Flexible Distributed Lag Models and their Application to Geophysical Data

Viola Obermeier

München 2014

Flexible Distributed Lag Models and their Application to Geophysical Data

Viola Obermeier

Dissertation
an der Fakultät für Mathematik, Informatik und Statistik
der Ludwig–Maximilians–Universität
München

vorgelegt von
Viola Obermeier
aus Freising

München, den 25.02.2014

Danksagung

Abstract

Regression models with lagged covariate effects are often used in biostatistical and geo-physical data analysis.

In the difficult and all-important subject of earthquake research, strong long-lasting rainfall is assumed to be one of many complex trigger factors that lead to earthquakes. Geophysicists interpret the rain effect with an increase of pore pressure due to the infiltration of rain water over a long time period. Therefore, a sensible statistical regression model examining the influence of rain on the number of earthquakes on day t has to contain rain information of day t and of preceding days $t-1$ to $t-L$.

In the first part of this thesis, the specific shape of lagged rain influence on the number of earthquakes is modeled. A novel penalty structure for interpretable and flexible estimates of lag coefficients based on spline representations is presented. The penalty structure enables smoothness of the resulting lag course and a shrinkage towards zero of the last lag coefficient via a ridge penalty. This additional ridge penalty offers an approach to another problem neglected in previous work. With the help of the additional ridge penalty, a suboptimal choice of the lag length L is no longer critical. We propose the use of longer lags, as our simulations indicate that superfluous coefficients are correctly estimated close to zero.

We provide a user-friendly implementation of our flexible distributed lag (FDL) approach, that can be used directly in the established R package *mgcv* for estimation of generalized additive models. This allows our approach to be immediately included in complex additive models for generalized responses even in hierarchical or longitudinal data settings, making use of established stable and well-tested algorithms. We demonstrate the performance and utility of the proposed flexible distributed lag model in a case study on (micro-) earthquake data from Mount Hochstaufen, Bavaria with focus on the specific shape of the lagged rain influence on the occurrence of earthquakes in different depths. The complex meteorological and geophysical data set was collected and provided by the Geophysical Observatory of the Ludwig-Maximilians University Munich. The benefit of flexible distributed lag modeling is shown in a detailed simulation study.

In the second part of the thesis, the penalization concept is extended to lagged non-linear covariate influence. Here, we extend an approach of Gasparrini et al. (2010), that was up to now unpenalized. Detailed simulation studies illustrate again the benefits of the penalty structure. The flexible distributed lag nonlinear model is applied to data of the volcano Merapi in Indonesia, collected and provided by the Geophysical Observatory

in Fürstenfeldbruck. In this data set, the specific shape of lagged rain influence on the occurrence of block and ash flows is examined.

Zusammenfassung

Regressionsmodelle mit verzögerten Kovariableneffekten werden in biostatistischen und geophysikalischen Datenanalysen häufig verwendet.

Starker, lang anhaltender Regenfall wird im schwierigen und äußerst wichtigen Gebiet der Erdbebenforschung als einer von vielen möglichen Auslösern von Erdbeben betrachtet. Geophysiker interpretieren den Regeneffekt mit einer Erhöhung des Porendruckes durch das Eindringen von Regenwasser über einen längeren Zeitraum. Ein vernünftiges statistisches Regressionsmodel, das den Einfluss von Regen auf die Anzahl von Erdbeben am Tag t untersucht, sollte daher neben der Regenmenge an Tag t auch diejenigen Regenmengen vorhergehender Tage $t-1$ bis $t-L$ beinhalten.

Im ersten Teil dieser Arbeit wird die genaue Gestalt von verzögertem Regeneinfluss auf die Anzahl von Erdbeben modelliert. Eine neue Penalisierungsstruktur für eine interpretierbare und flexible spline-basierte Schätzung von verzögerten Effekten wird vorgestellt. Diese Penalisierungsstruktur ermöglicht die Glattheit des resultierenden Lag-Verlaufs sowie eine Schrumpfung des letzten Lag-Koeffizienten gegen Null mithilfe einer Ridge-Penalisierung. Diese zusätzliche Ridge-Penalisierung bietet einen Lösungsansatz für ein weiteres Problem. Mit Hilfe der Ridge-Penalisierung ist eine suboptimale Wahl der Lag-Länge nicht mehr kritisch. Wir schlagen die Verwendung von längeren Lags vor, da unsere Simulationen zeigen, dass überflüssige Koeffizienten korrekterweise nahe Null geschätzt werden.

Wir stellen eine benutzerfreundliche Implementierung unseres "flexible distributed lag" Ansatzes vor, die direkt im etablierten R Paket *mgcv* zur Schätzung generalisierter additiver Modelle verwendet werden kann. Dadurch ist unser Ansatz sogleich in komplexe additive Modelle für generalisierte Zielgrößen hierarchischer oder longitudinaler Datenstrukturen unter Verwendung stabiler und getesteter Algorithmen eingebunden. Wir zeigen die Leistungsfähigkeit und Nützlichkeit des vorgeschlagenen "flexible distributed lag" Models anhand von (Micro-) Erdbebendaten am Hochstaufen, Bayern. Der Fokus hierbei liegt auf der genauen Gestalt des verzögerten Regeneinflusses auf das Auftreten von Erdbeben in verschiedenen Tiefenkategorien. Der komplexe meteorologische und geophysikalische Datensatz wurde vom Geophysikalischen Observatorium der Ludwig-Maximilians-Universität München erhoben. Der Nutzen der "flexiblen distributed lag" Modellierung wird anhand einer umfangreichen Simulationsstudie gezeigt.

Im zweiten Teil der Arbeit wird das Penalisierungskonzept auf verzögerten, nichtlinearen Einfluss von Kovariablen erweitert. Hierbei bauen wir auf einem Ansatz von Gasparrini et al. (2010) auf, der bislang unpenalisiert war. Detaillierte Simulationsstudien

zeigen auch hier wieder den Nutzen der Penalisierungsstruktur. Das "flexible distributed lag" Modell für nichtlineare Kovariableneffekte wird unter anderem auf Daten des Vulkanes Merapi in Indonesien angewandt. Diese Daten wurden vom Geophysikalischen Observatorium in Fürstenfeldbruck erhoben und zur Verfügung gestellt. Untersuchungsziel ist die spezifische Gestalt des verzögerten Regeneinflusses auf das Auftreten von Aktivitäten des Vulkans.

Contents

Chapter 1

Introduction

1.1 Motivation

1.1.1 Regression Models with Lagged Covariate Effects

In many settings where observations occur repeatedly over time, as in geophysical or econometric data, the association between a covariate $\mathbf{x}$ and a response variable $\mathbf{y}$ cannot reasonably be assumed to be only direct and instantaneous, but instead is often more likely to be cumulative over a certain time period, extending into the past of the respective response. In the difficult subject of earthquake research, strong long-lasting rainfall is assumed to be one of many complex and so far unknown trigger factors that lead to earthquakes (Roth et al., 1992). Thus, a sensible statistical regression model examining earthquake occurrence at time point t has to contain rain information of preceding time points, $t-1, t-2, \dots, t-L$. Overall, the coefficient β_l for the association between a covariate x_{t-l} (e.g., rain) and a response variable y_t (e.g., the number of quakes) is called the lag effect of $\mathbf{x}$ on $\mathbf{y}$ with time lag l. In this manuscript, we equivalently use the terms β_l and $\beta(l)$ describing the lag effect of $\mathbf{x}$ on $\mathbf{y}$ with time lag l.

1.1.2 Demands on Lag Curves

A reasonable modeling of lag coefficients poses the following challenges: One expects the effects of adjacent coefficients to be similar; so $\beta(l)$ is to be a smooth function. The resulting lag curve shall not be highly volatile and wiggly but has to take a smooth course. The agenda in that demand is, that the lag effect of $\mathbf{x}$ on $\mathbf{y}$ with time lag l is not expected to differ strongly from the lag effect with time lag $l+1$. On the other hand, a certain flexibility in the estimation is requested. Global minima and maxima or potential bimodal courses shall reliably be detected.
Second, the effects of the lagged covariates that are furthest in the past are assumed to be weak, so β_L should be close to zero. Of course, for this assumption to be reasonable, the maximal number of lag units L needs to be sufficiently large, so that $\beta(l)$ can capture all the effects of the covariate $\mathbf{x}$ on the response variable $\mathbf{y}$.

1.1.3   Example: Hochstaufen Data

The data set was collected and kindly provided by Dr. Joachim Wassermann from the Geophysical Observatory in Fürstenfeldbruck within the scope of a cooperation of the Department of Statistics and the Department of Earth- and Environmental Sciences of the Ludwig-Maximilians University Munich.

That above-average rates of rainfall may trigger earthquakes had been speculated since the start of detailed seismic observations in alpine areas (Roth et al., 1992). Several independent studies carried out at Mt. Hochstaufen (SE-Bavaria, Germany), Central Swiss and Mt. Hood (Oregon, USA) showed that the triggering or forcing of earthquake swarm occurrence is real and might be described by a simple 1D diffusion of rain into the subsurface structure (Roth et al., 1992; Saar and Manga, 2003; Kraft et al., 2006; Husen et al., 2007). Our analysis is based on data of the Mount Hochstaufen area (see Figure 1.1), a 1770 m high mountain massif at the south-east Bavarian, northern alpine front (see Kraft et al., 2006, for detailed geological description). Hainzl et al. (2013) examine Hochstaufen data and show that earthquake rates can be modeled via fluid diffusion and pore pressure diffusion, if the crust is near a critical state. This area has been known for its seismic activity since 1390 and has therefore been monitored starting in the 1970ies by the Geophysical Observatory in Fürstenfeldbruck with a changing number of seismometers. Since 1980, the swarm character of these earthquakes, which occur mainly in the summer months, has been investigated using digital mobile seismic equipment and one permanent analog ink recorder in Bad Reichenhall. From 2001 on, the area has been monitored by six digital short period seismic stations with permanent and continuous data transmission to the data center in Fürstenfeldbruck (Figure 1.2). During summer, this permanent network is extended by additional mobile stations (two - six) in order to decrease the magnitude of completeness (Mc) and increase the location accuracy. Additionally, meteorological stations (rain gauge, air pressure, air temperature, humidity) have been co-located at three sites since mid-2004 (RJOB = Reichenhall Jochberg, RNON = Reichenhall Nonn, RMOA = Reichenhall Moar Alp). After the construction of the permanent surveillance network, seven distinct earthquake swarms occurred at Mt. Hochstaufen (March and August 2002; May, July, August 2005; September 2007, April 2008). The maximum magnitude was reached in 2007, when an earthquake Ml=3.4 happened and was felt by humans in a wide area. The majority of the earthquakes, however, are small in magnitude (<1.5 Ml) and thus not felt. The largest earthquake swarm by number of events was recorded in August 2002 after a period of extraordinary heavy rainfall with more than 600 events/month. None of these earthquakes were strong enough to be felt.

We classify the depth of the earthquakes into depth categories. The interval sizes reflect the uncertainty of the hypocenter determination (see Svejdar et al., 2011, for the geophysical details of the hypocenter determination.). We assume, that the infiltration of rain water in the lower rock beds causes more lagged rain effects in the lower depth categories and more direct rain effects in the upper depth categories. We analyze this assumption by means of statistical regression methods in the following sections.

Figure 1.1: View of Mount Hochstaufen in Anger (Berchtesgadener Land, South-East Bavaria).
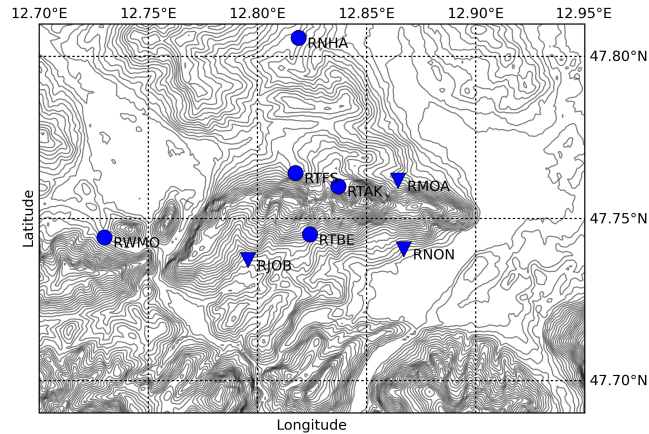
Figure 1.2: Seismic and meteorological network installed at Mt. Hochstaufen. The stations shown were recording during most of the time period analysed in this thesis. The circles represent seismic (3C 1 Hz seismic sensors) stations, the triangles mark seismic and additionally-installed meteorological stations (rain gauge, air temperature, ambient pressure, humidity), figure by Dr. Wassermann.

1.1.4 Example: Merapi Data

The second data set, that is worked on in this thesis, was also collected and kindly provided by Dr. Joachim Wassermann of the Department of Earth- and Environmental Sciences of the Ludwig-Maximilians University Munich.

Merapi is a strato-volcano on Java island in Indonesia. Figure 1.3 shows the former peak of Merapi, that was called "Puncak Garuda" which is the Indonesian term for "top of the eagle". The volcano is known to be one of the most dangerous volcanos in the world because of the frequently appearing (gravitational) dome collapses that lead to possibly devastating pyroclastic clouds sweeping down the flanks of the volcano. The volcano Merapi lastly erupted in the year 2010 with several thousands of evacuees and several hundreds of fatalities. Due to its dangerousness, the volcano is seismically monitored since 1924. Figure 1.4 shows rock falls in a South-Westward direction.

Geophysicists assume not only internal forcing, that brings the volcano close to the tipping point, but also external influences such as rain to be the cause of volcano eruptions.

Strong long-lasting rainfall is assumed to be one of many complex trigger factors that lead to seismic activity (Barclay et al., 2006). We are particularly interested in the specific shape of the lagged rain influence. Is the rain effect on the volcano activity decreasing with increasing lag? Or is the lagged rain effect rather unimodal or bimodal? The data set contains continuous recordings of daily seismic and meteorological data on 2474 days beginning from January 1st, 1995 to October 9th, 2001. The seismic data measuring the activity of the volcano basically consists of the binary response variable *Block and*

Figure 1.3: Former peak of Volcano Merapi with the so-called "Puncak Garuda" (Indonesian for "top of the eagle" (nonexistent since 2006), figure by Dr. Wassermann.

Figure 1.4: Rock falls (*Guguran*) at Volcano Merapi in a south-westward direction, figure by Dr. Wassermann.
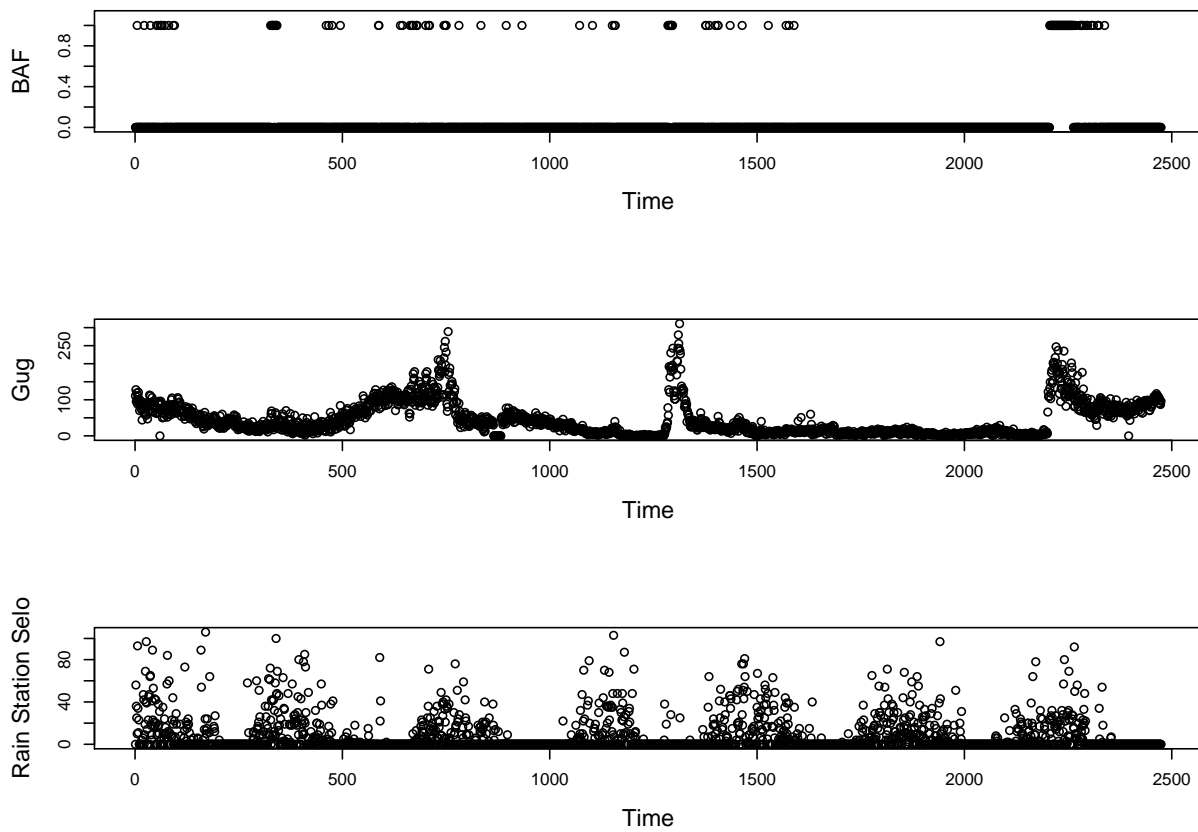
Figure 1.5: Block and ash flows (BAF, top), landslide of rock boulders (Gug, center) and rain on station Selo (bottom) on 2474 days.

Figure 1.6: Observation ward (station *Babadan*) at volcano Merapi, figure by Dr. Wasser-
             mann.

Ash Flows (*BAF*) and is coded with a one if there was a block and ash flow on the
respective day and zero if not. Furthermore, we have continuous recordings of the landslide
of rock boulders (local name: *Guguran*), given as a count variable. The meteorological data
contains the amount of rain in l/m^2 measured on different stations on the volcano. The
station names are: *Babadan*, *Jrakah*, *Kaliurang*, *Ngepos* and *Selo*. The stations *Babadan*
(see Figure 1.6), *Kaliurang* and *Selo* are closer to the events and therefore more suitable
for the analysis. Figure (1.5) shows the time dependent course of the block and ash flows,
the number of the landslides of the rock boulders and the amount of rain fallen on the 2474
days from January 1st, 1995 on.

1.1.5 Example: Geothermal Data

Geothermal energy is the thermal energy stored in the earth's crust. In geothermal energy
production, hot water is extracted and used to generate heat for use in district heating or
to produce electrical power. The cooled fluid subsequently has to be re-injected into the
reservoir. In a depth of about 3 to 4 kilometers below ground level, about 130° degree
celsius warm fluid is extracted. The cooled fluid (about 60° - 90° celsius) is then re-
injected in the earth's crust. Critics worry, that the re-injection of cooled water can cause
earthquakes. The "International Journal of Geothermal Research and its Applications"
attends to the research and development of geothermal energy. Reviews on the topic are
for example given by Majer et al. (2007) and Evans et al. (2012). We analyze confidential
data of operations at a geothermal power plant that can be used in this thesis under

the requirement, that the data source is not disclosed. Figure 1.7 shows on the top the response variable, the *summed moment magnitude* M_w, describing the seismic activity. Please note, that the *summed moment magnitude* is not directly summed (as M_w is a logarithmic measure of energy). We calculate for each seismic event the seismic moment out of the moment magnitude M_w, sum up the seismic moments and then calculate the M_w out of the sum.

Three parameters concerning the re-injection are available. The mass flow rate, the temperature of injected water, and the wellhead pressure of the injection well. In arrangement with our geophysical co-operation partners, we combine re-injection temperature and the mass flow rate to a synthetic covariate *energy* $= (130 - temperature) * rate$ as the measure of extracted thermal energy. Thus, we use two injection parameter sets in our analysis: The thermal energy removal rate described above and the wellhead pressure.

1.2   Literature Overview

In earlier work (Svejdar et al., 2011), we use two-dimensional Almon polynomials (Almon, 1965) to model the lagged rain effect on the number of earthquakes in four depth categories at Mt. Hochstaufen, SE-Bavaria between the years 2002 and 2007.
Almon Polynomials constrain the lag coefficients to form a polynomial of lower degree. That technique is often used and easily implemented by constructing synthetic covariates that represent cumulative sums of the first few polynomials of the covariate. Nevertheless the restrictiveness of the assumption that the lag coefficients of a covariate form a polynomial of lower degree calls for more flexible methods that are able to estimate lag effects with complex and not necessarily polynomial shapes.

Shiller (1973) uses a Bayesian approach and implicitly defines the polynomials for the distribution of the lags for a linear model. Welty et al. (2009) offer an empirical Bayesian approach. Prior information about the shape of the lag course is used to model the effect of particulate matter on mortality. Heaton and Peng (2012) use Gaussian processes to construct a prior model for the distributed lag function in order to model the effect of temperature on mortality. Heaton and Peng (2014) extend the models defined by Heaton and Peng (2012) and incorporate interactions between lagged covariates.

Rushworth et al. (2013) introduce flexible varying coefficient models to model the relationship between rainfall and river flow in river systems. Heagerty and Comstock (2013) propose a transition model to model lagged associations of longitudinal data illustrated by an example from anemia treatment of dialysis patients.

Zanobetti et al. (2000) examine mortality in dependence of air pollutants. They propose to extend polynomial lag modeling by means of truncated power series bases with the addition of a penalty term. This way of distributed lag estimation allows the estimation of flexible lag courses. Gasparrini et al. (2010) increase the model flexibility. In their B-spline based approach for the lag modeling, the effect of the covariate on the response variable is assumed to be nonlinear. In their application, the effect of temperature on daily mortality is examined and it is assumed that both, heat and cold increase mortality, whereas moderate
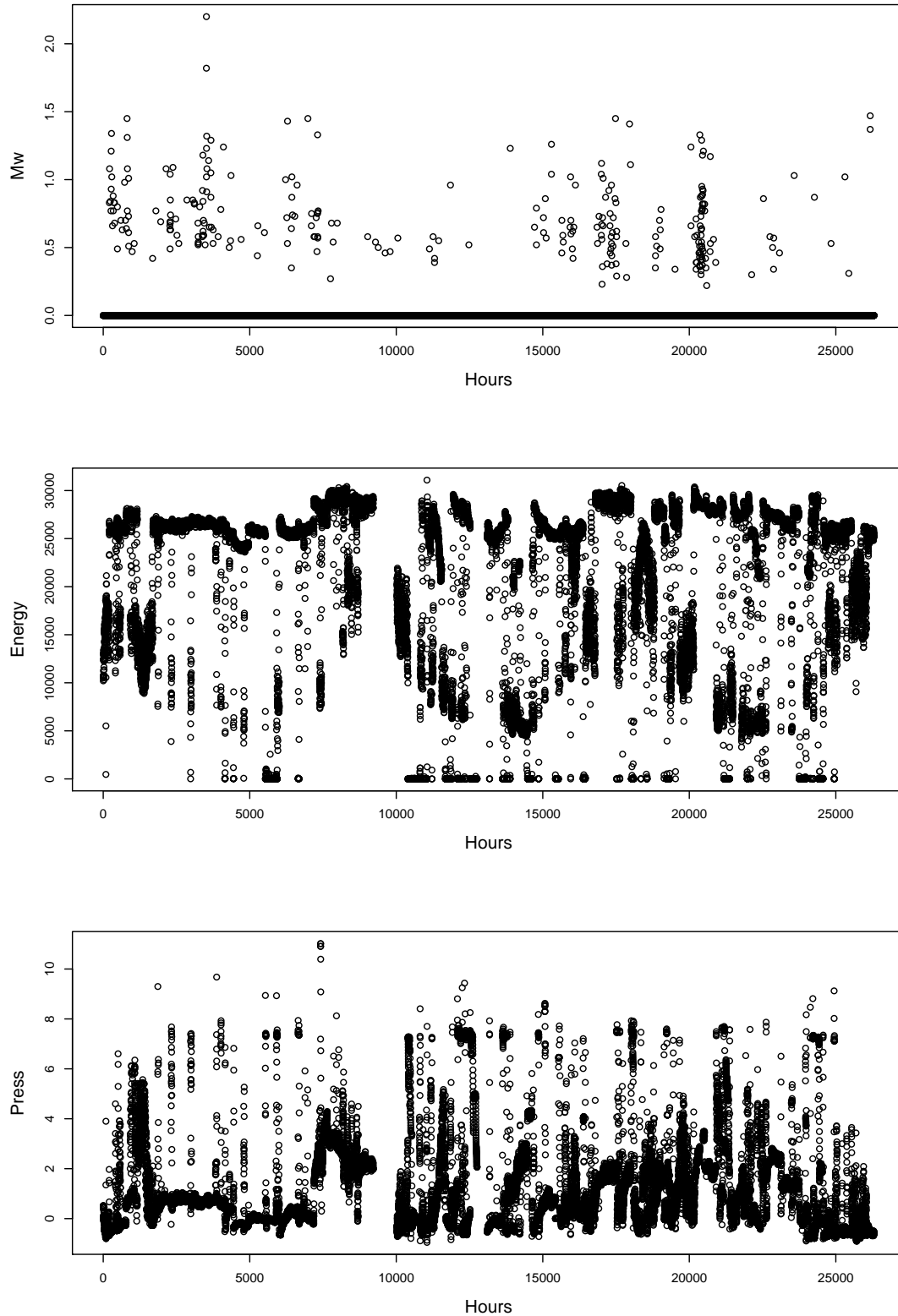
Figure 1.7: Geothermal data: Summed moment magnitude M_w (top), covariate energy (center), covariate press (bottom).

temperatures increase the wellbeing of human beings and lead to decreasing mortality rates. A crossbasis describes the distributed lag effect and simultaneously the shape of the relationship along the covariate. An additional penalty is not used. Muggeo (2008) models a v-shaped death-temperature relationship via a segmented break point model. Delayed temperature effects on the left (β_1 = cold related log risk) and on the right side (β_2 = heat related log risk) of the break point are modeled via doubly penalized B-splines. A difference penalty models the smoothness of the lag curves. Additional non decreasing hyper parameters $\nu_1, \ldots, \nu_L$ control the shrinkage of the distributed lag curves.

1.3 Content of the Thesis

In this thesis, we develop B-spline based lag modeling approaches for linear (flexible distributed lag, FDL) and nonlinear influences of lagged covariates on the response variable and carefully examine the impact of lag modeling via B-splines. The lag course is smoothed via a difference penalty and the last lag coefficient is shrunk towards zero via a ridge penalty, penalizing the basis functions related to the last lag coefficient. This additional ridge penalty offers an approach to another problem neglected in previous work. With the help of the additional ridge penalty, a suboptimal choice of the lag length L is no longer critical: We propose the use of longer lags as our simulations indicate that superfluous coefficients are correctly estimated close to zero. Furthermore, our simulation studies indicate that a too short lag length can be detected. The embedding in the *mgcv* (Wood, 2013a) context guarantees user friendly handling as well as stable, flexible and well-tested inference methods for the large group of generalized additive models (Wood, 2011; Marra and Wood, 2012; Wood, 2013b).

We transfer the penalization concept to the modeling of lagged nonlinear covariate effects. We use the approach of Gasparrini et al. (2010) and enlarge the approach by implementation of our penalizing concept.

1.4 Structure of the Thesis

The structure of the thesis is as follows. In Chapter 2, we introduce regression models with lags. We present penalized B-splines (Section 2.1) and the unconstrained lag model (Section 2.2) as an enlargement of generalized linear models. The problems of an unconstrained lag modeling are discussed in Section 2.3, followed by selected lag approaches in the literature (Section 2.4). Section 2.5 defines a lag model based on basis functions which will be penalized in Chapter 3.

Chapter 3 introduces the flexible distributed lag approach. We introduce the penalizing concept in Section 3.1 for a modeling without basis functions. That model is called the penalized lag model (PLM). In Section 3.2, the penalizing concept is transferred to the model with basis functions. Estimation and choice of the smoothing parameters are presented in Sections 3.4 and 3.5. A detailed simulation study (Section 3.6) shows a com-

prehensive performance comparison of the proposed flexible distributed lag model related to other approaches, followed by the presentation of the results of a data analysis of the Hochstaufen data. In this data example, the number of earthquakes in certain depth categories is related to the amount of rain fallen on preceding days. Section 3.8 delimits the proposed approach from previous approaches.

In Chapter 4, the penalization concept is translated to nonlinear lagged covariate effects. After introductions in the modeling of nonlinear covariate effects in general (Section 4.1) and a short summary of distributed lag non-linear models (DLNM, Section 4.2), we introduce the penalizing concept (Section 4.3). Section 4.4 shows the performance of the chosen penalty in comparison to the unpenalized approach of Gasparrini et al. (2010). We illustrate the penalized DLNM approach by means of volcano (Section 4.5) and geothermal energy data (Section 4.6).

1.5   Contributing Manuscripts

The present work is based on the following manuscripts:

- **Svejdar, V., Küchenhoff, H., Fahrmeir, L. and Wassermann, J. (2011): External forcing of earthquake swarms at Alpine regions: example from a seismic meteorogical network at Mt. Hochstaufen SE-Bavaria, Nonlin. Processes Geophys., 18(6), 849-860.**

  In this manuscript, we use two-dimensional Almon polynomials to model the lagged rain effect on the number of earthquakes in four depth categories at Mt. Hochstaufen, SE-Bavaria within the years 2002 until 2007.

  The introduction of the Hochstaufen data set in section 1.1.3 and the second part of section 2.4.1 are mainly based on contents of this manuscript.

  The article represents an enlargement of a report generated in the course "Statistisches Praktikum", where the cooperation between statisticians and geophysicists at the Ludwig-Maximilians University was developed further. I performed all analyses and wrote the sections 3 to 6 and parts of the discussion of the paper. The introductory sections describing the generation of the data set and the geophysical details were written by our geophysicist collaborator, Dr. Joachim Wassermann from the Geophysical Observatory of the Department of Earth- and Environmental Sciences of the Ludwig-Maximilians University Munich.

- Note: Viola Obermeier (b. Svejdar) is the first author of the paper.

- **Obermeier, V., Scheipl, F., Heumann, C., Wassermann, J. and Küchenhoff, H. (2014): Flexible Distributed Lags for Modeling Earthquake Data, Journal of the Royal Statistical Society, Series C, Submitted.**

This manuscript introduces a doubly penalized spline based approach for the inclusion of lagged linear covariate effects into a regression model context.

Chapter 3 is mainly based on contents of this manuscript.

The article contains contributions of Dr. Fabian Scheipl, Dr. Christian Heumann and Prof. Dr. Helmut Küchenhoff related to statistical concerns and Dr. Joachim Wassermann related to geophysical concerns. Dr. Fabian Scheipl was significantly involved in the embedding in *mgcv*. I performed all simulations and wrote the article.

- **Megies, T., Obermeier, V., Wassermann, J. and Küchenhoff, H. (2014): Correlating Geothermal Injection Parameters with Observed Seismicity using Penalized Distributed Lag Non-linear Models, Working Paper.**

This manuscript represents an application of flexible distributed lags for non-linear covariate effects to geothermal energy data.

Section 4.6 contains parts of this working paper.

I performed all analyses and wrote the statistical part of the manuscript.

1.6 Software

All calculations in this thesis were performed with the statistical software R (R Development Core Team, 2013). The estimation of the proposed lag approaches can be done via the function *gam* of the R-package *mgcv* (Wood, 2013a). In Chapter 4, the R-package *dlnm* (Gasparrini, 2011) is used.

Chapter 2

Regression Models with Lagged Covariate Effects

2.1 Penalized Splines based on B-Splines

Generalized linear models (McCullagh and Nelder, 1989) link the expectation $\boldsymbol{\mu} = E(\boldsymbol{y}|.)$ of a response variable $\boldsymbol{y}$ with the linear predictor $\boldsymbol{\eta}$ via a response function h and a link function $g = h^{-1}$, respectively, for response variables whose distribution is in the class of exponential families. This approach can be generalized to nonlinear covariate effects, where the influence of the covariate on the linked response variable can have any smooth form:

$$g(\boldsymbol{\mu}) = f(\mathbf{z})$$

for a single covariate $\mathbf{z}$. Additional covariates with (non-)linear influence on the (linked) response variable as well as an intercept can be added to the predictor. $f(\mathbf{z})$ can be represented by means of basis functions:

$$f(\mathbf{z}) = \sum_{i=1}^{I} \theta_i \mathbf{B}_i(\mathbf{z}).$$

with $\mathbf{B}_i(\mathbf{z})$ the i-th column of a matrix of (B-spline) basis functions evaluated on $\mathbf{z}$ and $\boldsymbol{\theta}$ a vector of unknown basis coefficients of length I. The number of basis coefficients I arises from the number m of knots $\kappa_j, j = 1, \ldots, m$ and from the degree d of the B-splines via $I = m + d - 1$. A small number of knots means smoothness of $\hat{f}(\mathbf{z})$, but possibly poor adjustment to the true function $f(\mathbf{z})$. A large number of knots can lead to a wiggly estimation of $f(\mathbf{z})$. Not only the number, but also the position of the knots play an important role for the estimation of $f(\mathbf{z})$. In addition to equidistant knots, quantile-based knots are a possibility to place the knots. In the approach of quantile-based knots, large numbers of knots are set in areas with a large number of observations. The visual knot choice based on a scatter plot can also be a possibility to place the knots. Nevertheless, the placement of the knots cannot solve the first problem, the question how many knots have to

be used in an estimation problem. Generally, two approaches can overcome this problem. One approach is the adaptive choice of the knots by means of model choice strategies. The second approach is the regularization of the estimation problem based on the introduction of a penalty term. This approach suggests the use of many equidistant knots. The resulting basis coefficients are smoothed via a penalty term. A quadratic penalty of the form $\lambda\boldsymbol{\theta}^T\mathbf{K}\boldsymbol{\theta}$ with smoothing parameter λ and penalty matrix $\mathbf{K}$ can be used in order to smooth and stabilize resulting estimates. The use of B-Spline bases (De Boor, 1978, see also the next paragraph) with discrete difference penalties has been popularized by Eilers and Marx (1996). The penalty matrix $\mathbf{K}$ is the product $\mathbf{D}_1^T\mathbf{D}_1$ of the difference matrix $\mathbf{D}_1$ (given for first order differences):

$$\mathbf{D}_1 = \begin{pmatrix} -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \end{pmatrix} \tag{2.1}$$

with

$$\mathbf{D}_1\boldsymbol{\theta} = \begin{pmatrix} \theta_2 - \theta_1 \\ \vdots \\ \theta_I - \theta_{I-1} \end{pmatrix}. \tag{2.2}$$

The $(I - 2) \times (I)$ difference matrix D_2 for differences of second order is:

$$\mathbf{D}_2 = \begin{pmatrix} 1 & -2 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & -2 & 1 \end{pmatrix} \tag{2.3}$$

Large values of the smoothing parameter λ mean large penalization of differences of adjacent basis coefficients and therefore lead to smooth functions. With the penalized likelihood:

$$l_{pen}(\boldsymbol{\theta}) = l(\boldsymbol{\theta}) - \frac{\lambda}{2}\boldsymbol{\theta}^T\mathbf{K}\boldsymbol{\theta}, \tag{2.4}$$

the estimation of the basis coefficients

$$\hat{\boldsymbol{\gamma}}_{pen} = \arg\max_{\boldsymbol{\gamma}} l_{pen}(\boldsymbol{\theta}) \tag{2.5}$$

can be done numerically via penalized iteratively reweighted least squares (P-IRLS, Marx and Eilers, 1998; Wood, 2006, see also Section 3.4).

B-splines

Polynomial pieces of low degree (d) are added sufficiently smooth on (equidistant) knots and build the numerical stable so-called B-Spline basis functions. The basis functions of

degree $d = 1$ are:

$$B_j^1(x) = \frac{x - \kappa_j}{\kappa_{j+1} - \kappa_j} \mathbb{1}_{[\kappa_j, \kappa_{j+1})}(x) + \frac{\kappa_{j+2} - x}{\kappa_{j+2} - \kappa_{j+1}} \mathbb{1}_{[\kappa_{j+1}, \kappa_{j+2})}(x). \tag{2.6}$$

B-Splines of higher degree d are recursively defined via the calculation rule:

$$B_j^d(x) = \frac{x - \kappa_j}{\kappa_{j+d} - \kappa_j} B_j^{d-1}(x) + \frac{\kappa_{j+d+1} - x}{\kappa_{j+d+1} - \kappa_{j+1}} B_{j+1}^{d-1}(z). \tag{2.7}$$

One B-Spline basis function consists of $(d + 1)$ polynomial pieces of degree d. B-Splines have several further good characteristics, that will be used and explained more in detail in Section 3.2.

2.2 Unconstrained Lag Model (ULM)

In the Hochstaufen data, one is interested in the influence of the amount of rain on the number of earthquakes in a time unit (e.g., a day). Assuming that the distribution of the number of earthquakes $y(t)$ on a day t depends on the rain on day t, x_t and on several (L) days before ($x_{t-1}, \ldots, x_{t-L}$), the resulting generalized linear model can be written as (intercept and further covariates suppressed):

$$y(t) \sim Po(\mu_t)$$
$$g(\mu_t) = \log(\mu_t) = \eta_t = x_t \beta_0 + x_{t-1} \beta_1 + \ldots + x_{t-L} \beta_L$$
$$= \sum_{l=0}^{L} x_{t-l} \beta_l \tag{2.8}$$

with $t = 1, \ldots, T$ the number of observed time units and $l = 0, \ldots, L$ the number of lagged time units included in the model.

In matrix notation one can write:

$$g(\boldsymbol{\mu}) = \underset{T \times (L+1)}{\mathbf{X}_{lag}} \underset{(L+1) \times 1}{\boldsymbol{\beta}} \tag{2.9}$$
$$\underset{T \times 1}{}$$

with

$$\underset{T \times (L+1)}{\mathbf{X}_{lag}} = \begin{pmatrix} x_1 & NA & \ldots & NA \\ x_2 & x_1 & & \vdots \\ \vdots & x_2 & & \vdots \\ & \vdots & & NA \\ & & & x_1 \\ & & & x_2 \\ & & & \vdots \\ x_T & x_{T-1} & \ldots & x_{T-L} \end{pmatrix}. \tag{2.10}$$
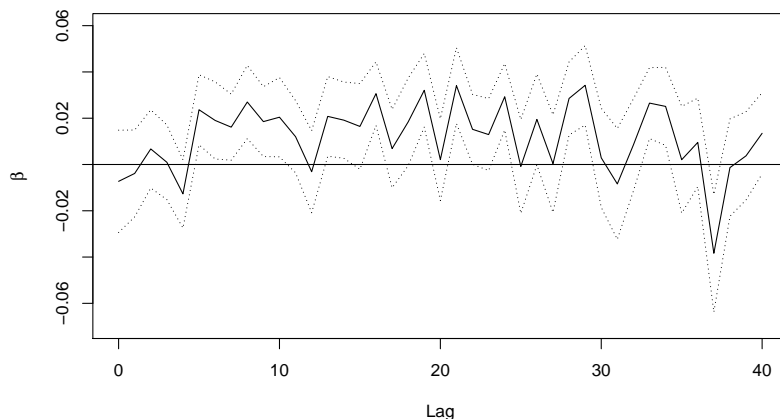
Figure 2.1: Unconstrained estimation and confidence intervals of lagged rain effects on the number of quakes in depth category 3 of the Hochstaufen data set.

2.3   Problems of the Unconstrained Lag Model

One can use a common generalized linear model with $(L + 1)$ covariates in order to incorporate the lag structure into the existing regression modeling technology. However, this simple, unrestricted modeling of the lagged effects of covariates entails two problems. First, a large number of parameters has to be estimated, if the lagged effect is assumed to be long. And second, the columns of $\mathbf{X}_{lag}$ can be highly collinear. These two problems together often cause instable estimates that are difficult to interpret. Figure 2.1 shows an example for an unconstrained estimation of lagged covariate effects. The aim is to estimate the lagged influence of rain on the number of earthquakes in the Hochstaufen data set. We see, that the estimated lag course $\hat{\boldsymbol{\beta}}$ is wiggly and that the confidence intervals are very high.

We can make the following two assumptions (see Welty et al., 2009) on interpretable and plausible shapes of the lag courses. First, we view β_l, $l = 1, \ldots, L$ as evaluations of a lag coefficient curve $\beta(l)$, and restrict the roughness of $\beta(l)$ by assuming a certain degree of regularity. One expects the effects of adjacent coefficients to be similar; so $\beta(l)$ is to be a smooth function. Second, the effects of the lagged covariates that are furthest in the past are assumed to be weak, so $|\beta(L)|$ should be close to zero. Of course, for this assumption to be reasonable, the maximal number of lag units L needs to be sufficiently large, so that $\beta(l)$ can capture all the effects of the covariate $\mathbf{x}$ on the response variable $\mathbf{y}$. In our application on the association between rainfall and earthquake activity, for example, geophysicists assume that lags of 20 to 40 days constitute the most likely time frame of relevant effects. It is implausible that the effect of rainfall on a particular day on future earthquakes is everlasting, so the last lag coefficient is expected to be close to zero.

2.4   Selected Approaches for Lagged Covariate Effects in the Literature

We deepen the understanding of distributed lag models and introduce some selected lag approaches in the literature. Some aspects of some approaches become relevant for further contents of this thesis.

2.4.1   Almon Polynomial Modeling (Almon, 1965)

Almon (1965) proposes an easy approach to give a structure to the lag coefficients β_l, $l = 0, \ldots L$. It is assumed that the $\boldsymbol{\beta}$ coefficients of the lagged covariates lie on a polynomial of lower degree d:

$$\beta_l = \gamma_0 + l^1\gamma_1 + \ldots + l^d\gamma_d = \tag{2.11}$$

$$= \sum_{j=0}^{d} l^j\gamma_j. \tag{2.12}$$

In doing so, one gets a smooth estimation of the course of the lag coefficients. The effect of adjacent lag coefficients is assumed to be similar. Insertion of Equation (2.11) in Equation (2.8) leads us to:

$$g(\mu_t) = \gamma_0 \underbrace{\left(\sum_{l=0}^{L} x_{t-l}\right)}_{\boldsymbol{\omega}_{\gamma_0 t}} + \gamma_1 \underbrace{\left(\sum_{l=0}^{L} l^1 x_{t-l}\right)}_{\boldsymbol{\omega}_{\gamma_1 t}} + \ldots +$$

$$+ \gamma_d \underbrace{\left(\sum_{l=0}^{L} l^d x_{t-l}\right)}_{\boldsymbol{\omega}_{\gamma_d t}}.$$

It can easily be seen, that after construction of the new design variables $(\boldsymbol{\omega}_{\gamma_0}, \ldots, \boldsymbol{\omega}_{\gamma_d})$ a known generalized linear model can be calculated. We achieve the pleasant spin-off, that the number of estimated parameters is clearly decreased from $L+1$ to $d+1$ coefficients. With

$$\begin{aligned}
\mathbf{u}_l &= \begin{pmatrix} 1 & l & l^2 & \ldots & l^d \end{pmatrix}' \quad \text{and} \\
\mathbf{V} &= \text{estimated covariance matrix of } \boldsymbol{\theta} = (\gamma_0, \ldots, \gamma_d)'
\end{aligned}$$

the standard error (se) of $\hat{\beta}_l$ can be calculated as

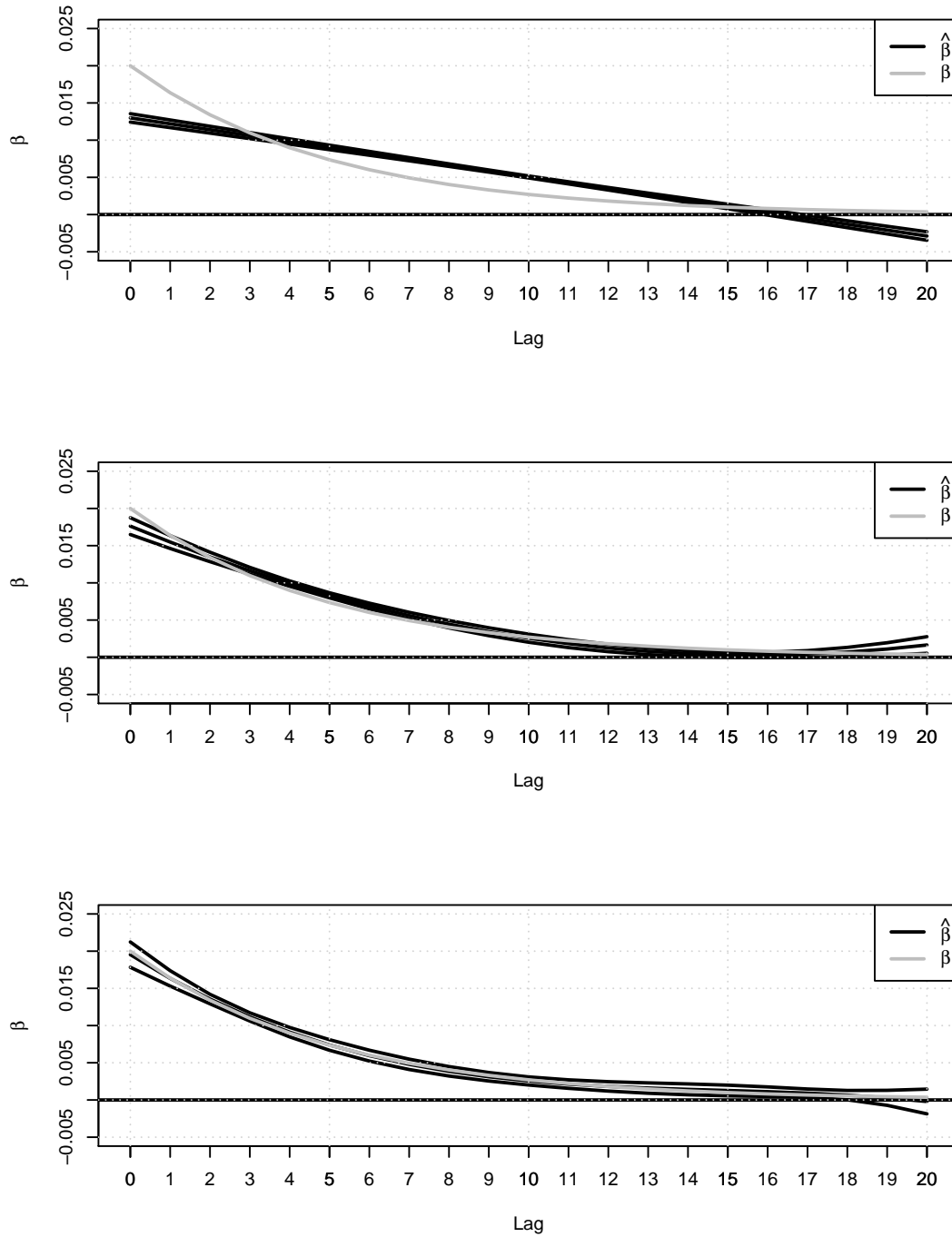$$se(\hat{\beta}_l) = \sqrt{\mathbf{u}_l'\mathbf{V}\mathbf{u}_l}.$$

Figure 2.2: Simulated (true) course of lag coefficients $\boldsymbol{\beta}$ (grey) vs. estimation $\hat{\boldsymbol{\beta}}$ via an Almon polynomial of first (top), second (center) and third degree (bottom).

Figure 2.2 shows, how well the estimation of distributed lag coefficients via the Almon polynomials can work. We create the course of lag coefficients β_l via an exponential function. Real temperature data acts as covariate for the simulated data and we draw 1000 random Poisson distributed response variables with the predictor as parameter of the Poisson distribution. We estimate the lag coefficients via Fisher Scoring with Almon polynomials of first (top), second (center) and third degree (bottom). We repeat this procedure 1000 times and build the average over the lag coefficients β_l and their confidence intervals. We compare averaged estimation to the underlying true lag function (grey). We see, that a linear modeling of the course of the lag coefficients (degree 1) is not sufficient for a good modeling of the exponential course of the lag coefficients. A polynomial of second degree (center) generates a good improvement and a polynomial of third degree already generates a very good fit to the data.

External Forcing of Earthquake Swarms at Alpine Regions: Example from a Seismic Meteorological Network at Mt. Hochstaufen SE-Bavaria

An example for polynomial modeling of the lag coefficients can be found in Svejdar et al. (2011). In this paper, we model the influence of (lagged) rainfall on the number of earthquakes in four different depth categories at Mt. Hochstaufen in Bavaria. We analyse amongst others daily values of rainfall and the number of earthquakes in four depth categories relative to the earth' s surface. The response variable is count data:

$$Y_{jt} \quad = \quad \text{Number of quakes in depth category } j \text{ at day t.}$$

Except the instant influence of the rain x_t, we also take the $l = 1, \ldots, 20$ lagged rain days into the predictor. The predictor in depth category j at day t is given by:

$$\eta_{jt} \quad = \quad \sum_{l=0}^{20} x_{t-l} \cdot \beta_{lj}. \tag{2.13}$$

where the β_{lj} are the coefficients for rain lag l in depth category j. Tide variables and temperature data are part of the analysis as well as a nonlinear time trend (see Svejdar et al., 2011, for further details). We use a two-dimensional polynomial of second degree to model the lagged rain-depth interaction

$$\beta_{lj} = \gamma_0 + \delta_1 l + \gamma_1 j + \delta_2 l^2 + \gamma_2 j^2 + \nu l j \tag{2.14}$$

for $l = 0, \ldots, 20$ and $j = 1, \ldots, 4$.

The two-dimensional polynomial implies the following reasonable assumptions:

1. within a fixed depth category, the effect of adjacent rain lags should be similar,

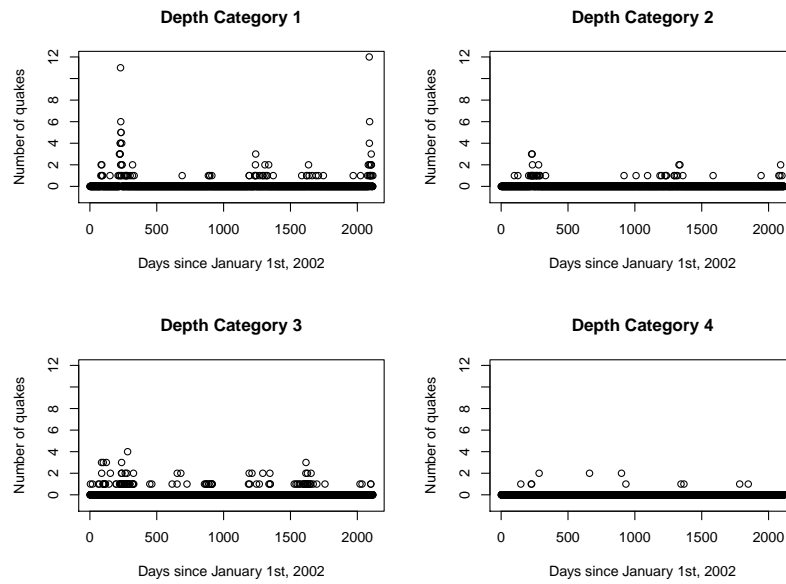2. for a fixed rain lag, the effect of adjacent depth categories should be similar.

Figure 2.3: Hochstaufen data: Daily counts of quakes in the sample period of 2100 days by depth categories.
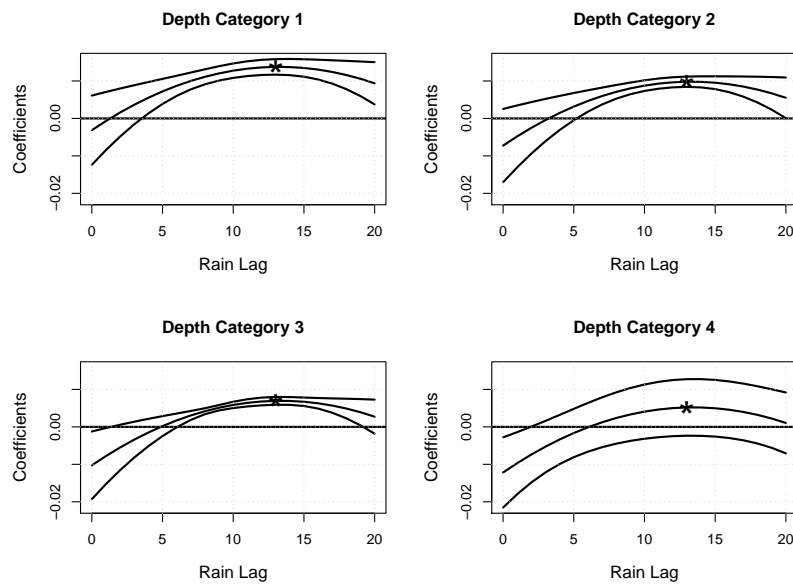
Figure 2.4: Hochstaufen data: Polynomial estimation.

Depth Category j	$j = 1$	$j = 2$	$j = 3$	$j = 4$
Depth in km	$]0, -1]$	$]-1, -2]$	$]-2, -4]$	$]-4, -\infty[$
Number of quakes	162	56	144	14
$l_{\max(j)}$	Lag13	Lag13	Lag13	Lag13
$\hat{\beta}_{l_{\max(j)}}$	0.014	0.010	0.007	0.005

Table 2.1: Hochstaufen data: Lags of prime importance on the number of quakes (Poisson regression based on daily values).

In Figure 2.3, the number of quakes on each of the $n = 2100$ days for each depth category is shown. The least quakes are counted in the deepest category 4. Figure (2.4) shows the resulting coefficient plots in each depth category j. In every depth category, the maximum effect of rain on the number of quakes is on lag 13 (see also Table 2.1). The coefficient plot also shows a decrease of the rain effect with increasing depth. The larger the distance of a depth class from the earth's surface (depth category 4), the smaller and less significant the respective coefficients. The strength of the maximum effect in each depth class decreases from 0.014 (depth category 1) to 0.005 (depth category 4), see Table 2.1. This corresponds to an increase of the expected number of earthquakes of the factor $\exp(10 \cdot 0.014) = 1.15$ and $\exp(10 \cdot 0.005) = 1.05$, respectively per 10 mm additional rain. We address the question of the accuracy of the determination of the maximum in each depth category. Focusing on the polynomial form of the rain-depth coefficients

$$\hat{\beta}_{lj} = \hat{\gamma}_0 + \hat{\delta}_1 l + \hat{\gamma}_1 j + \hat{\delta}_2 l^2 + \hat{\gamma}_2 j^2 + \hat{\nu} lj,$$

the maximal lag in every depth category j can be obtained after differentiating with respect to l and equating to zero via:

$$l_{\max(j)} = \frac{-\hat{\delta}_1 - \hat{\nu} j}{2\hat{\delta}_2}.$$

The estimated variance of the maximum in each depth category can then be calculated as:

$$\widehat{\mathrm{Var}(l_{\max(j)})} = \tag{2.15}$$

$$\left(-\frac{1}{2\hat{\delta}_2}, \quad \frac{\hat{\delta}_1 + \hat{\nu} j}{2\hat{\delta}_2^2}, \quad -\frac{j}{2\hat{\delta}_2} \right) \widehat{\mathrm{Var}(\hat{\delta}_1, \hat{\delta}_2, \hat{\nu})} \begin{pmatrix} -\frac{1}{2\hat{\delta}_2} \\ \frac{\hat{\delta}_1 + \hat{\nu} j}{2\hat{\delta}_2^2} \\ -\frac{j}{2\hat{\delta}_2} \end{pmatrix}.$$

For the Poisson model based on daily values, the estimated standard deviations of the maximum in each depth category can be seen in Table 2.2. The standard deviations of the maxima alternate poorly for the four depth categories and the maxima can be assumed to be specified relatively exact with standard deviations of about 0.7.

Although the modeling of distributed lag coefficients via Almon polynomials can have a quite good performance, we have to respond to the general problems of polynomial
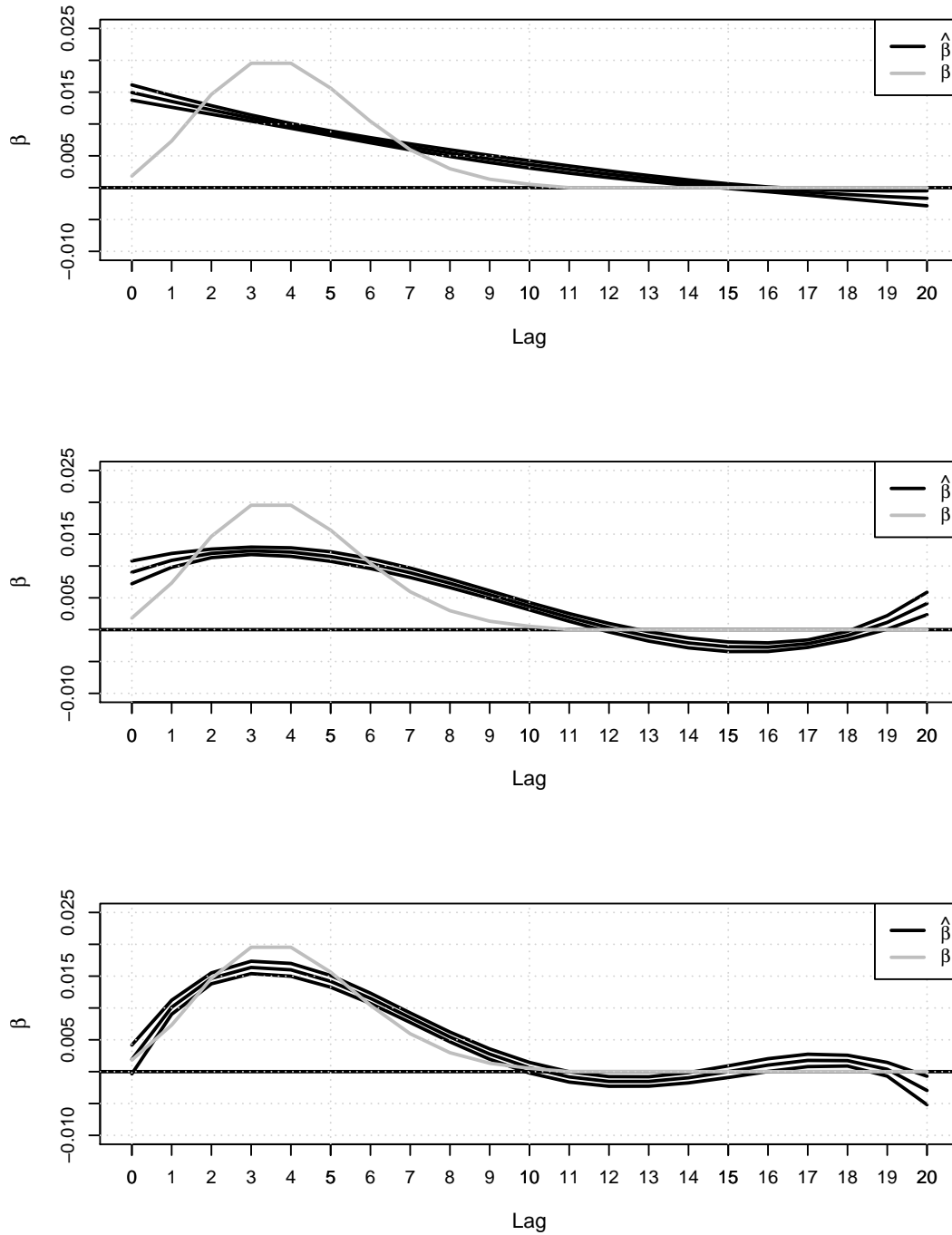
Figure 2.5: Simulated (true) course of lag coefficients $\boldsymbol{\beta}$ (grey) vs. estimation $\hat{\boldsymbol{\beta}}$ via an Almon polynomial of second (top), third (center) and forth (bottom) degree.

Depth Category j	$j = 1$	$j = 2$	$j = 3$	$j = 4$
$l_{\max(j)}$	Lag 13	Lag 13	Lag 13	Lag 13
$\sqrt{\mathrm{Var}(\widehat{i_{\max(j)}})}$	0.680	0.661	0.653	0.655

Table 2.2: Hochstaufen data: Accuracy of the maximum for the Poisson regression based on daily values.

modeling. While polynomials of higher degree can show the desired flexibility, they have the serious drawback, that they are highly oscillating functions. Figure 2.5 shows, similar to Figure 2.2, a true lag function (grey) and its estimation via Almon polynomials of second (top), third (center) and forth (bottom) degree (average over 1000 runs). We see that the unimodal shape of the lag function in general can be well discovered by polynomials of higher degrees. Polynomials of lower degrees have more problems to find the correct shape. Nevertheless, we encounter the problem, that polynomials of higher degrees result in oscillating functions. The lag coefficients from 10 to 20, that are close to zero, are over- and underestimated, respectively. Local polynomials could bring a large improvement in the estimation of a distributed lag function.

2.4.2 Bayesian Distributed Lag Models: Estimating Effects of Particulate Matter Air Pollution on Daily Mortality (Welty et al., 2009)

Welty et al. (2009) introduce an empirical Bayesian approach for the inclusion of lagged linear covariate effects into the regression context. Prior knowledge about the shape of the distributed lag function is incorporated into the model specification. The Bayesian distributed lag formulation is, corresponding to the authors, relevant to situations in which the lagged effects of an exposure on an outcome are unknown for the first few lags but are believed to dissipate with lag. Welty et al. (2009) formulate distributed lag models so that coefficients are constrained to approach zero smoothly with increasing lag. Early coefficients are relatively unconstrained. The application is limited to Poisson log-linear distributed lag models with the specific application of the examination of the effect of particulate matter air pollution on mortality counts. Welty et al. (2009) define generalized linear models with lags via $g(E(y_t \mid x_1, \ldots, x_t)) = \sum_{l=0}^{L} \beta_l x_{t-l}$ where L is the maximum lag and $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_L)'$ is the vector of the lag coefficients to be estimated. The goal is to specify a prior on $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_L)'$ that is uninformative on the lag coefficients for small l but that constrains the coefficients with larger l to be smoother and approach zero.

They assume

$$\boldsymbol{\beta} \ \sim \ N(\mathbf{0}, \boldsymbol{\Omega})$$

with $\boldsymbol{\Omega}$ so that $\mathrm{Var}(\beta_l) \to 0$ and $\mathrm{Cor}(\beta_{l-1}, \beta_l) \to 1$ via $\boldsymbol{\Omega} = \mathbf{ABA}$. $\mathbf{AA}$ is defined as

the diagonal matrix of the variances of the coefficients and $\mathbf{B}$ is the correlation matrix for the lag coefficients $\boldsymbol{\beta}$. $\mathbf{A}$ is constructed via $\mathbf{A} = \sigma \mathbf{V}(\eta_1)$. $\mathbf{V}(\eta_1)$ is a diagonal matrix with $\eta_1 < 0$. The entries of $\mathbf{V}(\eta_1)$ are $1, v_1(\eta_1)^{\frac{1}{2}}, \ldots, v_L(\eta_1)^{\frac{1}{2}}$ with $v_l(\eta_1) = \exp(\eta_1 l)$, a decreasing sequence of weights. σ^2 is the prior variance of θ_0. The correlation matrix $\mathbf{B}$ is constructed in a similar way via a hyperparameter η_2, see Welty et al. (2009) for detailed information. In the case of a normal distribution $\hat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}, \boldsymbol{\Sigma})$, the posterior for $\boldsymbol{\beta}$ conditional on $\boldsymbol{\eta}$ and σ^2 follows a normal distribution and can be written in closed form. The posterior distribution for $\boldsymbol{\beta}$ can be computed via Gibbs sampling (Carlin and Louis, 2000) for a general linear distributed lag model.

The choice of σ^2, of the hyperparameters η_1, η_2 and of the exponential function pose a critical point in that approach. A further examination of the effect of these parameters could be interesting.

2.4.3 Generalized Additive Distributed Lag Models: Quantifying Mortality Displacement (Zanobetti et al., 2000)

Zanobetti et al. (2000) examine the lagged effect of particulate matter on mortality with application to data from Milan, Italy for the years 1980 till 1989. Additionally, meteorological data such as temperature and relative humidity are part of the analysis. Zanobetti et al. (2000) firstly introduce generalized additive distributed lag models. As the name suggests, they combine generalized additive models and distributed lag models. The aim is to model in a parsimonious way the dependency of the number of daily deaths in dependence of lagged air pollutants and further covariates. The effect of particulate matter on mortality is assumed to take effect with some retard. Zanobetti et al. (2000) introduce a spline based distributed lag approach in order to examine long-term lagged effects. The approach is motivated via the so-called "mortality displacement". People, that are menaced by air pollutants due to possible pre-existing illness, are part of a risk pool according to the terminology of mortality displacement. High air pollution leads to higher mortality and to a depletion of that risk pool. Therefore, less people are under risk on subsequent days and a negative association of mortality with air pollution might be induced at those lags. This possible negative association is called the "rebound effect". Zanobetti et al. (2000) introduce a more flexible approach - more flexible in comparison to Almon polynomials - in order to fit long-term effects of particulate matter. The model is:

$$g\{E(y_t)\} = \sum_{l=0}^{L} \beta_l x_{t-l} + \text{further covariates}, \quad t = L+1, \ldots, T \qquad (2.16)$$

where g is a link function and β_l the lag coefficients, equivalent to Equation (2.8). In the paper, $\sum_{l=0}^{L} \beta_l$ is defined as the overall impact of a unit change in exposure over the next L days. Zanobetti et al. (2000) enlarge polynomial distributed lag models (PDLM, in this

thesis introduced as Almon polynomials) where the β_l is restricted to:

$$\beta_l = \sum_{j=0}^{p} \tau_l l^j, \quad l = 0, \ldots, L \tag{2.17}$$

to piecewise pth degree polynomial in l with knots at κ_k:

$$\beta_l = \sum_{j=0}^{p} \tau_l l^j + \sum_{k=1}^{K} \nu_k (l - \kappa_k)_+^p. \tag{2.18}$$

The truncated power series basis coefficients ν_l are restricted via a penalty term $\lambda \sum_{k=1}^{K} \nu_k^2$ in order to get estimates that are not too wiggly. The estimation is performed via iteratively reweighted least squares. The estimation of the smoothing parameters is done via GCV. Zanobetti et al. (2000, p 286, bottom) write that practical problems occured when selecting each of the smoothing parameters. The authors use GCV "to select *one* of the smoothing parameters if the remaining ones are fixed by other means".

2.4.4 Modeling Temperature Effects on Mortality: Multiple Segmented Relationships with Common Break Points (Muggeo, 2008)

Muggeo (2008) and Muggeo (2003) present a doubly penalized spline based distributed lag approach for a break point model. They examine lagged effects for one specific application within the context of a segmented break point model. Muggeo (2008) introduces a model for the effect of temperature on mortality with application to data from Milan, Italy of the years 1985 to 1989. Mortality is, according to the author, known to be minimal at some optimal value of temperature and to increase as temperature gets colder or hotter. Therefore, the doubly penalized spline based distributed lag approach within a segmented regression model represents a transition to nonlinearity of the covariate effect. The mortality-temperature relationship (under disregard of lagged effects) can be described via a 2-break point segmented model. An example is the "bathtub type"-shaped curve. The log-link model with two breakpoints Ψ_1 and Ψ_2 can be written as:

$$log E(Y_t) = \eta_t + \beta_1 (Z_t - \psi_1)_- + \beta_2 (Z_t - \psi_2)_+ \quad \text{with} \tag{2.19}$$
$$(Z_t - \psi_1)_- = (Z_t - \psi_1) \mathbb{1}(Z_t < \psi_1) \quad \text{and} \tag{2.20}$$
$$(Z_t - \psi_2)_+ = (Z_t - \psi_2) \mathbb{1}(Z_t > \psi_2). \tag{2.21}$$

η_t describes further covariates and Z_t is the temperature at time t. β_1 is the cold-related log risk if the temperature falls below the break point ψ_1 and β_2 the heat-related log risk if the temperature exceeds ψ_2. Mortality is constant and minimal in between the break points, in the optimal temperature margin. Including lagged variables and simplified for

a v-shaped relationship with only one break point ψ, the constrained multilag-segmented model can be written as:

$$logE(\mathbf{y}) = \boldsymbol{\eta} + \sum_{l_1=0}^{L_1-1} \boldsymbol{\beta}_{1l_1}(\mathbf{Z}_{l_1} - \psi)_- + \sum_{l_2=0}^{L_2-1} \boldsymbol{\beta}_{2l_2}(\mathbf{Z}_{l_2} - \psi)_+. \qquad (2.22)$$

The vectors described with $\mathbf{Z}_{l_1}$ and $\mathbf{Z}_{l_2}$ contain exposures of respective preceding days. Thus, $\boldsymbol{\beta}_1 = (\beta_{10}, \beta_{11}, \ldots, \beta_{1l_1}, \ldots, \beta_{1L_1-1})^T$ and $\boldsymbol{\beta}_2 = (\beta_{20}, \beta_{21}, \ldots, \beta_{2l_2}, \ldots, \beta_{2L_2-1})^T$ describe the distributed lag curves of cold and heat. To simplify the model, it is assumed that the break points are the same for every lag.

The distributed lag curves on the right and on the left curves of the break point are expressed by means of B-splines:

$$\boldsymbol{\beta}_1 = \mathbf{Cb}_1 \qquad (2.23)$$
$$\boldsymbol{\beta}_2 = \mathbf{Hb}_2 \qquad (2.24)$$

with $\mathbf{C} = (\mathbf{C}_1, \ldots, \mathbf{C}_{P1})$ and $\mathbf{H} = (\mathbf{H}_1, \ldots, \mathbf{H}_{P2})$ two B-spline bases and corresponding basis coefficients $\mathbf{b}_1$ and $\mathbf{b}_2$. Muggeo (2008) imposes two penalties for each of the distributed lag curve. A difference penalty $\lambda \sum_j (b_j - 2b_{j-1} + b_{j-2})^2$ in order to smooth each of the two lag curves and a ridge term $\omega \sum_l \beta_l^2 \nu_l$. Nondecreasing known weights $\nu_1, \ldots, \nu_L$ encourage decreasing temperature effects with increasing lags. Muggeo (2008) proposes $\nu_l = l$ or $\nu_l = l^2$ for the weights. The estimation is performed via maximization of the penalized log-likelihood and the smoothing parameter estimation is performed via the unbiased risk estimator (UBRE) score.

Software

Muggeo (2010) describes the R-package *modTempEff* (Muggeo, 2013) that implements the "Estimation of a Poisson log linear regression to model the temperature effect on health using the constrained segmented distributed lag parameterization which allows to account simultaneously for typical features of temperature effects: nonlinear effect delayed over several days."

2.4.5   Distributed Lag Non-Linear Models (Gasparrini et al., 2010)

Muggeo (2008, see Section 2.4.4) presents with his segmented break point model a first transition to the inclusion of lagged nonlinear covariate effects on a response variable. Muggeo (2008) modeles a v-shaped relationship between mortality and (lagged) temperature effects. Gasparrini et al. (2010) also models the relationship between mortality and lagged temperature effects and enlarges the model and accordingly the application range to common nonlinear lagged covariate effects. The approach is called "Distributed lag non-linear models" (DLNM). The specific application data originates from the National Morbidity, Mortality and Air Pollution Study (NMMAPS, Dominici et al., 2005) for New

York for the years 1987 to 2000. Gasparrini et al. (2010) introduce the lagged information of a covariate $\mathbf{x}$ via a $T \times (L+1)$ matrix $\mathbf{X}_{lag}$ (see Equation (2.10)). The maximum lag is defined by L and the first column of $\mathbf{X}_{lag}$ equals the covariate $\mathbf{x}$; in the current application $x_t, t = 1, \dots T$ is the temperature at day t. The shape of the lag effects is specified by basis functions, e.g., B-splines:

$$s(x; \boldsymbol{\gamma}) = \mathbf{X}_{lag}\mathbf{B}\boldsymbol{\gamma} \qquad (2.25)$$

with $\mathbf{B}$ a $(L+1) \times \nu_l$ matrix of basis variables and $\boldsymbol{\gamma}$ a vector of unknown basis coefficients. Gasparrini et al. (2010) define

$$\mathbf{Z} = \mathbf{X}_{lag}\mathbf{B}, \qquad (2.26)$$

where $\mathbf{Z}$ is the new synthetic design matrix performed of lagged information and a basis matrix. The estimated lag coefficients $\hat{\boldsymbol{\beta}}$ and their covariance matrix $\text{Var}(\hat{\boldsymbol{\beta}})$ can be performed by means of the estimated basis coefficients $\hat{\boldsymbol{\gamma}}$:

$$\hat{\boldsymbol{\beta}} = \mathbf{B}\hat{\boldsymbol{\gamma}} \qquad (2.27)$$
$$\text{Var}(\hat{\boldsymbol{\beta}}) = \mathbf{B}\text{Var}(\hat{\boldsymbol{\gamma}})\mathbf{B}^T \qquad (2.28)$$

DLNM is a modeling framework that can simultaneously represent nonlinear and delayed effects. A crossbasis is used to combine the shape of the relationship along the covariate and its distributed lag effects. A $T \times \nu_x \times (L+1)$ array $\dot{\mathbf{R}}$ represents the lagged occurrences of each of the basis variables in covariate direction. By means of $\mathbf{B}$, the matrix of basis variables in lag direction, a DLNM is specified by:

$$s(x_t, \boldsymbol{\gamma}) = \sum_{j=1}^{\nu_x} \sum_{k=1}^{\nu_l} \mathbf{r}_{tj.}^T \mathbf{b}_{.k} \gamma_{jk} = \mathbf{z}_{t.}^T \boldsymbol{\gamma}. \qquad (2.29)$$

Lagged exposures at time t transformed through basis function j are defined by $\mathbf{r}_{tj.}$. $\mathbf{z}_{t.}$ is the result of the application of the $\nu_x \cdot \nu_l$ cross-basis functions to x_t. The authors present $\mathbf{Z}$ as a tensor product and define

$$\dot{\mathbf{A}} = (\mathbf{1}^T \otimes \dot{\mathbf{R}}) \odot (\mathbf{1} \otimes P_{1,3}(\mathbf{B}) \otimes \mathbf{1}^T) \qquad (2.30)$$

where $P_{i,j}$ is an operator permuting the indices i and j. $\mathbf{1}$ indicate vectors of ones of appropriate length. The symbol $\otimes$ indicates a Kronecker product and the symbol $\odot$ a Hadamard product. The matrix of crossbasis functions $\mathbf{Z}$ is generated by re-arranging the $T \times (\nu_x \nu_l) \times (L+1)$ array $\dot{\mathbf{A}}$ and summing along the third dimension of lags, respectively. Gasparrini et al. (2010) interpret a DLNM by building a grid of predictions for lag and covariate values:

$$\mathbf{e}_{.l} = \mathbf{A}_{..l}^p \hat{\boldsymbol{\gamma}} \qquad (2.31)$$

$$\mathbf{e}_{.l}^{sd} = \sqrt{\mathrm{diag}(\mathbf{A}_{..l}^{p}\mathrm{Var}(\hat{\boldsymbol{\gamma}})\mathbf{A}_{..l}^{pT})}. \tag{2.32}$$

Gasparrini (2014) models exposure-lag-response associations with distributed lag non-linear models, extending beyond time series data. In this manuscript, Gasparrini (2014) refers to Sylvestre and Abrahamowicz (2009) and introduces constraints on B-splines bases of the DLNMs that forces lag courses at a maximum lag to be zero.

Software

Gasparrini (2011) describes the R-package *dlnm* (Gasparrini and Armstrong, 2013). It contains the mentioned construction of the crossbasis (see Section 2.4.5) as well as different plot commands for the presentation of distributed lag non-linear models. The crossbasis generated within the R-package *dlnm* can then be used as synthetic covariate in any appropriate statistical model, for example in the R-package *mgcv*.

2.5   Lag Model based on Basis Functions

We introduce a lag model based on basis functions (see Bayerstadler, 2010; Gasparrini et al., 2010, for nonlinear covariate influence). The lag coefficients β_L are developed in B-splines of degree d:

$$\begin{aligned}
\beta_l &= B_1(l)\gamma_1 + \ldots + B_m(l)\gamma_m(l) \quad , \quad l = 0, \ldots, L \\
&= \sum_{j=1}^{m} B_j(l)\gamma_j
\end{aligned} \tag{2.33}$$

or in matrix notation:

$$\underset{(L+1)\times 1}{\boldsymbol{\beta}} = \underset{(L+1)\times m}{\mathbf{B}} \quad \underset{m\times 1}{\boldsymbol{\gamma}} \tag{2.34}$$

with

$$\underset{(L+1)\times m}{\mathbf{B}} = \begin{pmatrix} B_1(0) & \ldots & B_m(0) \\ \vdots & & \vdots \\ B_1(L) & \ldots & B_m(L) \end{pmatrix} \tag{2.35}$$

with K equidistant knots κ_k and $m = d + K$ basis coefficients. A positive side-effect can be seen in potential reduction of the number of parameters from $L + 1$ to m parameters associated with the number of knots. Following Eilers and Marx (1996), we use a relatively large number of knots. A data-driven amount of penalization will be introduced in Section 3.2.1 to avoid overfitting and to smooth the estimated lag course. If we insert Equation (2.34) into the unrestricted model (Equation (2.8)), we get:

$$g(\mu_t) = \sum_{l=0}^{L} x_{t-l} \left(\sum_{j=1}^{m} B_j(l)\gamma_j \right) =$$

$$= \sum_{j=1}^{m} \underbrace{\left(\sum_{l=0}^{L} x_{t-l} B_j(l) \right)}_{z_{tj}} \gamma_j. \tag{2.36}$$

In matrix notation we have:

$$g(\boldsymbol{\mu}) = \underbrace{\underset{T\times 1}{\mathbf{X}_{lag}} \quad \underset{(L+1)\times m}{\mathbf{B}}}_{\mathbf{Z}} \quad \underset{m\times 1}{\boldsymbol{\gamma}} . \tag{2.37}$$

This has the structure of a conventional (penalized) generalized linear model with new synthetic covariates. Instead of the lagged covariates $\mathbf{x}, \mathbf{x}_{-1}, \dots, \mathbf{x}_{-L}$, the new design matrix now contains the product of $\mathbf{X}_{lag}$ with the basis function matrix $\mathbf{B}$. Instead of the original lag parameters $\beta_0, \dots, \beta_L$, the new basis coefficients $\gamma_1, \dots, \gamma_m$ are to be estimated.

Chapter 3

Flexible Distributed Lags (FDL) for Linear Effects

3.1 Penalized Lag Model

In Section 1.1.2, we make demands on lag curves $\beta(l)$. $\beta(l)$ is to be a smooth function and should in addition be able to reproduce complex courses. The second demand refers to the last lag coefficient β_L. If the maximal lag length is chosen large enough, β_L is assumed to be close to zero in order to satisfy the requirement of smoothness between β_L and β_{L+1}. β_{L+1} is not part of the model and is accordingly implicitly set to zero. In Section 2.3, we mention problems of an unconstrained lag model, which can cause collinearity problems and can have a large number of parameters if the chosen lag length is large.

We now take advantage of the two assumptions of Section 1.1.2 in order to construct a penalty, that stabilizes estimates of $\beta(l)$ and yields more plausible effects than the unconstrained lag model (Section 2.2).

3.1.1 Smoothing of the Lag Course

The objective of obtaining smooth curves can generally be pursued by penalizing adjacent coefficients using a difference penalty (see Whittaker, 1923). We smooth the unconstrained lag model of Section 2.2 by penalizing adjacent lag coefficients. We introduce penalizing terms based on differences for a penalized lag model. Analogously to first order differences, differences of higher order can be used in order to obtain more smooth functions $\beta(l)$. We visualize the use of a difference penalty for the simplified case of a a normally distributed response $\mathbf{y}$ and one lagged covariate $\mathbf{x}$:

$$y_t = x_t\beta_0 + x_{t-1}\beta_1 + \ldots + x_{t-L}\beta_L \tag{3.1}$$

$$= \sum_{l=0}^{L} x_{t-l}\beta_l \tag{3.2}$$

or in matrix notation:

$$\underset{T \times 1}{\mathbf{y}} = \underset{T \times (L+1)}{\mathbf{X}_{lag}} \quad \underset{(L+1) \times 1}{\boldsymbol{\beta}} . \tag{3.3}$$

with $\mathbf{X}_{lag}$ defined as in Equation (2.10). The penalized residual sum of squares (PRSS) with smoothing parameter ξ_d (index d stands for "differences") is:

$$PRSS(\xi_d) = \sum_{t=1}^{T} (y_t - \mathbf{x}_t \hat{\boldsymbol{\beta}})^2 + \xi_d \sum_{l=b+1}^{L} (\triangle^b \beta_l)^2 \tag{3.4}$$

with $\mathbf{x}_t = (x_t, x_{t-1}, \dots, x_{t-L})$ the t-th row of matrix $\mathbf{X}_{lag}$ and a smoothing parameter ξ_d that monitors the extent of smoothing. $\triangle^b$ describes b-th order differences that are recursively defined via:

$$\triangle^1 \beta_l = \beta_l - \beta_{l-1}, \tag{3.5}$$
$$\triangle^2 \beta_l = \triangle_1 \triangle_1 \beta_l = \triangle_1 \beta_l - \triangle_1 \beta_{l-1} = \beta_l - 2\beta_{l-1} + \beta_{l-2}, \tag{3.6}$$
$$\vdots \tag{3.7}$$
$$\triangle^b \beta_l = \triangle^{b-1} \beta_l - \triangle^{b-1} \beta_{l-1}. \tag{3.8}$$

The smoothing parameter ξ_d penalizes large differences of adjacent coefficients. The larger ξ_d is chosen, the more the resulting differences of lag coefficients are penalized. The lag curve β_l is very smooth for large smoothing parameters ξ_d. In the case that ξ_d is equal to zero, the penalized residual sum of squares is identical to the residual sum of squares and minimization of the RSS leads to the estimates of the unconstrained lag model (ULM, Section 2.2). We calculate the penalty matrix $\mathbf{J}_d = \mathbf{D}_2^T \mathbf{D}_2$ with $\mathbf{D}_2$ defined as in Equation (2.3) when using differences of second order. The use of differences of second order is recommended. Second order differences are the standard option in the R package *mgcv*. Higher differences would yield in over-smoothing.

In matrix notation the PRRS can be written as:

$$PRSS(\xi_d) = (\mathbf{y} - \mathbf{X}_{lag} \boldsymbol{\beta})^T (\mathbf{y} - \mathbf{X}_{lag} \boldsymbol{\beta}) + \xi_d \boldsymbol{\beta}^T \mathbf{J}_d \boldsymbol{\beta}. \tag{3.9}$$

3.1.2   Shrinkage of the Lag Course

In addition to the smoothing penalty, we define a ridge penalty for lagged covariate effects that penalizes the last lag coefficient β_L with intent to shrink that parameter towards zero. We explicitly shrink the last lag coefficient towards zero for several reasons in form and content. First of all, when choosing a maximum lag length L, we implicitly assume β_{L+1} to be zero. β_{L+1} will not be estimated and is therefore set to zero. The requirement of smoothness of $\beta(l)$ means smoothness between β_L and β_{L+1} as well as smoothness in the previous lag coefficients. Therefore, the call for β_L to be zero seems to be an obvious requirement. We also want the last lag coefficient to be close to zero for content reasons.

We want to catch all lagged effects of rain on the number of earthquakes in our Hochstaufen data set, e.g., but disbelieve that the effect of rain on the number of earthquakes could be everlasting. The demand on the tool, that shall realize a possible shrinkage is the following. The last lag coefficient β_L is to be shrunk towards zero with soft pressure. β_L shall be close to zero if and only if the data is compatible with such a decay. In the simplified case of a normally distributed response variable $\mathbf{y}$, see Equation (3.1), the penalized residual sum of squares to be minimized completes Equation (3.4) to

$$PRSS(\xi_d, \xi_r) = \sum_{t=1}^{T} (y_t - \mathbf{x}_t \hat{\boldsymbol{\beta}})^2 + \xi_d \sum_{l=b+1}^{L} (\triangle^b \beta_l)^2 + \xi_r \beta_L^2. \tag{3.10}$$

A large shrinkage parameter leads to high penalizing of large values of β_L and therefore results in small values of β_L. In the case that both penalizing parameters, the smoothing parameter ξ_d and the shrinkage parameter ξ_r are zero, the estimation is identical to the estimation of an unconstrained lag model. The corresponding ridge penalty matrix is the $(L+1) \times (L+1)$ matrix

$$\mathbf{J}_r = \begin{pmatrix} 0 & & & \\ & \ddots & & \\ & & 0 & \\ & & & 1 \end{pmatrix} \tag{3.11}$$

and the matrix notation of PRSS computes to:

$$PRSS(\xi_d) = (\mathbf{y} - \mathbf{X}_{lag}\boldsymbol{\beta})^T (\mathbf{y} - \mathbf{X}_{lag}\boldsymbol{\beta}) + \xi_d \boldsymbol{\beta}^T \mathbf{J}_d \boldsymbol{\beta} + \xi_r \boldsymbol{\beta}^T \mathbf{J}_r \boldsymbol{\beta}. \tag{3.12}$$

Conscient Ridge Penalty for $\hat{\beta}_L$ vs. increasing Weights

We also thought about increasing weights to reach an increasing penalty of the lag coefficients (see e.g., Welty et al., 2009 or Muggeo, 2008 in the context of segmented breakpoint models). We decided to explicitly shrink only the last lag coefficient β_L and to limit the previous lag coefficients only by the smoothing part of the penalization and not by further restriction. We explicitly shrink the last lag coefficient β_L towards zero while the remainder of the lag course is estimated without taking assumptions of monotonicity and without the need to manually specify a vector of hard to elicit control parameters. The first lag coefficients can take any smooth form that corresponds to the data.

3.1.3 Estimation of the Lag Curve

The estimation can be performed via maximization of the penalized likelihood $l_{pen}(\boldsymbol{\beta}) = l(\boldsymbol{\beta}) - \frac{1}{2}\boldsymbol{\beta}^T (\xi_d \mathbf{J}_d + \xi_r \mathbf{J}_r)\boldsymbol{\beta}$ (Fahrmeir and Tutz, 2001). The smoothing parameter $\xi_d > 0$ controls the trade-off between data fit and smoothness of the lag curve. ξ_d penalizes large differences in adjacent coefficients; large values of ξ_d result in very smooth estimates of the

lag course. The shrinkage parameter $\xi_r > 0$ penalizes large values in the last lag coefficient $\hat{\beta}_L$. Large values of ξ_r entail strong shrinkage of the last lag coefficient and decrease the estimated values of $\hat{\beta}_L$.

3.2 Flexible Distributed Lag Model (FDL)

In this section, we transfer the penalization concept proposed in Section 3.1 to a spline based approach. In the econometric literature, an established approach for the inclusion of lagged linear covariate effects is the use of Almon polynomials (Almon, 1965). The lag coefficients are assumed to lie on a polynomial of lower degree (see Section 2.4.1). The use of Almon polynomials encourages smooth estimates with a small number of parameters. Nevertheless, the assumption that the lag coefficients lie on a global polynomial is a strong assumption and more flexible shapes of the lag coefficients are hard to detect by means of Almon polynomials. Alternative to polynomials is the use of penalized splines. In Section 2.5, we introduce a lag model based on basis functions. Similar to the penalized lag model (Section 3.1), we add both a smoothing and a shrinkage penalty to the model based on basis coefficients presented in Section 2.5. We indirectly penalize the lag coefficients $\boldsymbol{\beta}$ via penalizing the basis coefficients $\boldsymbol{\gamma}$.

A penalized spline based approach represents an enlargement of the penalized lag model proposed in Section 3.1. The penalized lag model can be considered as penalized B-spline based approach with a B-spline basis of degree 0 with $L+1$ basis functions. We now present a penalized spline based approach, that includes the penalized lag model as a special case and enlarges the freedom of action of the user by the possibility of choosing B-spline bases of higher degree or less basis functions.

3.2.1 Smoothing of the Basis Coefficients

We reminisce about Equation (2.37) in Section 2.5, where we introduce a lag model based on basis functions:

$$g(\boldsymbol{\mu}) = \underbrace{\underset{T \times 1}{\mathbf{X}_{lag}} \underset{(L+1) \times m}{\mathbf{B}}}_{\mathbf{z}} \underset{m \times 1}{\boldsymbol{\gamma}}.$$

We penalize adjacent basis coefficients in order to penalize adjacent lag coefficients and therefore smooth the resulting lag course $\boldsymbol{\beta}$. The smoothing penalty matrix is the $m \times m$ matrix $\mathbf{K}_d = \mathbf{D}_2^T \mathbf{D}_2$, when using differences of second order. Please note, that $\mathbf{D}_2$ is defined as in Equation (2.3), except that the dimension of the matrix is different, usually smaller than in Equation (2.3) (Dimension $(L+1) \times (L+1)$). Figure 3.1 illustrates the effect of different smoothing parameters λ_d on the estimated lag course $\boldsymbol{\beta}$ in a simplified Hochstaufen data set for the model based on basis coefficients. We estimate the lagged rain effect on the number of earthquakes in one arbitrarily chosen depth category. For the illustration of the
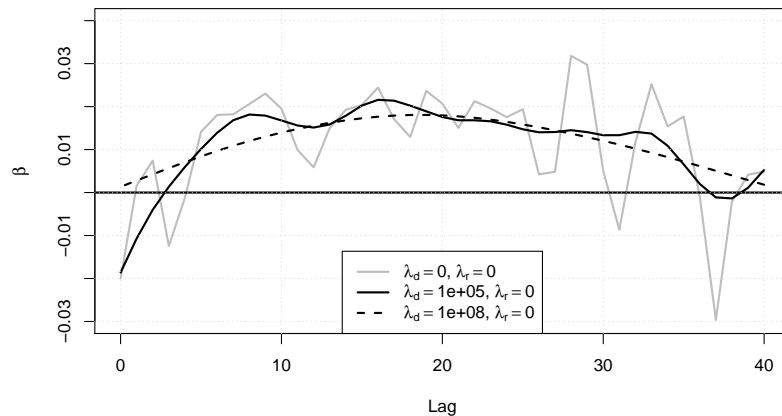
Figure 3.1: Effect of different smoothing parameters λ_d on the estimation of the influence of lagged rain on the number of earthquakes in the Hochstaufen data set.

effect of the smoothing parameter λ_d, we chose the lagged rain effect in depth category 3. The turbulent grey lag course shows the unpenalized course of the lag coefficients modeled via basis functions (see section 2.5). We use a relatively large number of knots and therefore basis functions, that is to say, 30, in order to maximize the flexibility of possible lag courses. The smoothing parameter λ_d then monitors the amount of smoothness of the lag curve. The larger the smoothing parameter λ_d is chosen, the smoother the resulting lag course is. For the manually chosen smoothing parameter $\lambda_d = 10^5$ (solid line), the lag course is clearly smoother than the unrestricted lag course. However, multiple local maxima occur. The first few lag coefficients are estimated with a negative sign which would mean decreased estimated numbers of quakes as an immediate effect of increased rain. Not before a certain lag, increased rain would mean increased estimated numbers of quakes, indicated with a positive sign of the lag coefficients. Just before the last lag coefficients $\hat{\beta}_{40}$, an anew slight slope of the lag course is detected according to the model with a smoothing parameter $\lambda_d = 10^5$. That slope seems to be an artefact and the possible elimination of that slope - if compatible with the data - will be the aim of the ridge penalty of the following subsection 3.2.2. When choosing an even larger smoothing penalty $\lambda_d = 10^6$ (dashed line), the local maxima give way to one global maximum and also the negative slope at the first few coefficients falls victim to the high amount of smoothing.

3.2.2 Shrinkage of the Basis Coefficients

A further desirable property of the distributed lag curve is that the last coefficient β_L is to be close to zero. This assumption seems to be plausible not only from content reasons but as well from formal reasons. With the choice of a certain maximum lag L, one implicitly
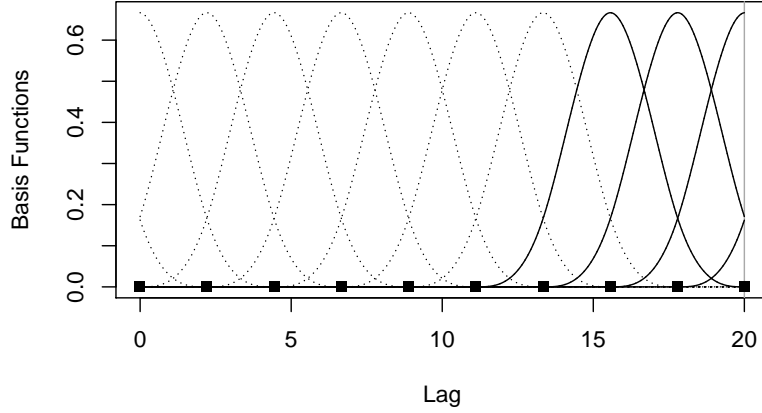
Figure 3.2: Third degree B-spline basis functions in the lag model, $L = 20$, $m = 12$, $K = 10$. Basis functions $\mathbf{B}_9(l), \mathbf{B}_{10}(l), \mathbf{B}_{11}(l)$ and $\mathbf{B}_{12}(l)$ are nonzero at location $l = L = 20$, that is marked by the vertical line. ■ marks the chosen inner knots.

assumes β_{L+1} to be zero, because β_{L+1} is no longer part of the estimation. Accordingly, - if the lag length L is correctly specified - for β_L should hold:

$$\beta_L = \sum_{j=1}^{m} B_j(L)\gamma_j \approx 0. \tag{3.13}$$

At any chosen point $l \in [0, L]$, exactly $(d+1)$ $d-$degree B-spline basis functions are nonzero. So, Equation (3.13) can be written as:

$$\beta_L = \sum_{j=m-d}^{m} B_j(L)\gamma_j \approx 0. \tag{3.14}$$

Figure 3.2 shows $m = 12$ B-splines (degree $d = 3$) $\mathbf{B}_j(l), j = 1 \ldots m$ for $L = 20$ and $K = 10$ equidistant inner knots. The $d + 1 = 4$ basis functions $\mathbf{B}_{m-d}(l), \ldots, \mathbf{B}_m(l) = \mathbf{B}_9(l), \ldots, \mathbf{B}_{12}(l)$ (solid lines) are nonzero at the interesting point $L = 20$. The remaining basis functions (dashed lines) are zero at that point of interest. The objective of shrinking the last lag coefficient β_L towards zero can be pursued by penalizing the last $(d + 1)$ $\boldsymbol{\gamma}$-coefficients via a $m \times m$ shrinkage matrix $\mathbf{K}_r$:

$$
\underbrace{\mathbf{K}_r}_{\text{m x m}} = \begin{pmatrix} 0 & & & & & \\ & \ddots & & & & \\ & & 0 & & & \\ & & & 1 & & \\ & & & & \ddots & \\ & & & & & 1 \end{pmatrix}. \tag{3.15}
$$

Accordingly, $\mathbf{K}_d$ and $\mathbf{K}_r$ form the penalizing matrices in the penalized likelihood maximization. In Figure 3.1, we showed the influence of different manually defined smoothing parameters on the estimated lag course $\hat{\boldsymbol{\beta}}$ in a simplified Hochstaufen data set. We observe, especially for the smoothing parameter $\lambda_d = 10^5$, that the last lag coefficients build a slight, but implausible slope. Geophysicists assume a maximum lag L of 20 to 40 days to be the underlying trigger scope of rain causing earthquakes. The question is, if this artefact can be corrected in a gentle way. We redo the analysis for the influence of lagged rain on the number of earthquakes within a simplified Hochstaufen data set and manually define for the lag courses with the two smoothing parameters $\lambda_d = 10^5$ and $\lambda_d = 10^8$ an additional shrinkage parameter $\lambda_r = 10^6$. Figure 3.3 shows the influence of the shrinkage parameter $\lambda_r = 10^6$ on the back part of the estimation. We see that the lag curve with $\lambda_d = 10^5$ (solid line) generally is responsive to the ridge penalty. A ridge penalty of $\lambda_r = 10^6$ yields a lag curve with $\hat{\beta}_L$ close to zero. For the smoother curve with $\lambda_d = 10^8$, $\hat{\beta}_L$ was already close to zero so that the differences of the dashed lines in Figure 3.1 and 3.3 are not enormous. We see, that a large shrinkage parameter λ_r can decrease the last lag coefficients close to zero. The choice of the shrinkage parameter λ_r will be discussed in Section 3.5. The choice of the smoothing parameter has to happen in that way, that $\hat{\beta}_L$ is to be close to zero if and only if the data is compatible with such a decay. Assuming that, in a data set, the maximum lag length L is shorter than necessary, a flexible ridge penalty should discover that in that case the last lag coefficient $\hat{\beta}_L$ cannot be shrunk close to zero as there is still information on and probably after $\hat{\beta}_L$.

3.3 Impact of Basis Coefficient Penalizing on the Lag Coefficients

Similar to Section 3.1, we indirectly penalize the lag coefficients $\boldsymbol{\beta}$ via penalizing the basis coefficients $\boldsymbol{\gamma}$. Concerning the smoothing, the question is how the penalization of large differences in $\gamma_j - \gamma_{j-1}$ causes a penalization of large differences in $\beta_l - \beta_{l-1}$. The answer lies in the special structure of the matrix $\mathbf{B}$ or rather in the local definition of B-splines. In any location $l \in [0, L]$, exactly $d + 1$ basis functions are nonzero. An exemplary basis function matrix $\mathbf{B}$ with B-splines of second degree $(d = 2)$, $L = 8$ and $m = 5$ basis functions

Figure 3.3: Effect of different shrinkage parameters λ_r on the estimation of the influence of lagged rain on the number of earthquakes in the Hochstaufen data set.

is:

$$
\mathbf{B} = \begin{pmatrix}
B_1(0) & B_2(0) & & & \\
B_1(1) & B_2(1) & B_3(1) & & \\
B_1(2) & B_2(2) & B_3(2) & & \\
& B_2(3) & B_3(3) & B_4(3) & \\
& B_2(4) & B_3(4) & B_4(4) & \\
& B_2(5) & B_3(5) & B_4(5) & \\
& & B_3(6) & B_4(6) & B_5(6) \\
& & B_3(7) & B_4(7) & B_5(7) \\
& & B_3(8) & B_4(8) & B_5(8)
\end{pmatrix}.
\tag{3.16}
$$

Another characteristic of B- Splines is, that every row sum in the band matrix is one. In any location $l \in [0, L]$, we have:

$$
\sum_{j=1}^{m} B_j(l) = 1.
\tag{3.17}
$$

With Equation (3.17), one can show that differences in β_l, β_{l-1} can be attributed to additive combinations of basis coefficient differences in γ_j, γ_{j-1}. We write down a common second

degree B-spline basis lag equation:

$$\boldsymbol{\beta} = \begin{pmatrix} \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & B_j(l-1) & B_{j+1}(l-1) & B_{j+2}(l-1) & 0 & \dots \\ \dots & 0 & B_{j+1}(l) & B_{j+2}(l) & B_{j+3}(l) & \dots \\ \dots & 0 & B_{j+1}(l+1) & B_{j+2}(l+1) & B_{j+3}(l+1) & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \end{pmatrix} \begin{pmatrix} \vdots \\ \gamma_j \\ \gamma_{j+1} \\ \gamma_{j+2} \\ \gamma_{j+3} \\ \vdots \end{pmatrix}$$

and get:

$$\beta_{l-1} = B_j(l-1)\gamma_j + B_{j+1}(l-1)\gamma_{j+1} + B_{j+2}(l-1)\gamma_{j+2} =$$
$$= B_j(l-1)\gamma_j + B_{j+1}(l-1)(\gamma_{j+1} - \gamma_j + \gamma_j) + B_{j+2}(l-1)(\gamma_{j+2} - \gamma_{j+1} + \gamma_{j+1} - \gamma_j + \gamma_j) =$$
$$= \underbrace{[B_j(l-1) + B_{j+1}(l-1) + B_{j+2}(l-1)]}_{=1} \gamma_j$$
$$+ B_{j+1}(l-1)(\gamma_{j+1} - \gamma_j) + B_{j+2}(l-1)((\gamma_{j+2} - \gamma_{j+1}) + (\gamma_{j+1} - \gamma_j)) =$$
$$= \gamma_j + B_{j+1}(l-1)(\gamma_{j+1} - \gamma_j) + B_{j+2}(l-1)((\gamma_{j+2} - \gamma_{j+1}) + (\gamma_{j+1} - \gamma_j))$$

$$\beta_l = B_{j+1}(l)\gamma_{j+1} + B_{j+2}(l)\gamma_{j+2} + B_{j+3}(l)\gamma_{j+3} =$$
$$= B_{j+1}(l)(\gamma_{j+1} - \gamma_j + \gamma_j) + B_{j+2}(l)(\gamma_{j+2} - \gamma_{j+1} + \gamma_{j+1} - \gamma_j + \gamma_j)$$
$$+ B_{j+3}(l)(\gamma_{j+3} - \gamma_{j+2} + \gamma_{j+2} - \gamma_{j+1} + \gamma_{j+1} - \gamma_j + \gamma_j) =$$
$$= \underbrace{[B_{j+1}(l) + B_{j+2}(l) + B_{j+3}(l)]}_{=1} \gamma_j + B_{j+1}(l)(\gamma_{j+1} - \gamma_j)$$
$$+ B_{j+2}(l)((\gamma_{j+2} - \gamma_{j+1}) + (\gamma_{j+1} - \gamma_j)) + B_{j+3}(l)((\gamma_{j+3} - \gamma_{j+2}) + (\gamma_{j+2} - \gamma_{j+1}) + (\gamma_{j+1} - \gamma_j)) =$$
$$= \gamma_j + B_{j+1}(l)(\gamma_{j+1} - \gamma_j) + B_{j+2}(l)((\gamma_{j+2} - \gamma_{j+1}) + (\gamma_{j+1} - \gamma_j))$$
$$+ B_{j+3}(l)((\gamma_{j+3} - \gamma_{j+2}) + (\gamma_{j+2} - \gamma_{j+1}) + (\gamma_{j+1} - \gamma_j))$$

$$\beta_{l+1} = B_{j+1}(l+1)\gamma_{j+1} + B_{j+2}(l+1)\gamma_{j+2} + B_{j+3}(l+1)\gamma_{j+3} = \dots$$
$$= \gamma_j + B_{j+1}(l+1)(\gamma_{j+1} - \gamma_j) + B_{j+2}(l+1)((\gamma_{j+2} - \gamma_{j+1}) + (\gamma_{j+1} - \gamma_j))$$
$$+ B_{j+3}(l+1)((\gamma_{j+3} - \gamma_{j+2}) + (\gamma_{j+2} - \gamma_{j+1}) + (\gamma_{j+1} - \gamma_j))$$

$$(3.18)$$

Differences of lag coefficients can be calculated as:

$$
\begin{aligned}
\beta_l - \beta_{l-1} &= [1 - B_{j+1}(l-1) - B_{j+2}(l-2)] & (\gamma_{j+1} - \gamma_j)+ \\
&\quad + [B_{j+2}(l) + B_{j+3}(l) - B_{j+2}(l-1)] & (\gamma_{j+2} - \gamma_{j+1}) \\
&\quad + B_{j+3}(l) & (\gamma_{j+3} - \gamma_{j+2}) \\
&= \sum_{a=1}^{3} w_a(\gamma_{j+a} - \gamma_{j-1+a}) \\
\beta_{l+1} - \beta_l &= [B_{j+2}(l+1) + B_{j+3}(l+1) - B_{j+2}(l) - B_{j+3}(l)] & (\gamma_{j+2} - \gamma_{j+1}) \\
&\quad + B_{j+3}(l+1) - B_{j+3}(l) & (\gamma_{j+3} - \gamma_{j+2}) \\
&= \sum_{a=1}^{2} w_a(\gamma_{j+1+a} - \gamma_{j+a})
\end{aligned}
\tag{3.19}
$$

Figure 3.4 shows for one arbitrarily chosen simulated lag course unpenalized basis coefficients courses (top left) and resulting unpenalized lag coefficient courses (top right). If we smooth basis coefficients, resulting lag coefficients take a smooth lag course.

3.4 Estimation

The estimation of the basis coefficients $\boldsymbol{\gamma}$ can be computed via maximization of the penalized likelihood of the estimation problem $g(\boldsymbol{\mu}) = \mathbf{Z}\boldsymbol{\gamma}$. The penalized likelihood with $l(\boldsymbol{\gamma}) = \sum_{t=1}^{T} \left[y_t(\mathbf{z}_t^T \boldsymbol{\gamma}) - \exp(\mathbf{z}_t^T \boldsymbol{\gamma}) \right]$ with $\mathbf{z}_t^T$ the t-th row of $\mathbf{Z}$ for a Poisson distributed response is:

$$
l_{pen}(\boldsymbol{\gamma}) = l(\boldsymbol{\gamma}) - \frac{1}{2}\boldsymbol{\gamma}^T \left(\lambda_d \mathbf{K}_d + \lambda_r \mathbf{K}_r \right) \boldsymbol{\gamma}.
\tag{3.20}
$$

The smoothing parameter λ_d penalizes large differences in adjacent basis coefficients. The shrinkage parameter λ_r penalizes the last basis coefficients. The estimation of the unknown $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$ coefficients respectively can be done with the function *gam* in the R-package (R Development Core Team, 2013) *mgcv* (see the call in Appendix A) via penalized iteratively reweighted least squares (P-IRLS, Marx and Eilers, 1998; Wood, 2006). In the FDL context (see Equation (2.37)) we have:

$$
g(\boldsymbol{\mu}) = \log(\boldsymbol{\mu}) = \mathbf{Z}\boldsymbol{\gamma}.
\tag{3.21}
$$

In the case of Poisson regression with canonical link function, penalized score function and penalized Fisher information can be written as:

$$
\mathbf{s}_p(\hat{\boldsymbol{\gamma}}) = \mathbf{Z}^T(\mathbf{y} - \boldsymbol{\mu}) - (\lambda_d \mathbf{K}_d + \lambda_r \mathbf{K}_r)\boldsymbol{\gamma}
\tag{3.22}
$$

$$
F_p(\hat{\boldsymbol{\gamma}}) = \mathbf{Z}^T \mathbf{W} \mathbf{Z} + \lambda_d \mathbf{K}_d + \lambda_r \mathbf{K}_r
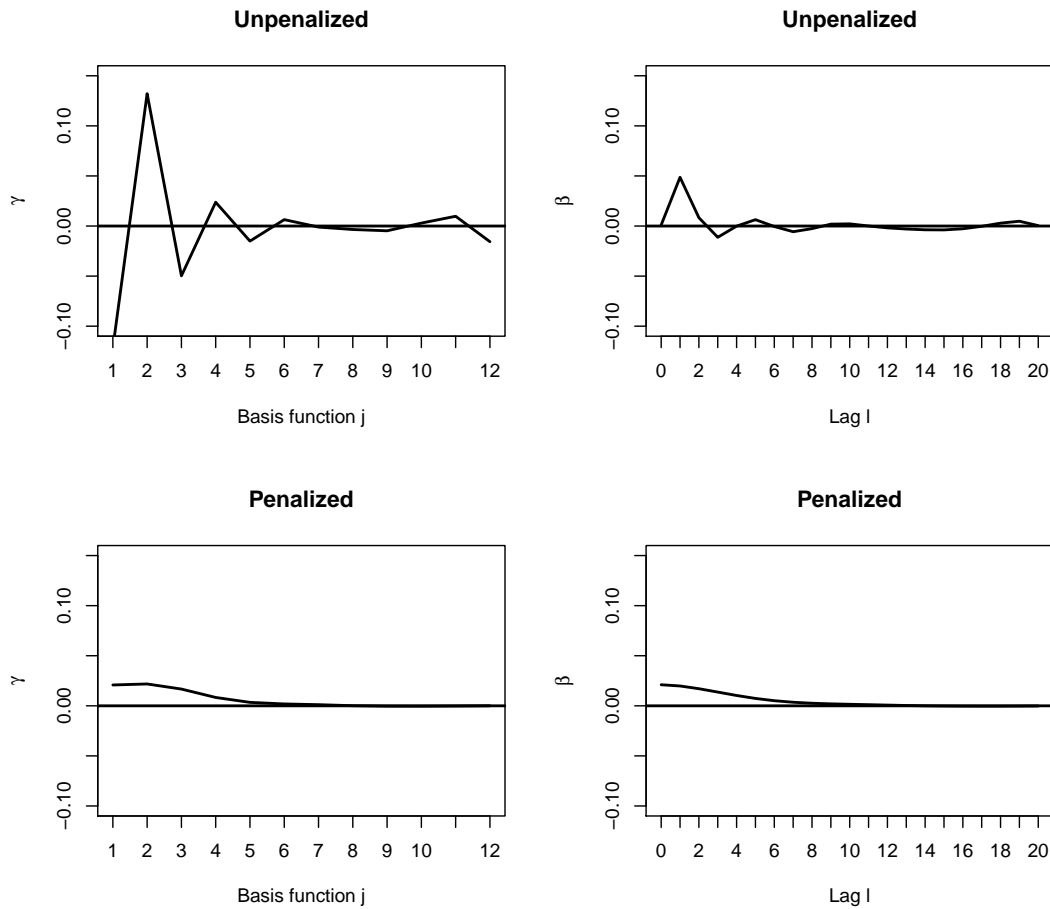\tag{3.23}
$$

Figure 3.4: Effect of the penalization of adjacent $\boldsymbol{\gamma}$- basis coefficients on the lag coeffcients $\boldsymbol{\beta}$. Figures at the top show unpenalized $\boldsymbol{\gamma}$- (left) and $\boldsymbol{\beta}$- (right) coefficients. At the bottom, $\boldsymbol{\gamma}$ is penalized with $\lambda_d = \lambda_r = 10^8$. (left). The effect on $\boldsymbol{\beta}$ is shown on the right side.

with $\mathbf{W}$ a diagonal matrix with entries $w_t = \mu_t$. Using penalized Fisher scoring with the definition of a stop criterion, we iterate:

$$\hat{\boldsymbol{\gamma}}^{(i+1)} = \hat{\boldsymbol{\gamma}}^{(i)} + \mathbf{F}_p^{-1}(\hat{\boldsymbol{\gamma}}^{(i)})\mathbf{s}_p(\hat{\boldsymbol{\gamma}}^{(i)}). \tag{3.24}$$

Iterations are stopped if some stop criterion is reached, for example if:

$$\frac{\|\hat{\boldsymbol{\gamma}}^{(i+1)} - \hat{\boldsymbol{\gamma}}^{(i)}\|}{\|\hat{\boldsymbol{\gamma}}^{(i)}\|} \leq \epsilon \tag{3.25}$$

for a small prechosen number ϵ.

3.5   Choice of the Smoothing Parameters

The smoothing parameter estimation is automatic in *gam* and can be performed by minimization of prediction errors such as the Generalized Cross Validation Criterion (GCV, Hastie and Tibshirani, 1999) and the Un-Biased Risk Estimator (UBRE) criterion or by restricted maximum likelihood (REML, Wahba, 1985) where the smooth functions are treated as random effects with λ as variance parameters.

For known scale parameter ϕ, the UBRE (Craven and Wahba, 1979), a rescaled version of Akaikes Information Criterion (AIC, Fahrmeir and Tutz, 2001) is used. It can be written as:

$$\frac{D}{T} + \frac{2\phi df}{T} - \phi \tag{3.26}$$

where T is the number of observations and D the deviance. *df* are the effective degrees of freedom of the model. Note that the scale parameter is 1 for binary and Poisson data.

Wood (2011) summarizes that REML or ML based smoothness selection is often preferable to prediction error based methods in practice, see also Reiss and Ogden (2009) for a similar conclusion. Wood (2011) proposes a Laplace approximation for an approximate REML criterion for smoothing parameter estimation. In each of the nested iterations, the selection criterion depends on $\boldsymbol{\gamma}$ only via $\hat{\boldsymbol{\gamma}}_{\boldsymbol{\lambda}}$. Accordingly, each iteratively proposed smoothing parameter needs a full P-IRLS step to find $\hat{\boldsymbol{\gamma}}_{\boldsymbol{\lambda}}$. Newton-Raphson iteration instead of Fisher Scoring is required. In the case of a canonical link function as presented in this manuscript, Newton-Raphson iterations and Fisher scoring coincide. Published results in Wood (2011) indicate that this yields a computationally stable algorithm.

3.6   Simulation: Performance of Flexible Distributed Lags

3.6.1   Design

We perform a simulation study to evaluate how well known courses of lag coefficients $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_L)$ can be recovered by means of the B-spline based flexible distributed

lag approach. To do so, we simulate several possible β courses. Real Bavarian daily temperature data measured at Mount Hochstaufen act as covariate x for the simulated data. We use a fairly long time series with $T = 5076$. The high autocorrelation of the temperature (0.83 for lag 1) means that unregularized estimates are unstable and imprecise. For a weakly autocorrelated covariate, the reconstruction of a smooth lag course is less problematic even without smoothing. 1000 Poisson distributed response variable vectors are constructed via $\boldsymbol{\mu} = \exp(\mathbf{X}_{lag}\boldsymbol{\beta})$ and $\mathbf{y} \sim Po(\boldsymbol{\mu})$. Figure 3.5 shows the used covariate *temperature*, one specific bimodal simulated lag course and one resulting Poisson distributed response variable.

3.6.2 MSE Comparison Almon-FDL-PLM

We compare the estimated lag courses to the simulated lag courses by means of root mean squared errors (RMSE, 1000 runs) with

$$RMSE = \sqrt{\frac{1}{1000}\sum_{i=1}^{1000}\frac{1}{L+1}\sum_{l=0}^{L}(\hat{\beta}_{li} - \beta_l)^2}. \qquad (3.27)$$

Table 3.1 gives a general overview of the performance of flexible distributed lags (FDL) in comparison to FDL without additional ridge penalty, the penalized lag model (PLM, Section 3.1) and 4th degree Almon polynomials (Almon, Equation (2.11)). For Courses 1 to 6, we use a simulated lag length (L) of 20 days, 15 knots, cubic B-splines and a second order difference penalty for the basis coefficients. For each setting and each method, we calculate the relative RMSE compared to the RMSE of the FDL (3rd column). Minima in each row are underlined. They represent the best estimation method for the respective simulated lag course. Except for Course 2, where PLM performs slightly better than FDL, all other courses are best represented by the flexible distributed lag approach. Lag Course 6 represents a constant effect of the covariate on the reponse variable. Both, FDL and FDL without additional ridge penalty perform quite well for lag Course 6. The ridge penalty acts flexible enough to discover, that the last lag coefficient cannot be shrunk towards zero and accordingly keeps the smoothing parameters small. The models with and without ridge penalty achieve the same performance quality even for the extreme case, where the assumption of a small last lag coefficient is wrong. For lag Courses 2 to 5, FDL without additional ridge penalty cannot reach the performance of FDL with ridge penalty. PLM without basis functions mostly performs worse than the FDL. The averaged degrees of freedom for the models compared in Table 3.1 are given in Table 3.2.

Figure 3.6 gives a more detailed overview of the performance of the single replications in the simulation study. It shows, that the Almon method performs worse on average (except for Course 1). However, the variance of the RMSE within the single replications can be larger within the FDL method (Courses 3 to 5).

Figure 3.7 (right side) shows exemplarily for Course 4 the estimated lag course (mean over 1000 runs) for FDL in comparison to the true underlying lag course (grey). Almon
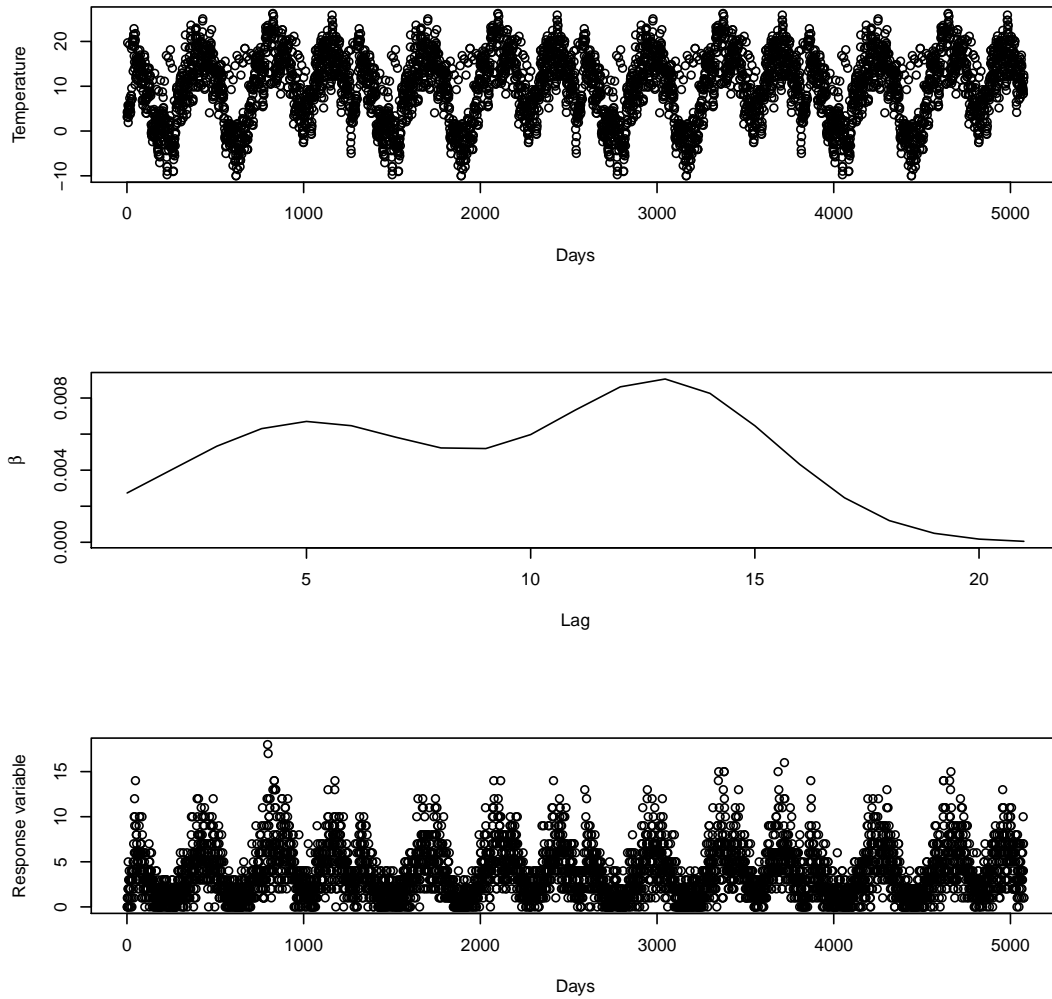
Figure 3.5: Covariate (temperature, top), one specific lag course (course 4, center) and resulting Poisson distributed response variable.
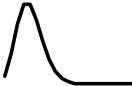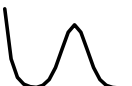
$\beta(l)$	Almon	FDL	FDL : no ridge	PLM
Course 1	1.02	<u>1.00</u>	1.04	1.07
Course 2	1.28	1.00	1.10	<u>0.99</u>
Course 3	2.33	<u>1.00</u>	1.13	1.15
Course 4	1.87	<u>1.00</u>	1.17	1.14
Course 5	1.39	<u>1.00</u>	1.09	1.11
Course 6	2.19	<u>1.00</u>	1.02	1.06

Table 3.1: $L = 20$: Relative RMSE (1000 runs) for several simulated lag courses (1st column) for four different estimation techniques: Almon polynomials (2nd column), flexible distributed lag model (FDL, 3rd column), flexible distributed lag model without additional ridge penalty (no ridge, 4th column) and penalized lag model without basis functions (PLM, 5th column). For each setting and each method, we calculate the relative RMSE compared to the RMSE of the FDL (3rd column). Minima in each row are underlined.

	Almon	FDL	FDL : no ridge	PLM
Course 1	5.00	3.89	4.99	6.18
Course 2	5.00	3.49	4.09	4.45
Course 3	5.00	6.35	8.19	9.92
Course 4	5.00	5.21	5.70	6.99
Course 5	5.00	5.24	6.97	8.58
Course 6	5.00	2.22	2.28	3.31

Table 3.2: Averaged (1000 runs) degrees of freedom for Courses 1 to 6.
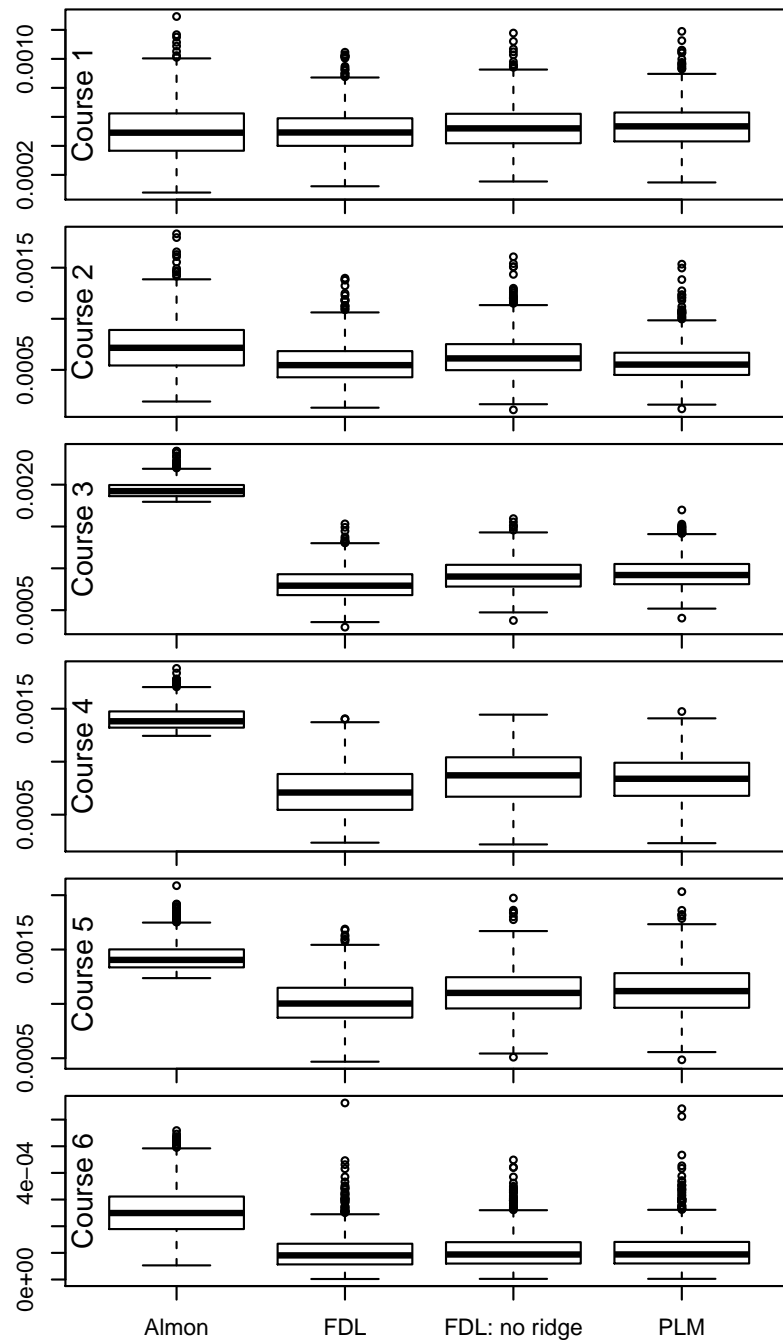
Figure 3.6: Boxplot of the distribution of the RMSE $= \sqrt{\frac{1}{L+1} \sum_{l=0}^{L} (\hat{\beta}_l - \beta_l)^2}$ for 1000 replications for the simulated lag courses for four different estimation techniques: Almon polynomials (2nd column), flexible distributed lag model (FDL, 3rd column), flexible distributed lag model without additional ridge penalty (no ridge, 4th column) and penalized lag model without basis functions (PLM, 5th column).
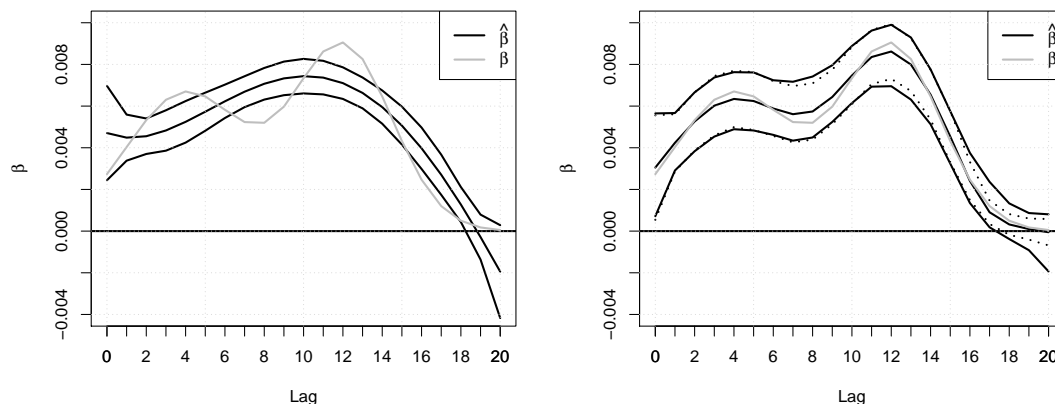
Figure 3.7: Simulation (Mean over 1000 runs): Left side: Estimation and pointwise confidence intervals of Course 4 via 4th degree Almon polynomials versus true (grey) $\boldsymbol{\beta}$ - course. Right side: Estimation, pointwise Bayesian confidence intervals (dotted lines) and bootstrap intervals of Course 4 via FDL versus true (grey) β - course.

polynomials (4th degree, left side) cannot replicate the bimodal shape of the underlying true lag course, that FDL can find quite well.

Figure 3.8 shows the estimation of Course 4 via FDL without additional ridge penalty. We already saw in Table 3.1, that the MSE of FDL without additional ridge penalty is 1.17 times larger than the MSE of our smoothed and shrinked FDL approach. We see that $\hat{\beta}_L$ is slightly underestimated in the setting without additional shrinkage. As already seen in Figure 3.7, the ridge penalty can correct that slight underestimation quite well.

3.6.3 MSE Comparison REML-UBRE

In Section 3.5, we already mentioned that Wood (2011) proposes a Laplace approximation to obtain a computationally stable REML. Generally he summarizes that REML or ML based smoothness selection is often preferable to GCV or AIC in practice, see also Reiss and Ogden (2009) for a similar conclusion. Within the framework of *mgcv*, we were easily able to compare smoothing parameter estimation via REML to smoothing parameter estimation via UBRE, a rescaled version of the AIC, and we also became aware of the superiority of REML in comparison to UBRE as Table 3.3 shows. It compares the RMSEs of UBRE and REML, related to the RMSE of REML for FDL. For each course, the RMSE of REML estimates is smaller than the RMSE of UBRE estimates. Accordingly, we suggest to manage smoothing parameter estimation via REML within the framework of FDL.
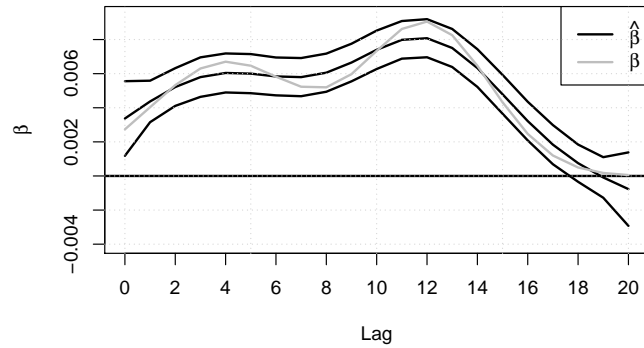
Figure 3.8: Simulation (Mean over 1000 runs): Estimation and pointwise confidence intervals of Course 4 via FDL without additional ridge penalty versus true (grey) $\boldsymbol{\beta}$ - course.

	UBRE	REML
Course 1	1.18	1.00
Course 2	1.21	1.00
Course 3	1.16	1.00
Course 4	1.03	1.00
Course 5	1.10	1.00
Course 6	2.12	1.00

Table 3.3: RMSE (1000 runs) for several simulated lag courses for UBRE and REML, related to the RMSE of REML.

3.6.4 Coverage

In addition to point estimates one wants to adequately describe the uncertainty in the estimates by confidence intervals. In our case, approximate pointwise confidence intervals may be constructed. (Empirical) Bayesian confidence intervals (Marra and Wood, 2012) are calculated in *mgcv*. We address the question how often the true parameters $\beta_l, l = 0, \ldots, L$ lie in their Bayesian 95% confidence intervals to judge the quality of the estimations via FDL. Please note, that we sometimes call the Bayesian intervals "classical" intervals in order to delimit them from the bootstrap intervals, that will be used in the following. Especially if the smoothing parameters are very large, we observe very small confidence intervals that lead to lower than nominal coverages. The upper plot of Figure 3.9 shows for lag Course 4 confidence intervals for the last lag parameter $\hat{\beta}_L = \hat{\beta}_{20}$ for the first 15 runs of our simulation study. On the bottom of Figure 3.9 we see the respective logarithmized shrinkage parameters $\hat{\lambda}_r$ for that 15 runs. Black color indicates that the true parameter β_{20} does not lie in the estimated confidence interval for the respective replication. This is exactly then the case if the shrinkage parameter $\hat{\lambda}_r$ is very large. In run 7, e.g., we observe a comparatively large shrinkage parameter $\hat{\lambda}_r$. The respective confidence interval is very small due to the disregard of the uncertainty of the estimated smoothing parameter. Therefore, the true parameter β_{20} does not lie within the classical confidence interval.

Approximate pointwise confidence intervals generally provide an undercoverage in certain areas of the flexible distributed lag courses due to large penalization parameters. Figure 3.10 exemplarily shows for lag Course 4 the proportion of coverage over 1000 runs for all lag coefficients for the classical confidence intervals (grey color). We address two critical regions of lags where the coverages are lower than nominal coverages of 95%: First, the regions around the two maxima of the demanding bimodal lag course that are the costs for the smooth estimation. Here we reach coverages of 80% - a result of the challenging bimodal shape of the lag course. Second, the last four lag coefficients reach only coverages of about 50% due to very small confidence intervals for the respective parameters.

A remedy for that observed undercoverage can be found in parametric bootstrap (Efron and Tibshirani, 1993). We draw 400 replicates from the estimated distribution of the response and determine bootstrap confidence intervals via empirical quantiles of the estimates on this artificial data. The number of replicates (400) is comparatively small, but for 1000 response variables, we already have to calculate $400 * 1000$ *gam* models. This computationally expensive parametric bootstrap procedure yields larger confidence intervals (see Figure 3.7). These bootstrap confidence intervals can solve the second critical coverage region and can raise the coverage proportion of the last four lag coefficients, so that at least the mean coverage properties approximate the nominal 95% limit. Even for the small number of 400 Bootstrap replicates, we obtain acceptable coverage properties.

3.6.5 Choice of Knots and Degree of B-Splines

We propose to use cubic B-splines and a large number of knots (Eilers and Marx, 1996) e.g., $\approx 0.75 L$. Subsequent roughness is balanced by the penalization and a slight reduction
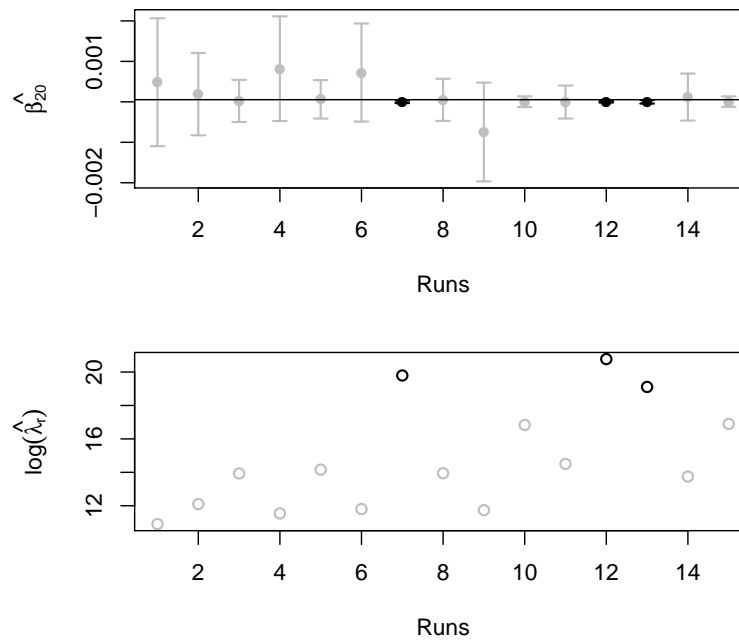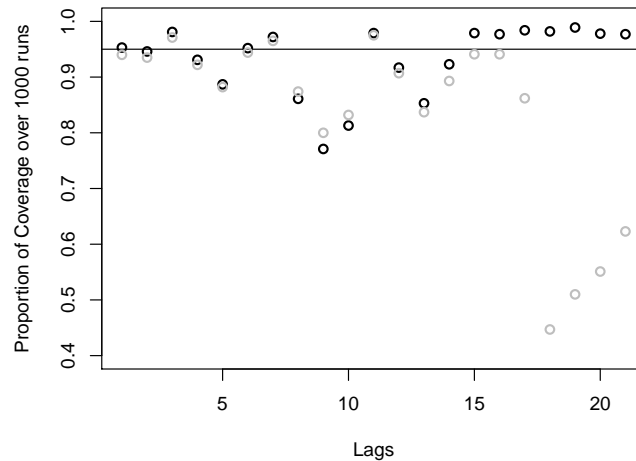
Figure 3.9: Simulation : Top: Estimation via FDL and pointwise classical 95% confidence intervals of the last lag coefficient $\hat{\beta}_{20}$ of Course 4 for the first 15 replications of the response variables. Bottom: Logarithmized estimated shrinkage parameters $\log(\hat{\lambda}_r)$ for the first 15 replications. Black color indicates that the true parameter β_{20} does not lie in the estimated confidence interval for the respective replication.

Figure 3.10: Simulation (1000 runs): Proportion of coverage over 1000 runs for the lag coefficients of Course 4 for classical pointwise confidence intervals (grey) and for bootstrap confidence intervals of the estimation via FDL.

of parameters is given as well.

3.6.6 Order of the Penalty

We use second order differences, as the simulations show better performance in comparison to lower or higher differences.

3.6.7 FDL as a Monitor of the Maximal Lag Length: A Detailed Look on the Ridge Penalty

The choice of the maximum lag L generally is to be justified based on domain knowledge. In case of doubt, we suggest the use of large lags L since the potentially superfluous lag coefficients for higher lags are shrunk towards zero and effectively drop out of the model if a shorter lag window than the originally specified one is compatible with the data. In our simulations, FDL is rather robust against the misspecification of L, as Figure 3.11 shows. On the left side, the effect of an excessively large lag choice is shown. Course 4* is estimated via FDL with a maximum lag L that is twice as large as necessary ($L = 40$ instead of $L = 20$). The number of knots is doubled from 15 to 30. One can see that the superfluous lag coefficients from 21 to 40 are correctly estimated close to zero. Only maxima and minima are slightly oversmoothed, but the essential features of the lag coefficient course are preserved well. Table 3.4 shows the RMSE comparison of the four estimation methods for Courses 1* until 5*. These courses are identical to the Courses 1 until 5 within the
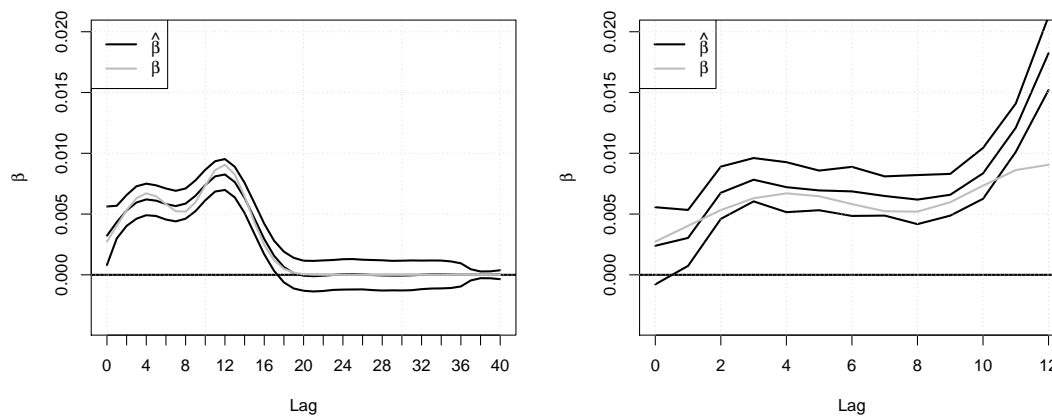
Figure 3.11: Simulation (Mean over 1000 runs): Left side: Estimation and classical confidence intervals of Course 4* via FDL with $L = 40$ versus true $\boldsymbol{\beta}$ (grey). Right side: Estimation and classical confidence intervals of Course 4 via FDL with $L = 12$ versus true β (grey).

first $L + 1 = 20 + 1$ lags and close to the zero line within further 20 superfluous coefficients (see Figure 3.11). The courses are estimated by the four methods with a lag length of $L = 40$. Comparing Table 3.1 and Table 3.4, we see that the performance of the Almon method worsens in comparison to the FDL method within an oversized lag choice (except for Course 1). While for Course 3, e.g., the RMSE of the Almon method is about two times higher than for the FDL (see Table 3.1), we have about 4 times higher RMSE for the Almon method in comparison to the FDL method in Course 3* (see Table 3.4).

Our simulation study indicates that the ridge penalty only takes effect if the data is compatible with the decay of $\hat{\beta}_L$. Figure 3.11 (right side) shows for Course 4 the estimation via FDL with a maximum lag $L = 12$. The FDL approach is flexible enough to discover that there is still information in $L = 12$ and keeps the shrinkage parameter λ_r small. If $\hat{\beta}_L$ is not shrunk to zero in spite of the ridge penalty, the user learns that the chosen L is smaller than necessary in comparison to the underlying true maximum lag. In the case of correlated data, $\hat{\beta}_L$ is overestimated in order to compensate the ignored coefficients for lags 13 to 20.

To compare the performance of FDL under lag length misspecification to Almon, PLM and FDL without additional ridge penalty, we refit Courses 1 to 5 with lag length $L = 12$ and 9 knots. Table 3.5 shows the RMSEs related to the RMSEs of the FDL. Except for Course 2 where PLM performs slightly better, FDL obtains better RMSEs. The additional ridge penalty of FDL prevents an even larger overestimation of β_{12} and FDL therefore performs even better than FDL without additional ridge penalty. Almon polynomials display (as well as FDL) the same overestimation of the last lag coefficient β_{12} due to the disregard of the lags 13 to 20, as Figure 3.12 shows.
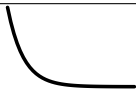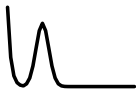
$\beta(l)$	Almon	FDL	FDL : no ridge	PLM
Course 1*	<u>0.86</u>	1.00	1.06	1.08
Course 2*	1.29	1.00	1.01	<u>0.94</u>
Course 3*	4.26	<u>1.00</u>	1.05	1.05
Course 4*	1.63	1.00	1.00	<u>0.96</u>
Course 5*	2.86	<u>1.00</u>	1.04	1.06

Table 3.4: $L = 40$: Relative RMSE (1000 runs) for several simulated lag courses (1st column) for four different estimation techniques: Almon polynomials (2nd column), flexible distributed lag model (FDL, 3rd column), flexible distributed lag model without additional ridge penalty (no ridge, 4th column) and penalized lag model without basis functions (PLM, 5th column). RMSEs are represented relating to the RMSEs of the proposed flexible distributed lag model (FDL, 3rd column). Minima in each row are underlined.

	Almon	FDL	FDL:no ridge	PLM
Course 1	1.33	1.00	1.08	1.10
Course 2	1.44	1.00	1.05	0.99
Course 3	1.80	1.00	1.09	1.11
Course 4	1.13	1.00	1.03	1.07
Course 5	1.05	1.00	1.04	1.06

Table 3.5: RMSE (1000 runs) for several simulated lag courses for the four estimation methods Almon, FDL, FDL: no ridge and PLM with $L = 12$, related to the MSE of FDL.
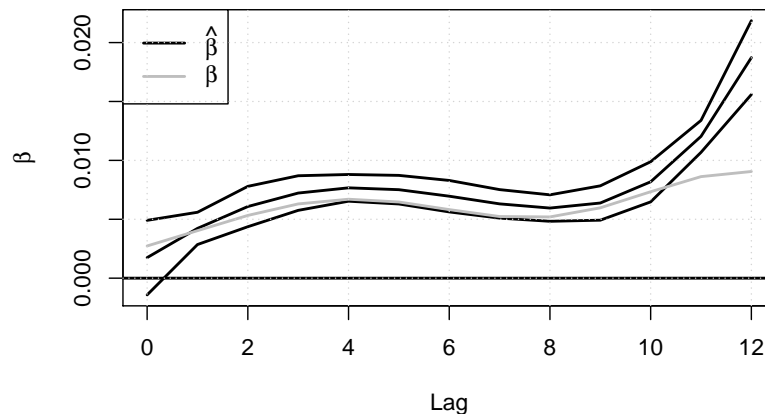
Figure 3.12: Simulation (Mean over 1000 runs): Estimation and confidence intervals of Course 4 ($L = 12$) via Almon polynomials of 4th degree.

Ridge Penalty vs. Constraint

In previous considerations, we thought about constraining the last lag coefficient $\hat{\beta}_L$ to zero in order to force the desired decreasing of the lag coefficients. This constraint can be implemented technically easy in *mgcv*. We rejected the idea, as we consider the operation to be too restrictive. In case the optimal lag length L cannot be defined exactly by domain knowledge or is chosen under wrong assumptions, the setting of the last coefficient $\hat{\beta}_L$ to zero is a mistake that can never be detected in a real data situation. For illustration, we again stress lag Course 4 and estimate the course with a too short lag length and with the help of a hard constraint instead of our data driven and therefore flexible ridge penalty. Figure 3.13 shows estimation and classical confidence intervals for the described scenario in comparison to the true underlying lag course. In addition to the miss-specification of $\hat{\beta}_{12}$ we observe the previously described overestimation of the preceeding coefficients acting as a counterbalance to the ignored but important last coefficients. This overestimation is particularly observed within the context of highly correlated covariates. In the temperature covariate used for the simulation tool, the correlation between the columns of $\mathbf{X}_{lag}$ is up to 0.83.

3.7   Example: Hochstaufen Data

We analyse a data set from Mount Hochstaufen, SE Bavaria of the year 2009. Similar to Section 2.4.1, we examine the dependency of the daily number of detected and located micro earthquakes measured in different depth categories on rain fall and in particular on lagged rainfall (see Svejdar et al., 2011, for geophysical details). Earthquakes with a magnitude
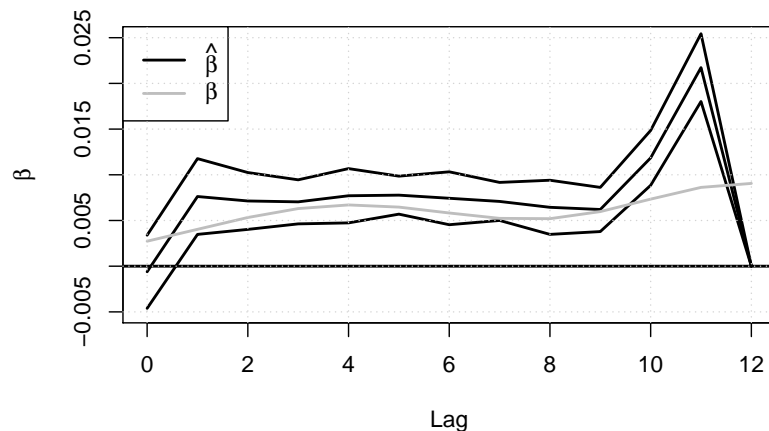
Figure 3.13: Simulation (Mean over 1000 runs): Right side: Estimation and classical confidence intervals of Course 4 via FDL with constraint at $L = 12$ versus true β (grey).

Depth Category h	$h = 1$	$h = 2$	$h = 3$	$h = 4$	$h = 5$	$h = 6$
Depth in km	[free surface, $-0.5[$	$[-0.5, 0.5[$	$[0.5, 1.5[$	$[1.5, 2.5[$	$[2.5, 3.5[$	$[3.5, \infty[$
Number of quakes	34	76	81	49	12	2

Table 3.6: Depth categories and numbers of quakes at Mt. Hochstaufen in 2009.

lower than the magnitude of completeness Mc $= 0.3$ are not part of this analysis. Table 3.6 shows the classification of the depth categories and the resulting numbers of quakes. The interval size reflects the uncertainty in depth estimation during the location procedure.

Figure 3.14 shows the corresponding amounts of rain fall on the 365 observed days. In the middle of July (around day 200), we observe an increased number of quakes. We observe increased rain fall several days before that period. We are particularly interested in the specific shape of the lagged rain effect using the proposed flexible distributed lag approach. In Section 3.2, we proposed a general approach for the inclusion of lagged linear covariates into the generalized linear model context. Nevertheless, our specific application requires an additional layer of complexity. Under the assumption that the rain effects in neighboring depth categories are more similar to each other than those in more distant depth categories, we define a Markov random field (Rue and Held, 2005) over the depth categories with precision matrix P given by the adjacency matrix of the depth categories with a first-order neighborhood structure. We fit the interaction of depth category and an FDL term for rain in the form of a tensor product of the Markov random field term with the FDL term to incorporate different, but similar, lagged rain effects for the different depth
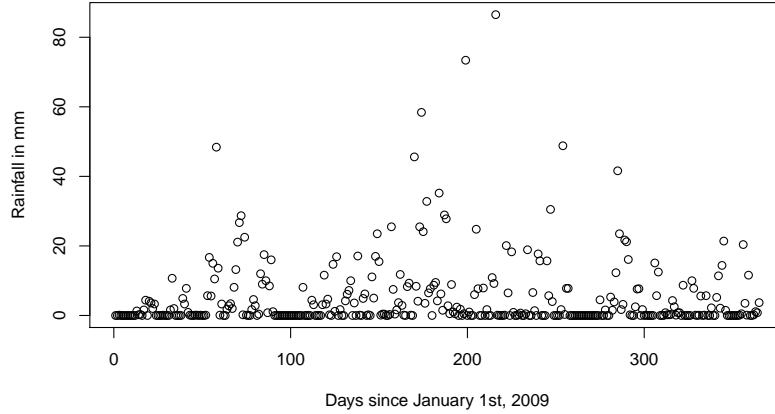
Figure 3.14: Rainfall in mm on 365 days at Mt. Hochstaufen.

categories. Under the assumption that the counts of earthquakes y_{ht} for depth categories h and days t are independently Poisson distributed, given the effects of rain (x_t) and depth, we fit the following model:

$$y_{ht} \sim Po(\mu_{ht})$$

$$\log(\mu_{ht}) = \alpha_0 + f(t) + \delta_h + \sum_{l=0}^{40} \beta_l x_{t-l} + \sum_{l=0}^{40} \nu_{h,l} x_{t-l}$$

$$\boldsymbol{\delta} \sim N_6(\mathbf{0}, \tau^{-1}\mathbf{P}_d^-)$$

where α_0 is a global intercept, $f(t)$ is a nonlinear time trend to capture effects of unobserved covariates, δ_h is the effect of depth category h, β_l are the global FDL effects and $\nu_{h,l}$ are the coefficients for the depth-FDL interaction in depth h for lag l. We use ten basis functions for the FDL term and four basis functions for each of the FDL-depth interaction effects. Since $\boldsymbol{\nu}_h = (\nu_{h,1}, \ldots, \nu_{h,L})^T = \mathbf{B}_h \boldsymbol{\gamma}_h$, where $\mathbf{B}_h$ and $\boldsymbol{\gamma}_h$ are defined as in Equation (2.37), the mathematical equivalence between quadratic penalties and Gaussian distributional assumptions on the coefficients allows us to re-write the penalty for the interaction as a distributional assumption on the vector of interaction coefficients $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1^T, \ldots, \boldsymbol{\gamma}_6^T)$ as $\boldsymbol{\gamma} \sim N_{24}(\mathbf{0}, \mathbf{S}(\tilde{\tau}, \tilde{\lambda}_d)^-)$. The penalty for the interaction is given by $\mathbf{S}(\tilde{\tau}, \tilde{\lambda}_d) = \tilde{\lambda}_d(\mathbf{I}_6 \otimes \tilde{\mathbf{K}}_d) + \tilde{\tau}(\mathbf{P}_d \otimes \mathbf{I}_4)$, where $\tilde{\mathbf{K}}_d$ is the 4×4 difference penalty matrix for the coefficients in each depth category. This sum-of-kronecker-products construction of the joint penalty (Currie et al., 2004; Wood, 2006) yields a penalty that enforces similarity of the lagged rain effects across neighboring depth categories controlled by $\tilde{\tau}$ and smoothness of the lagged rain effect within each category controlled by $\tilde{\lambda}_d$. To build some intuition for this, consider that $\mathbf{I}_6 \otimes \tilde{\mathbf{K}}_d$ yields a penalty matrix that simply repeats the difference penalty 6 times so that a difference penalty applies to each $\boldsymbol{\gamma}_h, h = 1, \ldots, 6$, yielding a smooth lag course within each
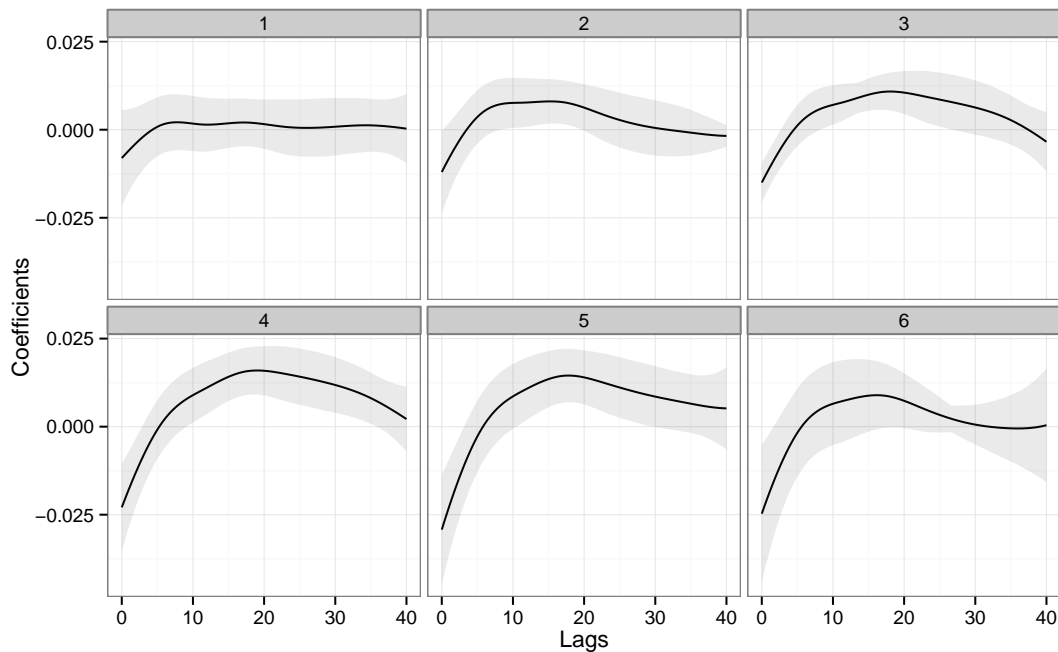
Figure 3.15: Estimated lag courses for rain, $\beta_l + \nu_{h,l}$ (main effects + interaction terms) in the six depth categories h.

depth category. In addition, $\mathbf{P}_d \otimes \mathbf{I}_4$ repeats the penalty matrix for the Markov random field so that it applies to the vectors of first spline coefficients $(\gamma_{1,1}, \ldots, \gamma_{6,1})^T$, second spline coefficients $(\gamma_{1,2}, \ldots, \gamma_{6,2})^T$, and so on, of the depth categories, encouraging similar shapes of these smooth lag courses in neighboring depths. The estimated time trend is shown in Figure 3.16. The wide confidence intervals due to the lack of information at the beginning of the year indicate that the strong minimum should not be (over-) interpreted. All terms in the Poisson model are significantly different from zero (p-values < 0.0001, see Wood, 2013b, for p-values of estimates of functions in generalized additive models). Figure 3.15 shows the estimated lag coefficients with approximate pointwise 95% confidence intervals for the depth categories. As intuitively expected, the maximum rain effects show an increasing lag time with increasing depth, supporting the (possibly) diffusive character of fluid penetration into the subsurface. In depth categories 1 and 6, we cannot detect significant rain lag coefficients $\hat{\beta}_l$. While the topography above sea level (depth category 1) may be very much influenced by the 3D structure, depth category 6 is sparsely populated with only two quakes. The model fit is fairly parsimonious, with about 4 effective degrees of freedom (edf) for each of the main effects of the lagged rainfall and the depth categories, and about 13 edf for the interaction of lagged rain fall and depth.

Figure 3.17 shows the estimated coefficients of the depth categories from depth category 1 (top of the figure) to depth category 6 (bottom).

Figure 3.18 shows the distribution of the quakes in the six depth categories and the
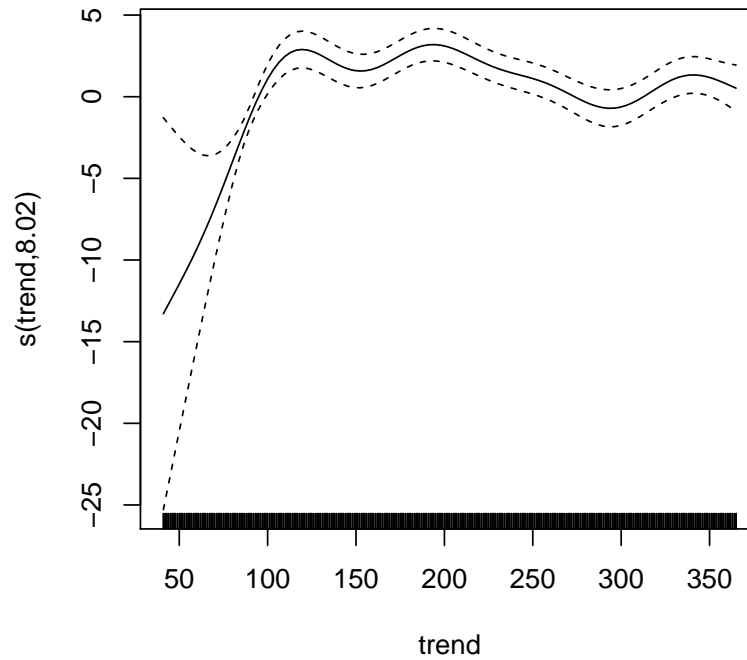
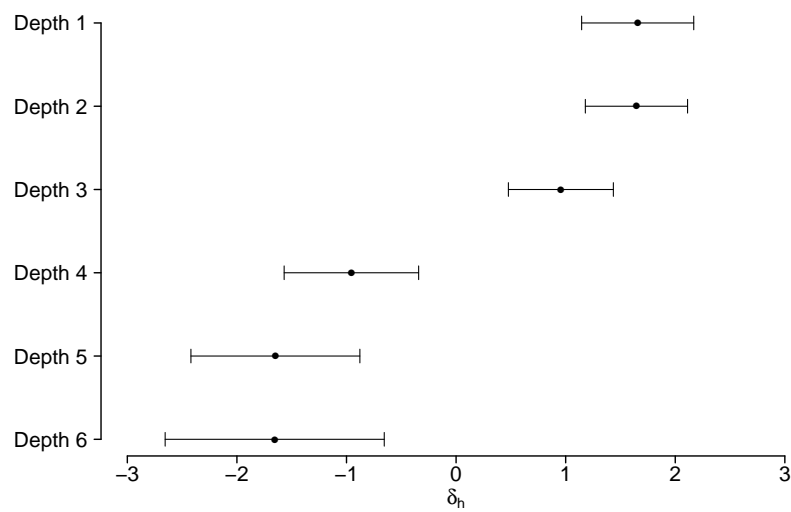Figure 3.16: Nonlinear time trend $f(t)$ in the Hochstaufen data.

Figure 3.17: Effects δ_h and 95% confidence intervals for the depth categories h from depth category 1 (top) to depth category six (bottom).
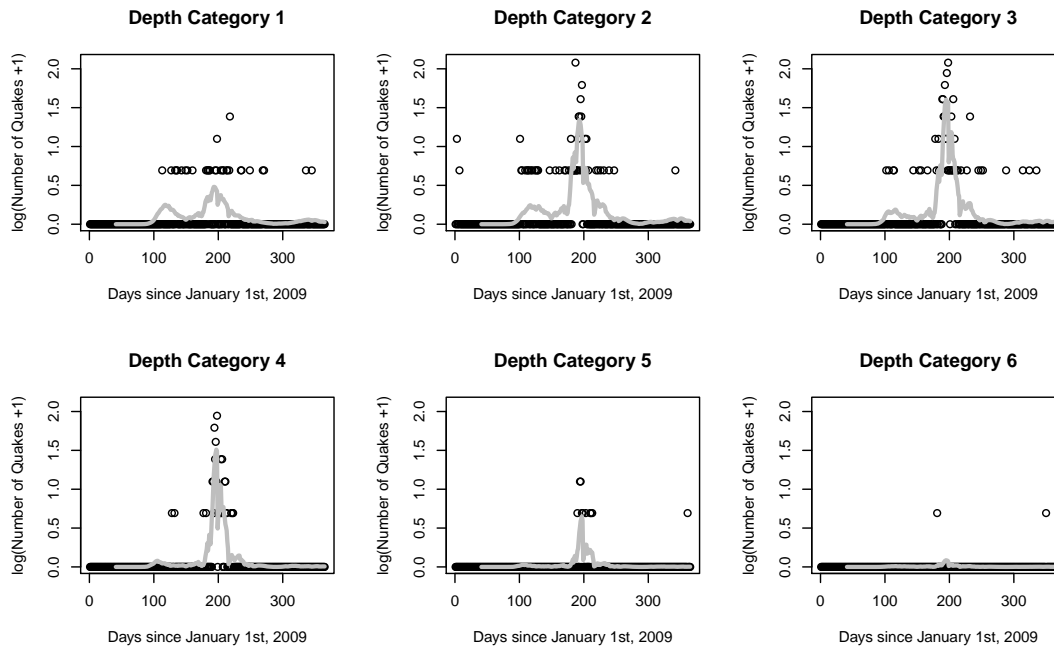
Figure 3.18: Number of quakes (black) in the six depth categories on 365 days at Mount Hochstaufen vs estimated number of quakes (grey) on the $\log(y+1)$ -scale.

estimated number of quakes in each depth category (grey).

The deviance residuals of the estimated model can be found in the Figure 3.19. An examination of the autocorrelations of the residuals is shown in the Figure 3.20. All in all we judge the quality of the modeling as quite acceptable.

3.8 Comparison with Other Lag Approaches

Strictly speaking, the strength of our method lies in the adequate combination of existing statistical components which allows us to get a general, easily applicable tool for reliable estimation of lagged linear covariate influence in the regression context. We use B-splines, that are already used by Muggeo (2008) and Gasparrini et al. (2010) in the distributed lags context. Nevertheless, Gasparrini et al. (2010) model linear and nonlinear lagged covariate influence without any penalization. Muggeo (2008) works on one explicit application with v-shaped covariate influence. In our simulation study, we compared our methods (FDL and PLM) to the use of Almon polynomials. The approaches of Zanobetti et al. (2000) and Gasparrini et al. (2010) represent an advance in comparison to the Almon polynomials. Zanobetti et al. (2000) use truncated power series bases. Further differences to the approach of Zanobetti et al. (2000) are our ridge penalty, which improves the estimates as the simulation study shows. Third, the estimates in the paper of Zanobetti et al. (2000) are
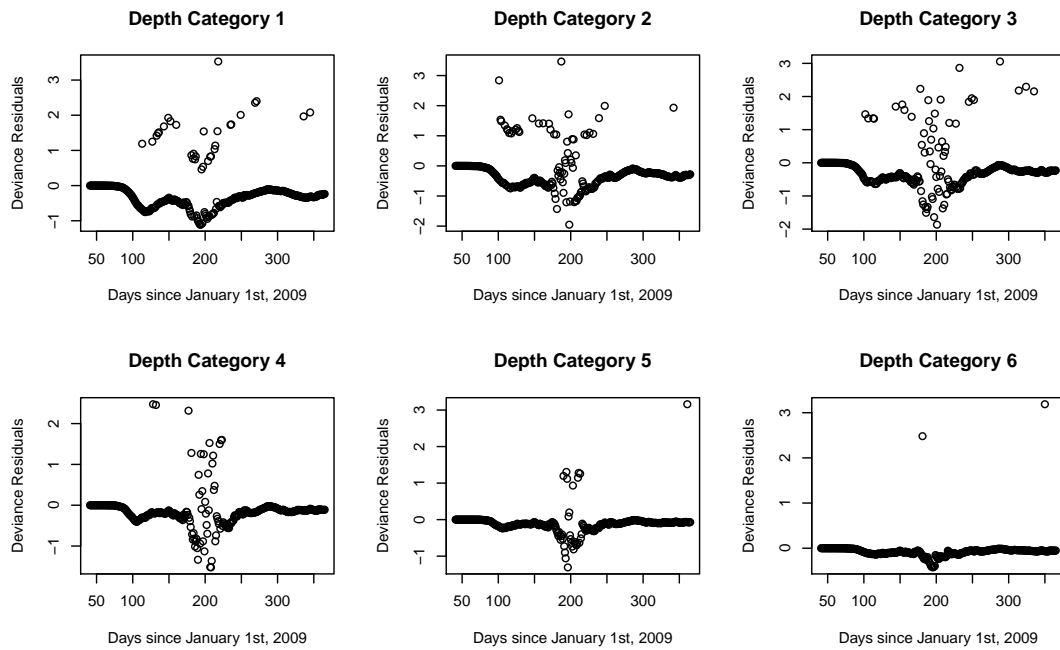
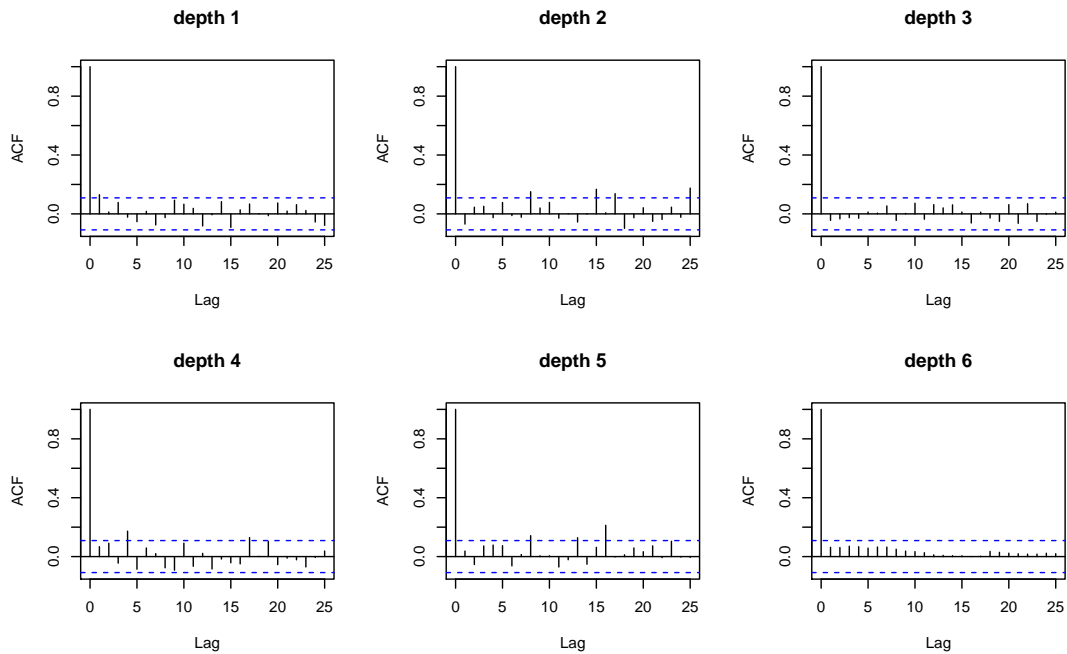Figure 3.19: Deviance residuals in the Hochstaufen data.

Figure 3.20: ACF of the residuals in the Hochstaufen data.

done via *gcv* which performed worse in our simulation study than REML (see also Wood (2011) and Reiss and Ogden (2009) for the benefits of REML based estimates in comparison to GCV-or UBRE-based estimates in similar contexts to ours). Zanobetti et al. (2000) write about problems occuring when selecting each of the smoothing parameters. The authors use GCV "to select *one* of the smoothing parameters if the remaining ones are fixed by other means." To our knowledge, Zanobetti et al. (2000) did not present any simulation study to show the benefits of their modeling approach.

Gasparrini et al. (2010) use locally defined B-splines with different numbers of knots for the modeling of nonlinear covariate influence. These B-splines are unpenalized. Also Muggeo (2008) and Muggeo (2010), respectively, propose smoothing of the lag curves and an additional ridge penalty within a segmented regression framework. Muggeo limits his method to Poisson log linear regression to model the v-shaped temperature effect on mortality. Nondecreasing known weights $\nu_1, \nu_2, \ldots, \nu_L$ with $\nu_l = l^2$, e.g., control the shrinkage of the lag curve. In our approach, we explicitly shrink the last lag coefficient β_L towards zero while the remainder of the lag course is estimated without taking assumptions of monotonicity and without the need to manually specify a vector of hard to elicit control parameters. The first lag coefficients can take any smooth form that corresponds to the data. Table 3.7 gives a general overview over distributed lag methods whose components or general ideas are (partly) used for the construction of our FDL method.

In our FDL approach, we provided a general, easily applicable tool for reliable estimation of lagged linear covariate influence in the generalized additive regression context while most previous proposals are implemented only for specific regression models.

	Welty et al. (2009)	Zanobetti et al. (2000)	Muggeo (2008)	Gasparrini et al. (2010)	FDL
Description	Empirical Bayesian approach. Specifies a prior series basis. that constrains the coefficients with larger l to be smooth and to approach zero.	Modeling of distributed lags via truncated power distributed lag parametrization within a segmented regression model.	Doubly penalized spline based parameterization along the space of the predictor and in the lag dimension by means of a crossbasis.	Models in a flexible way effects that vary both along the space of the predictor and in the lag dimension in a context.	Doubly penalized based B-spline approach for the inclusion of lagged linear covariates in the regression
Data	Effect of particulate matter on mortality	Effect of air pollutants on mortality	Effect of temperature on mortality	Effect of temperature on mortality	Effect of rain on the number of earthquakes.
Covariate influence	linear	linear	v-shaped	linear + nonlinear	linear
Type of Splines	( )	Truncated power series basis	B-splines	B-splines	B-splines
Smoothing	✓	✓	✓	✓	✓
Add. Shrinkage	✓	-	✓ (via known decreasing weights $\nu_l = l^2$, e.g.)	-	✓
Est. of λ	( )	GCV	UBRE	-	REML
Software	for $y \sim N, Po$	-	package *modTempEff* for Poisson log linear regression	crossbasis over package *dlnm* → *mgcv*	crossbasis over easy embedding in *mgcv*
Simulation Study	✓	-	-	-	✓

Table 3.7: Tabularly comparison of different distributed lag models

Chapter 4

Flexible Distributed Lags for Non-linear Effects

4.1 Nonlinear Covariate Effects

In many applications, the effect of a covariate on a response variable cannot reasonably assumed to be linear, but can take any functional form (see also Section 2.1). Gasparrini et al. (2010) and Muggeo (2008) model the influence of temperature on mortality, that is typically assumed to be v-shaped or u-shaped. Mortality is assumed to be lowest at a comfortable temperature or comfortable temperature range at round about 21 degree celsius and is assumed to be high at extremely cold and extremely hot temperature values. We want to model geothermal data (see Section 4.6). In this data, we are interested in the relationship between detracted energy, built from the amount of extracted earth liquid and its respective re-injection temperature and the occurrence of micro earthquakes. The influence of the detracted energy on the earthquake activity is clearly assumed to be nonlinear, assuming that high energy rates as well as extremely low energy rates, strictly speaking an immediate stop of energy extracting rates, could cause increased seismic activity.

In addition, we examine the effect of rain on volcanic activity. We illustrate nonlinear covariate effects by means of data of the volcano Merapi (see Section 1.1.4). Figure 4.1 shows the nonlinear (not lagged) effect of rain of the station *Selo* on the binary response variable *Block and Ash Flows*. We see a nonlinear relationship between rain of that station and the occurence of block and ash flows with increasing confidence intervals across the 90 liter per m^2 limit where less observations occur. It is assumed, that a certain time lag is needed, in order to observe the underlying rain effect. We need a distributed lag modeling approach, that can deal with nonlinear and lagged covariate effects.

4.2 Distributed Lag Non-Linear Models (DLNM)

DLNM (Gasparrini et al., 2010, see Section 2.4.5) is a modeling framework that can simultaneously represent nonlinear and lagged covariate effects. A tensor product (which is
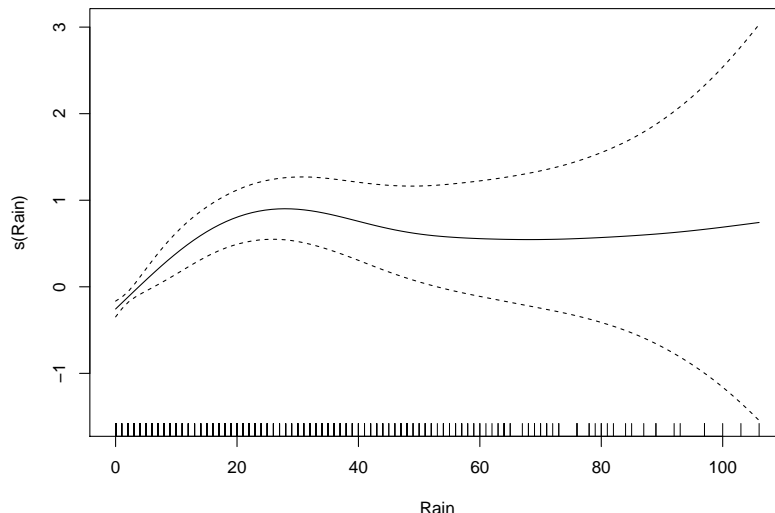
Figure 4.1: Nonlinear influence of rain measured at station Selo on the binary response
          variable Block and Ash Flows.

called crossbasis in the paper) is used, to combine the shape of the relationship along the
covariate and its distributed lag effects. A $n \times \nu_x \times (L+1)$ array $\mathbf{R}$ represents the lagged
occurrences of each of the basis variables in covariate direction. By means of $\mathbf{B}$, the matrix
of basis variables in lag direction, a DLNM is specified by (see Equation (2.29)):

$$s(x_t, \boldsymbol{\gamma}) = \sum_{j=1}^{\nu_x} \sum_{k=1}^{\nu_l} \mathbf{r}_{tj.}^T \mathbf{b}_{.k} \gamma_{jk} = \mathbf{z}_{t.}^T \boldsymbol{\gamma}. \tag{4.1}$$

The R-package *dlnm* generates a crossbasis for user-defined knots and degree of the B-
splines. Figure 4.2 shows some cubic B-spline basis functions. We only plot a subset of
basis functions in order to generate a demonstrative presentation of the basis functions. For
generalized additive models, as intended with the geothermal and the Merapi data set, the
estimations for γ_{jk} can be calculated via the function *gam* of the R-package *mgcv*. These
estimations are unpenalized. In the following sections, we penalize the basis coefficients
created via the package *dlnm* similar to the penalization concept proposed in Chapter 3,
in order to get smoothed and shrinked estimates for lagged nonlinear covariate effects.

4.3   Penalized Distributed Lag Non-Linear Models

The approach of Gasparrini et al. (2010) can model linear and nonlinear lagged covariate
effects. The concept of crossbasis boosts the number of parameters very quickly and the
user has to keep the number of knots small in order to get stable estimates with acceptable
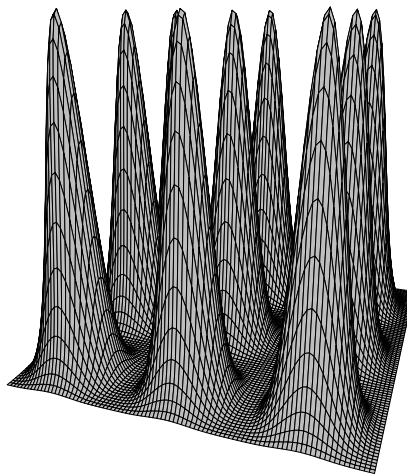
Figure 4.2: Partial tensor product basis obtained from cubic B-splines.

standard errors. The dimension of the cross-basis is the product of the dimensions in lag direction and in covariate direction. If one uses e.g., ten basis functions in lag direction and ten in covariate direction, ten $\times$ ten lag-covariate coefficients are to be estimated. So, the approach of Gasparrini et al. (2010), makes the user keep the number of knots small. On the other hand, the more knots are used, the more flexible lag courses can be detected. In Chapter 3, we propose to use more basis functions ($0.75 \cdot L$, with $L =$ maximum lag included into the model) and to stabilize estimates by means of smoothing of adjacent basis coefficients. The aim of this chapter is the transfer of the penalization concept for linear lagged covariate effects (Chapter 3) to nonlinear lagged covariate effects by extending the approach of Gasparrini et al. (2010). We call our penalized approach, based on the approach of Gasparrini et al. (2010) "Penalized Distributed Lag Non-Linear Models" (Penalized DLNM).

4.3.1 Smoothing

In the FDL for linear covariate effects (see Chapter 3), we use penalty terms via squared differences of neighbored basis coefficients in order to enable smooth estimates of the lag courses $\boldsymbol{\beta}$. In the case of bivariate basis functions, we first define suitable spatial neighborhoods (Fahrmeir et al., 2007) for differences of first order (see Figure 4.3) and differences of second order (see Figure 4.4). When using differences of first order, γ_{jk} has four neighbors whose differences of coefficients are penalized. We penalize $\gamma_{j-1,k}$ and $\gamma_{j+1,k}$ in variable direction and $\gamma_{j,k-1}$ and $\gamma_{j,k+1}$ in lag direction. Differences of second order lead to eight neighbors, indicated through the symbols | and $-$ in Figure 4.4. In Table 4.1, we exemplarily show a difference matrix for 4 neighbors (first order differences) for a simplified
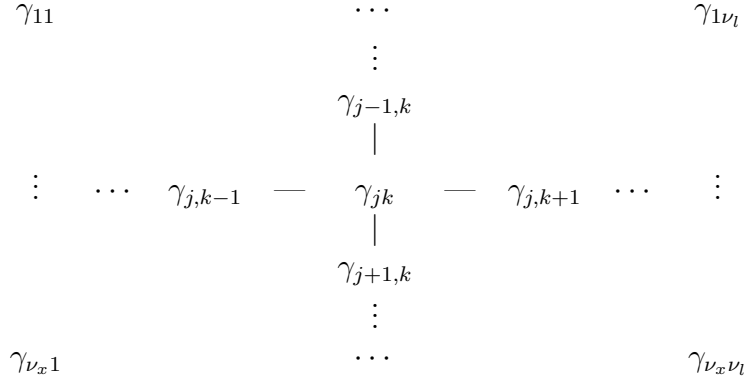
$$\gamma_{11} \qquad\qquad \cdots \qquad\qquad \gamma_{1\nu_l}$$

$$\vdots$$

$$\gamma_{j-1,k}$$

$$|$$

$$\vdots \quad \cdots \quad \gamma_{j,k-1} \quad - \quad \gamma_{jk} \quad - \quad \gamma_{j,k+1} \quad \cdots \quad \vdots$$

$$|$$

$$\gamma_{j+1,k}$$

$$\vdots$$

$$\gamma_{\nu_x 1} \qquad\qquad \cdots \qquad\qquad \gamma_{\nu_x \nu_l}$$

Figure 4.3: Spatial neighborhoods of the basis coefficient matrix $\boldsymbol{\gamma}$ (4 neighbors). The neighbors of γ_{jk} are marked with $|$ and $-$ respectively.

basis coefficient matrix with three basis coefficients in covariate direction and four basis coefficients in lag direction. The upper part of the matrix (the first eight rows) describes differences of basis coefficents in covariate direction. The lower part of the matrix describes differences of basis coefficients in lag direction.

We penalize the differences $\sum_{j=2}^{\nu_x} \sum_{k=1}^{\nu_l} (\gamma_{jk} - \gamma_{j-1,k})^2$ or rather $\sum_{j=3}^{\nu_x} \sum_{k=1}^{\nu_l} (\gamma_{jk} - 2\gamma_{j-1,k} + \gamma_{j-2,k})^2$ of the basis coefficients in variable direction and the differences $\sum_{j=1}^{\nu_x} \sum_{k=2}^{\nu_l} (\gamma_{jk} - \gamma_{j,k-1})^2$ or rather $\sum_{j=1}^{\nu_x} \sum_{k=3}^{\nu_l} (\gamma_{jk} - 2\gamma_{j,k-1} + \gamma_{j,k-2})^2$ of the basis coefficients in lag direction. The penalization matrix $\mathbf{K}_d$ for smoothing based on differences, can be constructed via Kronecker products. The analogon to the smoothing matrix for linear covariate effects (see Section 3.2.1) for the smoothing matrix for nonlinear covariate effects can be written as:

$$\mathbf{K}_d = \mathbf{I}_{\nu_l} \otimes \mathbf{K}_{d,\nu_x} + \mathbf{I}_{\nu_x} \otimes \mathbf{K}_{d,\nu_l} \tag{4.2}$$

with $\mathbf{I}_{\nu_l}$ and $\mathbf{I}_{\nu_x}$ identity matrices of dimensions $\nu_l \times \nu_l$ and $\nu_x \times \nu_x$, respectively and $\mathbf{K}_{d,\nu_x}$ and $\mathbf{K}_{d,\nu_l}$ the $\nu_x \times \nu_x$ and $\nu_l \times \nu_l$, respectively matrices $\mathbf{D}_1^T \mathbf{D}_1$ or $\mathbf{D}_2^T \mathbf{D}_2$, depending on the degree of differences, with $\mathbf{D}_1$ and $\mathbf{D}_2$ defined as in Equations (2.1) and (2.3) (except the dimension).

4.3.2 Shrinkage

In the case of linear lagged covariate effects in Chapter 3, we add a ridge penalty to the basis coefficients referring to the last lag coefficient β_L in order to enable the last lag coefficient to be close to zero if the data is compatible with that decay. We now proceed with lagged nonlinear effects just like with lagged linear effects. Similar to Section 3.2.2 for linear lagged covariate effects, we add a ridge penalty to the last $deg + 1$ basis coefficients in lag direction (with deg = degree of the B-Splines in lag direction) in order to support the decreasing effect of the covariate within the area of the maximum lag L. In the DLNM
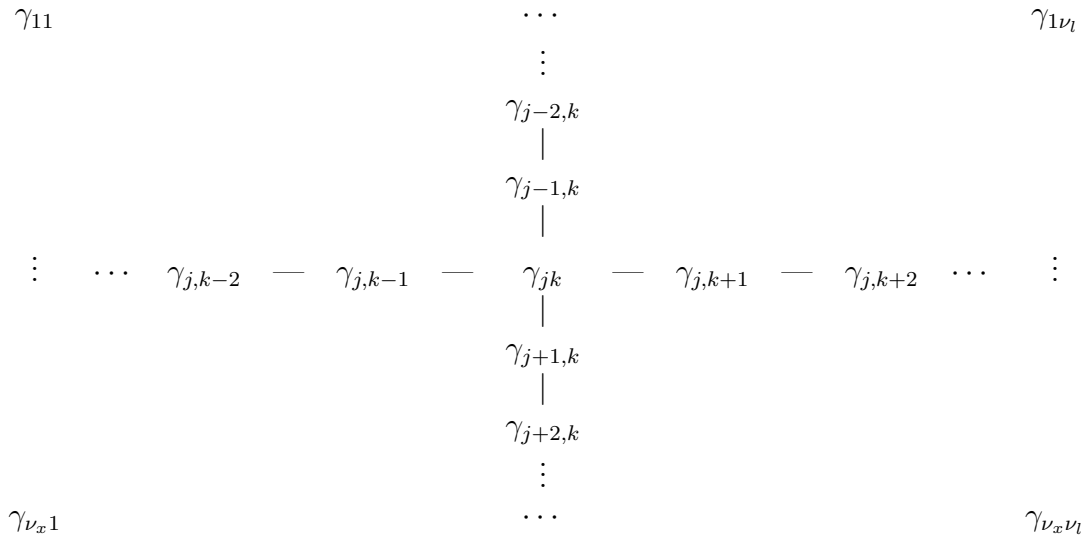
γ_{11} $\qquad\qquad\qquad\qquad\cdots\qquad\qquad\qquad\qquad$ $\gamma_{1\nu_l}$

$$\vdots$$

$$\gamma_{j-2,k}$$
$$|$$
$$\gamma_{j-1,k}$$
$$|$$

$\vdots\quad\cdots\quad\gamma_{j,k-2}\ -\ \gamma_{j,k-1}\ -\ \gamma_{jk}\ -\ \gamma_{j,k+1}\ -\ \gamma_{j,k+2}\quad\cdots\quad\vdots$

$$|$$
$$\gamma_{j+1,k}$$
$$|$$
$$\gamma_{j+2,k}$$
$$\vdots$$

$\gamma_{\nu_x 1}$ $\qquad\qquad\qquad\qquad\cdots\qquad\qquad\qquad\qquad$ $\gamma_{\nu_x\nu_l}$

Figure 4.4: Spatial neighborhoods of the basis coefficient matrix $\boldsymbol{\gamma}$ (8 neighbors). The neighbors of γ_{jk} are marked with $|$ and $-$ respectively.

γ_{11}	γ_{21}	γ_{31}	γ_{12}	γ_{22}	γ_{32}	γ_{13}	γ_{23}	γ_{33}	γ_{14}	γ_{24}	γ_{34}
1	-1	0	0	0	0	0	0	0	0	0	0
0	1	-1	0	0	0	0	0	0	0	0	0
0	0	0	1	-1	0	0	0	0	0	0	0
0	0	0	0	1	-1	0	0	0	0	0	0
0	0	0	0	0	0	1	-1	0	0	0	0
0	0	0	0	0	0	0	1	-1	0	0	0
0	0	0	0	0	0	0	0	0	1	-1	0
0	0	0	0	0	0	0	0	0	0	1	-1
1	0	0	-1	0	0	0	0	0	0	0	0
0	0	0	1	0	0	-1	0	0	0	0	0
0	0	0	0	0	0	1	0	0	-1	0	0
0	1	0	0	-1	0	0	0	0	0	0	0
0	0	0	0	1	0	0	-1	0	0	0	0
0	0	0	0	0	0	0	1	0	0	-1	0
0	0	1	0	0	-1	0	0	0	0	0	0
0	0	0	0	0	1	0	0	-1	0	0	0
0	0	0	0	0	0	0	0	1	0	0	-1

Table 4.1: Difference matrix for first order differences for 3 basis variables in covariate direction and 4 basis variables in lag direction.

$$
\begin{array}{ccc|ccc}
\gamma_{11} & \cdots & & \gamma_{1,\nu_l-deg} & \cdots & \gamma_{1\nu_l} \\
\vdots & & & \vdots & & \vdots \\
\gamma_{j1} & \cdots & & \gamma_{j,\nu_l-deg} & \cdots & \gamma_{j\nu_l} \\
\vdots & & & \vdots & & \vdots \\
\gamma_{\nu_x 1} & \cdots & & \gamma_{\nu_x,\nu_l-deg} & \cdots & \gamma_{\nu_x\nu_l}
\end{array}
$$

Figure 4.5: Basis coefficient matrix $\boldsymbol{\gamma}$: A ridge penalty is to be added to the last $deg + 1$ basis coefficients in lag direction (to the right of the dividing line).

framework, we have to penalize $\nu_x \times (deg + 1)$ coefficients, see also Figure 4.5 where all coefficients to the right of the dividing line have to be penalized in order to encourage a shrinked effect at the maximum lag. A corresponding ridge penalty matrix $\mathbf{K}_r$ for basis coefficients arranged in the same order like in Table 4.1, would be a $\nu_x\nu_l \times \nu_x\nu_l$ matrix, where the diagonal elements corresponding to the last $\nu_x \times (deg + 1)$ coefficients are ones and zero elsewhere.

4.4 Simulation: Comparison of Penalized DLNM and DLNM

4.4.1 Design

We restart a simulation study in order to find out, if the penalization concept proposed in Section 4.3 can improve the unpenalized estimation proposed by Gasparrini et al. (2010). The upper graphic of Figure 4.6 shows real bavarian temperature, already known from the previous chapter, that acts as covariate for the simulation study. We model the "true" temperature effect, that has to be detected via penalized and unpenalized estimation, as surface with a demanding bimodal temperature effect decreasing towards zero with increasing lag and a maximum lag length of $L = 7$ time units, e.g., days. Figure 4.7 shows the chosen lagged covariate effect $s(x, l)$ created via:

$$s(x, l) = 4 * (dnorm(x, sd = 3) + dnorm(x, mean = 15, sd = 2)) * (\exp(-t/2)/6) \quad (4.3)$$

The predictor is calculated as sum of the (lagged) temperature effects with x_{t-l} the amount of rain fallen l time units before time unit t:

$$\eta_t = \alpha_0 + \sum_{l=0}^{L=7} s(x_{t-l}, l) \tag{4.4}$$

$$\lambda_t = \exp(\eta_t) \tag{4.5}$$

$$y_t \sim Po(\lambda_t) \tag{4.6}$$

We generate 1000 Poisson distributed response variables out of the predictor. The lower graphic in Figure 4.6 shows one arbitrarily chosen resulting response variable.
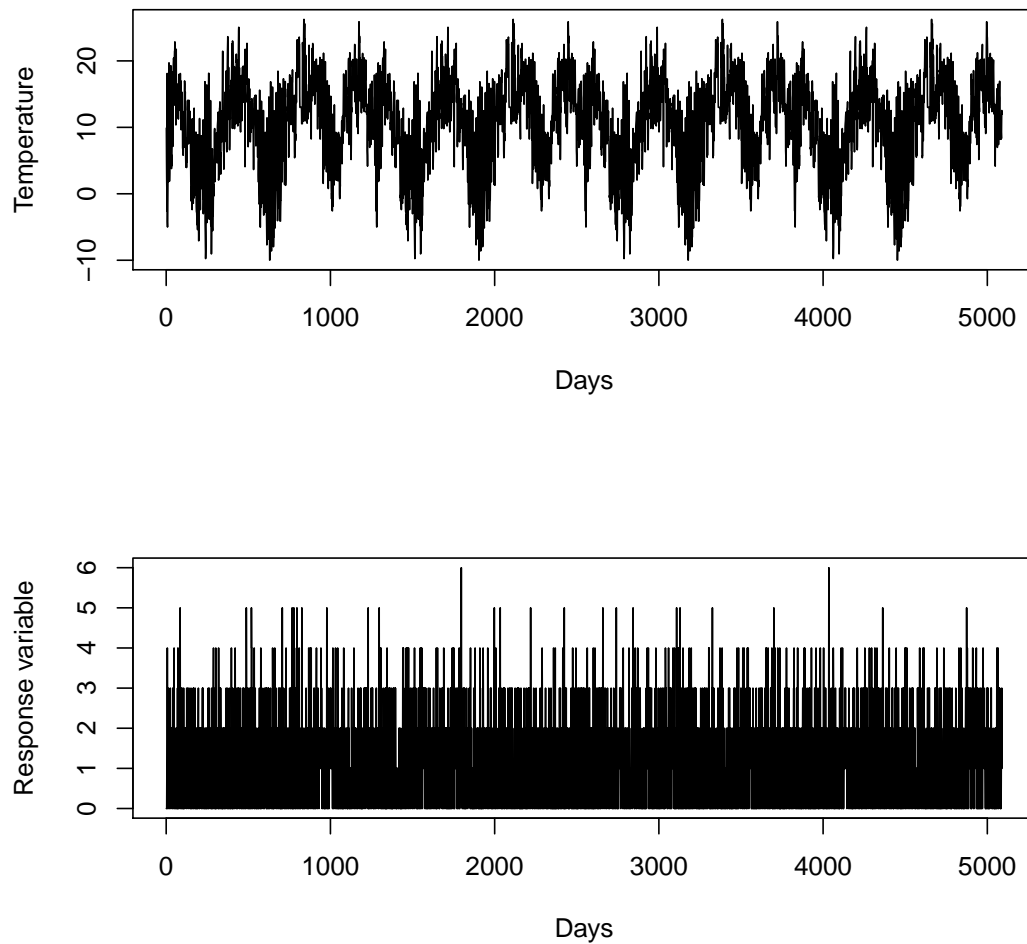
Figure 4.6: Simulation: Covariate temperature and resulting response variable for 1 simulation run.
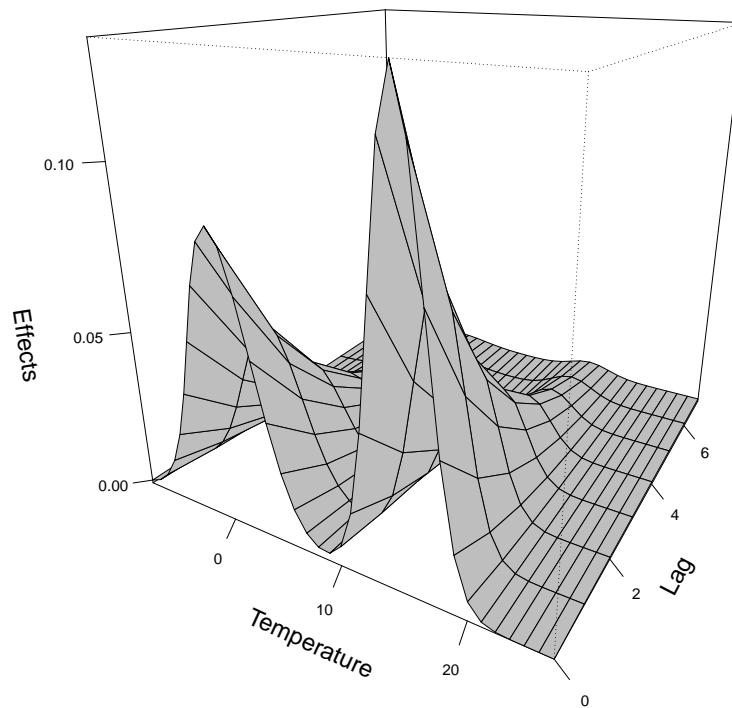
Figure 4.7: Simulated surface $s(x, l)$ of the effects of temperature (x) and lag (l):
$s(x, l) = 4*(dnorm(x, sd = 3) + dnorm(x, mean = 15, sd = 2))*(\exp(-t/2)/6)$.

4.4.2 Comparison of Unpenalized and Penalized Estimation for Varying Numbers of Knots

We estimate the lagged temperature effect $\hat{s}(x, l)$ by means of penalized and unpenalized REML- based estimation of the crossbasis coefficients for different numbers of equidistant inner knots. We compare estimated and true lagged covariate effects on a discrete two dimensional grid with $\sum_{x=-9}^{26} \sum_{l=0}^{L}(s(x,l) - \hat{s}(x,l))^2$, averaged over 1000 runs of the Poisson distribution.

Table 4.2 gives a performance comparison of unpenalized and penalized DLNM estimation depending on different numbers of knots in variable and in lag direction. Cubic B-splines and second order differences are used. All values are given relative to the best estimation (Setting A): a smoothed and shrinked estimation with one knot for every possible discrete temperature value and one knot for every lag unit. According to Table 4.2, the unpenalized estimation with only one inner knot in lag and one in variable direction (Setting D) has 23.75 higher summed squared errors than the estimation of Setting A. Concerning unpenalized estimates in general, the summed squared errors become larger with an increasing number of knots due to resulting wiggly and instable lag courses. Small numbers of knots, however cannot detect more flexible shapes of lag courses. We conclude that, if we use a suitable smoothing penalty, increasing numbers of knots lead to decreasing summed squared errors and therefore to an improvement of the estimation. The demanding bimodal course in covariate direction can best be discovered by penalized estimation with many knots. The second finding in this simulation study is about the ridge penalty. Even though the true lagged covariate influence was already close to zero within the area of the maximum lag $L = 7$ (see again Figure 4.7), the additional ridge penalty improves the estimation. The model without the additional ridge penalty (Setting B) has 2.56 times higher summed squared errors than the model with the additional ridge penalty for the setting with the many knots (Setting A).

Figures 4.8, 4.9 and 4.10 oppose sections of the true lagged temperature effects (see Figure 4.7) to selected estimates (Settings A, C, D in Table 4.2). All lag courses represent averaged estimates over 1000 runs. All effects are centered at a centering value of -9°. The centering is a common and necessary procedure in the modeling of nonlinear covariate effects. For all three figures, the left column shows estimated (black) and true (grey) temperature effects for selected lags. We show temperature effects for lag 0 (first row), lag 1 (second row), lag 3 (third row) and lag 7 (last row). Concerning the analysis in lag direction (right column), we show lag courses for the selected fixed temperature values -9°, 3°, 14° and 20°. Due to the centering at the temperature values at -9°, the estimates for Temperature = -9° (plot top right for all three figures) is indicated with a horizontal zero line. Please note, that the y-axes differ within the plots and within the figures in order to give an as detailed as possible look at the estimates. When comparing temperature effects at lag = 0 to those at lag = 7, where the absolute effects are much smaller (at least for the optimal estimation), same axes at both plots would lead to misrepresented estimates of lag = 7. The first of the three figures, Figure 4.8 shows the optimal estimation (Setting A) referring to Table 4.2, a smoothed and shrinked estimation with inner equidistant

	unpenalized	smoothed	smoothed + shrinked
knots x: 8 knots l: 3	23.75 (see Fig. 4.10) Setting D		
knots x: $0, 8, 16$ knots l: $2, 4, 6$	71.20		
knots x: $-7, -4, \ldots, 20, 23$ knots l: $1, 2, 3, 4, 5, 6$	158.85	3.56	1.59
knots x: $-9, -8, \ldots, 24, 25$ knots l: $1, 2, 3, 4, 5, 6$	649.35 (see Fig. 4.9) Setting C	2.56 (see Fig. 4.13) Setting B	1.00 (see Fig. 4.8) Setting A

Table 4.2: Summed squared errors, calculated on a discrete two dimensional grid (1000 runs, $\sum_{x=-9}^{26} \sum_{l=0}^{L} (s(x,l) - \hat{s}(x,l))^2$ for unpenalized and penalized (smoothed, smoothed and shrinked) estimation for different numbers of equidistant inner knots. We used cubic B-splines and second order difference penalties. All values are given relative to the optimal estimation (Setting A): a smoothed and shrinked estimation with one knot for every possible discrete temperature value and one knot for every lag.

knots $-9, -8, \ldots, 24, 25$ in temperature direction and knots $1, 2, 3, 4, 5, 6$ in lag direction. Interpreting the estimation in temperature direction (left column), we see that for all fixed lags, the bimodal shape of the true lag course (grey) can be discovered quite well. Due to the smoothing parameter, the local maxima are slightly underestimated. Interpreting the estimation in lag direction (right column), we see that for fixed temperature values of 3° and 14°, the decay of the influence of the temperature on the response variables with increasing lag can be detected with a slight underestimation of the maximum. The underestimation of the maximum is more serious for a temperature value of 20°. Nevertheless, the shape of the decreasing influence with increasing lag is detected. The estimated degrees of freedom within that setting are 9.44 degrees of freedom averaged over 1000 runs.

We compare these results optically with those of Figure 4.9. For that estimation the same number and position of knots are used but the estimates were left unpenalized (Setting C). Table 4.2 (last row) already showed, that the summed squared errors are round about 650 times higher than those of the penalized estimations. The averaged lag courses are quite smooth, especially for fixed lags zero and one, which must be a result of averaging, but very wiggly for the remaining courses in covariate and in lag direction. Figure 4.10 shows estimates for little knots (Setting D in Table 4.2). Due to the sparse knots, no penalization is used. We see, that the true shape in temperature direction cannot be discovered quite well.

All in all, the optimal penalized estimation (Setting A) with many knots yields quite acceptable performance characteristics in comparison to the alternative results and with

regard to a very demanding lag course of a highly correlated covariate. Figure 4.11 shows the comparison of the true simulated surface and its estimated surface (Setting A). We notice again, that the general shape of the surface is represented quite well.

4.4.3 Note on the Simulation Settings

In the previous section, we compare penalized and unpenalized estimates of DLNM effects with varying numbers of knots relating to their ability for the detection of a simulated true lag course.

We use cubic B-splines and REML-based estimation for second order difference penalties (if penalized).

We want to discuss the choice of the degree of the B-splines and the order of the differences as well as the choice of the smoothing parameter via REML. In the following paragraph, we refer to the "optimal" estimation (Setting A, smoothed and shrinked with one knot for any lag unit and one for any discrete temperature value, see Table 4.2).

Concerning the smoothing parameter choice, we detect, like in Chapter 3, a (slight) superiority of REML-based smoothing parameter estimation in comparison to smoothing parameter estimation based on the UBRE criterion. For the "optimal" setting, we observe REML-based summed squared errors of 0.00054 compared to UBRE-based summed squared errors of 0.00075, so round about 1.39 higher summed squared errors when using UBRE. Therefore, we concentrate on REML-based smoothing parameter estimation as in the FDL context.

Concerning the choice of the degree of the B-splines and the choice of the differences, we detect, that second degree B-splines with differences of first order lead to smaller summed squared errors in Setting A compared to cubic B-splines with second order differences. Nevertheless, we present results for cubic B-splines and second order differences for the following reasons. First of all, third degree B-splines can discover more flexible lag courses.

According to our knowledge adopted by the analysis of the mentioned geothermal energy data set and the Merapi data set (see Section 1.1.4), we cannot recommend the use of first order differences. Smoothing parameters are estimated rather small when using first order differences and the resulting surfaces of penalized DLNM effects are rather wiggly.

4.4.4 Centering and Interpretation

As already mentioned in Section 4.4.2, Gasparrini et al. (2010) center the estimates around a centering value x_0 for each lag $l = 0, \ldots, L$. It holds:

$$s(x_0, l) = 0, \quad l = 0, \ldots, L \tag{4.7}$$

The surface is centered around a line. This strong restriction has to be examined. On the one hand, that kind of restriction seems to be a very strong restriction. For one specific covariate value x_0, there is no interaction with the lag, while an interaction of lag and covariate is implied for all other values.
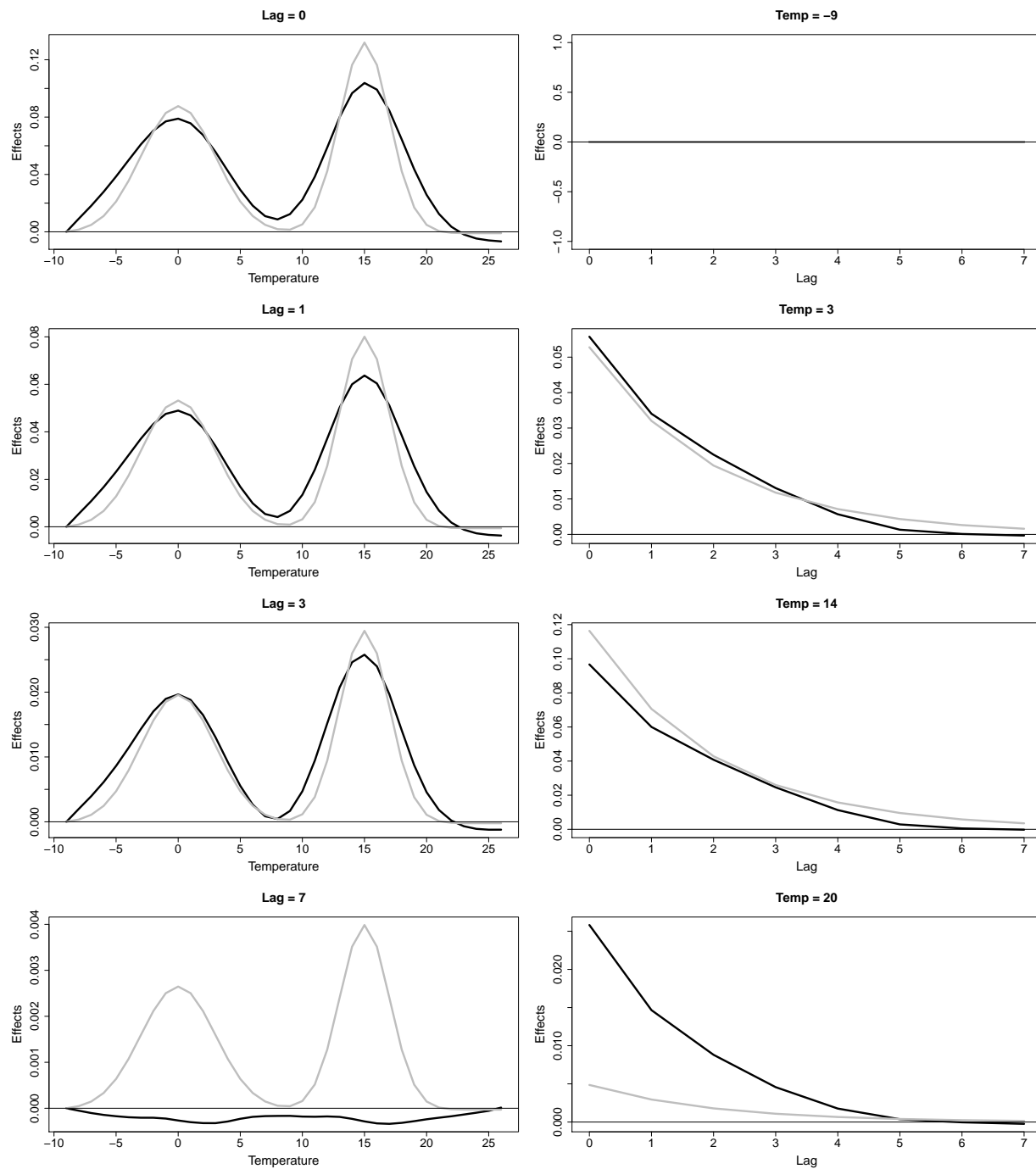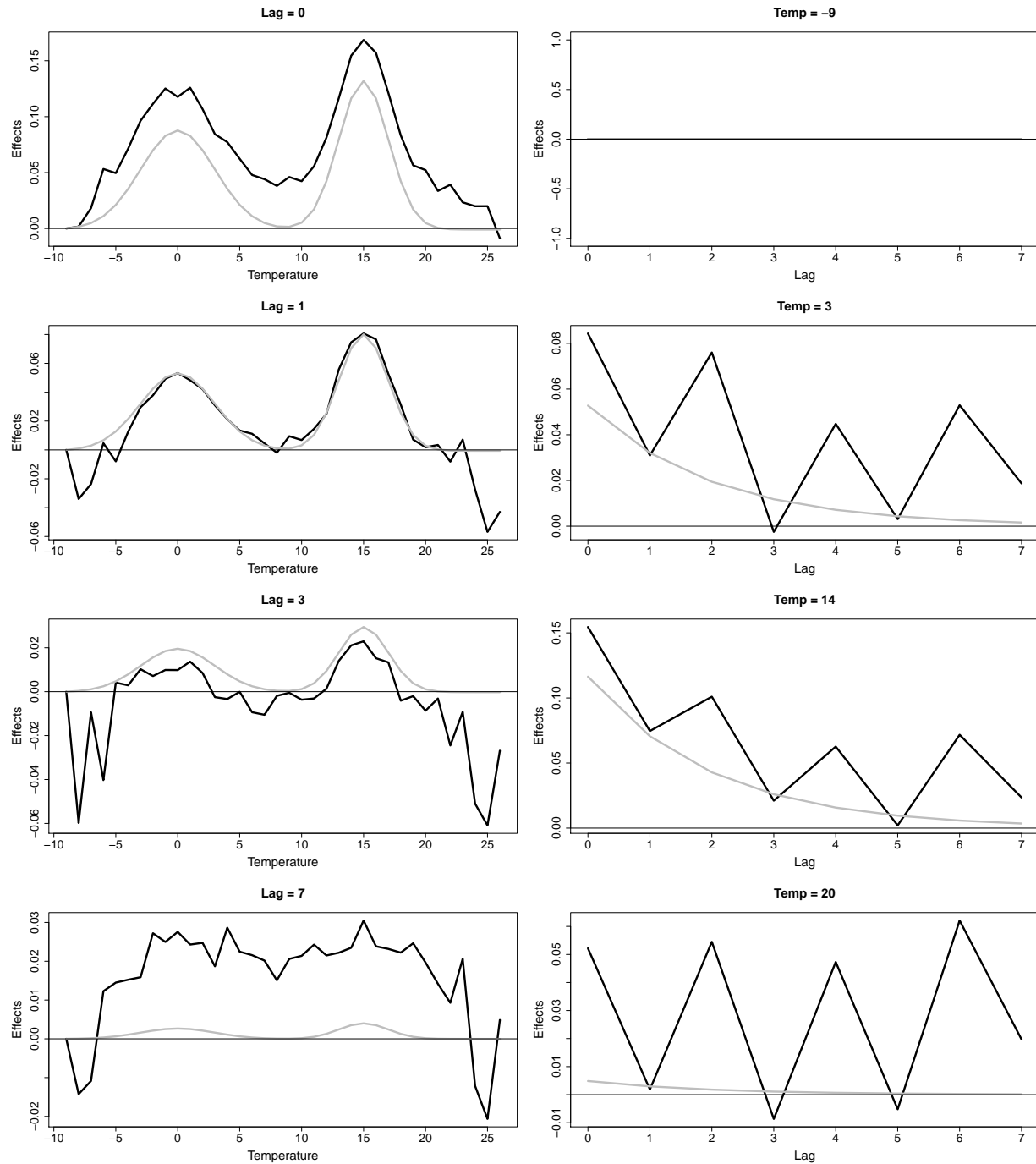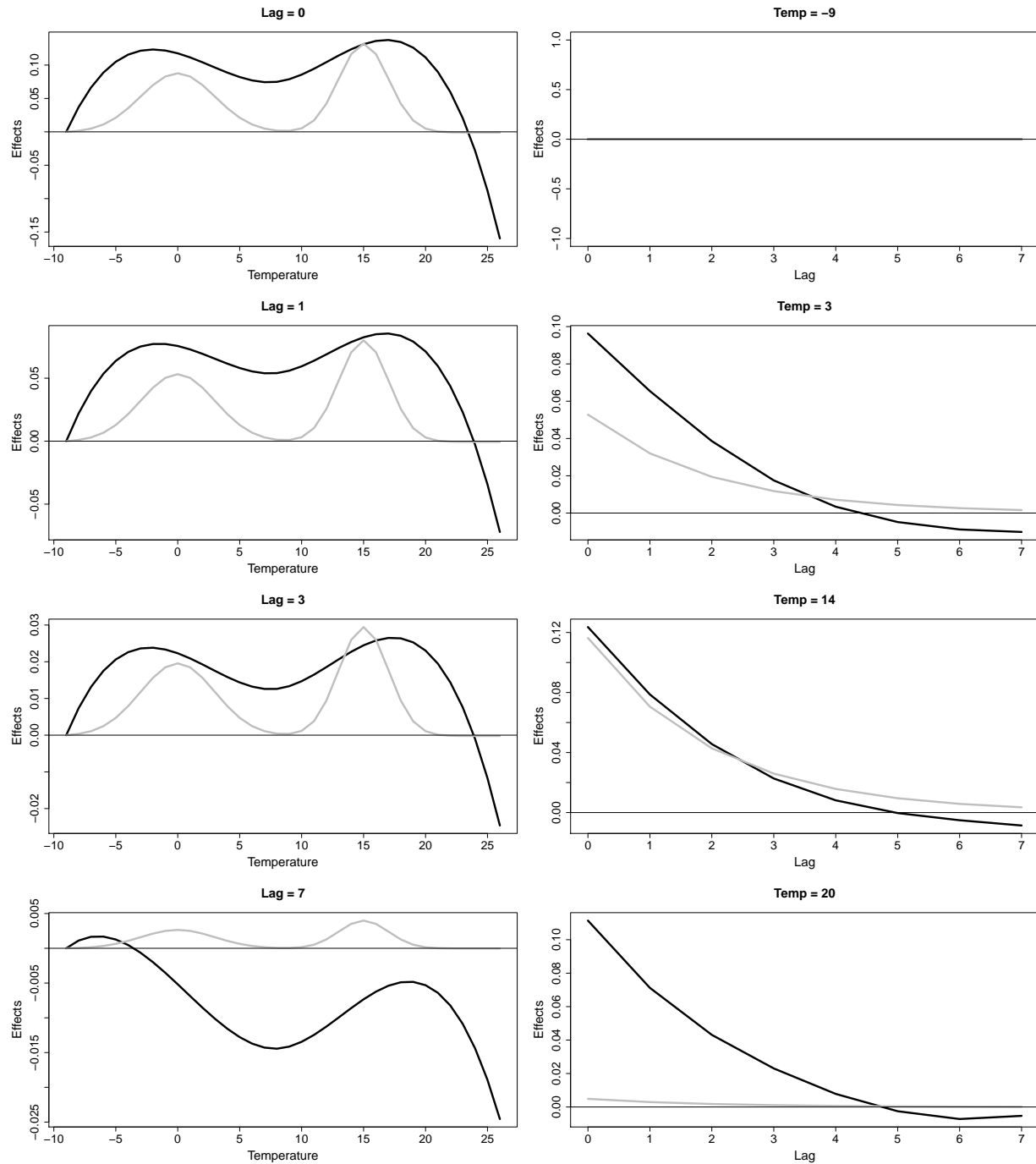
Figure 4.8: Simulation (Mean over 1000 runs): Specific estimated temperature and lag effects (black) in comparison to the true underlying effects (grey) for a smoothed and shrinked estimation (see Setting A in Table 4.2) with inner equidistant knots $-9, -8, \ldots, 24, 25$ in temperature direction and knots $1, 2, 3, 4, 5, 6$ in lag direction.

Figure 4.9: Simulation (mean over 1000 runs): Specific estimated temperature and lag effects (black) in comparison to the true underlying effects (grey) for an unpenalized estimation (see Setting C in Table 4.2) with inner equidistant knots $-9, -8, \ldots, 24, 25$ in temperature direction and knots $1, 2, 3, 4, 5, 6$ in lag direction.

Figure 4.10: Simulation (mean over 1000 runs): Specific estimated temperature and lag effects (black) in comparison to the true underlying effects (grey) for an unpenalized estimation (see Setting D in Table 4.2) with inner equidistant knot 8 in temperature direction and knots 3 in lag direction.
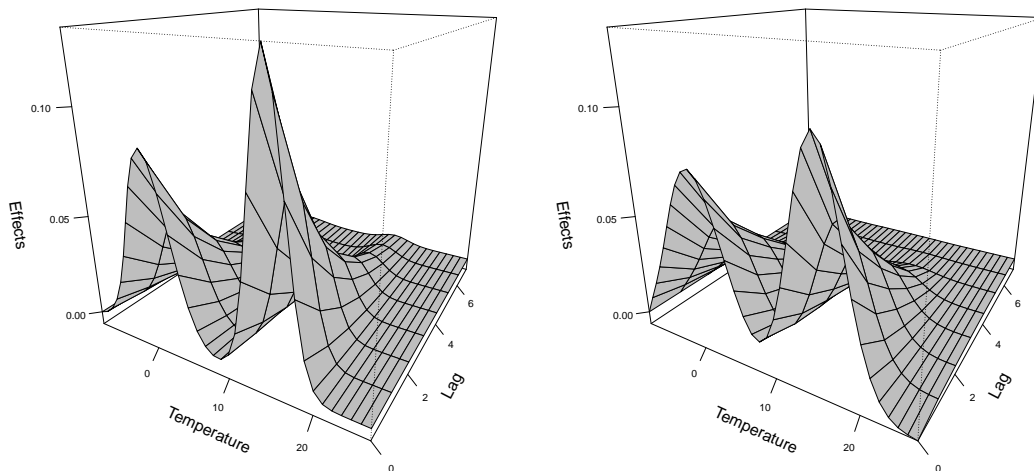
Figure 4.11: Simulated surface $s(x, l)$ (left) and estimated surface (mean over 1000 runs) $\hat{s}(x, l)$ via a smoothed and shrinked estimation (see Setting A in Table 4.2) (right).

We address the question, if the centering around a line is necessary to ensure the identifiability of the model.

Let $s_0, \ldots, s_L$ be the functions, that characterize lagged influence. Then, the linear predictor is given as:

$$\boldsymbol{\eta} = c_0 + \sum_{l=0}^{L} s_l(x) = c_0 + \sum_{l=0}^{L} s(x, l) \tag{4.8}$$

with a smooth function (surface) s. The surface is centered around the point $s(0, 0) = 0$. Due to the special model based on lags, the surface is only definded for whole numbers of l. We want to show, that the centering around the single point $s(0, 0)$ is not sufficient to ensure identifiability of the model. We construct a second surface s^*, that leads to the same liner predictor and that is similarly smooth. f^* is constructed with the same linear predictor and with the centering

$$s^*(0, l) = s_l^*(0) = 0 \quad \forall l \tag{4.9}$$

We define:

$$s^*(x, l) := s(x, l) - s(0, l). \tag{4.10}$$

It holds:

$$\boldsymbol{\eta} = c_0 + \sum_{l=0}^{L}(s_l(x) - s_l(0)) + \sum_{l=0}^{L} s_l(0) = c_1 + \sum_{l=0}^{L} s_l^*(x). \tag{4.11}$$

The functions s and s^* differ in

$$d(x, l) = s(0, l). \tag{4.12}$$

The function d is constant in x and smooth in l as $s(0, l)$ is smooth. Thus, s and s^* are similarly smooth and lead to the same linear predictor. We found two different surfaces with similar smoothness and the same linear predictor. We conclude, that the simple centering around one single point is not sufficient. We conclude from the second surface s^*, that the centering of each function s_l is necessary.

We see also advantages in the easy interpretation in the specific lag context. The main focus of lagged covariate effects lies in the presentation of the covariate effects for fixed lags (see e.g., left column of Figure 4.8). These are in the current representation easily interpretable effects.

It follows for the simulation scenarios, that the true underlying surface for the covariate lag effect has to be centered around a certain value x_0. We chose a point at the border of the surface and took the temperature value $x_0 = -9$ as centering value.

We examine the effects of a maladroit centering value. The simulated surface is simulated in a way, that the assumption of no interaction between lag and covariate is fulfilled for the reference value $x_0 = -9$. For that value, we have no interaction between lag and the covariate. For a reference value $x_0 = 15$, the assumption of no interaction between reference value and lag is clearly not fulfilled; the effect of the covariate decreases with increasing lag. We want to compare the presentation of estimated effects for estimation with the "right" reference value $x_0 = -9$ and the "wrong" reference value $x_0 = 15$. We chose arbitrarily the setting with smoothed and shrinked estimates for one inner knot for each discrete lag and covariate value (Setting A, please note, that the setting, the number of knots, and the application of a penalty does not play a role for the illustration of the influence of the centering value.) Figure 4.12 shows four plots. On the top, we present the effects of the covariate for lag 0 for the true reference value $x_0 = 0$ (left) and for the wrong reference value $x_0 = 15$ (right). We see, that the covariate effect for a fixed lag is only shifted along the y-axis depending on the reference value. The difference between the effects of two temperature values $x = 0$ and $x = 10$, e.g., is independent from the centering value. However, we see that the confidence intervals change extremely due to the constraint. For the centering value $x_0 = 15$, the intervals are much smaller. The intervals at the centering point are set to zero. We cannot solve the problems of the misleading intervals in the case of a maladroit centering value. Wood et al. (2013) write that "poor constraint choice can lead to practically useless intervals". Concerning the interpretation of the fixed covariate values along the lags (bottom), we again meet a problem. The true effect of $x_0 = 15$ is not constant. When centering around $x_0 = 15$, assuming that the effect of the covariate might be constant for all lags at that point, the effects of the other
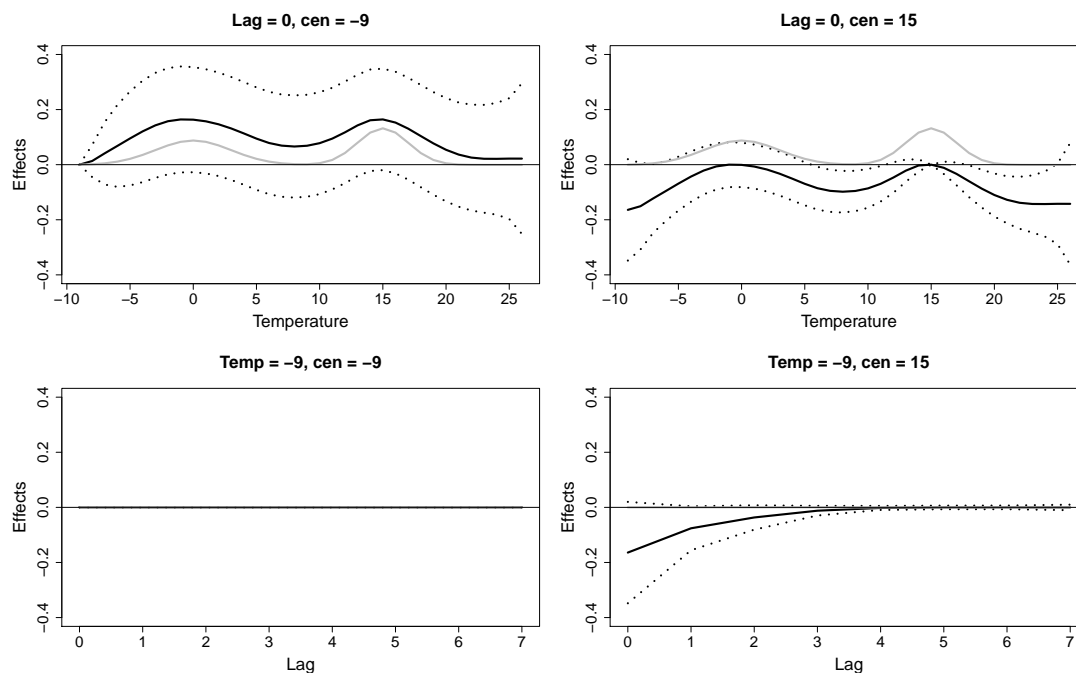
Figure 4.12: Effects of a maladroit reference value.

covariate values have to compensate that effect. Having a look at Figure 4.12, we see that for Temperature = -9 ° and a centering around $x_0 = 15$ one would interpret, that the effect of $x = -9$ has a negative effect, that diminishes with increasing lag. However, this is not correct. The effect of $x = -9$ is constant, but the difference from the effect for $x = 15$ decreases with increasing lag.

We summarize, that the choice of the centering value is a critical point within the approach of Gasparrini et al. (2010) and accordingly alike in a penalized Gasparrini approach. If one choses a maladroit centering value, confidence intervals are falsified and the interpretation of lag effects on fixed temperature values can be misleading.

The problem stays, that one does not know, whether the assumption of a certain centering value can represent the true relationship.

4.4.5   Effect of the Ridge Penalty

We focus on the effect of the ridge penalty on the quality of the estimation. Table 4.2 already showed, that under constant settings (third degree B-splines, smoothing via REML via second order differences, equidistant inner knots at $x = -9, -8, \ldots, 24, 25$ in temperature direction and equidistant inner knots at $l = 1, 2, 3, 4, 5, 6$ in lag direction) summed squared errors for smoothed, but not shrinked estimation (Setting B) are 2.56 higher than those for smoothed and shrinked estimation (Setting A). Having again a detailed look at the estimated temperature effect of Figure 4.8 for the last lag unit $L = 7$ (last row, first

	smooth (Setting B)	smooth and shrink (Setting A)
$\log_{10}(\hat{\lambda}_d)$(smoothing parameter)	6.59	3.38
$\log_{10}(\hat{\bar{\lambda}}_r)$(shrinkage parameter)	-	6.10

Table 4.3: Simulation DLNM: Comparison of logarithmized smoothing parameters $\bar{\hat{\lambda}}_d$ in the setting with (Setting A) and without ridge penalty (Setting B).

column), we locate, that the estimates for the last lag coefficients are close to zero, though below the zero level, whereas the true simulated lag curve (grey line) find its local maximum at 0.004.

We have a look at the estimate without the additional ridge penalty. Figure 4.13 shows transversal and longitudinal sections of the averaged effects for the setting without the additional ridge penalty (Setting B). Having a look at the first column, that presents the temperature effects for several fixed lags, several aspects attract the readers attention. First of all, the estimation for the temperature effect of the last lag unit $L = 7$ (last row) is far away from the real simulated lag course. For the last lag coefficient we discover large implausible negative effects that do not fit at all to the underlying true lag function. These negative effects seem to influence also the earlier lags of the surface. The temperature effects for lags 0 and 1 seem to be too smooth and drift to the negative scale for higher temperature values (see also Figure 4.14 for the estimated surface of Setting B, averaged over 1000 runs). Table 4.3 compares the averaged smoothing parameters for the setting without (Setting B) and the setting with the ridge penalty (Setting A). The table confirms that the (averaged) smoothing parameter λ_d is much larger, if there is no additional ridge penalty in the setting. The logarithmized (averaged over 1000 runs) smoothing parameter for the model without ridge penalty (first column) is 6.59 in comparison to 3.38 if an additional shrinkage is added ($\log_{10}(\bar{\hat{\lambda}}_r) = 6.10$).

We repeat the simulation study with a surface similar to the surface shown in Figure 4.7 in order to check the performance of the estimation without shrinkage in a situation where the effect for maximum lag is closer to zero. Figure 4.15 shows on the left side the new simulated lag course with longer lags. The effects have more time to taper and at the maximum lag $L = 12$, the true effects are closer to zero than in the original surface. The right side shows the averaged estimations over 1000 runs for the setting with ridge penalty and inner knots in lag direction for every lag (Setting A (L=12)). The summed squared errors (0.00040) are considerably smaller than in the estimation without ridge penalty (0.00075). Concerning the smoothing parameters, we encounter similar results to those represented in Table 4.3. Larger smoothing parameters are generated for estimation without additional shrinkage ($\log 10(\bar{\hat{\lambda}}_d) = 7.17$) than for estimation with additional shrinkage ($\log 10(\bar{\hat{\lambda}}_d) = 6.55$ and $\log 10(\bar{\hat{\lambda}}_r) = 7.22$). Instabilities in the border areas of the surface, in that case at the larger lags, are corrected via a larger amount of smoothing. This correction leads to oversmoothed lag effects at the beginning of the lag courses. The
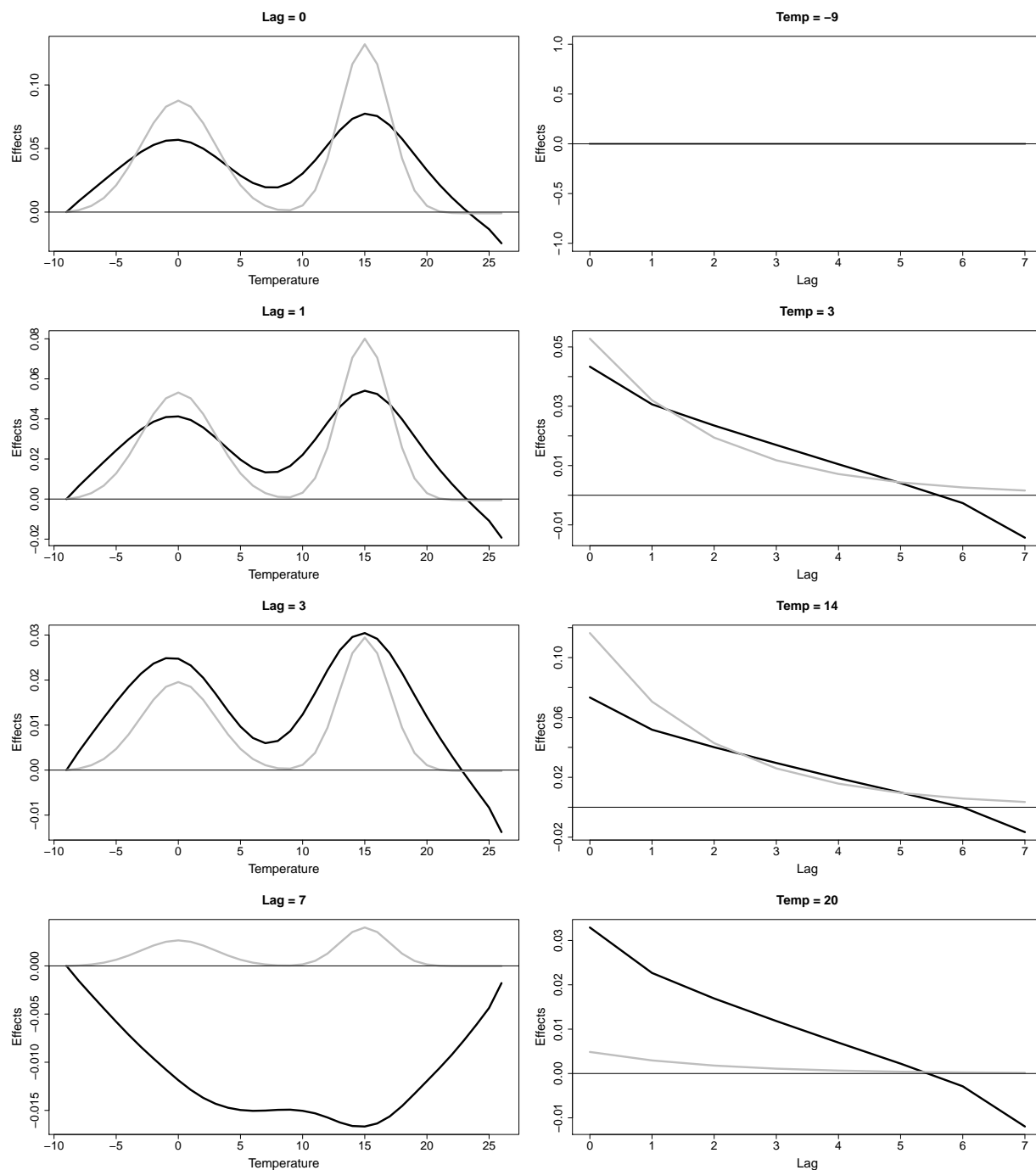
Figure 4.13: Simulation (mean over 1000 runs): Specific estimated temperature and lag effects (black) in comparison to the true underlying effects (grey) for a smoothed estimation (no shrinkage, see Setting B in Table 4.2) with inner equidistant knots $-9, -8, \ldots, 24, 25$ in temperature direction and knots $1, 2, 3, 4, 5, 6$ in lag direction.
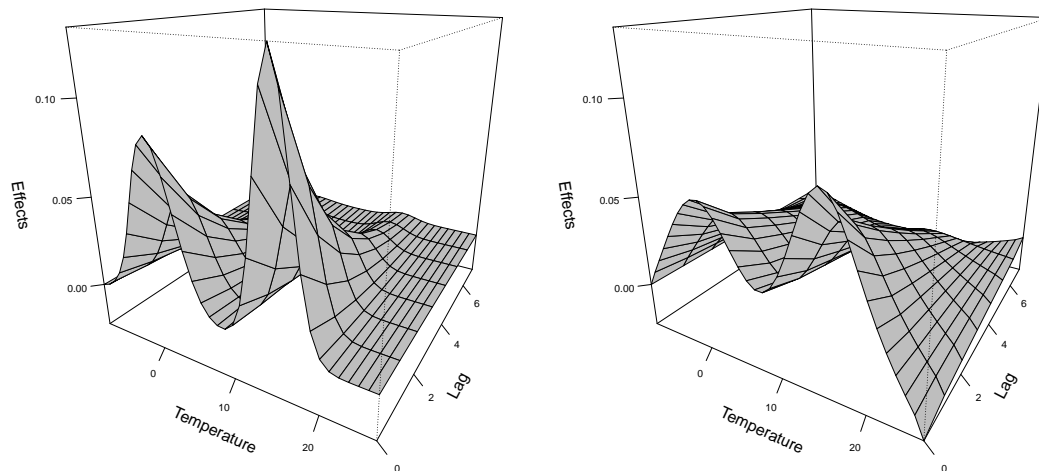
Figure 4.14: Simulated surface $s(x, l)$ (left) and estimated surface (mean over 1000 runs) $\hat{s}(x, l)$ via a smoothed but not shrinked estimation (Setting B, right).

ridge penalty can avoid these effects, as our simulation study indicated.

We try another simulation setting, as we want to find out, to which extent the high autocorrelations in the covariate temperature influence the estimation and in particular how the shrinkage affects the estimation in comparison to the estimation without shrinkage for an uncorrelated covariate. We simulate an artificial new covariate instead of the highly correlated temperature data set. We simulate an independent normally distributed random variable. To make the simulation comparable to the previous setting, mean, standard deviation and length of the new synthetic covariate correspond to mean, standard deviation and length of the covariate temperature; the mean is 9.81, the standard deviation is 7.54 and the length is 5096. We build response variables as in Subsection 4.4.1, and estimate the simulated surface shown in Figure 4.7 with one inner knot for each discrete variable value in lag and in covariate direction. Figure 4.16 shows the estimated surface in the case without shrinkage (comparable to Setting B, Setting B (ind)), averaged over 1000 runs (right) compared to the underlying true surface (left). We see, that the bimodal shape of the surface can be detected. Nevertheless the local maxima are rather underestimated. In the case of the correlated covariate temperature, higher peaks are easier to detect. We observe summed squared errors of 0.00128. Figure 4.17 shows for the independent covariate the estimation (right) of the surface with smoothing and shrinkage of the respective basis coefficients. We see, that the local maxima are better reached than for the setting without shrinkage. The shrinkage can balance instable estimates at the limits of the surface and stabilizes thereby the rest of the estimation. The summed squared errors in the smoothed and shrinked estimation are with 0.00057 considerably smaller than in the not shrinked
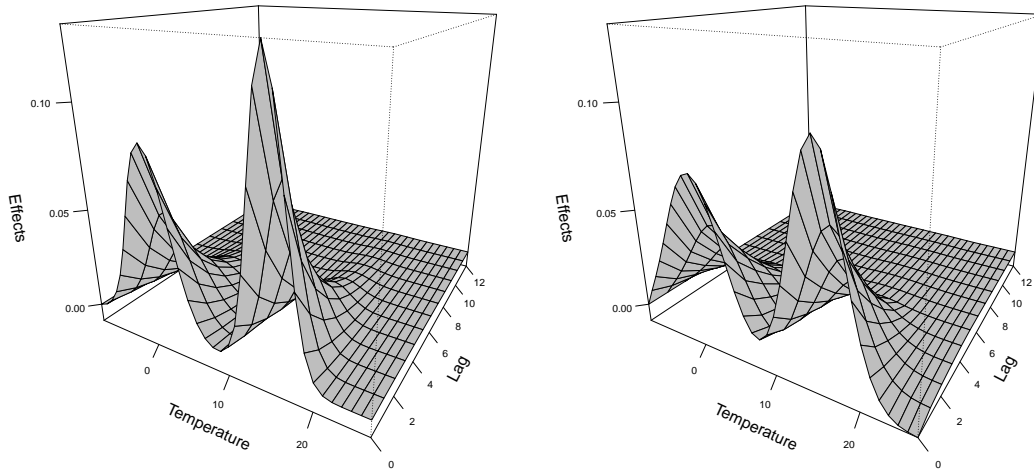
Figure 4.15: Simulated surface $s(x, l)$ (left) and estimated surface (mean over 1000 runs) $\hat{s}(x, l)$ via a smoothed and shrinked estimation (right, Setting A (L=12)) for a maximum lag $L = 12$.
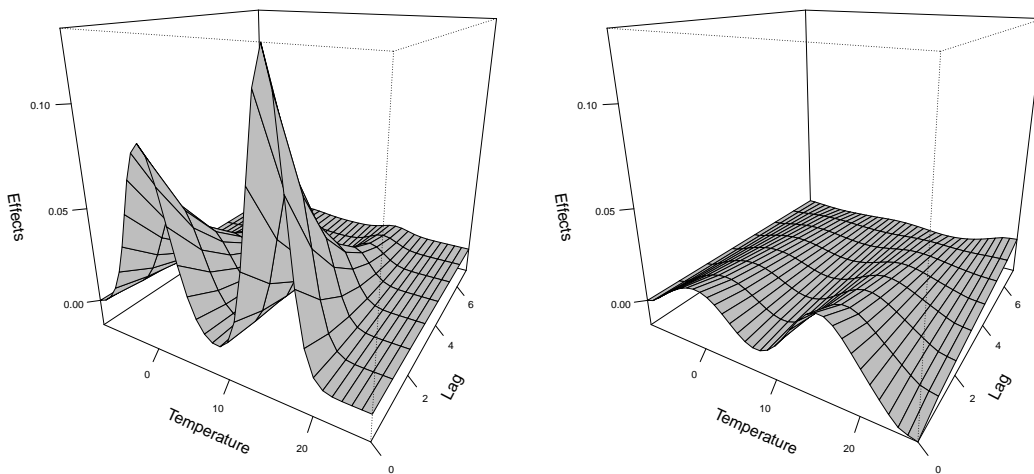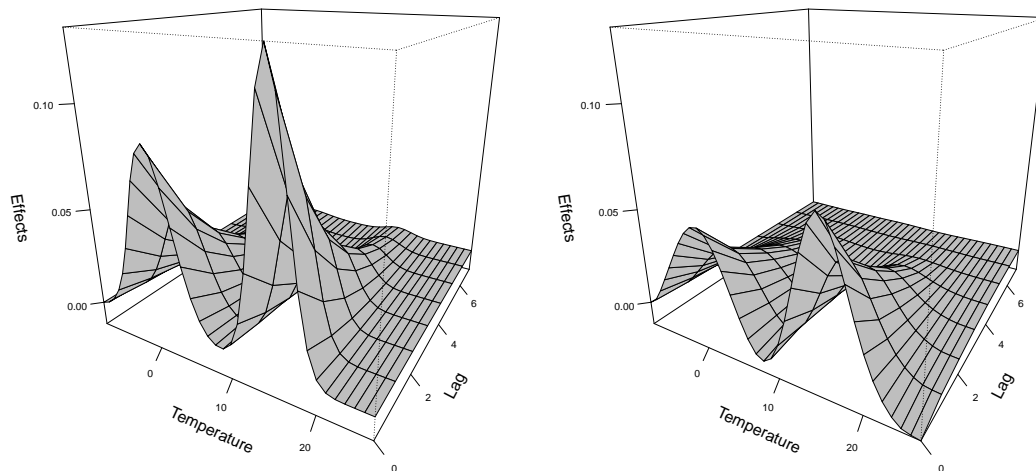

Figure 4.16: Independent synthetic covariate: Simulated surface $s(x, l)$ (left) and estimated surface (mean over 1000 runs) $\hat{s}(x, l)$ via a smoothed but not shrinked estimation (right, Setting B (ind)).

Figure 4.17: Independent synthetic covariate: Simulated surface $s(x, l)$ (left) and estimated surface (mean over 1000 runs) $\hat{s}(x, l)$ via a smoothed and shrinked estimation (right, Setting A (ind)).

estimation setting.

4.4.6 Confidence Intervals

We examine coverage properties in order to judge confidence intervals of the proposed estimation method. For computational reasons, (see Section 4.4.8) we examine in detail a penalized version with less knots than the optimal estimation (Setting A) with nevertheless acceptable mean squared errors and a good result based on an optical judgement. We examine an estimation with inner knots at $1, 3, 5$ in lag direction and at $x = -8, -6, \ldots, 22, 24$ in covariate direction. The reduction of coefficients from $37 * 9 = 333$ to $20 * 7 = 140$ coefficients reduces the computational costs at round 75% coming along with an enlargement of the summed squared errors from 0.00054 to 0.00069. The left plot of Figure 4.19 shows the already known true surface and the right plot shows the estimation (averaged over 1000 runs). Black color in the grid indicates coverage problems of the estimations. It means that these points of the grid reach lower than nominal coverages. In more than 5% of the cases, the true parameter does not lie in the confidence interval. Overall, we observe a mean coverage of 92% (mean over 1000 runs for 36*8=288 grid points of the temperature - lag surface). We observe coverage problems at 80 of the 288 grid points. In these 80 grid points, more than the nominal 5% of the runs, thus more than 50 of the 1000 runs yield intervals that do not contain the corresponding grid point of the simulated surface. We compare the coverage performance of our smoothed and shrinked estimation
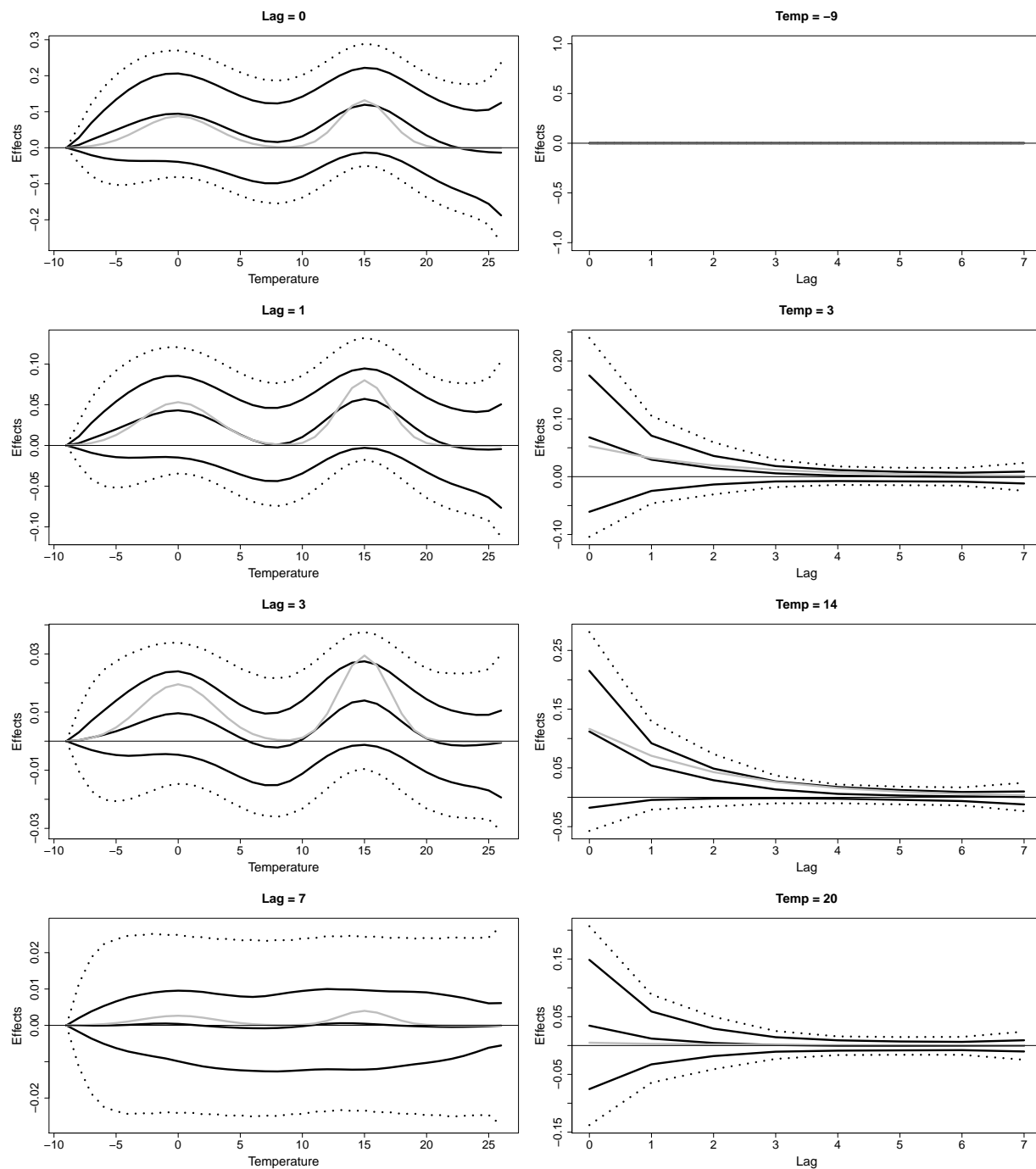
Figure 4.18: Simulation (mean over 1000 runs): Specific estimated temperature and lag effects (black), confidence intervals (dotted) and bootstrap intervals in comparison to the true underlying effects (grey) for Setting B.
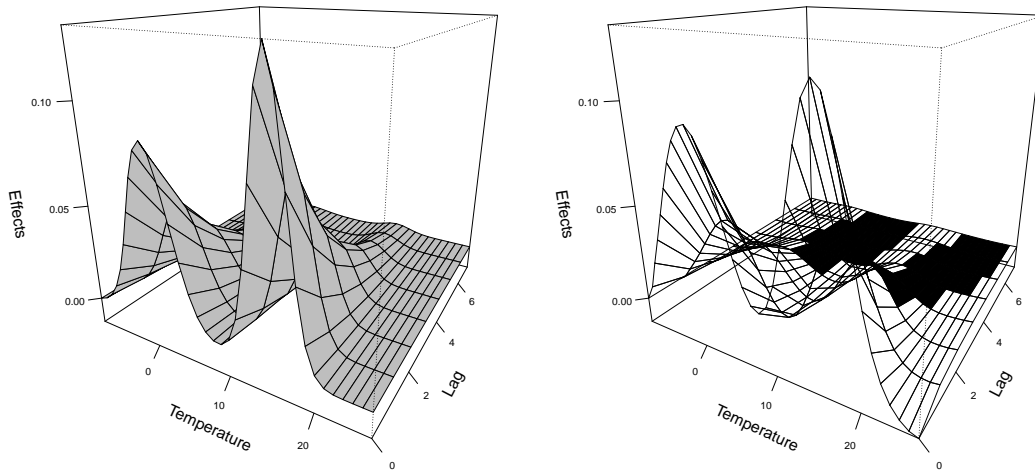
Figure 4.19: Simulated surface $s(x,l)$ (left) and estimated surface (mean over 1000 runs) $\hat{s}(x,l)$ via a smoothed and shrinked estimation (right) (inner knots $= 1, 3, 5$ in lag direction and $-8, -6, \ldots, 24$ in covariate direction). Black color in the right a plot indicates coverage problems with a lower coverage than nominal. Mean coverage $= 92\%$ on a grid of $36 * 8 * 1000$ points, maximal undercoverage rate $= 49.4\%$.
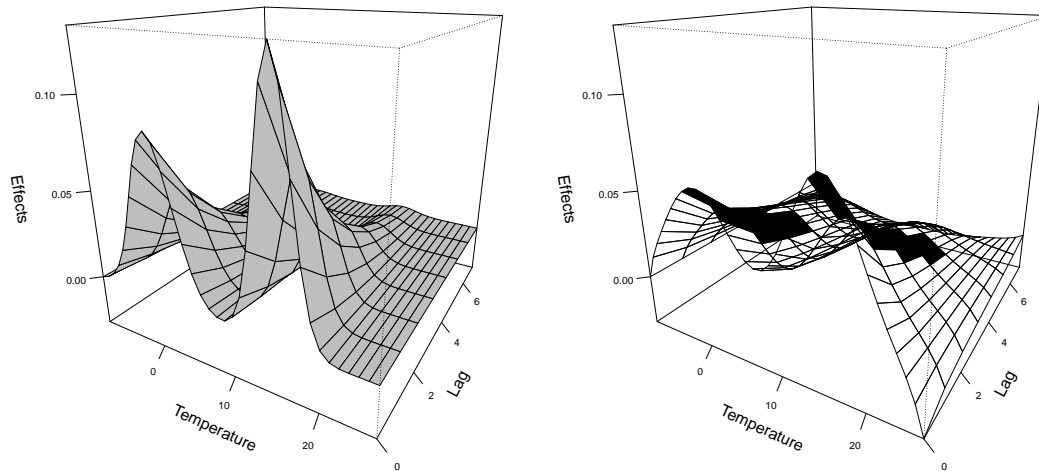
Figure 4.20: Simulated surface $s(x, l)$ (left) and estimated surface (mean over 1000 runs) $\hat{s}(x, l)$ via a smoothed but not shrinked estimation (right) (inner knots = 1,3,5 in lag direction and $-8, -6, \ldots, 24$ in covariate direction). Black color in the right a plot indicates coverage problems with a lower coverage than nominal. Mean coverage = 97% on a grid of 36*8*1000 points.

setting to the smoothed and not shrinked one. Figure 4.20 shows on the left side the true simulated surface and on the right side the averaged estimation over 1000 runs. Black color indicates grid points of critical coverage values. The mean coverage is with 97% higher than the nominal coverage of 95%. Thus, the interval size is larger than necessary in the not shrinked version. Table 4.4 recapitulates the performance comparison for the setting (inner knots = 1, 3, 5 in lag direction and $-8, -6, \ldots, 24$ in covariate direction) with and without shrinkage. The summed squared errors for the not shrinked setting are larger than those for the shrinked setting. While the confidence intervals for the setting without shrinkage tend to result in a certain overcoverage, with confidence intervals that are larger than necessary, the setting with shrinkage suffers from a certain undercoverage with lower coverages than the nominal 95%.

We refer to Section 3.6.4, where we detected for linear lagged covariate effects, that the confidence intervals for certain regions of penalized courses are very small due to the disregard of the uncertainty of the smoothing parameters. These small intervals lead to lower than nominal coverages.

We fit bootstrap confidence intervals for the prediction grid in the following way. For each of the 1000 response vectors we do the following procedure. We first fit a generalized additive model with the crossbasis as covariate. We take the resulting fitted response vector and draw 400 Poisson distributed response variables with the fitted response vector

	smooth	smooth and shrink
Summed Squared Errors	0.00129	0.00069
Coverage CI	97%	92%
Coverage bootstrap CI	-	93%

Table 4.4: Simulation:DLNM: Comparison of the performance in the setting with and without ridge penalty (inner knots = 1,3,5 in lag direction and $-8, -6, \ldots, 24$ in covariate direction).
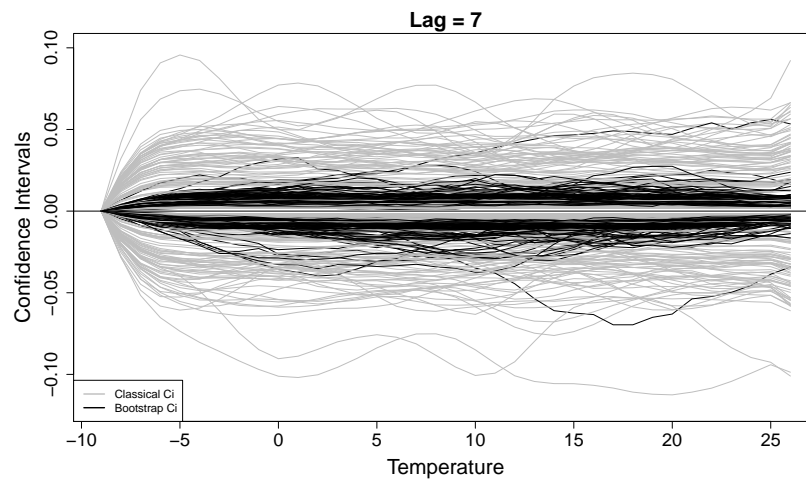

Figure 4.21: Simulation: Comparison of confidence intervals (grey) and bootstrap confidence intervals for the first 100 runs for fixed rain lag $L = 7$.

as parameter. We build the 2.5% and 97.5%-quantiles respectively as lower and upper bootstrap confidence intervals for each of the 36*8 points of the discrete prediction grid. We have a more detailed look at the performance of confidence interavals and bootstrap confidence intervals. Table 4.4 already shows that the coverage of bootstrap intervals are higher than for the "classical" ones, whereas Figure 4.18 indicates that bootstrap confidence intervals are smaller than the classical ones. Figure 4.21 shows for the first 100 of 1000 runs lower and upper interval borders (grey) and the bootstrapped borders for fixed lag $L = 7$. We see, that the variance in the confidence intervals is very large whereas the bootstrap interval borders show smaller variances. Classical confidence intervals are sometimes clearly larger than the bootstrap ones.

Nevertheless, we sporadically see interval borders close to the zero line. We learn in Section 3.6.4, that large shrinkage parameters lead to very small confidence intervals. The previously mentioned interval borders, that are close to zero, are due to large shrinkage parameters. Due to the chosen surface, where the last lag coefficient are not completely close to zero, the shrinkage parameters are comparably small (compared to Section 3.6.4) and the averaged intervals are larger than the bootstrap intervals. Even though the aver-
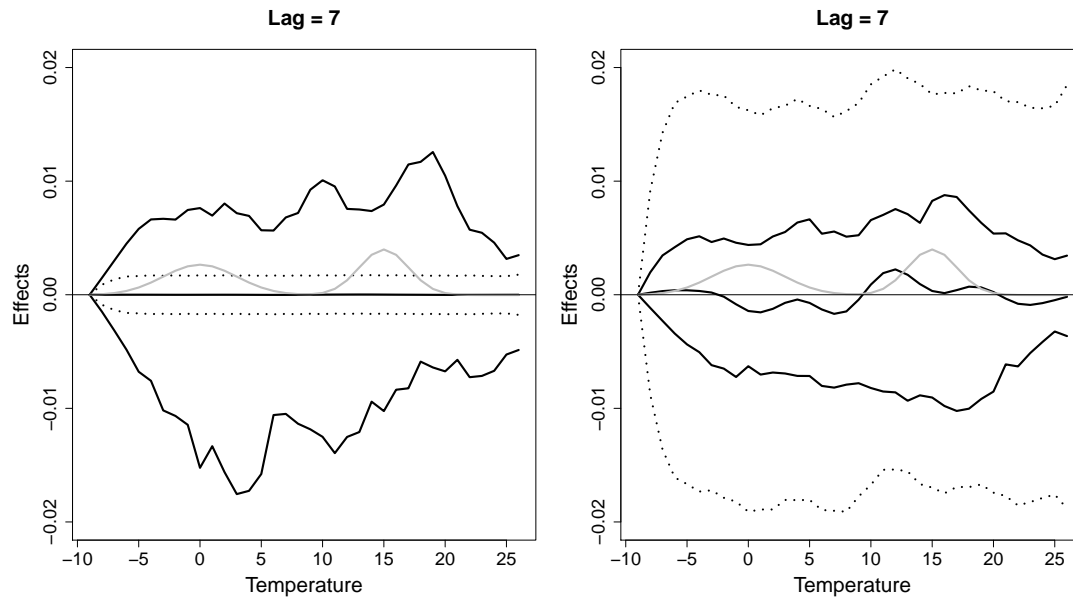
Figure 4.22: Exemplary performance of bootstrap and classical (dotted) confidence intervals in comparison to the true underlying effect (grey) for the first 2 runs for fixed rain lag $L = 7$.

aged bootstrap confidence intervals are smaller than the classical ones, we reach a better coverage (93% vs. 92%) for the bootstrap intervals.

Figure 4.22 shows for fixed lag $L = 7$ intervals and bootstrap intervals of arbitrarily chosen two runs in comparison to the true underlying lag course. The left side shows a run, where the confidence intervals are very small, the right side shows a run, where the confidence interval is much larger than the bootstrap interval. We see, that the interval sizes for the classical confidence intervals are quite different. While for the first run (left), we observe very small classical confidence intervals that do not contain the true effects for the local maxima, we observe very large classical confidence intervals for the temperature effects at the maximum lag $L = 7$ for the second run. The bootstrap intervals for both runs show similar intervals sizes for the two runs and contain in both runs the true covariate effects.

4.4.7   Simulation Results for a Binary Response Variable

In the previous sections, we show simulation results for the performance of penalized DLNMs for Poisson distributed response variables. In Section 4.5, we analyze the influence of lagged rain effects on a binary response variable, the occurrence of *block and ash flows* at volcano Merapi in Indonesia. In preparation to the data analysis, we want to check out, if penalized DLNMs can detect similarly the shape of simulated lag covariate surfaces for binary response variables. To do so, we simulate, similar to Equation (4.4) to (4.6), a
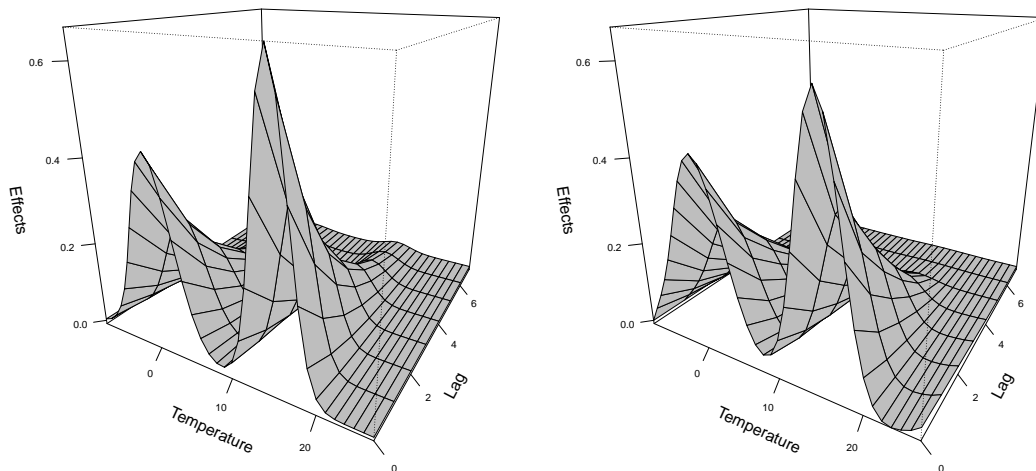
Figure 4.23: Binary response variable (mean over 1000 runs): Simulated surface $s(x,l)$ (left) and estimated surface $\hat{s}(x,l)$ via a smoothed and shrinked estimation (right, Setting A).

predictor and resulting response variables. For the lag covariate surface, we use the same shape as in Equation (4.3) with all values multiplied by the factor 5 in order to generate strong varying probabilities. Probabilities and response variables are modeled via:

$$\eta_t = \alpha_0 + \sum_{l=0}^{L=7} s(x_{t-l}, l) \tag{4.13}$$

$$\pi_t = \frac{\exp(\eta_t)}{1 + \exp(\eta_t)} \tag{4.14}$$

$$y_t \sim B(\pi_t) \tag{4.15}$$

1000 response variabels are simulated. We estimate the simulated lag covariate surface by means of smoothed and shrinked estimations and the use of many knots (Setting A). Figure 4.23 shows the true simulated surface left side and estimations averaged over 1000 runs. We see, that bimodal shape of the response variable can be discovered quite well in the proposed scenario.

4.4.8   Running Times

We calculate penalized DLNM models via the function *gam* of the package *mgcv*. In the simulation study, the calculation of one gam with $37 * 9$ penalized coefficients (Setting A) for $T = 5096$ and $L = 7$ lasts about 90 seconds (Intel(R) Core(TM) 2 Quad CPU Q 9550

2.83 GHz with Windows XP 32-bit). The calculation of 400 gams for the bootstrap refits for 1000 simulation run poses a considerable computational effort. The reduction to inner knots at $1, 3, 5$ in lag direction and at $x = -8, -6, \ldots, 22, 24$ in covariate direction reduces the length of time of one gam to 20 seconds.

4.5   Example: Merapi Data

We model a lagged nonlinear influence of rain on the occurrence of block and ash flows (BAF) in the Merapi data set (see Section 1.1.4). The response variable is binary. Accordingly, we model a logistic regression with a penalized cross-basis containing the rain effect. We use cubic B-Splines in covariate and in lag direction. We model the expectation π via the link function g:

$$E(y|..) = P(y = 1|..) = \pi \tag{4.16}$$

$$g(\pi_t) = \log(\frac{\pi_t}{1 - \pi_t}) = \zeta_t = \alpha_0 + s(x_t, \boldsymbol{\gamma}) + f(t). \tag{4.17}$$

The predictor ζ_t contains the tensor product describing the rain-lag interaction defined as in Equation (4.1). The basis coefficients $\boldsymbol{\gamma}$ are smoothed and shrinked as described in Sections 4.3.1 and 4.3.2. $f(t)$ describes a nonlinear time trend. Figure 4.24 shows a smoothed and shrinked modeling of a DLNM with knots in covariate direction at $5, 10, \ldots, 105$ and at every second lag for a maximum lag length of $L = 20$ days. We use a second degree difference penalty and estimate the smoothing parameters via REML. We center the surface around the line $x_0 = 0$. That means, we assume, that there is no interaction between rain and lags for the specific rain amount of 0 liters. We interpret estimated effects by means of appropriate sections of Figure 4.24, see Figure 4.25. The left column shows estimated rain effects for fixed lags and their classical (dotted lines) and bootstrap confidence intervals. The bootstrap intervals are generated as described in Section 4.4.6 with 400 bootstrap replications. For earlier lags, e.g., for lag 5 (first column, second row), the models predict increased effects of rain until an amount of 40 liters compared to the reference value $x_0 = 0$ and decreasing but positive effects from 40 to 60 liters in comparison to the reference value $x_0 = 0$ (no rain). From 60 liters on, bootstrap and classical confidence intervals become very large, probably due to the little numbers of days with rain amounts larger than 60 liters (see also Figure 1.5). For the last lag coefficient, ($L = 20$, first column, third row), estimated rain effects are very small with small bootstrap intervals; the uncertainty of the smoothing parameter sees to be very small. The small smoothing parameter $\lambda_r = 15.2$ indicates in this case, that the lag length is definitely sufficient. The interpretation of the confidence intervals is only reasonable, if the centering value is a good choice. The interpretation of the lag effects for fixed rain values - under the reserve, that the assumption of the reference value is correct - indicates decreasing rain effects with increasing lags. Figure 4.26 shows the estimated time trend $f(t)$. The trend indicates, that the rain effect is not the only factor that influences the occurrence of block and ash flows. We analyze the effect of rain
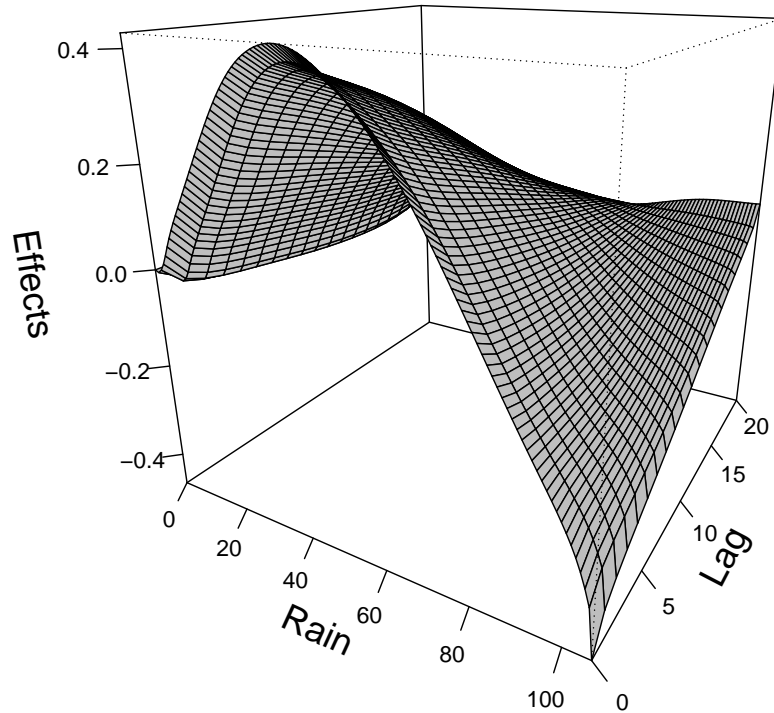
Figure 4.24: Merapi data: Estimated rain - lag effects on the binary response variable *block and ash flows.*

on the response variable but of course we now, that rain can only be one of many complex but unknown factors that can cause volcanic activity.

4.6   Example: Geothermal Data

In the geothermal data (see Section 1.1.5), we are interested in the relationship between detracted energy, built from the amount of extracted earth fluids and its respective re-injection temperature and the occurrence of micro earthquakes. Both covariates, *energy* and *pressure* shall be involved into the regression context with lagged effects and both covariates are assumed to have a nonlinear influence on the response variable. Accordingly, we use a penalized DLNM approach as proposed in Section 4.3. We use a large number
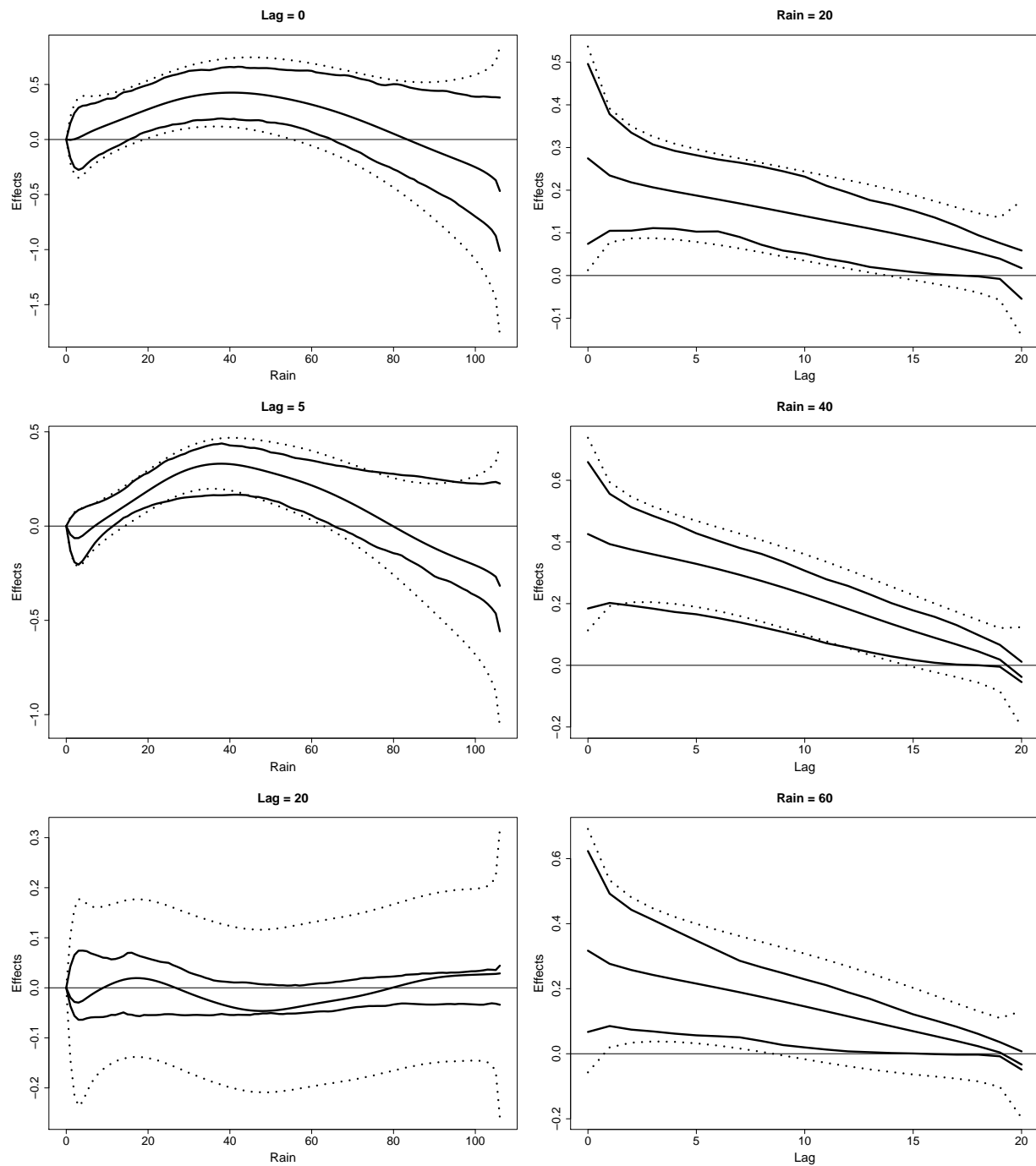
Figure 4.25: Merapi data: Estimated rain - lag effects, confidence intervals (dotted) and bootstrap intervals on the binary response variable *block and ash flows*.
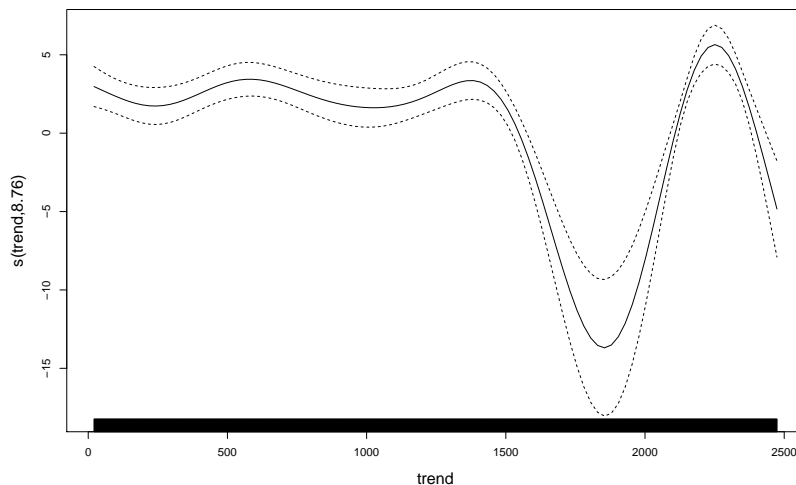
Figure 4.26: Merapi data: Nonlinear time trend f(t).

of knots and penalize basis coefficients in order to smooth and shrink them. In agreement with our consultative geophysicists, we chose a reference value of 5400 for the extracted energy and of 0 for pressure. For these covariate values, the effect on the response variable is supposed to be constant over the lags. We chose a low reference value of 5400 (instead of a reference value of $x_0 = 0$) for the extracted energy, as we suspect a direct short-term effect after a stop of energy extraction. We take the reference value of 5400, as this corresponds to a cooling down of $30°$ and a rate of 180 mg/h as a moderate value, taken in many geothermal generating plants where no quakes are observed.

The response variable can be simplified by reduction to a binary response variable *Activity*, indicating whether seismic activity was observed or not. This procedure leads us to a common logistic regression model, similar to the previous subsection. It can be written as:

$$E(y|..) = P(y = 1|..) = \pi \tag{4.18}$$

$$g(\pi_t) = \log(\frac{\pi_t}{1 - \pi_t}) = \zeta_t = \alpha_0 + s(x_t, \boldsymbol{\gamma}) + s(z_t, \boldsymbol{\gamma}) + f(t) \tag{4.19}$$

with $Y = 1$ if *Activity = yes*. The predictor ζ_t contains the tensor product describing the energy-lag interaction and the tensor product describing the pressure-lag interaction, both defined as in Equation (4.1). The basis coefficients $\boldsymbol{\gamma}$ are smoothed and shrinked as described in Sections 4.3.1 and 4.3.2 for both possible covariates. $f(t)$ describes a nonlinear time trend. Figure 4.27 shows the estimated surface of the lagged energy effects on the binary response variable. According to our model, the chance for seismic activity increases with increasing extracted energy rates, compared to a reference value of 5400 within the first hours (about 50). Figure 4.28 shows estimates and confidence intervals for fixed lags
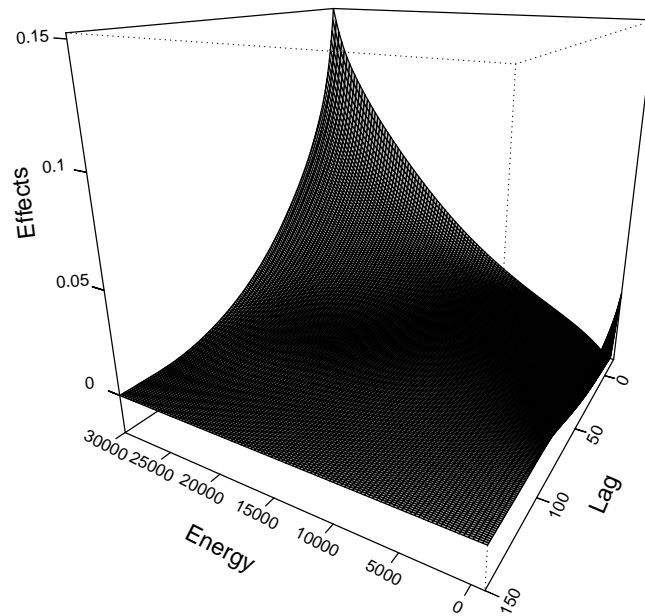
Figure 4.27: Geothermal data: Estimated energy - lag effects on the binary response variable *activity*.

(left) and for fixed energy rates (right). According to the confidence intervals, the short-term effect of an interruption of the energy abstraction cannot be proved. Concerning the pressure effect, we could not detect any significances in our model. (figure not shown). Figure 4.29 shows the estimated nonlinear time trend.
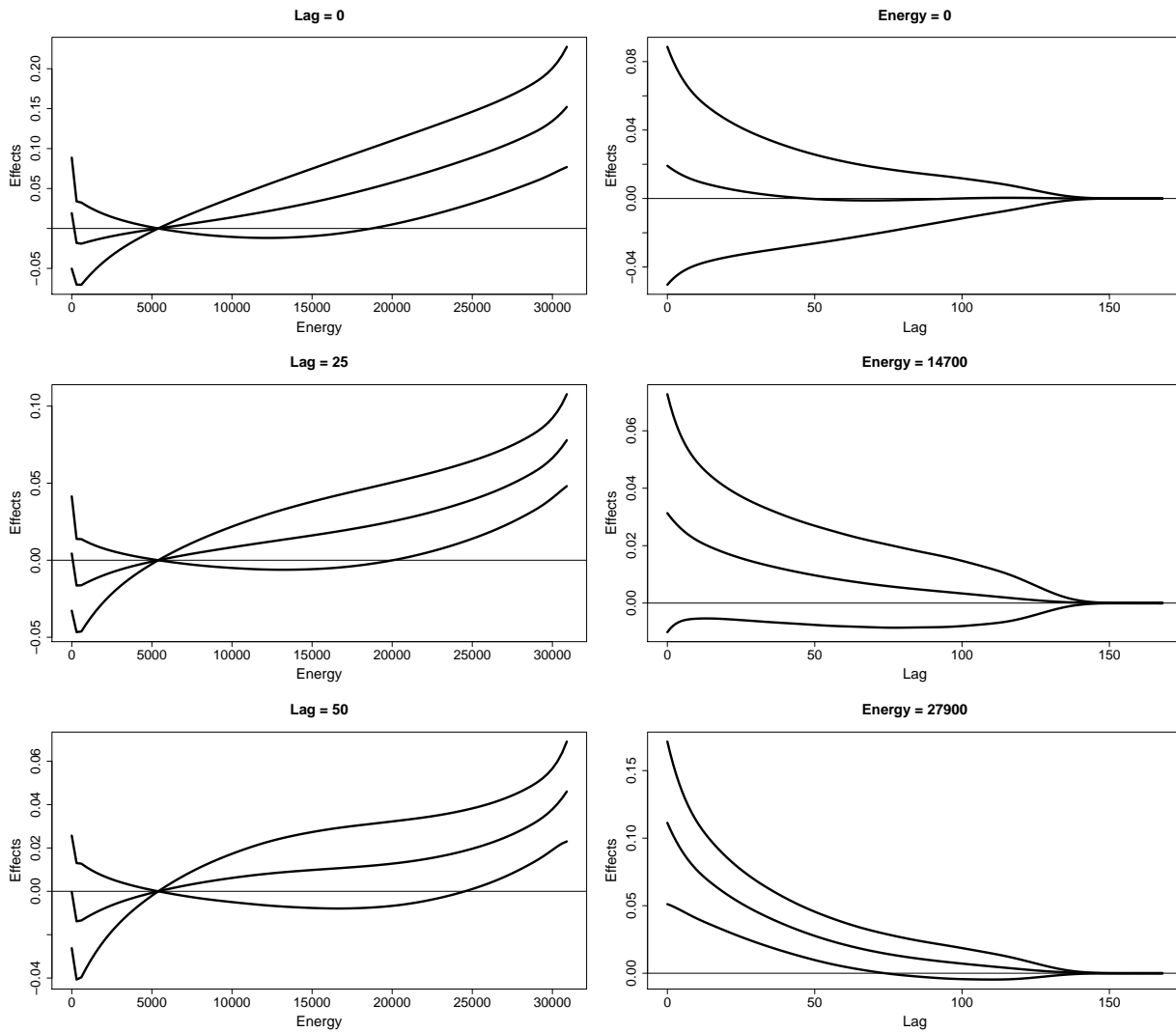
Figure 4.28: Geothermal data: Estimated energy - lag effects and confidence intervals on the binary response variable *activity*.
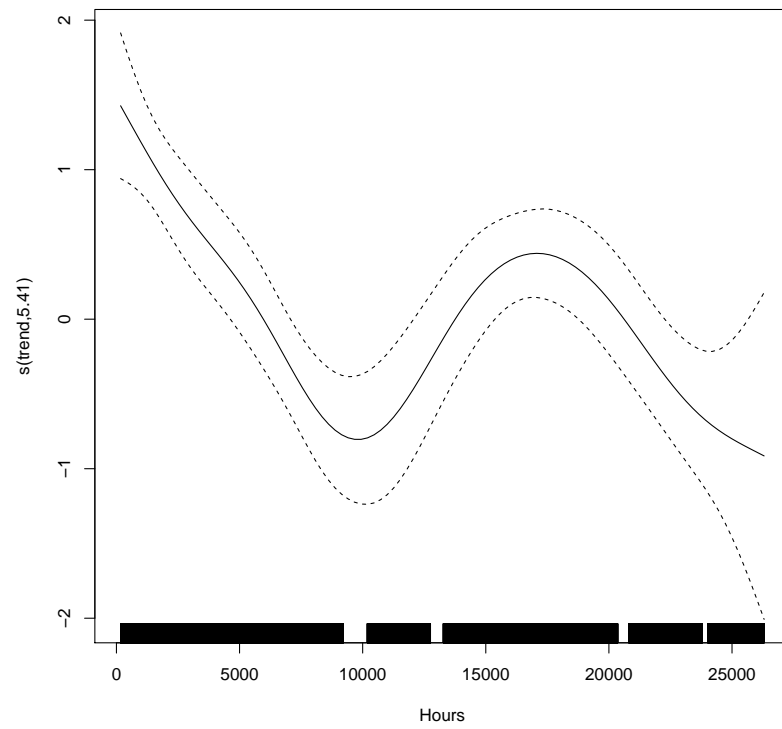
Figure 4.29: Geothermal data: Nonlinear time trend $f(t)$.

Chapter 5

Discussion

We have extended, implemented and evaluated a penalized spline based approach (Zanobetti et al., 2000) for the inclusion of lagged linear covariate effects into a regression model context. We theoretically justified an approach based on B-spline expansions and penalties motivated by reasonable assumptions of smoothness over time and decaying influence of the highest lags. In a detailed simulation study, we demonstrated the very good performance characteristics of the proposed flexible distributed lag model in comparison to the well known Almon polynomial method. The embedding in *mgcv* guarantees the use of established stable and well-tested algorithms as well as the inclusion of the distributed lag covariate in complex additive models for generalized responses, as seen in our application on the Hochstaufen earthquake data.

A major advantage of our approach is that the additional ridge penalty reduces sensitivity of results to the maximal lag length. We propose the use of longer lags as the simulation studies indicate that superfluous lags are correctly estimated close to zero. On the other hand, we have some evidence that a too short lag length can be discovered. The user has the possibility to choose longer lags if $\hat{\beta}_L$ cannot be shrunk towards zero despite the additional ridge penalty.

In the Hochstaufen data set, we modeled the lagged rain covariate via our proposed FDL approach including an interaction of the FDL term with the depth categories. The resulting lag curves in the different depth categories support the theory of a larger delay of rain effects with increasing depth. The FDL approach suggests to use a maximal number of lag units of $L = 40$ days to be included in the model.

In our simulation study, we compared our methods (FDL and PLM) to the use of Almon polynomials. Almon polynomials enjoy great popularity in the areas of econometric and medical research. The idea that the lag coefficients lie on a polynomial of lower degree is easy to understand even for non-statisticians.

The approaches of Gasparrini et al. (2010) and Zanobetti et al. (2000) represent an advance in comparison to the Almon polynomials. Gasparrini et al. (2010) use locally defined B-splines with different numbers of knots for the modeling of nonlinear and linear lagged covariate effects. These B-splines are unpenalized. Complex model selection procedures exist for non-penalized models (Baccini et al., 2007). A thorough comparison of penalized

and non-penalized models is beyond the scope of that thesis.

We advance the approach of Zanobetti et al. (2000) with an additional shrinkage penalty. Further differences to Zanobetti et al. (2000) can be seen in the REML-based smoothing parameter selection, the use of B-splines and the validation through a simulation study. In addition to the double penalty, we see the benefit of our approach in the embedding in *mgcv*, which allows the inclusion of lagged covariate effects in complex additive models for generalized responses.

Muggeo (2008) and Muggeo (2010), respectively, propose smoothing of the lag curves and an additional ridge penalty within a segmented regression framework. Muggeo limits his method to Poisson log linear regression to model the v-shaped temperature effect on mortality. Nondecreasing known weights $\nu_1, \nu_2, \ldots, \nu_L$ with $\nu_l = l^2$, e.g., control the shrinkage of the lag curve. In our approach we explicitly shrink the last lag coefficient β_L towards zero while the remainder of the lag course is estimated without assumptions of monotonicity and without the need to manually specify a vector of hard to elicit control parameters. The first lag coefficients can take any smooth form that corresponds to the data.

Future research should focus on further developments of the proposed FDL approach. The ridge penalty can give a good orientation concerning the maximum lag length L. A next step could be the automatic determination of the optimal lag length. Furthermore, we mentioned coverage problems due to the penalization and to a great extent due to the ridge penalty. We solve the coverage problem via Bootstrap confidence intervals, which are computationally intensive and not directly implemented. Future research should focus on the implementation of suitable intervals. Further simulation studies might assess the inferential properties of the model in more general cases.

In our FDL-approach, we provided a general, easily applicable tool for reliable estimation of lagged linear covariate influence in the generalized additive regression context while most previous proposals are implemented only for specific regression models.

We transfered the penalization concept to nonlinear covariate effects by enlarging an already existing concept of tensor product splines, the distributed lag nonlinear models provided in Gasparrini et al. (2010) and Gasparrini and Armstrong (2013) .

Simulation studys indicate that the penalization concept increases the performance quality of that approach. The smoothing penalty enables the use of more knots in lag and in covariate direction and enlarges the possibilty to detect complex covariate lag surfaces.

The simulation studies indicate, that the additional ridge penalty can stabilize resulting estimation surfaces and can improve the performance of distributed lag non-linear models even in the case of correlated data. The effects of synthetically generated independent covariates benefit from an additional shrinkage, that stabilizes the borders of the estimation and thereby stabilizes the complete covariate lag surface.

Nevertheless, the choice of the centering value is a critical point within the approach of Gasparrini et al. (2010) and accordingly alike in a penalized Gasparrini approach. If one choses a maladroit centering value, confidence intervals are falsified and the interpretation of lag effects on fixed temperature values can be misleading.

Future research should focus on the question, how to overcome this problem.

In the Merapi data set and in the geothermal data set, we modeled lagged nonlinear covariate influences on binary response variables via our penalized DLNM approach and produced plausible shapes of lagged covariate surfaces.

The proposed penalizing concept enables flexible and smooth estimation of linear and nonlinear lagged covariate effects for generalized responses.

Appendix A

Embedding in MGCV

A Flexible Distributed Lag model can be calculated via a *gam* with the R package *mgcv* (Wood, 2013a). Some short preparation steps are necessary. First of all, the class of constructor functions for smooth terms in a gam is extended to the *fdl* (Flexible Distributed Lag) class:

```
> smooth.construct.fdl.smooth.spec<-function(object,data,knots){
+
+      #modify object so that its fitted as a p-spline signal regression term:
+      object$bs <- "ps"
+      object <- smooth.construct.ps.smooth.spec(object,data,knots)
+
+      if(!is.null(object$xt$ridge) && object$xt$ridge){
+          #add ridge penalty to first <order of B-spline>+1 (=m+2) basis functions
+          # penalty = coef (lambda_1*DiffPen + lambda_2*RidgePen) coef
+          object$S[[2]] <- matrix(0, object$bs.dim, object$bs.dim)
+          object$S[[2]][cbind(1:(object$m[1]+2), 1:(object$m[1]+2))] <- 1
+          object$rank <- c(object$rank, object$m[1]+2)
+      }
+      return(object)
+
+ }
>
```

A Flexible Distributed Lag can be called via a known *gam* with basis *bs="fdl"*:

```
gam1 <- gam(y ~ 1 + s(lags, K=15, by=X, bs="fdl",
xt=list(ridge=TRUE), data=simul, family="poisson", method="REML")
```

with $X \mathrel{\hat{=}}$ the lagged covariate matrix (Equation (2.10)) and $lags \mathrel{\hat{=}}$ a $(T \times (L+1))$ matrix indicating the lags l with the identical rows $(-L, \ldots, 1, 0)$ and $K =$ the number of knots. The *by*- argument ensures the matrix multiplication of the resulting B-Spline basis matrix with X_{lag} (see *linear.functional.terms*, package *mgcv*).

Bibliography

Almon, S. (1965). The Distributed Lag Between Capital Appropriations and Expenditures. *Econometrica 33*(1), 178–196.

Baccini, M., A. Biggeri, C. Lagazio, A. Lertxundi, and M. Saez (2007). Parametric and semi-parametric approaches in the analysis of short-term effects of air pollution on health. *Computational Statistics & Data Analysis 51*, 4324–4336.

Barclay, J., J. E. Johnstone, and A. J. Matthews (2006). Meteorological monitoring of an active volcano: Implications for eruption prediction. *Journal of Volcanology and Geothermal Research 150*(4), 339–358.

Bayerstadler, A. (2010). *Analyse potenzieller Einflußgrößen für das Auftreten pneumologischer Erkrankungen und Entwicklung eines zeitlich-räumlichen Prognosemodells*. Diploma Thesis, Ludwig-Maximilians-Universität München.

Carlin, B. and T. A. Louis (2000). *Bayes and Empirical Bayes Methods for Data Analysis*. Boca Raton: Chapman & Hall.

Craven, P. and G. Wahba (1979). Smoothing Noisy Data with Spline Functions. *Numerische Mathematik 31*, 377–403.

Currie, I. D., M. Durban, and P. H. C. Eilers (2004). Smoothing and forecasting mortality rates. *Statistical Modelling 4*(4), 279–298.

De Boor, C. (1978). *A Practical Guide to Splines*. Berlin: Springer.

Dominici, F., A. McDermott, M. Daniels, and S. L. Zeger (2005). Revised Analyses of the National Morbidity, Mortality, and Air Pollution Study: Mortality Among Residents Of 90 Cities. *Journal of Toxicology and Environmental Health: Part A 68*(13-14), 1071–1092.

Efron, B. and R. J. Tibshirani (1993). *An introduction to the bootstrap*. Boca Raton: Chapman & Hall/CRC.

Eilers, P. H. C. and B. D. Marx (1996). Flexible smoothing with B -splines and penalties. *Statistical Science 11*(2), 89–121.

Evans, K. F., A. Zappone, T. Kraft, N. Deichmann, and F. Moia (2012). A survey of the induced seismic responses to fluid injection in geothermal and CO2 reservoirs in Europe. *Geothermics 41*, 30–54.

Fahrmeir, L., T. Kneib, and S. Lang (2007). *Regression: Modelle, Methoden und Anwendungen* (1st ed.). Berlin, Heidelberg: Springer-Verlag.

Fahrmeir, L. and G. Tutz (2001). *Multivariate Statistical Modelling Based on Generalized Linear Models* (2nd ed.). New York, Berlin, Heidelberg: Springer-Verlag.

Gasparrini, A. (2011). Distributed Lag Linear and Non-Linear Models in R: The Package dlnm. *Journal of statistical software 43*(8), 1–20.

Gasparrini, A. (2014). Modeling exposure-lag-response associations with distributed lag non-linear models. *Statistics in medicine 33* (5), 881–899.

Gasparrini, A. and B. Armstrong (2013). *dlnm: Distributed Lag Non-linear Models.* R package version 2.0.3.

Gasparrini, A., B. Armstrong, and M. G. Kenward (2010). Distributed lag non-linear models. *Statistics in medicine 29*(21), 2224–2234.

Hainzl, S., Y. Ben-Zion, C. Cattania, and J. Wassermann (2013). Testing atmospheric and tidal earthquake triggering at Mt. Hochstaufen, Germany. *Journal of Geophysical Research: Solid Earth 118*(10), 5442–5452.

Hastie, T. J. and R. J. Tibshirani (1999). *Generalized Additive Models.* Boca Raton: Chapman & Hall/CRC.

Heagerty, P. J. and B. a. Comstock (2013). Exploration of lagged associations using longitudinal data. *Biometrics 69*(1), 197–205.

Heaton, M. J. and R. D. Peng (2012). Flexible Distributed Lag Models using Random Functions with Application to Estimating Mortality Displacement from Heat-Related Deaths. *Journal of agricultural, biological, and environmental statistics 17*(3), 313–331.

Heaton, M. J. and R. D. Peng (2014). Extending distributed lag models to higher degrees. *Biostatistics 15*, 398–412.

Husen, S., C. Bachmann, and D. Giardini (2007). Locally triggered seismicity in the central Swiss Alps following the large rainfall event of August 2005. *Geophysical Journal International 171*, 1126–1134.

Kraft, T., J. Wassermann, and H. Igel (2006). High-precision relocation and focal mechanism of the 2002 rain-triggered earthquake swarms at Mt Hochstaufen, SE Germany. *Geophysical Journal International 167*(3), 1513–1528.

Kraft, T., J. Wassermann, E. Schmedes, and H. Igel (2006). Meteorological triggering of earthquake swarms at Mt. Hochstaufen, SE-Germany. *Tectonophysics 424*(3-4), 245–258.

Majer, E. L., R. Baria, M. Stark, S. Oates, J. Bommer, B. Smith, and H. Asanuma (2007). Induced seismicity associated with Enhanced Geothermal Systems. *Geothermics 36*(3), 185–222.

Marra, G. and S. N. Wood (2012). Coverage Properties of Confidence Intervals for Generalized Additive Model Components. *Scandinavian Journal of Statistics 39*(1), 53–74.

Marx, B. D. and P. H. Eilers (1998). Direct generalized additive modeling with penalized likelihood. *Computational Statistics & Data Analysis 28*(2), 193–209.

McCullagh, P. and J. Nelder (1989). *Generalized Linear Models* (2nd ed.). New York: Chapman & Hall.

Muggeo, V. M. R. (2003). Estimating regression models with unknown break-points. *Statistics in medicine 22*(19), 3055–3071.

Muggeo, V. M. R. (2008). Modeling temperature effects on mortality: multiple segmented relationships with common break points. *Biostatistics 9*(4), 613–620.

Muggeo, V. M. R. (2010). Analyzing Temperature Effects on Mortality Within the R Environment : The Constrained Segmented Distributed Lag Parameterization. *Journal of Statistical Software 32*(12), 1–17.

Muggeo, V. M. R. (2013). *modTempEff: Modelling temperature effects using time series data*. R package version 1.5.1.

R Development Core Team (2013). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Reiss, P. T. and R. T. Ogden (2009). Smoothing parameter selection for a class of semiparametric linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 71*(2), 505–523.

Roth, P., N. Pavoni, and N. Deichmann (1992). Seismotectonies of the eastern Swiss Alps and evidence for precipitation-induced variations of seismic activity. *Tectonophysics 207*, 183–197.

Rue, H. and L. Held (2005). *Gaussian Markov random fields: Theory and applications*. Boca Raton: Chapman & Hall/CRC.

Rushworth, A. M., A. W. Bowman, M. J. Brewer, and S. J. Langan (2013). Distributed lag models for hydrological data. *Biometrics 69*(2), 537–44.

Saar, M. O. and M. Manga (2003). Seismicity induced by seasonal groundwater recharge at Mt. Hood, Oregon. *Earth and Planetary Science Letters 214*(3-4), 605–618.

Shiller, R. (1973). A Distributed Lag Estimator Derived from Smoothness Priors. *Econometrica 41*(4), 775–788.

Svejdar, V., H. Küchenhoff, L. Fahrmeir, and J. Wassermann (2011). External forcing of earthquake swarms at Alpine regions: example from a seismic meteorological network at Mt. Hochstaufen SE-Bavaria. *Nonlinear Processes in Geophysics 18*(6), 849–860.

Sylvestre, M.-P. and M. Abrahamowicz (2009). Flexible modeling of the cumulative effects of time-dependent exposures on the hazard. *Statistics in medicine 28*(28), 3437–3453.

Wahba, G. (1985). A Comparison of GCV and GML for Choosing the Smoothing Parameter in the Generalized Spline Smoothing Problem. *The Annals of Statistics 13*(4), 1378–1402.

Welty, L. J., R. D. Peng, S. L. Zeger, and F. Dominici (2009). Bayesian distributed lag models: estimating effects of particulate matter air pollution on daily mortality. *Biometrics 65*(1), 282–291.

Whittaker, E. (1923). On a new method of graduation. *Proc. Edinburgh Math. Soc 41*, 63–75.

Wood, S. N. (2006). *Generalized Additive Models: An Introduction with R*. Boca Raton: Chapman & Hall/CRC.

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 73*(1), 3–36.

Wood, S. N. (2013a). *mgcv: Mixed GAM Computation Vehicle with GCV/AIC/REML smoothness estimation*. R package version 1.7-24.

Wood, S. N. (2013b). On p-values for smooth components of an extended generalized additive model. *Biometrika 100*(1), 221–228.

Wood, S. N., F. Scheipl, and J. J. Faraway (2013). Straightforward intermediate rank tensor product smoothing in mixed models. *Statistics and Computing 23*(3), 341–360.

Zanobetti, A., M. P. Wand, J. Schwartz, and L. M. Ryan (2000). Generalized additive distributed lag models: quantifying mortality displacement. *Biostatistics (Oxford, England) 1*(3), 279–292.

Eidesstattliche Versicherung

(Gemäß Promotionsordnung vom 12.07.11, § 8, Abs. 2, Pkt. 5)

Hiermit erkläre ich an Eidesstatt, dass die Dissertation von mir selbstständig, ohne unerlaubte Beihilfe angefertigt ist.

München, 25.02.2014
Ort, Datum

Viola Obermeier