i - \hat{f}_\lambda(x_i)}{1-p_{ii}}\right)^2 . \tag{3.73}$$

Craven and Wahba (1979) replaced p_{ii} by the average, i.e. $\sum_{i=1}^{n} p_{ii}/n = \mathrm{tr}\,(\mathbf{P})/n = df(\lambda)/n$ such that the generalized cross-validation will be defined as

$$\mathrm{GCV}(\lambda) := \frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i - \hat{f}_\lambda(x_i)}{1-df(\lambda)/n}\right)^2 . \tag{3.74}$$

Unfortunately, a general recommendation of which criterion should be preferred is not possible. Usually, the AIC, CV, and GCV methods lead to different "optimal" smoothing parameters, so one should not trust the automatic smoothing parameter selectors blindly.

3.4 Random Effects and (Generalized) Linear Mixed Models

Laird and Ware (1982) extended the linear model for Gaussian data and introduced random effects which vary by group. An illustrative example and introduction is given by Wand (2003), where the random effects have been introduced to analyze group-specific intercepts. Furthermore, random effects can be used to allow for group-specific covariates effects, which would be interpreted as group-specific slopes.

The response variable y_{ij} is the j-th out of n_i observations for the i-th out of m groups. The data set might either be balanced, i.e. $n_1 = \ldots = n_m$, or unbalanced if there are at least two groups i_1 and i_2 for which $n_{i_1} \neq n_{i_2}$ holds. For each observation the p fixed effect regressors $x_{ij1}, \ldots, x_{ijp}$ are given which usually include the intercept, i.e. $x_{ij1} \equiv 1 \;\forall\; i, j$. The unknown

fixed-effect coefficients are identical for each group, i.e. $\beta_1, \ldots, \beta_p$ do not depend on i. The random-effect regressors will be denoted as $z_{ij1}, \ldots, z_{ijq}$ and the corresponding random-coefficients $u_{i1}, \ldots, u_{iq}$ have to be predicted for all groups $i = 1, \ldots, m$. Usually, the random effects vary by group and are unlike the fixed-effects parameters $\beta_1, \ldots, \beta_p$ no parameters which need to be estimated, but random variables which need to be predicted.

The (linear) mixed effect model, which consists of fixed and random effects, can be written as

$$y_{ij} = x_{ij1}\beta_1 + \ldots + x_{ijp}\beta_p + z_{ij1}u_{i1} + \ldots + z_{ijq}u_{iq} + \epsilon_{ij}\,. \tag{3.75}$$

Defining

$$\mathbf{y} = \begin{pmatrix} y_{11} \\ \vdots \\ y_{1n_1} \\ y_{21} \\ \vdots \\ y_{2n_2} \\ \vdots \\ y_{m1} \\ \vdots \\ y_{mn_m} \end{pmatrix}, \boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_{11} \\ \vdots \\ \epsilon_{1n_1} \\ \epsilon_{21} \\ \vdots \\ \epsilon_{2n_2} \\ \vdots \\ \epsilon_{m1} \\ \vdots \\ \epsilon_{mn_m} \end{pmatrix}, \mathbf{X} = \begin{pmatrix} x_{111} & \ldots & x_{11p} \\ \vdots & & \vdots \\ x_{1n_11} & \ldots & x_{1n_1p} \\ x_{211} & \ldots & x_{21p} \\ \vdots & & \vdots \\ x_{2n_21} & \ldots & x_{2n_2p} \\ \vdots & & \vdots \\ x_{m11} & \ldots & x_{m1p} \\ \vdots & & \vdots \\ x_{mn_m1} & \ldots & x_{mn_mp} \end{pmatrix}, \mathbf{u} = \begin{pmatrix} u_{11} \\ \vdots \\ u_{1q} \\ u_{21} \\ \vdots \\ u_{2q} \\ \vdots \\ u_{m1} \\ \vdots \\ u_{mq} \end{pmatrix},$$

$$\mathbf{Z}_i = \begin{pmatrix} x_{i11} & \ldots & x_{i1q} \\ \vdots & & \vdots \\ x_{in_i1} & \ldots & x_{in_iq} \end{pmatrix} \forall\, i = 1, \ldots, m \text{ and } \mathbf{Z} = \mathbf{Z}_1 \oplus \ldots \oplus \mathbf{Z}_m$$

with the operator $\mathbf{A}_{n\times m} \oplus \mathbf{B}_{p\times q} := \begin{pmatrix} \mathbf{A} & \mathbf{0}_{n\times q} \\ \mathbf{0}_{p\times m} & \mathbf{B} \end{pmatrix}$, the equation (3.75) can be written in matrix-vector form as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\epsilon} \tag{3.76}$$

with

$$\begin{pmatrix} \mathbf{u} \\ \boldsymbol{\epsilon} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Psi} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma} \end{pmatrix} \right) \tag{3.77}$$

with $\boldsymbol{\Psi}$ as (co-)variance matrix for the random effects and $\boldsymbol{\Sigma}$ as (co-)variance matrix for the residuals.

To estimate the mixed effect model, the unknown parameters $\boldsymbol{\beta}$ and the (co-)variance matrices $\boldsymbol{\Psi}$ and $\boldsymbol{\Sigma}$ need to be estimated using the Restricted Maximum Likelihood regression technique (REML) and the random effects $\mathbf{u}$ will be predicted using best prediction (BP), i.e. $\hat{\mathbf{u}} = \mathrm{E}(\mathbf{u}|\,\mathbf{y})$.

The (log) likelihood for the model in equation (3.77) is given by

$$l(\boldsymbol{\beta},\, \boldsymbol{\Psi},\, \boldsymbol{\Sigma}) = -\frac{1}{2}\left[n\log(2\pi) + \log|\mathbf{V}| + (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right] \tag{3.78}$$

with the (co-)variance matrix

$$\mathbf{V} = \mathrm{Cov}(\mathbf{y}) = \mathbf{Z}\boldsymbol{\Psi}\mathbf{Z}^T + \boldsymbol{\Sigma} \tag{3.79}$$

remembering equations (3.76) and (3.77).

Solving equation (3.78) for $\boldsymbol{\beta}$ the best linear unbiased estimator (BLUE) is given by

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{y} \tag{3.80}$$

which is the standard maximum likelihood estimator.

Plugging equation (3.80) into the likelihood function in equation (3.78) gives the profile (log) likelihood for $\mathbf{V}$, i.e.

$$l_P(\mathbf{V}) = -\frac{1}{2}\left[n\log(2\pi) + \log|\mathbf{V}| + \mathbf{y}^T\mathbf{V}^{-1}(\mathbf{I} - \mathbf{X}(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1})\mathbf{y}\right]$$

for which the maximum likelihood estimation will give $\boldsymbol{\Psi}$ and $\boldsymbol{\Sigma}$ and thus $\mathbf{V}$. However, Searle, Casella, and McCulloch (1992) prefer the restricted (log) likelihood, which can be written as

$$l_R(\mathbf{V}) = l_P(\mathbf{V}) - \frac{1}{2}\log\left|\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}\right| \tag{3.81}$$

to account for the degrees of freedom for the fixed effects in the model, which is more accurate for small sample sizes. Given $\boldsymbol{\beta}$, $\boldsymbol{\Psi}$, and $\boldsymbol{\Sigma}$, the best linear predictor (BLP) for the random effects is given by

$$\tilde{\mathbf{u}} = \boldsymbol{\Psi}\mathbf{Z}^T\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\,. \tag{3.82}$$

Alternatively, defining the combined design matrix $\mathbf{C} = \begin{pmatrix}\mathbf{X} & \mathbf{Z}\end{pmatrix}$ and the matrix $\mathbf{D} = \mathbf{0} \oplus \boldsymbol{\Psi}$, the BLUE can be written as

$$\begin{pmatrix}\tilde{\boldsymbol{\beta}} \\ \tilde{\mathbf{u}}\end{pmatrix} = \left(\mathbf{C}^T\boldsymbol{\Sigma}^{-1}\mathbf{C} + \mathbf{D}\right)^{-1}\mathbf{C}^T\boldsymbol{\Sigma}^{-1}\mathbf{y} \tag{3.83}$$

which is from a notational point of view similar to equation (3.55) in which the (co-)variance matrix is set to the identity matrix.

Using equation (3.83), the covariance of the predictions is given as

$$\mathrm{Cov}\left(\begin{pmatrix}\tilde{\boldsymbol{\beta}} \\ \tilde{\mathbf{u}}\end{pmatrix}\right) = \left(\mathbf{C}^T\boldsymbol{\Sigma}^{-1}\mathbf{C} + \mathbf{D}\right)^{-1} \tag{3.84}$$

and the conditional covariance matrix as

$$\mathrm{Cov}\left(\begin{pmatrix}\tilde{\boldsymbol{\beta}} \\ \tilde{\mathbf{u}}\end{pmatrix} \mid \mathbf{u}\right) = \left(\mathbf{C}^T\boldsymbol{\Sigma}^{-1}\mathbf{C} + \mathbf{D}\right)^{-1}\mathbf{C}^T\boldsymbol{\Sigma}^{-1}\mathbf{C}\left(\mathbf{C}^T\boldsymbol{\Sigma}^{-1}\mathbf{C} + \mathbf{D}\right)^{-1}\,. \tag{3.85}$$

The main advantage of the mixed model solution is to be able to treat a spline model as a mixed model. Rewriting the model from subsection 3.2 for the Gaussian case as (penalized) truncated spline basis of first order as

$$y_i = \beta_0 + x_i\beta_1 + \sum_{k=1}^{K}(x_i - \kappa_k)_+ u_k + \epsilon_i \tag{3.86}$$

with the residuals $\epsilon_i \sim \mathcal{N}(0, \sigma_\epsilon^2)$ and the coefficients are assumed to be normally distributed, i.e. $u_k \sim \mathcal{N}(0, \sigma_u^2)$, then equation (3.86) can be written in matrix-vector form as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\epsilon} \tag{3.87}$$

with

$$\begin{pmatrix}\mathbf{u}\\ \boldsymbol{\epsilon}\end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix}\mathbf{0}\\ \mathbf{0}\end{pmatrix}, \begin{pmatrix}\sigma_u^2\mathbf{I} & \mathbf{0}\\ \mathbf{0} & \sigma_\epsilon^2\mathbf{I}\end{pmatrix}\right). \tag{3.88}$$

For the model above in equation (3.87) and (3.88), the covariance matrices are given by $\boldsymbol{\Psi} = \sigma_u^2\mathbf{I}$ and $\boldsymbol{\Sigma} = \sigma_\epsilon^2\mathbf{I}$. The amount of smoothness will be determined by $\lambda = \frac{\sigma_\epsilon^2}{\sigma_u^2}$.

Solving the model in equations (3.87) and (3.88) via maximum likelihood, is equivalent to the penalized spline approach; see

$$\begin{pmatrix}\hat{\boldsymbol{\beta}}\\ \hat{\mathbf{u}}\end{pmatrix} = \arg\min_{\boldsymbol{\beta},\,\mathbf{u}} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u}) + \lambda\mathbf{u}^T\mathbf{u} \tag{3.89}$$

$$= \left(\mathbf{C}^T\mathbf{C} + \lambda\mathbf{D}\right)^{-1}\mathbf{C}^T\mathbf{y} \tag{3.90}$$

with the combined design matrix $\mathbf{C} = \begin{pmatrix}\mathbf{X} & \mathbf{Z}\end{pmatrix}$ and the penalty matrix $\mathbf{D} = \mathbf{0} \oplus \mathbf{I}$.

The extended generalized linear mixed model (GLMM) will be used if the response y belongs to the distribution of exponential family such that the conditional density is given by

$$f(\mathbf{y}|\,\mathbf{u}) \;=\; \exp\left(\mathbf{y}^T(\mathbf{X}\boldsymbol{\beta}+\mathbf{Z}\mathbf{u}) - \mathbf{1}^T b(\mathbf{X}\boldsymbol{\beta}+\mathbf{Z}\mathbf{u}) + \mathbf{1}^T c(\mathbf{y})\right) \quad (3.91)$$

and the random effects are assumed to still be normally distributed, i.e.

$$\mathbf{u} \sim \mathcal{N}(\mathbf{0},\, \boldsymbol{\Psi})\,. \quad (3.92)$$

Similar to the GLM, the linear predictor is given by

$$\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} \quad (3.93)$$

and the the conditional expected value is given by

$$\mathrm{E}\,(\mathbf{y}|\,\mathbf{u}) = \boldsymbol{\mu} = h(\boldsymbol{\eta}) \quad (3.94)$$

given a known link function h.

From a practical point of view, the calculation of $\boldsymbol{\beta}$, $\boldsymbol{\Psi}$, and $\boldsymbol{\Sigma}$ involves the integration over $\mathbb{R}^q$ for the random-effects coefficients $\mathbf{u}$ such that a Laplace approximation is needed; see McCulloch and Searle (2001). The simultaneous calculation of the parameters $\boldsymbol{\beta}$, $\boldsymbol{\Psi}$, $\boldsymbol{\Sigma}$, and random effects $\mathbf{u}$ are based on posterior modes from a Bayesian view. In this case, the estimation equation for the fixed and random effects is given by

$$\begin{pmatrix} \mathbf{X}^T\mathbf{W}\mathbf{X} & \mathbf{X}^T\mathbf{W}\mathbf{Z} \\ \mathbf{Z}^T\mathbf{W}\mathbf{X} & \mathbf{Z}^T\mathbf{W}\mathbf{Z}+\boldsymbol{\Psi}^{-1} \end{pmatrix} \begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \end{pmatrix} = \begin{pmatrix} \mathbf{X}^T\mathbf{H}^T\boldsymbol{\Sigma}^{-1}\,(\mathbf{y}-\boldsymbol{\mu}+\mathbf{H}\boldsymbol{\eta}) \\ \mathbf{Z}^T\mathbf{H}^T\boldsymbol{\Sigma}^{-1}\,(\mathbf{y}-\boldsymbol{\mu}+\mathbf{H}\boldsymbol{\eta}) \end{pmatrix} \quad (3.95)$$

with the definitions $\mathbf{H} = \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\eta}^T}$ and $\mathbf{W} = \mathbf{H}^T\boldsymbol{\Sigma}^{-1}\mathbf{H}$. The vector $\left(\hat{\boldsymbol{\beta}}^T \quad \hat{\mathbf{u}}^T\right)^T$ will be calculated iteratively and for the general case no close-form solution can be provided.

3.5 Generalized Additive Models

In subsections 3.2 and 3.3, the response variable was explained by just one single variable. In this subsection, a multi-dimensional approach, known as (Generalized) Additive Model (GAM), will be introduced. For the sake of simplicity only two variables and a Gaussian response will be assumed, but the extension with more variables is straight forward given the guidance below and a generalization for response variables which belong to the family of exponential distributions can easily be done if one recalls the previous subsections.

Similar to equation (3.45), the expected value is assumed to be given by

$$\mathrm{E}\left(y_i\right) = \beta_0 + f_1(x_{i1}) + f_2(x_{i2}) \tag{3.96}$$

with the unknown functions f_1 and f_2. Analog to the previous one-dimensional version, the functions f_1 and f_2 are assumed to be smooth. Furthermore, assume that the unknown function can be modeled by (truncated) linear splines of first order, i.e.

$$f_1(x) = \beta_1 x + \sum_{k=1}^{K_1} (x - \kappa_{k,1})_+ u_{k,1} \tag{3.97}$$

and

$$f_2(x) = \beta_2 x + \sum_{k=1}^{K_2} (x - \kappa_{k,2})_+ u_{k,2} \tag{3.98}$$

but higher orders or a different spline basis can be used. The knots for the first and second variables are $\kappa_{1,1} < \ldots < \kappa_{K_1,1}$ and $\kappa_{1,2} < \ldots < \kappa_{K_2,2}$, respectively, which can be selected independently and the number of knots can differ, i.e. $K_1 \neq K_2$ is allowed.

Using the equations (3.97) and (3.98), the equation (3.96) can be written in matrix-vector form as

$$\mathrm{E}\left(\mathbf{y}\right) = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_1\mathbf{u}_1 + \mathbf{Z}_2\mathbf{u}_2 \tag{3.99}$$

with $\mathbf{X} = \begin{pmatrix} 1 & x_{i1} & x_{i2} \end{pmatrix}_{i=1,\ldots,n}$, $\boldsymbol{\beta} = \begin{pmatrix} \beta_0 & \beta_1 & \beta_2 \end{pmatrix}^T$, $\mathbf{u}_j = \begin{pmatrix} u_{1,j} & \ldots & u_{K_j,j} \end{pmatrix}^T$ and $\mathbf{Z}_j = \begin{pmatrix} (x_{ij} - \kappa_{1,j})_+ & \ldots & (x_{ij} - \kappa_{K_j,j})_+ \end{pmatrix}_{i=1,\ldots,n}$ for $j = 1, 2$. Defining $\mathbf{Z} = \begin{pmatrix} \mathbf{Z}_1 & \mathbf{Z}_2 \end{pmatrix}$, $\mathbf{u} = \begin{pmatrix} \mathbf{u}_1^T & \mathbf{u}_2^T \end{pmatrix}^T$, the equation (3.99) can be written in the well-known form

$$\mathrm{E}\left(\mathbf{y}\right) = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} \tag{3.100}$$

which was already analyzed in the previous subsections.

Assuming furthermore

$$\mathrm{E}\left(\begin{pmatrix} \mathbf{u} \\ \boldsymbol{\epsilon} \end{pmatrix}\right) = \mathbf{0} \tag{3.101}$$

and

$$\mathrm{Cov}\left(\begin{pmatrix} \mathbf{u} \\ \boldsymbol{\epsilon} \end{pmatrix}\right) = \begin{pmatrix} \sigma_{u,1}^2\mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \sigma_{u,2}^2\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \sigma_{\epsilon}^2\mathbf{I} \end{pmatrix}, \tag{3.102}$$

the maximum likelihood estimation is given by

$$\hat{\mathbf{y}} = \mathbf{C}\left(\mathbf{C}^T\mathbf{C} + \mathbf{D}(\boldsymbol{\lambda})\right)^{-1}\mathbf{C}^T\mathbf{y} \tag{3.103}$$

for the Gaussian case with $\mathbf{C} = \begin{pmatrix} \mathbf{X} & \mathbf{Z} \end{pmatrix}$, $\mathbf{D}(\boldsymbol{\lambda}) = \lambda_1\mathbf{D}_1 + \lambda_2\mathbf{D}_2$, $\mathbf{D}_1 = \mathbf{0} \oplus \mathbf{I} \oplus \mathbf{0}$, $\mathbf{D}_2 = \mathbf{0} \oplus \mathbf{0} \oplus \mathbf{I}$, and λ_1 and λ_2 as smoothing parameters for the functions f_1 and f_2, respectively.

Interpreting the model as a mixed model, the smoothing parameters will be given by $\lambda_1 = \frac{\sigma_\epsilon^2}{\sigma_{u,1}^2}$ and $\lambda_2 = \frac{\sigma_\epsilon^2}{\sigma_{u,2}^2}$ such that the degrees of freedom for the REML estimation are needed. Similar to the previous subsections the total degrees of freedom, recalling equation (3.103), are given by

$$\begin{aligned} df_{fit} &= \operatorname{tr}\left(\mathbf{C}\left(\mathbf{C}^T\mathbf{C} + \mathbf{D}(\boldsymbol{\lambda})\right)^{-1}\mathbf{C}^T\right) && (3.104)\\ &= \operatorname{tr}\left(\left(\mathbf{C}^T\mathbf{C} + \mathbf{D}(\boldsymbol{\lambda})\right)^{-1}\mathbf{C}^T\mathbf{C}\right) && (3.105)\end{aligned}$$

and for this model, the degrees of freedom for the function f_j are given by

$$df_j = \operatorname{tr}\left(\mathbf{D}_j\left(\mathbf{C}^T\mathbf{C} + \mathbf{D}(\boldsymbol{\lambda})\right)^{-1}\mathbf{C}^T\mathbf{C}\right) + 1 \qquad (3.106)$$

for $j = 1, 2$ and $df_0 = 1$ accounting for the intercept. Remember that the degrees of freedom are additive, i.e. $df_{fit} = df_0 + df_1 + df_2$.

Using equation (3.84), the covariance matrix can be written as

$$\operatorname{Cov}\left(\begin{pmatrix}\hat{\boldsymbol{\beta}}\\ \hat{\mathbf{u}} - \mathbf{u}\end{pmatrix}\right) = \sigma_\epsilon^2\left(\mathbf{C}^T\mathbf{C} + \mathbf{D}(\boldsymbol{\lambda})\right)^{-1} \qquad (3.107)$$

and using equation (3.85), the conditional covariance matrix can be written as

$$\operatorname{Cov}\left(\begin{pmatrix}\hat{\boldsymbol{\beta}}\\ \hat{\mathbf{u}}\end{pmatrix} \mid \mathbf{u}\right) = \sigma_\epsilon^2\left(\mathbf{C}^T\mathbf{C} + \mathbf{D}(\boldsymbol{\lambda})\right)^{-1}\mathbf{C}^T\mathbf{C}\left(\mathbf{C}^T\mathbf{C} + \mathbf{D}(\boldsymbol{\lambda})\right)^{-1}. \qquad (3.108)$$

3.6 Varying Coefficients Model

Given the observations $(t_i,\, x_i,\, y_i)$ for $i = 1, \ldots, n$, a model will be analyzed in which the response variable y will be described by the covariate x depending on the variable t. Instead of the static model $\mathrm{E}(y|\, x) = \beta_0 + \beta_1 x$, a model will be assumed in which the coefficients will vary with the variable t. In this example, a simple model with varying coefficients of the form

$$\mathrm{E}(y|\, x,\, t) = \beta_0(t) + \beta_1(t)x \qquad (3.109)$$

will be described. More complicated as well as even less complicated models can obviously be handled with this example, but here for the sake of simplicity just the model in equation (3.109) with Gaussian response data will be analyzed. Following Hastie and Tibshirani (1993), for each value t, the linear relationship between the regressor y and the variable x will hold. If the variable t represents time, then the model can be called a time-varying coefficients model. In this case, the intercept and the slope of the linear relationship varies over time, but it is assumed that $\beta_0(\cdot)$ and $\beta_1(\cdot)$ are smooth functions over the variable t.

Using a (truncated) spline basis of first order, the model in equation (3.109) can be written as

$$\begin{aligned} \mathrm{E}\,(y|\,x,\,t) &= \beta_0 + \beta_{0,\,t}t_i + \sum_{k=1}^{K} u_{k,\,0}(t_i - \tau_k)_+ \\ &\quad +\beta_1 x_i + \beta_{1,\,t}t_i x_i + \sum_{k=1}^{K} u_{k,\,1}(t_i - \tau_k)_+ \cdot x_i \end{aligned} \tag{3.110}$$

with the knots $\tau_1 < \ldots < \tau_K$ build over the support of the variable t.

Defining

$$\begin{aligned} \mathbf{y} &= \begin{pmatrix} y_1 & \ldots & y_n \end{pmatrix}^T, \mathbf{X} = \begin{pmatrix} 1 & t_i & x_i & x_i t_i \end{pmatrix}_{i=1,\ldots,n}, \\ \mathbf{Z} &= \begin{pmatrix} (t_i - \tau_1)_+ & \ldots & (t_i - \tau_K)_+ & (t_i - \tau_1)_+ x_i & \ldots & (t_i - \tau_k)_+ x_i \end{pmatrix}_{i=1,\ldots,n}, \\ \boldsymbol{\beta} &= \begin{pmatrix} \beta_0 & \beta_{0,\,t} & \beta_1 & \beta_{1,\,t} \end{pmatrix}^T, \mathbf{u} = \begin{pmatrix} u_{1,0} & \ldots & u_{K,0} & u_{1,1} & \ldots & u_{K,1} \end{pmatrix}^T, \\ \mathbf{D}_0 &= \mathbf{I}_{K\times K} \oplus \mathbf{0}_{K\times K} \text{ and } \mathbf{D}_1 = \mathbf{0}_{K\times K} \oplus \mathbf{I}_{K\times K} \end{aligned}$$

allows the model in equation (3.110) to be rewritten in matrix-vector form as

$$\mathrm{E}\,(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} \tag{3.111}$$

which can be estimated maximizing the penalized likelihood

$$-\frac{1}{2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Z}\mathbf{u})^T(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Z}\mathbf{u})-\frac{1}{2}\lambda_0\mathbf{u}^T\mathbf{D}_0\mathbf{u}-\frac{1}{2}\lambda_1\mathbf{u}^T\mathbf{D}_1\mathbf{u} \quad (3.112)$$

with the smoothing parameters λ_0 and λ_1 which might be selected via REML.

4 Application of GLM, GLMM, GAM, and Penalized Splines Regression Techniques

4.1 Cyclicality of Fiscal Policies in OECD Countries

The financial crisis in 2008 led to a domino effect in which smaller economies like Iceland but also bigger European countries like Greece, Portugal, Spain, Ireland, and Italy tumbled (or are still tumbling) and only a collective intervention of the ECB and EU with the help of the IMF stopped the bigger dominos from falling. Given the model in this chapter, the baseline model allows us to measure how interest rate changes like for the European countries might change the fiscal condition of the states, or how a GDP shock or how long-run GDP growth rates might affect the debt situation. Furthermore, the models might measure which countries will be effected the most by a GDP shock, if and how well austerity measures have been implemented, and if an equilibrium GDP growth rate is achievable.

If one dares to cry wolf, the numbers crunched should be based on reliable and dated numbers. Like the canary in the mine, the canary should not be left at the entrance but should be placed at the point which is most critical. Although Lane (2003) argued that recent research has received increasing attention, most of the estimations are based on static parameters, so the question arises whether these fancy models are reliable. For instance, Lane (2003), who examined the cyclicality of fiscal policies in OECD countries, just used ordinary least square regressions, which by definition do not account for structural breaks. Any political change would only effect the next year's regression by a small proportion due to the averaging effect. This would mean that the canary would live until there is on average not enough oxygen. Calderon and Schmidt-Hebbel (2003) confirmed a structural break

for the Latin American countries and used rolling windows as an alternative. Persson and Tabellini (2003) discuss not only country-by-country differences but also changes over time. However, it was Aghion and Marinescu (2006) who started to emphasize the time-varying effect with much more effort.

However, a penalized spline model was used by Ernst and Teuber (2006) to replicate and expand the model of Aghion and Marinescu (2006), who analyzed the relationship between the growth of an economy and the cyclicality of government debt. Aghion and Marinescu (2006) stated that the fiscal policy should affect the economic growth primarily in the short run, and in the long run it will depend upon structural characteristics of the economy. The model focuses mainly on the cyclicality of fiscal deficit of various OECD countries over time. From a statistical point of view, the analysis might involve a (time-) varying coefficient model with random effects to account for the structural difference of the countries. Unfortunately, the model presented by Aghion and Marinescu (2006) has not been formulated in any way to test for time-varying coefficients nor to test for random (or even fixed) effects which can measure the structural differences of countries. Due to the insufficiency of the estimation technique of Aghion and Marinescu (2006), Ernst and Teuber (2006) used in their joint work for the ESEMK project another approach to overcome the problem, which will be presented below.

4.1.1 The Baseline Model

For the empirical investigation of the kind of cyclicality of fiscal policies, Aghion and Marinescu (2006) started with the baseline model for public debt growth as a function of the output gap given by the tax-rate smoothing model by Barro (1979).

It will be assumed that the new debt less interest payments measured in

GDP units can be explained by the temporary deviation derived from an economic shock (measured by the output gap) which influences directly the tax and expenditure side due to less collected taxes and/or more transfer payments. Deviations from the "normal" expenditure path (for example austerity measures) are expected to influence the debt level directly. Furthermore, the new debt level (measured in GDP) might be influenced by the real interest burden, which can be modelled as a (country-specific) constant growth of debt. For the estimation it is useful to define the variables as (percentage) values of GDP to avoid heterogeneous residuals (which might be increasing proportionally with the GDP) such that the regressor will be defined as $\frac{b_t - b_{t-1} - i_t}{y_t}$ with b_t as government debt, i_t as interest payments and y_t as GDP at time t. The first covariate describes the deviation of the debt path due to a deviation from the "normal" GDP growth path measured by the output gap, i.e. $x_{1t} := y_{gap} \frac{\overline{g}_t}{y_t}$, with g_t as government expenditure at time t. The impact of the deviation from the normal expenditure path will be measured by the second covariate for which the normal level $\overline{g}_t$ will be set to the Hodrick-Prescott filtered trend of the expenditure. Using the definition above, the second covariate can be written as $x_{2t} := (\log(g_t) - \log(\overline{g}_t)) \frac{\overline{g}_t}{y_t}$. The new debt amount will be influenced by the interest payment represented by the third covariate $x_{3t} := \frac{b_{t-1}}{y_t}$ which is the last period's debt level measured in GDP. The country-specific structural characteristics will be represented by a constant coefficient, such that the corresponding parameter is different for each country, either fixed or random.

The model presented by Aghion and Marinescu (2006) can be written as

$$\begin{aligned} \frac{b_{j,t} - b_{j,t-1} - i_{j,t}}{y_{j,t}} &= \beta_{1jt} y_{gap,j,t} \frac{g_{j,t}}{y_{j,t}} + \beta_{2jt} \left(\log(g_{j,t} - \log(\overline{g}_{j,t})) \right) \frac{\overline{g}_{j,t}}{y_{j,t}} \\ &\quad + \beta_{3jt} \frac{b_{j,t-1}}{y_{j,t}} + \beta_{4jt} + \epsilon_{j,t} \end{aligned} \tag{4.1}$$

with the country,specific residuals $\epsilon_{j,t} \sim \mathcal{N}(0, \sigma^2_{\epsilon,j})$ for the country j.

Note that the coefficients β_{1jt}, β_{2jt}, β_{3jt}, and β_{4jt} depend not only on the country j, but also on the time t. In standard panel data analysis, the coefficients would not vary over time. Furthermore, in this context, the coefficients might be treated either as fixed effects or random effects. However, a more stringent model set $\beta_{\cdot jt} \equiv \beta_{\cdot t}$, such that each country would react to the covariate in the same way and the structural differences would either be expressed by those coefficients which might differ across countries or by the different values of the covariates, for instance level of expenditure and or debt burden.

Aghion and Marinescu (2006) remind us that the case $\beta_{1jt} = -1$ and $\beta_{2jt} = 1$ would correspond with tax smoothing and the coefficient β_{3jt} is expected to be the growth rate of real GDP less the interest rate. If the coefficient is less than the expected value, then a mean-reverting effect to a target debt-to-GDP value is wanted by the government which might differ from zero.

If the coefficient $\beta_{1\cdot\cdot}$ is above zero, then a procyclical deficit policy had been implemented by the government which is far from tax-rate smoothing. It is more interesting to differentiate if the government is implementing a more counter-cyclical deficit policy, i.e. $\beta_{1\cdot\cdot} < -1$, or if the coefficient is less than the tax smoothing level, i.e. if $-1 < \beta_{1\cdot\cdot} < 0$.

4.1.2 Regression Techniques

A linear regression technique would treat each time point t as an equally weighted observation point and would lead to constant estimated coefficients over time t. The model might be differentiated if country-specific coefficients are wanted and explicitly defined or if identical coefficients are assumed.

Rewriting the model in equation (4.1) as

$$y_{jt} = \beta_{1jt}x_{1jt} + \beta_{2jt}x_{2jt} + \beta_{3jt}x_{3jt} + \beta_{4jt} + \epsilon_{jt} \tag{4.2}$$

for the sake of clarification with covariates defined in subsection 4.1.1. Assuming Gaussian response variables, the expected value can be written as

$$\mu_{jt} = \beta_{1jt}x_{1jt} + \beta_{2jt}x_{2jt} + \beta_{3jt}x_{3jt} + \beta_{4jt}\,. \tag{4.3}$$

Aghion and Marinescu (2006) proposed weighting the time points $\tau \in (t-5,\ t+4)$ equally to estimate the coefficients $\beta_{\cdot\cdot t}$ at time t. Using the notation in equation (4.2), the weighted sum of the squared residuals will be minimized for each point in time $t \in (t_1, \ldots, t_n)$ and each country j to estimate the parameters $\beta_1, \ldots, \beta_4$, i.e.

$$arg \min_{\beta_{1jt},\beta_{2jt},\beta_{3jt},\beta_{4jt}} \sum_{\tau=t_1}^{t_n} \omega_\tau (y_{jt} - \mu_{jt})^2 \tag{4.4}$$

with the weights $\omega_\tau = 1$ if $\tau \in (t-5,\ t+4)$ and otherwise zero. Although the choice of using only ten time points might be in an economic view reasonable, the (subjective) choice of the bandwidth for the estimation is in a statistical sense doubtful and an automatic data-driven bandwidth should be preferred over a choice out of the blue.

For the second estimation approach Aghion and Marinescu (2006) propose using a Gaussian distribution to weight the time points. Furthermore, the standard deviation for the weighting is set to $\sigma = 5$, such that the weight is defined as $\omega_\tau = \frac{1}{\sqrt{2\pi 5^2}} \exp\left(-0.5\left(\frac{\tau-t}{5}\right)^2\right)$ used in equation (4.4). The two approaches can be treated as kernel estimators, see Nadaraya (1964)and Watson (1964) for more information, with a uniform and Gaussian kernel function, respectively. Obviously, the two bandwidths do not correspond to the same

amount of smoothness, such that the estimation of the second proposed kernel is by definition smoother. But, as in the first approach, the main shortcoming of the approach is the subjective and the more than questionable choice of smoothing parameters, which do not correspond to the smoothness of the first choice, meaning that nearly the same result might have resulted with an adjustment of the bandwidth of the kernel estimator.

Furthermore, Aghion and Marinescu (2006) assume in their third proposal that the coefficients can be described by an AR(1) process, i.e.

$$\beta_{cjt} = \beta_{cjt-1} + u_{cjt} \tag{4.5}$$

with the residuals $u_{cjt} \sim \mathcal{N}(0, \sigma^2_{u,c})$ for all coefficients $c = 1, \ldots, 4$. A maximum likelihood approach carries out the variances $\sigma^2_{u,1}$, $\sigma^2_{u,2}$, $\sigma^2_{u,3}$, $\sigma^2_{u,4}$ and $\sigma^2_{\epsilon,j}$ for each country j.

The recursive notation in equation (4.5) can be written as

$$\beta_{cjt} = \beta_{cj0} + \sum_{t=1}^{t} u_{cjt} \tag{4.6}$$

with an initial start value β_{cj0} and the cumulative sum of residuals $u_{cj.}$.

The model in equation (4.6) can be written as a random effect model with a spline basis of 0-th order, either truncated linear splines or B-splines. Re-

calling chapter 3, the model in equation (4.2) can be written as

$$\begin{aligned} y_{jt} &= \beta_{1j0}x_{1jt} + \sum_{k=1}^{K} \mathbf{1}_{\{t\geq\tau_k\}} u_{1k}x_{1jt} \\ &\quad + \beta_{2j0}x_{2jt} + \sum_{k=1}^{K} \mathbf{1}_{\{t\geq\tau_k\}} u_{2k}x_{2jt} \\ &\quad + \beta_{3j0}x_{3jt} + \sum_{k=1}^{K} \mathbf{1}_{\{t\geq\tau_k\}} u_{3k}x_{3jt} \\ &\quad + \beta_{4j0} + \sum_{k=1}^{K} \mathbf{1}_{\{t\geq\tau_k\}} u_{4k} + \epsilon_{jt} \end{aligned} \tag{4.7}$$

with the indicator function $\mathbf{1}_{\{x\}}$ which is 1 if the condition x is true and zero otherwise. The knots are given for this approach by $\tau_i = t_{0i+1}$ for $i = 1, \ldots, n_j - 1$ and thus $K_j = n_j - 1$.

Defining

$$\begin{aligned} \mathbf{y}_j &= \begin{pmatrix} y_{j1} & \ldots & y_{jn_j} \end{pmatrix}^T, \mathbf{X}_j = \begin{pmatrix} x_{1jt} & x_{2jt} & x_{3jt} & 1 \end{pmatrix}_{t=1,\ldots,n_j}, \\ \boldsymbol{\beta}_j &= \begin{pmatrix} \beta_{1j0} & \ldots \beta_{4j0} \end{pmatrix}^T, \\ \mathbf{u}_j &= \begin{pmatrix} u_{11} & \ldots & u_{1K} & u_{21} & \ldots & u_{2K} & \ldots & u_{41} & \ldots & u_{4K} \end{pmatrix}^T, \\ \mathbf{Z}_j &= \begin{pmatrix} \mathbf{1}_{\{t\geq\tau_1\}}x_{1jt} & \ldots & \mathbf{1}_{\{t\geq\tau_K\}}x_{1jt} & \ldots & \mathbf{1}_{\{t\geq\tau_1\}}x_{4jt} & \ldots & \mathbf{1}_{\{t\geq\tau_K\}}x_{4jt} \end{pmatrix}_{t=1,\ldots,n_j} \\ &\quad \text{with } x_{4\cdot\cdot} = 1 \text{ and} \\ \boldsymbol{\epsilon}_j &= \begin{pmatrix} \epsilon_{j1} & \ldots & \epsilon_{jn_j} \end{pmatrix}^T \end{aligned}$$

allows the writing of equation (4.7) as

$$\mathbf{y}_j = \mathbf{X}_j\boldsymbol{\beta}_j + \mathbf{Z}_j\mathbf{u}_j + \boldsymbol{\epsilon}_j \tag{4.8}$$

for the countries $j = 1, \ldots, m$.

For the univariate case, each country can be estimated using equation (4.8) and the penalty matrices $\mathbf{D}_1 = \mathbf{I}_k \oplus \mathbf{0}_k \oplus \mathbf{0}_k \oplus \mathbf{0}_k$, $\mathbf{D}_2 = \mathbf{0}_k \oplus \mathbf{I}_K \oplus \mathbf{0}_k \oplus \mathbf{0}_k$, $\mathbf{D}_3 = \mathbf{0}_k \oplus \mathbf{0}_k \oplus \mathbf{I}_K \oplus \mathbf{0}_k$ and $\mathbf{D}_4 = \mathbf{0}_k \oplus \mathbf{0}_k \oplus \mathbf{0}_k \oplus \mathbf{I}_K$ with the corresponding smoothing parameters $\lambda_1, \ldots, \lambda_4$.

The linear mixed model can be estimated recalling $\lambda_c = \frac{\sigma_\epsilon^2}{\sigma_{u,c}^2}$ for $c = 1, \ldots, 4$ which would preferably be selected via (RE)ML.

The combined design matrix in equation (4.8), namely $\mathbf{C}_j = \begin{pmatrix} \mathbf{X}_j & \mathbf{Z}_j \end{pmatrix}$, is of dimension $n \times 4n$, such that the dimension is too large for an unrestricted estimation. Either the smoothing parameters should be large enough so that $df_{fit} \leq n$ or another spline basis should be used. Recalling the last chapter, a spline basis built over less knots and a different order might be used. Ernst and Teuber (2006) used in their work for the ESEMK project a B-spline basis of second order with $K_j = \min(n_j/5,\, 40)$ knots equally spaced over time t and a first order penalty matrix, i.e. $\mathbf{D} = \mathbf{L}^T\mathbf{L}$ with $\mathbf{L} = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & \ddots & \ddots & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix}$. In the extreme case of no restriction on the parameters, K_j degrees of freedom will be used and in the case of the most stringent restriction, i.e. $\lambda \to \infty$, the degrees of freedom will be 1 which coincides with a constant parameter over time. Furthermore, due to the maximum likelihood approach, the hypothesis of $\sigma_{u,c}^2 = 0$ can be tested given the asymptotic distribution of the maximum likelihood parameters.

The AR model written in the form of the penalized spline model is too overloaded with respect to the number of knots so that fewer knots and/or different spline bases were used by Ernst and Teuber (2006), which is shown in the following section.

4.1.3 Empirical Results

The type of cyclicality of the public debt is the main point of interest in the research of Aghion and Marinescu (2006) and Ernst and Teuber (2006), meaning that the estimation of the coefficient $\beta_{1..}$ is in focus. Ernst and Teuber (2006) use a penalized spline model shown in equation (4.8) with a B-spline basis of second order, penalty matrices of first order, and equally spaced knots over time interval with $K_j = \min(n_j/5, 40)$ knots.

Figure 4.1 shows the time-dependent estimates for the OECD countries with its confidence bands. It is worth noting that for most countries the coefficients are (nearly) constant over time. A close look at the confidence bands reveals, that from a statistical point of view even some of the non-linear curvatures are not likely to withstand a statistical test. Furthermore, the estimated $\hat{\beta}_{1..}$ coefficients vary around the value -1, such that on average tax smoothing is targeted. Even though the conclusion is true on average, the confidence bands of some countries do not include the value -1. The analysis reveals that some countries have a pro-cyclical public debt policy while most countries implement counter-cyclical policies. Looking at the confidence bands more closely, one can see that there is not a single country which has its confidence bands completely above zero at all points in time. Furthermore, the analysis shows that the coefficients vary over time and between countries. Some countries implemented a pro- and counter-cyclical fiscal debt policy over the period which can be verified looking at their confidence bands. It seems remarkable that no major trend is visible, as there is no single time-variant function which differs only slightly over the period. Even in the later years, there is no common trend recognizable for the EMU countries.

In figure 4.2, the time structure of the cyclical public debt coefficient is shown for Canada (black curve), France (red curve), and Germany (green

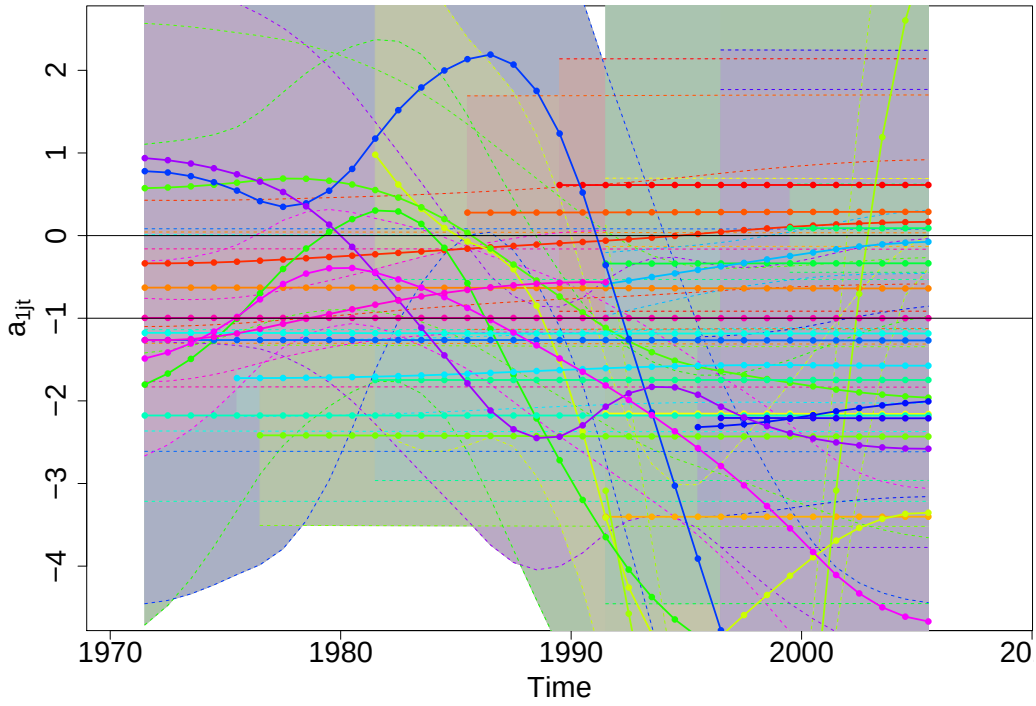

Figure 4.1: Estimation of the time-varying policy debt cyclicality coefficient a_{1jt} for the OECD countries over time.

---

curve) which is from a curvature point of view nearly identical and differs only on the level (intercept). All three countries are characterized by a counter-cyclical public debt policy which was the strongest in the '70s. At the beginning of the '80s, the counter-cyclicality was becoming weaker and has been slightly increasing since.

Figure 4.3 reveals that the curvature of the cyclicality of the UK (black curve), Japan (red curve), and the US (green curve) is nearly identical. Looking at the confidence bands this hypothesis is hard to reject. The cyclicality

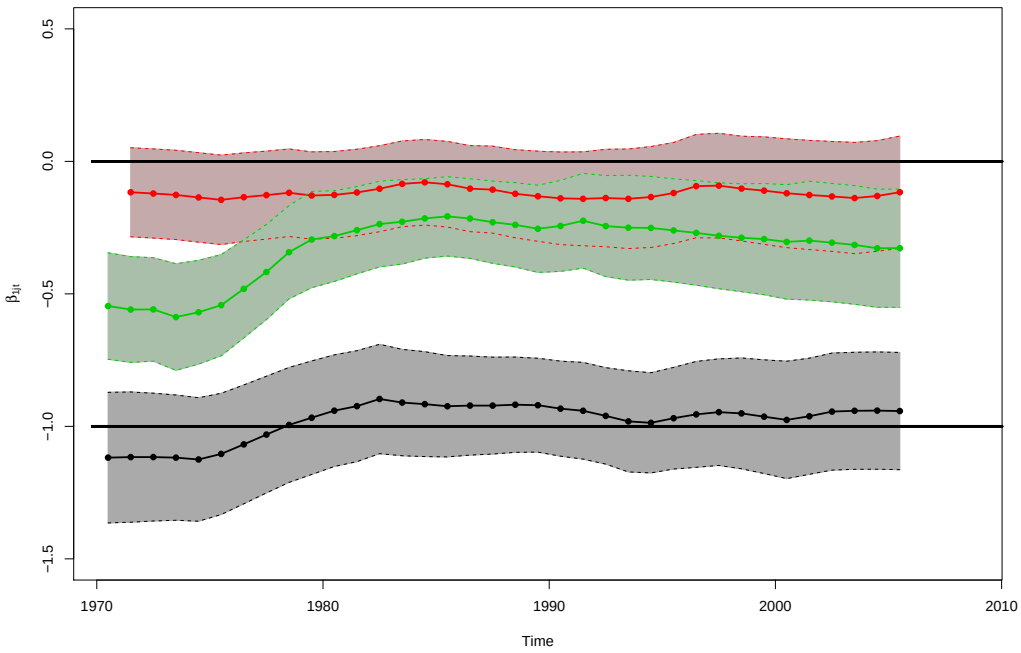

Figure 4.2: Estimation of the time-varying policy debt cyclicality coefficient a_{1jt} for selected OECD countries over time, namely Canada (black line), France (red line), and Germany (green line).

---

in the UK and Japan became more counter-cyclical from the '80s to the beginning of the '90s and has been nearly constant since. In the US, the cyclicality was nearly constant in the '70s nearly constant, then started to get more counter-cyclical until the beginning of the new century and has nearly been on a high counter-cyclical level since then. It is remarkable that all three countries seem to follow the same trend, which starts at different time points. Furthermore, the public debt policy in the UK was pro-cyclical in the '70s and has been counter-cyclical since the '90s. Japan and the US were both on a weaker counter-cyclical level in the '70s ($\beta_{1..} > -1$) than nowadays

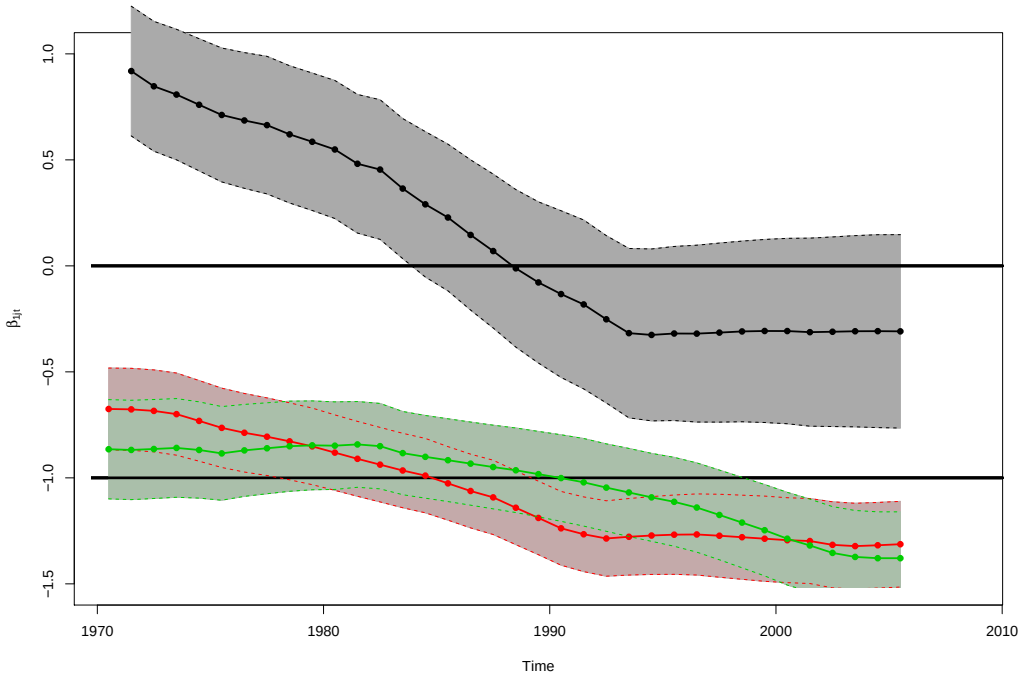

Figure 4.3: Estimation of the time-varying policy debt cyclicality coefficient a_{1jt} for selected OECD countries over time, namely the United Kingdom (black line), Japan (red line), and the United States (green line).

---

and have been becoming more counter-cyclical since the '90s ($\beta_{1..} < -1$).

4.1.4 Extension of the basic model

The model in equations (4.1) and (4.2) does not differentiate which kind of government expenditure is the reason for the change in the public debt. A model in which the expenditures will be differentiated in more detail is preferred, so the expenditure $g_{j,t}$ will be split into three variables, namely the non-wage expenditures $g_{1,j,t}$, the wage expenditures $g_{2,j,t}$ and the investment

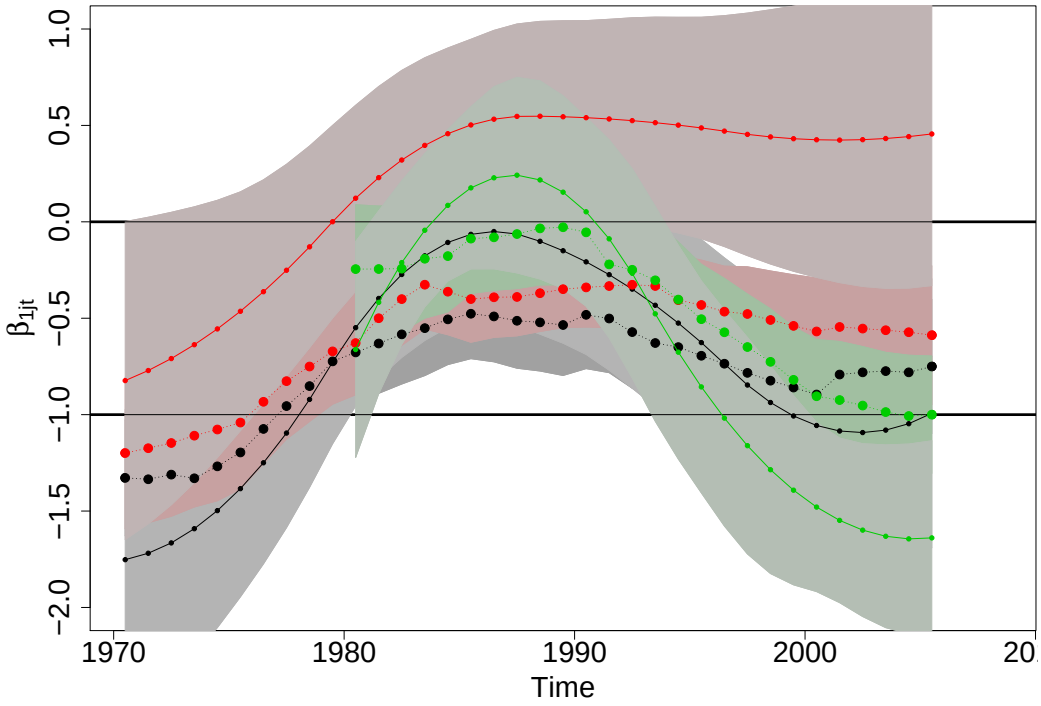

Figure 4.4: Estimation of the time-varying policy debt cyclicality coefficient a_{1jt} for selected OECD countries over time, namely Germany (black line), Italy (red line), and the Netherlands (green line). Solid line with small dots: Penalized spline regression; broken line with big dots: Kalman filter.

---

of the government $g_{3,j,t}$ at each time t and for each country j. Therefore, the variables $x_{5jt} = \big(\log(g_{1,j,t}) - \log(\overline{g}_{1,j,t})\big)\,\frac{\overline{g}_{1,j,t}}{y_{j,t}}$, $x_{6jt} = \big(\log(g_{2,j,t}) - \log(\overline{g}_{2,j,t})\big)\,\frac{\overline{g}_{2,j,t}}{y_{j,t}}$ and $x_{7jt} = \big(\log(g_{3,j,t}) - \log(\overline{g}_{3,j,t})\big)\,\frac{\overline{g}_{3,j,t}}{y_{j,t}}$ will be defined to replace $x_{2,j,t}$ in the equation (4.2), i.e.

$$y_{jt} = \beta_{1jt}x_{1jt} + \beta_{5jt}x_{5jt} + +\beta_{6jt}x_{6jt} + \beta_{7jt}x_{7jt} + \beta_{3jt}x_{3jt} + \beta_{4jt} + \epsilon_{jt} \quad (4.9)$$

which is the extended model analyzed by Ernst and Teuber (2006).

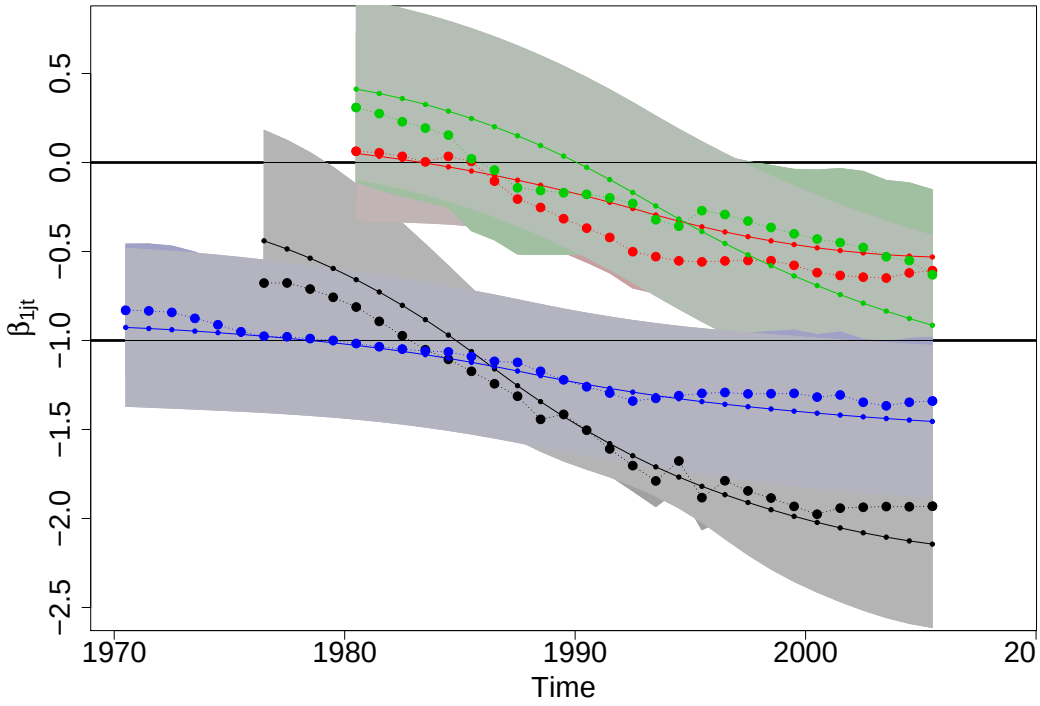

Figure 4.5: Estimation of the time-varying policy debt cyclicality coefficient a_{1jt} for selected OECD countries over time, namely Finland (black line), Austria (red line), Iceland (green line) and Japan (blue line). Solid line with small dots: Penalized spline Regression; broken line with big dots: Kalman filter.

The estimation of the extended model in equation (4.9) reveals no major trend over time or between the countries for the cyclicality of the public debt policy, either. However, some countries form a cluster with the same curvature of the function $\beta_{1\cdot\cdot}$ over time, as already seen in the previous section.

Figure 4.4 shows the estimated coefficients of Germany (black curve), Italy (red curve), and the Netherlands (green curve) which follow nearly the same structure over time. The counter-cyclicality of the three countries was around

the tax smoothing level of -1 and got closer to zero from the end of '70s towards the end of '80s and has had a more counter-cyclical pattern since the mid-1990s. Italy is the only country which implemented a pro-cyclical public debt policy and not followed the trend to a (more) counter-cyclical fiscal policy from the '90s onwards. However, apart from the different levels, the curvature is roughly the same, which is confirmed by looking more closely at the confidence bands.

In figure 4.5, another cluster of countries is shown, namely Finland (black curve), Austria (red curve), Iceland (green curve), and Japan (blue curve). All the countries share nearly the same trend over time, which is negative sloping. Austria and Iceland had been pro-cyclical at the beginning of the '80s and downward sloped to a counter-cyclical policy with an estimated coefficient slightly above the current -1 level. In contrast, Finland and Japan had been in the '70s at a level close to -1 and been stronger counter-cyclical since then, and are now at a level significantly below -1. The same is true for this group as for the group, shown in figure 4.4, that apart from the level, the curvature could be the same over time, indicated by overlapping confidence regions at any point in time (if adjusted correctly for the level).

4.2 Structuralist Model of the Wage-Price Spiral with Non-Linear Demand Pressure Terms

Flaschel, Tavani, Taylor, and Teuber (2008) introduced the results of a non-parametric estimation of the wage Phillips curve into a simplified version of the model by Flaschel and Krolzig (2006). The resulting non-linearity in the wage inflation employment relation translates into a non-linearity in the reduced form of the model, namely the wage-price spiral.

4.2.1 Cross-over Wage-Price Dynamics

Lets start with a simplified version of the wage-spiral model from Flaschel and Krolzig (2006) with the structural form

$$\frac{\dot{w}}{w} = \beta_w \left(\overline{U}^l - U^l\right) + \kappa_w \left(\frac{\dot{p}}{p} + \eta_x\right) + (1-\kappa_w)(\pi + \eta_x) \tag{4.10}$$

$$\frac{\dot{p}}{p} = \beta_p \left(\overline{U}^c - U^c\right) + \kappa_p \left(\frac{\dot{w}}{w} - \eta_x\right) + (1-\kappa_p)\pi \tag{4.11}$$

with U^l as labor unemployment rate, U^c as capital unemployment rate, $\overline{U}^l$ and $\overline{U}^c$ as NAIRU values of labor and capital unemployment rate, respectively, η_x as Harrod-neutral technical change rate and π as inflationary climate, $\frac{\dot{w}}{w}$ as wage inflation, and $\frac{\dot{p}}{p}$ as price inflation. For the sake of simplification, the growth rate of technology is set to zero, i.e. $\eta_x \equiv 0$, which would not affect the structure of the model because it will only affect the level of intercept in this model; see Flaschel, Tavani, Taylor, and Teuber (2008).

The model in the equations (4.10) and (4.11) can be written as

$$\begin{aligned} \frac{\dot{w}}{w} &= \beta_w \left((1-\overline{e}) - (1-e)\right) + (1-\kappa_w)\pi + \kappa_w \frac{\dot{p}}{p} \\ &= \beta_w (e - \overline{e}) + (1-\kappa_w)\pi + \kappa_w \frac{\dot{p}}{p} \end{aligned} \tag{4.12}$$

$$\begin{aligned} \frac{\dot{p}}{p} &= \beta_p \left((1-\overline{u}) - (1-u)\right) + (1-\kappa_p)\pi + \kappa_p \frac{\dot{w}}{w} \\ &= \beta_p (u - \overline{u}) + (1-\kappa_p)\pi + \kappa_p \frac{\dot{w}}{w} \end{aligned} \tag{4.13}$$

with the labor and capital employment rate $e := 1 - U^l$ and $u := 1 - U^c$, respectively, and the NAIRU of labor and capital employment as $\overline{e} := 1 - \overline{U}^l$ and $\overline{u} := 1 - \overline{U}^c$, respectively.

In matrix-vector form, the model in the equations (4.12) and (4.13) can be

written as

$$\begin{pmatrix} 1 & -\kappa_w \\ -\kappa_p & 1 \end{pmatrix} \begin{pmatrix} \frac{\dot{w}}{w} - \pi \\ \frac{\dot{p}}{p} - \pi \end{pmatrix} = \begin{pmatrix} \beta_w(e - \overline{e}) \\ \beta_p(u - \overline{u}) \end{pmatrix}. \tag{4.14}$$

Defining $\kappa = (1 - \kappa_p\kappa_w)^{-1}$, the solution of the model in equation (4.14) is given by

$$\begin{pmatrix} \frac{\dot{w}}{w} - \pi \\ \frac{\dot{p}}{p} - \pi \end{pmatrix} = \kappa \begin{pmatrix} 1 & \kappa_w \\ \kappa_p & 1 \end{pmatrix} \begin{pmatrix} \beta_w(e - \overline{e}) \\ \beta_p(u - \overline{u}) \end{pmatrix}. \tag{4.15}$$

Remembering the absence of productivity growth, the growth of wage share $\psi = w/p/x$ is given by

$$\frac{\dot{\psi}}{\psi} = \frac{\dot{w}}{w} - \frac{\dot{p}}{p} = \kappa\left((1 - \kappa_p)\beta_w(e - \overline{e}) - (1 - \kappa_w)\beta_p(u - \overline{u})\right) \tag{4.16}$$

which shows the wage share response to both utilization rates, namely of labor and capital.

4.2.2 A Generalized Additive Model to Estimate the Phillips Curve

In the model presented by Flaschel, Tavani, Taylor, and Teuber (2008), the wage Phillips curve is described by the price inflation, the (log) wage share, the employment rate, and the price inflation climate. Therefore, the wage Phillips curve as a Generalized Additive Model is given by

$$\mathrm{E}\left(\frac{\dot{w}}{w}\right) = f_1\left(\frac{\dot{p}}{p}\right) + f_2\left(\psi\right) + f_3\left(e\right) + f_4\left(\pi\right) \tag{4.17}$$

with the unknown and unspecified functions $f_1, \ldots, f_4$. It is easily justifiable that the wage inflation can be treated as Gaussian distributed, so that the response variable can be modelled as

$$y_t = f_1(x_{1t}) + f_2(x_{2t}) + f_3(x_{3t}) + f_4(x_{4t}) + \epsilon_t \tag{4.18}$$

with $y = \frac{\dot{w}}{w}$, $x_1 = \frac{\dot{p}}{p}$, $x_2 = \log \psi$, $x_3 = e$, $x_4 = \pi$ and the normal residuals ϵ. The model in equation (4.18) had been estimated using a penalized spline approach, i.e.

$$l(\boldsymbol{\theta}) = -(\mathbf{y} - \mathbf{C}\boldsymbol{\theta})^T(\mathbf{y} - \mathbf{C}\boldsymbol{\theta})/\sigma^2 - \lambda_1\boldsymbol{\theta}^T\mathbf{D}_1\boldsymbol{\theta} - \ldots - \lambda_m\boldsymbol{\theta}^T\mathbf{D}_m\boldsymbol{\theta} \quad (4.19)$$

with the combined design matrix $\mathbf{Z} = \begin{pmatrix}\mathbf{X} & \mathbf{Z}\end{pmatrix}$ which contains the fixed effect design matrix $\mathbf{X} = \begin{pmatrix}1 & x_{1t} & \ldots & x_{1t}^{q_1} & \ldots & x_{mt} & \ldots & x_{mt}^{q_m}\end{pmatrix}_{t=1,\ldots,n}$ and the random effect matrix $\mathbf{Z} = \begin{pmatrix}\mathbf{Z}_1 & \ldots & \mathbf{Z}_m\end{pmatrix}$ which are constructed using a linear spline basis of q_j-th order for the j-th covariate, i.e. $\mathbf{Z}_j = \begin{pmatrix}(x_{jt} - \kappa_{1,j})_+^{q_j} & \ldots (x_{jt} - \kappa_{K_j,j})_+^{q_j}\end{pmatrix}_{t=1,\ldots,n}$ with the knots $\kappa_{1,j} < \ldots < \kappa_{K_j,j}$ over the support of x_j.
In this example, different orders of the truncated polynomial have been used. On the one hand, the degrees of freedom for some functions should not be too high, so that for those functions an order of the first degree is used. On the other hand, for some coefficients more degrees of freedom are justifiable and a higher order will be used to estimate a smooth first derivative. To keep the model simple, the same number of knots for all covariates has been used, i.e. $K_1 = \ldots = K_m \equiv K$. The main diagonal of the penalty matrix $\mathbf{D}_l$ contains a 1 if the index belongs to the truncated spline basis $\mathbf{Z}_l$, and otherwise the elements are zero, i.e. $\mathbf{D}_l = (d_{ij})_{i,j=1,\ldots,mK+q+1}$ with $q = \sum_{i=1}^m q_i$ and $d_{ij} = \mathbf{1}_{\{i=j\}}\mathbf{1}_{\{i\in\{q+2+(l-1)K,\ldots,q+1+lK\}\}}$.

The smoothing parameters which control the complexity of the structure of the unknown functions $f_.$ will be selected via REML to avoid misleading parameters in the case of misspecified autocorrelated errors; see Krivobokova and Kauermann (2007). In the case of no penalization of the j-th covariate, i.e. $\lambda_j = 0$, the function f_j contains $q_j + K$ degrees of freedom. If the function is fully penalized, i.e. $\lambda_j \to \infty$, the function f_j is a function of order q_j.

In a first step, Flaschel, Tavani, Taylor, and Teuber (2008) have set the orders of the spline bases to one, i.e. $q_1 = \ldots = q_4 \equiv 1$ to avoid misleading estimators due to too many degrees of freedom even if $\lambda_j = \infty$. The estimation shows that the price inflation and the wage share are linearly linked to the wage inflation. The employment rate and the price climate are linked in a non-linear way, in particular the function f_3 has more than three degrees of freedom, such that in the next step a higher polynomial order will be used, here $q_3 = 2$, to estimate smooth first derivatives of the function f_3. The corresponding estimation is nearly similar to the first one, such that in the latter model, the first derivative can be estimated be referring back to equation (4.18).

Given the model in equation (4.18), the first derivative with respect to the j-th covariate is given by

$$\frac{\partial y}{\partial x_j} = \frac{\partial f_j(x_j)}{\partial x_j} \tag{4.20}$$

and the first derivative of the estimated function is given by

$$\frac{\partial \hat{y}}{\partial x_j} = \frac{\partial \hat{f}_j(x_j)}{\partial x_j} \tag{4.21}$$

with $\hat{f}_j(x_j) = \hat{c} + x_j\hat{\beta_{j1}} + \ldots + x_j^{q_j}\hat{\beta}_{jq_j} + \mathbf{Z}_j\hat{\boldsymbol{\theta}}_j$ given the predicted coefficient $\hat{\boldsymbol{\theta}}_j = \left(\hat{\theta}_{q+2+(j-1)K}, \ldots, \hat{\theta}_{q+1+jK}\right)^T$ and the estimated coefficients $\hat{\boldsymbol{\beta}}$. Given the construction of the design matrices, the equation (4.21) can be written as

$$\frac{\partial \hat{\mathbf{y}}}{\partial x_j} \quad = \quad \mathbf{X}_{j,p}\hat{\boldsymbol{\beta}} + \mathbf{Z}_{j,p}\hat{\boldsymbol{\theta}}_j \tag{4.22}$$

with the first derivative of the design matrix for the fixed and random effect design matrices with respect to the j-th covariate which is given by

$$\mathbf{X}_{j,p} := \frac{d\,\mathbf{X}}{d\,x_j} = \begin{pmatrix} 0 & \ldots & 0 & 1 & 2x_{jt} & \ldots & q_j x_{j,t}^{q_j-1} & 0 & \ldots & 0 \end{pmatrix}_{t=1,\ldots,n} \tag{4.23}$$

and

$$\mathbf{Z}_{j,p} := \frac{d\,\mathbf{Z}_j}{d\,x_j} = \left(q_j(x_{j,t} - \kappa_{1,j})_+^{q_j-1} \quad \dots \quad q_j(x_{j,t} - \kappa_{K,j})_+^{q_j-1}\right)_{t=1,\dots,n} \tag{4.24}$$

respectively. Remember that $\frac{d\,\mathbf{Z}_j}{d\,x_i} = \mathbf{0}_{n\times K}$ if $i \neq j$.

Also remembering, that the advantage of being able to write the (first) derivative of the estimated function in the format in equation (4.22) is that confidence bands can easily be estimated, referring to the asymptotic distribution of maximum likelihood estimations.

4.2.3 Non-linearities in the Wage Demand-Pressure Term

The estimation in the previous section gives rise to the non-linear relationship between wage inflation and demand pressure on the labor market shown in figure 4.6: The curve is increasing up to an employment rate of roughly 92%, then has an almost flat or at most slightly decreasing region, and increases for employment levels above 94%.

The estimated first derivative of the estimated function shows a non-linear relation. The first increasing portion is virtually linear; at an employment rate of about 91%, the curve becomes concave until an inflexion point around 93.5% employment rate, after which the curve increases and becomes convex again. Eventually, there is another inflexion point around an employment rate of roughly 95.5%, and the final portion of the curve is increasing but concave.

Although the US labor market is not internationally known for the strength of its labor unions, a standard economic institution behind the behavior of the curve could lay on a bargaining power argument relative to labor supply. For high levels of unemployment, the workers' bargaining power is small:

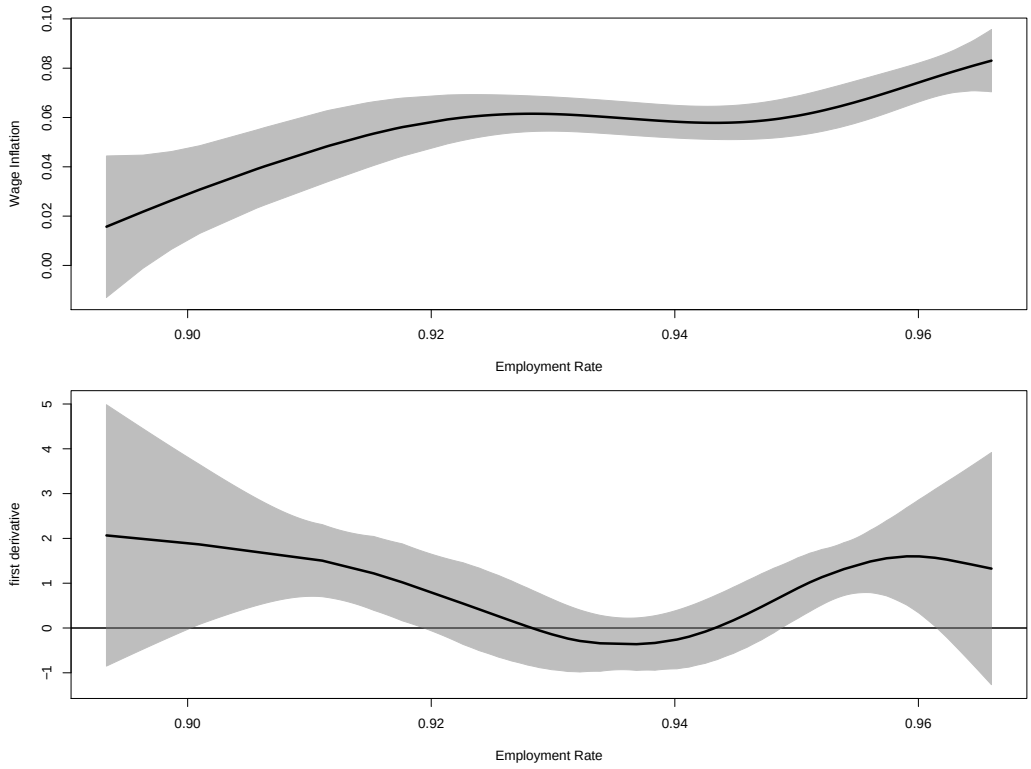

Figure 4.6: Estimated wage inflation for different levels of the employment rate: Upper plot shows the estimated wage inflation (black line) and the confidence regions (grey-shaded area) given the employment rate. The lower plot shows the estimated first derivative of the wage inflation (black line) with respect to the employment rate; the shaded region represents the corresponding confidence regions.

They (or the labor union representing them) will be satisfied with only small increases or even decreases in the nominal wage in order to increase the employment rate. Corresponding to the center of the curve, there is a flat region where labor is resisting wage inflation and decreases at the given expected price inflation, a situation widely familiar through Keynes' discussion of it. Finally, as soon as the economic activity is above a NAIRU-type full employ-

ment rate, workers will exercise their increased bargaining power in requiring significantly more than proportional increases in wage inflation (as compared to price inflation).

Similarly, for the price Phillips curve, Flaschel, Tavani, Taylor, and Teuber (2008) are described the price inflation (y) by the covariates wage inflation (x_1), the log of the wage share (x_2), the utilization rate (x_3) and the price inflation climate (x_4). But, in contrast to the wage Phillips curve the functional shape of the price Phillips curve with respect to the utilization rate is not distinctively different from a linear curve (as was the functional shape of the wage Phillips curve with respect to the employment rate), which is the reason why the visualization of the price Phillips curve estimation has been omitted.

4.3 The Dutch Tax Benefit System and Life Cycle Employment

Kotlikoff and Rapson (2006) asks in their work if it pays at the margin to work and save because highly advanced computer technology and software to unfold the incredibly complex US fiscal system. Their research is not only important from an individual perspective to calculate an optimal path given the current state, but also from a political point of view because the "amazing complexity of the fiscal system" and the various components of the systems are being developed with little or no thought to their interaction.

The effective marginal tax rate is enormously high for low income individuals because at a given threshold not only wage taxes have to paid, but more importantly welfare transfers will be cut. At first glance, a flat tax might be favored just by high-income earners. Conesa and Krueger (1999) analyzed the

voting distribution for new tax regimes vs. an already implemented regime for different scenarios. It should be noted that the favoring of a tax regime does not only depend on the age (or time to retirement), but also within cohorts different tax regimes might be favored due to accumulated capital stock, productivity level, and income. In economies with social welfare and progressive tax regimes, the voting distribution within cohorts is not necessarily linear in the meaning of just one income (or wealth) threshold to split the cohort into two groups favoring one or the other tax regime.

Bloemen (2007) investigated the wealth effect on the early retirement rates by elderly workers which can be observed empirically and was confirmed in his model. However, French (2003) observed that social security benefits, health, and borrowing constraints are less important determinants of job exit at older ages. Furthermore, Määttänen and Poutvaara (2007) analyzed the effect of earnings test for old-age benefits because in some countries the earnings test and/or wealth-test is implemented with respect to federal retirement, social, and unemployment benefits.

For the analysis of the effects of new tax regimes, it is unavoidable to use an as realistic as possible tax regime for the economy. This means that not only tax breaks but also earning/income and wealth-tests for social, unemployment, and retirement welfare should replicate the current laws. Furthermore, the level of welfare transfers are crucial to the analysis. However, most research stops at this level when it comes to calibrating their model. It is clear that the more time points (cohorts) are used in the model, the better the results should be. From a computational point of view yearly time steps can be handled and from a reliability point of view the grid is fine enough. Survival probabilities, productivity over the working age, as well as probabilities of losing the job can be approximated quite well on a yearly basis. However,

the calibration of most models normally lacks a good estimation technique of survival probabilities, the age productivity, transition probabilities of productivity, as well as the probabilities of suffering unemployment or finding a new job based on the productivity and age.

Given the conclusion by French (2003), the complex analysis of the implemented tax regimes is sensitive to small changes, such that realistic modelling/calibration is crucial. Ernst and Teuber (2008) and Ernst and Teuber (2012) calibrated an overlapping generation model with search unemployment for the Netherlands to assess the impact of tax-benefit reforms on labor supply. For the calibration the DNB Household Survey with data over 14 years was used to estimate the age-productivity profiles and transition probabilities for income levels the better calibration allowing more light to be shone on the tax reform discussion.

4.3.1 The Overlapping Generation Model

The overlapping generation model by Ernst and Teuber (2008) and Ernst and Teuber (2012) is based on the model by Auerbach and Kotlikoff (1987) and Kotlikoff and Rapson (2006) in which each household optimizes their consumption, working hour, and labor market participation decisions. The economic life of an individual starts at age 18 and ends with certain death at age 100 if the individual does not pass away before this age. The maximum number of periods an individual lives is therefore 83 periods and the probability of surviving the current period is given by the survival probability $0 \leq s(t) \leq 1$, which was estimated through the use of the Dutch death statistics. The life span of each individual is divided into working life (age 18-65) and retirement (age 66-100) with a certain time of retirement.

Each cohort consists of three groups, namely the inactive group, which only

has access to social assistance, the unemployed group, which receives earning-related unemployment benefits, and employed individuals, who receive a wage depending on their individual profitability. The profitability depends on their individual profitability and an age dependent cohort productivity. At the age of 65, each cohort will become inactive and will receive a basic pension. Furthermore, each household can consume or save for future periods via accumulated wealth. If a household with positive accumulated wealth dies before certain death at age 100, the wealth will be distributed across cohorts.

Each household maximizes its intertemporal utility by making an optimal arbitrage between work and leisure over its life cycle. The optimal labor supply comprises the decision to participate on the labor market, the number of hours worked, and participation itself, which is costly.

Households have to pay wealth and consumption taxes as well as income taxes and receive transfers in the form of a basic pension, social assistance, and unemployment benefits.

During the working life the households decide in the first step if participation is wanted or if the wealth stock and/or transfers will be used for consumption, which determines the next period's wealth stock. Participating workers optimize the labor input conditional on being employed, the consumption, and the next period's wealth stock.

Households which are retired need to optimize their consumption to determine the next period's wealth stock. Note that each households needs to optimize expected utility due to the uncertainty of surviving the current period, except for the retired household at age 100 in which the wealth stock will be fully consumed due to the certain death at the end of the period.

Wealth accumulated by the household sector is used by firms for production investment. Moreover, firms will decide upon total hours worked by opening vacancies to fill available jobs, taking the decision on average hours as given. Jobs are filled through a search and matching process on the labor market, leaving some markets unrealized and thereby generating unemployment even in steady state. In order to maximize the net present value of their profits, firms select their flows of investments and vacancies.

Wages are negotiated at the firm level. As a first approximation to a Nash-bargaining distribution of profits, wages are determined as a weighted average between the marginal contribution of an additional worker to the firms' profits and the worker's fall-back option, i.e. social assistance.

The schedule for marginal effective tax rates on labor income is taken from OECD (2006), such that statutory tax rates for different income brackets and various social benefits have been included. General tax credits and work-related tax credits have been used which have been calibrated for a single household and a married couple with two children to identify the burden of different tax-benefit reforms on these groups. The VAT of 19% and the wealth tax of 1.2% represent the current tax situation in the Netherlands. In the Beveridge-type social security system of the Netherlands, benefits are exclusively financed out of tax revenues.

Social assistance is available for non-participating households subject to a capital income test, whereby all capital income above 15% of the average wage is deducted from social assistance. In order to integrate the idea that younger people face a higher risk of inactivity without any replacement income, a relative wage profile has used to adjust the replacement income

accordingly.

Unemployment insurance is granted for periods during which the individual is participating but without a job. The replacement income covers 70% of the last salary but is limited to 30% of the average salary.

The Beveridge-style state pension amounts to 30% of the average wage, which constitutes a social safety net and represents 34% of pensioners' average earnings; see OECD (2006).

Reforms of tax benefits and/or the pension system impose a balanced-budget rule. The government pays out state pensions and social benefit exclusively by levying taxes. In addition, the government finances (unproductive) government spending at an arbitrary rate of GDP. The model abstracts from budget deficits and public debt. Should a budget surplus occur, these additional budgetary resources made available by lowering replacement rates or raising the marginal income tax rates or raising the marginal tax of pensioners and are redistributed to households via lower consumption taxes.

4.3.2 Calibrating the Model and Determining the Equilibrium Distribution of Work and Consumption

In order to obtain a reliable estimation of the life cycle tax burden and a household's labor supply decisions, the model has been calibrated using the DNB Household Survey. In particular, in order to properly reflect distributional consequences of different tax-benefit systems, a Markov transition matrix with ten different income deciles and an unemployment state has been estimated with the use of the DNB Household Survey. This transition matrix indicates the probabilities with which individuals in different income strata persist at their current (relative) income level or move up or down

from one period to the other. As the transition probability changes over time, the matrix has been estimated time-dependently over the entire working life span. Moreover, age productivity profiles have been estimated for an average worker and used to modulate the wage over each cohort's working age. Finally, the survival probabilities for each age group have been taken directly from Statistic Netherlands - and for the last year (at age 100) has been set to zero.

The estimation (calibration) of the individual and age productivity will be discussed in more detail in the next subsection. The remaining parameters have been calibrated by Ernst and Teuber (2008) and Ernst and Teuber (2012) to measure the Dutch economy.

Given the parameters as well as the tax and benefit regimes, a finite value approach has been chosen to calculate the equilibrium distribution of consumption and working hours due to the fact that life ends with certainty at age 100. For each age, the optimal path of consumption and worked hours for each value out of a large, but finite number of capital stocks has been calculated recursively. The equilibrium value is then determined by selecting the path with the highest net present value of the individual's utility.

4.3.3 Estimation of the Productivity over the Life Cycle

The productivity of a worker depends on several covariates which might be fixed or random. The age and therefore the work history as well as the skills and education are the most important factors to determine the productivity of a worker. Both, Ernst and Teuber (2008) and Ernst and Teuber (2012), used the log of wage per hour to get a proxy of the productivity of a worker which implies that the (hourly) wage measures the productivity of a worker or at least its expected productivity. Although this proxy is obviously not a

perfect linkage of an individual level, on the aggregate level the simplification might hold to a certain extent.

For the age-dependent function over time, different approaches have been used by economists so far: Määttänen and Poutvaara (2007) have blindly copied the estimation of Floden and Lindé (2001), who proposed a polynomial of second order to estimate the age-dependent productivity function. Börsch-Supan, Ludwig, and Winter (2006) copied one-by-one the estimation function of Fitzenberger, Hujer, MaCurdy, and Schnabel (2001), who preferred a polynomial of third order; French (2003) used a polynomial of fourth order. From an economic point of view, the more complex functions might be preferable due to the fact that the other models are nested in them. The (estimated) second order model has some shortcomings in the design due to the fact that the productivity of a worker does not peak at the age of 50 and is symmetrical around that age. From a statistical point of view, the estimation of polynomial functions of higher orders are not reliable due to the high estimation errors of the coefficients, which leads to wide confidence bands of the estimated productivity function. It is worrisome that none of the authors have shown that their models and/or the additional parameters are significant in a statistical sense.

Ernst and Teuber (2008) and Ernst and Teuber (2012) used a (Generalized) Additive Mixed Model approach to estimate the productivity, i.e.

$$y_{it} = f(t) + \mathbf{x}_{it}\boldsymbol{\beta} + u_i + \epsilon_{it} \tag{4.25}$$

with y_{it} as the productivity of the individual i at time t measured by the log wage per hour, $\mathbf{x}_{it}$ as covariates for the individual i at time t, $\boldsymbol{\beta}$ as unknown parameters, u_i as idiosyncratic productivity level of the individual i, and ϵ_{it} as residuals for each individual $i = 1, \ldots, m$ at each point in time (age)

$t = \underline{t}_i, \ldots, \bar{t}_i$.

The idiosyncratic level of productivity u_i is a random covariate which determines individual productivity which accounts for example for the skill of a worker, the wage premium/discount due to the company, and a wage difference due to location and/or industry of the job. The fixed covariates account for observable indicators which (might) influence the productivity and/or wage of a worker, which are in this case the education, which can vary over time, and the year to account for (wage) inflation by introducing dummy variables.

The age-dependent productivity function f should measure the wage/productivity structure of a worker over time without accounting for a different education, to circumvent the fade-in effect of better educated workers at the beginning of the (working) life and to avoid a misspecification of the function due to effects of early retirement of workers towards the end of the (working) life.

The calendar year could also be modelled as a random effect, but has been modelled solely as a fixed effect for the sake of simplification.

Ernst and Teuber (2008) and Ernst and Teuber (2012) assumed that the unknown function f can be described by a truncated spline basis of first order, i.e.

$$f(t) = a_0 + a_1 t + \sum_k = 1^K (t - \tau_k)_+ \alpha_k \tag{4.26}$$

with K equidistant knots $18 \leq \tau_1 < \ldots < \tau_K \leq 65$ over the support of the age and used an REML approach to predict the random effect coefficients $\alpha_k \sim \mathcal{N}(0, \sigma_\alpha^2)$. The degrees of freedom will be determined by the variance σ_α^2 which is two if $\sigma_\alpha^2 = 0$, and $K + 2$ if $\sigma_\alpha^2 \to \infty$.

The estimated function f shown in figure 4.7 is from a statistical point of view significantly different to the estimated functions of the functions mentioned above. It is worth noting, that the strong increase of productivity at a young age is not due to a better (scholarly) education nor due to the fact that better educated workers start working later and thus receive higher salaries. The productivity measured by the salary is increasing the strongest at the beginning of the working life, which might be due to the effect that the workers might learn the most (or become more productive) when they start their jobs, which represents a steep learning curve.

The estimation shows that learning on the job (or at least aging) will be rewarded by higher salaries and/or fewer working hours for the same salary. The drop towards the end of the (working) life is remarkable, although it is questionable if the productivity is decreasing or if just the wage per hour is decreasing.

The idiosyncratic shock of a randomly chosen individual is given by the sum of the individual random level of productivity u and the residual ϵ, i.e. $s_{it} = u_i + \epsilon_{it}$. Ernst and Teuber (2008) and Ernst and Teuber (2012) defined a discretized version of the idiosyncratic shock in their overlapping generation model. In the general form the variable s will be discretized to a variable ds with c different classes with equal mass such that

$$ds_{it} = j \qquad \text{if} \qquad \frac{(j-1)N}{c} \leq \sum_{t'=\underline{t}_i}^{\bar{t}_i} \mathbf{1}_{\{s_{i',t'} \leq s_{i,t}\}} < \frac{jN}{c} \tag{4.27}$$

with N as the total number of observations and the classes $j = 1, \ldots, c$. The 0-th class will be introduced to account for unemployment, i.e. $ds_{it} = 0$ if the individual i is unemployed at time t.

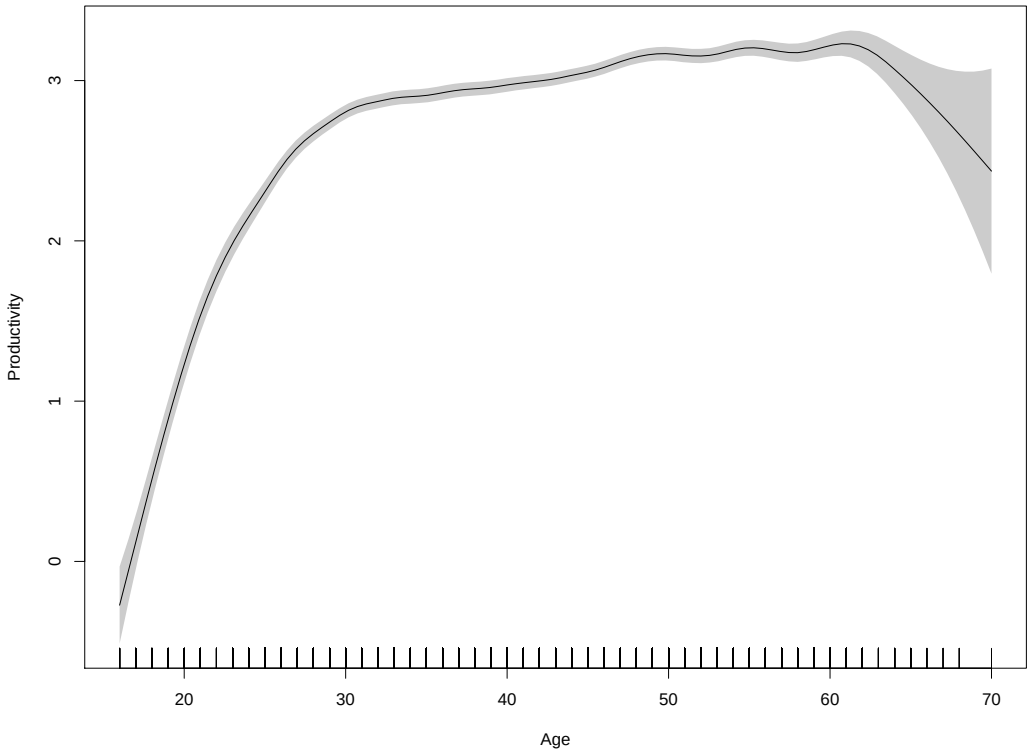

Figure 4.7: Estimation of the age-productivity function: The estimated productivity (black line) over the working life with confidence regions (grey-shaded area).

It is worth noting that the empirical analysis has shown that the idiosyncratic shock of an individual will most likely be the same one as the one just one period earlier. With the exception of unemployment (class zero), the change of the idiosyncratic shock from one period to another is in most cases moderate. This phenomenon increases in strength, the older the individual is. The unemployment state is a special case: A worker with a high idiosyncratic shock (for instance $ds_{it} = c$) is more likely to become unemployed than to be in the worst decile (i.e. $ds_{it+1} = 1$).

The persistence in the same shock category is increasing with the age of the worker and the probabilities for extreme changes is higher for younger workers. An age-dependent transitory matrix was therefore preferred.

Due to the definition of the idiosyncratic shock ds, a Generalized Linear Model will be used to model the age-dependent relationship so that the probability of the worker i being in the l-th class in the next period if the worker was at time t in class k is given by

$$P\left(ds_{it+1} = l | ds_{it} = k,\, t\right) = \frac{\exp(\beta_{kl0} + \beta_{kl1} t)}{\sum_{q=0}^{c} \exp(\beta_{ql0} + \beta_{ql1} t)} \tag{4.28}$$

with the unknown coefficients $\beta_{..0}$ and $\beta_{..1}$. For the reason of normalization, the coefficients β_{0l0} and β_{0l1} will be set to zero without the loss of generality, i.e. $\beta_{0l0} \equiv 0 \equiv \beta_{0l1}$ for all $l = 0, \ldots, c$. The model has been estimated via Fisher scoring, presented in chapter 3.1.

4.3.4 Conclusion

The analysis concludes that for people with life-time careers, the well functioning second pillar pension funds provide adequate old age retirement income. A reduction in the state pension is likely to increase participation through postponed (early) retirement, without aggravating poverty among pensioners. More generally, a budget-neutral reduction of both the pension entitlement and income tax rates would substantially reduce the net present value of the tax burden over the working life, in particular for lower-income earners.

Interestingly, the introduction of a flat tax would help to decrease the tax burden over the life cycle substantially both for low and for high income groups,

while it would increase it only moderately for medium income households. Such a shift in the life cycle tax burden would raise total hours worked by 6-8%. Furthermore, the introduction of the flat tax would change the shape of the life cycle tax burden with respect to the income, such that the function would be monotonically increasing instead of being u-shaped. This would be from an idealistic point of view a much "fairer" tax scheme. The introduction of a 34% to 37% flat tax rate, would both increase the participation rate and increase the hours worked, and a VAT increase of 2.9 percentage points to 0.5 percentage points would balance the federal budget.

4.3.5 Outlook

In the model by Ernst and Teuber (2008) and Ernst and Teuber (2012), the productivity age profile was exogenously given, although estimated with a state-of-the-art regression technique estimated. However, an individual might be incentivized by the tax regime to increase his own productivity either by education, on-the-job training, basic and advanced training, or re-training. Obviously, education is competing with labor and free-time which lowers not only the wage but also the utility due to less free time. Basic training might be the first choice for young individuals to increase the educational level and therefore the human capital. Otherwise, the individual will be furthermore rewarded by wage. However, one should remark that students might "finance" their life not only by intertemporal borrowing but, also with wages earned by low-skilled jobs. Advanced training, however, might be an additional accelerator of human capital to outpace the peers and climb up the productivity ladder. The decision to be lazy at the job, on the other hand, leads to a relative decrease of productivity although not necessarily on the absolute level. Unemployed people might not only choose to participate in the working life again via a job search, but to invest in their human capital either via second-chance education or re-training. However, the tax regime,

the accumulated capital stock, and accumulated human capital/productivity is critical to the optimal path of consumption, education, and work time.

Ludwig and Vogel (2010) showed in their two-period overlapping generation model that education effort increases with aging, such that the question arises if this is equivalent to a postponed (early) retirement in the case of an introduction of a flat tax regime. Peterman (2011) analyzes two models in which education is endogenously chosen: In the first model, the individuals are learning-by-doing while in the second model they are learning-or-doing. In the learning-by-doing model, the shift from the current US tax scheme to a flat capital and wage tax leads to on average higher capital taxes (21.5% vs. 17.6%) and on average lower labor taxes (23.3% vs. 24.4%), such that the GDP increases by 0.6%, the capital stock by 2.5%, and average hours by 0.2%. In the case of learning-or-doing, the shift to a flat tax would lead to on average lower capital taxes (14.3% vs. 15.0%) and on average higher labor taxes (24.3% vs. 23.5%), but the GDP increases by 2.7% while the capital stock increases by 5.4% and average worked hours by 0.6%.

The question about the optimal tax regimes in the presence of endogenously education, however, increases the complexity, as one can see in the models by Ludwig and Vogel (2010) and Peterman (2011). A change of the tax regime does not only have an effect on the output, accumulated capital stock, and labor market, but also on the human capital due to the effects on educational efforts. This means that an imprudent tax regime shift might negatively impact the economy even though the new law was implemented for the best.

5 Long-Term Centers, Bivariate Cycles, and Generalized Loops

In this chapter an approach to model multivariate time series will be discussed. For this purpose it will be assumed that the trajectory of the time series follows a circular or elliptical path over time. For detrended time series the trajectory path should be centered around the point of origin, i.e. (0, 0), for a two-dimensional time series analysis. In the more general form, it will be assumed that the circular function $\mathbf{g}$ fluctuates around the center $\boldsymbol{\gamma}$. Instead of modelling the observed time series $\mathrm{E}(\tilde{\mathbf{y}}) = \boldsymbol{\gamma} + \mathbf{g}$, the detrended time series $\mathrm{E}(\mathbf{y}) = \mathrm{E}(\tilde{\mathbf{y}} - \boldsymbol{\gamma}) = \mathbf{g}$ will be preferred.

In chapter 5.1, the two-dimensional model will be discussed in more depth similar to the work of Kauermann, Teuber, and Flaschel (2012). The model will be extended in chapter 5.2 to be able to describe elliptical trajectories. This model can be used to extend the basic model in a multivariate approach so that more than two time series can be modelled simultaneously. In chapter 5.3, the extension of the two-dimensional loop model will be extended to a three-dimensional approach. Similar to chapter 5.2, the three-dimensional model can be generalized to model more than three time series simultaneously.

5.1 Basic Bivariate Loops Model

In this chapter two time series will be modelled simultaneously, for which it will be assumed that the two-dimensional trajectory follows a circular path over time. For the sake of simplicity a model with the center $\boldsymbol{\gamma} = \mathbf{0}$ will be discussed first. Rewriting the unknown functions in polar coordinates leaves the angle and the radius as unknown functions over time, such that

a penalized spline approach will be used to allow the circles to vary over time in terms of duration and magnitude. Due to the fact that a (penalized) likelihood function will be maximized, the estimation and properties of estimates can be calculated with reference to the maximum likelihood theory. Linking the penalized spline model with the Generalized Linear Mixed Model approaches allows us to determine the smoothing parameters via REML estimation, shown in chapter 3. The confidence regions and the handling of correlated errors will be discussed. Furthermore, a center will be introduced that allows the long-term trend of the time series to be estimated simultaneously. The approach for the long-term trend can be treated as a generalized form of the Hodrick-Prescott filter, but this approach is even more flexible. Two simulations will show the capability of the bivariate cycle approach which is quite promising for the analysis of business cycles as one can see in chapter 6.

5.1.1 Modelling Bivariate Cycles using Penalized Splines

Assume the data points $(y_{t1},\, y_{t2})$ will be observed in pairs with t as index referring to the time point. Furthermore, the data points are assumed to be noisy observations of a smooth two-dimensional function $\mathbf{g}(t) = (g_1(t),\, g_2(t))^T$, where $\mathbf{g}$ is smooth in the following sense. The trajectory $\mathbf{g}$ follows loops or circles around the origin, and both the velocity as well as the radius have no rapid changes. In particular this means, that locally and ignoring the implicit role of t, $g_1(\cdot)$ is a smooth function of $g_2(\cdot)$ and vice versa. More precisely we formulate $\mathbf{g}(t)$ in the polar coordinate functions

$$\begin{aligned} \text{radius:} \quad \rho(t) &= \sqrt{g_1(t)^2 + g_2(t)^2} && (5.1) \\ \text{angle:} \quad \phi(t) &= \arctan\left(\frac{g_2(t)}{g_1(t)}\right) \\ &+ 1_{\{g_1(t)<0\}}\left[1_{\{g_2(t)>0\}} - 1_{\{g_2(t)<0\}}\right]\pi && (5.2) \end{aligned}$$

which are smooth functions in t, where $1_{\{.\}}$ is the indicator function. Clearly, smoothness of $\rho(t)$ refers to smooth changes of the radius while smoothness of $\phi(t)$ means circular smoothness with jumps every 2π. Retransformation allows the writing of $\mathbf{g}(t)$ as

$$\mathbf{g}(t) = \Big(\rho(t)\cos\phi(t) \quad \rho(t)\sin\phi(t)\Big)^T \tag{5.3}$$

Assuming that $\mathbf{y}_t = (y_{t1}, y_{t2})^T$ are noisy observations of $\mathbf{g}(t)$, i.e.

$$\mathbf{y}_t \;=\; \mathbf{g}(t) + \boldsymbol{\epsilon}_t \tag{5.4}$$

with $\boldsymbol{\epsilon}_t = (\epsilon_{t1},\, \epsilon_{t2})^T$ as residuals. For simplicity, and to make the machinery of estimation running, the residuals $\boldsymbol{\epsilon}_t$ are assumed to be independent over time, but it seems necessary to allow for correlation between ϵ_{t1} and ϵ_{t2}. With normality assumed, denote $\boldsymbol{\epsilon}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}_\epsilon)$ with $\boldsymbol{\Sigma}_\epsilon$ as covariance matrix.

The functions $\rho(t)$ and $\phi(t)$ are estimated using a penalized spline approach in the style of Eilers and Marx (1996) and Ruppert, Wand, and Carroll (2003), which is shown in chapter 3.2 and 3.3, respectively. Setting therefore

$$\rho(t) = \exp\left(\tilde{\rho}(t)\right) = \exp\left\{\mathbf{B}_\rho(t)\mathbf{b}_\rho\right\} \tag{5.5}$$

where $\mathbf{B}_\rho(t)$ is a spline basis built over the support of t. The $\exp\{\cdot\}$ link in (5.5) is used for technical reasons to ensure a positive radius and $\tilde{\rho}(t)$ is the linear combination of the splines, i.e. $\tilde{\rho}(t) = \mathbf{B}_\rho(t)\mathbf{b}_\rho$. The spline basis is chosen in a rich manner with knots for spline functions placed every 4-5 observed time points. A more theoretical investigation on how many spline functions should be chosen asymptotically is provided for example in Ruppert (2002). In principle, the choice of the basis functions in $\mathbf{B}_\rho(\cdot)$ is left to the user and any spline shape function could be used. For simplicity, both in terms of numerical behavior and notation, a B-spline basis of third order as introduced in de Boor (1978) and in chapter 3.3 will be preferred. The B-spline basis is

built from piecewise polynomials connected at knots $\tau_0 < \tau_1 < \ldots < \tau_K$. In this work, equidistant knots covering the support of t have been used.

The change of the angle $\phi(t)$ will be modelled analog to the radius which is accommodated by setting

$$\phi(t) = \text{mod}\left(\tilde{\phi}(t)\right) = \text{mod}\Big(\mathbf{B}_\phi(t)\mathbf{b}_\phi\Big) \tag{5.6}$$

where $\text{mod}(x) = 2\pi\left(\frac{x}{2\pi} - \left\lfloor\frac{x}{2\pi}\right\rfloor\right)$ and $\lfloor x \rfloor$ returns the smallest integer value of x. Again $\tilde{\phi}(t)$ is the linear combination of the splines. Note that $\text{mod}(\cdot)$ is used for graphical reasons only and the discontinuity is not a technical problem. In fact, we have for instance $\sin(\phi(t)) = \sin(\tilde{\phi}(t))$. The spline basis $\mathbf{B}_\phi(t)$ in equation (5.6) can in principle be chosen differently from $\mathbf{B}_\rho(t)$, but to keep the procedure simple $\mathbf{B}_\phi(t) = \mathbf{B}_\rho(t)$ will be preferred.

Assuming normality for the residuals, the log likelihood can be written as

$$l(\mathbf{b}, \mathbf{\Sigma}_\epsilon) = -\frac{n}{2}\,\log|\mathbf{\Sigma}_\epsilon| - \frac{1}{2}\sum_{i=1}^{n}\left(\mathbf{y}_i - \mathbf{g}(t_i)\right)^T \mathbf{\Sigma}_\epsilon^{-1}\left(\mathbf{y}_i - \mathbf{g}(t_i)\right) \tag{5.7}$$

with $\mathbf{b} = (\mathbf{b}_\rho^T, \mathbf{b}_\phi^T)$ and $\mathbf{g}(\cdot)$ as defined in equation (5.3).

Simple parameter maximization of the likelihood in equation (5.7) would provide unsatisfactory estimates since bases $\mathbf{B}_\rho(t)$ and $\mathbf{B}_\phi(t)$ were chosen to be high dimensional and the corresponding estimates would be jagged. Therefore, a penalized fit is instead pursued, by imposing a penalty on $\mathbf{b}_\rho$ and $\mathbf{b}_\phi$. This is achieved by maximizing the penalized likelihood

$$l_P(\mathbf{b}, \mathbf{\Sigma}_\epsilon; \lambda_b) = l(\mathbf{b}, \mathbf{\Sigma}_\epsilon) - \frac{1}{2}\lambda_\rho \mathbf{b}_\rho^T \mathbf{D}_\rho \mathbf{b}_\rho - \frac{1}{2}\lambda_\phi \mathbf{b}_\phi^T \mathbf{D}_\phi \mathbf{b}_\phi \tag{5.8}$$

with $\lambda_b = (\lambda_\rho, \lambda_\phi)$ as penalty parameters and $\mathbf{D}_\rho$ and $\mathbf{D}_\phi$ as penalty matrices.

As has been suggested in Eilers and Marx (1996), a smooth fit is achieved if spline coefficients of adjacent B-splines are of the same order. This is achieved by imposing a penalty on first or higher order differences of the elements on $\mathbf{b}_\rho$ and $\mathbf{b}_\phi$. In the simplest case we penalize $b_{\rho l} - b_{\rho l-1}$, which can be written in matrix form as $\mathbf{L}\mathbf{b}_\rho$ with $\mathbf{L}$ as $(p-1) \times p$ dimensional contrast matrix where p is the dimension of $\mathbf{b}_\rho$. Setting now $\mathbf{D}_\rho = \mathbf{L}_\rho^T \mathbf{L}_\rho$ leads to the penalty matrix in equation (5.8). The same applies to the construction of $\mathbf{D}_\phi$.

Statistical properties of the estimate as well as optimization with respect to the smoothing parameter λ_ρ and λ_ϕ are given in the following sections.

5.1.2 Estimation and Properties of Estimates

Reformulating the model in equation (5.4) by defining $\mathbf{B}(t) = \text{diag}(\mathbf{B}_\rho(t), \mathbf{B}_\phi(t))$ with $\text{diag}(\cdot)$ as block diagonal matrix, this yields the linear predictor as $\boldsymbol{\eta}(t) = \mathbf{B}(t)\mathbf{b}$. We denote the derivative of $\mathbf{g}(\cdot)$ with respect to $\boldsymbol{\eta}$ by

$$\nabla \mathbf{g}(t) = \rho(t) \begin{pmatrix} \cos\phi(t) & \sin\phi(t) \\ -\sin\phi(t) & \cos\phi(t) \end{pmatrix}. \tag{5.9}$$

which allows us to write the first derivative of equation (5.8) as

$$\frac{\partial l_P(\mathbf{b}, \lambda_b)}{\partial \mathbf{b}} = \sum_{i=1}^{n} \mathbf{B}^T(t_i) \nabla \mathbf{g}(t_i) \boldsymbol{\Sigma}_\epsilon^{-1} (\mathbf{y}_{t_i} - \mathbf{g}(t_i)) - \mathbf{D}(\lambda_b)\mathbf{b} = \mathbf{0}, \tag{5.10}$$

where $\mathbf{D}(\lambda_b)$ is a block diagonal of the form $\text{diag}(\lambda_\rho \mathbf{D}_\rho, \lambda_\phi \mathbf{D}_\phi)$. Solving $\partial l_P(\cdot)/\partial \mathbf{b} = \mathbf{0}$ provides the penalized estimate which can be calculated in the ordinary way using a Newton Raphson procedure (or Fisher scoring).

Accordingly, the penalized Fisher matrix results to

$$\begin{aligned} \mathbf{I}(\mathbf{b}, \lambda_b) &= -\text{E}\left(\frac{\partial^2 l_P(\mathbf{b}, \lambda_b)}{\partial \mathbf{b} \partial \mathbf{b}^T} \right) \\ &= \sum_{i=1}^{n} \mathbf{B}^T(t_i) \nabla \mathbf{g}(t_i) \boldsymbol{\Sigma}_\epsilon^{-1} \nabla \mathbf{g}(t_i)^T \mathbf{B}(t_i) + \mathbf{D}(\lambda_b). \end{aligned} \tag{5.11}$$

Moreover, conventional likelihood theory can be used to derive asymptotic properties for $\hat{\mathbf{b}}$, keeping the penalty parameter fixed. Under the technical assumption that design points t_i at which observations are collected become dense the variance of the estimate is asymptotically given by

$$\mathrm{Var}\begin{pmatrix}\hat{\tilde{\rho}}(t_0)\\ \hat{\tilde{\phi}}(t_0)\end{pmatrix} = \mathbf{B}(t_0)\mathbf{I}\left(\hat{\mathbf{b}}, \lambda_b\right)^{-1}\mathbf{I}\left(\hat{\mathbf{b}}; \lambda_b = \mathbf{0}\right)\mathbf{I}\left(\hat{\mathbf{b}}, \lambda_b\right)^{-1}\mathbf{B}^T(t_0). \quad (5.12)$$

Based on this result, the variance for radius and angle estimates is then obtained by the delta rule

$$\mathrm{Var}\left(\hat{\rho}(t), \hat{\phi}(t)\right) \approx \hat{\mathbf{C}}(t)\mathrm{Var}\left(\hat{\tilde{\rho}}(t), \hat{\tilde{\phi}}(t)\right)\hat{\mathbf{C}}(t)^T \quad (5.13)$$

with $\mathbf{C}(t) = \mathrm{diag}\left(\exp\left(\hat{\tilde{\rho}}(t)\right), 1\right)$, where, as motivated above, the mod$(\cdot)$ function is ignored. Accordingly, the variance for the cycle estimate $\mathbf{g}(t)$ results as

$$\mathrm{Var}(\hat{\mathbf{g}}(t)) \approx \hat{\mathbf{G}}(t)\mathrm{Var}(\hat{\tilde{\rho}}(t), \hat{\tilde{\phi}}(t))\hat{\mathbf{G}}(t)^T \quad (5.14)$$

with

$$\hat{\mathbf{G}}(t) = \hat{\rho}(t)\begin{pmatrix}\cos\hat{\phi}(t) & -\sin\hat{\phi}(t)\\ \sin\hat{\phi}(t) & \cos\hat{\phi}(t)\end{pmatrix} \quad (5.15)$$

where the results follow directly in the line of Ruppert, Wand, and Carroll (2003).

5.1.3 Numerical and Practical Adjustments

For each point in time, the estimates and confidence intervals for the fitted functions $g_1(t)$ and $g_2(t)$ will be obtained. However, the confidence regions for the fitted two-dimensional curves $(g_1(t), g_2(t))$ are of greater interest. These are achieved using the asymptotic arguments from above and constructing a confidence ellipse at time point t through

$$\mathrm{CR}_g(t) \;=\; \left\{\mathbf{y} \in \mathbb{R}^2 | (\mathbf{y} - \hat{\mathbf{g}}(t))^T \mathrm{Var}(\hat{\mathbf{g}}(t))^{-1}(\mathbf{y} - \hat{\mathbf{g}}(t)) \leq \chi^2_{2,\,0.95}\right\} (5.16)$$

with $\chi^2_{2,\,0.95}$ as 95% Quantile of a χ^2 distribution with two degrees of freedom.

One should note that the confidence ellipses are constructed pointwise and a global confidence level for CR is therefore not easily available. This is, however, a standard problem in smoothing. Moreover, the confidence ellipse does not mirror the variability due to the estimation of the smoothing parameter. For simplicity, these two issues will not be investigated in more depth here, but see also Mao and Zhao (2003), or Härdle and Marron (1991) for a more theoretical consideration of these points.

For time-dependent data it is generally difficult to distinguish between trend and correlation. For P-spline smoothing it has been shown in Krivobokova and Kauermann (2007) that residual correlation in a normal smoothing model has only a weak influence on the resulting fitted trend of Maximum Likelihood or Restricted Maximum Likelihood (REML) smoothing parameter selection. Kauermann, Teuber, and Flaschel (2012) conjecture that this result also holds for the non-normal model fitted here, but don't have formal proof. Instead, they exemplify the point with some simulations. In general, of course, a unique decomposition of trend and correlation is impossible. It should also be noted that in principle a two-step fitting can be pursued. First, a mean structure can be fitted which is then used to estimate the temporal correlation from the residuals. This is again used to recalculate both, the fit as well as the smoothing parameter.

5.1.4 Generalized Linear Mixed Models and Laplace Approximation

It has shown to be advantageous, both in terms of numerics and theory, to link spline smoothing with linear mixed models; see for instance chapter 3. For penalized spline smoothing this connection has been demonstrated in Wand

(2003), as shown in chapter 3. This idea was extended here by formulating the penalization as an *a priori* distribution on the spline coefficients. This is available with the following reformulation:

$$\mathbf{y}_t|\mathbf{b} \sim N\left(\mathbf{g}(t), \mathbf{\Sigma}_\epsilon\right) \qquad \mathbf{Lb} \sim N\left(\mathbf{0}, \mathbf{\Lambda}_b\right) \tag{5.17}$$

with $\mathbf{g}(t)$ as defined in equation (5.3) and $\mathbf{L} = \text{diag}\left(\mathbf{L}_\rho, \mathbf{L}_\lambda\right)$ and $\mathbf{\Lambda}_b = \mathbf{\Lambda}_b(\lambda_b)$ as block diagonal having $\mathbf{I}_{k_\rho}\lambda_\rho^{-1}$ and $\mathbf{I}_{k_\phi}\lambda_\phi^{-1}$ on the diagonal where k_ρ and k_ϕ are the dimensions of spline bases $\mathbf{B}_\rho$ and $\mathbf{B}_\phi$, respectively.

Now, penalty parameter $\lambda = (\lambda_\rho, \lambda_\phi)$ expresses the *a priori* precision, that is $1/\lambda$ gives the a priori variance for spline coefficients treated as random coefficients. Integrating out $\mathbf{Lb}$, the marginal log likelihood is obtained based on the mixed model in equation (5.17), i.e.

$$l_{mm}\left(\mathbf{\Sigma}_\epsilon, \lambda_b\right) = \log \int \frac{1}{|\mathbf{\Lambda}_b|^{1/2}} \exp\{l_p\left(\mathbf{b}, \mathbf{\Sigma}_\epsilon; \lambda_b\right)\} d\mathbf{Lb} \tag{5.18}$$

The objective is now to maximize the marginal likelihood in (5.18) with respect to λ_b and $\mathbf{\Sigma}_\varepsilon$ and predict the spline coefficients $\mathbf{b}$ to achieve a smooth fit.

Note that maximization with respect to λ_b provides an estimate for the amount of penalization required. Apparently, due to the non-linear link used for the mean structure, the likelihood in equation (5.18) does not yield an analytic solution. Instead, a Laplace approximation can be used in the line of Breslow and Clayton (1993) or Lindstrom and Bates (1990). In the following, it is shown that the penalized fit from above is equivalent to a posterior mode estimate in the mixed model.

First, the equivalence between penalized spline smoothing with B-splines and mixed models for the simple smoothing model

$$\mathrm{E}\left(y|\, t\right) = g\left(\mathbf{B}(t)\mathbf{b}\right) \tag{5.19}$$

with $\mathbf{B}(t)$ as B-spline basis of dimension k and order p, will be derived. Extensions to the bivariate fitting routine described in section 5.1.1 are straight forward. Let $\mathbf{u} := \mathbf{Lb}$ with $\mathbf{L}$ as difference matrix of order p such that $\mathbf{L} \in \mathbb{R}^{k\times(k-p)}$. Completing $\mathbf{L}$ by adding linearly independent rows such that

$$\tilde{\mathbf{u}} = \begin{pmatrix} \mathbf{u}_0 \\ \mathbf{u} \end{pmatrix} = \begin{pmatrix} \mathbf{L}_0 \\ \mathbf{L} \end{pmatrix} \mathbf{b} = \tilde{\mathbf{L}}\mathbf{b} \tag{5.20}$$

with $\tilde{\mathbf{L}}$ invertible. For fitting the penalty $\lambda \mathbf{b}^T \mathbf{D} \mathbf{b}$ will be imposed on $\mathbf{b}$, where $\mathbf{D} = \mathbf{L}^T\mathbf{L}$. This penalty will be comprehended as *a priori* normal distribution for $\mathbf{u} = \mathbf{Lb} \sim N(\mathbf{0}, \mathbf{I}_{(k-p)}/\lambda)$ with $\mathbf{I}_{k-p}$. Integrating out $\mathbf{u}$ leaves the remaining (unpenalized) parameter $\mathbf{u}_0$. This leads to the marginal likelihood for $\mathbf{u}_0$ and λ given through

$$l_{mm}\left(\lambda, \mathbf{u}_0\right) = \log \int \exp\left(l_p(\mathbf{u}, \lambda)\right) d\mathbf{u} \tag{5.21}$$

with $l_p(\mathbf{u}, \lambda)$ as penalized log likelihood defined through

$$\log\left\{ f\left(y|\mathbf{u}\right) \varphi\left(\mathbf{u}, I_{(k-p)}/\lambda\right)\right\} \tag{5.22}$$

where $f(y|\mathbf{u})$ is the density of y given the linear predictor $\mathbf{B}(t)\mathbf{b} = \mathbf{B}(t)\tilde{\mathbf{L}}^{-1}\tilde{\mathbf{u}}$ and $\varphi(\cdot)$ is the normal density.

Laplace approximation with respect to $\mathbf{u}$ and maximization with respect to the remaining parameter $\mathbf{u}_0$ leads to the maximized marginal log likelihood

$$\begin{aligned} l_{mm}(\lambda, \hat{\mathbf{u}}_0) &\approx l_p\left(\hat{\mathbf{b}}, \lambda\right) - \frac{1}{2}\log\left|\frac{I_{(k-p)}}{\lambda}\right| \\ &\quad -\frac{1}{2}\log\left|\left(0, I_{(k-p)}\right)\left(\tilde{\mathbf{L}}^{-1}\right)^T \frac{\partial^2 l_p\left(\hat{\mathbf{b}}, \lambda\right)}{\partial \mathbf{b} \partial \mathbf{b}^T} \tilde{\mathbf{L}}^{-1} \begin{pmatrix} 0 \\ I_{(k-p)} \end{pmatrix}\right| \\ &\approx l_p(\hat{\mathbf{b}}, \lambda) - \frac{1}{2}\log\left|\mathbf{F}\left(\hat{\mathbf{b}}, \lambda\right) \frac{I_{(k-p)}}{\lambda}\right| \end{aligned} \tag{5.23}$$

with

$$\mathbf{F}\left(\mathbf{b},\lambda\right)=\left(0,I_{(k-p)}\right)\left(\tilde{\mathbf{L}}^{-1}\right)^{T}\mathbf{I}\left(\mathbf{b},\lambda\right)\tilde{\mathbf{L}}^{-1}\begin{pmatrix}0\\I_{(k-p)}\end{pmatrix}. \tag{5.24}$$

The latter simplification results by replacing the second order derivative by the Fisher information in equation (5.11). It is also not difficult to show that $\hat{\mathbf{b}}$ is the solution to the penalized score given in equation (5.10) with $\hat{\mathbf{u}}=\left(\hat{\mathbf{u}}_0^T,\hat{\mathbf{u}}^T\right)=\tilde{\mathbf{L}}\hat{\mathbf{b}}$.

Differentiating equation (5.23) with respect to λ yields

$$0=-\hat{\mathbf{b}}^T\mathbf{L}^T\mathbf{L}\hat{\mathbf{b}}+\frac{(k-p)}{\lambda}-\operatorname{tr}\left(\mathbf{F}^{-1}\left(\hat{\mathbf{b}},\lambda\right)\right). \tag{5.25}$$

To simplify numerics and since $\mathbf{F}^{-1}\left(\mathbf{b},\lambda\right)$ is of order $O(n^{-1})$ the formula will be approximated with

$$\operatorname{tr}\left(\mathbf{F}^{-1}\left(\mathbf{b},\lambda\right)\right)\approx\operatorname{tr}\left(\left(\left(\tilde{\mathbf{L}}^{-1}\right)^{T}\mathbf{I}\left(\mathbf{b},\lambda\right)\tilde{\mathbf{L}}^{-1}\right)^{-1}\right)=\operatorname{tr}\left(\mathbf{I}^{-1}\left(\mathbf{b},\lambda\right)\mathbf{L}^T\mathbf{L}\right) \tag{5.26}$$

which suggests the estimating equation

$$\hat{\lambda}=\frac{\hat{\mathbf{b}}^T\mathbf{D}\hat{\mathbf{b}}+\operatorname{tr}\left(\mathbf{I}^{-1}\left(\hat{\mathbf{b}},\lambda\right)\mathbf{D}\right)}{k-p} \tag{5.27}$$

with $\mathbf{D}=\mathbf{L}^T\mathbf{L}$. It is thereby worth pointing out that equation (5.29) does not provide an analytic solution since the right-hand-side also depends on λ_ρ through $\hat{\mathbf{b}}_\rho$. Equation (5.29) can however be used in a backfitting style such that $\hat{\mathbf{b}}$ will be estimated through equation (5.10) by keeping λ fixed. In the next step, $\hat{\mathbf{b}}$ is considered as fixed and update λ through equation (5.29). Iteration between these two steps mirrors the backfitting iterations; see also Krivobokova and Kauermann (2007) for a justification of this algorithm as Newton procedure.

Applying equation (5.29), the smoothing parameter $\lambda = (\lambda_\rho, \lambda_\phi)$ can be chosen to maximize the (Laplace approximated) marginal likelihood such that the smoothing parameter for the radius is implicit given by

$$\frac{1}{\hat{\lambda}_\rho} = \frac{\operatorname{tr}\left\{(\mathbf{I}(\boldsymbol{b}, \lambda)^{-1})_{\rho\rho}\, \mathbf{D}_\rho\right\} + \hat{\mathbf{b}}_\rho^T \mathbf{D}_\rho \hat{\mathbf{b}}_\rho}{k_\rho} \tag{5.28}$$

and the smoothing parameter for the angle by

$$\frac{1}{\hat{\lambda}_\phi} = \frac{\operatorname{tr}\left\{(\mathbf{I}(\boldsymbol{b}, \lambda)^{-1})_{\phi\phi}\, \mathbf{D}_\phi\right\} + \hat{\mathbf{b}}_\phi^T \mathbf{D}_\phi \hat{\mathbf{b}}_\phi}{k_\phi} \tag{5.29}$$

respectively.

Finally, based on the Laplace approximation likelihood an estimate for $\boldsymbol{\Sigma}_\epsilon$ is defined through

$$\hat{\boldsymbol{\Sigma}}_\epsilon = \frac{\sum_{i=1}^{n} \{\mathbf{y}_i - \hat{\mathbf{g}}(t_i)\}\{\mathbf{y}_i - \hat{\mathbf{g}}(t_i)\}^T}{n} + O(n^{-1}). \tag{5.30}$$

which is well-known.

5.1.5 Short-Term Fluctuations and Long-Term Trends

In the previous sections, the center for cycles had been set to zero, which implies that the series y_{t1} and y_{t2} are stationary without any long-term trend. Apparently this is a stringent assumption which will be weakened now to a more practical situation. To do so, the model in equation (5.4) will be replaced by

$$\mathbf{y}_t = \boldsymbol{\gamma}(t) + \mathbf{g}(t) + \boldsymbol{\epsilon}_t \tag{5.31}$$

where $\boldsymbol{\gamma}(t) = (\gamma_1(t),\, \gamma_2(t))^T$ is the long-term trend around which $\mathbf{g}(t)$ is oscillating.

In equation (5.31), the mean structure had been decomposed into a long-term trend $\boldsymbol{\gamma}(t)$ and the shorter phase oscillation $\mathbf{g}(t)$. In time series analysis the decomposition of trends and seasonal effects is well established; see for instance Brockwell and Davis (1987). Unlike in classical time series, the phase lengths of these cycles are unknown and the objective is to estimate these from the data.

The canonical candidate for long-term trend estimation is the Hodrick and Prescott (1997) filter. It leaves, however, the unsatisfactory requirement of choosing a penalty parameter λ, which Hodrick and Prescott are setting to $\lambda = 1600$. From a statistical point of view fixing the smoothing parameter in advance is unsatisfactory and a data-driven criterion seems preferable. Therefore, a smooth approach is pursued by fitting $\boldsymbol{\gamma}(t)$ again using penalized spline fitting, such that the long-term trend will be written as

$$\boldsymbol{\gamma}(t) = \begin{pmatrix} \mathbf{B}_{\gamma,1}(t)\boldsymbol{\theta}_{\gamma,1} \\ \mathbf{B}_{\gamma,2}(t)\boldsymbol{\theta}_{\gamma,2} \end{pmatrix} =: \mathbf{B}_{\gamma}(t)\boldsymbol{\theta}_{\gamma} \tag{5.32}$$

where $\mathbf{B}_{\gamma,l}(t)$ are spline bases chosen complex enough to capture the long-term trends for $l = 1$ and 2. Using a B-spline basis for $\mathbf{B}_{\gamma,l}(t)$ with the same knots as for the estimation of $\mathbf{g}(t)$, the spline coefficient $\boldsymbol{\theta}_{\gamma}$ is now estimated in a penalized form with penalty $\boldsymbol{\theta}_{\gamma,l}^T \mathbf{L}_{\gamma}^T \mathbf{L}_{\gamma} \boldsymbol{\theta}_{\gamma,l}$ for $l = 1$, 2, and $\mathbf{L}_{\gamma}$ as difference matrix.

In principle, the penalty could be formulated as a priori normality and the resulting structured mixed model can be fitted. To keep the numerics simple and understandable, a hybrid two-step procedure is undertaken. This means that the long-term trend $\boldsymbol{\gamma}(t)$ will be estimated first, and then the business cycle structure $\mathbf{g}(t)$ will be fitted to the residuals $\tilde{\mathbf{y}}_t = \mathbf{y}_t - \boldsymbol{\gamma}(t)$. This hybrid approach appears justifiable since the objective is the estimation of the

shorter phase structure $\mathbf{g}(t)$ in its dependence on the longer cycle. Therefore, the long-term trend $\boldsymbol{\gamma}_l(t)$ will be fitted component-wise with given penalty parameters $\boldsymbol{\lambda}_\gamma = (\lambda_1, \lambda_2)$, given by

$$\boldsymbol{\gamma}_l(t) = \mathbf{B}_{\gamma,l}(t) \left(\mathbf{B}_{\gamma,l}^T \mathbf{B}_{\gamma,l} + \mathbf{D}_\gamma(\lambda_l)\right)^{-1} \mathbf{B}_{\gamma,l}^T \mathbf{Y}_l = \mathbf{S}_l(\lambda_l)\mathbf{Y}_l \tag{5.33}$$

for $l = 1, 2$ with $\mathbf{Y}_l = (y_{1l}, \ldots, y_{nl})^T$ and $\mathbf{B}_{\gamma,l}$ as matrix built from $b_{\gamma,l}(t_i)$, for $i = 1, \ldots, n$ and $\mathbf{S}_l(\lambda_l)$ as smoothing matrix.

In particular, for the long-term trend any possible correlation among the components of $\mathbf{y}$ will be ignored. The resulting residuals $\tilde{\mathbf{y}}_t$ are assumed to be distributed according to equation (5.17) with $\mathbf{y}_t$ replaced by $\tilde{\mathbf{y}}_t$. In a second step estimate $\hat{\mathbf{g}}(t)$ is obtained based on the observations corrected by the long-term trend.

It remains for us to select appropriate penalty parameters for $\boldsymbol{\lambda}_\gamma$ for which an Akaike-based criterion is proposed. This allows a grid search for $\boldsymbol{\lambda}_\gamma$ to be used such that the minimum value for

$$\begin{aligned} \mathrm{AIC}(\boldsymbol{\lambda}_\gamma) &= n \log \left| \sum_{i=1}^{n} \left(\mathbf{y}_i - \hat{\boldsymbol{\gamma}}(t_i) - \hat{\mathbf{g}}(t_i)\right) \left(\mathbf{y}_i - \hat{\boldsymbol{\gamma}}(t_i) - \hat{\mathbf{g}}(t_i)\right)^T \right| \\ &\quad + 2df\gamma + 2dfg \end{aligned} \tag{5.34}$$

has to be found where $\hat{\boldsymbol{\gamma}}(\cdot)$ is the penalized fit with penalty parameters $\boldsymbol{\lambda}_\gamma$ and $\hat{\mathbf{g}}(\cdot)$ is the fit based on $\tilde{\mathbf{y}}_t$ using the Mixed Model formulation from above. The degrees of freedom of the fits are calculated from

$$df\gamma = \mathrm{tr}\left(\mathbf{B}_\gamma \left(\mathbf{B}_\gamma^T \mathbf{B}_\gamma + \mathbf{D}_\gamma(\boldsymbol{\lambda}_\gamma)\right)^{-1} \mathbf{B}_\gamma^T\right) \tag{5.35}$$

and

$$dfg = \mathrm{tr}\left(\mathbf{I}\left(\hat{\mathbf{b}}, \hat{\boldsymbol{\lambda}}_b\right)^{-1} \mathbf{I}\left(\hat{\mathbf{b}}, \boldsymbol{\lambda}_b = \mathbf{0}\right)\right). \tag{5.36}$$

similar to chapter 3.

5.1.6 Simulation

The performance and capability of the routine described above will be demonstrated with a simulation study. For this purpose an artificial time series will be simulated which is given by the noisy observations of the trajectory around the unit circle, i.e.

$$\tilde{y}_{t1} = \sin(2\pi t) + \epsilon_{t1} \tag{5.37}$$

$$\tilde{y}_{t2} = \cos(2\pi t) + \epsilon_{t2} \tag{5.38}$$

with the independent residuals $\epsilon_{t1}, \epsilon_{t2}$ which are normally distributed, i.e. $\epsilon_{t1} \sim \mathcal{N}(0,\, 0.25^2)$ and $\epsilon_{t2} \sim \mathcal{N}(0,\, 0.25^2)$, and t ranging from 0 to 1 in $n = 200$ equidistant steps. The long-term fluctuations are overlaid with a short-term trend such that

$$y_{t1} = \tilde{y}_{t1} + 0.5\cos(10\pi t) \tag{5.39}$$

$$y_{t2} = \tilde{y}_{t2} + 0.5\sin(10\pi t). \tag{5.40}$$

For the numerical estimations the same B-spline bases of order 3, i.e. $\mathbf{B}_\rho = \mathbf{B}_\phi = \mathbf{B}_{\gamma,1} = \mathbf{B}_{\gamma,2}$, and the same penalty matrices of order 2, i.e. $\mathbf{D}_\rho = \mathbf{D}_\phi = \mathbf{D}_\gamma$ have been used.

In figure 5.1 the simulated data and the corresponding estimates are shown. This figure is therefore organized as follows: The first two plots show the two time series y_{t1} and y_{t2}. The resulting long-term trend estimate is superimposed as a solid line. The final estimate $\hat{\boldsymbol{\gamma}}(t) + \hat{\mathbf{g}}(t)$ is shown with confidence bands for both series. The bottom row shows the observations $(y_{t1},\, y_{t2})$ with the long-term trend $\hat{\boldsymbol{\gamma}}(t)$ (bottom left plot) and the residuals $(\tilde{y}_{t1},\, \tilde{y}_{t2}) = (y_{t1} - \hat{\boldsymbol{\gamma}}_1(t),\, y_{t2} - \hat{\boldsymbol{\gamma}}_2(t))$ with the fitted shorter phase structure $\hat{\mathbf{g}}(t)$. Finally, the two right-hand-side plots show the fitted radius $\hat{\rho}(t)$ (upper right-hand-side plot) and the fitted angle $\hat{\phi}(t)$ (bottom right-hand-side plot).

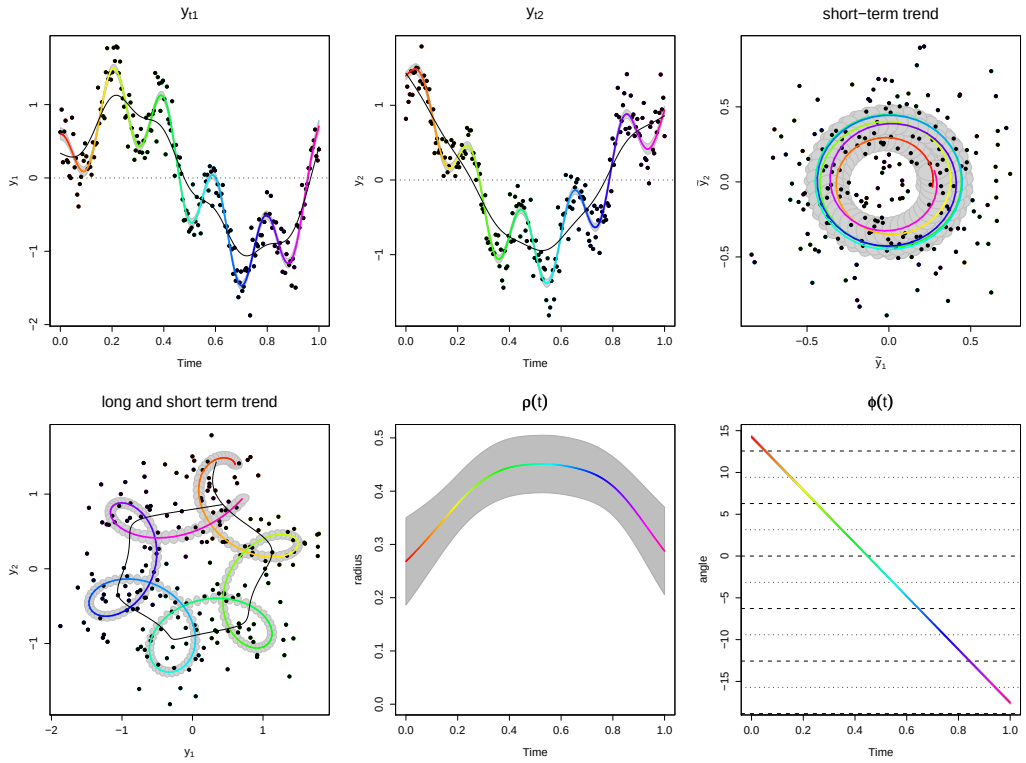

Figure 5.1: Estimation of simulated time series using a circular regression approach: The upper left-hand and middle plot: Observations (black dots), estimated long-term trend (black line), and estimated short-term fluctuations around long-term trend (colored line) with confidence regions (grey-shaded area) over time for first and second time series, respectively. Upper right hand plot: Detrended observations (black dots), trajectory of short-term trend (colored line), and their confidence regions (grey shaded area). The lower left-hand plot: Observations (black dots), long-term trend (black line), trajectory of short-term trend around long-term trend (colored line), and confidence regions (grey-shaded area). Middle and right-hand plot: Radius and angle over time (colored line), and their confidence regions (grey-shaded area).

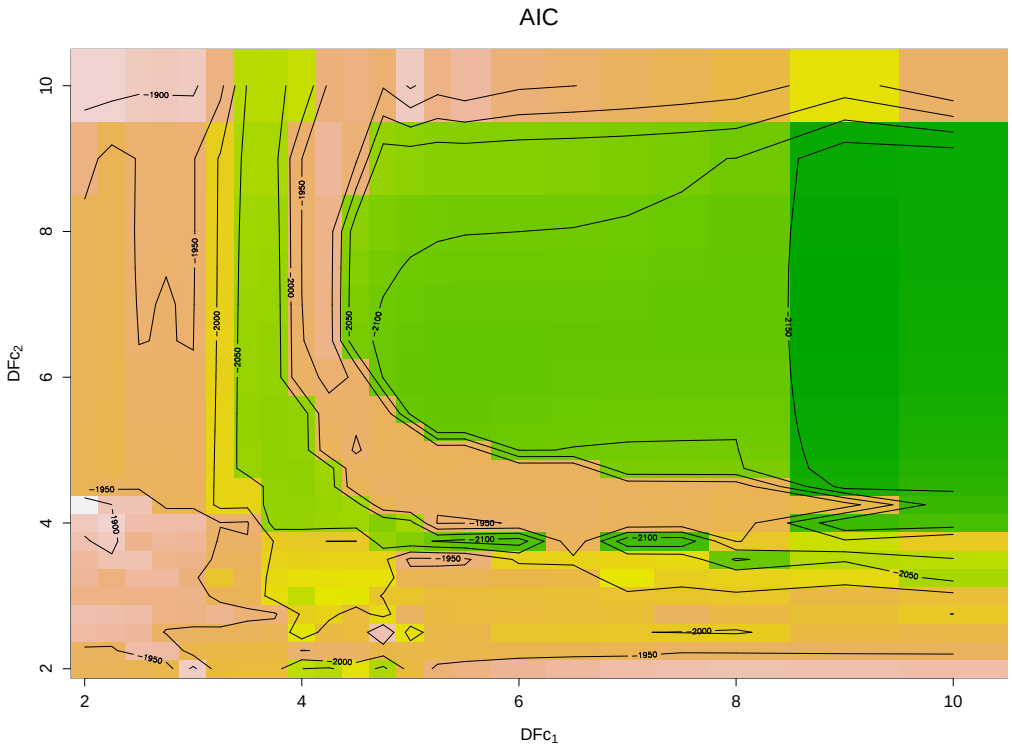

Figure 5.2: Colored contour plot of AIC function given the degrees of freedom of long-term trend functions: For each pair of degrees of freedoms for the first and second long-term trend of the time series the AIC function value is shown in the contour plot. Color scheme: Green (low range), orange to brown (middle range), brown to white (high range) based on geographical contour maps.

Given the estimation in figure 5.1, the separation of long-term trend and the short-term fluctuation seems adequate for the data. The smoothing parameters for the long-term trend are thereby selected following the Akaike criterion proposed above. The corresponding shape of AIC(λ_a) is shown ex-

emplarily in figure 5.2. Clearly the existence of two phases is visible in the AIC function, where the value of AIC is given for the resulting degrees of freedom for $\gamma_1(t)$ and $\gamma_2(t)$.

Furthermore, the angle is estimated nearly linearly, indicating a constant rotation speed for $\mathbf{g}(t)$. In the same way, the radius is about constant over time. It can be seen that the true function is reproduced for this simulation, which is promising for the application of real world data.

5.2 Extension of the Bivariate Loops Model

The basic model of the bivariate loop approach will be extended in this chapter. For this purpose, the main findings will be briefly recapitulated in the next section so that the extension to a more general approach is straightforward. The introduction of two additional parameters allows not only the fitting of two-dimensional circular trajectories but also elliptical oscillations, such that the radius and the angle might be from a degrees of freedom perspective less complex functions. A reformulation of this extended model allows us to generalize the model so that multivariate time series can be estimated using the two-dimensional loop approach. Similar to the previous model, the properties of the model and the selection of the smoothing parameters will be discussed before a numerical simulation concludes this chapter.

5.2.1 The Basic Model

Let the observations (y_{i1}, y_{i2}, t_i) with index $i = 1, \ldots, n$ be ordered by the latter time series which represents the time points of the observation, i.e. $\underline{t} := t_1 < \ldots < t_i < \ldots < t_n =: \bar{t}$.

It will be assumed that the first two time series $\mathbf{y}_{\bullet 1} := (y_{11}, \ldots, y_{n1})^T$ and $\mathbf{y}_{\bullet 2} := (y_{12}, \ldots, y_{n2})^T$ are noisy observations of an unknown two-dimensional function $\tilde{\mathbf{g}}$, such that the structural part of the model can be written as

$$\mathrm{E}\left(\begin{pmatrix} y_{i1} \\ y_{i2} \end{pmatrix} | \, t_i\right) \;=\; \tilde{\mathbf{g}}(t_i) := \begin{pmatrix} g_1(t_i) \\ g_2(t_i) \end{pmatrix}. \tag{5.41}$$

In accordance to Kauermann, Teuber, and Flaschel (2012), it will be assumed that the trajectory $\tilde{\mathbf{g}}$ follows a loop over time around the center $\boldsymbol{\gamma}(t) := (\gamma_1(t), \gamma_2(t))^T$, such that the function can be written in polar coordinates as

$$\tilde{\mathbf{g}}(t) \;:=\; \rho(t) \begin{pmatrix} \cos\phi(t) \\ \sin\phi(t) \end{pmatrix} + \begin{pmatrix} \gamma_1(t) \\ \gamma_2(t) \end{pmatrix} \tag{5.42}$$

with ρ representing the radius and ϕ the angle. Although the center plays a key role in describing a loop, the center is set to be the point of origin for the sake of simplicity. One can think of using the transformed data $\tilde{y}_{i1} = y_{i1} - \gamma_1(t_i)$ and $\tilde{y}_{i2} = y_{i2} - \gamma_2(t_i)$ instead of the original data without the loss of generality.

Assuming that the polar coordinate functions can be expressed by the functions

$$\rho(t) = \exp(\mathbf{B}_\rho(t)\boldsymbol{\theta}_\rho) \tag{5.43}$$

and

$$\phi(t) = \mathbf{B}_\phi(t)\boldsymbol{\theta}_\phi \tag{5.44}$$

with $\mathbf{B}_\rho(t)$ and $\mathbf{B}_\phi(t)$ as B-spline bases built over the support of the time $\left[\underline{t}; \bar{t}\right]$. Even though the number of knots and/or the order of the basis can differ, for the practical estimations the same basis will be used.

For the estimation of the parameters $\boldsymbol{\theta}_\rho$ and $\boldsymbol{\theta}_\phi$, the (log) likelihood has to be derived and is given by

$$l(\mathbf{y};\, \boldsymbol{\theta}_\rho,\, \boldsymbol{\theta}_\phi) \;=\; \sum_{i=1}^{n} \log f_Y(\mathbf{y}_{i\bullet};\, \boldsymbol{\theta}_\rho,\, \boldsymbol{\theta}_\phi) \tag{5.45}$$

with $\mathbf{y}_{i\bullet} := (y_{i1},\, y_{i2})^T$, $\mathbf{y} := (\mathbf{y}_{1\bullet}, \ldots, \mathbf{y}_{n\bullet})^T$ and $f_Y(\cdot)$ as the assumed density function for the observations. But instead of maximizing the likelihood directly, the penalized likelihood

$$l_p(\mathbf{y};\, \boldsymbol{\theta}_\rho,\, \boldsymbol{\theta}_\phi,\, \lambda_\rho,\, \lambda_\phi) \;:=\; l(\mathbf{y};\, \boldsymbol{\theta}_\rho,\, \boldsymbol{\theta}_\phi) - \frac{1}{2}\lambda_\rho \boldsymbol{\theta}_\rho^T \mathbf{D}_\rho \boldsymbol{\theta}_\rho - \frac{1}{2}\lambda_\phi \boldsymbol{\theta}_\phi^T \mathbf{D}_\phi \boldsymbol{\theta}_\phi \tag{5.46}$$

is preferred with $\mathbf{D}_\rho$ and $\mathbf{D}_\phi$ as appropriate penalty matrices and λ_ρ and λ_ϕ as the resulting smoothing parameters. Similar to the B-spline basis, the same penalty matrices for the angle and the radius are used for the estimations.

Assuming that the observations $\mathbf{y}_{\bullet 1}$ and $\mathbf{y}_{\bullet 2}$ are normally distributed, then one gets for $f_Y(\cdot)$ the bivariate normal density function, i.e.

$$f_Y(\mathbf{y}) = \frac{1}{(2\pi)^{m/2} |\det \boldsymbol{\Sigma}|^{1/2}} \times \exp\left[-\frac{1}{2}(\mathbf{y} - \tilde{\mathbf{g}})^T \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \tilde{\mathbf{g}})\right] \tag{5.47}$$

with $m = 2$ for this case.

Defining $\mathbf{B}(t_i) := \mathbf{B}_\rho(t_i) \oplus \mathbf{B}_\phi(t_i)$, $\nabla\tilde{\mathbf{g}}(t_i) := \rho(t_i)\mathbf{T}(\phi(t_i))$, $\mathbf{T}(\phi(t_i)) := \begin{pmatrix} \cos\phi(t_i) & -\sin\phi(t_i) \\ \sin\phi(t_i) & \cos\phi(t_i) \end{pmatrix}$, $\tilde{\mathbf{D}}(\boldsymbol{\lambda}) := (\lambda_\rho \mathbf{D}_\rho) \oplus (\lambda_\phi \mathbf{D}_\phi)$ with $\boldsymbol{\lambda} := (\lambda_\rho,\, \lambda_\phi)$ and the operator defined as $\mathbf{A}_{n\times m} \oplus \mathbf{B}_{p\times q} := \begin{pmatrix} \mathbf{A} & \mathbf{0}_{n\times q} \\ \mathbf{0}_{p\times m} & \mathbf{B} \end{pmatrix}$, the score function for $\boldsymbol{\theta} := (\boldsymbol{\theta}_\rho^T,\, \boldsymbol{\theta}_\phi^T)^T$ is given by

$$\mathbf{s}_\theta(\boldsymbol{\theta},\, \boldsymbol{\lambda},\, \boldsymbol{\Sigma}_\epsilon) \;:=\; \sum_{i=1}^{n} \left(\nabla\tilde{\mathbf{g}}(t_i)\mathbf{B}(t_i)\right)^T \boldsymbol{\Sigma}_\epsilon^{-1} \left(\mathbf{y}_{i\bullet} - \tilde{\mathbf{g}}(t_i)\right) - \tilde{\mathbf{D}}(\boldsymbol{\lambda})\boldsymbol{\theta}. \tag{5.48}$$

This leads to the expected second derivative which is equal to the negative (penalized) Fisher information matrix given by

$$\mathbf{F}(\boldsymbol{\theta},\, \boldsymbol{\lambda},\, \boldsymbol{\Sigma}_\epsilon) \;:=\; \sum_{i=1}^{n} \left(\nabla \tilde{\mathbf{g}}(t_i)\mathbf{B}(t_i)\right)^T \boldsymbol{\Sigma}_\epsilon^{-1} \left(\nabla \tilde{\mathbf{g}}(t_i)\mathbf{B}(t_i)\right) + \tilde{\mathbf{D}}(\boldsymbol{\lambda}) \tag{5.49}$$

and a Newton-Raphson procedure can be applied to derive the maximum likelihood estimation for $\boldsymbol{\theta}$, similar to chapter 5.1.

5.2.2 An Extension of the Basic Model

Instead of using the basic model an extension will be presented which can lead to "better" estimations. For this purpose, different intercepts for the radius and the angle for the two time series will be introduced, such that the model is given by

$$\begin{aligned} \mathbf{g}(t) &= \begin{pmatrix} g_1(t) \\ g_2(t) \end{pmatrix} := \begin{pmatrix} \exp(\mathbf{B}_\rho(t)\boldsymbol{\theta}_\rho + \rho_{01})\cos(\phi(t) + \phi_{01}) \\ \exp(\mathbf{B}_\rho(t)\boldsymbol{\theta}_\rho + \rho_{02})\sin(\phi(t) + \phi_{02}) \end{pmatrix} \\ &= \begin{pmatrix} \exp(\mathbf{B}_\rho(t)\boldsymbol{\theta}_\rho)\cos(\phi(t)) \\ \exp(\mathbf{B}_\rho(t)\boldsymbol{\theta}_\rho + \rho_0)\sin(\phi(t) + \phi_0) \end{pmatrix} =: \begin{pmatrix} \rho_1(t)\cos\phi_1(t) \\ \rho_2(t)\sin\phi_2(t) \end{pmatrix} \end{aligned} \tag{5.50}$$

with $\rho_0 := \rho_{02} - \rho_{01}$ as the shift in the radius and $\phi_0 := \phi_{02} - \phi_{01}$ as the shift in the angle.

The main advantage of the model in equation (5.50) in contradiction to the model in equation (5.42) is that the basic model leads to wiggly estimations even though the true radius and the true angle are linear functions, if

(a) the shift in the angle is unequal to zero. A graphical illustration is shown in figure 5.3 for a true shift $\phi_0 = -\frac{\pi}{4}$. The basic model leads to an overly complex fit because the radius and the angle change rapidly over time. The formulation in equation (5.50) describes the elliptic model in a less complex way.

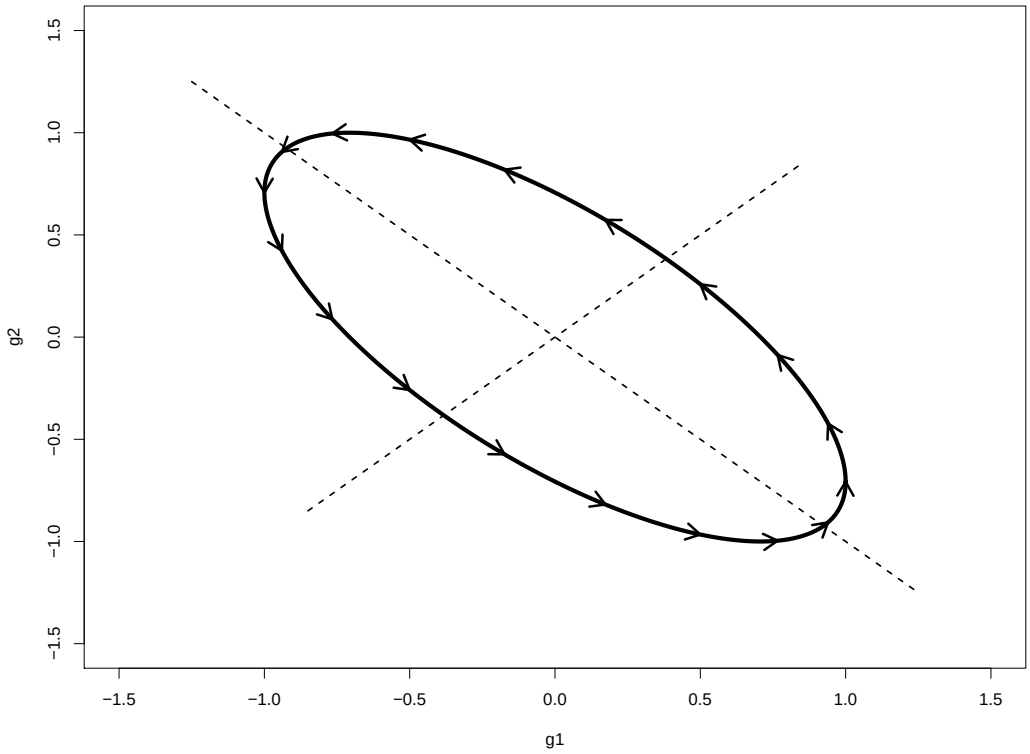

Figure 5.3: Trajectory of shifted circular function explained in example (a): Trajectory follows an elliptical pattern instead of a circular one.

(b) the shift in the radius is unequal to zero, which is illustrated in figure 5.4, where $\rho_0 = \log 2$ is chosen. Similar to (a) the basic model is too complex and the more flexible model leads to a better fit with fewer degrees of freedom.

These two examples above show that a more flexible model is desired resulting by a "validation-by-eye", because the fit seems to be smoother and therefore results in narrower confidence regions.

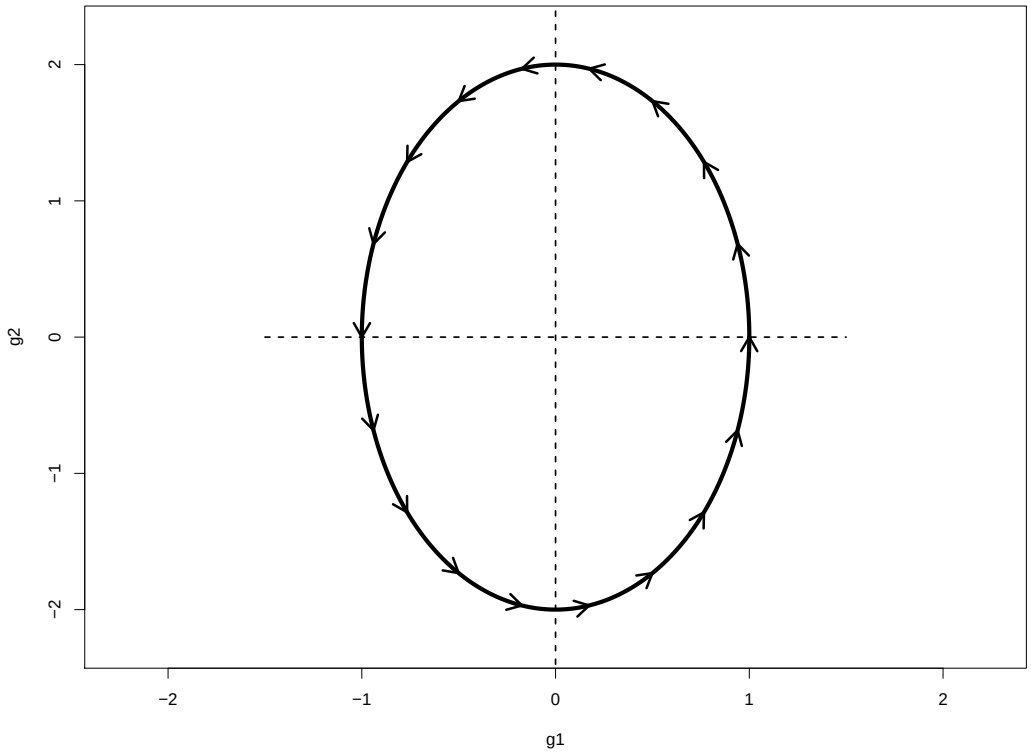

Figure 5.4: Trajectory of shifted circular function explained in example (b): Trajectory follows a prolate circle instead of a round one.

In accordance to the definitions given in section 5.2.1, the definitions

$$\nabla \mathbf{g}(t) := (\rho_1(t) \oplus 0)\, \mathbf{T}(\phi_1(t)) + (0 \oplus \rho_2(t))\mathbf{T}(\phi_2(t)),$$

$$\mathbf{C}(t_i) := \left(\nabla \mathbf{g}(t_i)\mathbf{B}(t_i)\,,\ (0 \oplus 1)\, \nabla \mathbf{g}(t_i)\right) \qquad \text{and}$$
$$\mathbf{D}(\boldsymbol{\lambda}) := \tilde{\mathbf{D}}(\boldsymbol{\lambda}) \oplus \mathbf{0}_{2\times 2}$$

are given with the operator defined as $\mathbf{A}_{n\times m} \oplus \mathbf{B}_{p\times q} := \begin{pmatrix} \mathbf{A} & \mathbf{0}_{n\times q} \\ \mathbf{0}_{p\times m} & \mathbf{B} \end{pmatrix}$. The first derivative with respect to $\boldsymbol{\theta} := \left(\boldsymbol{\theta}_\rho^T,\, \boldsymbol{\theta}_\phi^T,\, \rho_0,\, \phi_0\right)^T$ can be expressed as

$$\mathbf{s}_\theta(\boldsymbol{\theta},\, \boldsymbol{\lambda},\, \boldsymbol{\Sigma}_\epsilon) \;:=\; \sum_{i=1}^{n} \mathbf{C}(t_i)^T \boldsymbol{\Sigma}_\epsilon^{-1} \left(\mathbf{y}_{i\bullet} - \mathbf{g}(t_i)\right) - \mathbf{D}(\boldsymbol{\lambda})\boldsymbol{\theta}. \tag{5.51}$$

and the Fisher matrix is equal to

$$\mathbf{F}(\boldsymbol{\theta},\, \boldsymbol{\lambda},\, \boldsymbol{\Sigma}_\epsilon) \;:=\; \sum_{i=1}^{n} \mathbf{C}(t_i)^T \boldsymbol{\Sigma}_\epsilon^{-1} \mathbf{C}(t_i) + \mathbf{D}(\boldsymbol{\lambda}) \tag{5.52}$$

which are needed for the Fisher scoring.

5.2.3 Generalized Two-Dimensional Loops

Once modelling bivariate time series by loops it seems natural to extend the model to higher dimensions so that more than two time series can be estimated simultaneously. The first suggestion is to project a two-dimensional loop into a higher dimension.

Observing $m+1$ different time series $(y_{i1}, \ldots, y_{im},\, t_i)$ with index $i = 1, \ldots, n$ and $t_i \in [\underline{t},\, \overline{t}]$ as the corresponding time points, for the structural part of the model it will be assumed that

$$\mathrm{E}\left(\begin{pmatrix} y_{i1} \\ \vdots \\ y_{im} \end{pmatrix} \mid t_i \right) = \mathbf{A}\tilde{\mathbf{g}}(t_i) = \mathbf{A}\rho(t_i) \begin{pmatrix} \cos\phi(t_i) \\ \sin\phi(t_i) \end{pmatrix} \tag{5.53}$$

holds.

Without loss of generality, the linear combination can be set for all $i = 1, \ldots, m$ to

$$\mathbf{A}_i \;=\; (a_{i1},\, a_{i2}) = (\exp(\rho_{0i})\cos\phi_{0i}\,,\; -\exp(\rho_{0i})\sin\phi_{0i}) \tag{5.54}$$

such that equation 5.53 can be written as

$$\begin{aligned}
\mathbf{A}\tilde{\mathbf{g}}(t_i) &= \begin{pmatrix} \exp(\rho_{01})\rho(t)\left[\cos\phi_{01}\cos\phi(t) - \sin\phi_{01}\sin\phi(t)\right] \\ \vdots \\ \exp(\rho_{0m})\rho(t)\left[\cos\phi_{0m}\cos\phi(t) - \sin\phi_{0m}\sin\phi(t)\right] \end{pmatrix} \\
&= \begin{pmatrix} \exp(\rho_{01} + \mathbf{B}_\rho(t)\boldsymbol{\theta}_\rho)\cos\left(\phi_{01} + \mathbf{B}_\phi(t)\boldsymbol{\theta}_\phi\right) \\ \vdots \\ \exp(\rho_{0m} + \mathbf{B}_\rho(t)\boldsymbol{\theta}_\rho)\cos\left(\phi_{0m} + \mathbf{B}_\phi(t)\boldsymbol{\theta}_\phi\right) \end{pmatrix} \\
&=: \begin{pmatrix} \rho_1(t)\cos\phi_1(t) \\ \vdots \\ \rho_m(t)\cos\phi_m(t) \end{pmatrix} =: \begin{pmatrix} g_1(t) \\ \vdots \\ g_m(t) \end{pmatrix} =: \mathbf{g}(t)\,. \qquad (5.55)
\end{aligned}$$

Obviously for $m = 2$ time series, the generalized model in equation (5.55) is equal to the extended model in equation (5.50).

The advantage of formulating the projection matrix $\mathbf{A}$ as in equation (5.54) is the possibility of both an appropriate interpretation and a simultaneous estimation.

For the likelihood estimation the definitions $\mathbf{D}(\boldsymbol{\lambda}) := \tilde{\mathbf{D}}(\boldsymbol{\lambda}) \oplus \mathbf{0}_{2m\times 2m}$ and $\nabla\mathbf{gB}(t_i) := \left(\mathbf{A}\nabla\tilde{\mathbf{g}}(t_i)\mathbf{B}(t_i)\,,\ \oplus_{j=1}^{m} g_j(t_i)\,,\ \oplus_{j=1}^{m}(-\rho_j(t_i)\sin\phi_j(t_i))\right)$ are given, such that the first derivative with respect to $\boldsymbol{\theta} := \left(\boldsymbol{\theta}_\rho^T,\ \boldsymbol{\theta}_\phi^T,\ \rho_{01},\ldots,\rho_{0m},\ \phi_{01},\ldots,\phi_{0m}\right)^T$ is given by

$$\mathbf{s}_\theta(\boldsymbol{\theta},\,\boldsymbol{\lambda},\,\boldsymbol{\Sigma}_\epsilon) := \sum_{i=1}^{n} \nabla\mathbf{gB}(t_i)^T\boldsymbol{\Sigma}_\epsilon^{-1}\left(\mathbf{y}_{i\bullet} - \mathbf{g}(t_i)\right) - \mathbf{D}(\boldsymbol{\lambda})\boldsymbol{\theta} \qquad (5.56)$$

and the negative (expected) second derivative as

$$\mathbf{F}(\boldsymbol{\theta},\,\boldsymbol{\lambda},\,\boldsymbol{\Sigma}_\epsilon) := \sum_{i=1}^{n} \nabla\mathbf{gB}(t_i)^T\boldsymbol{\Sigma}_\epsilon^{-1}\nabla\mathbf{gB}(t_i) + \mathbf{D}(\boldsymbol{\lambda}), \qquad (5.57)$$

which can be applied for the Fisher scoring to derive the maximum likelihood estimations. To avoid any confusion given the definitions and the solutions defined in equations (5.56) and (5.57), the dimensions of the variables will be given. Assuming that the B-spline matrix $\mathbf{B}_\phi(t_i)$ and $\mathbf{B}_\rho(t_i)$, are given as $K_\phi \times 1$ and $K_\rho \times 1$, respectively, then the dimension of the square matrix $\mathbf{D}(\boldsymbol{\lambda})$ is $K_\rho + K_\phi + 2m \times K_\rho + K_\phi + 2m$, of the parameter vector $\boldsymbol{\theta}$ is $K_\rho + K_\phi + m + m \times 1$, of the matrix $\nabla \mathbf{g}\mathbf{B}(t_i)$ is $m \times K_\rho + K_\phi + m + m$, of the co-variance matrix $\boldsymbol{\Sigma}$ is $m \times m$ and of the i-th observation $\mathbf{y}_{i\bullet}$ and their expected trend $\mathbf{g}(t_i)$ are $m \times 1$. The dimension of the first derivative with respect to $\boldsymbol{\theta}$, defined in equation (5.56), is therefore $K_\rho + K_\phi + 2m \times 1$ and the dimension of the negative (expected) second derivative, defined in equation (5.57), is therefore $K_\rho + K_\phi + 2m \times K_\rho + K_\phi + 2m$, as one can easily see.

5.2.4 Variance of Estimations

One of the many advantages of penalized splines is the possible linkage to Generalized Linear Mixed Models shown in the previous chapters. Treating the coefficients $\boldsymbol{\theta}$ as random variables, which are normally distributed, one can derive by a Laplace approximation a maximum likelihood estimation for the smoothing parameters; see Kauermann, Teuber, and Flaschel (2012) and chapter 5.1.4. The resulting estimations are given for the radius function as

$$\hat{\lambda}_\rho = \frac{\operatorname{tr}\left(\mathbf{D}\left((1,\,0)^T\right)^- \mathbf{D}\left((1,\,0)^T\right)\right)}{\operatorname{tr}\left(\mathbf{F}\left(\hat{\boldsymbol{\theta}},\,\hat{\boldsymbol{\lambda}},\,\hat{\boldsymbol{\Sigma}}_\epsilon\right)\mathbf{D}\left((1,\,0)^T\right)\right) + \hat{\boldsymbol{\theta}}^T\mathbf{D}\left((1,\,0)^T\right)\hat{\boldsymbol{\theta}}} \tag{5.58}$$

and for the angle function as

$$\hat{\lambda}_\phi = \frac{\operatorname{tr}\left(\mathbf{D}\left((0,\,1)^T\right)^- \mathbf{D}\left((0,\,1)^T\right)\right)}{\operatorname{tr}\left(\mathbf{F}\left(\hat{\boldsymbol{\theta}},\,\hat{\boldsymbol{\lambda}},\,\hat{\boldsymbol{\Sigma}}_\epsilon\right)\mathbf{D}\left((0,\,1)^T\right)\right) + \hat{\boldsymbol{\theta}}^T\mathbf{D}\left((0,\,1)^T\right)\hat{\boldsymbol{\theta}}}, \tag{5.59}$$

with $\mathbf{D}^{-}$ as the generalized inverse of the matrix $\mathbf{D}$. The maximum likelihood estimation for the (co-) variance matrix is given by

$$\hat{\mathbf{\Sigma}}_{\epsilon} = \frac{\sum_{i=1}^{n} \left(\mathbf{y}_{i\bullet} - \hat{\mathbf{g}}(t_i)\right) \left(\mathbf{y}_{i\bullet} - \hat{\mathbf{g}}(t_i)\right)^T}{n} \tag{5.60}$$

similar to the estimation in the previous chapter.

With these maximum likelihood estimators the variance of the estimated functions can be obtained by defining the term

$$\Delta_{\eta_\rho}(t) := (\mathbf{1}_{m\times 1} \otimes \mathbf{B}_\rho(t)\,,\ \mathbf{0}_{m\times 1} \otimes \mathbf{B}_\phi(t)\,,\ \mathbf{I}_m\,,\ \mathbf{0}_{m\times m}) \tag{5.61}$$

for the linear predictor of the radius and the term

$$\Delta_{\eta_\phi}(t) := (\mathbf{0}_{m\times 1} \otimes \mathbf{B}_\rho(t)\,,\ \mathbf{1}_{m\times 1} \otimes \mathbf{B}_\phi(t)\,,\ \mathbf{0}_{m\times m}\,,\ \mathbf{I}_m) \tag{5.62}$$

for the linear predictor of the angle, such that the term

$$\Delta_{\phi}(t) := (\oplus_{i=1}^{n} \rho_i(t))\, \Delta_{\eta_\phi}(t) \tag{5.63}$$

for the radius and the term

$$\Delta_{\phi}(t) := \Delta_{\eta_\phi}(t) \tag{5.64}$$

for the angle, can be defined to get the term

$$\Delta_{\mathbf{g}}(t) := \nabla \mathbf{g}\mathbf{B}(t) \tag{5.65}$$

for the loop function.
The variance for the loop function is then given by

$$\mathrm{Var}\left(\hat{\mathbf{g}}(t)\right) = \Delta_{\mathbf{g}}(t)\mathbf{F}\left(\hat{\boldsymbol{\theta}},\, \hat{\boldsymbol{\lambda}},\, \hat{\mathbf{\Sigma}}_{\epsilon}\right)^{-1} \mathbf{F}\left(\hat{\boldsymbol{\theta}},\, \mathbf{0},\, \hat{\mathbf{\Sigma}}_{\epsilon}\right) \mathbf{F}\left(\hat{\boldsymbol{\theta}},\, \hat{\boldsymbol{\lambda}},\, \hat{\mathbf{\Sigma}}_{\epsilon}\right)^{-1} \Delta_{\mathbf{g}}(t)^T \tag{5.66}$$

for the radius given by

$$\mathrm{Var}\left(\hat{\rho}(t)\right) \;=\; \Delta_\rho(t)\mathbf{F}\left(\hat{\boldsymbol{\theta}},\,\hat{\boldsymbol{\lambda}},\,\hat{\boldsymbol{\Sigma}}_\epsilon\right)^{-1}\mathbf{F}\left(\hat{\boldsymbol{\theta}},\,\mathbf{0},\,\hat{\boldsymbol{\Sigma}}_\epsilon\right)\mathbf{F}\left(\hat{\boldsymbol{\theta}},\,\hat{\boldsymbol{\lambda}},\,\hat{\boldsymbol{\Sigma}}_\epsilon\right)^{-1}\Delta_\rho(t)^T \quad (5.67)$$

and for the angle given by

$$\mathrm{Var}\left(\hat{\phi}(t)\right) \;=\; \Delta_\phi(t)\mathbf{F}\left(\hat{\boldsymbol{\theta}},\,\hat{\boldsymbol{\lambda}},\,\hat{\boldsymbol{\Sigma}}_\epsilon\right)^{-1}\mathbf{F}\left(\hat{\boldsymbol{\theta}},\,\mathbf{0},\,\hat{\boldsymbol{\Sigma}}_\epsilon\right)\mathbf{F}\left(\hat{\boldsymbol{\theta}},\,\hat{\boldsymbol{\lambda}},\,\hat{\boldsymbol{\Sigma}}_\epsilon\right)^{-1}\Delta_\phi(t)^T \quad (5.68)$$

respectively, for all $t \in [\underline{t},\, \overline{t}]$.

The resulting (95%-) confidence region based on the asymptotic of the maximum likelihood theory yields

$$\mathrm{CR}_{\mathbf{g}}(t) \;=\; \left\{\mathbf{y} \in \mathbb{R}^m \middle|\, (\mathbf{y}-\hat{\mathbf{g}}(t))^T\mathrm{Var}(\hat{\mathbf{g}}(t))^{-1}(\mathbf{y}-\hat{\mathbf{g}}(t)) \le \chi^2_{m,\,0.95}\right\} \quad (5.69)$$

with $\chi^2_{m,\,0.95}$ as the 95% quantile of the χ^2 distribution with m degrees of freedom.

5.2.5 Simulations

Similar to the simulations in section 5.2.2, more simulations will be presented here where each simulation contains $n = 200$ equidistant time points over the support $[\underline{t},\, \overline{t}] = [0,\, 1]$.

The first simulation *S1* is given by the data described by the model

$$\begin{pmatrix} y_{i1} \\ y_{i2} \end{pmatrix} \;=\; \begin{pmatrix} \sin(3\pi t_i) \\ \sin(3\pi t_i + \frac{\pi}{4}) \end{pmatrix} + \begin{pmatrix} \epsilon_{i1} \\ \epsilon_{i2} \end{pmatrix} \quad (5.70)$$

for $i = 1, \ldots, n$ and with the residuals $(\epsilon_{i1},\, \epsilon_{i2})^T \overset{iid}{\sim} \mathcal{N}\left(\mathbf{0}, 0.25^2\mathbf{I}_2\right)$, which is the example (a) given in section 5.2.1.

The estimated functions are shown in figure 5.5. The left-hand-side plots

are estimated by the basic model and the right-hand-side plots by the generalized (or extended) model. In the first row the functions g_1 and g_2 are illustrated over time. The second row shows the trajectory of the loop function $\mathbf{g}$. In the third and fourth rows the radius (ρ_1 and ρ_2) and the angle (ϕ_1 and ϕ_2) are plotted against time.

The second simulation *S2* is given by the data described by the model

$$\begin{pmatrix} y_{i1} \\ y_{i2} \end{pmatrix} = \begin{pmatrix} \sin(3\pi t_i) \\ 2\cos(3\pi t_i) \end{pmatrix} + \begin{pmatrix} \epsilon_{i1} \\ \epsilon_{i2} \end{pmatrix} \tag{5.71}$$

for $i = 1, \ldots, n$ and with the residuals $(\epsilon_{i1},\, \epsilon_{i2})^T \overset{iid}{\sim} \mathcal{N}(\mathbf{0}, 0.25^2\mathbf{I}_2)$, which is the example (b) given in section 5.2.1.

The estimated functions are presented in figure 5.6 in the same way as the simulation *S1* above.

The third simulation *S3* is an example of the generalized model which can be used for multivariate time series, generated by the model

$$\begin{pmatrix} y_{i1} \\ y_{i2} \\ y_{i3} \end{pmatrix} = \mathbf{A}_3 \begin{pmatrix} \sin(2\pi t_i) - 0.5\cos(10\pi t_i) \\ \cos(2\pi t_i) - 0.5\sin(10\pi t_i) \end{pmatrix} + \begin{pmatrix} \epsilon_{i1} \\ \epsilon_{i2} \\ \epsilon_{i3} \end{pmatrix} \tag{5.72}$$

for $i = 1, \ldots, n$ and with the residuals $(\epsilon_{i1},\, \epsilon_{i2},\, \epsilon_{i3})^T \overset{iid}{\sim} \mathcal{N}(\mathbf{0}, 0.25^2\mathbf{I}_3)$ and $\mathbf{A}_3 = \begin{pmatrix} 1\cos 0 & 1\sin 0 \\ \frac{3}{2}\cos\frac{\pi}{2} & \frac{3}{2}\sin\frac{\pi}{2} \\ \frac{11}{6}\cos\frac{5\pi}{6} & \frac{11}{6}\sin\frac{5\pi}{6} \end{pmatrix}$.

The estimation of the simulation *S3* is shown in figure 5.7. The upper left-hand-side plot shows the functions g_1, g_2, and g_3 over time. The other figures show the trajectories of g_1 vs g_2 (upper right-hand-side plot), g_1 vs g_3 (lower

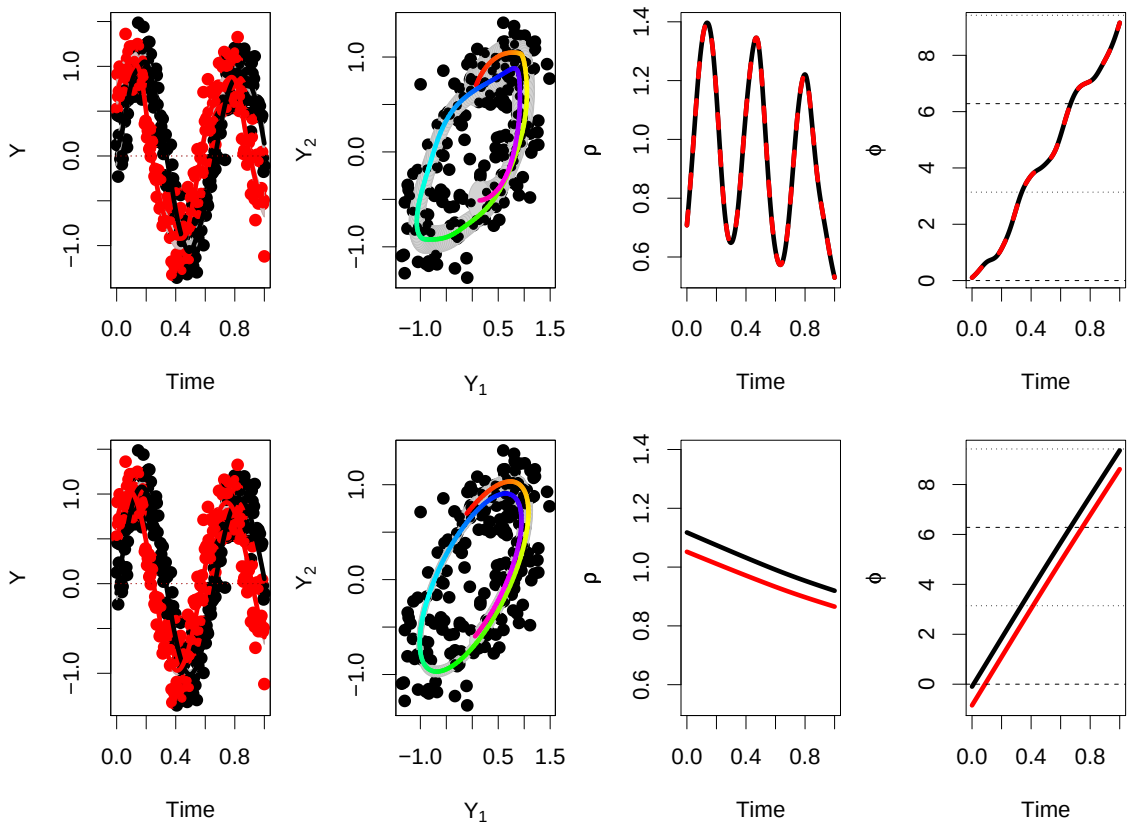

Figure 5.5: Estimations of simulation *S1* given in equation 5.70: Upper row: Basic circular model; lower row: Extended circular model. First column: Observation of first (black dots) and second time series (red dots), and their estimations (first time series: black line, second time series: Red line) and their confidence regions (grey-shaded areas). Second column: Observations (dots), estimation (colored lines), and confidence regions (grey-shaded areas). Third column: Estimated radius for first (black line) and second (red line) time series. Fourth column: Estimated angle for first (black line) and second (red line) time series.

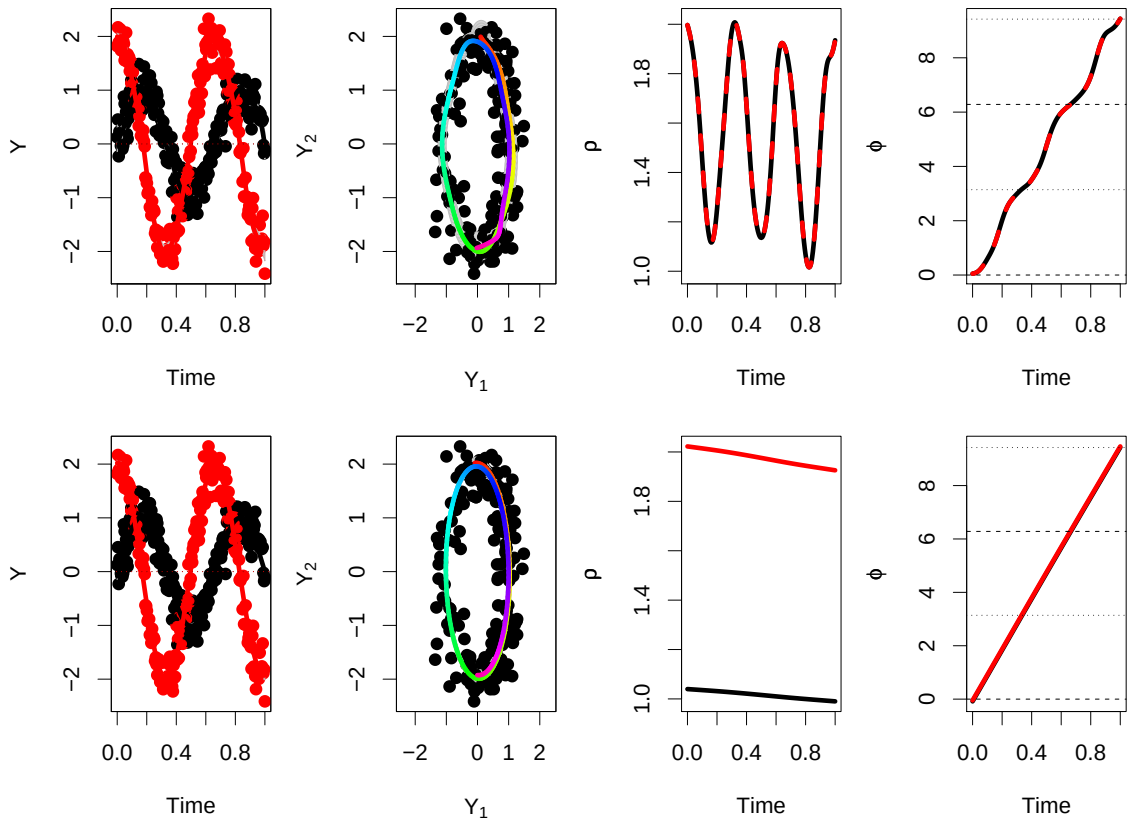

Figure 5.6: Estimations of simulation *S2* given in equation 5.71: Upper row: Basic circular model; lower row: Extended circular model. First column: Observation of first (black dots) and second time series (red dots), and their estimations (first time series: Black line, second time series: Red line) and their confidence regions (grey-shaded areas). Second column: Observations (dots), estimation (colored lines), and confidence regions (grey-shaded areas). Third column: Estimated radius for first (black line) and second (red line) time series. Fourth column: Estimated angle for first (black line) and second (red line) time series.

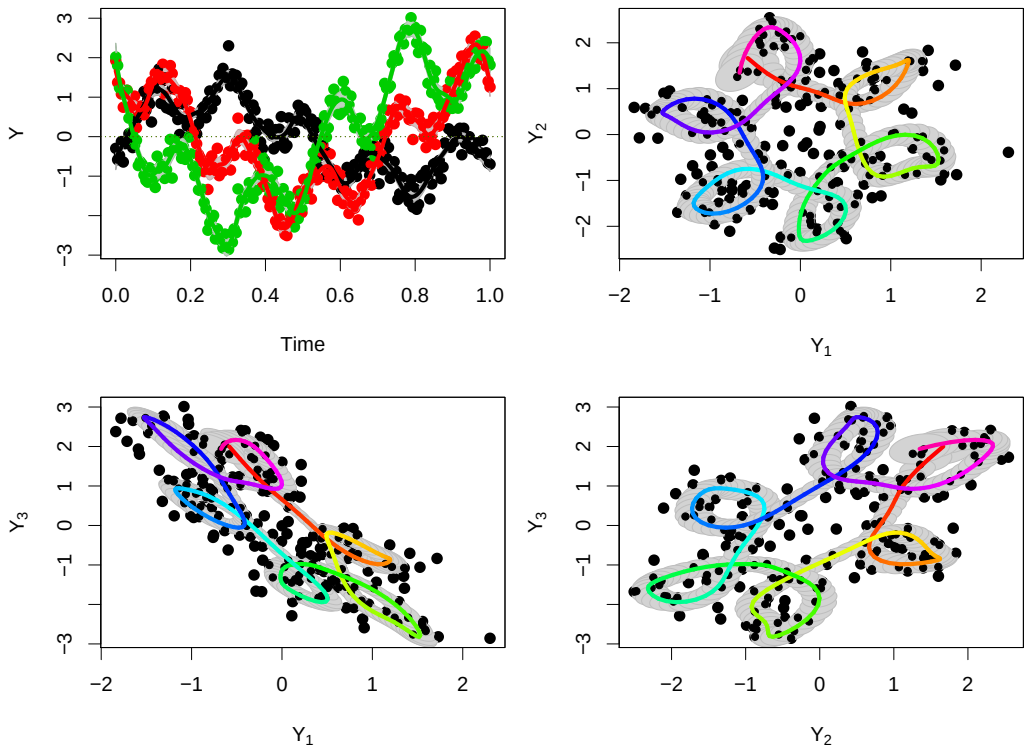

Figure 5.7: Estimations of simulation *S3* given in equation 5.72: Observations (dots), estimation (colored lines), and confidence regions (grey-shaded areas). Upper left-hand plot: Time series over time (black: y_1, red: y_2, green: y_3). Upper right-hand plot: y_1 vs y_2; lower left-hand plot: y_1 vs y_3; lower right-hand plot: y_2 vs y_3.

---

left-hand-side plot), and g_2 vs g_3 (lower right-hand-side plot).

The inclusion of more time series is shown in simulation *S4* given by the

model

$$\begin{pmatrix} y_{i1} \\ \vdots \\ y_{i6} \end{pmatrix} = \mathbf{A}_6 \begin{pmatrix} \sin(2\pi t_i) - 0.5\cos(10\pi t_i) \\ \cos(2\pi t_i) - 0.5\sin(10\pi t_i) \end{pmatrix} + \begin{pmatrix} \epsilon_{i1} \\ \vdots \\ \epsilon_{i6} \end{pmatrix} \tag{5.73}$$

for $i = 1, \ldots, n$ and with the residuals $(\epsilon_{i1}, \ldots, \epsilon_{i6})^T \overset{iid}{\sim} \mathcal{N}(\mathbf{0}, 0.25^2\mathbf{I}_6)$ and

$$\mathbf{A}_6 = \begin{pmatrix} \mathbf{A}_3 \\ \frac{25}{12}\cos\frac{13\pi}{12} & \frac{25}{12}\sin\frac{13\pi}{12} \\ \frac{137}{60}\cos\frac{77\pi}{60} & \frac{137}{60}\sin\frac{77\pi}{60} \\ \frac{147}{60}\cos\frac{87\pi}{60} & \frac{147}{60}\sin\frac{87\pi}{60} \end{pmatrix}.$$

For the fourth simulation only a few graphics are shown: In figure 5.8 the same time series as for the simulation *S3* are shown and in figure 5.9 some additional time series are plotted. In figure 5.9 on the first row on the left-hand-side all functions $g_4, \ldots, g_6$ are plotted against time. Furthermore, the trajectories of g_3 vs g_4 (first row, right-hand-side), g_4 vs g_5 (second row, left-hand-side), and g_5 vs g_6 (second row, right-hand-side) are presented.

Obviously, one sees that the first three estimated time series of the simulation *S4* are of "better" quality than in the simulation *S3* because more information (here via more time series) is available, meaning that the confidence regions are narrower.

The simulations *S3* and *S4* have been repeated $N = 500$ times and the coverage probabilities are graphically illustrated in figure 5.10. The upper figure shows the coverage probability for *S3*, the middle figure of the first three time series of *S4*, and the lower figure of all six time series of *S4*, which leads to the conclusion that the estimation will be improved if more time series are included.

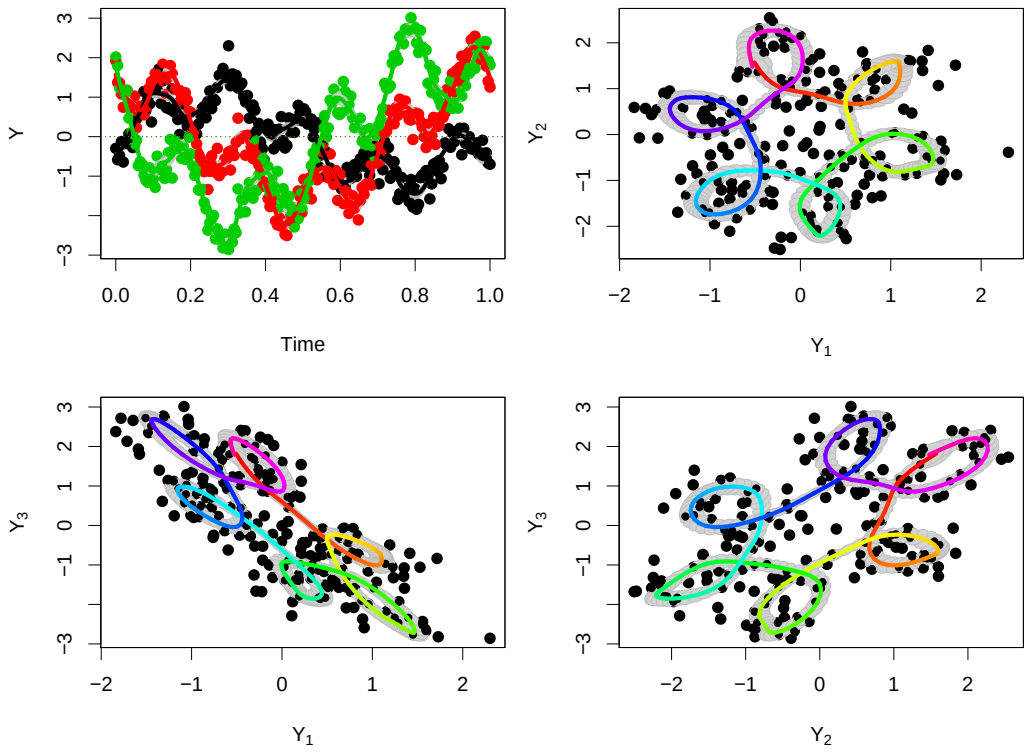

Figure 5.8: Estimations of simulation *S4* given in equation 5.73: Observations (dots), estimation (colored lines), and confidence regions (grey-shaded areas). First row left-hand plot: Time series over time (black: y_1, red: y_2, green: y_3). First row right-hand plot: y_1 vs y_2; second row left-hand plot: y_1 vs y_3; second row right-hand plot: y_2 vs y_3. Third row left-hand plot: Time series over time (black: y_4, red: y_5, green: y_6).

5.3 The Basic and the Extended Three-Dimensional Loops Model

The basic model presented in chapter 5.1 and the extended as well as the generalized model presented in chapter 5.2 for the multivariate time series

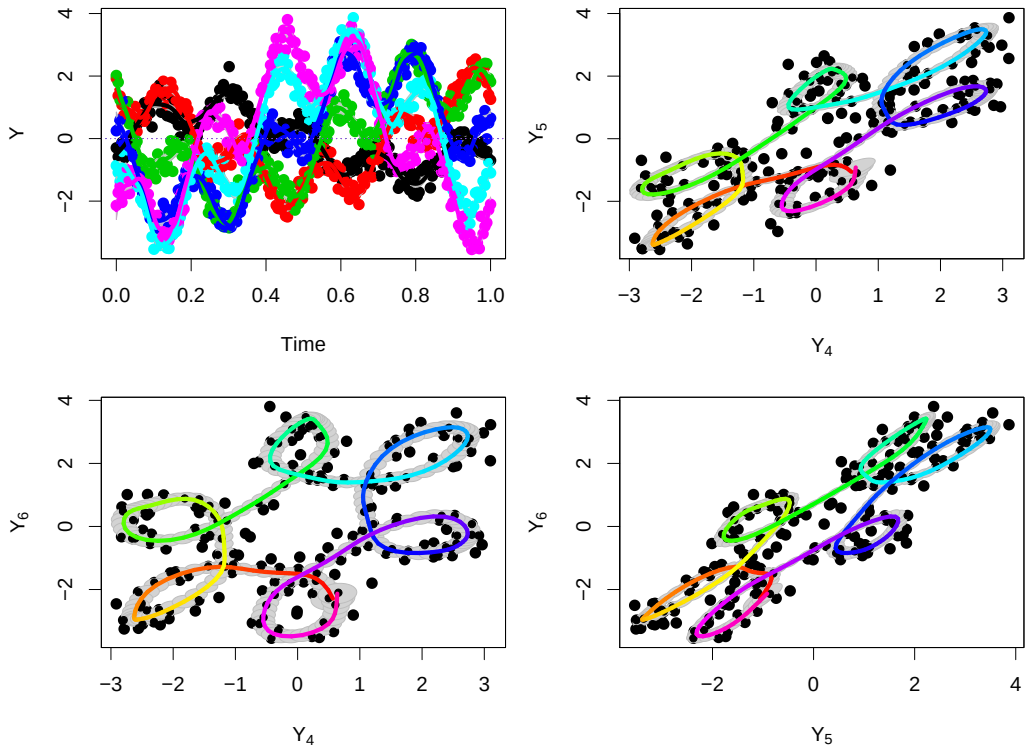

Figure 5.9: Estimations of simulation *S4* given in equation 5.73: Observations (dots), estimation (colored lines), and confidence regions (grey-shaded areas). First row left-hand plot: Time series over time (black: y_4, red: y_5, green: y_6). First row right-hand plot: y_4 vs y_5; second row left-hand plot: y_4 vs y_6; second row right-hand plot: y_5 vs y_6.

---

can be extended to be formulated to capture a higher dimensional ball movement. The complexity of the three-dimensional circular movement model is strongly increasing compared to the two-dimensional model. It is worth noting that from a mathematical (theoretical) exercise, the model can easily be extended to higher dimension, but the numerical applicability of the model decreases with each new dimensions included. Nonetheless, the extension

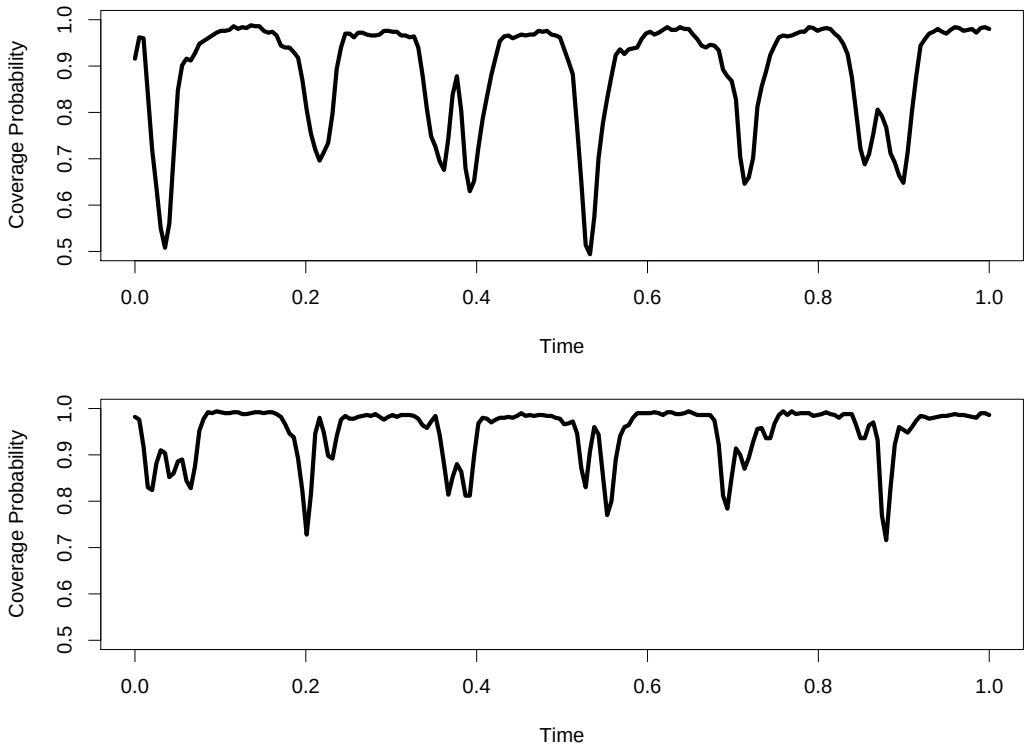

Figure 5.10: Coverage probabilities for different simulations: Coverage probabilities for simulation *S3* (upper plot) and *S4* (lower plot), based on $N = 500$ repeated simulations.

---

will be presented for the sake of completeness so that the user can choose the preferred model depending on the data at hand.

Therefore, the model will be introduced with the basic model for three time series. Similar to the two-dimensional model, the extension of the three-dimensional model shown in section 5.3.2 is straightforward and can be used for the generalized version for more than the three time series shown in section 5.3.3. The properties of the estimation and the smoothing parameters

are analogously given as in the previous models.

5.3.1 Basic Three-Dimensional Model

Another possibility to describe $m = 3$ time series is a three-dimensional loop, whereby the observations will be described using three-dimensional polar coordinates. The structural part of the observations $(y_{i1},\, y_{i2},\, y_{i3},\, t_i)$ with index $i = 1, \ldots, n$ will be written as

$$\mathrm{E}\left(\begin{pmatrix} y_{i1} \\ y_{i2} \\ y_{i3} \end{pmatrix} \mid t_i\right) = \tilde{\mathbf{g}}(t_i) := \begin{pmatrix} g_1(t_i) \\ g_2(t_i) \\ g_3(t_i) \end{pmatrix} \tag{5.74}$$

for which it will be assumed that the three-dimensional loop function can be expressed in spherical polar coordinates, i.e.

$$\tilde{\mathbf{g}}(t) := \rho(t) \begin{pmatrix} \cos \upsilon(t) \cos \phi(t) \\ \cos \upsilon(t) \sin \phi(t) \\ \sin \upsilon(t) \end{pmatrix} + \begin{pmatrix} \gamma_1(t_i) \\ \gamma_2(t_i) \\ \gamma_3(t_i) \end{pmatrix} \tag{5.75}$$

around the center $\boldsymbol{\gamma}(t) := (\gamma_1(t),\, \gamma_2(t),\, \gamma_3(t))^T$ which is set for the sake of simplicity to the point of origin, i.e. $\boldsymbol{\gamma}(t) \equiv (0,\, 0,\, 0)^T$ without the loss of generalization. The radius function ρ is the (Euclidean) distance between $\tilde{\mathbf{g}}$ and the center. The (azimuth) angle function ϕ represents the angle between the function $\tilde{\mathbf{g}}$ projected onto the xy-plane and the positive x-axis. The (zenith) angle υ represents the angle between the line from the center and the function $\tilde{\mathbf{g}}$ projected onto the xy-plane and the line from the center to function $\tilde{\mathbf{g}}$; see figure 5.11.

Similar to the previous model it will be assumed that the polar coordinate functions can be expressed by

$$\rho(t) = \exp(\mathbf{B}_\rho(t)\boldsymbol{\theta}_\rho) \tag{5.76}$$

$$\phi(t) = \mathbf{B}_\phi(t)\boldsymbol{\theta}_\phi \tag{5.77}$$

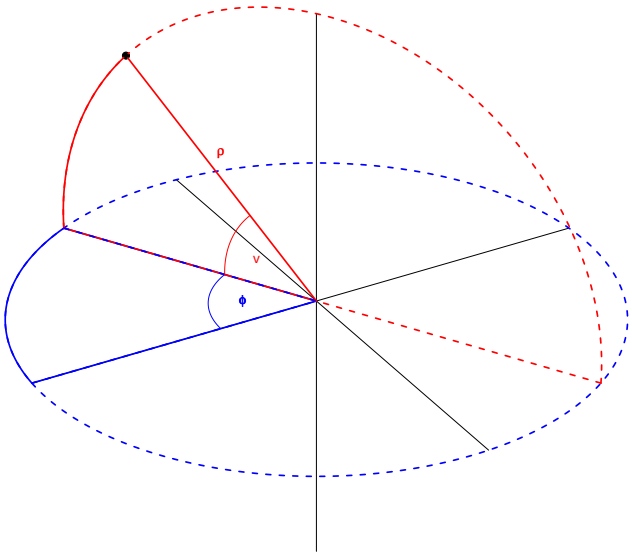

Figure 5.11: Three-dimensional point measured with spherical coordinates.

---

and

$$\upsilon(t) \quad = \quad \mathbf{B}_\upsilon(t)\boldsymbol{\theta}_\upsilon \tag{5.78}$$

with the B-spline bases $\mathbf{B}_\rho(\cdot)$, $\mathbf{B}_\phi(\cdot)$, and $\mathbf{B}_\upsilon(\cdot)$ over the support of the time $[\underline{t}, \overline{t}]$.

The unknown parameters $\boldsymbol{\theta} := (\boldsymbol{\theta}_\rho^T, \boldsymbol{\theta}_\phi^T, \boldsymbol{\theta}_\upsilon^T)^T$ are estimated using the maximum likelihood approach, but instead of using the (log) likelihood like in equation (5.45) the penalized form

$$l_p(\mathbf{y}; \boldsymbol{\theta}, \boldsymbol{\lambda}) \quad := \quad l(\mathbf{y}; \boldsymbol{\theta}) - \frac{1}{2}\lambda_\rho \boldsymbol{\theta}_\rho^T \mathbf{D}_\rho \boldsymbol{\theta}_\rho - \frac{1}{2}\lambda_\phi \boldsymbol{\theta}_\phi^T \mathbf{D}_\phi \boldsymbol{\theta}_\phi - \frac{1}{2}\lambda_\upsilon \boldsymbol{\theta}_\upsilon^T \mathbf{D}_\upsilon \boldsymbol{\theta}_\upsilon \tag{5.79}$$

will be preferred with the smoothing parameters $\boldsymbol{\lambda} := (\lambda_\rho, \lambda_\phi, \lambda_\upsilon)$ and the penalty matrices $\mathbf{D}_\rho$, $\mathbf{D}_\phi$, and $\mathbf{D}_\upsilon$.

Defining,

$$\begin{aligned}
&\mathbf{B}(t) := \mathbf{B}_\rho(t) \oplus \mathbf{B}_\phi(t) \oplus \mathbf{B}_\upsilon(t) \\
&\tilde{\mathbf{D}}(\boldsymbol{\lambda}) := (\lambda_\rho \mathbf{D}_\rho) \oplus (\lambda_\phi \mathbf{D}_\phi) \oplus (\lambda_\upsilon \mathbf{D}_\upsilon) \\
\nabla\tilde{\mathbf{g}}(t) \quad := \quad & \rho(t) \begin{pmatrix} \cos\upsilon(t)\cos\phi(t) & -\cos\upsilon(t)\sin\phi(t) & -\sin\upsilon(t)\cos\phi(t) \\ \cos\upsilon(t)\sin\phi(t) & \cos\upsilon(t)\cos\phi(t) & -\sin\upsilon(t)\sin\phi(t) \\ \sin\upsilon(t) & 0 & \cos\upsilon(t) \end{pmatrix}
\end{aligned}$$

given the operator defined as $\mathbf{A}_{n\times m} \oplus \mathbf{B}_{p\times q} := \begin{pmatrix} \mathbf{A} & \mathbf{0}_{n\times q} \\ \mathbf{0}_{p\times m} & \mathbf{B} \end{pmatrix}$.

Under the assumption that the observations are normally distributed, the score equation for $\boldsymbol{\theta}$ is given by

$$\mathbf{s}_\theta(\boldsymbol{\theta}, \boldsymbol{\lambda}, \boldsymbol{\Sigma}_\epsilon) := \sum_{i=1}^{n} \left(\nabla\tilde{\mathbf{g}}(t_i)\mathbf{B}(t_i)\right)^T \boldsymbol{\Sigma}_\epsilon^{-1} \left(\mathbf{y}_{i\bullet} - \tilde{\mathbf{g}}(t_i)\right) - \tilde{\mathbf{D}}(\boldsymbol{\lambda})\boldsymbol{\theta} \quad (5.80)$$

and the Fisher matrix can be expressed as

$$\mathbf{F}(\boldsymbol{\theta}, \boldsymbol{\lambda}, \boldsymbol{\Sigma}_\epsilon) := \sum_{i=1}^{n} \left(\nabla\tilde{\mathbf{g}}(t_i)\mathbf{B}(t_i)\right)^T \boldsymbol{\Sigma}_\epsilon^{-1} \left(\nabla\tilde{\mathbf{g}}(t_i)\mathbf{B}(t_i)\right) + \tilde{\mathbf{D}}(\boldsymbol{\lambda}). \quad (5.81)$$

Obviously the estimation depends on the order of the three time series, namely which is the first, second, and third time series, due to the different functional forms given by the definition of the model in equation (5.75), such that either one has to be sure of the order of the time series or one has to estimate and compare the three different combinations of the time series and then select the best model. Because of this problem the extension of the basic model is an advantage.

5.3.2 Extended Model

Similar to section 5.2 in which the equivalency of the extended model and the generalized model is shown, here a linear combination will be introduced such that

$$\mathrm{E}\left(\begin{pmatrix} y_{i1} \\ y_{i2} \\ y_{i3} \end{pmatrix} \mid t_i\right) = \mathbf{A}\tilde{\mathbf{g}}(t_i) =: \mathbf{g}(t_i) := \begin{pmatrix} g_1(t_i) \\ g_2(t_i) \\ g_3(t_i) \end{pmatrix} \tag{5.82}$$

represents the extended model with $\tilde{\mathbf{g}}(\cdot)$ given in equation (5.75). Each row of the projection matrix is set without loss of generality to

$$\mathbf{A}_i = (a_{i1},\, a_{i2},\, a_{i3}) = \exp(\rho_{0i})\,(\cos\upsilon_{0i}\cos\phi_{0i},\; -\cos\upsilon_{0i}\sin\phi_{0i},\; \sin\upsilon_{0i}) \tag{5.83}$$

for $i = 1, 2, 3$ such that the equation (5.82) can be written as

$$\mathbf{g}(t) = \begin{pmatrix} \rho(t)\exp(\rho_{01})\left[\cos\upsilon_{01}\cos\upsilon(t)\left(\cos\phi_{01}\cos\phi(t) - \sin\phi_{01}\sin\phi(t)\right) + \sin\upsilon_{01}\sin\upsilon(t)\right] \\ \rho(t)\exp(\rho_{02})\left[\cos\upsilon_{02}\cos\upsilon(t)\left(\cos\phi_{02}\cos\phi(t) - \sin\phi_{02}\sin\phi(t)\right) + \sin\upsilon_{02}\sin\upsilon(t)\right] \\ \rho(t)\exp(\rho_{03})\left[\cos\upsilon_{03}\cos\upsilon(t)\left(\cos\phi_{03}\cos\phi(t) - \sin\phi_{03}\sin\phi(t)\right) + \sin\upsilon_{03}\sin\upsilon(t)\right] \end{pmatrix} \tag{5.84}$$

$$= \begin{pmatrix} \exp(\mathbf{B}_\rho(t)\boldsymbol{\theta}_\rho + \rho_{01})\left[\cos\upsilon_{01}\cos\upsilon(t)\cos(\phi_{01} + \phi(t)) + \sin\upsilon_{01}\sin\upsilon(t)\right] \\ \exp(\mathbf{B}_\rho(t)\boldsymbol{\theta}_\rho + \rho_{02})\left[\cos\upsilon_{02}\cos\upsilon(t)\cos(\phi_{02} + \phi(t)) + \sin\upsilon_{02}\sin\upsilon(t)\right] \\ \exp(\mathbf{B}_\rho(t)\boldsymbol{\theta}_\rho + \rho_{03})\left[\cos\upsilon_{03}\cos\upsilon(t)\cos(\phi_{03} + \phi(t)) + \sin\upsilon_{03}\sin\upsilon(t)\right] \end{pmatrix} \tag{5.85}$$

$$=: \begin{pmatrix} \rho_1(t)\left[\cos\upsilon_{01}\cos\upsilon(t)\cos\phi_1(t) + \sin\upsilon_{01}\sin\upsilon(t)\right] \\ \rho_2(t)\left[\cos\upsilon_{02}\cos\upsilon(t)\cos\phi_2(t) + \sin\upsilon_{02}\sin\upsilon(t)\right] \\ \rho_3(t)\left[\cos\upsilon_{03}\cos\upsilon(t)\cos\phi_3(t) + \sin\upsilon_{03}\sin\upsilon(t)\right] \end{pmatrix} \tag{5.86}$$

Recalling the score function and the Fisher matrix in the previous section and remembering that the extended model is a simple linear combination of the basis model, the score function and Fisher matrix can be straightforwardly

calculated, such that these steps will be omitted in this section. However, the generalized form presented below can also be used for $m = 3$ time series, which is then equivalent to the extended model presented here.

5.3.3 Generalized Three-Dimensional Model

In the generalized model, the $m + 1$ time series $(y_{i1}, \ldots, y_{im}, t_i)$ with $i = 1, \ldots, n$ and $m \geq 3$ are observed. Using the definition for the projection matrix $\mathbf{A}$ in equation (5.83) for $i = 1, \ldots, m$ the structural part of the model should be written in the form of the extended model, i.e.

$$\begin{aligned} \mathrm{E}\left(\begin{pmatrix} y_{i1} \\ \vdots \\ y_{im} \end{pmatrix} | \, t_i\right) &= \mathbf{A}\tilde{\mathbf{g}}(t_i) = \mathbf{g}(t_i) \qquad (5.87) \\ &= \begin{pmatrix} \rho_1(t)\left[\cos\upsilon_{01}\cos\upsilon(t)\cos\phi_1(t) + \sin\upsilon_{01}\sin\upsilon(t)\right] \\ \vdots \\ \rho_m(t)\left[\cos\upsilon_{0m}\cos\upsilon(t)\cos\phi_m(t) + \sin\upsilon_{0m}\sin\upsilon(t)\right] \end{pmatrix} \end{aligned}$$

Defining

$$\begin{aligned} \boldsymbol{\theta} &:= (\boldsymbol{\theta}_\rho^T, \boldsymbol{\theta}_\phi^T, \boldsymbol{\theta}_\upsilon^T, \rho_{01}, \ldots, \rho_{0m}, \phi_{01}, \ldots, \phi_{0m}, \upsilon_{01}, \ldots, \upsilon_{0m})^T \qquad (5.88) \\ & \mathbf{D}(\boldsymbol{\lambda}) := \tilde{\mathbf{D}}(\boldsymbol{\lambda}) \oplus \mathbf{0}_{3m\times 3m} \qquad (5.89) \end{aligned}$$

and

$$\begin{aligned} \nabla\mathbf{gB}(t) &:= \Big(\mathbf{A}\nabla\tilde{\mathbf{g}}(t)\mathbf{B}(t)\,,\ \oplus_{j=1}^m g_j(t)\,,\ -\oplus_{j=1}^m \rho_j(t)\cos\upsilon_{0j}\cos\upsilon(t)\sin\phi_j(t)\,, \\ &\quad \oplus_{j=1}^m(\cos\upsilon_{0j}\sin\upsilon(t) - \sin\upsilon_{0j}\cos\upsilon(t)\cos\phi_j(t))\Big), \qquad (5.90) \end{aligned}$$

such that the first derivative of the likelihood with respect to $\boldsymbol{\theta}$ can be written as

$$\mathbf{s}_\theta(\boldsymbol{\theta}, \boldsymbol{\lambda}, \boldsymbol{\Sigma}_\epsilon) := \sum_{i=1}^n \nabla\mathbf{gB}(t_i)^T\boldsymbol{\Sigma}_\epsilon^{-1}\left(\mathbf{y}_{i\bullet} - \mathbf{g}(t_i)\right) - \mathbf{D}(\boldsymbol{\lambda})\boldsymbol{\theta} \qquad (5.91)$$

and the negative (expected) second derivative

$$\mathbf{F}(\boldsymbol{\theta},\, \boldsymbol{\lambda},\, \boldsymbol{\Sigma}_\epsilon) \;:=\; \sum_{i=1}^{n} \nabla \mathbf{gB}(t_i)^T \boldsymbol{\Sigma}_\epsilon^{-1} \nabla \mathbf{gB}(t_i) + \mathbf{D}(\boldsymbol{\lambda}), \tag{5.92}$$

which are necessary for the Fisher scoring algorithm to get the maximum likelihood estimations. To avoid any confusion given the definitions and the solutions defined in equations (5.91) and (5.92), the dimensions of the variables will be given. Assuming that the B-spline matrix $\mathbf{B}_\phi(t)$, $\mathbf{B}_\rho(t)$, and $\mathbf{B}_\upsilon(t)$ are given as $K_\phi \times 1$, $K_\rho \times 1$, and $K_\upsilon \times 1$, respectively, then the dimension of the square matrix $\mathbf{D}(\boldsymbol{\lambda})$ is $K_\rho + K_\phi + K_\upsilon + 3m \times K_\rho + K_\phi + K_\upsilon + 3m$, of the parameter vector $\boldsymbol{\theta}$ is $K_\rho + K_\phi + K_\upsilon + m + m + m \times 1$, of the matrix $\nabla \mathbf{gB}(t_i)$ is $m \times K_\rho + K_\phi + K_\upsilon + m + m + m$, of the co-variance matrix $\boldsymbol{\Sigma}$ is $m \times m$, and of the i-th observation $\mathbf{y}_{i\bullet}$ and their expected trend $\mathbf{g}(t_i)$ are $m \times 1$. The dimension of the first derivative with respect to $\boldsymbol{\theta}$, defined in equation (5.91), is therefore $K_\rho + K_\phi + K_\upsilon + 3m \times 1$ and the dimension of the negative (expected) second derivative, defined in equation (5.92), is therefore $K_\rho + K_\phi + K_\upsilon + 3m \times K_\rho + K_\phi + K_\upsilon + 3m$.

Similar to the two-dimensional model, the smoothing parameters are given by the Laplace approximation, which yields for the radius

$$\hat{\lambda}_\rho \;=\; \frac{\operatorname{tr}\left(\mathbf{D}\left((1,\,0,\,0)^T\right)^{-}\mathbf{D}\left((1,\,0,\,0)^T\right)\right)}{\operatorname{tr}\left(\mathbf{F}\left(\hat{\boldsymbol{\theta}},\,\hat{\boldsymbol{\lambda}},\,\hat{\boldsymbol{\Sigma}}_\epsilon\right)\mathbf{D}\left((1,\,0,\,0)^T\right)\right) + \hat{\boldsymbol{\theta}}^T\mathbf{D}\left((1,\,0,\,0)^T\right)\hat{\boldsymbol{\theta}}} \tag{5.93}$$

for the azimuth angle

$$\hat{\lambda}_\phi \;=\; \frac{\operatorname{tr}\left(\mathbf{D}\left((0,\,1,\,0)^T\right)^{-}\mathbf{D}\left((0,\,1,\,0)^T\right)\right)}{\operatorname{tr}\left(\mathbf{F}\left(\hat{\boldsymbol{\theta}},\,\hat{\boldsymbol{\lambda}},\,\hat{\boldsymbol{\Sigma}}_\epsilon\right)\mathbf{D}\left((0,\,1,\,0)^T\right)\right) + \hat{\boldsymbol{\theta}}^T\mathbf{D}\left((0,\,1,\,0)^T\right)\hat{\boldsymbol{\theta}}} \tag{5.94}$$

and for the zenith angle

$$\hat{\lambda}_\upsilon \;=\; \frac{\operatorname{tr}\left(\mathbf{D}\left((0,\,0,\,1)^T\right)^{-}\mathbf{D}\left((0,\,0,\,1)^T\right)\right)}{\operatorname{tr}\left(\mathbf{F}\left(\hat{\boldsymbol{\theta}},\,\hat{\boldsymbol{\lambda}},\,\hat{\boldsymbol{\Sigma}}_\epsilon\right)\mathbf{D}\left((0,\,0,\,1)^T\right)\right) + \hat{\boldsymbol{\theta}}^T\mathbf{D}\left((0,\,0,\,1)^T\right)\hat{\boldsymbol{\theta}}}. \tag{5.95}$$

Analog to the previous models, the variance of the fitted function is given by

$$\text{Var}\left(\hat{\mathbf{g}}(t)\right) \;=\; \Delta_{\mathbf{g}}(t)\mathbf{F}\left(\hat{\boldsymbol{\theta}},\,\hat{\boldsymbol{\lambda}},\,\hat{\boldsymbol{\Sigma}}_{\epsilon}\right)^{-1}\mathbf{F}\left(\hat{\boldsymbol{\theta}},\,\mathbf{0},\,\hat{\boldsymbol{\Sigma}}_{\epsilon}\right)\mathbf{F}\left(\hat{\boldsymbol{\theta}},\,\hat{\boldsymbol{\lambda}},\,\hat{\boldsymbol{\Sigma}}_{\epsilon}\right)^{-1}\Delta_{\mathbf{g}}(t)^{T} \tag{5.96}$$

for all $t \in [\underline{t};\,\overline{t}]$ with $\Delta_{\mathbf{g}}(t) := \nabla\mathbf{g}\mathbf{B}(t)$ defined above and the (co-) variance of the residuals estimated by

$$\hat{\boldsymbol{\Sigma}}_{\epsilon} \;=\; \frac{\sum_{i=1}^{n}\left(\mathbf{y}_{i\bullet}-\hat{\mathbf{g}}(t_i)\right)\left(\mathbf{y}_{i\bullet}-\hat{\mathbf{g}}(t_i)\right)^{T}}{n} \tag{5.97}$$

similar to the previous models.

Given the variance in equation (5.96), the (95%-) confidence region based on the asymptotic of the maximum likelihood theory yields

$$\text{CR}_{\mathbf{g}}(t) \;=\; \left\{\mathbf{y}\in\mathbb{R}^{m}\middle|\,(\mathbf{y}-\hat{\mathbf{g}}(t))^{T}\text{Var}(\hat{\mathbf{g}}(t))^{-1}(\mathbf{y}-\hat{\mathbf{g}}(t)) \le \chi^{2}_{m,\,0.95}\right\} \tag{5.98}$$

with $\chi^{2}_{m,\,0.95}$ as the 95% quantile of the χ^{2} distribution with m degrees of freedom.

5.3.4 Simulations

This section will be concluded with a simulation to show how the model behaves and how the time series might look if the underlying trend follows a three-dimensional loop structure. The simulation contains $n = 200$ equidistant time points t over the support from zero to one. For the simulation the three-dimensional loop functions

$$\begin{pmatrix} g_1(t) \\ g_2(t) \\ g_3(t) \end{pmatrix} \;=\; \rho(t)\begin{pmatrix} \cos\upsilon(t)\sin\phi(t) \\ \cos\upsilon(t)\cos\phi(t) \\ \sin\upsilon(t) \end{pmatrix} \tag{5.99}$$

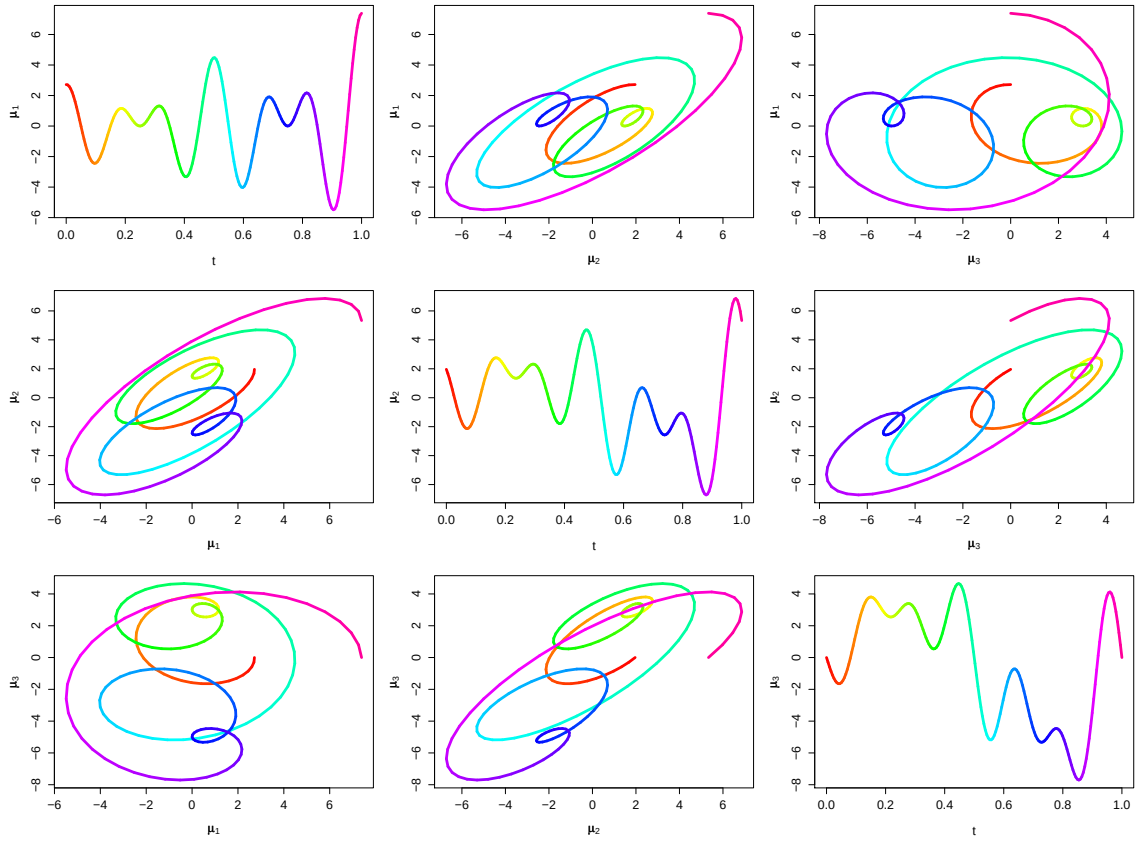

Figure 5.12: Trend of simulation given in equation 5.99: On the main diagonals the trend is plotted over time. In the others plots, the trajectories of the trend is plotted over time.

---

with $\rho(t) = \exp(1 + t)$, $\upsilon(t) = 2\pi t$, and $\phi(t) = 10\pi t$ are given. Furthermore, defining the shifts $\rho_{01} = 0$, $\rho_{02} = 0.1$, $\rho_{03} = 0.2$, $\upsilon_{01} = 0$, $\upsilon_{02} = 0.125\pi$, $\upsilon_{03} = 0.25\pi$, $\phi_{01} = 0$, $\phi_{02} = 0.25\pi$, and $\phi_{03} = 0.5\pi$ to build the matrix $\mathbf{A}$

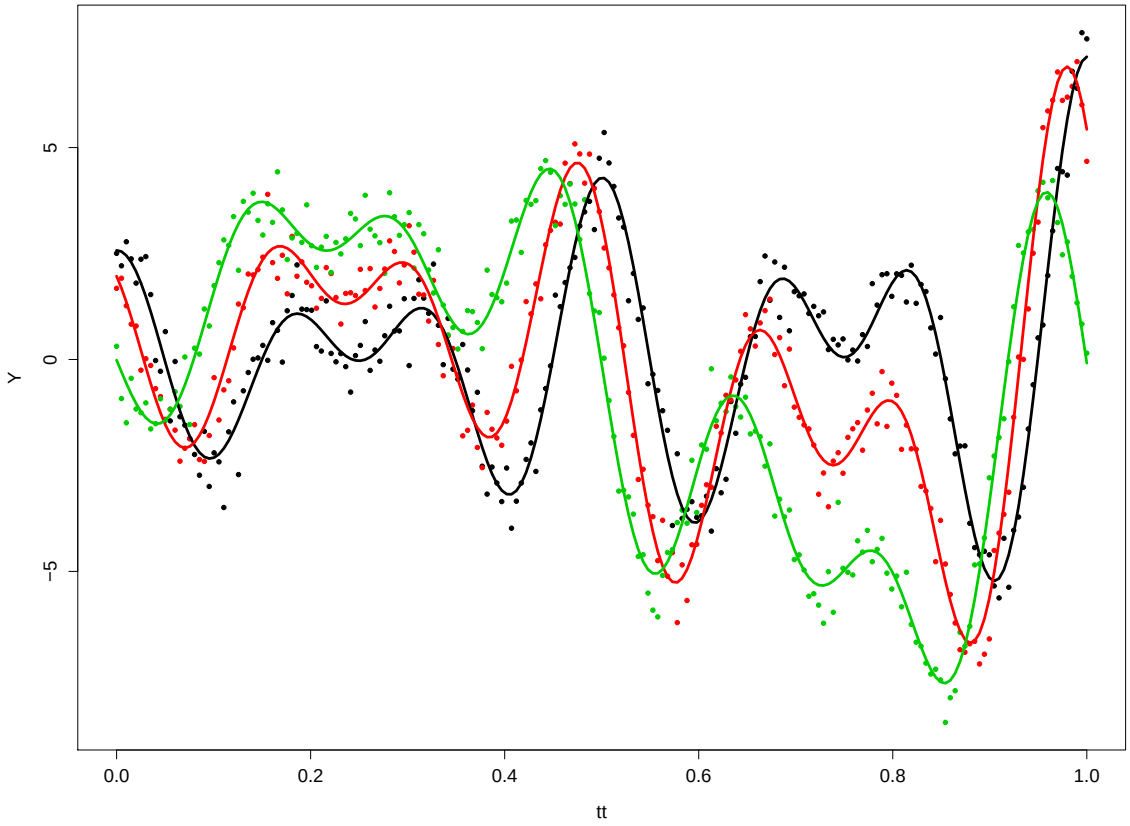

Figure 5.13: Estimation of the simulated data given in equation 5.99. Raw data are shown as dots whilst the lines represent the expected values over time. Black: First function; red: Second function; green: Third function.

---

given as

$$\mathbf{A} = \begin{pmatrix} \exp\rho_{01}\cos\upsilon_{01}\cos\phi_{01} & \exp\rho_{01}\cos\upsilon_{01}\sin\phi_{01} & \exp\rho_{01}\sin\upsilon_{01} \\ \exp\rho_{02}\cos\upsilon_{02}\cos\phi_{02} & \exp\rho_{02}\cos\upsilon_{02}\sin\phi_{02} & \exp\rho_{02}\sin\upsilon_{02} \\ \exp\rho_{03}\cos\upsilon_{03}\cos\phi_{03} & \exp\rho_{03}\cos\upsilon_{03}\sin\phi_{03} & \exp\rho_{03}\sin\upsilon_{03} \end{pmatrix} \quad (5.100)$$

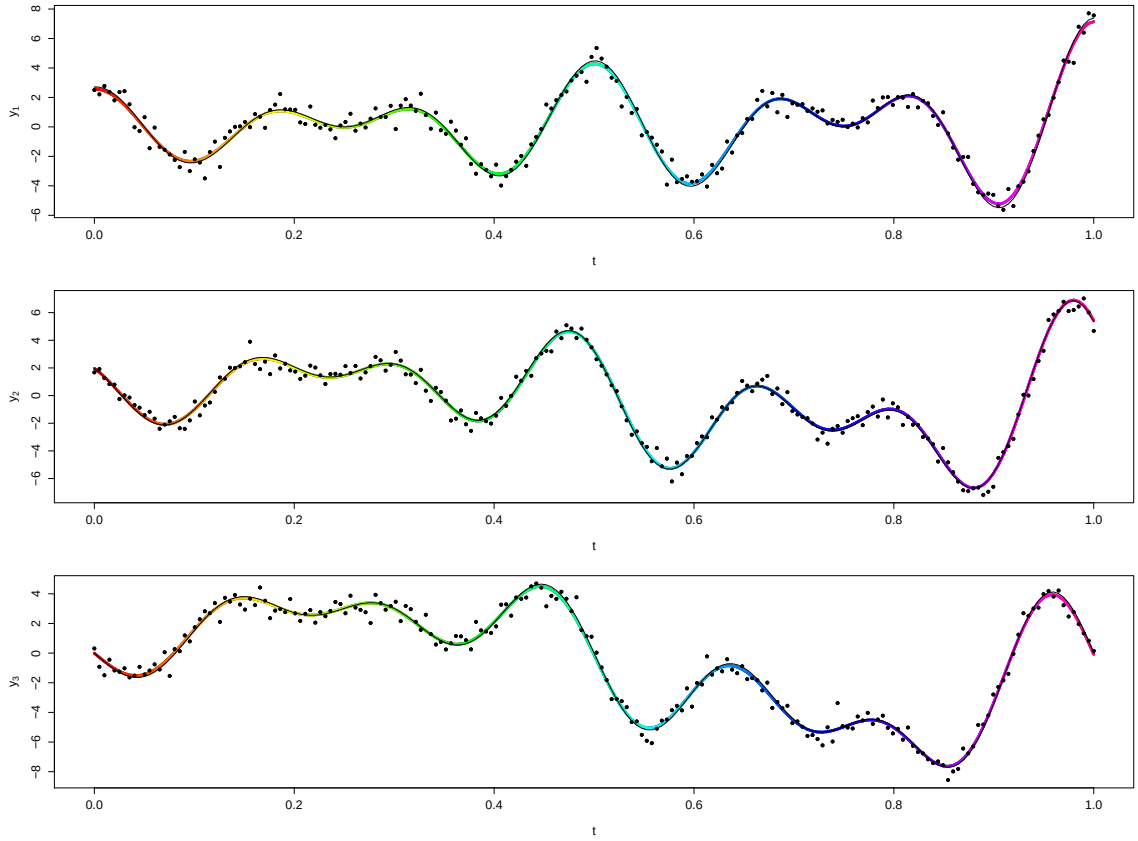

Figure 5.14: Estimation of the simulated data given in equation 5.99 one by one. Raw data are shown as dots whilst the colored lines represent the expected values over time. The thin black line shows the true trend.

---

to get the trend as a linear combination of the three basic three-dimensional loop structure functions, i.e.

$$\begin{pmatrix} \mu_1(t) \\ \mu_2(t) \\ \mu_3(t) \end{pmatrix} = \begin{pmatrix} g_1(t) \\ g_2(t) \\ g_3(t) \end{pmatrix} \mathbf{A}^T . \tag{5.101}$$

The noisy observations are then given by

$$\begin{pmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \end{pmatrix} = \begin{pmatrix} \mu_1(t) \\ \mu_2(t) \\ \mu_3(t) \end{pmatrix} + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \\ \epsilon_{3t} \end{pmatrix} \tag{5.102}$$

with iid normal errors $\epsilon_{it} \sim \mathcal{N}(0,\, 0.25)$ for each time point t and all three functions $i = 1, 2, 3$.

In figure 5.12, the trend is shown graphically. On the main diagonal the trend of each function is plotted over time. In the other sub-figures, the trajectory over time is plotted such that one can see that the trend has short-term fluctuations and a long-term trend, although in this simulation the center of the three-dimensional ball structure is constant, namely the point of origin. This means that the long-term trend is implicitly given by the zenith angle function.

The artificial time series have fitted quite well, as one can see for the three time series together in figure 5.13 and separately in figure 5.14. In the latter figure, the trend is shown as a solid black line whilst the estimated function is shown as a colored line. Given the complex structure, the estimation seems to fit the model quite well, showing that an empirical estimation might be promising at this stage.

6 Interpreting the Business Cycles Analysis

6.1 Business and Long-Phase Cycles in Inflation and Income Distribution

Econometric studies often focus on the methodological level as well as in empirical research on the problem of how to separate the business cycle from the trend in macroeconomic time series. However, economic growth theory in its advanced form provides us with insights into which economic ratios may exhibit a secular trend (like capital intensity when not measured in efficiency units) and which ones will not (like the output-capital ratio or the rate of employment as two measures of macroeconomic factor utilization). In contrast to a variety of econometric studies, macrodynamic growth theory therefore generally uses appropriate ratios or growth rates in its analytical investigations. In particular, ratios are used that allow for the determination of steady state positions and which therefore should not exhibit a trend in the very long run.

In applying the methodology developed in this work we will in fact concentrate on secularly trendless magnitudes, namely the employment rate on the external labor market, the wage share in national income, and the inflation rate (here of producers' prices). There are a variety of smaller as well as larger macrodynamic models in the tradition of Friedman (1968) and Goodwin (1967) which show the existence of persistent cycles in the interaction between the employment rate and the wage share on the one hand and the employment rate and the inflation rate on the other hand which tend to be long-phased when simple constant parameter estimates are used for their numerical investigation (see also Atkinson 1969). In these models the ordinary business cycle fluctuations must therefore be explained by something

else, namely by systematic variations in the parameters of the model which then add cycles of period lengths of about eight years to the fifty-years cycles these models are generating when used with average or constant parameter values. Based on earlier work (see Flaschel, Kauermann, and Teuber 2005) we now investigate the working hypothesis that there are long-phase cycles interacting with business cycles in the data as far as employment, income distribution, and inflation are concerned.

The method developed in this work now in fact allows us to check this hypothesis in a way much more refined than just using the Hodrick-Prescott filters with an arbitrarily given λ parameter. Moreover, we pursue the hypothesis in the spirit of the two-dimensional phase plots of the employment–inflation cycle and the employment–income distribution cycle of the literature on the Friedman inflation cycle and the Goodwin growth cycle.

Applying the technique developed, leads to the estimates shown in figure 6.1 and figure 6.2. First focus on inflation dynamics, figure 6.1. One can see that the unemployment rate is leading compared to the inflation rate in the long-phase cycle (the solid lines in the two time series plots top-left). In the bottom figure showing angle estimate $\hat{\phi}(t)$ we see moreover that there are approximately six business cycles surrounding these long-phase cycles, as $\hat{\phi}(t)$ crosses about six times the 2π full circle, marked as horizontal dashed lines. This finding is in line with Chiarella, Flaschel, and Franke (2005) and other work. The fitted angle also shows that the anti-clockwise rotation of the long-phase cycle is by and large also characterized by the business cycles surrounding it, though there are exceptions to this rule (periods at the beginning and the end of the considered time span); see also the figure top-right. Note that we follow the tradition here which uses the unemployment rate in place of the employment rate on the horizontal axis (the latter would give

rise to an anti-clockwise orientation of the business and the long-phase cycles shown in these figures).

The long-phase cycle (bottom left plot of figure 6.1) indicates that 50 years of data are indeed needed in order to get the indication of the existence of such a cycle. We observe that long periods where unemployment and inflation are both rising (i.e., where stagflation occurs) and also periods where the opposite takes place and therefore falling unemployment rates do not lead to rising inflation rates immediately. We stress again that our extraction of the business cycle component as shown in figure 6.1, through a phase as well as a radius plot, is an integral part of our treatment of the long-phase evolution of the economy.

With respect to the other long-phase cycle model, the Goodwin (1967) growth cycle model, we now have to look at figure 6.2. As far as the evolution of the wage share (top-left plot) is concerned we now have more volatility, as was the case with the inflation rate. This may be due to the involvement of labor productivity as a constituent part of the definition of the wage share. Nevertheless one can see a single long-phase cycle in the solid line shown in the time series presentation of the wage share. Again, the employment rate is leading with respect to this long-phase cycle in the wage share. We know from Goodwin (1967) and the numerous articles that followed his approach that the interaction of the employment rate with the wage share is generating a clockwise motion. In this regard, we can confirm that the cycles of business cycle frequency are moving in a clockwise fashion, as is suggested by the again basically downward sloping angle line shown bottom-right in figure 6.2. To the right of this figure we see again (if minor cycles are ignored) now by and large seven business cycles overlaid over the long-phase cycles, as also shown in the figure bottom-right. Looking at the long-phase cycle (bottom-middle

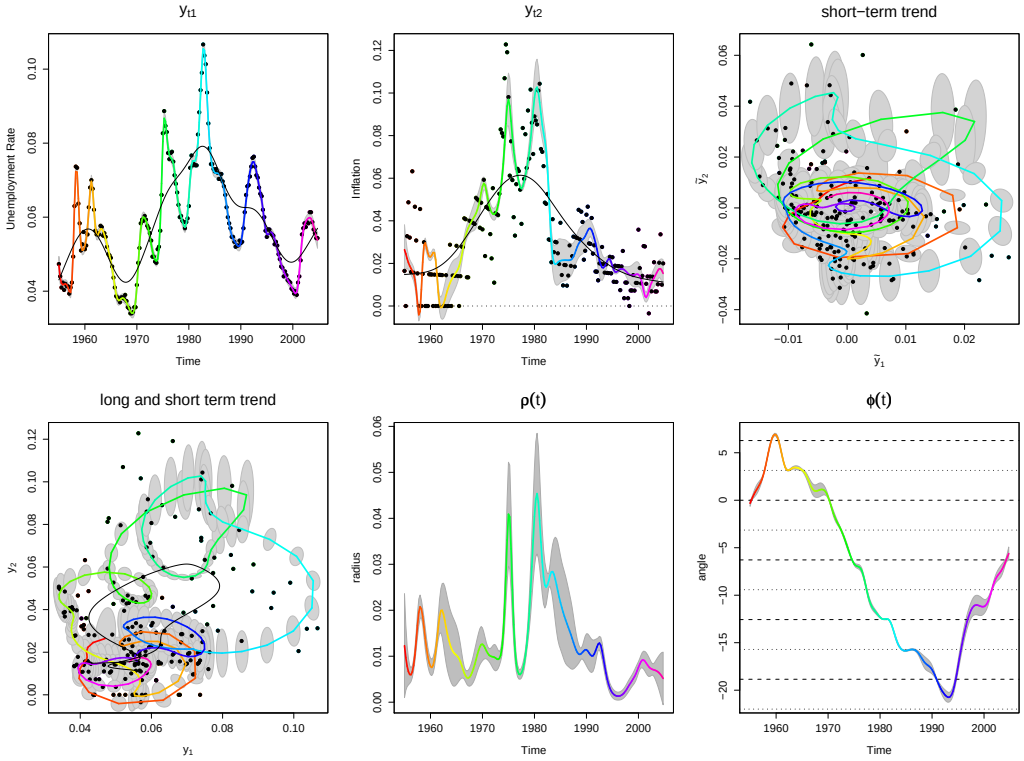

Figure 6.1: Estimation of unemployment rate and price inflation using a circular regression approach: The upper left-hand and middle plot: Observations (black dots), estimated long-term trend (black line), and estimated short-term fluctuations around long-term trend (colored line) with confidence regions (grey-shaded area) over time for first and second time series, respectively. Upper right hand plot: Detrended observations (black dots), trajectory of short-term trend (colored line), and their confidence regions (grey-shaded area). The lower left-hand plot: Observations (black dots), long-term trend (black line), trajectory of short-term trend around long-term trend (colored line), and confidence regions (grey-shaded area). Middle and right-hand plot: Radius and angle over time (colored line) and their confidence regions (grey-shaded area), respectively.

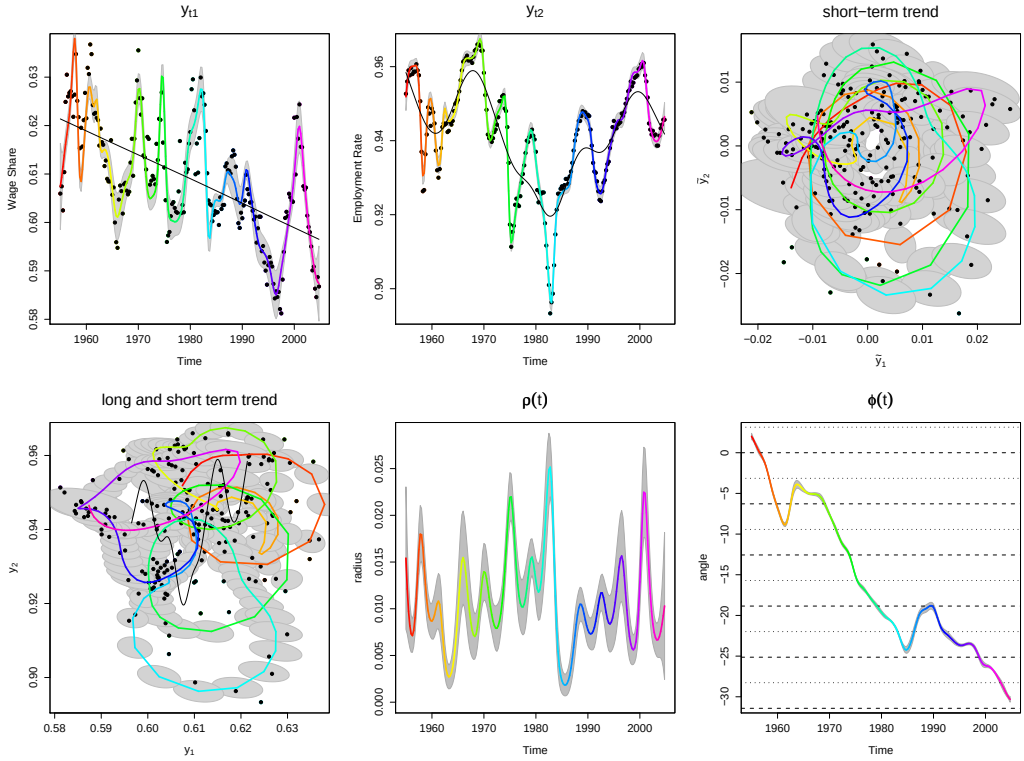

Figure 6.2: Estimation of wage share and employment rate using a circular regression approach: The upper left-hand and middle plot: Observations (black dots), estimated long-term trend (black line), and estimated short-term fluctuations around long-term trend (colored line) with confidence regions (grey-shaded area) over time for first and second time series, respectively. Upper right hand plot: Detrended observations (black dots), trajectory of short-term trend (colored line), and their confidence regions (grey-shaded area). The lower left-hand plot: Observations (black dots), long-term trend (black line), trajectory of short-term trend around long-term trend (colored line), and confidence regions (grey-shaded area). Middle and right-hand plot: Radius and angle over time (colored line) and their confidence regions (grey-shaded area), respectively.

plot) we indeed see a cycle that is nearly closed (and thus approximately of fifty years' length) and that is moving clockwise as suggested by the simple Goodwin (1967) growth cycle model (see Solow 1990 for early comments on an empirical phase plot of this cycle) and its many extensions.

We conclude that the method developed in this paper provides a helpful approach to the separation of long-phased cycles that describe the evolution from high to low inflation regimes and from high to low wage share regimes from cycles of business cycle frequency. This method therefore allows in a distinct way the discussion of long waves in inflation and income distribution in modern market economies after World War II.

6.2 Are the US Business Cycles Real Cycles?

In this section the focus is primarily on the short-term fluctuations of the economic time series. Although, as has been shown in the previous sections, the specification of the long-term trend is essential for the estimation of the short-term trend, here the focus lies on the discussion of the short-term trend, with the long-term trend taken as given. The short-term trend is presented in the figures 6.3 and 6.4. One should remember that the scale of the short-term trend and the residuals of the long-term trend (i.e. observation minus long-term trend) does not account for absolute values but for the deviation from the secular trend. This would mean that the plots in figure 6.3 measure the percentage point deviation from the moving long-term trend, which was already shown in the previous section.

Focussing on the short-term trend of the price Phillips curve, namely figure 6.3, one can see eight different time frames in which the 50 year history of data points and estimations have been separated. Unfortunately, the separa-

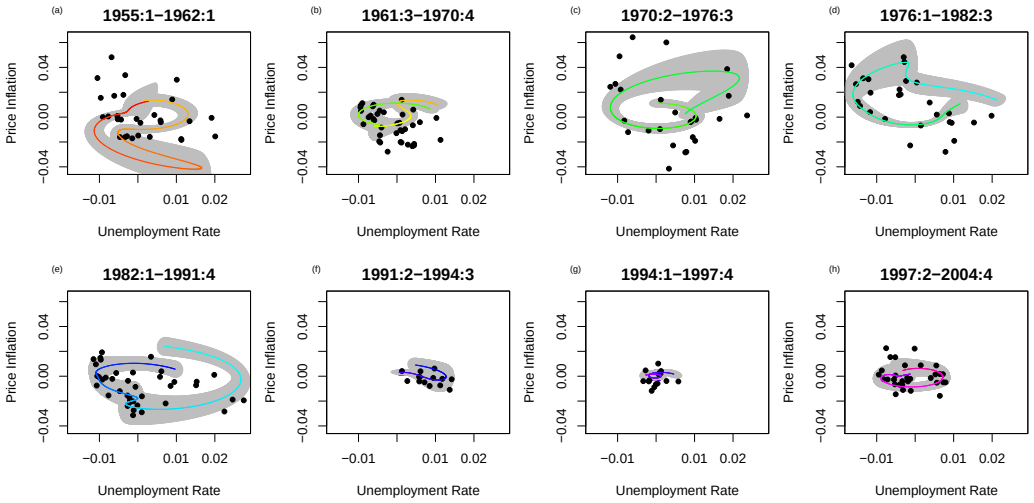

Figure 6.3: Detrended price inflation and unemployment rate and estimated short-term fluctuations over time: Detrended observations (dots), estimated short-term trend (colored lines), and confidence regions (grey-shaded areas) separated into different time periods: (a) 1955Q1-1962:Q2, (b) 1961Q3-1970Q4, (c) 1970Q2-1976Q3, (d) 1976Q1-1982Q3, (e) 1982Q1-1991Q4, (f) 1991Q2-1994Q3, (g) 1994Q1-1997Q4, and (h) 1997Q2-2004Q4.

---

tion into the time frames is arbitrary, or better said subjective, because the business cycles are, as already mentioned and shown in this work, not perfect circles, meaning such that the end points are not connected. Although with the goodwill to imagine a circle and accounting for estimation error, it is hard to deny (even in a statistical sense) for most of the time frames that the cycles are connected or at least that the confidence regions of the end points do overlay.

One way to start separating the history of data into different time frames could be to select predefined economic states, for instance ”start of the boom

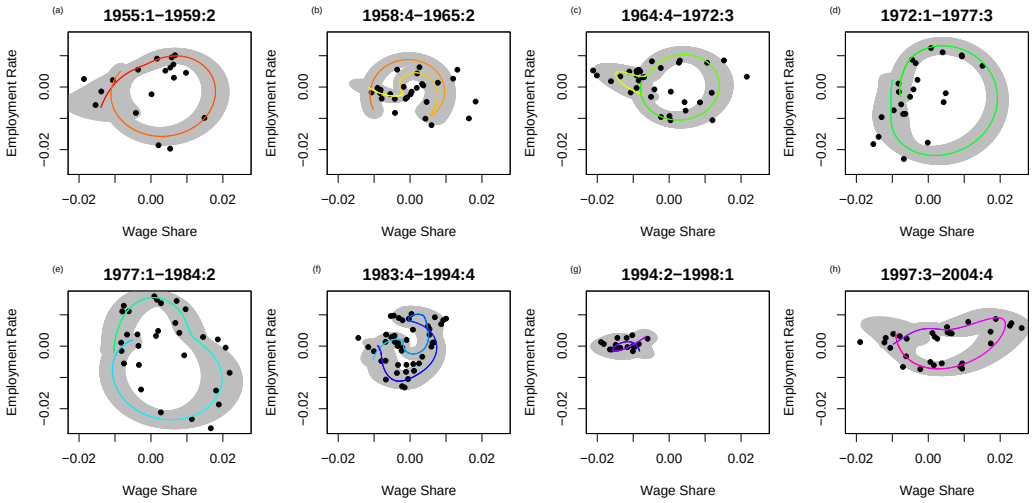

Figure 6.4: Detrended wage share and employment rate and estimated short-term fluctuations over time: Detrended observations (dots), estimated short-term trend (colored lines), and confidence regions (grey-shaded areas) separated into different time periods: (a) 1955Q1-1959:Q2, (b) 1958Q4-1965Q2, (c) 1964Q4-1972Q3, (d) 1972Q1-1977Q3, (e) 1977Q1-1984Q2, (f) 1983Q4-1994Q4, (g) 1994Q2-1998Q1, and (h) 1997Q3-2004Q4.

---

period", "end of the boom period", or "start of recession". However, the objective definition of each such economic state would be problematic, taking into account that most researchers and research institutes define the economic states with the help of the real GDP growth, which has been ignored in this example. Furthermore, due to the fact that the cycles are not perfect in the sense of repeating the same trajectory from one cycle to the other cycle, the separation using the predefined economic states might be inconclusive. Nevertheless, for the sake of progress, the separation was done by eye, meaning that those time points which seemed to fit to describing a cycle - at least from the eye (of the author) - were selected. But it is still

debatable how the time frames were selected and even if the selected cycles qualify as a "full business cycle". Looking at plot (a) in figure 6.3, one could argue that the shown plot could qualify to account for one or even two cycles.

However, taking the cycle shown in plot (a) in figure 6.3 as given, one can see that the one and only cycle is turning counter-clockwise, meaning that the unemployment rate is leading the price inflation. Interestingly, comparing plot (a) with plot (b), the time frame for the second cycle is much longer although the magnitude (radius) is much lower. Normally one might expect, as in plots (c), (d), and (e) which show three different, but nearly similar, circular patterns with a larger radius, that these trajectories would take much longer to be formed than for those cycles which are narrower. It is doubtful if the circles in plots (f) and (g) would qualify as "full business cycles", due to their length and magnitude. However, the last cycle in plot (h) would at least qualify as a full cycle for which the confidence regions of the end points overlay.

Looking at the different cycles in figure 6.3, one can see that the rotation of the cycles changes over time. In the first period from 1955 to 1962, shown in plot (a), the cycles are turning counter-clockwise, which means that the unemployment rate is leading price inflation. In the second period from the end of 1961 to 1994, shown in plots (b) to (f), the cycles now turn clockwise, which means that the price inflation is leading the unemployment rate. In the last time frame from 1994 until now, shown in plots (g) and (h), the cycles are turning counter-clockwise, which means that the unemployment rate is leading the price inflation again.

Focussing on the wage share/employment rate short-term fluctuations shown in figure 6.4, one can determine eight circles, similar to the previous case. It

should be remarked that the cycles for the wage share/employment rate differ in terms of the pattern, magnitude, timing, and turning rotation compared with price inflation/unemployment rate short-term cycles. The cycles shown in plots (a), (c), (d), (e), and (h) are nearly perfect cycles with overlapping end point confidence regions. However, the magnitudes and lengths of cycles do differ from four to eight years. Although the cycle in plot (g) is complete in the sense of overlapping confidence regions, it is debatable if the cycle would qualify as a "full business cycle" from an economic as well as a graphical and statistical point of view. Then again, plots (b) and (f) are not perfect cycles from a pattern point of view, but would qualify, with greater conviction, as business cycles compared to the previously described ones, as shown in plot (g). Another aspect which makes the cycle in plot (g) unique compared to the other ones is the turning rotation: Only the cycle in plot (g) is turning counter-clockwise, while all other cycles are turning clockwise. This means that the employment rate is leading the wage share.

6.3 Estimating the Leading, Coincident, and Lagging Indicators Using Generalized Two-Dimensional Loops

In this chapter, the analysis is focussing on the leading, coincident, and lagging indicators defined by the Conference Board and described in chapter 2.4. Due to the fact that three time series are estimated simultaneously, Teuber (2012b) used the generalized two-dimensional loop model presented in chapter 5.2.3.

For the estimation of the model in equation (5.55) the growth rates of the three leading, coincident, and lagging indicators, defined by the Conference Board, from January 1960 to December 2010 have been used. Here, the growth rates are defined as the 12-month differences of the logarithmic val-

ues of each indicator.

Instead of using an ordinary Hodrick-Prescott filter, a penalized spline regression with autocorrelated residuals is preferred. Obviously, the residuals resulting from an Hodrick-Prescott filter are autocorrelated in this case such that the Hodrick-Prescott filter should not be used from a statistical point of view if the autocorrelation exceeds a value of roughly 0.3; see Krivobokova and Kauermann (2007) for a discussion on the effect of misspecified autocorrelation structures for penalized splines and therefore the Hodrick-Prescott filter.

For the estimation, several long-term centers γ have been estimated on a grid of different degrees of freedoms and then the generalized loops regression technique was used to fit the remaining deviations from the long-term trend. Here, the best choice will be shown, remembering that the long-term fluctuations do not come out of the blue, but have been selected using a hybrid approach.

The process of separating the long-term trend out of the given data is delicate because of the trade-off between short-term and long-term fluctuations. The more structure one is allowing for the long-term trend, the less profound the short-term fluctuations might be. Assuming a normal Hodrick-Prescott filter, then more or less the average values of the observations around the current time points will be used to estimate the best fit. This means that it will be assumed that the short-term trend values around these observation time points cancels each other out. However, if the short-term trend is fluctuating around the long-term trend in waves, a misspecified window for the time points used, might lead to completely misspecified long-term trends, such that the short-term fluctuations might not be estimated correctly any-

more. Therefore, firstly an autocorrelated residual structure will be assumed and secondly a judgemental degree of freedom for the selection process is important, meaning that the structure of the long-term trend - as well as the Hodrick-Prescott filter - should not be used blindly.

In figure 6.5, the growth rates for the indicators are shown as dots and the corresponding long-term fluctuations as solid lines (leading indicators: Green; coincident indicators: Black; lagging indicators: Red). It should be noted that the long-term fluctuations differ from each other but the cyclicality character can still be observed. Furthermore, fluctuations of longer frequencies can be seen in the long-term trends; for example the long-term trend of growth rates is higher in the '60s and '90s and lower around the late '70s/early '80s and starts declining at the beginning of the new century. This means that similar to the work of Kauermann, Teuber, and Flaschel (2012), here the long-term fluctuations form a full cycle with a length of roughly 40 to 50 years.

Focussing now on the short-term fluctuations calculated using the "detrended" growth rates, the resulting short-term fluctuations are shown in figure 6.6 (leading indicators: Green; coincident indicators: Black; lagging indicators: Red). It is worth noting that the two-dimensional loop structure is capable of fitting the three economic time series. This means that the underlying business cycles can be described using polar coordinates and, furthermore, this means that the time series follow each other on this cyclical pattern, meaning that the terms "leading", "coincident", and "lagging" are well chosen not just from a timely point of view but also from a cyclical point of view. Figure 6.7 shows how well each individual time series has been fitted using the generalized two-dimensional loops approach.

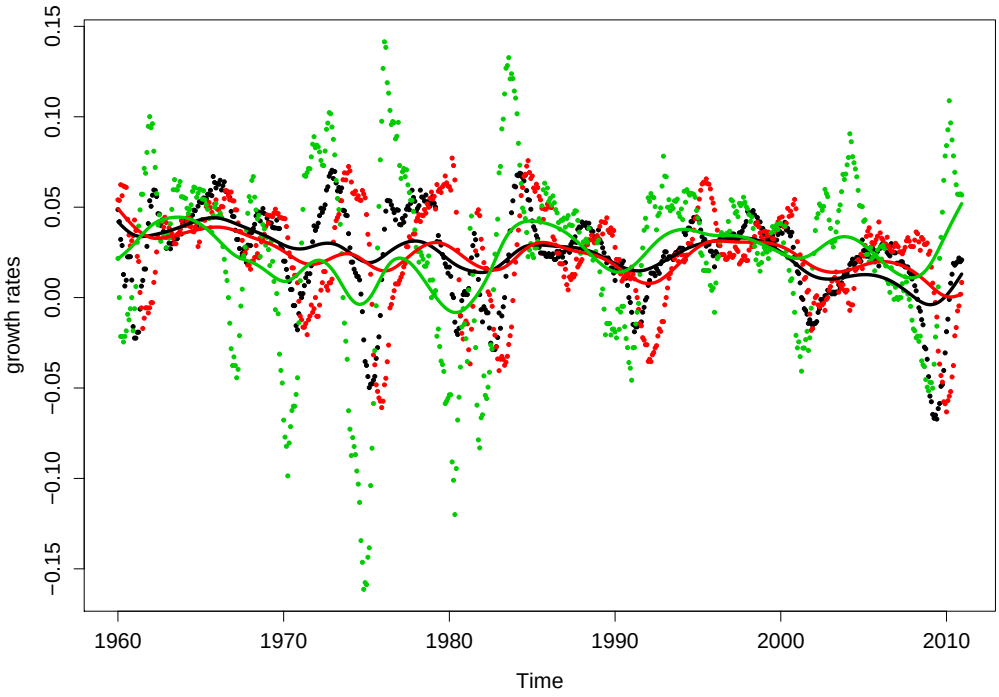

Figure 6.5: Raw growth rates (12-month logarithmic differences) as dots and estimated long-term trend as solid lines. Green: Leading indicators; black: Coincident indicators; red: Lagging indicators. Source: Conference Board, January 1960 - December 2010.

The shift of the angle for the leading indicators compared to the coincident indicators has been estimated to be $\hat{\phi}_{02} = -1.170$ with an estimated standard error of 0.0339. The lagging indicators lag the coincident indicators, based on the shift in the angle by $\hat{\phi}_{03} = 1.324$ with an estimated standard error of 0.0372. This would mean that if a full cycle takes five years, then the leading indicators lead roughly by 0.93 years and the lagging indicators lag the coincident indicators by roughly 1.05 years. However, the shift in the

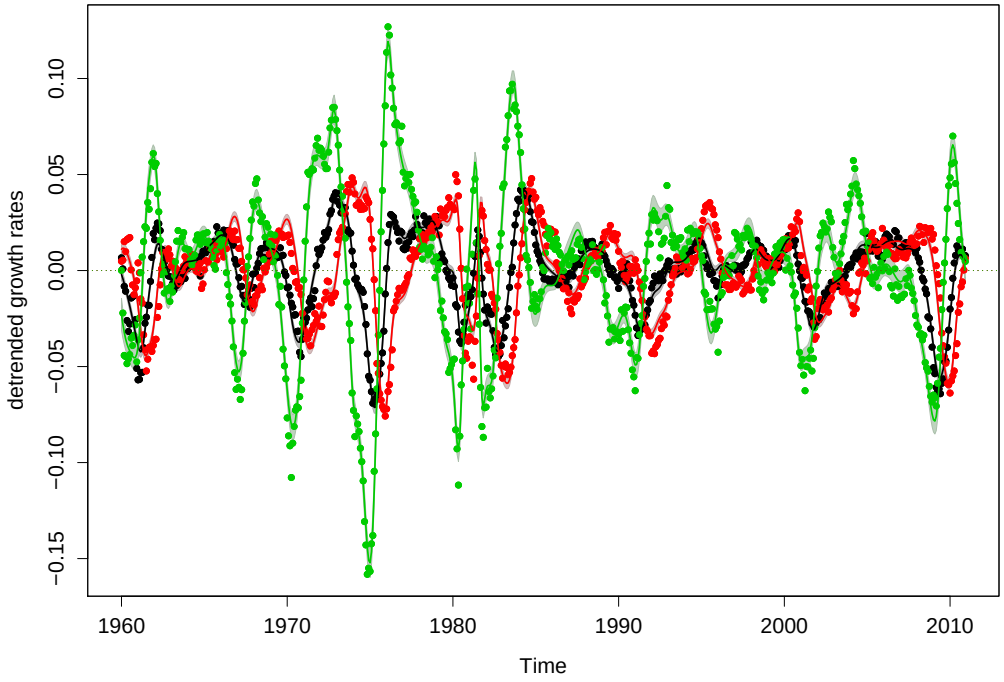

Figure 6.6: Detrended growth rates (dots) and estimated short-term trends (solid lines). Green: Leading indicators; black: Coincident indicators; red: Lagging indicators.

---

angle is set to be fixed while the frequency/speed of the cycles varies over time.

In figure 6.8, the trajectories of the two-dimensional short-term trends are shown for each combination of the three time series. The main structure, i.e. the angle and radius, can be seen in each sub-figure, however the plots differ slightly due to the shift in the angle and the radius as well as the non-constant cyclicality pattern. The shift of the angle can be seen by the tilt of

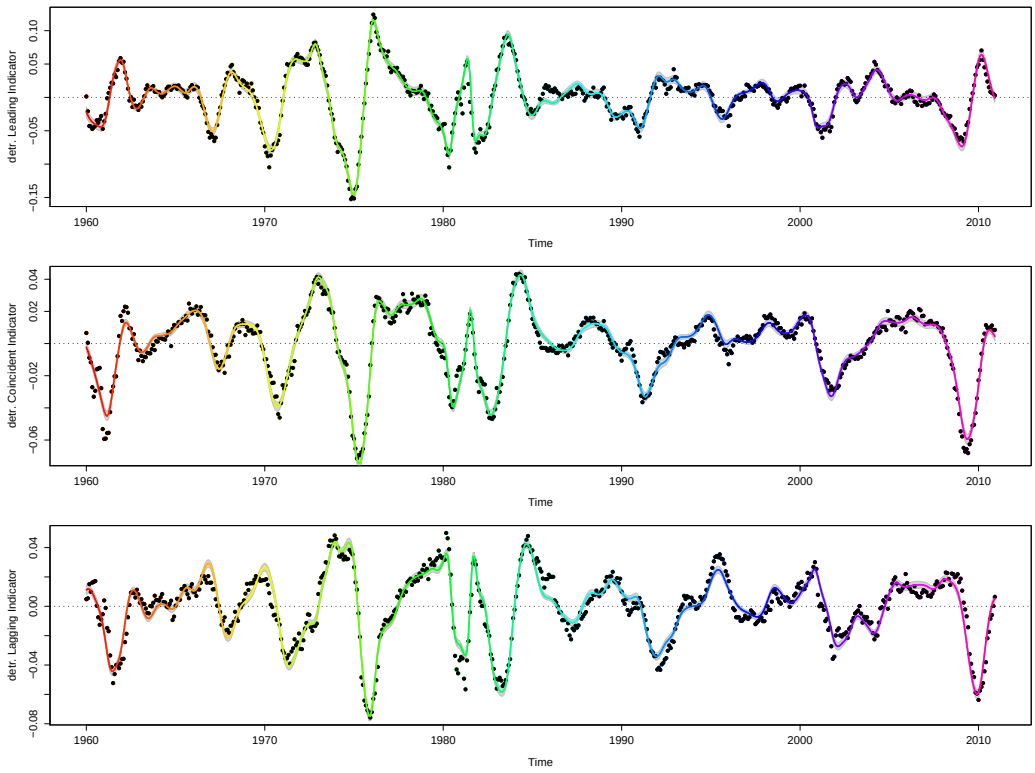

Figure 6.7: Detrended growth rates (dots) and estimated short-term trend (solid lines). First panel: Leading indicators; second panel: Coincident indicators; third panel: Lagging indicators.

---

the different cycles while the shift of the radius can only be seen by looking at the scale of the axes. It is interesting that the lagging indicators and the leading indicators both have a positive radius shift, i.e. $\hat{\rho}_{03} = 0.4803$ with an estimated standard error of 0.0311 and $\hat{\rho}_{02} = 1.0751$ with an estimated standard error of 0.0273. This means that the fluctuations of the growth rates for the lagging indicators have a 61.6% stronger magnitude than those of the coincident indicators, and the magnitude of the leading indicators growth rates fluctuations are roughly 193% stronger than those of the coincident indica-

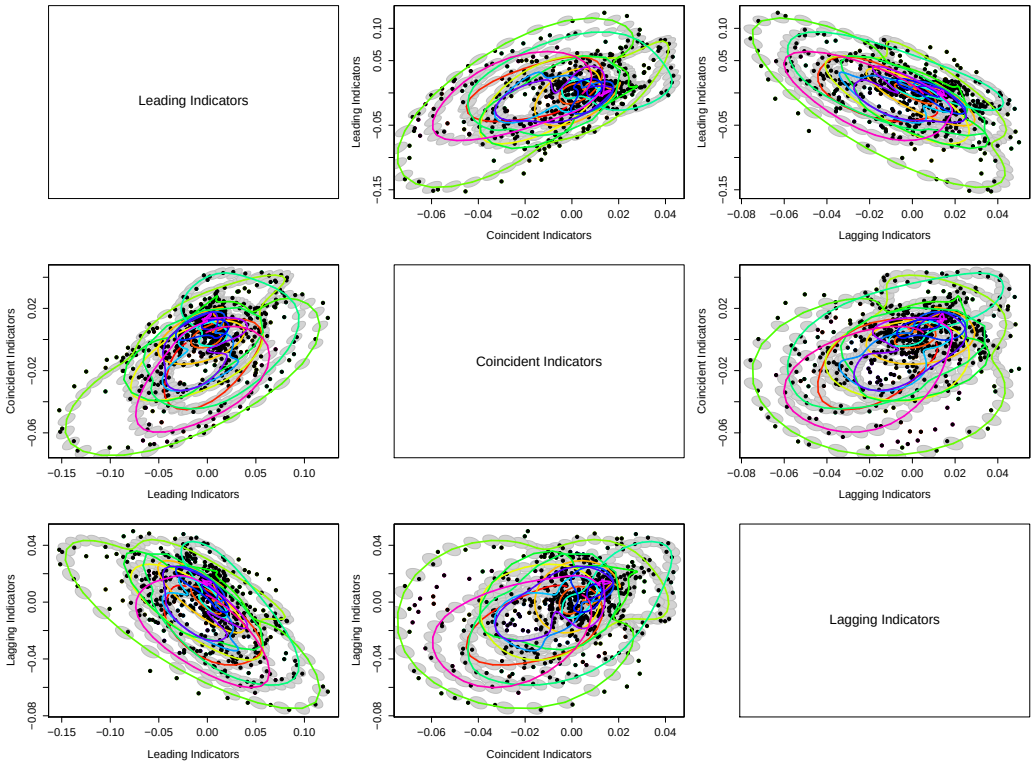

Figure 6.8: Two-dimensional trajectories of leading, coincident, and lagging indicators over time. Estimated function: Rainbow-colored lines; confidence region: Grey-shaded ellipse.

---

tors. The fact that the leading indicators fluctuate from a magnitude point of view more strongly than the coincident indicators comes as no surprise, however it is not the naive supposition that the lagging indicators fluctuate more strongly than the coincident indicators. However, the same pattern can be observed for the estimated standard deviations of the residuals, i.e. $\hat{\sigma}_1 = 0.004650$ for the coincident indicators, $\hat{\sigma}_2 = 0.009567$ for the leading indicators, and $\hat{\sigma}_3 = 0.007459$ for the lagging indicators.

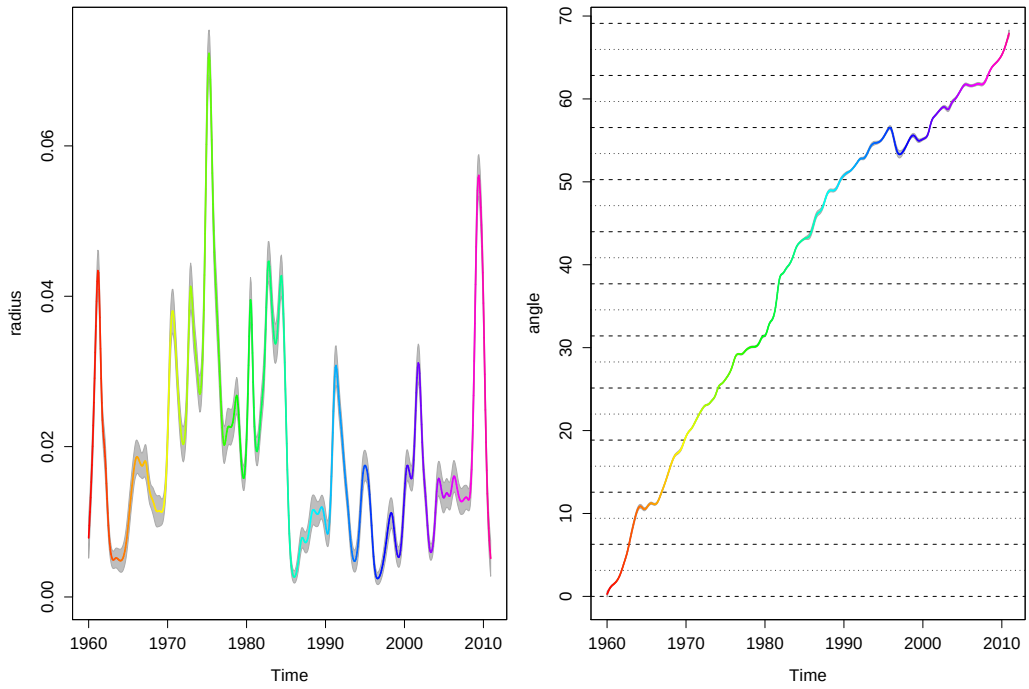

Figure 6.9: Left-hand-side: Estimated radius function over time. Right-hand-side: Estimated angle over time.

Looking at figure 6.9, on the left hand side the fitted radius and on the right hand side the fitted angle have been plotted and are both wiggly, meaning that no cycle is alike nor is any circle a perfect cycle with constant radius or constant speed. This means that a non-linear radius and angle function was necessary to fit the three time series using the generalized two-dimensional loops approach. The National Bureau of Economic Research (NBER) cites the business cycle dates for the last 150 years, meaning that reference dates for peaks and troughs as well as the duration of business cycles are available. The Business Cycle Dating Committee normally determines business cycle

peaks and troughs with a publication lag of roughly one year, so at the moment eight cycles have been identified by the committee for the data at hand.

In figure 6.10, the time intervals, starting from the current NBER peak until reaching the corresponding NBER trough, has been shaded grey. The figure shows that the main downward slowing trends can be measured quite well using the coincident indicators. However, not all steep declines have been determined by the NBER Business Cycle Dating Committee. Normally, most of the academic research on business cycles would stop at this point, but the approach presented here might shed more light on the insights of the business cycles. Therefore, not only the fitted short-term fluctuations and raw data might be used to determine the stage of the cycles but also the fitted angle.

Due to the fact that the sinus and cosine functions have a periodicity of 2π, the angle function is invariant to the addition of multipliers of 2π. Therefore, in figure 6.11 the angle function will be plotted on the $[0, 2\pi]$ interval.

Focussing on the angle at the recession start dates (reported NBER peak) and on the angle at the recession end dates (reported NBER trough), the estimated angles and their confidence regions are plotted in figure 6.12. Looking at the plot on the left hand side, one can see that the angles at the NBER peaks vary a lot. However, the average upper confidence region is slightly negative (shown as upper broken line in the figure) and the average lower confidence regions is slightly negative (shown as lower broken line in the figure). Furthermore, the statistical null hypothesis that the angle at NBER peaks times is zero cannot be rejected at a 95% significance level. Looking at the angle during reported NBER trough times, one can see that the estimated angles and their confidence regions are much narrower and fluctuating around 0.5π. The average upper confidence level is slightly above 0.5π while

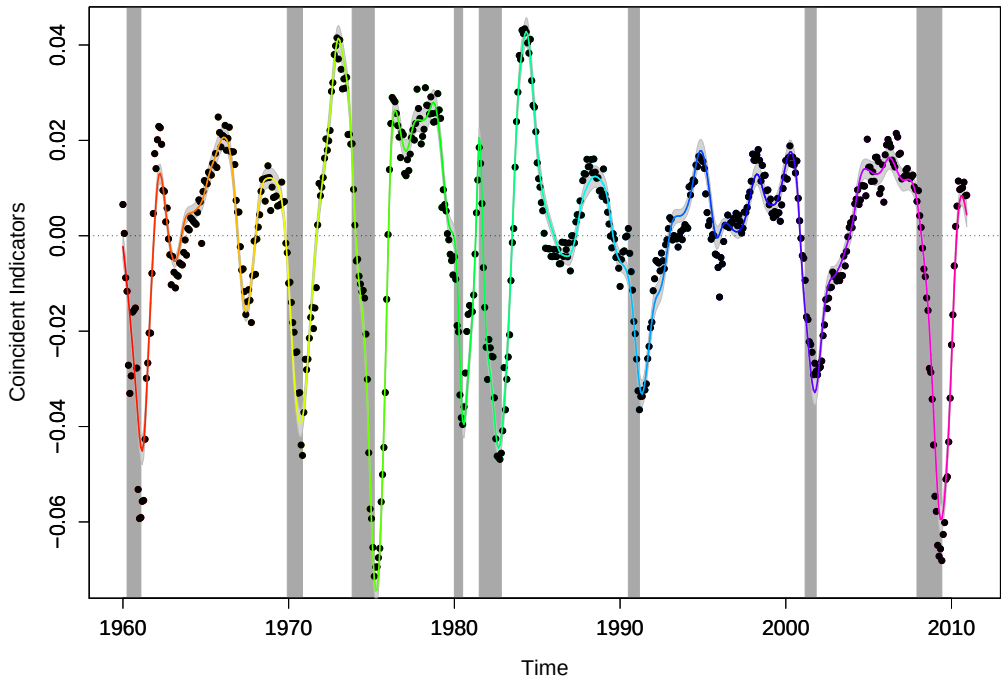

Figure 6.10: Detrended growth rates of coincident indicators. Dots: raw data; rainbow-colored line: Estimated short-term trend; grey-shaded area: NBER recession times.

---

the average lower confidence region is slightly below 0.5π. The statistical null hypothesis that the angle at reported NBER troughs coincides with an angle of 0.5π cannot be rejected at a 95% significance level.

Following the results, this would mean that an angle between 0 and 0.5π would suggest a recession. Going one step further, one might conclude that an angle between 0.5π and π would suggest a recovery, an angle between π and 1.5π a boom time, and an angle between 1.5π and 2π a contraction time.

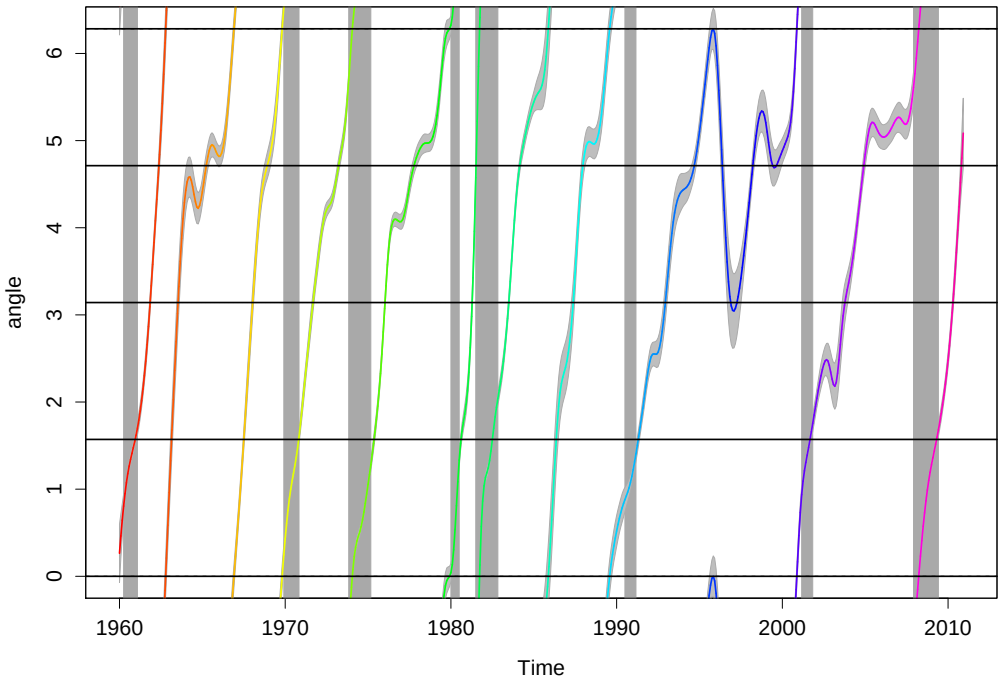

Figure 6.11: Estimated angle over time on the $[0; 2\pi]$ interval. Grey-shaded area: NBER recession times.

This means that, given the model, one might not only classify the current stage of the economy into the four business cycle phases, but one can even instantly classify within the class the stage of the cycle; for instance one can differentiate just by the angle if at the current time point the business cycle is in an early recovery or in a late recovery phase.

Now, let us go one step further: If the thresholds defined above are correct, then one could easily classify the different business cycle stages just by using the estimated angle. Therefore, with the help of figure 6.11 and the

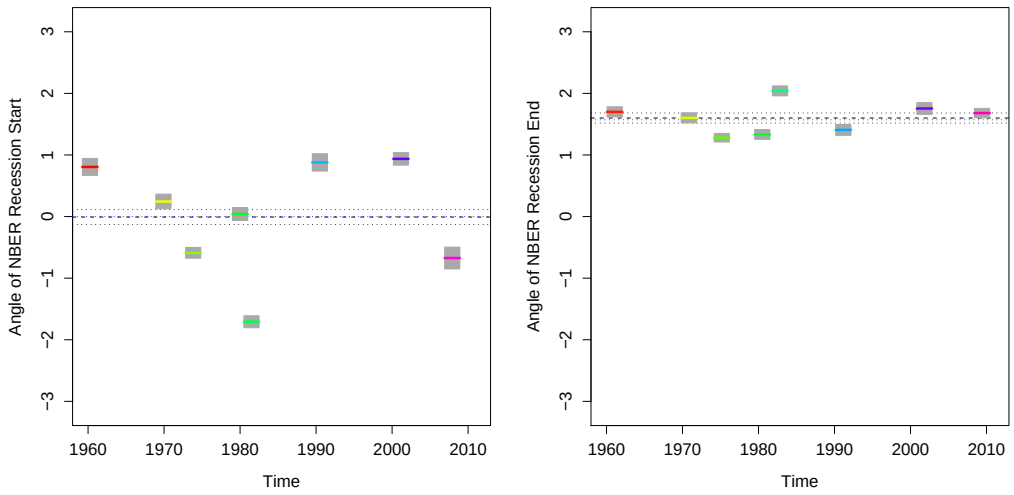

Figure 6.12: Left-hand-side: Estimated angle at NBER recession start dates. Right-hand-side: Estimated angle at NBER recession end dates. Rainbow-colored line: Estimated angle; grey-shaded area: Confidence regions. Upper broken lines: Average upper confidence region; lower broken lines: Average lower confidence regions.

---

data at hand, one would get the classifications shown in table 6.1.

It should be noted that cycles 2, 3, and 8 in table 6.1 have not been determined by the NBER Business Cycle Dating Committee. Furthermore, the cycles ”9-1” and ”9-2” show an irregular pattern in that the angle is turning the rotation twice, meaning that the regular stages ”recession-recovery-boom-

Cycle	recession	recovery	boom	contraction
1	1960:1	1960:12-1961:1	1961:11	1962:6
2	1962:10-1962:11	1963:2-1963:3	1963:7-1963:8	1964:2-1966:2
3	1966:12	1967:7	1968:1-1968:2	1968:10-1969:2
4	1969:11	1970:11	1971:9	1973:3-1973:4
5	1974:2	1975:5	1977:9	1977:12
6	1980:8	1981:4	1981:7	1981:10
7	1982:7	1983:7	1984:3	1985:10-1986:1
8	1986:4-1986:6	1987:5-1987:6	1987:12-1988:2	1989:7-1989:9
9-1	1991:4-1991:5	1992:12-1993:1	1994:8-1994:10	1995:9-1996:1
9-2		1996:10-1997:7	1998:3-1999:12	2001:1
10	2001:9-2001:10	2003:9-2003:12	2004:12-2005:1	2008:4
11	2009:4-2009:5	2010:5	2010:10-2010-12	
	$\hat{=}$ NBER trough			$\hat{=}$ NBER peak

Table 6.1: Start dates of various business cycles stages. Recession start corresponds to NBER troughs. contraction start corresponds to NBER peak.

contraction" does not hold and instead the pattern "recession-recovery-boom-contraction-boom-contraction-recession" had been observed. This means that the cycles "9-1" and "9-2" might qualify as only one full cycle. In figure 6.13, the nine cycles are separately plotted for the coincident vs. lagging indicators. One can see directly that no cycle is alike, in terms of speed, magnitude, duration, location, and "completeness". However, the pattern can be observed quite well and it seems worth noting that the model captures the cyclicality and sheds new light onto the business cycle discussion.

Another aspect which will be captured by the NBER is the duration of a business cycle. Given their data, one can see not only that each dura-

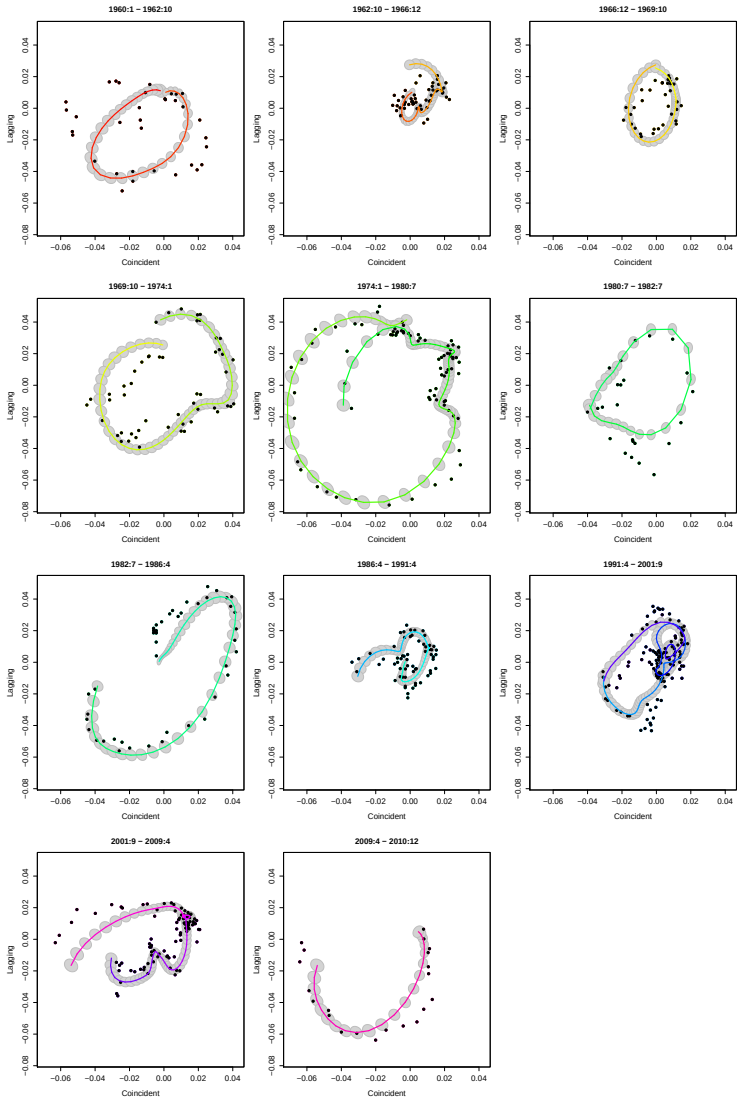

Figure 6.13: Two-dimensional trajectories of estimated short-term fluctuations of the coincident and lagging indicators. Colored line: Estimated short-term trend; grey-shaded area: Confidence ellipse; dots: Detrended growth rates. Plots are split to cover the eleven business cycles defined in table 6.1.

tion differs from cycle to cycle but also that the long-term trend changes over time. For instance, the recession duration post-World War II cycles is shorter on average than in the older days (1854-1919: 22 months; 1919-1945: 18 months; 1945-2009: 11 months). However, the cycle length is expanding based on peak-to-peak (1854-1919: 49 months; 1919-1945: 53 months; 1945-2009: 66 months) and trough-to-trough (1854-1919: 48 months; 1919-1945: 53 months; 1945-2009: 73 months) measurements. It is interesting to note, that the lengths of cycles differ based on the measurement, which is plausible if one remembers the functional form of the estimated angle.

Given the possibility of classifying any stage of the business cycles with the help of the estimated angle functions, one could easily approximate the duration of a cycle for each time point. However, there is a much more elegant way to calculate the instantaneous frequency of a business cycle. Assuming a cycle with a linearly increasing/decreasing angle, for instance $g_1(t) = \rho(t)\cos\phi(t)$ with $\phi(t) = p \cdot t$ and $\rho(t) > 0$, the corresponding duration of a cycle is $\frac{2\pi}{|p|}$. More generally, the instant business cycle duration will be defined as

$$\text{duration}(t) = \frac{2\pi}{|\phi'(t)|} \quad \text{if } \phi'(t) \neq 0 \tag{6.1}$$

for any point in time t.

Given the B-spline structure, the estimation of the first derivative can be calculated using the B-splines basis $\mathbf{B}'(t) = \frac{\partial \mathbf{B}(t)}{\partial t}$ and the estimated coefficients $\hat{\boldsymbol{\theta}}$ such that one can estimate the duration by

$$\widehat{\text{duration}}(t) = 2\pi \cdot |\mathbf{B}'(t)\hat{\boldsymbol{\theta}}|^{-1} \quad \text{if } \mathbf{B}'(t)\hat{\boldsymbol{\theta}} \neq 0 \tag{6.2}$$

for any point in time t.

Unfortunately, the estimation of the angle function is too wiggly to get any

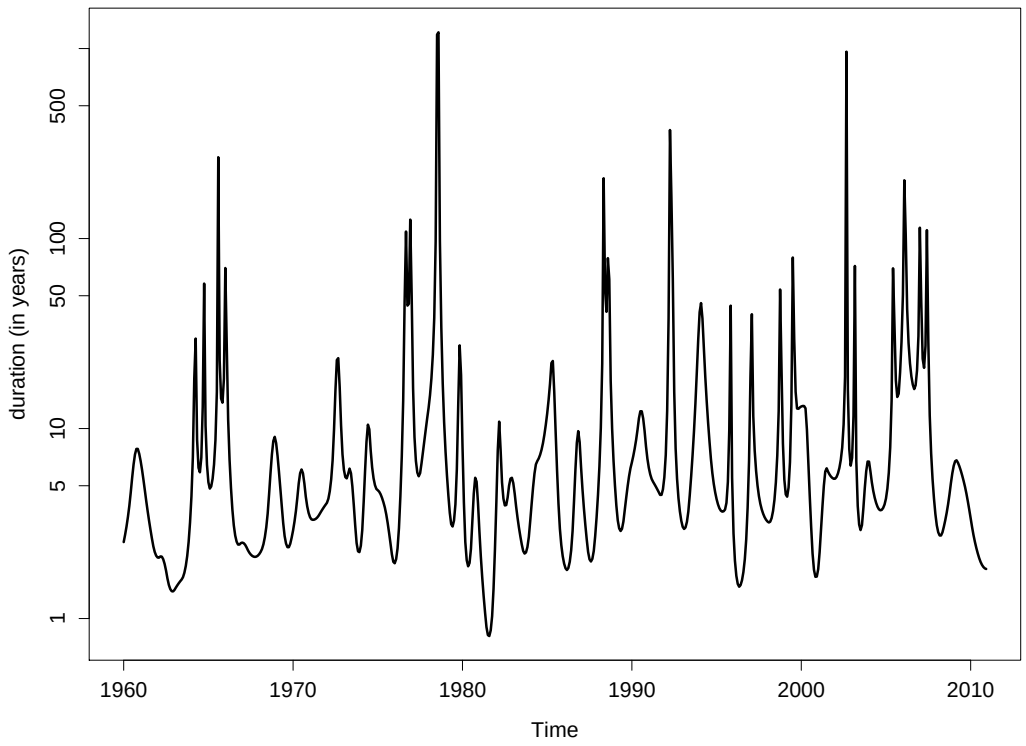

Figure 6.14: Estimated instantaneous duration of business cycles using the estimated angle.

---

useful information out of the function to approximate the duration; see therefore figure 6.14. But if the angle function is approximated using a function with fewer degrees of freedom, then the corresponding estimated duration function will be less wiggly, as one can see in figure 6.15 for various degrees of freedom (black: df=8, red: df=6, green: df=4, blue: df=3, cyan: df=2). As one can see, the main trend for the duration estimates changes over time: The duration of the business cycles is lower in the '60s to the late '80s and then increasing from the early '90s until today. On average a cycle length is about five years, but in earlier times (1960-1990) a cycle length is more like

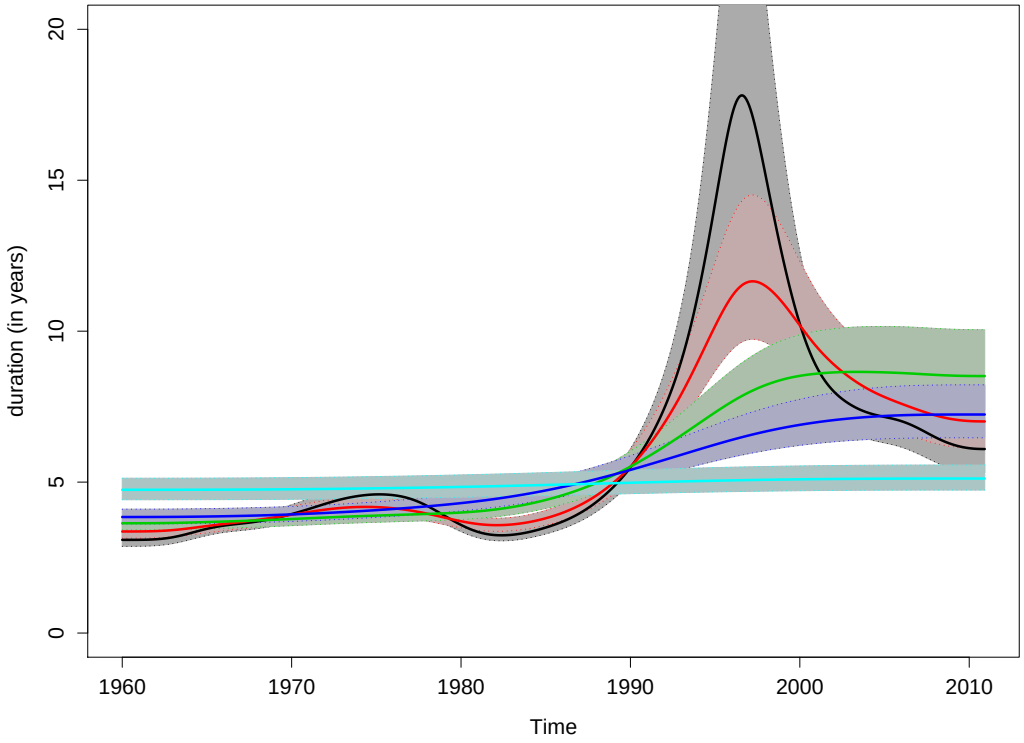

Figure 6.15: Estimated instantaneous duration of business cycles using an angle function with fewer degrees of freedom. Black: df=8; red: df=6; green: df=4; blue: df=3; cyan: df=2.

---

four years, while nowadays a full business cycle takes more like six to eight years. However, the "real" duration varies from cycle to cycle. Using the NBER classification, one can see that the estimated duration fits quite well into the pattern calculated by the NBER Business Cycle Dating Committee determination using the peak-to-peak (1960-1990: 4.78 years and 1990-2009: 8.81 years) and trough-to-trough (1960-1990: 4.92 years & 1990-2009: 8.86 years) measurements.

6.4 Estimating the Leading, Coincident, and Lagging Indicators Using Three-Dimensional Loops

Another way to estimate the leading, coincident, and lagging indicators defined by the Conference Board was shown by Teuber (2012a) who uses the generalized three-dimensional loop model presented in section 5.3, and in particular equation (5.87).

Instead of using an ordinary Hodrick-Prescott filter, a hybrid approach has been used. The generalized loops will be fitted given a long-term trend, and the deviation of the raw from the short-term trend, i.e. $\mathbf{y}_t - \mathbf{g}(t) = \boldsymbol{\gamma}(t) + \boldsymbol{\epsilon}_t$, will then be used to fit the long-term trend $\boldsymbol{\gamma}$. These two steps will be repeated until both functions, namely the short-term ($\mathbf{g}$) and long-term trend ($\boldsymbol{\gamma}$), converge. The main advantage is that a misspecified long-term trend in the first step will not have an irreparable effect.

In figure 6.16, the growth rates for the indicators are shown as dots and in the upper figure the estimated long-term fluctuations are shown as solid lines (leading indicators: green; coincident indicators: black, lagging indicators: red). In the lower figure, in addition to the estimated long-term trend, the "longer" three-dimensional fluctuation defined by the zenith angle function, i.e. $\hat{\rho}(t) \sin \hat{\upsilon}(t)$, has been added such that the resulted long-term trend looks much more moderate. However it should be noted that the long-term trend differs from the Hodrick-Prescott filter a lot.

Focussing on the three-dimensional loop structure, for which the estimation is shown in figure 6.17, one can see that the detrended growth rates have been fitted quite well, though the estimation might be improved. However, it is remarkable that a three-dimensional loop structure with few degrees

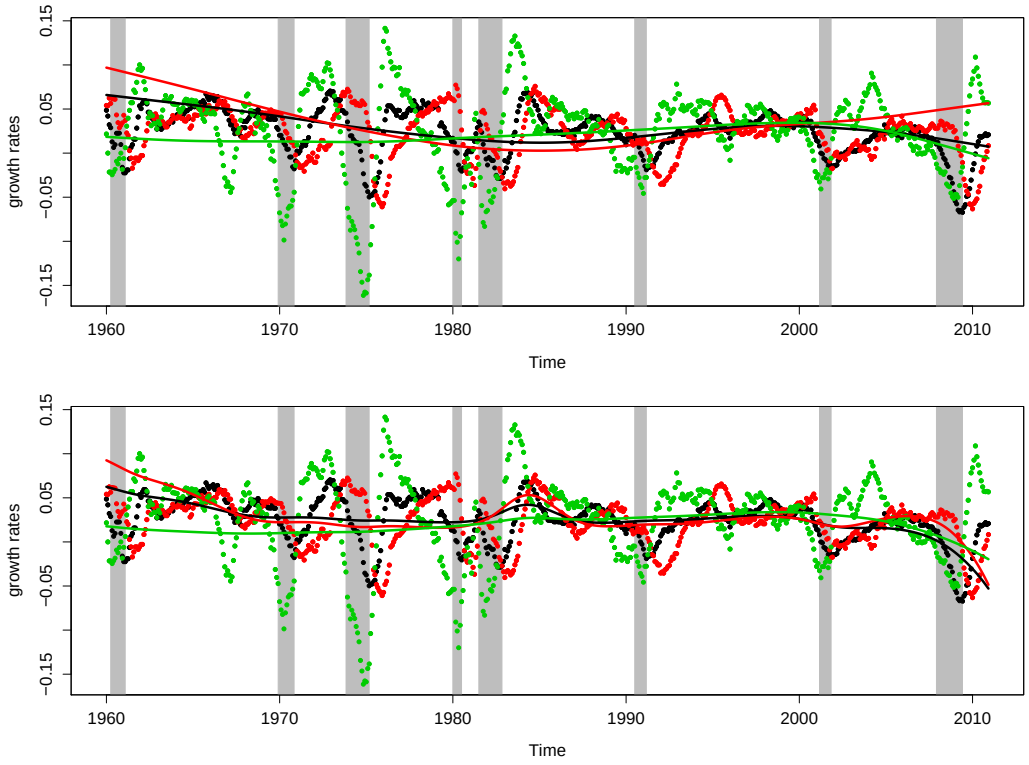

Figure 6.16: Raw growth rates (12-month logarithmic differences) as dots and estimated long-term trend as solid lines. Green: Leading indicators; black: Coincident indicators; red: Lagging indicators. Source: The Conference Board, January 1960 - December 2010.

---

of freedom describes the complex data structure at hand quite impressively. Furthermore, the solid black lines in figure 6.17 show the "longer-term" part of the three-dimensional loop structure without the short-term fluctuation part resulting solely from the zenith angle function, i.e. $\hat{\rho}(t) \sin v(t)$. It is worth noting that this long-term structure had been already observed by Kauermann, Teuber, and Flaschel (2012) for two more pairs of economic time series using their two-dimensional loop structure. And now, this structure has been

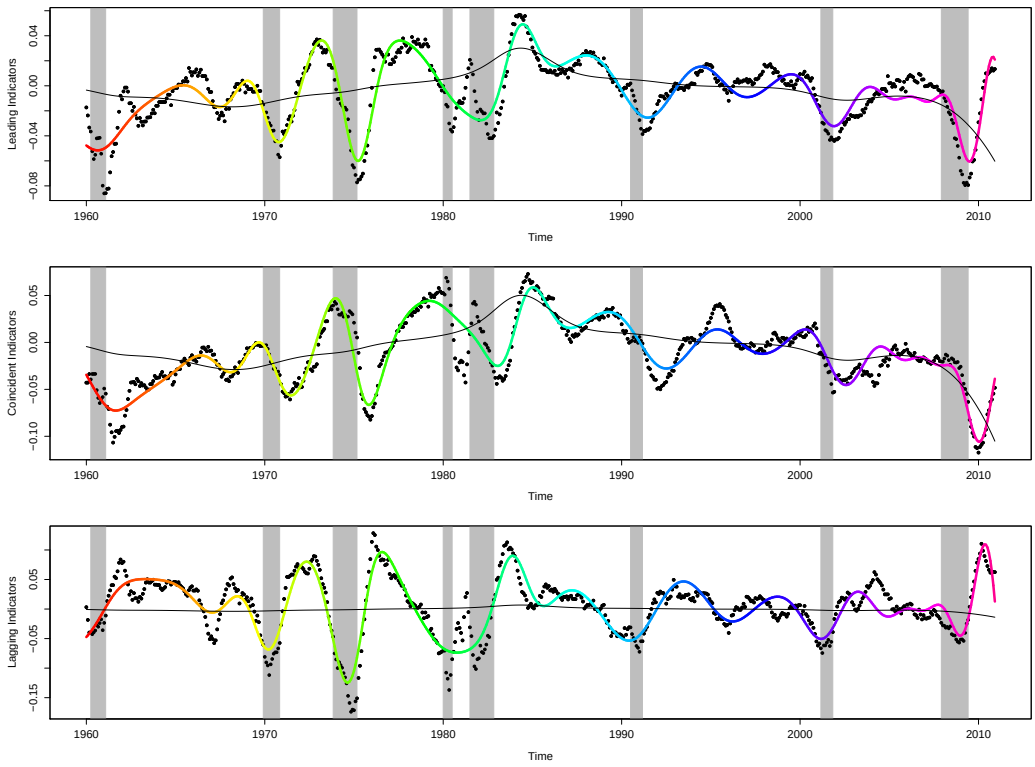

Figure 6.17: Detrended growth rates (dots) and estimations of the leading indicators (first row), coincident indicators (second row), and lagging indicators (third row). The colored line represents the estimated function using the three-dimensional loop structure whilst the black solid line represents the "longer-term" trend derived from the zenith angle function. Grey-shaded areas indicate time frames of NBER defined recession time points.

---

confirmed by the estimation of the three Conference Board indicators, too.

The long-term cycle length is roughly 40 years. However, given the esti-

mation of the angle functions, one can define the cycle duration by

$$\text{duration}_{\upsilon}(t) = \frac{2\pi}{|\hat{\upsilon}'(t)|} \tag{6.3}$$

and

$$\text{duration}_{\phi}(t) = \frac{2\pi}{|\hat{\phi}'(t)|}\,. \tag{6.4}$$

The upper plot in figure 6.18 shows the duration of the zenith cycle, estimated at any given time point. One can see that until the new century, the cycle duration was slightly below 40 years and then suddenly increased strongly, which might be an inaccuracy due to the boundary problem. The average duration over the last 50 years has been roughly 40 years and confirms the findings of Kauermann, Teuber, and Flaschel (2012).

The lower plot in figure 6.18 shows the estimated duration of the azimuth cycle, which is more or less decreasing over time. Over the last 50 years, the average cycle length was about five years, which is nearly the same duration which the National Bureau of Economic Research has calculated for the post-World War II business cycles.

Given the estimation of the three Conference Board indicators, the leading/lagging character has been clearly confirmed. The estimated shift in the (azimuth) angle of the leading indicator to the coincident indicator is $\hat{\phi}_{02} - \hat{\phi}_{01} = -1.0321$, and the estimated shift of the lagging indicator to the coincident indicator is $\hat{\phi}_{03} - \hat{\phi}_{01} = 1.0415$. This would translate into a 9.86 month lead by the leading indicators and a 9.95 month lag by the lagging indicators compared to the coincident indicators, if the cycle length is five years. However, the cycle length - as one can see in figure 6.18 - is varying over time and so is the lagging/leading shift measured in time.

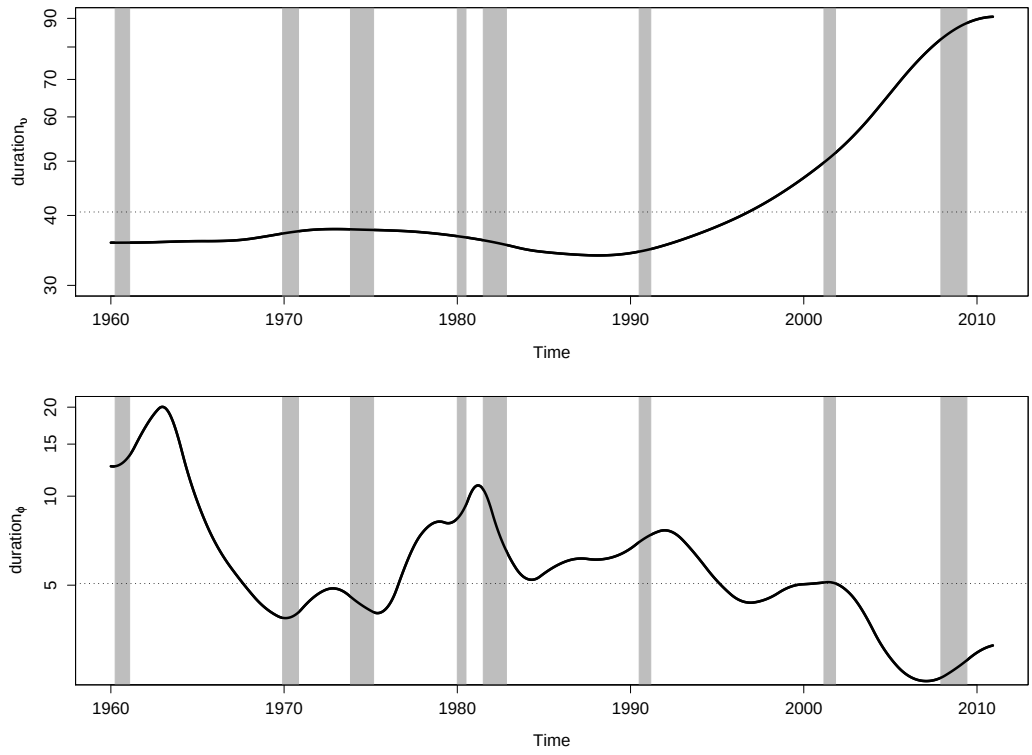

Figure 6.18: Instantaneous business cycle duration. Upper plot duration of the longer-term zenith cycle. Lower plot duration of the short-term azimuth cycle. Grey-shaded areas indicate time frames of NBER defined recession time points.

Furthermore, it is interesting that the shift in the radius has nearly the same ratio for the three time series as the ratio for the standard deviations of the residuals. The estimated standard deviations of the three time series are $\hat{\sigma}_1 = 0.01146$ for the coincident indicators, $\hat{\sigma}_2 = 0.02184$ for the leading indicators, and $\hat{\sigma}_1 = 0.01193$ for the lagging indicators. Given the shift in the radius of the leading vs. the coincident indicators, i.e. $\hat{\rho}_{02} - \hat{\rho}_{01} = 0.7692$, and of the lagging vs. coincident indicators, i.e. $\hat{\rho}_{03} - \hat{\rho}_{01} = 0.2206$, this translates into roughly a 2.158× and a 1.247× stronger magnitude of short-term fluc-

tuations of the leading and lagging vs. the coincident indicators, respectively.

However, from a statistical point of view a difference in the shift of the zenith angle could not be confirmed, suggesting that the leading and lagging character of the time series are more of a short-term character.

Unfortunately, the estimated azimuth angle function cannot be used to reverse-engineer the peak and trough dates provided by the National Bureau of Economic Research Business Cycle Dating Committee. The angles and their confidence regions are shown in figure 6.19, which shows the estimated angles in a "business cycle clock". Unfortunately, only a main trend but no meaningful generalization can be drawn from this analysis, so it is not (at least from a statistical point of view) possible to use the estimation to date or define the stages of the economy using this estimated three-dimensional loop structure.

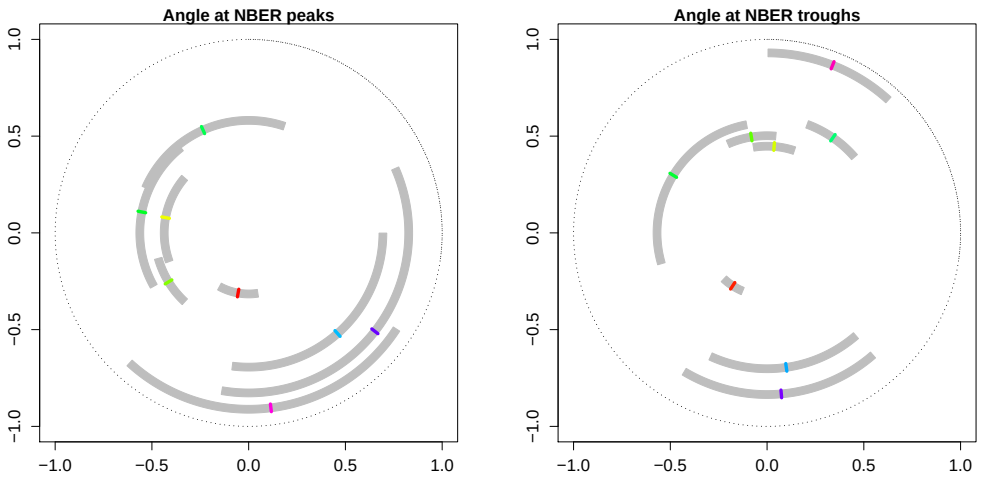

Figure 6.19: Business cycle clock set by the estimated azimuth angle function. Left plot shows the point estimation (colored lines) and the confidence regions (grey-shaded area) of the angle at NBER provided peak dates. Right plot shows the point estimation (colored lines) and the confidence regions (grey-shaded area) of the angle at NBER provided trough dates.

7 Summary

In this work several non-parametric estimation techniques have been introduced and developed to shed light on several economic problems. First, based on the brief history of business cycles in chapter 2, their definition, their evolvement over time and several empirical studies were shown. Furthermore, some problems were highlighted and discussed in that chapter. In chapter 3 several modern statistical models were discussed which build the groundwork for a sound modelling and testing of the economic hypothesis presented in chapter 4. It should be noted that without the statistical models in chapter 3, the modelling and estimation of the circular models in chapter 5 would have been more a theoretical exercise and the findings presented in chapter 6 would not have been possible.

It is remarkable that even though the statistical groundwork has been known for a long time, many economists avoid using these techniques to bring their data sets and models to life. The benefits of the models, presented in chapter 3, are enormous, especially considering to their easy accessibility in most statistical packages. The effort of an OLS analysis compared to a GLM or a GAM analysis is negligible. Furthermore, the even more complex and well-known Kalman filter and Hodrick-Prescott filter are nested models, as already shown in chapter 3, so these tools should be preferred.

The Generalized Linear Models shown in chapter 3.1 scale the usability of linear models not only for normal data but also to the whole family of exponential distributions, among them for instance the binomial, multinomial, Poisson, gamma, and exponential distributions. This means that even more data sets can be handled with the same well-sounded model.

Penalized splines, shown in chapter 3.2, and the Generalized Additive Models, shown in chapter 3.5, are handy tools to test for non-linearities in data sets. Most real world relationships are not linear or static over time, so a suitable framework to estimate these relationships is needed. In chapter 3.6 an example of time-varying coefficients is provided, allowing one to easily test if the relationship of variables is dynamically changing over time.

Ernst and Teuber (2006), as shown in chapter 4.1, used a penalized spline model to model time-varying coefficients to test for and estimate the dynamic debt policy of OECD countries. Another example which had been modelled based on the introduced statistical models is shown by Flaschel, Tavani, Taylor, and Teuber (2008) and in chapter 4.2. Here, the non-linearity of the models was used not only to test for non-linear relationships, but also to get as a by-product the first derivative of the estimated function. The confidence regions of the first derivative were analyzed which is bearly possible using most other techniques. Furthermore, Ernst and Teuber (2008) and Ernst and Teuber (2012), as shown in chapter 4.3, used a penalized spline approach with random effects to estimate the non-linear age-productivity function. Without this approach it is nearly impossible to predict the random effects which were essential for the estimation of transitory probabilities of the idiosyncratic productivity shocks of workers. Here, the multinomial variables were estimated using the Generalized Linear Model introduced in chapter 3.1.

The circular model, shown in chapter 5, which has been extensively discussed by Kauermann, Teuber, and Flaschel (2012), Flaschel, Kauermann, and Teuber (2005), Flaschel, Tavani, Taylor, and Teuber (2008), Proaño, Flaschel, Diallo, and Teuber (2008), Flaschel, Groh, Kauermann, and Teuber (2008), Teuber (2012b), and Teuber (2012a) would not have been possible to model

adequately without the statistical tools shown in chapter 3. As already mentioned in chapter 2, no business cycle is alike and they differ in timing and magnitude making a static model useless. Furthermore, as shown in chapter 5, the center (long-term trend) is essential for the estimation of the short-term fluctuations, such that at least a test or, even better, a modelling of a moving center should be preferred. The model developed by Kauermann, Teuber, and Flaschel (2012) lacks the possibility to model more than two time series, while Teuber (2012b) improved the two-dimensional model to fit any arbitrary number of time series. Furthermore, the way Teuber (2012b) interpreted the modelling is much more elegant in terms of a reduction of complexity. Teuber (2012a) extended the model and used spherical coordinates instead of polar coordinates, which was an elegant way to account for long-term cycles confirmed in the previous works.

This work concludes with the estimation and discussion of the long-term and short-term fluctuations of different time series. In chapter 6, the dynamic of the unemployment rate/price inflation trajectories and wage share/employment rate trajectories was analyzed. Here, not only were the short-term cycles verified, but also long-term cycles with a length of 40-50 years were observed. The transformation of the Cartesian coordinates of the observations to polar coordinates was crucial in modelling the short-term fluctuations. The magnitude, measured by the radius, the speed/persistence of business cycles, measured by the angle, and the leading/lagging character, modelled by a separate coefficient (shift of angle), can easily be examined within this framework and had shed light on the discussion of business cycles, as one can see in chapters 6.3 and 6.4, and in the work by Teuber (2012b) and Teuber (2012a).

Furthermore, another handy framework was established to separate the long-term trend from the short-term fluctuations, such that the clumsy Hodrick-

Prescott filter with a predefined smoothing parameter was avoided due to the fact that this model is already nested in the more elegant penalized spline model. In the light of the empirical studies presented in chapter 6, it seems that the Hodrick-Prescott filter (with the parameter $\lambda = 1600$) uses too many degrees of freedom, and a less complex long-term component should be preferred. Although the estimations in chapters 6.2, 6.3, and 6.4 use slightly different approaches to determine the long-term and short-term trend and handle different data sets, in all models a long-term trend with a cyclicality of roughly 40 to 50 years was observed. Furthermore, all estimations show that the short-term cycles do vary from one cycle to another in terms of magnitude, speed, duration, and location. However, an average cycle length of five years was confirmed with the models in chapter 6. The dating of business cycle peaks and troughs, however, differs slightly from the dates provided by the NBER Business Cycle Dating Committee and, furthermore, more business cycles were observed. It should be noted that the provided methodology seems to handle the dating procedure much better due to the fact that the model in chapter 5 provides confidence regions which allow not only a point estimate of a business cycle date to be given, but also a confidence region.

References

Aghion, P. and I. Marinescu (2006). Cyclical budgetary policy and economic growth: What do we learn from oecd panel data? Technical report, WCFIA Working Paper.

Akaike, H. (1974). A new look at the statistical model identification. In AC (Ed.), *IEEE Transactions of Automatic Control*, Volume 19 of *716–723*.

Arrow, K. J. (1963). Uncertainty and the welfare economics of medical care. *The American Economic Review 53*, 941–973.

Arrow, K. J. and G. Debreu (1954). Existence of an equilibrium for a competitive economy. *Econometrica 22*, 265–290.

Atkinson, A. (1969). The timescale of economic models: How long is the long run? *Review of Economic Studies 36*, 137–152.

Auerbach, A. and L. Kotlikoff (1987). *Dynamic Fiscal Policy*. MIT Press.

Backus, D., P. Kehoe, and F. Kydland (1992). International real business cycles. *Journal of Political Economy 100*(4), 745–775.

Barro, R. (1979). On the determination of public debt. *Journal of Political Economy 87*, 940–971.

Baxter, M. and R. King (1999). Measuring business cycles: Approximate band-pass filters for economic time series. *The Review of Economics and Statistics 81*(4), 575–593.

Bloemen, H. (2007). The Impact of Wealth on Job Exit Rates of Elderly Workers. Technical report, Central Bank of Chile Working Papers, TI 2007-002/3.

Breslow, N. E. and D. G. Clayton (1993). Approximate inference in generalized linear mixed model. *Journal of the American Statistical Association. 88*, 9–25.

Brockwell, P. J. and R. A. Davis (1987). *Time Series: Theory and Methods*. Springer-Verlag, Berlin, New York.

Börsch-Supan, A., A. Ludwig, and J. Winter (2006, November). Ageing, pension reform and capital flows: A multi-country simulation model. *Economica 73*, 625–658.

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

Calderon, C. and K. Schmidt-Hebbel (2003). Macroeconomic Policies and Performance in Latin America. Technical report, Central Bank of Chile Working Papers, No. 217.

Chiarella, C., P. Flaschel, and R. Franke (2005). *Foundations for a Disequilibrium Theory of the Business Cycle. Qualitative Analysis and Quantitative Assessment.* Cambridge: Cambridge University Press.

Conesa, J. C. and D. Krueger (1999). Social Security Reform with Heterogeneous Agents. *Review of Economic Dynamics 2*, 757–795.

Craven, P. and G. Wahba (1979). Smoothing noisy data with spline functions. *Numer. Math. 31*, 377–403.

de Boor, C. (1978). *A Practical Guide to Splines*. Berlin: Springer.

Debreu, G. (1959). *Theory of Value: An Axiomatic Analysis of Economic Equilibrium*. Princeton, NJ: Cowles Foundation Monographs Series.

Eilers, P. and B. Marx (1996). Flexible smoothing with b-splines and penalties. *Statistical Science 11*(2), 89–121.

Ernst, E. and T. Teuber (2006). Macroeconomic policy regimes: A new database on the evolution of macroeconomic policies in the oecd. working paper.

Ernst, E. and T. Teuber (2008). The dutch tax-benefit and life-cycle employment. OECD ECO Working Paper.

Ernst, E. and T. Teuber (2012). The Dutch Tax-Benefit System and Life-Cycle Employment Outcomes and Reform Options. *AUCO Czech Economic Review*. submitted.

Eubank, R. L. (1989). *Spline smoothing and nonparametric regression*. Dekker: New York.

Fan, J. and I. Gijbels (1996). *Local Polynomial Modelling and its Applications*. London: Chapman & Hall.

Fitzenberger, B., R. Hujer, T. E. MaCurdy, and R. Schnabel (2001). Ageing, pension reform and capital flows: A multi-country simulation model. *Empirical Ecomomics 26*, 41–86.

Flaschel, P., G. Groh, G. Kauermann, and T. Teuber (2008). The classical growth cycle after fifteen years of new observations. In *Mathematical Economics and the Dynamics of Capitalism.* London: Routledge: P. Flaschel and M. Landesmann.

Flaschel, P., G. Kauermann, and T. Teuber (2005). Long cycles in employment, inflation and real unit wage costs. Qualitative analysis and quantitative assessment. *American Journal of Applied Sciences, Special Issue on: Economic Sciences, Dynamics, Expectations and Learning*, 69–77.

Flaschel, P. and H.-M. Krolzig (2006). Wage-price philips curves and macroeconomic stability: Basic structural form, estimation and analysis. In C. Chiarella, P. Flaschel, R. Franke, and W. Semmler (Eds.), *Quantitative and Empirical Analysis of Nonlinear Dynamic Macromodels*, Contributions to Economic Analysis (Series Editors: B. Baltagi, E. Sadka and D. Wildasin), Elsevier, Amsterdam.

Flaschel, P., D. Tavani, L. Taylor, and T. Teuber (2008). A structuralist model of the wage-price spiral with non-linear demand pressure

terms. working paper, Schwarz Center for Economic Policy Analysis. New School University, New York.

Floden, M. and J. Lindé (2001). Idiosyncratic risk in the united states and sweden: Is there a role for government insurance? *Review of Economic Dynamics 4*, 406–437.

French, E. (2003). The effects of health, wealth, and wages on labor supply and retirement behavior. Technical report, Federal Reserve Bank of Chicago: Working Papers Series Research Department, WP 2000-02.

Friedman, M. (1957). *A Theory of the Consumption Theory*. Princeton, NJ: Princeton University Press.

Friedman, M. (1968). The role of monetary policy. *American Economic Review 58*, 1–17.

Friedman, M. (2008). The optimum quantity of money. In *The Optimum Quantity of Money and Other Essays*. Chicago: Aldine Publishing Company: Friedman, M.

Goodwin, R. (1967). A growth cycle. In C. Feinstein (Ed.), *Socialism, Capitalism and Economic Growth*, pp. 54–58. Cambridge: Cambridge University Press.

Hamilton, J. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica 57*(2), 357–384.

Hamilton, J. (2005). What's real about the business cycle? *Federal Reserve Bank of St. Louis, Review 87*(4), 435–452.

Härdle, W. and J. S. Marron (1991). Bootstrap simultaneous error bars for nonparametric regression. *The Annals of Statistics 19*(2), 778–796.

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

Hastie, T. and R. Tibshirani (1993). Varying–coefficient models. *Journal of the Royal Statistical Society, Series B 55*, 757–796.

Hodrick, R. and E. Prescott (1997). Postwar u.s. business cycles: An empirical investigation. *Journal of Money, Credit, and Banking 29*, 1–16.

Kauermann, G., T. Teuber, and P. Flaschel (2012). Exploring US Business Cycles with Bivariate Loops using Penalized Spline Regression. *Computational Economics 39(4)*, 409–427.

Keynes, J. M. (1936). *The General Theory of Employment, Interest and Money*. New York: Macmillan.

Kondratieff, N. (1984). *Long Wave Cycle*. Contains the translated Russian work first published in 1925: Richardson & Snyder.

Kotlikoff, L. and D. Rapson (2006). Does it pay, at the margin, to work and save? Measuring effective marginal taxes on Americans labor supply and saving. Technical report, NBER Working Paper, No. 12533.

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

Kuznets, S. (1930). *Secular Movements in Production and Prices*. Cambridge, Massachusetts: Houghton Mifflin.

Kuznets, S. (1934, November). Gross capital formation, 1919 - 1933. In S. Kuznets (Ed.), *Gross Capital Formation, 1919 - 1933*, NBER, pp. 1–20.

Kydland, F. and E. Prescott (1982). Time to build and aggregate fluctuations. *Econometrica 50*(6), 1345–1370.

Kydland, F. E. and E. C. Prescott (1990). Business cycles: Real facts and a monetary myth. *Federal Reserve Bank of Minneapolis Quarterly Review Spring*, 3–18.

Laird, N. and J. Ware (1982). Random effects models for longitudinal data. *Biometrika 38*, 963–974.

Lane, P. R. (2003). The Cyclical Behaviour of Fiscal Policy: Evidence from the OECD. *Journal of Public Economics 87*, 2661–2675.

Lindstrom, M. and D. Bates (1990). Nonlinear mixed-effects models for repeated measures data. *Biometrics 46*, 673–687.

Long, J. and C. Plosser (1983). Real business cycles. *Journal of Political Economy 91*(1), 39–69.

Lucas, R. E. J. (1977). Understanding business cycles. In K. Brunner and A. H. Meltzer (Eds.), *Stabilization of the domestic and international economy*, Volume 5 of *Carnegie-Rochester Conference Series on Public Policy*, pp. 7–29. Amsterdam: North-Holland.

Lucas, R. E. J. (1980). Methods and problems in business cycle theory. *Journal of Money, Credit, and Banking 12*(4), 696–715.

Ludwig, A. and E. Vogel (2010). Mortality, fertility, education and capital accumulation in a simple OLG economy. *Journal of Population Economics 23*, 703–735.

Mao, W. and L. Zhao (2003). Free-knot polynomial splines with confidence intervals. *Journal of the Royal Statistical Society, Series B 65*, 901–919.

McCullagh, P. and J. A. Nelder (1989). *Generalized Linear Models* (second ed.). New York: Chapman and Hall.

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

Määttänen, N. and P. Poutvaara (2007). Should old-age benefits be earnings-tested? Technical report, Institute for the Study of Labor, Discussion Paper No. 2616.

Nadaraya, E. (1964). On estimating regression. *Theory of Probability and Its Applications 10*, 186–190.

OECD (2006). Taxing wages. Technical report, OECD, Paris.

O'Sullivan, F. (1988). Nonparametric estimation of relative risk using splines and cross-validation. *SIAM J. Sci. Statist. Comput. 9*, 531–542.

Parker, R. and J. Rice (1985). Discussion of "Some aspects of the spline smoothing approach to nonparametric curve fitting" by B.W. Silverman. *Journal of the Royal Statistical Society, Series B 47*, 40–42.

Persson, T. and G. Tabellini (2003). *The Economic Effects of Constitutions: What do the data say?* Cambridge, MA: MIT Press.

Peterman, W. B. (2011). The effect of endogenous human capital accumulation on optimal taxation. unpublished.

Proaño, C., P. Flaschel, M. B. Diallo, and T. Teuber (2008). Distributive cycles, business fluctuations and the wage-led/profit-led debate. working paper.

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

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

Ruppert, D., M. Wand, and R. Carroll (2009). Semiparametric Regression During 2003-2007. *Electronic Journal of Statistics 3*, 1193–1256.

Searle, S., G. Casella, and C. McCulloch (1992). *Variance Components*. Wiley.

Silverman, B. W. (1985). Some aspects of the spline smoothing approach to non-parametric regression curve fitting. *Journal of the Royal Statistical Society, Series B 47*, 1–21.

Smith, A. (1776). *An Inquiry into the nature and causes of the wealth of nations, Reprinted 1966*. New York: Kelley.

Solow, R. (1990). Goodwin's growth cycle: Reminiscence and rumination. In K. Velupillai (Ed.), *Nonlinear and multisectoral macrodynamics. Essays in Honour of Richard Goodwin*, pp. 31–41. Macmillan.

Stock, J. H. (1987). Measuring business cycle time. *Journal of Political Economy 95*(6), 1240–1261.

Stock, J. H. and M. W. Watson (1999). Business cycle fluctuations in us macroeconomic time series. In J. B. Taylor and M. Woodford (Eds.), *Handbook of Macroeconomics*, Volume 1A, pp. 3–64. Elsevier.

Teuber, T. (2012a). Interpreting Business Cycles as Generalized Three-Dimensional Loops Using Penalized Splines Regression Techniques. *Singapore Economic Review*. to be awaited for resubmission.

Teuber, T. (2012b). Interpreting Business Cycles as Generalized Two-Dimensional Loops Using Penalized Splines Regression Techniques. *Applied Economics Quarterly*. submitted.

Wahba, G. (2003). *Spline Models for Observational Data*. Siam.

Wand, M. (2003). Smoothing and mixed models. *Computational Statistics 18*, 223–249.

Watson, G. (1964). Smooth regression analysis. *Sankhy, Series A 26*, 359–372.