
Time-varying coefficient models and measurement error

Susanne Breitner



München, 23. Februar 2007

Time-varying coefficient models and measurement error

Susanne Breitner

Dissertation

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

in Zusammenarbeit mit dem

Institut für Epidemiologie
des GSF-Forschungszentrums
für Umwelt und Gesundheit in Neuherberg

vorgelegt von

Susanne Breitner
aus Wolnzach

München, 23. Februar 2007

Erstgutachter: Prof. Dr. Helmut Küchenhoff
Zweitgutachter: PD. Dr. Annette Peters
Drittgutachter: Prof. Dr. Joel Schwartz
Rigorosum: 27.07.2007

Danksagung

An dieser Stelle möchte ich mich für die Unterstützung aller bedanken, die wesentlich zum Gelingen dieser Arbeit beigetragen haben.

Zu allererst möchte ich meine Doktoreltern Helmut Küchenhoff und Annette Peters erwähnen. Für die zahlreichen weiterführenden Diskussionen, die fachliche Unterstützung, aber auch für manches Gespräch über den Tellerrand hinaus gilt ihnen ein ganz besonderes Dankeschön.

Herrn Prof. Dr. Dr. H.-E. Wichmann, Leiter des Instituts für Epidemiologie am GSF-Forschungszentrum für Umwelt und Gesundheit, danke ich herzlich für die Möglichkeit, diese Doktorarbeit an seinem Institut anzufertigen und für die großzügige Unterstützung. My special thanks to Prof. Dr. Joel Schwartz for being the external referee as well as for the discussions and his hospitality during my stay at the Harvard School of Public Health, Boston MA.

Allen Mitarbeitern am Institut für Epidemiologie sowie am Department für Statistik an der LMU München ein Dankeschön für die angenehme Atmosphäre während der letzten Jahre. Insbesondere möchte ich mich hier bei Stefan Krieger, Matthias Schmid und Matthias Stölzel bedanken, die das manchmal zweifelhafte Vergnügen hatten, (fast) während der gesamten Zeit ihre Büros mit mir zu teilen. Matthias Stölzel und Josef Cyrus möchte ich auch für die sehr konstruktive Projektarbeit danken. Es hat wirklich Spaß gemacht! Ein herzliches Dankeschön Thomas Kneib für die vielen, sehr hilfreichen Diskussionen und, zusammen mit Michael Höhle, für die zahlreichen "Revival"-Mittagessen, die stets eine sehr amüsante Abwechslung für mich darstellten. Stefan Lang und Andreas Brezger danke ich für die Hilfestellung bei der BayesX-Implementierung. Auf keinen Fall möchte ich Alexandra Schneider, Regina Rückerl und Hanna Kirchmair unerwähnt lassen, die mir bei allen Problemen stets mit Rat und Tat zur Seite standen.

Mein Dank gilt dem Institut für Epidemiologie der GSF sowie dem Statistischen Beratungslabor des Departments für Statistik an der LMU München für finanzielle Unterstützung. Die der Arbeit zugrunde liegende Projektarbeit wurde finanziell durch das Health Effects Institute in Boston gefördert.

Nicht zuletzt möchte ich meiner Familie für die stete Unterstützung, das Vertrauen und den großen Rückhalt danken. Mein allergrösster Dank gilt meinem Freund, der mich mit viel Geduld und Liebe durch alle 'Aufs' und 'Abs' der letzten Jahre begleitet hat. Danke, dass es Dich in meinem Leben gibt und Danke für die vielen schönen gemeinsamen Momente!

Zusammenfassung

In der vorliegenden Arbeit werden Modellansätze vorgestellt und entwickelt, die einen zeitlich variierenden Effekt von Kovariablen durch die Verwendung von zeitvariierenden Koeffizienten erlauben. Die Methoden werden in ausführlichen Simulationsstudien miteinander verglichen. Dabei wird untersucht, wie gut verschiedene Komponenten in den simulierten Modellen erkannt werden. Im Anschluss werden die Ansätze, die sich in den Simulationsstudien als am Besten geeignet gezeigt haben, auf Daten angewendet, die im Rahmen der Studie "Improved Air Quality and its Influences on Short-Term Health Effects in Erfurt, Eastern Germany" über den Zeitraum von 1991 bis 2002 erhoben wurden. In dieser Analyse wird auch auf statistisches Testen bezüglich des zeitlich variierenden Einflusses eingegangen.

Des Weiteren beschäftigt sich diese Arbeit mit der Schätzung von zeitvariierenden Koeffizienten Modellen mit fehlerbehafteten Expositionsvariablen. Hierzu wird ein geeignetes Messfehlermodell spezifiziert. Zudem werden Methoden vorgestellt, die die Schätzung der Modellparameter und -varianzen unter Berücksichtigung von zeitlicher Korrelation in der latenten Variablen sowie in den Messfehlern erlauben. Im Anschluss werden zwei Korrekturverfahren entwickelt. Eine dieser Methoden basiert dabei auf einem hierarchischen Bayesianischen Modell und einem Bayesianischen Messfehler-Korrekturansatz. Außerdem wird das im Zusammenhang mit messfehlerbehafteten Kovariablen gängige Verfahren der Regressionskalibrierung für den Fall von zeitlicher Korrelation erweitert. Beide Korrekturansätze werden zum Schluss wiederum auf Daten aus Erfurt angewendet und verglichen.

Abstract

This thesis is concerned with presenting and developing modeling approaches which allow for a time-varying effect of covariates by using time-varying coefficients. The different approaches are compared in simulation studies. Thereby, we investigate how well different components of the simulated models can be identified. The models performing best in the simulation study are then applied to data collected within the study "Improved Air Quality and its Influences on Short-Term Health Effects in Erfurt, Eastern Germany". One specific aspect in this analysis is to assess the necessity of a time-varying estimate compared to a more parsimonious, time-constant fit.

A further topic is the estimation of time-varying coefficient models in the presence of measurement errors in the exposure variable. We specify a measurement error model and present methods to estimate parameters and measurement error variances of the model in the case of autocorrelated latent exposure as well as measurement errors. Furthermore, two methods adjusting for measurement errors in the context of time-varying coefficients are developed. The first one is based on a hierarchical Bayesian model and the Bayesian error correction principle. The second method is an extension of the well-known regression calibration approach to the case of autocorrelated data. The obtained estimated true values can then be included into the main model to assess the effect of the variable of interest. Finally, the approaches are again applied to the Erfurt data.

Contents

1	Introduction	1
1.1	Background and Motivation	2
1.1.1	Historical background and important findings of time series studies on air pollution and mortality	2
1.1.2	A motivating example	4
1.1.3	Time-varying coefficient models in air pollution analysis	5
1.1.4	Time-varying coefficient models and measurement error	5
1.2	Overview of the thesis	7
2	Generalized models	9
2.1	Generalized linear models	9
2.2	Generalized additive models	11
2.3	Varying coefficient models	13
2.4	Time-varying coefficient models	13
3	Modeling approaches for TVC	15
3.1	TVCM using regression splines	15
3.2	Bayesian TVCM with P-splines	17
3.2.1	Observation model	17
3.2.2	Prior assumptions	17
3.2.3	MCMC inference	24
3.3	TVCM with penalized linear splines	25
3.3.1	Mixed model representation	26
3.3.2	Estimation	28
3.3.3	Estimation algorithm	32
3.4	TVCM based on empirical Bayes inference	32
3.4.1	Observation model and prior assumptions	33
3.4.2	Mixed model representation	33
3.4.3	Empirical Bayes Inference	34

3.5	Adaptive generalized TVC models	36
3.5.1	Estimation	37
4	Simulation Study	43
4.1	Simulation setup	43
4.2	Results	49
5	Analysis of the Erfurt data	61
5.1	Data and confounder model description	61
5.2	Results	67
5.2.1	Sensitivity within MCMC estimation	73
5.2.2	Time-varying effect or dose-response?	74
5.2.3	Inference	78
5.3	Discussion of the (time-varying) air pollution effects	79
6	A measurement error model for TVCM	83
6.1	Measurement error model formulation	83
6.1.1	Model for the latent exposure variable	84
6.1.2	Disease model	85
6.1.3	Measurement model	85
6.2	Identifiability of parameters	86
6.3	Estimating the measurement error model parameters	87
6.3.1	Simple Estimation	87
6.3.2	Models allowing for an AR structure of the observations	88
7	Measurement error correction methods	95
7.1	A Bayesian measurement error correction model	95
7.1.1	Model formulation	96
7.1.2	Prior specifications for model parameters and variances	98
7.1.3	MCMC Estimation	101
7.1.4	Sampling Algorithm	108
7.2	A measurement error calibration model	110
8	Measurement error analysis of the Erfurt data	113
8.1	Description of the data sources	113
8.2	Application of the measurement error models	116
8.2.1	Estimating the model parameters and measurement error variances	117
8.2.2	Results of the correction methods - time-varying effect of $PM_{2.5}$.	120
8.2.3	Conclusion	122

9	Summary and Outlook	125
A	Time-varying coefficient models and measurement error analysis with BayesX	129
A.1	Getting started	129
A.2	Dataset objects	130
A.3	Bayesreg and remlreg objects	131
A.4	Time-varying coefficients and measurement error	133
A.5	Post-estimation commands and visualization of the results	134
A.5.1	Post-estimation commands	134
A.5.2	Visualization of the results	134
A.6	Convergence diagnostics	135
B	Proof of the ARMA(p,p) representation of AR(p) models plus white noise	139
	Bibliography	141

List of Figures

3.1	Illustration of B-spline basis functions of degree 1 (top), degree 2 (middle) and degree 3 (bottom). Each Figure exemplarily shows the different polynomials of one B-spline.	20
3.2	Trends induced by first (top) and second order (bottom) random walk priors.	21
4.1	Simulated effect of $f_{CO}(t)$ was curvilinear (top) or followed a cosine function (bottom).	45
4.2	Simulated effect of the confounder trend $f(t)$ (upper panel), of temperature $f(temp)$ (middle panel) and simulated time-varying effect of temperature $f_{temp}(t)$ (lower panel) for the example CO	46
4.3	Case 1): RMSE of $f_{CO}(t)$ (top) and averaged estimates (bottom) for the example CO	54
4.4	Case 2): RMSE of $f_{CO}(t)$ and averaged estimates (bottom) for the example CO	55
4.5	Case 3): RMSE of $f_{CO}(t)$ and averaged estimates (bottom) for the example CO	56
4.6	Case 4): RMSE of $f_{CO}(t)$ (top left) and averaged estimates (top right), and RMSE of $f(t)$ (bottom left) and averaged estimates (bottom right), for the example CO	57
4.7	Case 5): RMSE of $f_{CO}(t)$ (top left) and averaged estimates (top right), RMSE of $f(t)$ (middle left) and averaged estimates (middle right), and RMSE of $f(temp)$ (bottom left) and averaged estimates (bottom right), for the example CO	58
4.8	Case 6): RMSE of $f_{CO}(t)$ (top left) and averaged estimates (top right), RMSE of $f(t)$ (middle left) and averaged estimates (middle right), and RMSE of $f(temp)$ (bottom left) and averaged estimates (bottom right), for the example CO	59

4.9	Case 7): RMSE of $f_{CO}(t)$ (top left) and averaged estimates (top right), RMSE of $f(t)$ (middle left) and averaged estimates (middle right), and RMSE of $f_{temp}(t)$ (bottom left) and averaged estimates (bottom right), for the example CO	60
5.1	Time series of mortality counts in Erfurt, Germany, from October 1991 to March 2002 (a), old city limits) and September 1995 to March 2002 (b), new city limits).	63
5.2	Time series of CO concentration (a) from October 1991 to March 2002, UFP number concentration (imputed) (b) and $PM_{2.5}$ mass concentration (imputed) (c) from September 1995 to March 2002.	64
5.3	Doctor's practice index for all study winters	65
5.4	Posterior mean estimates (top) and posterior mode estimates (bottom) for the nonlinear effects of trend (left panel) and the doctor's practice index (right panel). All estimates are shown with pointwise 95% credible intervals.	68
5.5	Posterior mean estimates (top) and posterior mode estimates (bottom) for the nonlinear effects of the temperature of the same day (left panel) and relative humidity with a two-days lag (right panel).	69
5.6	Time-varying effect of carbon monoxide (CO) with a four-day lag estimated using the MCMC (top) and the mixed model based EB method (bottom).	70
5.7	Time-varying effect of ultrafine particles (UFP) with a four-day lag estimated using the MCMC (top) and the mixed model based EB method (bottom).	71
5.8	Time-varying effect of fine particle mass ($PM_{2.5}$) of the same day estimated using the MCMC (top) and the mixed model based EB method (bottom).	72
5.9	Time-varying association of carbon monoxide (CO) at lag 4 with different choices of the hyperparameters: $a = b = 0.001$ (top, standard choice); $a = 0.00001, b = 0.00001$ (middle) and $a = 1, b = 0.005$ (bottom).	75
5.10	Exposure-response-relationships of CO concentrations with lag 4 (a), of UFP concentrations with lag 4 (b) and of $PM_{2.5}$ concentrations of the same day (c).	76
5.11	Exposure-response-relationship of UFP with lag 4 (top); Time series of UFP concentrations with lag 4 and time-varying association of UFP , lag 4 (middle); and scatterplot of UFP concentrations versus the product of the time-varying coefficient times the UFP concentrations (bottom).	77

8.1	Time series of the two $PM_{2.5}$ measurement devices (top) and the corresponding autocorrelation functions (bottom).	115
8.2	Autocorrelation functions (ACF) of the standardized residuals obtained by ARMA(1,1) models.	119
8.3	Autocorrelation functions (ACF) of the standardized residuals obtained by ARMA(2,1) models.	121
8.4	Time-varying coefficients of $PM_{2.5}$ series. Top row: 'Naive' time-varying coefficients of the imputed $PM_{2.5}$ series (left panel) and of the mean of the two observed $PM_{2.5}$ series (right panel). Middle row: Time-varying coefficients of $PM_{2.5}$ obtained by the (Bayesian) regression calibration model (left) and by the Bayesian measurement error model (right). Underlying assumption of ARMA(1,1) models. Bottom row: Time-varying coefficients of $PM_{2.5}$ obtained by the (Bayesian) regression calibration model (left) and by the Bayesian measurement error model (right). Underlying assumption of ARMA(2,1) models.	124
A.1	Sampling paths for the first $PM_{2.5}$ values of the Bayesian measurement error model.	136

Chapter 1

Introduction

Parametric generalized linear regression models play a fundamental role in statistics. Although these models allow for simple interpretation and estimation, the assumption of a parametric linear predictor may often be too restrictive in applications. Therefore, various non- and semiparametric extensions have been proposed.

One such extension is the class of varying coefficient models. In these models, a response variable is allowed to depend linearly on some regressors, with coefficients as smooth functions of some other predictor variables, called effect modifiers. The varying coefficient model is useful for exploring complicated non-linear interactions between covariates while avoiding the 'curse of dimensionality' problem. A special case of the varying coefficient model is given for time series data, where the effect modifier variable usually is calendar time, hence, resulting in time-varying coefficient models (TVCM). No restrictions other than smoothness are placed on the coefficient functions in order to allow for enough flexibility.

Time series data arise in a variety of applications including modeling the effects of ambient air pollution on human health. These time-series analysis may, for example, assess whether daily concentrations of ambient air pollutants are associated with daily numbers of deaths in a certain geographical region. In such applications, the common assumption of a time-constant effect of environmental pollution may be too restrictive and simplifying, and more general forms are required. It seems, therefore, natural to analyze regression models that allow for a time-varying association between response and the covariates of interest.

Furthermore, there often exist covariate measurement errors in such applications. For example, it has been well documented in the literature that exposure to pollutants is often subject to measurement errors, and covariate measurement errors may cause difficulties and complications in conducting statistical analysis.

In this introduction, we first give a short historical overview of the statistical approaches used in the context of time series studies of air pollution and mortality. Additionally, the most important findings regarding the association between air pollution and mortality are resumed. We further present the motivating example for this thesis, a study conducted in Erfurt, Germany over a 10.5-year period, and give an overview over the extensions and topics considered in this thesis.

1.1 Background and Motivation

The potential health effects of ambient air pollution are a major public health issue that has received much attention over the past several decades. Results of many epidemiological time series studies have suggested an association between air pollution and daily mortality. The relative risks of the observed effects are small; however, as exposure affects a large population, the public health impact is substantial.

1.1.1 Historical background and important findings of time series studies on air pollution and mortality

Much of the concern about the health effects of air pollution originated from dramatic and severe air pollution episodes. Three prominent episodes occurred in Meuse Valley, Belgium, in 1930 (Friket (1931)); Donora, PA, USA, in 1948 (Schrenk et al. (1949); Ciocco & Thompson (1961)); and London, England, in 1952 (Ministry of Health 1954). For these episodes, analysis consisted of a simple comparison of death counts for several days or weeks before, during, and after the episodes. Nevertheless, exposures to high peak concentrations of air pollutants such as sulfur dioxide (SO_2) and total suspended particles (TSP) were consistently found to be strongly associated with mortality.

"In the 1970s and 1980s, research on air pollution and health largely involved the use of cross-sectional designs comparing the morbidity and mortality rates across cities with different ambient air pollutant concentrations" (Bell, Samet & Dominici (2004)). Further, researchers used linear models with transformations or linear autoregressive models (Özkaynak & Spengler (1985)). Adjustment for confounding effects was mainly done by using moving averages of confounder variables or stratification. Despite the different analysis methods as well as different measures of air pollutants, generally these studies found associations between air pollution and mortality, with higher levels of mortality accompanying higher levels of air pollution (Bell et al. (2004)).

At the beginning of the 1990s, several studies used more advanced statistical modeling techniques. The primary statistical approach was formal time-series modeling of mortality data using Poisson regression. These models further allowed for potential confounders such as trend, season, and meteorological variables using trigonometric functions, polynomials and, later, nonparametric smoothers (Fairly (1990); Schwartz & Dockery (1992); Pope, Schwartz & Ransom (1992); Schwartz (1993); Spix et al. (1993) Schwartz (1994)). Since the mid-90s, most of the studies use generalized linear models (GLM) with polynomials or natural cubic regression splines as well as generalized additive models (GAM) with nonparametric smooth functions (smoothing splines or loess smoothers), or, more recently, penalized splines as statistical approaches for time-series analysis. Alternative approaches used are case-crossover methods (Neas, Schwartz & Dockery (1999), Bateson & Schwartz (1999), Janes, Sheppard & Lumley (2005), Forastiere et al. (2005)).

Although the air pollution concentrations were much lower in the last decade compared to the 1930s to 1950s, there have been hundreds of single-city studies that have found associations between mortality and air pollutants. The most common and consistent association has been found with particulate matter (PM) (see for example Pope, Dockery & Schwartz (1995) or Health Effects Institute (2003)). But associations have also been reported with gaseous pollutants such as carbon monoxide (CO) (Burnett et al. (1998), Wichmann et al. (2000), Forastiere et al. (2005)).

Moreover, in studies where ambient particle mass concentrations below aerodynamic diameters of 10 micrometres (PM_{10} , inhalable particles which are deposited in the upper airways) as well as below 2.5 μm ($PM_{2.5}$, particles mainly deposited in the lung) were available, there were indications that $PM_{2.5}$ was more strongly associated with mortality than PM_{10} (Dockery, Schwartz & Spengler (1992); Schwartz, Dockery & Neas (1996)). Larger particles, also called coarse particles ($PM_{10} - PM_{2.5}$), showed no effect.

Currently, ultrafine particles (UFP ; smaller than 0.1 micrometres) are discussed to be another important fraction of ambient particles. However, evidence on associations of UFP number concentrations and mortality is still limited. A study conducted in Erfurt, Germany, from 1995 to 1998 showed comparable effects of $PM_{2.5}$ and UFP (Wichmann et al. (2000)). Forastiere et al. (2005) and Stölzel et al. (in press) also find significant associations between UFP and mortality.

More recently, large multi-center studies have been conducted using uniform methods for assembling and analyzing data. Examples are the *Air Pollution and Health: A European Approach (APHEA I and II)* projects (Katsouyanni et al. (2001)) and the *US National Morbidity and Mortality Air Pollution Study (NMMAPS)* (Samet et al. (2000a); Samet et al. (2000b)). These multi-city studies have confirmed the findings of earlier studies in

single cities of an association between daily mortality and elevated concentrations of PM and other pollutants. A $10 \mu\text{g}/\text{m}^3$ increase in the concentration of PM_{10} is associated with an increase in mortality by 0.27% in the US (Health Effects Institute (2003)) or by 0.6 % in Europe (Katsouyanni et al. (2002)), respectively.

A number of studies have furthermore allowed for an air pollution effect as a non-linear function (Daniels et al. (2000), Schwartz et al. (2001), Dominici et al. (2002), Samoli et al. (2005)).

Yet, time series studies relating short-term changes in air pollution levels to daily mortality counts have, so far, typically assumed that the effects of air pollution on the log relative risk of mortality are constant over time. The following section introduces a motivating example for the necessity of more flexible regression models.

1.1.2 A motivating example

Drastic improvements of air quality have been observed in Eastern Germany comparing the mid-80s to today. These improvements are mainly the consequences of the unification of East and West Germany in 1990, which brought major social and political changes particularly affecting East Germany. The restructuring process resulted in a reorganization of East German industrial structures and in the implementation of air pollution controls between 1992 and 1996. These processes further led to remarkable changes in emission sources of air pollutants. Sources that have undergone significant changes in East Germany since 1990 include energy production by power plants and local heating. Lignite and coal have been exchanged for natural gas in power plants and in domestic heating. Concurrent with the changes in energy production, mobile sources have undergone transitions. The old car fleet was gradually replaced by vehicles with modern engine technology including three-way catalysts, while the overall number of vehicles increased (Peters et al. (2007)). All these processes resulted in major changes in the amount of emissions as well as in the concentrations of measured air pollutants.

It has further been suggested that ambient concentrations of specific air pollutants may be surrogate markers of exposure to other agents or of specific pollution sources and that those are in fact responsible for the observed effects. An example are ultrafine particles (UFP) which are mainly released from vehicle exhausts. UFP are supposed to be one of the most likely causes of adverse health effects, but they are usually not measured. Instead, studies have shown effects of CO as well as nitrogen dioxide (NO_2). However, in many cities all three are being emitted by the same source, namely combustion processes and, hence, are highly correlated. This high correlation, though, makes it impossible to

distinguish their effects in epidemiological studies. Seaton & Dennekamp (2003), for example, argued that "*if NO_2 ... is measured as an index of pollution and is shown to be associated with health effects, these effects could equally be due to the numbers of (ultra-fine) particles*". The observed effects of CO and NO_2 are, therefore, often interpreted as potential surrogate effects of unmeasured ultrafine particles. This view is supported by some source apportionment analyses including gaseous pollutant measurements. However, along with changing pollution sources and emissions there could also be a changing surrogate status of the measured pollutants.

The main objective of the study "Improved Air Quality and its Influences on Short-Term Health Effects in Erfurt, Eastern Germany" (Peters et al. (2007)), which is the motivating example for this thesis, was to assess the impact of such changes in ambient air pollution on mortality during the period 1991 to 2002 in Erfurt, East Germany.

1.1.3 Time-varying coefficient models in air pollution analysis

There are only a few publications that allowed the effects of air pollution to change smoothly over time. Peng et al. (2005), for example, allowed the effect of particulate matter to change seasonally forcing the seasonal effect to be a sinusoidal function with a period of one year. This constraint ensures a smooth time-varying effect, but the size and location of the seasonal effect are forced to be constant over the years of the study. A more flexible approach is proposed by Chiogna & Gaetan (2002) who apply a dynamic generalized linear model (DGLM) based on a likelihood analysis using the iteratively re-weighted Kalman filter and smoother. They allow the effect of air pollution to change, in principle, freely over time; however, to enforce smoothness of the time-varying effect, the amount in which the effect could change is restricted by constraining it to follow a first order random walk. This approach is extended by Lee & Shaddick (2005), who replace the assumption of a first order random walk with an autoregressive process of higher order and base their inference on Markov Chain Monte Carlo (MCMC) simulation methods.

1.1.4 Time-varying coefficient models and measurement error

Measurement error models are appropriate in cases when the variable of interest ξ is not measured directly, but for which information is available through recording of a surrogate X . It is well-known that ignoring measurement errors on the explanatory variable X in regression analysis may induce bias in the effect estimates and can, therefore, be misleading.

The effect of measurement errors in predictor variables on regression analysis have been carefully studied in the statistical and epidemiological literature for several decades. Among many others, Fuller (1987) summarized early research on linear regression with so-called *errors-in-variables*. Carroll et al. (2006) provide recent illustrations of statistical approaches to measurement error in epidemiologic research.

The measurement error problem that we consider in this thesis is concerned with the situation of air pollution time-series studies. A systematic overview of measurement errors in such kind of studies is given by Zeger et al. (2000). They find three sources of measurement error to be relevant. The first one is the instrument error, that means the accuracy and precision of the monitoring instrument. The second one is the error resulting from the non-representativeness of a monitoring site (reflected by the spatial variability of the pollutant measured). The third source is given by differences between the average personal exposure to a pollutant and the true ambient pollutant level. Most of the literature in the air pollution context is concerned with errors resulting from using centrally measured data as a surrogate for personal exposure (see Sheppard & Damian (2000), for example). Dominici, Zeger & Samet (2000) proposed an approach correcting for the consequences of this kind of measurement error.

Recently, Schwartz & Coull (2003) have developed an approach that deals with multiple exposures and, under certain conditions, is resistant to measurement error. An application of this approach is presented by Zeka & Schwartz (2004).

To frame the general problem discussed here, consider a situation in which two environmental monitoring systems are collecting data on the same variable ξ_t . We define the variable of interest, ξ , as the (true) mean ambient level of an air pollutant in a city. It is further assumed that ξ_t may be temporally correlated. Instead of ξ_t , one now observes perturbed versions X_{tj} from the two different measurement systems at regular time points

$$\begin{aligned} X_{t1} &= \xi_t + \epsilon_{t1} & t = 1, \dots, T, \\ X_{t2} &= \xi_t + \epsilon_{t2}, \end{aligned}$$

where ϵ_{tj} is the corresponding measurement error of the two monitoring devices. Our aim is to estimate the time-varying association between ξ and health outcomes such as mortality.

Varying coefficient models and, in particular, time-varying coefficient models without measurement error have been studied in Hastie & Tibshirani (1993), for example.

Regarding measurement error, Liang, Wu & Carroll (2003) studied a longitudinal linear model with random intercepts and slopes which are allowed to depend on time. Regres-

sion splines are used to approximate the coefficient functions, and the error correction algorithm is similar to regression calibration. Recently, You, Zhou & Chen (2006) propose a corrected local linear estimator for varying coefficient models with measurement errors.

1.2 Overview of the thesis

The outline of this thesis is as follows. The second chapter gives a short introduction into the basic concepts of generalized linear and additive models. This is followed by an overview of varying coefficient models, in general, as well as of one of their special cases, time-varying coefficient models.

In the third chapter, we present five approaches for estimating time-varying coefficients. The five approaches are: Time-varying coefficient models using regression splines; Bayesian time-varying coefficient models with P-splines; Time-varying coefficient models with penalized linear splines based on a generalized linear mixed model framework; Time-varying coefficient models with P-splines based on empirical Bayes inference; and Adaptive generalized TVC models. We will describe the models and discuss inference. The proposed methods are investigated in an extensive simulation study, which is described in Chapter 4. This is done by re-estimating various models typically used in the air pollution analysis context. The simulated response is based on real data obtained in the study "Improved Air Quality and its Influences on Short-Term Health Effects in Erfurt, Eastern Germany".

In Chapter 5, the main results of an air pollution data analysis are presented. Especially, the results of the application of those methods selected based on the simulation study are shown and contrasted. This chapter is also concerned with an extensive sensitivity analysis. We further present a bootstrap test procedure for investigating whether a time-varying coefficient term is indeed required.

Chapter 6 focusses on the formulation of a measurement error model in the context of time-varying coefficient models. As the observed time series are assumed to be temporally correlated, a further aspect is given by an appropriate selection of the latent variable's distribution. In addition, we deal with methods for estimating the measurement error model parameters in this special context.

Chapter 7 introduces two methods adjusting for measurement errors in the context of time-varying coefficients. One is a hierarchical Bayesian model based on MCMC inference. The second approach is a variant of the well-known regression calibration. Applications of these correction methods on air pollution data are illustrated in Chapter

8, including an comparison with the results obtained by the 'naive' analysis.

Finally, in Chapter 9, a summary of the presented work leads to possible directions for future research.

Chapter 2

Generalized models

The general aim of regression models is to explore an association between dependent and independent variables to identify the impact of these covariates on the response.

The classical linear regression model is an important statistical tool, but its use is limited to those settings where the normal distribution is valid and the assumption of a linear function relating the response to the predictors is given.

Generalized linear models (GLM), introduced by Nelder & Wedderburn (1972), expand the well known linear model to accommodate non-normal response variables in a single unified approach. It is common to find response variables which do not fit the standard assumptions of the linear model (normally distributed errors, constant variance, etc.). Examples are count data, dichotomous variables and truncated data. In GLMs, the influence of each covariate is modeled parametrically and the covariate effects are additively combined in the so-called linear predictor.

However, the identification of appropriate polynomial terms and transformations of the predictors to improve the fit of a linear model can be tedious and imprecise. Generalized additive models (GAM, Hastie & Tibshirani (1990)) extend generalized linear models by replacing the linear predictor with an additive sum of smooth functions. The following sections are intended to give an informative overview of generalized linear and generalized additive models. A detailed overview of these two types of models is given by McCullagh & Nelder (1989) and Hastie & Tibshirani (1990).

2.1 Generalized linear models

Let y_1, \dots, y_T denote observations on a response, measured over T independent time points. We treat y_t as a realization of the random variable Y_t , $t = 1, \dots, T$. The p

dimensional vector z_t of independent variables contains continuous or discrete values, which are either fixed by design or which are realizations of sets of random variables Z_t . A generalized linear model is composed of three parts:

- A random component specifying the response distribution:

In a generalized linear model it is typically assumed that, conditional on the covariates, the responses y_t are independent. The conditional distribution of the responses y_t follows a one-dimensional exponential family with density

$$f(y_t|z_t, \theta_t, \phi) = \exp \left\{ \frac{y_t \theta_t - b(\theta_t)}{\phi} + c(y_t, \phi) \right\}, \quad (2.1)$$

where the functions $b(\cdot)$ and $c(\cdot)$ are known and specific for the distribution of the exponential family. ϕ denotes an additional scale or dispersion parameter, whereas θ_t is the so-called natural parameter.

- A systematic component η_t specifying the expectation of y_t accounted for by known covariates. The linear predictor η_t combines the covariate effects

$$\eta_t = z_t' \beta = \beta_0 + \beta_1 z_{t1} + \cdots + \beta_p z_{tp}. \quad (2.2)$$

- A monotonic differentiable link function which links the mean $E(y_t|z_t) = \mu_t$ and the linear predictor η_t by

$$g(\mu_t) = \eta_t$$

The inverse function $h = g^{-1}$ is referred to as response function such that $\mu_t = h(\eta_t)$.

The function $b(\theta)$ is important because it relates the natural parameter to the mean μ_t as well as the variance of y_t : $\text{Var}(y_t|z_t) = \sigma^2(\mu_t)$. These parameters can be explicitly determined by

$$\begin{aligned} \mu_t &= b'(\theta_t) = \partial b(\theta_t) / \partial \theta, \\ \sigma^2(\mu_t) &= \phi v(\mu_t), \\ v(\mu_t) &= b''(\theta_t) = \partial^2 b(\theta_t) / \partial \theta^2. \end{aligned}$$

Therefore, the natural parameter θ_t can be expressed as a function of the mean, i.e. $\theta_t = \theta(\mu_t)$. If $g(\mu_t) = \theta(\mu_t)$, $g(\cdot)$ is called the natural or canonical link function. This is a special, but important case of a link function which yields $\theta_t = \eta_t$. It is frequently the link function of choice, because it generally has mathematical advantages.

In the following, we consider the example of a Poisson distributed response. Due to the count nature of the data considered in the motivating example, this is the most appropriate distribution from the exponential family. This leads to

$$\begin{aligned} y_t|z_t &\sim \text{Poisson}(\mu_t) \\ f(y_t|z_t) &= \exp\{y_t \log(\mu_t) - \mu_t - \log(y_t!)\}, \end{aligned}$$

where

$$\theta_t = \log(\mu_t), \quad b(\theta_t) = \exp(\theta_t) = \mu_t, \quad \phi = 1.$$

As the mean μ_t has to be strictly positive, the logarithmic function is chosen as link function

$$\log(\mu_t) = \eta_t,$$

which also means that we do not need further restrictions on the parameters β . The present model is also referred to as a log-linear model. Note that the logarithmic function is the natural link function for the Poisson distribution.

To make inference, the vector $\hat{\beta}$ that maximizes the whole likelihood of the model has to be found. In practice, $\hat{\beta}$ is the solution of the estimating equations obtained by differentiating the log-likelihood in terms of β and solving them to zero. Nelder & Wedderburn (1972) proposed a solution using a modified Newton-Raphson technique. This is based on the idea of using a linear approximation to these equations using the second derivatives of the log-likelihood. The modification is that these second derivatives are replaced by their expected values (using the current estimate of the distribution), which results in a simpler form. This general procedure is referred to as Fisher scoring. For a GLM, this procedure can be framed as equivalent to weighted multiple regression with a constructed dependent variable and weights that change at each iteration. Thus, this is often called Iterative (Re-) Weighted Least Squares (IWLS; see Fahrmeir & Tutz (2001), Section 2.2).

2.2 Generalized additive models

Identifying appropriate polynomial terms or transformations of the predictor variables to improve the fit of a (generalized) linear model can be difficult and imprecise. Therefore, the introduction of models that automatically identify appropriate transformations was a further important step forward in regression analyses. It led to an extension of GLM

known as generalized additive models (GAM, Hastie & Tibshirani (1990)). The only underlying assumptions made are the additivity of the functions and that the components are smooth. Like a GLM, a GAM uses a link function to establish the relationship between the mean of a response variable and a 'smoothed' function of the explanatory variable(s). The strength of these models is their ability to deal with highly non-linear and non-monotonic relationships between the response and the explanatory variables. The predictor is then specified as

$$\eta_t = \beta_0 + f_1(z_{t1}) + \cdots + f_p(z_{tp}) = \beta_0 + \sum_j f_j(z_{tj}) \quad (2.3)$$

where each f_j , $j = 1, \dots, p$ is an arbitrary smooth function that is to be estimated.

Estimation of the additive terms for generalized additive models is accomplished by replacing the weighted linear regression for the adjusted dependent variable by the weighted backfitting algorithm, essentially fitting a weighted additive model. The algorithm used in this case is called the *local scoring algorithm*. It is also an iterative algorithm and starts with initial estimates of the f_j . During each iteration, an adjusted dependent variable and a set of weights are computed, and the smoothing components are estimated using a weighted backfitting algorithm. The scoring algorithm stops when the deviance of the estimates ceases to decrease. Overall, the estimating procedure for generalized additive models consists of two separate iterative operations which are usually labeled the outer and inner loop. Inside each step of the local scoring algorithm (outer loop), a weighted backfitting algorithm (inner loop) is used until convergence. Then, based on the estimates from this weighted backfitting algorithm, a new set of weights is calculated and the next iteration of the scoring algorithm starts (see Hastie & Tibshirani (1990) and Fahrmeir & Tutz (2001) for details).

Suppose now that we have data (y_t, z'_t, x'_t) , $t = 1, \dots, T$, and vectors $x_j = (x_{1j}, \dots, x_{Tj})'$ and $z_j = (z_{1j}, \dots, z_{Tj})'$, which are column vectors of continuous and categorical covariates, respectively. The predictor is then a sum of linear combinations of the observed categorical covariates z_j and some unknown parameters, denoted by β_j , and some non-linear smooth functions, denoted by f_j of x_j . One obtains a semiparametric model with the following predictor

$$\eta = \beta_0 + Z\beta + \sum_j f_j(x_j). \quad (2.4)$$

2.3 Varying coefficient models

In contrast to GLMs or GAMs, where the regression coefficients are assumed to be constant, varying coefficient models (VCM), systematically introduced by Hastie & Tibshirani (1993), accommodate situations in which the effects of one or more of the covariates are allowed to vary (interact) more or less smoothly over the values of other variables. The additive-linear structure enables simple interpretation and avoids the so-called 'curse of dimensionality' problem in high dimensional cases.

Varying coefficient models (VCM) have, in general, a predictor of the form

$$\log(\mu) = f_1(r_1) x_1 + \dots + f_p(r_p) x_p, \quad (2.5)$$

where r_1, \dots, r_p are additional covariates and f_1, \dots, f_p are unspecified functions to be estimated. Model (2.5) says that the so-called effect modifiers r_1, \dots, r_p change the coefficients of the x_1, \dots, x_p through the (unspecified) functions f_1, \dots, f_p . The dependence of $f_j(\cdot)$ on r_j implies a special kind of interaction between each r_j and x_j . Typically, the effect modifiers r_1, \dots, r_p are continuous covariates, whereas the interacting variables x_j can be either continuous or categorical.

Varying coefficient models arise from various statistical contexts in slightly different forms. The vast amount of literature includes, among many others, approaches based on local regression (Kauermann & Tutz (1999), Fan & Zhang (1999), Fan & Zhang (2000)), smoothing spline regression (Eubank et al. (2004)), functional data analysis (Ramsay & Silverman (1997), Ramsay & Silverman (2002)) and generalized linear models with varying coefficients (Cai, Fan & Li (2000), Galindo et al. (2001))

2.4 Time-varying coefficient models

As an important special case of the general varying coefficient formulation, time-varying coefficient models (TVCM) are considered in the following, with the effect modifier being time t .

Most of the literature about time-varying coefficient models has been on models for longitudinal data. For example, Hoover et al. (1998) consider a time-varying coefficient model for continuous longitudinal data and use smoothing splines and local polynomial estimators. Huang, Wu & Zhou (2002) propose a basis function approximation method to estimate the time-varying coefficients, whereas Tutz & Kauermann (2003) use generalized local likelihood estimation. Another type of models that has been suggested is

the dynamic generalized linear model (DGLM). For example, West, Harrison & Migon (1985) consider this model and use a linear Bayes estimation and the discount method. More recently, fully Bayesian analysis of DGLMs based on Markov Chain Monte Carlo approaches is proposed, among others, by Gamerman (1998) and Lee & Shaddick (2005).

The basic model to be considered here assumes that we have

$$\log(\mu) = f_0(t) + f_1(t) x_1 + \dots + f_p(t) x_p. \quad (2.6)$$

Further, we assume that some of the functions $f_j(\cdot)$ are constant, that means $f_j(\cdot) = \beta_j$, thus, resulting in terms with simple linear effects. Another specification is given by the fact that the confounders are typically modeled as nonlinear time-invariant functions. In this case the j -th term is simply $f_j(x_j)$, an unspecified function in x_j . Altogether, this leads to a semiparametric time-varying coefficient model with a predictor of the following form:

$$\log(\mu) = \eta = \beta_0 + \sum_{j=1}^l \beta x_j + \sum_{j=l+1}^k f_j(x_j) + \sum_{j=k+1}^p f_j(t) x_j. \quad (2.7)$$

Chapter 3

Modeling approaches for time-varying coefficients

In this chapter, five different approaches for fitting time-varying coefficient models are introduced and discussed. The five approaches are:

- a) Time-varying coefficient models using regression splines,
- b) Bayesian time-varying coefficient models with P-splines,
- c) Time-varying coefficient models with penalized linear splines based on a generalized linear mixed model framework,
- d) Time-varying coefficient models with P-splines based on empirical Bayes inference, and
- e) Adaptive generalized TVC models.

3.1 TVCM using regression splines

One possibility, mentioned in Hastie & Tibshirani (1993), is estimating an interaction term, $f_j(t) x_j$, where the unknown function $f_j(t)$ is approximated by a polynomial spline in time.

Polynomial (regression) splines are piecewise polynomials with the polynomial pieces joining together smoothly at a set of interior knot points. More precisely, suppose that the domain of a variable x is divided by a set of (equally spaced) knots

$$\xi_0 < \xi_1 < \dots < \xi_{M-1} < \xi_M,$$

where ξ_0 and ξ_M are the two end points of x , that means $\xi_0 = x_{(min)}$ and $\xi_M = x_{(max)}$. Then, a polynomial spline of degree l , $l \in \mathbb{N}$ is a function that

- consists of a polynomial of degree l within each of the intervals $[\xi_m, \xi_{m+1})$, $0 \leq m \leq M - 1$, and
- has $(l - 1)$ -times continuous derivatives.

The collection of spline functions of a particular degree and knot sequence forms a linear function space and it can be shown that a spline with the above properties can be represented in terms of a linear combination of $D = M + l$ basis functions B_d in the following way

$$f(x) = \sum_{d=1}^S \gamma_d B_d(x), \quad (3.1)$$

where $\gamma = (\gamma_1, \dots, \gamma_S)'$ corresponds to the vector of unknown regression coefficients. There are several popular basis functions, for example *truncated power series* bases (Ruppert, Wand & Carroll 2003) or *B-spline* bases functions (Eilers & Marx 1996). Here, the focus is on truncated power series bases. A polynomial spline for $f_j(t)$ can be constructed using the truncated polynomials

$$\begin{aligned} B_1(t) &= 1, B_2(t) = t, B_3(t) = t^2, \dots, B_{l+1}(t) = t^l, \\ B_{l+2}(t) &= (t - \xi_1)_+^l, \dots, B_{M+l}(t) = (t - \xi_{M-1})_+^l, \end{aligned}$$

where l is the degree of the spline, $(t - \xi_m)_+^l = \max(0, (t - \xi_m))^l$, and ξ_1, \dots, ξ_{M-1} are fixed knots. Thus, a time-varying coefficient term can be modeled by

$$f_j(t) x_j = \left(\gamma_{j,1} + \gamma_{j,2} t + \dots + \gamma_{j,l+1} t^l + \sum_{m=1}^{M-1} \gamma_{j,l+1+m} (t - \xi_m)_+^l \right) x_j, \quad \text{for } l = 0, 1, 2, \dots \quad (3.2)$$

One special case, which will be considered in this work, is the cubic regression spline. The functions $f_j(t)$ can be parameterized in terms of a linear combination of $M + 3$ basis functions, which can be denoted as

$$\Phi(t) = [1, t, t^2, t^3, (t - \xi_1)_+^3, \dots, (t - \xi_{M-1})_+^3].$$

Likewise, the corresponding coefficients can be denoted by $\gamma_j = (\gamma_{j,1}, \dots, \gamma_{j,M+3})'$. Then the time-varying coefficient term can be written in matrix formulation as

$$f_j(t) x_j = \text{diag}(x_j) \Phi(t)' \gamma_j.$$

If polynomial splines are also used for all of the other nonlinear functions f_j , then the analysis is completely parametric, so that estimation can be done by using the methods considered for generalized linear models (see Chapter 2 or Section 2.2 of Fahrmeir & Tutz (2001)).

3.2 Bayesian TVCM with P-splines

A second approach is the use of Bayesian varying-coefficient models with penalized splines (Lang & Brezger (2004)).

In the frequentist paradigm, parameters in a statistical model (for example the regression coefficients) are treated as unknown, but fixed quantities. In Bayesian statistics, however, all these unknown parameters are considered as random variables which follow a certain probability distribution. Another important difference is the use of the posterior distribution, that means the conditional distribution of the parameters given the observed data y .

Let, in general, ϕ denote the vector of all unknown parameters. Then, the posterior distribution $p(\phi|y)$ is, by Bayes theorem, proportional to the likelihood $L(\phi)$ times the prior distribution $p(\phi)$

$$p(\phi|y) \propto L(\phi|y) p(\phi).$$

All statistical inference can be deduced from the posterior distribution by reporting appropriate summaries. For example, point estimates for unknown parameters are given by the posterior means.

3.2.1 Observation model

Bayesian-type models are defined hierarchically: in the first step, an observation model for responses given covariates is specified. Here, we consider the predictor (2.7) as observation model, that is

$$\log(\mu) = \eta = \beta_0 + \sum_{j=1}^l \beta x_j + \sum_{j=l+1}^k f_j(x_j) + \sum_{j=k+1}^p f_j(t) x_j.$$

3.2.2 Prior assumptions

The Bayesian model formulation is completed by assumptions about priors for parameters and functions. Prior distributions should account for available information and

reflect our prior knowledge about the parameters. Often, these priors will depend on further parameters, called hyperparameters, for which additional hyperpriors have to be defined.

3.2.2.1 Continuous covariates

Priors for the unknown functions f_{l+1}, \dots, f_p depend on the type of the covariates as well as on prior beliefs about the smoothness of f_j . Several alternatives have been proposed for specifying smoothness priors for continuous covariates or time scales. These are random walk priors or, more generally, autoregressive priors (see Fahrmeir & Lang (2001)), Bayesian smoothing spline (Hastie & Tibshirani (2000)) and Bayesian P-spline priors (Lang & Brezger (2004)), which are assumed here. For notational simplicity we will drop the index j in the following discussion.

Analogous to Section 3.1, we assume here that an unknown function f of a covariate x can be approximated by a polynomial spline of degree l and with knots

$$\xi_0 = x_{(min)} < \xi_1 < \dots < \xi_{M-1} < \xi_M = x_{(max)}$$

within the domain of x . However, the basic idea of P-splines (Eilers & Marx 1996) is that the polynomial spline can be constructed in terms of a linear combination of $D = M + l$ B(sic)-spline basis functions (de Boor (1978), Dierckx (1993)). Denoting the d -th B-spline basis function by B_d^l , one then obtains

$$f(x) = \sum_{d=1}^S \gamma_d B_d^l(x). \quad (3.3)$$

The vector of function parameters now contains the regression coefficients or weights of the individual B-spline basis functions: $\gamma = (\gamma_1, \dots, \gamma_S)'$.

The B-spline basis functions of degree l can be defined using the Cox-de Boor recursion formula:

$$B_d^0(x) := \left\{ \begin{array}{ll} 1, & \text{if } \xi_d \leq x < \xi_{d+1} \\ 0 & \text{otherwise} \end{array} \right\}$$

$$B_d^l(x) := \frac{x - \xi_d}{\xi_{d+l} - \xi_d} B_d^{l-1}(x) + \frac{\xi_{d+l+1} - x}{\xi_{d+l+1} - \xi_{d+1}} B_{d+1}^{l-1}(x), \quad l \geq 1. \quad (3.4)$$

For equidistant knots, equation (3.4) simplifies to

$$B_d^l(x) = \frac{1}{l \cdot \Delta\xi} [(x - \xi_d) B_d^{l-1}(x) + (\xi_{d+l+1} - x) B_{d+1}^{l-1}(x)],$$

since $\xi_{d+l} - \xi_d = \xi_{d+l+1} - \xi_{d+1} = l \cdot \Delta\xi$, where $\Delta\xi$ is the distance between two adjacent knots. Note that one needs an extension beyond the domain of x for l knots in each direction to generate a complete B-spline basis of degree l . More information about B-splines can be found in the mentioned literature. Figure 3.1 shows a small set of B-spline basis functions for the degrees $l = 1, 2$, and 3 . B-spline basis functions of degree one, for example, are non-zero within an area spanned by $1 + 2 = 3$ knots. Further, they are generated by linear functions within each interval $[\xi_d, \xi_{d+1})$, as shown in the top row of Figure 3.1. Correspondingly, B-splines of degree two are only positive over the range of four subsequent knots and are polynomials of degree two within each interval (Figure 3.1, middle row), whereas cubic B-spline functions (Figure 3.1, bottom row) are generated by four polynomials of degree three and are non-zero within the range of five knots.

Knot specification and penalization

A crucial problem in spline theory is the appropriate choice of the number of knots. For a small number of knots, the resulting spline may not be flexible enough to capture the variability of the data, whereas for a large number of knots, the estimated curve tends to overfit the data; as a result, a wigglier function is obtained. To meet the opposite requirements, that is enough flexibility without large overfitting, Eilers & Marx (1996) suggest the use of a moderately large number of equally spaced knots (between 20 and 40). Additionally, they introduce a roughness penalty on adjacent B-spline coefficients in their frequentist approach to reduce high variation of the curves leading to a penalization of the likelihood with difference penalty terms

$$\lambda \sum_{d=u+1}^S (\Delta^u \gamma_d)^2, \quad (3.5)$$

where λ is a smoothing parameter and Δ^u is the difference operator of order u .

Usually, first or second order differences are enough. In general, first order differences penalize too big jumps between successive parameters $\gamma_d - \gamma_{d-1}$, whereas second order differences penalize deviations from the linear trend $2\gamma_{d-1} - \gamma_{d-2}$. Therefore, a second order difference imposes a smoother function f than a first order difference does.

The smoothness of the function is now regulated by the smoothing parameter λ . The method recommended by Eilers & Marx (1996) is to minimize the Akaike information criterion (AIC). Details about this criterion can be found in Hastie & Tibshirani (1990). However, the computation of AIC for many values of λ is very time-consuming and becomes intractable in higher dimensions. Moreover, this procedure often fails in practice, since no optimal solutions for the λ can be found (Lang & Brezger 2004).

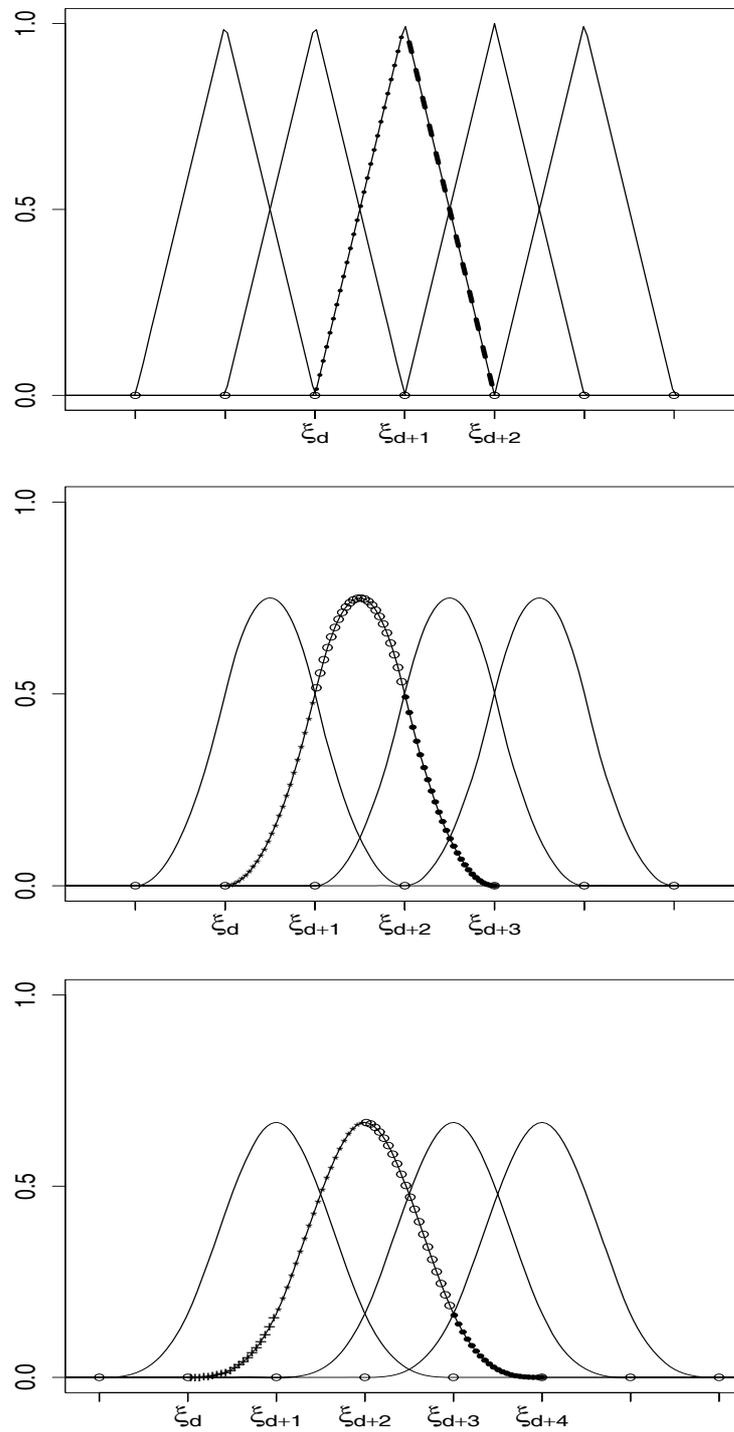


Figure 3.1: Illustration of B-spline basis functions of degree 1 (top), degree 2 (middle) and degree 3 (bottom). Each Figure exemplarily shows the different polynomials of one B-spline.

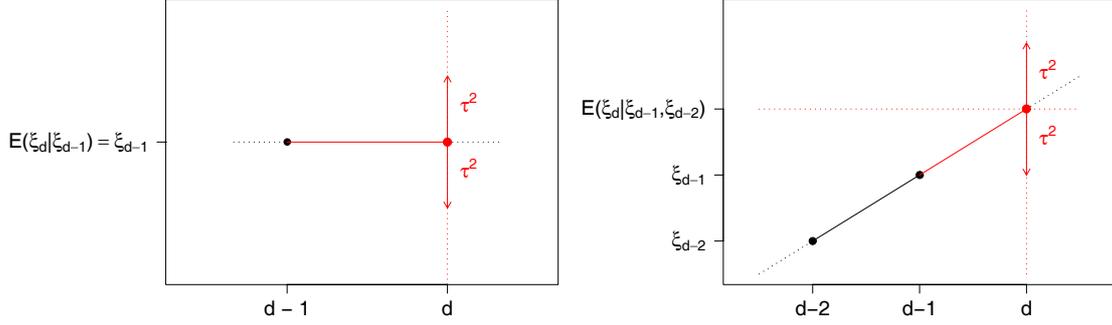


Figure 3.2: Trends induced by first (top) and second order (bottom) random walk priors.

To solve the problem of the choice of an optimal value for the smoothing parameter, a Bayesian approach to P-splines is suggested. Following Lang & Brezger (2004), the Bayesian version of P-splines allows for considering the smoothing parameters as random and, hence, for simultaneous estimation together with the other model parameters. Starting from the general formulation in (3.3), the difference penalty terms on adjacent B-spline coefficients given by (3.5) are replaced by their stochastic analogues: first order random walks (RW1) and second order random walks (RW2). Thus, we obtain

$$\gamma_d = \gamma_{d-1} + u_d \quad (3.6)$$

and

$$\gamma_d = 2\gamma_{d-1} - \gamma_{d-2} + u_d \quad (3.7)$$

respectively, with Gaussian errors $u_d \sim N(0, \tau^2)$. Diffuse priors $p(\gamma_1) \propto \text{const}$ and $p(\gamma_2) \propto \text{const}$ are chosen for initial values.

Alternatively, the above assumptions can be replaced by using the conditional distributions of γ_d given the preceding neighboring parameters

$$\gamma_d | \gamma_{d-1} \sim N(\gamma_{d-1}, \tau^2)$$

and

$$\gamma_d | \gamma_{d-1}, \gamma_{d-2} \sim N(2\gamma_{d-1} - \gamma_{d-2}, \tau^2).$$

The amount of smoothness or penalization is controlled by the additional variance parameters $\tau^2 = \text{Var}(u_d)$. These parameters correspond to the inverse smoothing parameters in the classical approach.

An illustration of this concept is given in Figure 3.2. In case of a first order random walk,

coefficient γ_d is restricted to deviate by τ^2 from the preceding coefficient γ_{d-1} (Figure 3.2 left panel), or alternatively from the interpolating line between γ_{d-2} and γ_{d-1} , in the case of a second order random walk (Figure 3.2 right panel).

Matrix formulation

By defining the $T \times S$ design matrices X_j , where the element in row t and column d is given by $X_j(t, d) = B_{jd}(x_{tj})$, that means

$$X_j = \begin{pmatrix} X_{11} & \dots & X_{1S} \\ \vdots & X_{td} & \vdots \\ X_{T1} & \dots & X_{TS} \end{pmatrix},$$

we can rewrite the P-spline formulation in matrix notation

$$f_j(x_j) = X_j \gamma_j. \quad (3.8)$$

As the splines are non-zero functions in a defined interval and zero elsewhere, the matrix X_j has some sort of band structure that can be used to improve the computational implementation.

3.2.2.2 Time-varying coefficients

The models discussed so far are not suitable for modeling interactions between covariates. As introduced above, in a TVCM, the interaction terms are of the form

$$f_j(t, x_j) = f_j(t)x_j,$$

where the interacting variables x_j are continuous or categorical. Regarding the varying coefficient terms, we can use the priors already defined above for the nonlinear functions f_j , adapting them to the variable *time* t :

$$f_j(t) = X_j^* \gamma_j, \quad j = k + 1, \dots, p. \quad (3.9)$$

Only the design matrices X_j^* in (3.9) have to be redefined by multiplying each element in row t of X_j^* with x_{tj} . Hence, the overall matrix formulation for the varying terms is given by

$$f_j(t, x_j) = f_j(t)x_j = \text{diag}(x_{1j}, \dots, x_{Tj}) X_j^* \gamma_j = X_j \gamma_j,$$

where X_j^* is the usual design matrix for splines composed of the basis functions evaluated at the observations t as given above.

3.2.2.4 Hyperparameters

Additionally, for a fully Bayesian inference, the variance parameters τ_j^2 are also considered as random. Therefore, priors for these variances have to be adopted in a further stage of the hierarchy by highly dispersed but proper inverse gamma priors $p(\tau_j^2) \sim IG(a_j, b_j)$. This allows for simultaneous estimation of the unknown functions and the amount of smoothness. The probability density of a $IG(a_j, b_j)$ distribution is given by

$$\tau_j^2 \propto (\tau_j^2)^{-a_j-1} \exp\left(-\frac{b_j}{\tau_j^2}\right).$$

The parameters a_j and b_j have to be chosen appropriately. A common choice is $a_j = 1$ and b_j equaling a small value, e. g. $b_j = 0.005$, as proposed by Besag et al. (1995). The choice of such highly vague but proper priors prevents problems associated with noninformative priors. Alternatively, we may set $a_j = b_j$, e.g. $a_j = b_j = 0.001$, leading to an almost noninformative prior for τ_j^2 . For a detailed discussion about the propriety of the posteriors see Fahrmeir & Kneib (2006). Based on experience from extensive simulation studies, we use $a_j = b_j = 0.001$ as our standard choice.

3.2.3 MCMC inference

Let $\gamma = (\gamma'_{l+1}, \dots, \gamma'_p)'$ denote the vector of all regression coefficients of the B-spline bases, β the vector of all fixed effects, and $\tau^2 = (\tau_{l+1}^2, \dots, \tau_p^2)$ the vector of all variance parameters. As mentioned above, full Bayesian inference is based on the posterior distribution which is given by

$$p(\gamma, \tau^2, \beta | \text{data}) \propto L(\gamma, \tau^2, \beta) p(\gamma, \tau^2, \beta),$$

where $L()$ denotes the likelihood which is the product of individual likelihood contributions. Under usual conditional independence assumptions for the parameters, the joint prior $p(\gamma, \tau^2, \beta)$ can be factorized, yielding

$$p(\gamma, \tau^2, \beta | \text{data}) \propto L(\gamma, \tau^2, \beta) \prod_{j=l+1}^p (p(\gamma_j | \tau_j^2) p(\tau_j^2)) p(\beta), \quad (3.11)$$

where the last factor can be omitted, as we have chosen diffuse priors for the fixed effects.

In many practical applications, this high dimensional posterior will not have a known closed form, but rather a complicated high dimensional density only known up to the proportionality constant. Estimation of the unknown parameters is therefore done by simulation of the posterior (3.11) with Markov Chain Monte Carlo (MCMC) techniques.

The main idea of MCMC is to establish a Markov chain whose stationary distribution is the posterior distribution of interest, and then to collect samples from that chain. In this way, a sample of dependent random numbers from the posterior is obtained.

Full Bayesian inference via MCMC simulation is based on updating full conditionals of single parameters or blocks of parameters, given the rest of the data. For updating the parameter vectors γ_j and the fixed effects β , a Metropolis Hastings (MH)-algorithm based on IWLS proposals is used. The concept of IWLS updates has been introduced by Gamerman (1997) in the context of the estimation of generalized linear mixed models and extended to a more general setting by Brezger & Lang (2006). A more detailed description of this algorithm will be given in the context of measurement error correction methods (Section 7.1).

3.3 TVCM with penalized linear splines based on a generalized linear mixed model framework

Recent applications in spline smoothing make use of a convenient connection between penalized splines and mixed models (Brumback, Ruppert & Wand 1999).

In contrast to the fully Bayesian approach, where the unknown functions were approximated by penalized B-spline basis functions, we consider here the truncated power series basis, similar to Section 3.1.

Mixed model approaches to penalized splines based on truncated power series representation have been described in Coull, Schwartz & Wand (2001), Wand (2003), Ruppert et al. (2003) and Ngo & Wand (2004). The latter also describes the estimation of varying coefficient models, but only for the case of Gaussian distributed responses and two covariates.

In this section, we extend this estimation approach to a generalized setting; however, we follow the suggestions of Ngo & Wand (2004) and use penalized linear splines.

Before addressing the estimation of regression coefficients as well as variance parameters in a (generalized) mixed model in detail, we will first show how to reformulate model (2.7) as generalized linear mixed model.

3.3.1 Mixed model representation

The linear spline estimator for a function f_j is of the form

$$f_j(x_j) = \alpha_{j0} + \alpha_{j1}x_j + \sum_{m=1}^{M_j-1} u_{j,m}(x_j - \xi_m^j)_+,$$

where

$$(x_j - \xi_m^j)_+ = \begin{cases} 0, & x_j \leq \xi_m^j \\ x_j - \xi_m^j, & x_j > \xi_m^j \end{cases},$$

and $\xi_1, \dots, \xi_{M_j-1}$ are fixed knots. Equivalently, a linear spline version of a (time-) varying coefficient term is given by

$$f_j(t) x_j = \left(\alpha_{j0} + \alpha_{j1} t + \sum_{m=1}^{M_t-1} u_{j,m}(t - \xi_m)_+ \right) x_j,$$

where $\xi_1, \dots, \xi_{M_t-1}$ are fixed knots over the range of t .

Using these two formulae, model (2.7) can be written as

$$\begin{aligned} \eta_t = \beta_0 &+ \sum_{j=1}^l \beta x_{tj} + \sum_{j=l+1}^k \left(\alpha_j x_{tj} + \sum_{m=1}^{M_j-1} u_{j,m}(x_{tj} - \xi_m^j)_+ \right) \\ &+ \sum_{j=k+1}^p \left(\alpha_{j0} + \alpha_{j1} t + \sum_{m=1}^{M_t-1} u_{j,m}(t - \xi_m)_+ \right) x_{tj}, \quad t = 1, \dots, T. \end{aligned} \quad (3.12)$$

To implement spline smoothing in practice, the coefficient vectors $\alpha = (\alpha_{l+1}, \dots, \alpha_p)'$ and $u = (u_{l+1,1}, \dots, u_{p,M_t-1})$ have to be estimated. In principle, we could apply the usual methods as described in Section 3.1, but we expect that this tends to result in a rather rough estimate of the function. A greater degree of smoothing can be achieved by shrinking the estimated coefficients towards zero. This is now achieved by regarding the $u_{j,m}$ as random coefficients, distributed independently as

$$u_{l+1,1}, \dots, u_{l+1,M_{l+1}-1} \text{ iid } N(0, \sigma_{l+1}^2), \quad \dots \quad u_{p,1}, \dots, u_{p,M_t-1} \text{ iid } N(0, \sigma_p^2),$$

as suggested by Brumback et al. (1999), with σ_j^2 controlling the amount of smoothing. Constraining the coefficients $u_{j,m}$ to come from a common distribution has the effect of damping changes in the gradient of fitted line segments from one knot point to the next. Thus, the resulting estimator is a type of penalized spline smoother.

Now, we set the design matrix X to take the following form

$$X = \begin{bmatrix} 1 & x_{11} & \dots & x_{1,k-1} & 1 & x_{1,k+1} & 1 & x_{1,k+1} & \dots & x_{1p} & 1 & x_{1p} \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{T1} & \dots & x_{T,k-1} & T & x_{T,k+1} & T & x_{T,k+1} & \dots & x_{Tp} & T & x_{Tp} \end{bmatrix}.$$

The random effects matrix Z is constructed as

$$Z = [(x_{t,l+1} - \xi_m^{l+1})_+ \cdots (t - \xi_m^k)_+ \quad x_{t,k+1} (t - \xi_m^{k+1})_+ \cdots x_{tp} (t - \xi_m^p)_+]_{1 \leq t \leq T}.$$

By defining

$$b = [u_{l+1,1}, \dots, u_{l+1, M_{l+1}-1}, \dots, u_{p,1}, \dots, u_{p, M_t-1}]' \quad \text{and}$$

$$D = \text{Cov}(b) = \begin{pmatrix} \sigma_{l+1}^2 I & 0 & \dots & 0 \\ 0 & \sigma_{l+2}^2 I & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_p^2 I \end{pmatrix},$$

we can then write model (3.12) in matrix notation as

$$\eta = X\beta + Zb.$$

Having divided the functions $f(x_j)$ and the varying coefficient terms $f(t)x_j$ into a fixed effects and into a random effects part, we get the following generalized linear mixed model:

$$\begin{aligned} y|b &\sim \text{Poisson}(\mu) \\ \log(\mu) &= X\beta + Zb + \epsilon \\ b &\sim N(0, D). \end{aligned}$$

Fitting the linear splines therefore amounts to minimization of the penalized spline criterion.

Knot specification

The location of the knots must be pre-specified before fitting the penalized spline model. However, following Wand (2003) knot specification is "*very much a minor detail for penalized splines*". We therefore follow Wand's suggestion that a reasonable default rule for the location of the knots is given by

$$u_s = \frac{(s+1)}{S+2} \text{th sample quantile of the unique } x_t, \quad 1 \leq s \leq S,$$

where $S = \min \{ \frac{1}{4} \times \text{number of unique } x_t; 35 \}$. Some additional algorithms, empirical results and further comments on the topic of knot selection are supplied in Ruppert (2002).

3.3.2 Estimation

The marginal likelihood is given by

$$\begin{aligned} L(\beta, \theta, y) &= \int f(y|b)f(b)db \\ &= \int \left[\prod_{t=1}^T -\mu_t + y_t \log(\mu_t) - \log(y_t!) \right] \left[\prod_{j=l+1}^p (2\pi\sigma_j)^{-\frac{s_j}{2}} \exp \left\{ -\frac{b'_j b_j}{2\sigma_j^2} \right\} \right] db, \end{aligned} \quad (3.13)$$

where $f(y|b)$ is the conditional distribution of y and $f(b)$ is the multivariate normal density of b . In contrast to linear mixed models, the likelihood above involves intractable integrals whose dimension depends on the specific problem at hand.

Therefore, solutions are proposed where the integrand in equation (3.13) is approximated by a normal density such that the integral can still be calculated analytically (see Wolfinger & O'Connell (1993) and Breslow & Clayton (1993), who arrived independently at an IWLS procedure for fitting generalized linear mixed models). The process of computing the linear approximation must be repeated several times until some criterion stabilizes. Fitting methods based on linearizations are usually doubly iterative. The generalized linear mixed model is approximated by a linear mixed model based on current values of the covariance parameter estimates. The resulting linear mixed model is then fit, which is itself an iterative process. On convergence, the new parameter estimates are used to update the linearization, which results in a new linear mixed model.

3.3.2.1 Linearization

The linearization method first constructs a pseudo-model and pseudo-data. Following Wolfinger & O'Connell (1993) and Breslow & Clayton (1993), we suppose that the conditional mean satisfies

$$E(y|b) = \mu = h(\eta) = h(X\beta + Zb)$$

with $b \sim N(0, D)$ and

$$\text{Var}(y|b) = V_\mu^{1/2} R V_\mu^{1/2},$$

where $V_\mu^{1/2}$ is a diagonal matrix with elements equal to the variance functions. R contains the dispersion parameter, if one exists.

For the first analytic approximation, let $\hat{\beta}$ and \hat{b} be known estimates of β and b . Then, we define $\hat{\mu} = h(X\hat{\beta} + Z\hat{b})$, which is a vector with elements consisting of evaluations of

the conditional mean μ at each component of $X\hat{\beta} + Z\hat{b}$. Further, we have a diagonal matrix consisting of evaluations of the first derivatives of the conditional mean

$$\tilde{\Delta} = \left(\frac{\partial h(\eta)}{\partial \eta} \right)_{\hat{\beta}, \hat{b}}.$$

Now, let

$$\tilde{\epsilon} = y - \hat{\mu} - \tilde{\Delta}(X\beta - X\hat{\beta} + Zb - Z\hat{b}),$$

which implies that $\tilde{\epsilon}$ is a Taylor series approximation to $\epsilon = y - \mu$ about $\hat{\beta}$ and \hat{b} .

Following Lindstrom & Bates (1990), the conditional distribution of $\tilde{\epsilon}$ given β and b is approximated with a Gaussian distribution having the same first two moments as $\epsilon|\beta, b$; in particular, we assume that $\epsilon|\beta, b \sim N(0, V)$.

A further approximation is the substitution of $\hat{\mu}$ for μ in the variance matrix. Since

$$\tilde{\Delta}_t = \frac{1}{g'(\hat{\mu}_t)}$$

for each component t , where $g'(\hat{\mu}_t) = \frac{\partial g(\hat{\mu}_t)}{\partial \mu}$, one can then write

$$\tilde{\Delta}^{-1}(y - \hat{\mu})|\beta, b \sim N(X\beta - X\hat{\beta} + Zb - Z\hat{b}, \tilde{\Delta}^{-1}V\tilde{\Delta}^{-1}).$$

Using the definition $\tilde{y} = X\hat{\beta} + Z\hat{b} + \tilde{\Delta}^{-1}(y - \hat{\mu})$ results in the specification

$$\tilde{y}|\beta, b \sim N(X\beta + Zb, \tilde{\Delta}^{-1}V\tilde{\Delta}^{-1}),$$

which is a weighted linear mixed model with pseudo-response \tilde{y} , (unknown) fixed effects β , random effects $b \sim N(0, D)$, and pseudo-error $\tilde{\epsilon} \sim N(0, \tilde{\Delta}^{-1}V\tilde{\Delta}^{-1})$. The weight matrix in this linear mixed model takes the form $\tilde{W} = \tilde{\Delta}V^{-1}\tilde{\Delta}$, where V is defined as above. For canonical link functions, as is the case in our Poisson models, $\tilde{W} = V_\mu$.

In this linearized model, we observe \tilde{y} , X , and Z , while β and b as well as D and W are generally unknown. Thus, mixed-model analysis involves two complementary estimation issues: (1) estimation of the vectors of fixed and random effects, β and b , and (2) estimation of the covariance matrices W and D .

Now, we define $\tilde{\Lambda}(\theta^*) = \tilde{\Lambda} = W^{-1} + ZD^*Z'$, as the marginal variance in the linearized mixed model for y , where θ^* is a vector of variance-component parameters consisting of all unknown parameters in D^* and V^* . Here, θ^* , D^* and V^* are re-parameterized versions of θ , D and V in terms of ratios with the scale or dispersion parameter ϕ . W is the diagonal matrix of weights. Then, the following Gaussian log-likelihood function is corresponding to the linearized mixed model for \tilde{y} :

$$l(\beta, \phi, \theta^*) = -\frac{T}{2} \log(2\pi) - \frac{1}{2} \log |\phi \tilde{\Lambda}| - \frac{1}{2} \phi^{-1} (\tilde{y} - X\beta)' \tilde{\Lambda}^{-1} (\tilde{y} - X\beta). \quad (3.14)$$

By including ϕ , a quasi-likelihood-style extension of the generalized linear model is provided, which can be omitted if desired.

3.3.2.2 Estimation of regression coefficients

As shown above, the new pseudo-response \tilde{y} is a Taylor series approximation to the linked response $g(y)$. It is analogous to the modified dependent variable used in the IWLS algorithm of Nelder & Wedderburn (1972).

To solve for $(\hat{\beta}, \hat{b})$, the Fisher scoring algorithm is employed which leads to iteratively solving the system of equations

$$\begin{pmatrix} X'W^{-1}X & X'W^{-1}Z \\ Z'W^{-1}X & Z'W^{-1}Z + D^{-1} \end{pmatrix} \begin{pmatrix} \hat{\beta} \\ \hat{b} \end{pmatrix} = \begin{pmatrix} X'W^{-1}\tilde{y} \\ Z'W^{-1}\tilde{y} \end{pmatrix}. \quad (3.15)$$

The equation system (3.15) is Henderson's mixed model equation from the normal theory model but with (pseudo-) error $\tilde{\epsilon}$ and (pseudo-) dependent variable \tilde{y} . This is equivalent to the best linear unbiased estimation of β and best linear unbiased prediction of b based on the linearized mixed models

$$\hat{\beta} = (X'\Lambda X)^{-1}X'\Lambda\tilde{y} \quad (3.16)$$

$$\hat{b} = D^{-1}Z'\Lambda(\tilde{y} - X\beta). \quad (3.17)$$

A relatively straightforward extension of Henderson's mixed-model equations provides estimates of the standard errors of the fixed and random effects. Let the inverse of the leftmost matrix in Equation (3.15) be

$$H = \begin{pmatrix} X'W^{-1}X & X'W^{-1}Z \\ Z'W^{-1}X & Z'W^{-1}Z + D^{-1} \end{pmatrix}^{-1}.$$

Using the definition

$$H_0 = \begin{pmatrix} X'W^{-1}X & X'W^{-1}Z \\ Z'W^{-1}X & Z'W^{-1}Z \end{pmatrix},$$

the approximate covariance matrix of $\hat{\beta}$ and \hat{b} is obtained as

$$\text{Cov} \begin{pmatrix} \hat{\beta} \\ \hat{b} \end{pmatrix} = H H_0 H.$$

These expressions allow us to compute approximate covariance matrices for the estimates $\hat{f}_j(x_j)$ and $\hat{f}_j(t)$

$$\text{Cov}(\hat{f}) = (X_j, Z_j) \text{Cov} \begin{pmatrix} \hat{\beta}_j \\ \hat{b}_j \end{pmatrix} (X_j, Z_j)', \quad (3.18)$$

where $\text{Cov}(\hat{\beta}'_j, \hat{b}'_j)$ can be obtained from the corresponding blocks of $\text{Cov}(\hat{\beta}', \hat{b}')$.

3.3.2.3 Estimation of variance parameters

Denoting the estimates β and b by $\hat{\beta}$ and \hat{b} and plugging them back into the log-likelihood (3.14) yields the so-called profile log-likelihood

$$l(\theta^*) = \text{constant} - \frac{1}{2} \log |\phi \tilde{\Lambda}| - \frac{1}{2} \log(r' \tilde{\Lambda}^{-1} r),$$

where $r = \tilde{y} - X(X' \tilde{\Lambda}^{-1} X)^{-1} X' \tilde{\Lambda}^{-1} \tilde{y}$.

However, it is well known that maximization of the log-likelihood with respect to θ^* yields biased estimates, since the estimation does not take into account the loss of degrees of freedom due to the estimation of β . Therefore, for the estimation of variance parameters the usage of *restricted maximum likelihood* (REML) is preferred. The restricted log-likelihood is given by

$$l_r(\theta^*) = \text{constant} - \frac{1}{2} \log |\phi \tilde{\Lambda}| - \frac{T-p}{2} \log(r' \tilde{\Lambda}^{-1} r) - \frac{1}{2} \log |X' \tilde{\Lambda}^{-1} X|, \quad (3.19)$$

where p denotes the rank of X . Hence, the restricted log-likelihood mainly differs from the log-likelihood by an additional component; in particular, one has

$$l_r(\theta^*) \approx l(\theta^*) - \frac{1}{2} \log |X' \tilde{\Lambda}^{-1} X|.$$

To obtain REML-estimates for the variance components, we have to maximize (3.19) over the parameters in θ^* . This requires numerical methods such as the Fisher-Scoring algorithm. A detailed description of REML-estimation by Fisher-Scoring can be found in Kneib (2006).

Upon obtaining θ^* , estimates for ϕ are computed as

$$\hat{\phi} = (r' \tilde{\Lambda}^{-1} r) / (T - p). \quad (3.20)$$

3.3.3 Estimation algorithm

As described in Wolfinger & O'Connell (1993), the estimation algorithm for the restricted pseudo-likelihood (REPL) technique is as follows:

1. Obtain an initial estimate $\hat{\mu}$ of μ .
2. Compute

$$\tilde{y} = X\hat{\beta} + Z\hat{b} + g'(\hat{\mu})(y - \hat{\mu}).$$

3. Fit a weighted linear mixed model using REML with pseudo-response \tilde{y} , fixed effects model matrix X , random effects model matrix Z and weight matrix

$$\tilde{W} = \tilde{\Delta} V^{-1} \tilde{\Delta}.$$

Convert the yielding estimates \hat{D}^* and \hat{V}^* to \hat{D} and \hat{V} by using an estimate $\hat{\phi}$ of ϕ as given in (3.20).

4. The old estimates of \hat{D} and \hat{V} are then compared to the new ones. If there is a sufficiently small difference, stop; otherwise, go to the next step.
5. Obtain updated estimates $\hat{\beta}$ and \hat{b} for the regression coefficients given the current variance parameters by solving the mixed model equation system

$$\begin{pmatrix} X'\hat{W}^{-1}X & X'\hat{W}^{-1}Z \\ Z'\hat{W}^{-1}X & Z'\hat{W}^{-1}Z + \hat{D}^{-1} \end{pmatrix} \begin{pmatrix} \hat{\beta} \\ \hat{b} \end{pmatrix} = \begin{pmatrix} X'\hat{W}^{-1}\tilde{y} \\ Z'\hat{W}^{-1}\tilde{y} \end{pmatrix}.$$

6. Compute a new estimate of $\hat{\mu}$ by substituting $\hat{\beta}$ and \hat{b} in the expression

$$\hat{\mu} = h(X\hat{\beta} + Z\hat{b}).$$

Then go back to step 2.

3.4 TVCM with P-splines based on empirical Bayes inference

An alternative approach is the use of varying coefficient models with P-splines based on empirical Bayes inference (Fahrmeir, Kneib & Lang (2004) and Kneib (2006)). This approach is similar to the one of Section 3.3, the differences being that, firstly, smooth estimation of the unknown functions is achieved here using penalized splines based on B-spline bases and, secondly, this approach is seen from a Bayesian perspective.

3.4.1 Observation model and prior assumptions

Similar to the fully Bayesian approach, we have to specify an observation model, for which we suppose again a model as defined in (2.7), that is

$$\log(\mu) = \eta = \beta_0 + \sum_{j=1}^l \beta x_j + \sum_{j=l+1}^k f_j(x_j) + \sum_{j=k+1}^p f_j(t) x_j.$$

As noted above, in a Bayesian approach, the unknown functions $f_j(x_j)$ and $f_j(t)$ in the observation model as well as the parameter vectors β and γ are considered as random variables and must be supplemented by appropriate prior assumptions. We choose here the same prior distributions as given in Section 3.2.

3.4.2 Mixed model representation

As shown in Section 3.2.2, the functions $f_j(x_j)$, $j = l + 1, \dots, k$ and $f_j(t)x_j$, $j = k + 1, \dots, p$ can be expressed as the matrix product of a design matrix X_j and a vector of unknown parameters γ_j , i.e.

$$f_j(x_j) = X_j \gamma_j \quad \text{and} \quad f_j(t)x_j = \text{diag}(x_{1j}, \dots, x_{Tj}) X_j^* \gamma_j = X_j \gamma_j,$$

respectively.

Similar to Section 3.3, we want to express the functions $f_j(x_j)$ and $f_j(t)$ as the sum of a fixed and a random component. To achieve this, we need to decompose the vectors of regression coefficients γ_j , $j = l + 1, \dots, p$ into an unpenalized and a penalized part. Consider, therefore, the precision matrices K_j corresponding to P-splines (see Section 3.2.2.1). As the matrix K_j is rank deficient with rank r_j , the decomposition of γ_j is of the form

$$\gamma_j = \psi_j^{\text{unpen}} \gamma_j^{\text{unpen}} + \psi_j^{\text{pen}} \gamma_j^{\text{pen}}, \quad (3.21)$$

with the dimensions of ψ_j^{unpen} and ψ_j^{pen} given as $S_j \times (S_j - r_j)$ and $S_j \times r_j$, respectively. Following Fahrmeir et al. (2004) and Kneib (2006), the matrices ψ_j^{unpen} contain a $(S_j - r_j)$ -dimensional basis of the null space of K_j . More specifically, we define new matrices ψ_j^{unpen} , depending on the dimension of K_j , by

$$\begin{pmatrix} 1 & \xi_1 & \dots & \xi_1^{r_j-1} \\ \vdots & \vdots & & \vdots \\ 1 & \xi_{S_j} & \dots & \xi_{S_j}^{r_j-1} \end{pmatrix}.$$

For example, for P-splines with a second order random walk prior, ψ_j^{unpen} is specified as a two-column matrix, where the first column is the identity vector and the second column is composed of the (equidistant) knots of the spline:

$$\begin{pmatrix} 1 & \xi_1 \\ \vdots & \vdots \\ 1 & \xi_{S_j} \end{pmatrix}.$$

Matrices ψ_j^{pen} can be constructed as $\psi_j^{pen} = L'(LL')^{-1}$, where L is a full column rank matrix with $K_j = L_j L_j'$. For P-splines, a suitable choice is given by $L_j = D_j'$, where D_j is the corresponding difference matrix.

Using the decomposition (3.21), we get for the general prior (3.10)

$$p(\gamma_j | \tau_j^2) \propto \exp\left(-\frac{1}{2\tau_j^2} \gamma_j' K_j \gamma_j\right) = \exp\left(-\frac{1}{2\tau_j^2} \gamma_j^{pen'} \gamma_j^{pen}\right), \quad (3.22)$$

which implies that γ_j^{unpen} has a flat prior, $p(\gamma_j^{unpen}) \propto \text{const}$, and the random components $p(\gamma_j^{pen})$ are assumed to be

$$\gamma_j^{pen} | \tau_j^2 \propto N(0, \tau_j^2 I_{r_j}). \quad (3.23)$$

In the next step, we define the matrices $X_j^* = X_j \psi_j^{unpen}$ and $Z_j^* = X_j \psi_j^{pen}$, $j = l+1, \dots, p$. Using these matrices, equation (2.7) can be rewritten as

$$\eta = X \beta + \sum_{j=l+1}^p (X_j^* \gamma_j^{unpen} + Z_j^* \gamma_j^{pen}).$$

This, however, can be further simplified by defining the matrix $X^* = (X_{l+1}^*, \dots, X_p^*, X)$ as well as the vector $\gamma^{unpen} = ((\gamma_{l+1}^{unpen})', \dots, (\gamma_p^{unpen})', \beta)'$. Similarly, we make the specifications $Z^* = (Z_{l+1}^*, \dots, Z_p^*)$ and $\gamma^{pen} = ((\gamma_{l+1}^{pen})', \dots, (\gamma_p^{pen})')'$. Using these specifications, we obtain a generalized linear mixed model

$$\eta = X^* \gamma^{unpen} + Z^* \gamma^{pen}, \quad (3.24)$$

with fixed effects γ^{unpen} and random effects γ^{pen} , where $\gamma^{pen} \sim N(0, \Lambda)$, and $\Lambda = \text{diag}(\tau_{l+1}^2, \dots, \tau_{l+1}^2, \dots, \tau_p^2, \dots, \tau_p^2)$.

3.4.3 Empirical Bayes Inference

Bayesian inference is based on the posterior of the model. The analytic form of the posterior depends on the specific parameterization of the model. Based on the generalized

mixed model representation (3.24), we get the following posterior distribution:

$$p(\gamma^{unpen}, \gamma^{pen} | y) \propto L(y, \gamma^{unpen}, \gamma^{pen}) \prod_{j=l+1}^p (p(\gamma_j^{pen} | \tau_j^2)), \quad (3.25)$$

where $p(\gamma_j^{pen} | \tau_j^2)$ is defined in (3.23) and $L()$ denotes the likelihood which is the product of individual likelihood contributions. Note that for empirical Bayes inference, no priors $p(\tau_j^2)$ for the variances are specified, since the variances τ_j^2 are considered to be fixed; but the τ_j^2 are estimated from the data.

Using the approximation (3.22), the posterior distribution (3.25) can be transformed to

$$p(\gamma^{unpen}, \gamma^{pen} | y) \propto L(y, \gamma^{unpen}, \gamma^{pen}) \exp\left(-\frac{1}{2} \gamma^{pen'} \Lambda^{-1} \gamma^{pen}\right).$$

The corresponding log-posterior is then given by

$$l_p(\gamma^{unpen}, \gamma^{pen} | y) \propto l(y, \gamma^{unpen}, \gamma^{pen}) - \sum_{j=l+1}^p \frac{1}{2\tau_j^2} \gamma^{pen'} \gamma^{pen}.$$

Updated estimates for the unknown functions and covariate effects are obtained using IWLS, resulting in posterior mode estimates. Updated estimates for the variance parameters τ_j^2 given the regression coefficients are obtained by a Fisher-Scoring-type algorithm that maximizes the (approximate) marginal likelihood of the variances. Note that this marginal likelihood is equivalent to the REML in Section 3.3.2.3. Consequently, we have again a data driven smoothing parameter selection.

Estimation begins by assigning arbitrary starting values to the model parameters. These values are iteratively updated by the following steps:

1. Assuming the variance parameters to be known and equal to their current estimates, $\tilde{\tau}_{l+1}^2, \dots, \tilde{\tau}_p^2$, the fixed and random effects γ^{unpen} and γ^{pen} are updated by maximizing Green's (1987) PQL criterion. Applying Fisher scoring, the estimates are computed as the iterative solution to the following system of equations (see also Section 3.3.2.2):

$$\begin{pmatrix} X^{*'} W X^* & X^{*'} W Z^* \\ Z^{*'} W X^* & Z^{*'} W Z^* + \Lambda^{-1} \end{pmatrix} \begin{pmatrix} \hat{\gamma}^{unpen} \\ \hat{\gamma}^{pen} \end{pmatrix} = \begin{pmatrix} X^{*'} W \tilde{y} \\ Z^{*'} W \tilde{y} \end{pmatrix}. \quad (3.26)$$

The symbol $\tilde{y} = (\tilde{y}_1, \dots, \tilde{y}_T)$ in the above equation represents the vector of usual working observations. The $T \times T$ diagonal matrix of working weights is denoted by W , where the working weights are, except for the dispersion parameter, given similar to Section 3.3.2.2.

2. Assuming the fixed and random effects to be equal to their current estimates, $\hat{\gamma}^{unpen}$ and $\hat{\gamma}^{pen}$, estimators of the variance parameters $\tau^2 = (\tau_{l+1}^2, \dots, \tau_p^2)$ are obtained. To accomplish this, the (approximate) restricted log likelihood

$$l_r(\tau^2) = -\frac{1}{2} \log(|\Sigma|) - \frac{1}{2} \log(|X^{*\prime} \Sigma^{-1} X^*|) \\ - \frac{1}{2} (\tilde{y} - X^* \hat{\gamma}^{unp})' \Sigma^{-1} (\tilde{y} - X^* \hat{\gamma}^{unp})$$

is maximized with respect to the variance parameters $\tau_{l+1}^2, \dots, \tau_p^2$. Here, $\Sigma = W^{-1} + Z^* \Lambda Z^{*\prime}$ is an approximation to the marginal covariance matrix of $\tilde{y} | \gamma^{pen}$. Updated estimates $\hat{\tau}^2$ are then obtained by

$$\hat{\tau}^2 = \tilde{\tau}^2 + \tilde{F}(\tilde{\tau}^2)^{-1} \tilde{s}(\tilde{\tau}^2),$$

where \tilde{F} is the expected Fisher information and \tilde{s} is the score vector (see Kneib (2006) for the exact formulas).

The two estimation steps are iterated until convergence. A more detailed description of the empirical Bayes inference can be found in Fahrmeir et al. (2004) or Kneib (2006).

Similar to Section 3.3.2.2, the system of equations (3.26) can be used to derive standard errors for the functions f_j and $f_j(t)$. Again, we denote the inverse of the matrix on the left hand side of (3.26) by H^{-1} . The approximate Bayesian covariance matrix of $\hat{\gamma}^{unpen}$ and $\hat{\gamma}^{pen}$ may be derived as

$$\text{Cov} \begin{pmatrix} \hat{\gamma}^{unpen} \\ \hat{\gamma}^{pen} \end{pmatrix} = H, \quad (3.27)$$

which has a simpler form than its frequentist counterpart in equation (3.18) (see also Lin & Zhang (1999)). The covariance matrices of \hat{f}_j and $\hat{f}_j(t)$ are then given by

$$\text{Cov}(\hat{f}_j, \hat{f}_j(t)) = (X_j^*, Z_j^*) \text{Cov}(\hat{\gamma}_j^{unpen}, \hat{\gamma}_j^{pen}) (X_j^*, Z_j^*), \quad (3.28)$$

where $\text{Cov}(\hat{\gamma}_j^{unpen}, \hat{\gamma}_j^{pen})$ can be obtained from the corresponding blocks of (3.27).

3.5 Adaptive generalized TVC models

In contrast to the approaches described in the last sections, which were based on spline estimation, in this chapter a method based on local likelihood estimation will be considered.

Fan, Yao & Cai (2003) propose a class of adaptive varying-coefficient linear models of the form

$$Y_t = \sum_{j=0}^p g_j(\beta_0^T X_t) x_{t,j} + \epsilon_t, \quad (3.29)$$

where $X_t = (x_0, \dots, x_{tp})$ and $x_0 = 1$. In such models, the parameter vector β_0 is unknown, and the functions g_j are also unknown. Such a model is called "adaptive" by Fan et al. (2003) to indicate "that the coefficients are functions of an unknown index $\beta_0^T X$ ", in contrast to, for example, the generalized varying coefficient models proposed by Cai et al. (2000). Formally, model (3.29) also includes the popular single-index model as well as the partially linear single-index models as special cases; see Chapter 8 of Fan & Yao (2003) and the references therein.

To be applicable in a more general context, this approach is extended to allow, for example, for a Poisson response. Then, the following predictor is obtained:

$$\log(\mu) = \sum_{j=0}^p g_j(\beta_0^T X) x_j = g_0(\beta_0^T X) + X^T g(\beta_0^T X). \quad (3.30)$$

The estimation procedure can be formally split into two parts: Estimation of functions g_j with given β_0 and estimation of the index coefficient β_0 with given functions g_j ; unlike Fan et al. (2003), however, we do not apply backward deletion to choose locally significant variables.

The algorithm for practical implementation will be summarized at the end of this section.

3.5.1 Estimation

Model (3.30) is not identifiable, as we may replace the (g_0, g) by $(g_0 + c\beta_0^T X, g - cg_0)$ for any $c \in \mathbb{R}$. Therefore, the model may be represented in a reduced form

$$\sum_{j=0}^{p-1} g_j(\beta_0^T X) x_j, \quad (3.31)$$

see Theorem 1 of Fan et al. (2003). Note that (3.30) may always be expressed in the form of (3.31), provided the last component of β_0 is non-zero.

More comments about the identifiability of models such as (3.31) are given in Fan et al. (2003) and Lu, Tjøstheim & Yao (2007).

We now turn to the estimation of the functions $f_j(\cdot)$. The method described here uses a given estimator of β_0 , which is called $\hat{\beta}_0$. We therefore write for the index $\hat{Z} = \hat{\beta}_0^T X$.

3.5.1.1 Estimation of the functions $g_j(\cdot)$ with given \mathbf{Z}

Estimation of the varying coefficients $g_j(\cdot)$ is done by local smoothing techniques. Local polynomial smoothers have become one method of choice in nonparametric regression

in recent years. We refer to Fan & Gijbels (1996) and Hastie & Tibshirani (1990) for a detailed review.

The general idea of local regression is the following: Suppose that the $(k+1)$ -th derivative of a function $g_j(\cdot)$ at point z exists. We can approximate the unknown function g_j locally at z by a polynomial of order l . The theoretical justification is that we can approximate $g_j(\cdot)$ for data points Z in a neighborhood of z via a Taylor expansion by a polynomial of degree l :

$$g_j(\cdot) \approx \sum_{k=0}^l a_{jk} (\hat{Z} - z)^k,$$

where

$$a_{jk} = \frac{f^{(k)}(z)}{k!}.$$

For various practical applications, Fan & Gijbels (1996) recommend using the local linear modeling scheme, that means $l = 1$. This specification is also used here to get an estimate for the functions $g_j(\hat{Z})$, $j = 0, \dots, p-1$. Suppose, therefore, that the second derivative of g_j exists and is continuous. For each given z , $g_j(\hat{Z})$ is approximated locally by a linear function $g_j(\hat{Z}) \approx a_{j0} + a_{j1}(\hat{Z} - z)$ for \hat{Z} in a neighborhood of z . Note that a_{j0} and a_{j1} depend on z .

Then, for given β_0 with $\beta_{0p} \neq 0$, the local likelihood is given by

$$l(a_0, a_1) = \frac{1}{T} \sum_{t=1}^T l \left[g^{-1} \left\{ \sum_{j=0}^{p-1} (\hat{a}_{j0} + \hat{a}_{j1}(\hat{Z} - z)) x_{tj} \right\}, Y_t \right] K_h(\hat{Z}) w(\hat{Z}) \quad (3.32)$$

where h is the so-called bandwidth, which is a non-negative number controlling the size of the local neighborhood

$$I_h(z) = [z - h, z + h],$$

and, hence, the degree of smoothing. $K_h(\cdot) = h^{-1}K(\cdot/h)$, where K is the kernel function. We opted for using the Epanechnikov kernel

$$K(u) = \frac{3}{4}(1 - u^2)I_{[-1,1]}(u).$$

Other kernel functions that are widely used are the triangle kernel

$$K(u) = (1 - |u|)I_{[-1,1]}(u)$$

or the Gaussian kernel

$$K(u) = 2\pi^{-\frac{1}{2}} \exp\left(-\frac{u^2}{2}\right).$$

Further, in expression (3.32), $w(\cdot) = I_{[-L,L]}(\cdot)$ is a bounded weight function controlling the edge effect in the estimation. Maximizing the local likelihood function $l(a_0, a_1)$ gives estimates $\hat{a}_{MLE}(z)$, where $\mathbf{a}(z) = (a_{0,0}, \dots, a_{p-1,0}, a_{0,1}, \dots, a_{p-1,1})$.

Unlike the least-squares setting used in Fan et al. (2003), the solution for (3.32) generally does not have a closed form. Hence, one usually uses an iterative algorithm to obtain the solution, such as the Newton-Raphson method. Computing the local maximum likelihood estimate using an iterative method can, however, be very time-consuming. This is especially true for more complex models such as varying coefficient models as one needs to maximize the local likelihood (3.32) for many distinct values of z , with each maximization requiring an iterative algorithm. To reduce the computational cost, we follow Cai et al. (2000) who proposed to replace iterative local maximum likelihood estimation by an explicit non-iterative estimator, the so-called one-step Newton-Raphson estimator, which has been frequently used in parametric models.

Suppose that \hat{a}_{MLE} is the maximum likelihood estimator and $\hat{a}_0 = \hat{a}_0(z) = (\hat{a}_0(z)', \hat{a}_1(z)')'$ is an initial estimate with a reasonable good precision. Let, further, $l'(\mathbf{a})$ and $l''(\mathbf{a})$ be the gradient and Hessian matrix of the local likelihood $l(\mathbf{a})$, respectively. Then, using a Taylor expansion yields

$$0 = l'(\hat{a}_{MLE}) \approx \hat{a}_{OSE} = l'(\hat{a}_0) + l''(\hat{a}_0) (\hat{a}_{MLE} - \hat{a}_0). \quad (3.33)$$

Hence, the one-step local maximum likelihood estimator is obtained as

$$\hat{a}_{OSE} = \hat{a}_0 - [l''(\hat{a}_0)]^{-1} l'(\hat{a}_0). \quad (3.34)$$

We can further estimate the standard error of the resulting estimator by using the sandwich formula

$$\text{cov}(\hat{a}_{OSE}) = [l''(\hat{a}_0)]^{-1} \text{cov}(l'(\hat{a}_0)) [l''(\hat{a}_0)]^{-1},$$

following conventional techniques in the likelihood setting. The accuracy of this formula is tested by Cai et al. (2000).

In the case of a Poisson-distributed response, the updated estimator is then obtained by

$$\hat{a}_{OSE} = \hat{a}_0 + \begin{pmatrix} H_{t,0} & H_{t,1} \\ H_{t,1} & H_{t,2} \end{pmatrix}^{-1} + \begin{pmatrix} u_{t,0} \\ u_{t,1} \end{pmatrix}, \quad (3.35)$$

where the Hessian matrix has the form

$$H_{t,l} = \sum_{t=1}^T K_h(\hat{Z}) w(\hat{Z}) \hat{\mu}_{t0} (\hat{Z} - z)^l X_t X_t',$$

with $\hat{\mu}_{t0} = \exp\{\sum_{j=0}^{p-1} [\hat{a}_{j0} + \hat{a}_{j1}(\hat{Z} - z)] X_{tj}\}$ for $l = 0, 1, 2$.

Furthermore, $u_{t,j}$ is given by

$$u_{t,l} = \sum_{t=1}^T K_h(\hat{Z}) w(\hat{Z}) (Y_t - \exp\{\sum_{j=0}^{p-1} [\hat{a}_{j0} + \hat{a}_{j1}(\hat{Z} - z)] X_{tj}\}) (\hat{Z} - z)^l X_t,$$

for $l = 0, 1$.

A problem associated with the local quasi-likelihood estimator is that the matrix in (3.35) can be ill-conditioned. This may occur in certain local neighborhoods, if there are only a few data points. A commonly used technique to deal with this singularity problem is via ridge regression, as proposed by Seifert & Gasser (1996) or Fan & Chen (1999). In this case, an issue arises as how to choose the ridge parameter. Here, we follow the suggestions of Cai et al. (2000) and use the following ridge parameter,

$$r_{l,e} = \left(\frac{1}{T} \sum_{t=1}^T x_{te}^2 \right) \exp(\hat{g}'_0 \bar{x}) h^{l-1} \int z^l K(z) dz,$$

for the e -th diagonal element of the matrix $H_{t,l}$, $l = 0$ or $l = 2$ to attenuate a potential singularity.

We now turn to the choice of the initial estimator. Cai et al. (2000) show that such an estimator can be found as follows: Let z_g , $g = 1, \dots, T_{grid}$ be a number of grid points. Further, we specify by z_o distinct points within the range of the grid points. At these distinct points, the local maximum likelihood estimators $\hat{a}_{MLE}(z_o)$ are computed. The corresponding estimates are used as initial estimates for the points z_{o+1} . Applying the one-step algorithm (3.34) leads to $\hat{a}_{OS}(z_{o+1})$, which is then used as the initial estimate at point z_{o+2} , and so on. Likewise, $\hat{a}_{OS}(z_{o-1})$, ... can be computed.

3.5.1.2 Estimation of the index β_0 with $g_j(\cdot)$ fixed

Again, a one-step estimation scheme is used to estimate the index coefficient β_0 . We search for β_0 to minimize

$$R(\beta_0) = \frac{1}{T} \sum_{t=1}^T \left[Y_t - \exp \left(\sum_{j=0}^{p-1} g(\beta_0^T X - z) x_{tj} \right) \right]^2 w(\beta_0^T X).$$

For any initial value $\beta_0^{(0)}$ close to $\hat{\beta}_0$, one has the approximation

$$R'(\hat{\beta}_0) \approx R'(\beta_0^{(0)}) + R''(\beta_0^{(0)}) (\hat{\beta}_0 - \beta_0^{(0)}), \quad (3.36)$$

where R' is the derivative and R'' is the Hessian matrix of $R(\cdot)$.

Thus, we get the one-step iterative estimator

$$\beta_0^{(1)} = \beta_0^{(0)} - R''(\beta_0^{(0)})^{-1} R'(\beta_0^{(0)}). \quad (3.37)$$

Similar to the matrix in (3.35), the matrix $R''(\cdot)$ can be singular or nearly so. If this is the case, one has to add again a ridge regression parameter (see Fan et al. (2003)).

3.5.1.3 Estimation algorithm

The estimation algorithm is outlined as follows:

1. Standardize the data X_t such that they have sample mean 0 and the sample variance and covariance matrix I_d . This ensures that $\beta_0^T X$ has sample mean 0 and sample variance 1 for any β_0 . Specify an initial value of β_0 .
2. For each prescribed bandwidth value h_k , $k = 1, \dots, q$, the following two steps
 - a) For a given index β_0 , the functions $g_j(\cdot)$ are estimated by (3.35),
 - b) For given functions $g_j(\cdot)$, index coefficients β_0 are searched using algorithm (3.37),
 are iterated until the change in two successive values of $R(\beta_0)$ is smaller than the pre-specified measure of tolerance ($\epsilon = 0.00001$).
3. For $k = 1, \dots, q$ the optimal bandwidth value h_{opt} is calculated using the GCV criterion (see Fan et al. (2003) for more details).
4. For $h = \hat{h}_{opt}$ selected in 3., the steps (a) and (b) of 2. are repeated until convergence of $R(\beta_0)$.

Remarks To make the search for β_0 in part b) of step 2. more stable, an estimate of $g_j(\cdot)$ on a grid point is replaced by a weighted average on its five nearest neighbors. The edge points are adjusted correspondingly.

Finally, we make some remarks about the properties of the one-step algorithms. Cai et al. (2000) carefully study the properties of the local one-step estimator for the unknown functions g_j . They demonstrate on a theoretical basis that the one-step estimator and the fully iterative local maximum likelihood estimator share the same asymptotic distribution, provided that the initial estimate used for the one-step algorithm is good

enough. Consequently, the local one-step estimator does not lose efficiency compared to the local maximum likelihood estimator. Regarding the one-step estimator for the unknown index, convergence aspects require further research. Fan et al. (2003) only state that "*In practice, we may detect whether an estimated β_0 is likely to be the global minimum by using multiple initial values*".

Chapter 4

Simulation Study

To evaluate which of the approaches considered in Chapter 3 performs best in the context of air pollution data, we conducted a simulation study. Thereby, we focused on scenarios comparable to situations found in real data. We investigated models with one or several, but not all coefficients time-varying. Further, a model under the null hypothesis was estimated by time-varying estimation methods, that means the true effect was, in fact, time-constant. We further examined how well different components in the predictor can be identified and separated from one another.

4.1 Simulation setup

Our simulated time series of mortality counts were based on data collected within the study "Improved Air Quality and its Influences on Short-Term Health Effects in Erfurt, Eastern Germany". Data were available for the period October 1, 1991 to March 31, 2002 (the data will be described in more detail in Chapter 5).

The simulated time series were constructed using on the following general model:

$$Y_t \sim Po(\mu_t)$$
$$\log(\mu_t) = \log(\mu_0) * (\beta_1 relhum_t + f(t) + f(temp_t) + f_{CO}(t)CO_t). \quad (4.1)$$

The variables used in the simulated model consisted of time series of daily 24-h average relative humidity (*relhum*), trend (*t*), daily 24-h average temperature (*temp*) and daily 24-h average air pollution measurements (*poll*), respectively. As air pollution measurement, we considered carbon monoxide (*CO*) with a lag of four days. The lag structure as well as the range of the pollutant effects were chosen based on the results of a previous study in Erfurt from 1995-1998 (Wichmann et al. 2000). The average daily mortality

count, μ_0 , in Erfurt over the study period was found to be 4.62.

Relying on simulated data, we investigated different cases:

- 1 Constant effect of $f_{CO}(t) = 0.1$ and no confounders were included. Hence, this case can be regarded as a model under the null hypothesis.
- 2 The effect of $f_{CO}(t)$ is assumed to be curvilinear. We make the following specification (Bateson & Schwartz (1999)):

$$f_{CO}(t) = 0.05 \cdot \left(1 + \frac{2t}{T} - \frac{t^2}{T^2}\right). \quad (4.2)$$

The corresponding plot is shown in the top row of Figure 4.1. No additional confounders are considered in this case.

- 3 The effect of $f_{CO}(t)$ is modeled as a cosine function, which is specified as

$$f_{CO}(t) = 0.12 \cdot \cos(0.002t). \quad (4.3)$$

The simulated effect is displayed in the bottom row of Figure 4.1. No additional confounders are included.

Another question is whether the methods are able to distinguish time-varying from time-invariant coefficients, when several, but not all coefficients are time-varying. Further, we investigated a time-constant model with additional confounder variables. These questions are considered in the following cases.

- 4 $f_{CO}(t)$ is assumed to be curvilinear, as in (4.2). Additionally, the trend $f(t)$, a typical confounder variable, is simulated as the product of a sine function with constant amplitude and a linear pattern:

$$f(t) = 0.04 \cdot \left[\left(1 + \frac{t}{T}\right) \left(1 + \left\{0.6 \cdot \sin\left(\pi \cdot \frac{t}{365}\right)\right\}\right) \right], \quad (4.4)$$

which is a slightly modified version of the formula used by Bateson & Schwartz (1999). The simulated trend effect is shown in Figure 4.2.

- 5 The effect of $f_{CO}(t)$ is assumed to be constant. The effect of calendar time $f(t)$ is again simulated as a sine function, similar to (4.4). Further, effects of temperature and relative humidity are included. The effect of temperature is defined as sine function:

$$f(temp) = -0.02 \cdot \sin(0.1 temp), \quad (4.5)$$

as shown in Figure 4.2, middle row. A linear, constant-coefficient effect $relhum = -0.0015$ is considered for relative humidity.

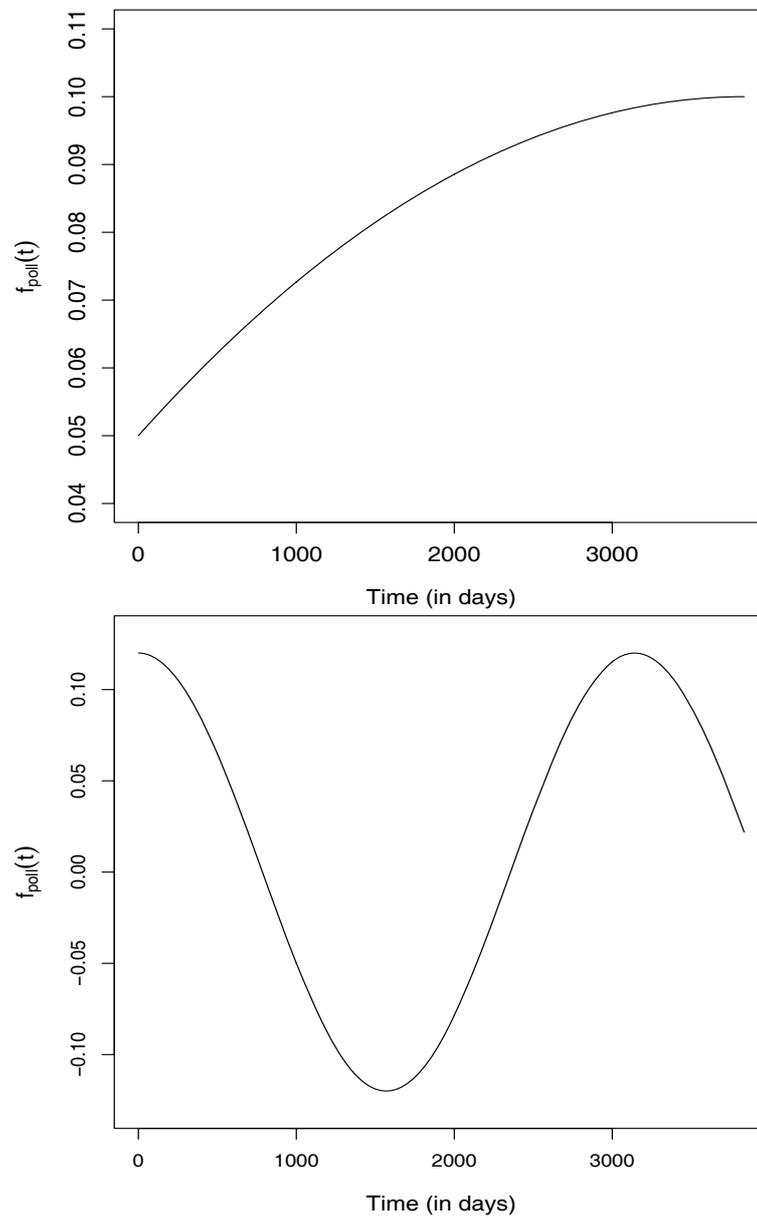


Figure 4.1: Simulated effect of $f_{CO}(t)$ was curvilinear (top) or followed a cosine function (bottom).

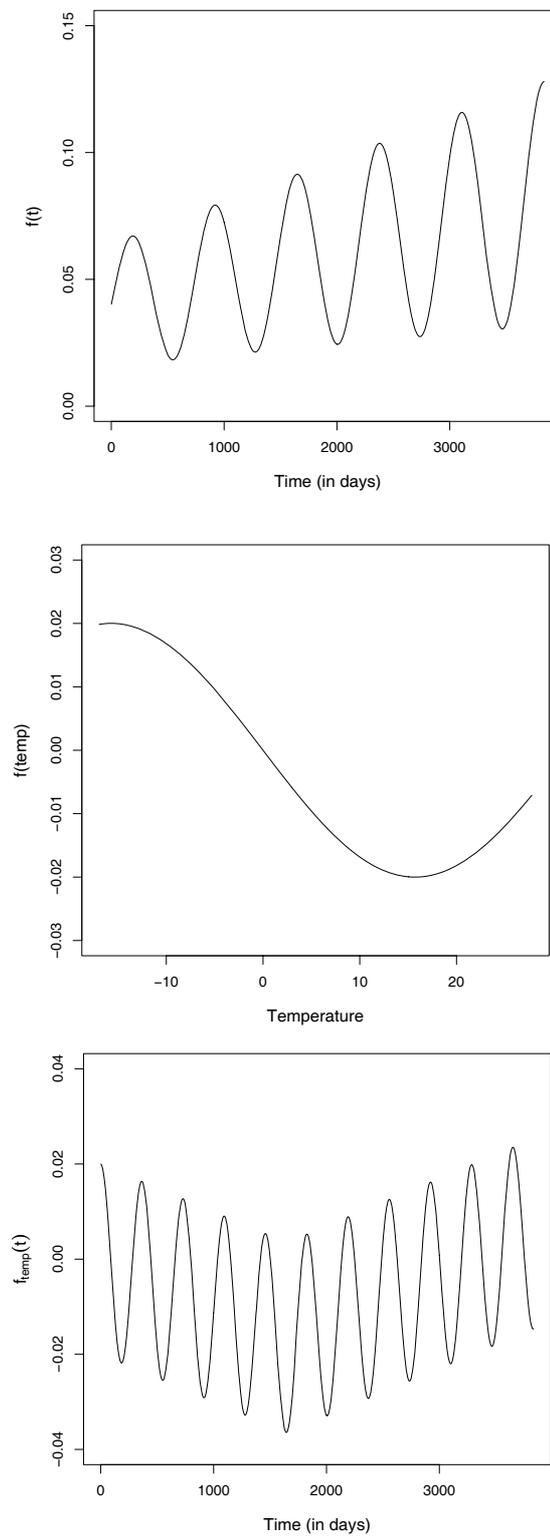


Figure 4.2: Simulated effect of the confounder trend $f(t)$ (upper panel), of temperature $f(temp)$ (middle panel) and simulated time-varying effect of temperature $f_{temp}(t)$ (lower panel) for the example *CO*.

- 6 The effect of $f_{CO}(t)$ is specified as in (4.3), the effect of calendar time $f(t)$ is similar as in (4.4) and both, simulated temperature and relative humidity effects are the same as in Case 5.
- 7 In the last case, the effect of $f_{CO}(t)$ is specified as in (4.3), the effect of calendar time $f(t)$ is assumed as in (4.4) and the effect of relative humidity is again linear and time-invariant. Additionally, we include a time-varying effect of temperature, specified as

$$f_{temp}(t) = \begin{cases} -0.00001 t + \left[0.02 \cdot \cos \left(2\pi \cdot \frac{t}{365}\right)\right], & \text{for } t < 1720, \\ -0.033 + 0.00001 t + \left[0.02 \cdot \cos \left(2\pi \cdot \frac{t}{365}\right)\right], & \text{for } t \geq 1720. \end{cases} \quad (4.6)$$

The seven simulated scenarios are summarized in Table 4.1. For each of the cases listed above, we generated counts $Y_t^{(n)}$, $t = 1, \dots, 3835$, using expression (4.1) for simulation runs $n = 1, \dots, N = 250$. We assessed the performance of the different time-varying estimation procedures using the following specifications (see also Chapter 3):

- a) Time-varying coefficient models using regression splines. For estimation, we first chose the optimal degrees of freedom of the splines by using the AIC.
- b) Bayesian time-varying coefficient models with P-splines. The MCMC specification for each model was 20000 iterations, whereby the first 4000 iterations were discarded as burn-in. Every 15th subsequent sample point was saved for estimation of posterior means. We specified the number of knots for the unknown functions to be 20.
- c) Time-varying coefficient models with penalized linear splines based on a generalized linear mixed model framework.
- d) Time-varying coefficient models with P-splines based on empirical Bayes inference. In case of no convergence in the empirical Bayes approach, we used the final values after the maximum number of iterations (400) to compute empirical root mean squared errors. Fahrmeir et al. (2004) found in their simulation analysis that "*a closer inspection of estimates with and without convergence showed that differences in terms of MSE are negligible and the choice of the final values leads to reasonable estimates*". Comparable to the fully Bayesian approach, the number of knots was specified to be 20.
- e) Adaptive generalized TVC models. For estimation, an Epanechnikov kernel was used. Further, the number of the bandwidth values was chosen to be $q = 15$ and

Table 4.1: Simulation scenarios for the example *CO*.

Case	$f_{poll}(t)$	$f(t)$	$f(temp)$	$\beta_1 relhum$
1	constant (= 0.1)	–	–	–
2	curvilinear	–	–	–
3	cosine function	–	–	–
4	curvilinear	sine function	–	–
5	constant	sine function	sine function	constant (= -0.0015)
6	cosine function	sine function	sine function	constant (= -0.0015)
7	cosine function	sine function	$f_{temp}(t)$	constant (= -0.0015)

the range was $h_k = 0.2 \times 1.2k - 1$. To speed up computation, the functions $f_j(\cdot)$ were estimated on 200 grid points.

As there were some highly oscillating functions under consideration in case 7, we allowed for a larger number of knots, in the approaches a) to d). Further, the bandwidth value in approach e) was changed accordingly.

All approaches were compared based on the empirical root mean squared error (RMSE) defined as

$$RMSE(\hat{f}_{CO}(t)) = \sqrt{\frac{1}{T} \sum_{t=1}^T (f_{CO}(t) - \hat{f}_{CO}(t))^2},$$

for the time-varying air pollution effect,

$$RMSE(\hat{f}_{temp}(t)) = \sqrt{\frac{1}{T} \sum_{t=1}^T (f_{temp}(t) - \hat{f}_{temp}(t))^2},$$

for the time-varying temperature effect,

$$RMSE(\hat{f}(t)) = \sqrt{\frac{1}{T} \sum_{t=1}^T (f(t) - \hat{f}(t))^2},$$

for the trend effect $f(t)$, and

$$RMSE(\hat{f}(temp)) = \sqrt{\frac{1}{T} \sum_{t=1}^T (f(temp_t) - \hat{f}(temp_t))^2},$$

for the time-constant, but non-linear temperature effect, where T is the number of observations.

Note that our simulation framework did not address the issue of measurement error in the pollutant variable. This problem will be investigated in more detail in Chapters 6 to 8.

4.2 Results

In this section, we present the results of the comparison of the five approaches. Each Figure shows boxplots of the empirical root mean squared errors (RMSE) of approaches

- a) TVCM with an interaction term between the pollutant and a regression spline in time (REGSPLINE);
- b) Bayesian TVCM with P-splines (MCMC);
- c) TVCM with penalized linear splines based on a mixed model framework (MIXEDLIN);
- d) TVCM with Bayesian P-splines based on empirical Bayes inference (MIXEDB) and
- e) Adaptive varying coefficient generalized linear models (AGVCM).

Additionally, the corresponding averaged estimates are shown for each case. For comparison the true functions are always included in the plots (solid black lines).

Case 1 The first case investigated the consequences of estimating a time-varying coefficient, although the true effect is constant. This case can be regarded as a model under the null hypothesis. If one wrongly estimates a time-varying coefficient model, there is the danger of deriving time-variation in the coefficients; however, they are in fact time-constant.

The boxplots of the RMSE for this case are given in the top row of Figure 4.3. The time-varying estimates for $f_{poll}(t)$ were relatively stable for approaches a) (REGSPLINE), b) (MCMC) and d) (MIXEDB). Hence, estimating a time-varying coefficient instead of the true time-invariant model defined in Case 1 did not seem to produce relevant coefficient variation for these three estimation methods. The RMSE values of approaches c) (MIXEDLIN) and e) (AGVCM) suggest that the estimates of these methods are more biased.

The bottom row of Figure 4.3 shows the corresponding plot of the averaged estimated

time-varying coefficients. It can be seen that averaged estimated coefficients of approaches c) (MIXEDLIN) and e) (AGVCM) deviate from the true function (solid black line) especially at the end of the study period. These deviations might also have caused the larger RMSE values for the two methods.

Cases 2 and 3 The simulated time-varying coefficients in the cases 2 and 3 were assumed to be curvilinear or to follow a cosine function, respectively.

Boxplots of the RMSE values of the five estimation methods for these cases are shown in Figures 4.4 and 4.5. While the results obtained for method a) (REGSPLINE) in case 2 were comparable to those in case 1, this approach had the largest RMSE values in case 3. This finding is corroborated by the corresponding averaged estimated time-varying coefficients, shown at the bottom row of Figures 4.4 and 4.5. Figure 4.5 shows considerable bias in the averaged estimated time-varying coefficients obtained by approach a).

Case 4 In this case, we examined whether the methods are able to distinguish time-varying from time-invariant coefficients. In particular, we assumed $f_{poll}(t)$ to be curvilinear and additionally included a time trend, which was simulated as a sine function (see Figure 4.2, upper panel).

Boxplots of the RMSEs of $f_{CO}(t)$ for this case are shown in Figure 4.6, top left panel. The two Bayesian approaches showed the smallest RMSE values.

For method c) (MIXEDLIN), there were large convergence problems in this case, which is reflected in the large RMSE values as well as the averaged time-varying estimates (Figure 4.6, top right). Further, approaches a) (REGSPLINE) and e) (AGVCM) showed considerable deviations from the true simulated curve (Figure 4.6, top right). For approach a) (REGSPLINE), the time-varying estimates even seemed to absorb the sine pattern of the trend.

The corresponding boxplots of the RMSE values for the time trend $f(t)$ are shown on the bottom row of Figure 4.6. Again approaches b) (MCMC) and d) (MIXEDB) had the smallest RMSE values. However, the plot of the averaged trend estimates revealed that all approaches estimated much smoother effects than assumed and, hence, had large bias.

Case 5 This case extends the model of case 1, as we further included a number of confounders. The aim of this case was to see if the considered approaches were able to detect the time-invariant air pollution effect in the presence of other time-invariant covariates.

The plots on the left-hand side of Figure 4.7 show the RMSE of the different simulated functions. Regarding the time-constant air pollution effect, approach d) (MIXEDB)

showed the smallest RMSE values, followed by method b) (MCMC). Again, approaches a) (REGSPLINE) and e) (AGVCM) showed considerable deviations from the true simulated curve (Figure 4.7, top right). Further, the time-varying estimates obtained by approach a) (REGSPLINE) again seemed to absorb the sine pattern of the trend, as can be seen in the wiggly averaged estimate.

Averaged estimates of the trend revealed that, again, all approaches estimated a smoother curve compared to the true simulated trend effect (Figure 4.7, middle right). However, the two Bayesian approaches obtained the average of the true effect, and therefore revealed the smallest RMSE values. Regarding the time-constant temperature effect, method d) (MIXEDB) gave the smallest RMSE values (Figure 4.7, bottom left), whereas approaches a) (REGSPLINE) and e) (AGVCM) showed some deviations from the true simulated function (Figure 4.7, bottom right).

Case 6 The sixth case describes a model which is of particular interest in applied work of air pollution analysis, since it contains trend and meteorological confounders.

With respect to the time-varying coefficient $f_{CO}(t)$, the estimates from methods b) (MCMC) and d) (MIXEDB) seemed to exploit the variation quite well, although there were a number of outliers for the latter approach (4.8, top left).

The boxplot of the RMSE for the time trend $f(t)$ is given in the middle row of Figure 4.8. Method d) (MIXEDB) performed best in terms of RMSE, followed by approach b) (MCMC). The corresponding averaged estimates show again that all approaches estimated a much smoother curve for the time trend (Figure 4.8).

Finally, the boxplots of the RMSE and the averaged estimates for $f(\text{temp})$ are given on the bottom panels of Figure 4.8. Again, method d) (MIXEDB) performed best.

Case 7 Finally, case 7 considers a model with more than one time-varying coefficient term. As the functions $f(t)$ and $f_{temp}(t)$ were both assumed to be highly oscillating, we allowed for more knots or a smaller bandwidth, and, hence, more variability in the functional form.

Boxplots of the RMSE as well as averaged estimates of the time-varying effect $f_{CO}(t)$ and of the trend showed results comparable to case 6 (see Figure 4.9, top and middle row). Overall, method d) (MIXEDB) performed slightly better than approach b) (MCMC) in terms of RMSE; however, the empirical Bayes method showed a number of spikes. In general, all approaches besides method a) (REGSPLINE) detected the functional form of the simulated time-varying temperature effect (Figure 4.9, bottom right). The estimates obtained by approach c) (MIXEDLIN) show some deviations from the true curve caused by the modeling with linear splines (Figure 4.9, bottom right).

Summary

Taken together, the simulations suggested that

- Fully Bayesian models with P-splines (method b)) and models with P-splines based on empirical Bayes inference (method d)) performed quite well in terms of RMSE within all simulated cases and mostly outperformed the other approaches.
- The two methods furthermore seemed to be able to distinguish time-varying from time-invariant coefficients when some, but not all coefficients were time-varying.
- Method a) (REGSPLINE), which does not use any form of shrinkage, mostly performed inferior compared to the approaches using a penalty (methods b) to d)). This corroborates the well known observation that, on general principles, shrinkage is desirable in estimation problems with many parameters.
- The approaches using B-spline basis functions outperformed the two approaches based on truncated power series bases, as could have been expected. Reasons might be the use of penalized linear splines for method c) (MIXEDLIN), but also better numerical properties of the B-spline bases. For example, Eilers & Marx (2004) show in the case of generalized additive models that penalized B-splines have improved numerical properties compared to splines based on the truncated power series basis.
- The assumption of global variances (or smoothing parameters) may be inappropriate in cases with rapidly-varying curvature and might be replaced by locally adaptive variances (or smoothing parameters) to improve the estimation of functions. This aspect is discussed in detail for Gaussian responses by Lang, Fronk & Fahrmeir (2002), for binary responses by Jerak & Lang (2005) and in a general context by Krivobokova, Crainiceanu & Kauermann (2006). A further possibility could be given by increasing the number of knots for the corresponding functions (as it was done in case 7).
- Based on the simulation results, we additionally investigated the coverage of point-wise credible intervals obtained by the two Bayesian approaches. We therefore computed empirical coverage probabilities, that means, we calculated relative frequencies indicating how often the true simulated functions were covered by the credible intervals of the corresponding estimates. In the fully Bayesian approach credible intervals are obtained by computing the respective quantiles of the sampled function evaluations. For empirical Bayes estimates, credible intervals are computed as described in Section 3.4.2. Considering the coverage properties of

Table 4.2: Average empirical coverage probabilities for a nominal level of 95% obtained for the two Bayesian approaches.

Case	Method	$f_{poll}(t)$	$f(t)$	$f(temp)$ or $f_{temp}(t)$	$\beta_1relhum$
1	MCMC	0.968	–	–	–
1	EB	0.948	–	–	–
2	MCMC	0.956	–	–	–
2	EB	0.930	–	–	–
3	MCMC	0.956	–	–	–
3	EB	0.916	–	–	–
4	MCMC	0.984	0.968	–	–
4	EB	0.924	0.914	–	–
5	MCMC	0.976	0.988	0.952	0.951
5	EB	0.956	0.860	0.984	0.960
6	MCMC	0.972	0.968	0.952	0.953
6	EB	0.924	0.916	0.976	0.940
7	MCMC	0.98	0.972	0.976	0.952
7	EB	0.936	0.928	0.968	0.960

pointwise credible intervals for a nominal level of 95%, average coverage rates obtained for the fully Bayesian approach were always at least slightly above the nominal level, ranging from 95.1% up to 98.8%, see Table 4.2. This indicates that the fully Bayesian approach yields rather conservative credible intervals, an observation already found in other Bayesian applications (see for example Fahrmeir et al. (2004) or Lang & Brezger (2004)). While the average coverage probabilities obtained by the empirical Bayes approach revealed coverage rates above the nominal level of 95% for temperature and relative humidity effects, they are in part slightly below the nominal level for the time-varying effect of CO and, in particular, for the trend.

As a consequence, we considered the use of fully Bayesian models with P-splines (method b)) and time-varying coefficient models with P-splines based on empirical Bayes inference (method d)) for the analysis presented in the next chapter.

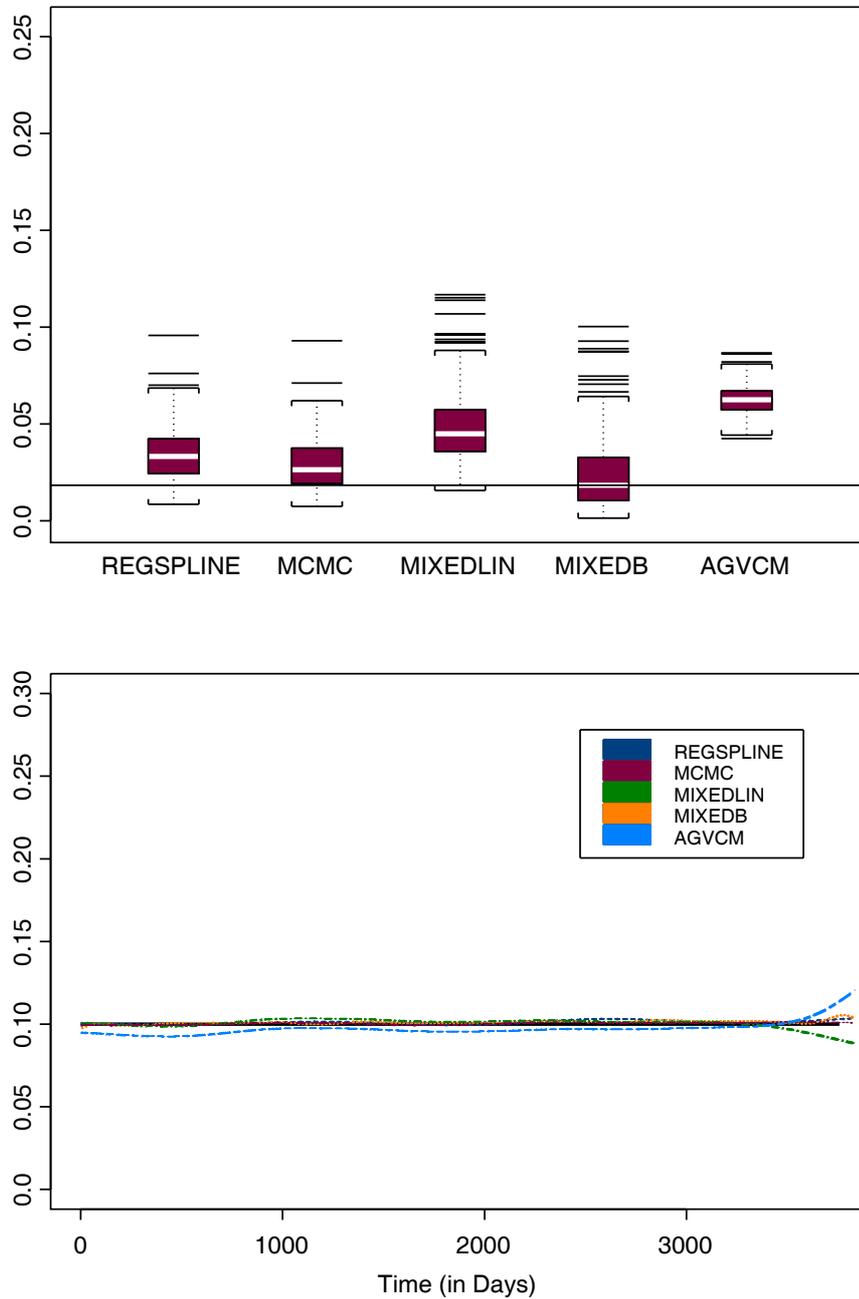


Figure 4.3: Case 1): RMSE of $f_{CO}(t)$ (top) and averaged estimates (bottom) for the example CO .

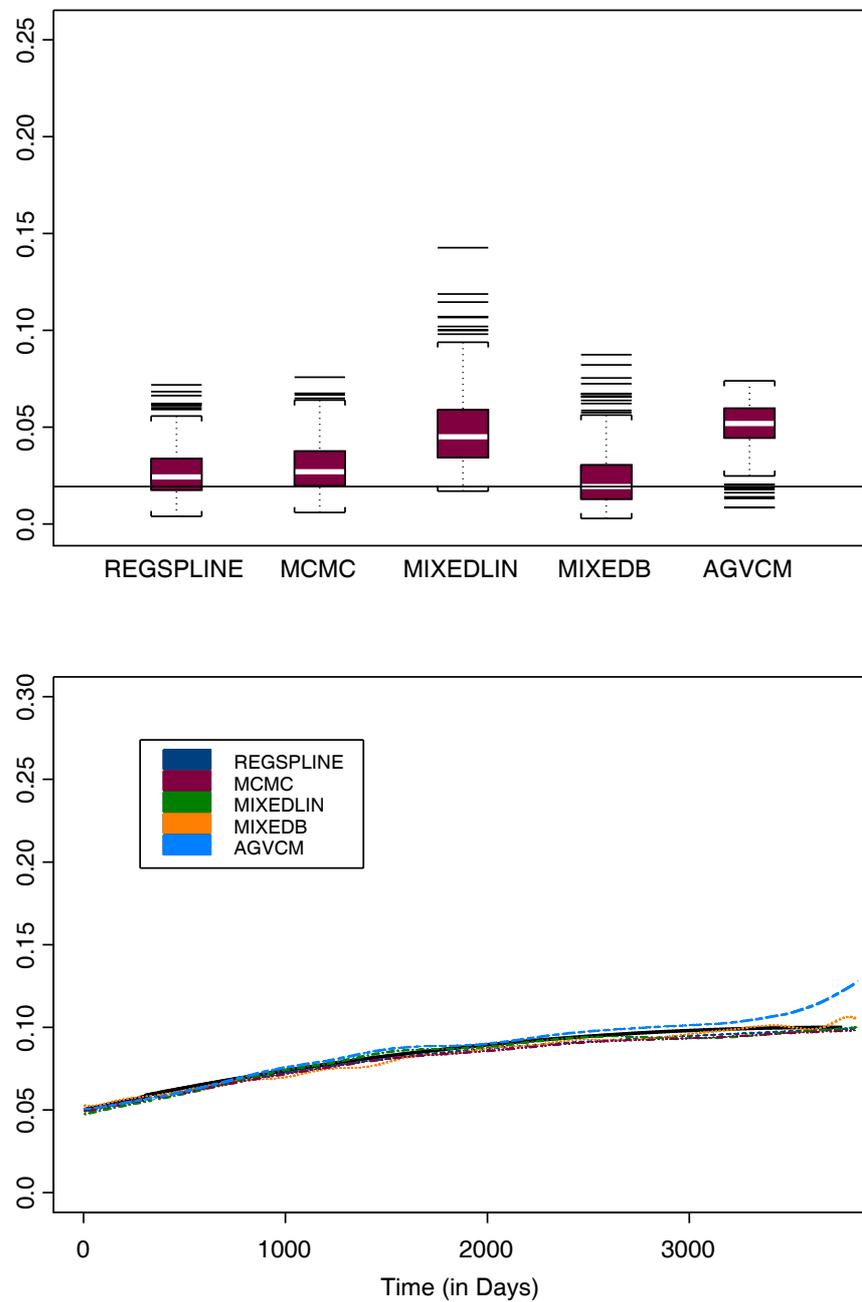


Figure 4.4: Case 2): RMSE of $f_{CO}(t)$ and averaged estimates (bottom) for the example CO .

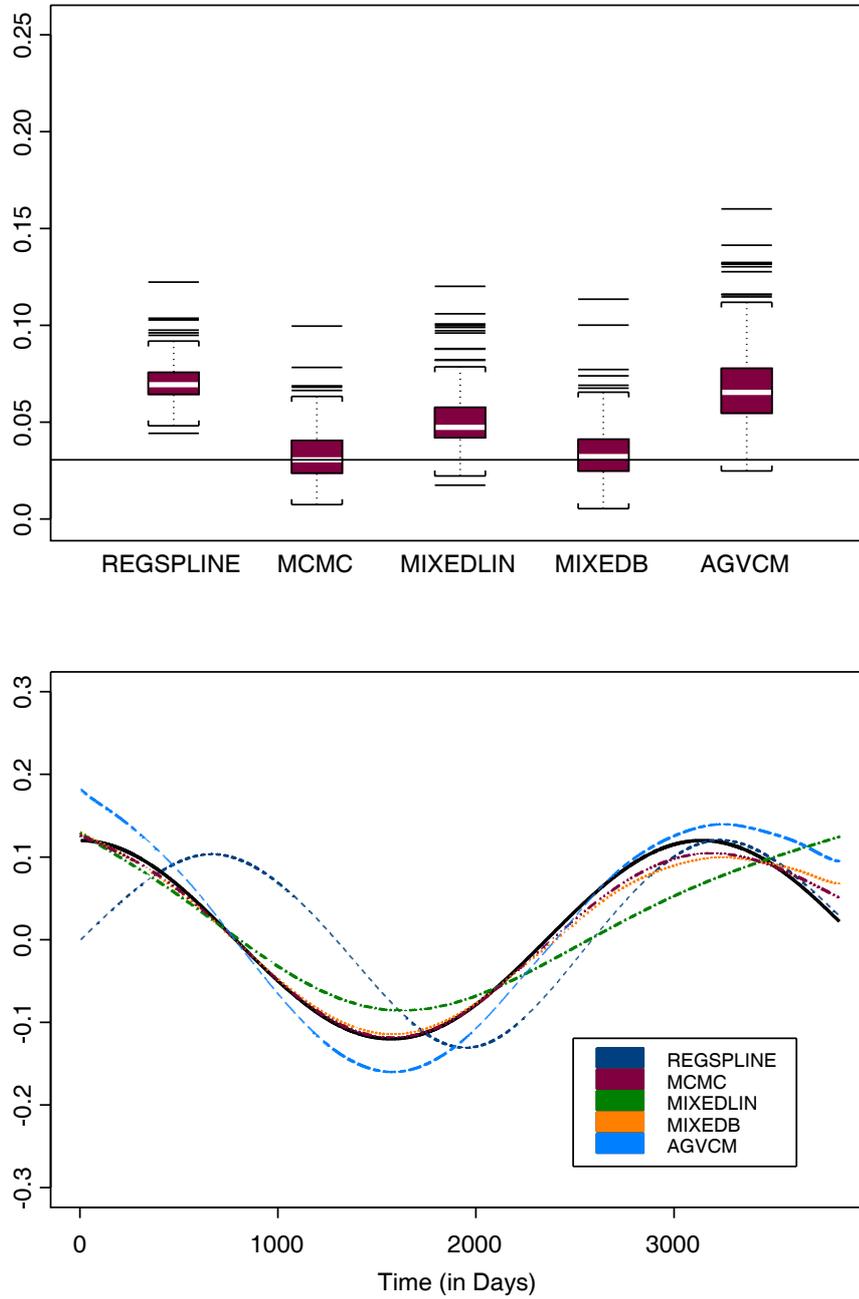


Figure 4.5: Case 3): RMSE of $f_{CO}(t)$ and averaged estimates (bottom) for the example CO .

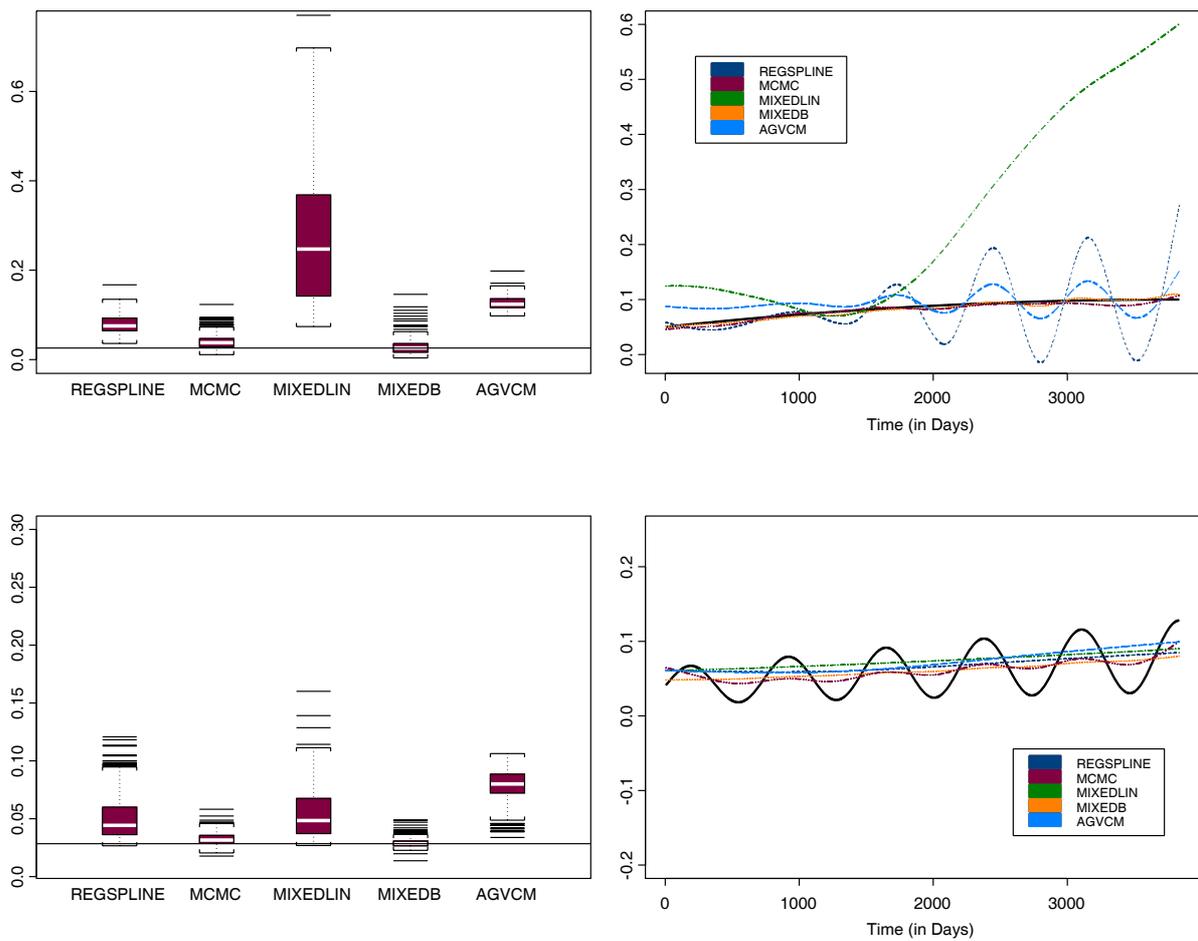


Figure 4.6: Case 4): RMSE of $f_{CO}(t)$ (top left) and averaged estimates (top right), and RMSE of $f(t)$ (bottom left) and averaged estimates (bottom right), for the example CO .

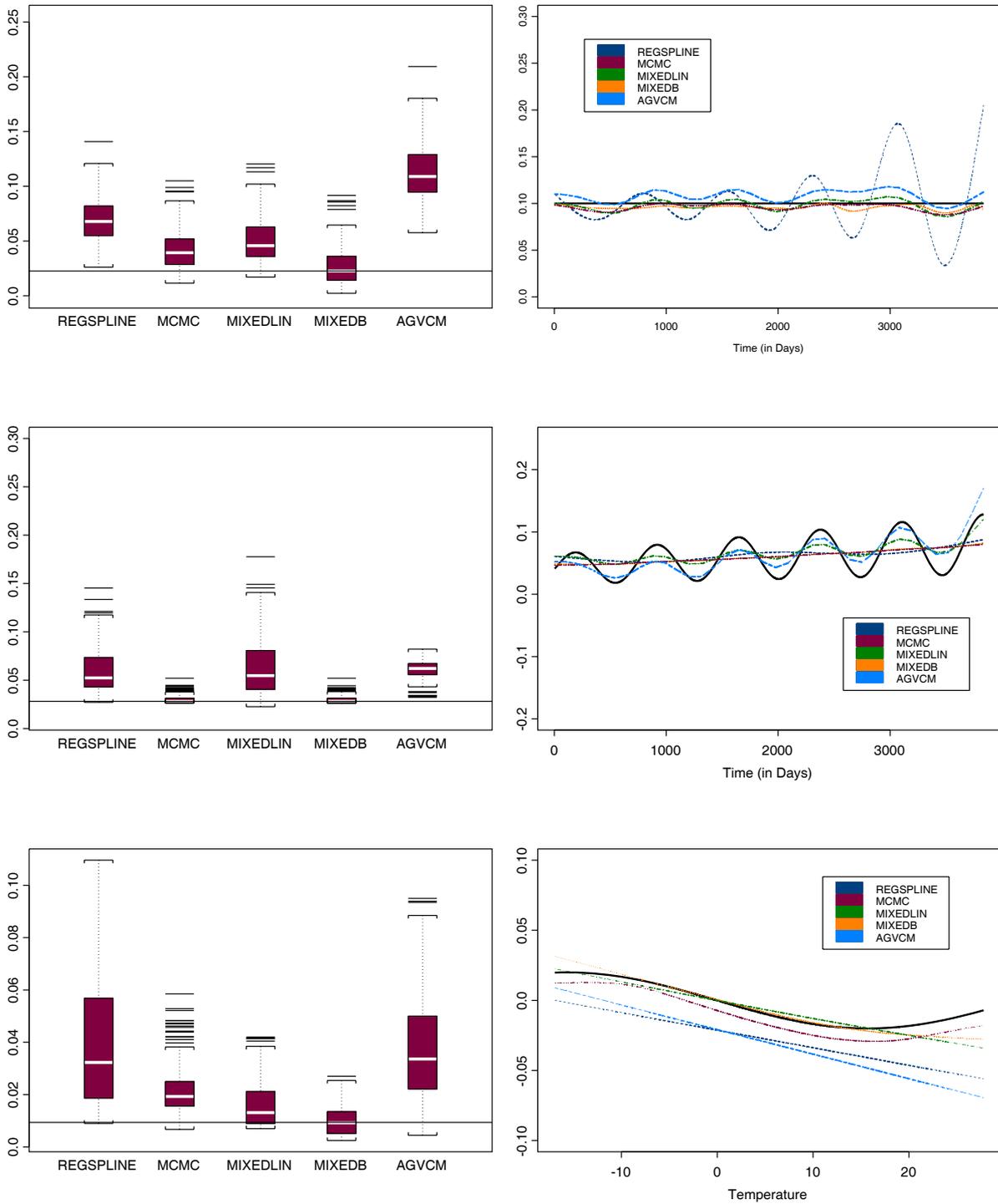


Figure 4.7: Case 5): RMSE of $f_{CO}(t)$ (top left) and averaged estimates (top right), RMSE of $f(t)$ (middle left) and averaged estimates (middle right), and RMSE of $f(temp)$ (bottom left) and averaged estimates (bottom right), for the example CO .

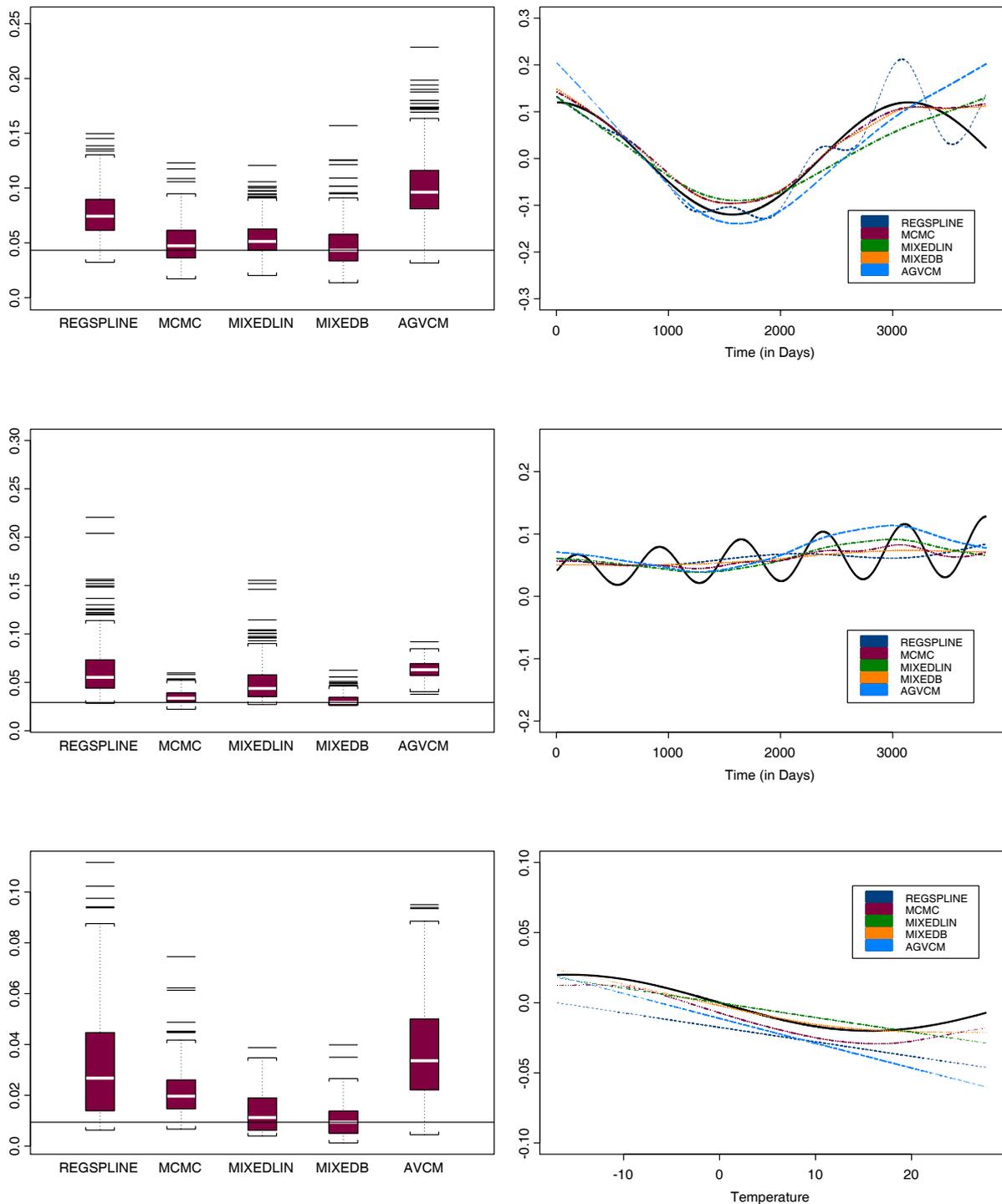


Figure 4.8: Case 6): RMSE of $f_{CO}(t)$ (top left) and averaged estimates (top right), RMSE of $f(t)$ (middle left) and averaged estimates (middle right), and RMSE of $f(temp)$ (bottom left) and averaged estimates (bottom right), for the example CO .

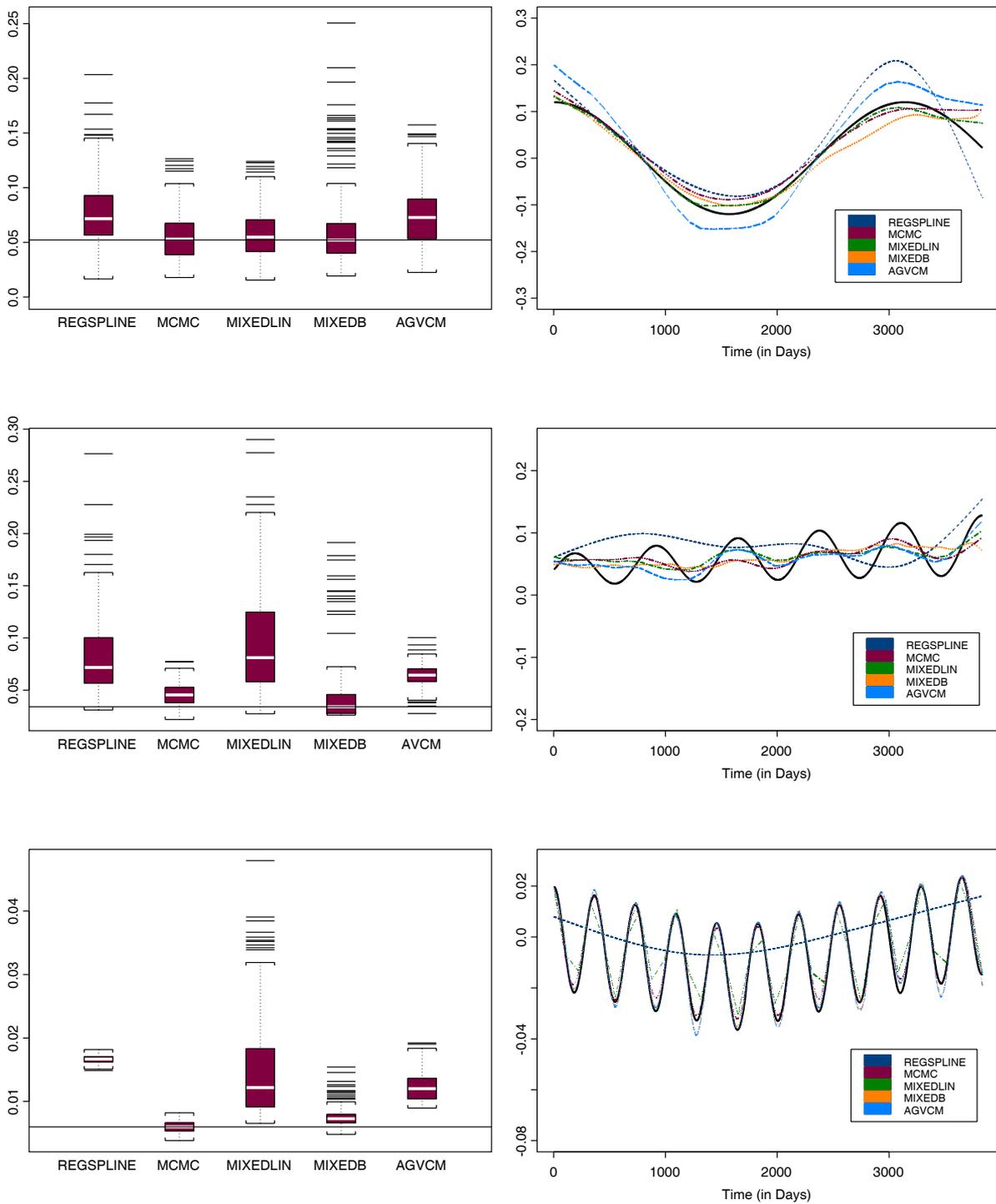


Figure 4.9: Case 7): RMSE of $f_{CO}(t)$ (top left) and averaged estimates (top right), RMSE of $f(t)$ (middle left) and averaged estimates (middle right), and RMSE of $f_{temp}(t)$ (bottom left) and averaged estimates (bottom right), for the example CO .

Chapter 5

Analysis of the Erfurt data

In this Chapter, those two methods selected based on the simulation study (Chapter 4), that is Bayesian varying coefficient models with P-splines (method b)) and varying coefficient models with P-splines based on empirical Bayes inference (method d)) are applied to time-series data. The analysis is based on data collected within the study "Improved Air Quality and its Influences on Short-Term Health Effects in Erfurt, Eastern Germany". We analyze the time-varying dependence of the daily counts of deaths on air pollution. However, it is known that air pollution effect estimates on mortality could be affected by observed and unobserved confounders such as weather variables or season that vary in a similar manner as the air pollution and mortality time series. To account for these, smooth functions of calendar time, influenza epidemics, and weather are also included.

In the following, we first present and describe the data, as well as the models applied. Secondly, we describe the most important results and draw some conclusions.

5.1 Data and confounder model description

Daily mortality (daily counts of deaths), 24-h average air pollution concentrations as well as additional confounder variables were collected in Erfurt for the time period October 1991 to March 2002.

Daily mortality Copies of death certificates without the names and addresses of the deceased were obtained from the local health authorities to comply with the rules of the German data privacy law. Daily counts of deaths of Erfurt residents, who died within the city, were calculated for total non-trauma deaths, that means for ICD-9 (International Classification of Diseases, 9th revision) codes below 800, and for ICD-10 (10th revision)

Table 5.1: Summary statistics for daily mortality, air pollution concentrations and key meteorology.

Variable	Period	Mean (SD)	Min	25%	Median	75 %	Max
Daily counts of death	1991 - 2002	4.6 (2.2)	0	3	4	6	15
Daily counts of death	1995 - 2002	4.90 (2.3)	0	3	5	6	15
CO (mg/m^3)	1991 - 2002	0.63 (0.54)	0.08	0.29	0.48	0.78	5.77
UFP (cm^{-3})	1995 - 2002	12888 (8661)	717	6800	10400	16218	63181
$PM_{2.5}$ ($\mu g/m^3$)	1995 - 2002	21.9 (18.7)	1.6	10.5	16.0	26.8	212.0
Temperature ($^{\circ}C$)	1991 - 2002	8.3 (7.6)	-16.9	2.7	8.4	14.3	27.8
Relative humidity (%)	1991 - 2002	79.5 (11.2)	34.3	72.4	80.7	87.5	100

codes less than S00. Infants (less than one year) were excluded.

Due to an administrative reform in 1994, a number of autonomous communities were incorporated into the city of Erfurt. Before this date, all death certificates of the inhabitants of these communities were stored in these villages and were therefore not available. More details can be found in Peters et al. (2007). For the time series analysis considering the gaseous pollutants as exposure, data were available from 1991 onwards. Therefore, only deaths occurring within the old city limits of Erfurt were considered. For the analysis of particle number and mass concentrations, which were available from 1995 onwards, also deaths in the incorporated communities were included.

Ambient air pollution Ambient gaseous pollutant concentrations were obtained from a state-run network monitoring station in Erfurt for the whole study period from October 1991 to March 2002. Data were available for CO , NO , NO_2 , SO_2 and ozone (O_3), from which CO was used in this application. Particles in different size ranges were measured at a research site from September 1995 to March 2002. Data were available for mass and number concentration of ultrafine particles (UFP), fine particles (less than 2.5 microns aerodynamic diameter ($PM_{2.5}$)), and coarse particles (less than 10 microns aerodynamic diameter (PM_{10})). In this application, we considered UFP number concentrations and $PM_{2.5}$ mass concentrations. Missing pollution values were imputed, where possible (see Peters et al. (2007)), and otherwise discarded.

Confounder variables Data on influenza epidemics were obtained from the "Arbeitsgemeinschaft Influenza" (German Influenza Working Group, AGI 2005). They consist of a weekly doctor's practice index (p_{winter}^i) for each winter season (October through April) indicating the relative deviation of the number of doctor visits due to acute respiratory symptoms in comparison to a background level (of value 100), averaged for the whole of Germany. We used the doctor's practice index for the whole of Germany since only very few doctor's practices in Thuringia participated in the survey system, especially during the first winters.

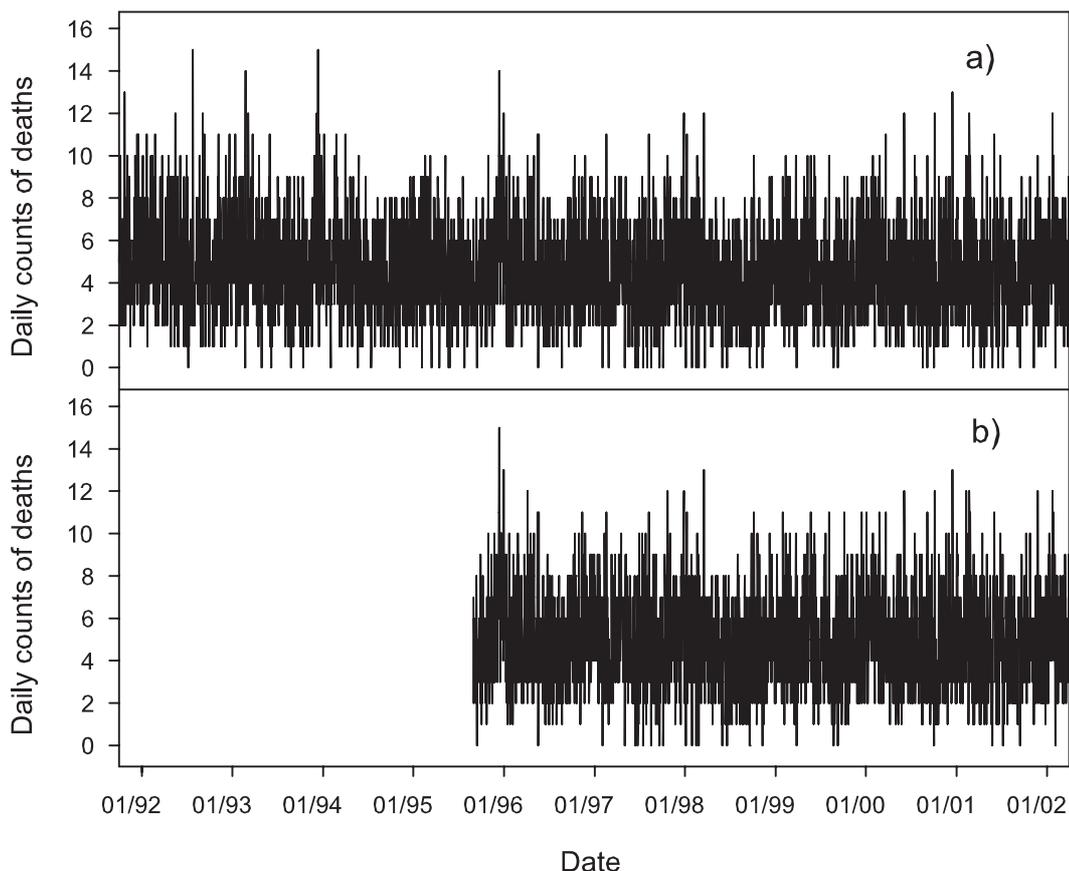


Figure 5.1: Time series of mortality counts in Erfurt, Germany, from October 1991 to March 2002 (a), old city limits) and September 1995 to March 2002 (b), new city limits).

Daily means of air temperature (*temp*) and relative humidity (*rh*) were obtained from a site of the German Meteorological Service (DWD, Deutscher Wetterdienst) located at the airport 5 km west of the research measurement station.

Table 5.1 includes summary statistics for daily counts of deaths, air pollution concentrations of carbon monoxide (*CO*), ultrafine particle counts (*UFP*) and $PM_{2.5}$ as well as key meteorology during the study period.

The differences between the two mortality series were due to the administrative reform in 1994 (as described above), which can also be seen in Figure 5.1. *CO* concentrations were available on 3761 days (98.1%), *UFP* number concentrations on 2403 days (91.3%) and $PM_{2.5}$ on 2351 days (97.8%). Daily average concentrations for the pollutants are shown in Figure 5.2.

A clear seasonal pattern could be observed for all these pollutants; peak pollutant concentrations occurred in winter. Further, a decreasing trend in the pollutant concentration

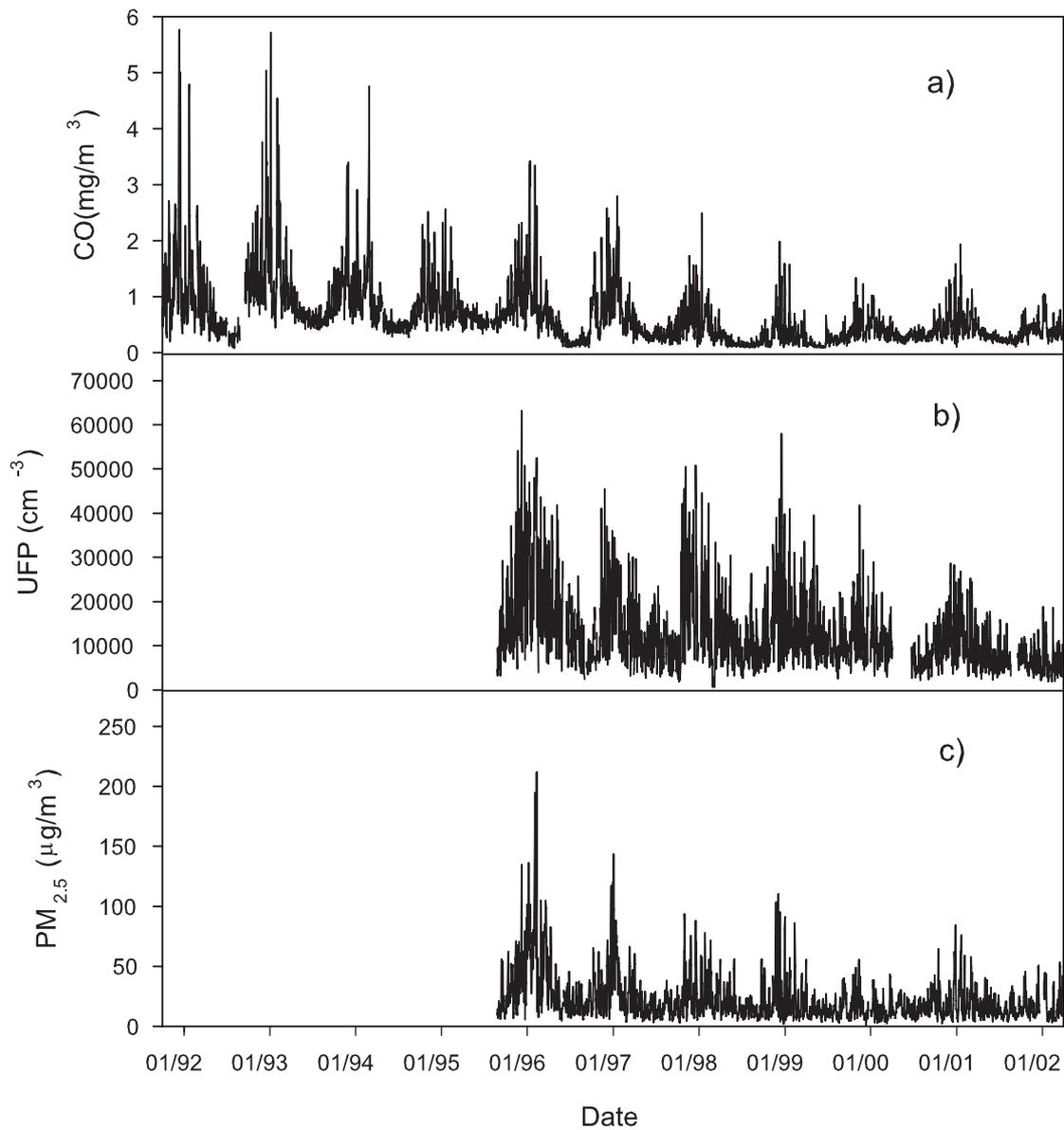


Figure 5.2: Time series of CO concentration (a) from October 1991 to March 2002, UFP number concentration (imputed) (b) and $PM_{2.5}$ mass concentration (imputed) (c) from September 1995 to March 2002.

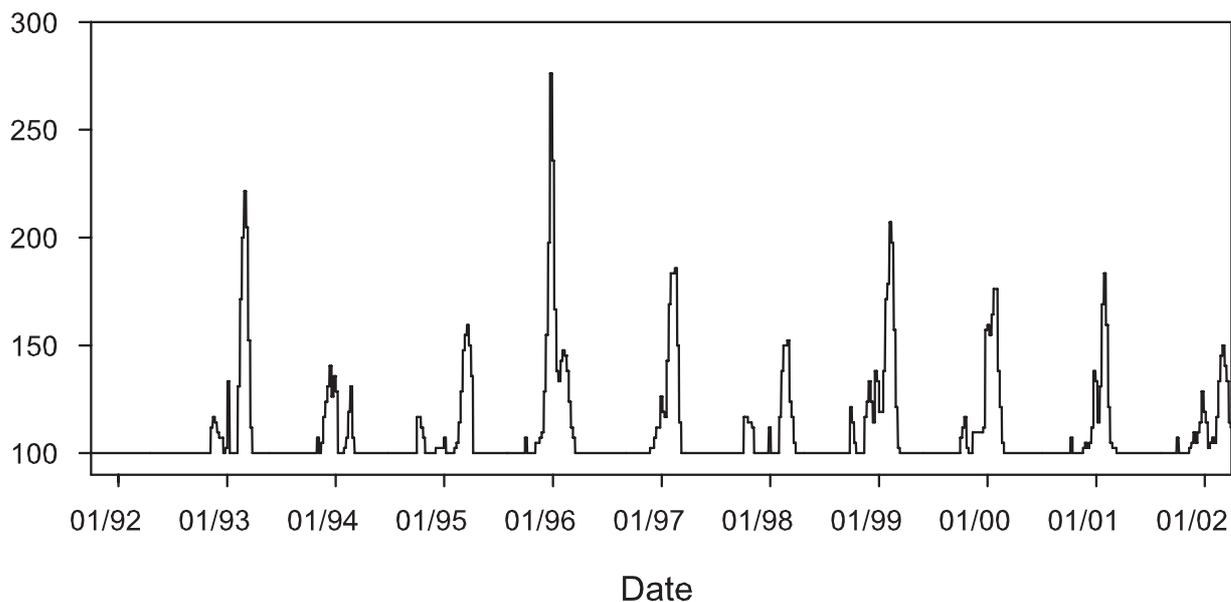


Figure 5.3: Doctor's practice index for all study winters

could be observed. For example, ambient CO levels decreased by 55% over the study period.

Figure 5.3 shows the doctor's practice index for all study winters except 1991/92, where no data were available. The relative deviations ranged from 100 to a maximum of 276 in the winter of 1995/1996.

Model building The models used in the application are given by

$$\begin{aligned}
 Y_t &\sim Po(\mu_t) \\
 \log(\mu_t) &= \log(pop) + confounder + f_{poll}(t) poll.
 \end{aligned}
 \tag{5.1}$$

Confounder model building was done in a similar way as in the APHEA II study (Touloumi et al. (2004) and Touloumi et al. (2006)) and is described in detail in Peters et al. (2007), Stölzel et al. (in press) and Stölzel et al. (submitted).

We fitted the confounders trend, season, major influenza epidemics, air temperature, and relative humidity by smooth functions. These terms were included step-by-step into the model, which, in turn, was assessed by the partial autocorrelation function (PACF) and the AIC. Furthermore, only terms with plausible dose-response-relationships were selected; for example, we assumed a priori a positive association of the doctor's practice index indicating the severity of an influenza epidemic and mortality. Lags of up to plus or minus three weeks were tested, because the peak of an influenza epidemic may hit Erfurt at another date than the whole of Germany. In the last step, potential calendar

effects were considered. Table 5.2 contains a description of all variables that were pursued in the regression analysis.

As *UFP* number concentrations and *PM_{2.5}* mass concentrations were only available between September 1995 and March 2002, we built separate models for the time periods from 1991 to 2002 and from 1995 to 2002.

Following this procedure, a final confounder model was built. In this final model the shape of the effect of temporal trend was double checked. This is necessary after other confounders, particularly meteorological variables, have been included into the model. Many of these variables exhibit seasonal patterns themselves and, thus, capture part of the observed seasonal trends in the outcome.

The final confounder model for the whole study period consisted of

$$\begin{aligned} \log(\mu_t) = & \log(pop) + f(t) + f(pi9394_{l_0}) + f(pi9495_{l_{-3}}) + f(pi9596_{l_{-1}}) \\ & + f(pi9900_{l_2}) + f(pi0001_{l_3}) + so + f(temp_{l_0}) + temp_{l_1} + f(rh_{l_2}), \end{aligned} \quad (5.2)$$

whereas the confounder model for the period 1995 to 2002 was given by

$$\begin{aligned} \log(\mu_t) = & \log(pop) + f(t) + f(pi9596_{l_{-1}}) + pi9899_{l_2} + pi9900_{l_2} \\ & + f(pi0001_{l_3}) + so + f(temp_{l_0}) + temp_{l_1} + f(rh_{l_2}). \end{aligned} \quad (5.3)$$

Cubic P-splines with 20 knots were estimated for calendar time (t) and depending on the model for the doctor's practice index during the winters 1993/94 with lag 0 ($pi9394_{l_0}$), 1994/95 with a lag minus three weeks ($pi9495_{l_{-3}}$), 1995/96 with lag minus one week ($pi9596_{l_{-1}}$), 1999/2000 with lag 2 weeks ($pi9900_{l_2}$), and 2000/01 with a lag of three weeks ($pi0001_{l_3}$). Further, splines were estimated for the meteorological confounders temperature with lag 0 days ($temp_{l_0}$) and relative humidity, lag 2 days (rh_{l_2}).

Air pollution models In the last step of the model building process, air pollution variables were included as in (5.1). Air pollution was first assessed linearly to select the best time lag (see Stölzel et al. (submitted)). For this analysis, a lag of four days was chosen for *CO* as well as a four-days lag for *UFP* number concentrations and same-day *PM_{2.5}* mass concentrations.

For the fully Bayesian approach, the first 10,000 iterations were discarded as burn-in. Further 50,000 iterations were then completed for the chain, which were thinned by 40 to reduce autocorrelation.

Table 5.2: Erfurt air pollution data: List of variables.

Variable	Description
y	daily counts of deaths
pop	population numbers per year
$trend$	calendar time (continuous)
pi_{winter}	doctor's practice index per winter (continuous)
so	indicator for Sundays
$temp$	temperature in °C (continuous)
rh	relative humidity in % (continuous)
CO	carbon monoxide concentrations in mg/m^3 (continuous)
UFP	ultrafine particle number concentrations (cm^{-3}) (continuous)
$PM_{2.5}$	particulate matter (continuous)

5.2 Results

We first present some exemplary results of the associations between confounders and mortality. Figure 5.4 displays the estimates of the time trend and one typical exposure-response-relationship for the doctor's practice index of the winter 1995/1996 obtained with the fully Bayesian (top panel) and the mixed model based empirical Bayes (lower panel) approach. Both methods lead to comparable estimates for the time trend, with a somewhat smoother trend for the empirical Bayes (EB) estimates. There was a rapidly decreasing trend in mortality until the beginning of 1996, followed by a slight increase up to mid 2000. In contrast, quite different results were obtained for the doctor's practice index. With the EB method a much smoother, basically linearly increasing effect was obtained. With the fully Bayesian approach, however, the pi_{winter} effect for the winter 1995/1996 was slightly increasing up to a relative deviation of 140, followed by a sharp increase until values of 200. Afterwards, it was slightly decreasing, and finally remained constant from a relative deviation of 230 upwards. Regarding the meteorology effects, again, much smoother posterior mode effects were obtained by applying the EB approach, resulting, for example, in a linearly increasing temperature effect. The posterior mean estimate of the nonlinear temperature effect also showed an overall increasing effect, however with some wiggles between temperatures of 0 and 15 °C. The nonlinear relative humidity effects exhibited a different shape. Starting from a positive value, they decreased over most of the humidity range before showing a slight increase from a relative humidity of 90 % onwards.

The time-varying estimated effects of CO with a four-days lag, $f_{CO_{14}}(t)$, using the two Bayesian methods are shown in Figure 5.6. The plot on the top shows the time-varying

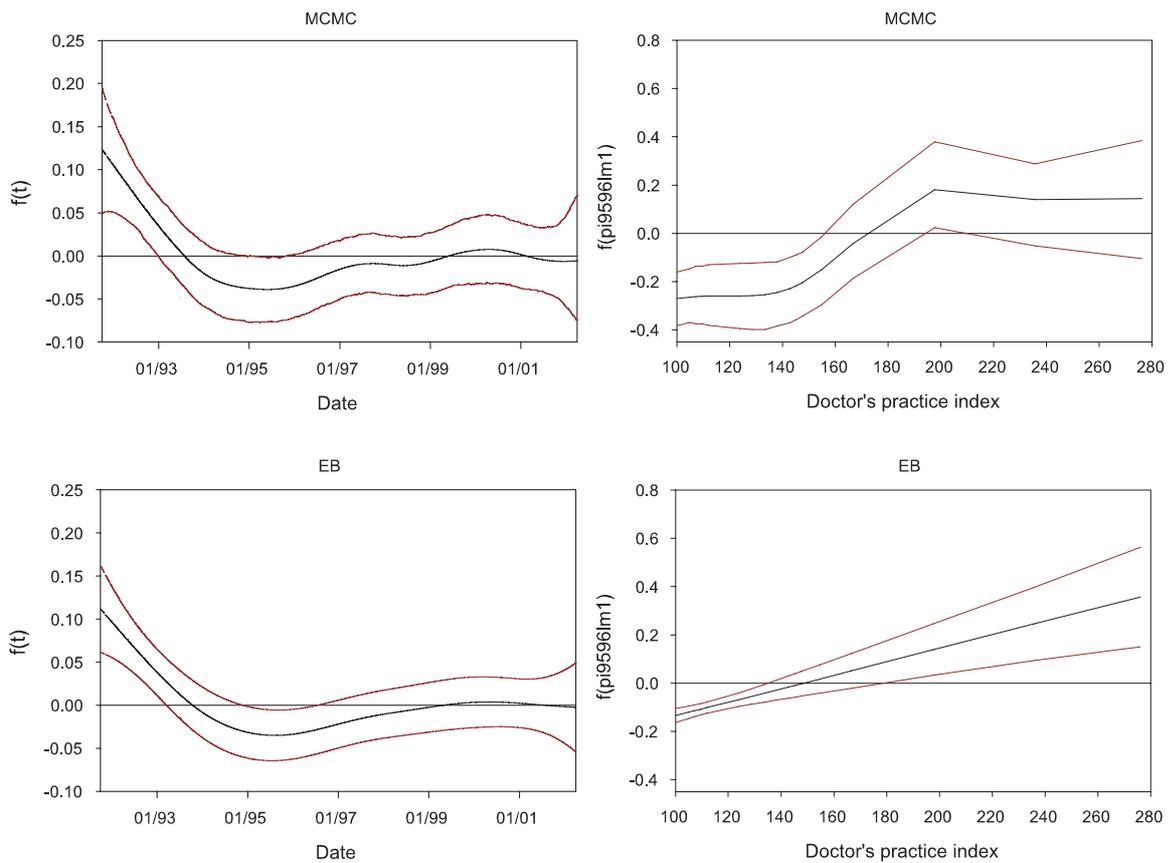


Figure 5.4: Posterior mean estimates (top) and posterior mode estimates (bottom) for the nonlinear effects of trend (left panel) and the doctor's practice index (right panel). All estimates are shown with pointwise 95% credible intervals.

posterior mean estimated using fully Bayesian method b) (MCMC) together with 95% pointwise credible intervals, whereas the plot for empirical Bayes method d) (EB, bottom) gives the time-varying posterior mode, again with 95% pointwise credible intervals. Both estimates have very similar shapes, showing a long-term trend with an effect that is increasing until mid 1997 and is afterwards decreasing. The MCMC estimate shows slightly more curvature over the study period.

The quantiles of the estimated time-varying CO effects for the two methods are shown in Table 5.3. It can be seen that the posterior mode of the empirical Bayes approach has a larger value than the posterior mean of the fully Bayesian model. This also holds true for the medians. The MCMC estimates have a greater range of values, which corroborates the visual assessment of more curvature in the time-varying coefficients.

In contrast to the long-term trend in the association of CO and daily mortality, the time-varying effect estimates of ultrafine particles showed some seasonal patterns (Figure 5.7). The increases at the end of the study period should be evaluated in view of the wide 95% credible intervals. Furthermore, especially on days with missing values, the 95% credible

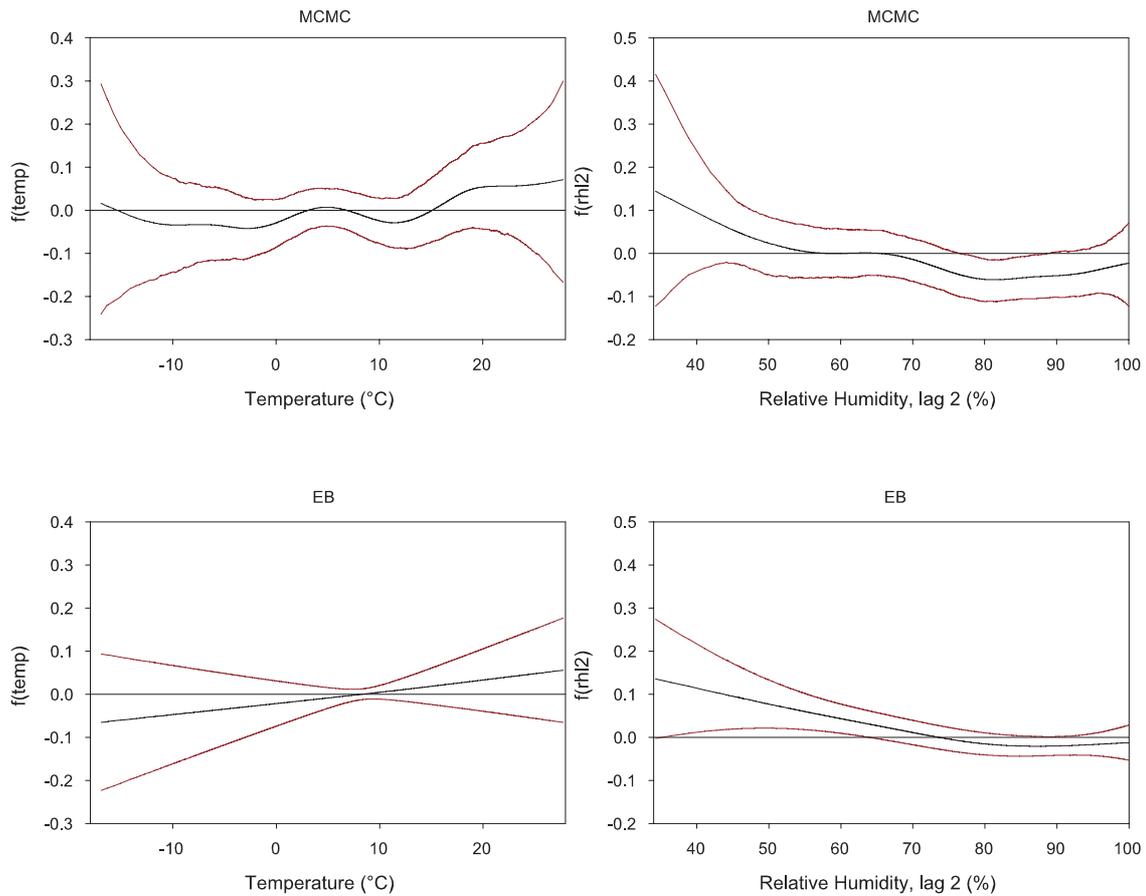


Figure 5.5: Posterior mean estimates (top) and posterior mode estimates (bottom) for the nonlinear effects of the temperature of the same day (left panel) and relative humidity with a two-days lag (right panel).

intervals are wider due to the uncertainty incorporated into corresponding estimates. Again, both estimates have very similar functional shapes, where the MCMC estimates have a greater range of curvature.

Evaluating the quantiles of the estimated time-varying effects, the posterior mode of the empirical Bayes approach has, similar to the results for CO , a larger value than the posterior mean (Table 5.3).

Figure 8.4 shows the time-varying association of $PM_{2.5}$ for lag 0. The largest time-varying coefficients were observed for the years 1998 and 1999. Afterwards, the effect estimates decreased until mid 2001 and increased again hereafter.

In contrast to CO and UFP , EB estimates for $PM_{2.5}$ showed slightly more curvature, which is also reflected in Table 5.3. Further, applying the EB method to $PM_{2.5}$ resulted in more wiggly time-varying coefficients compared to the ones obtained by MCMC.

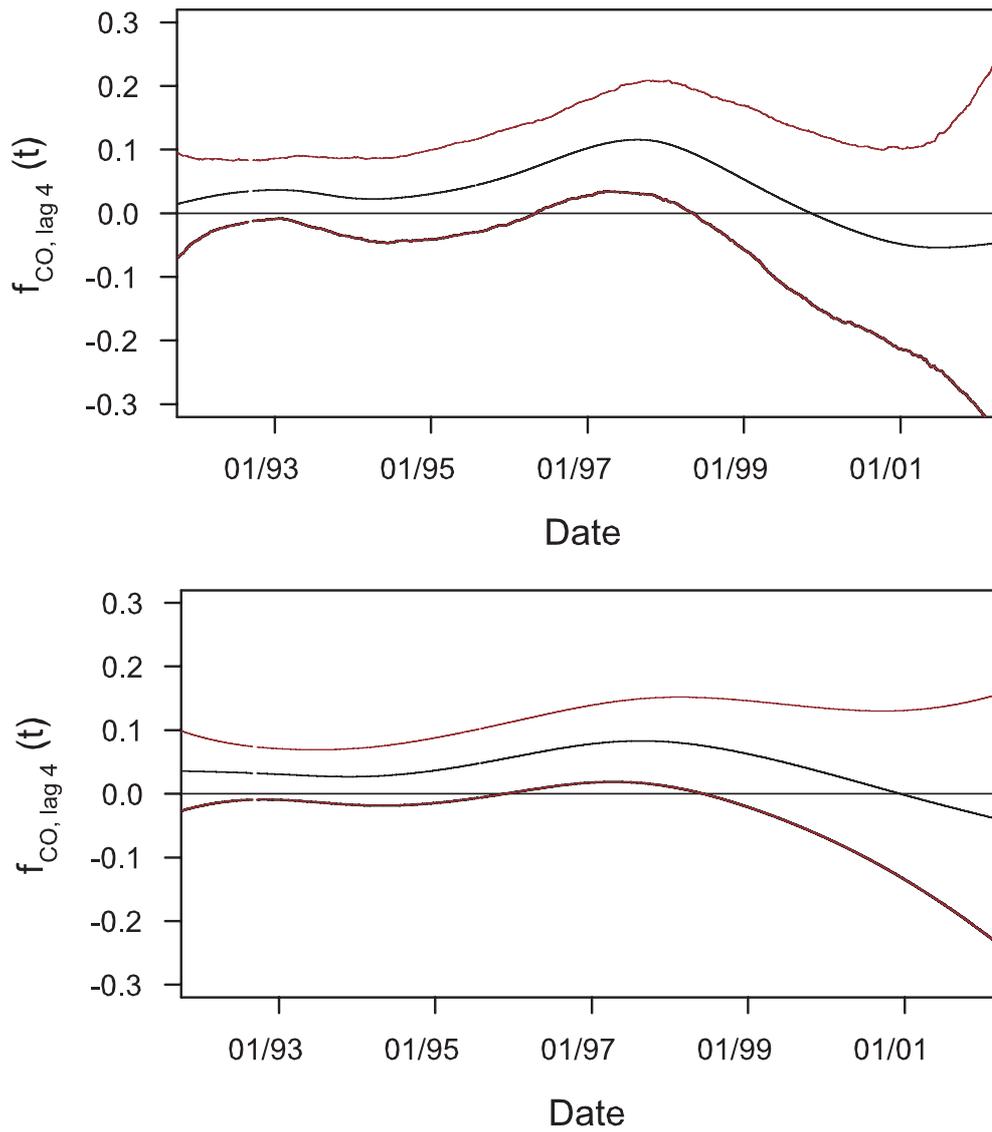


Figure 5.6: Time-varying effect of carbon monoxide (CO) with a four-day lag estimated using the MCMC (top) and the mixed model based EB method (bottom).

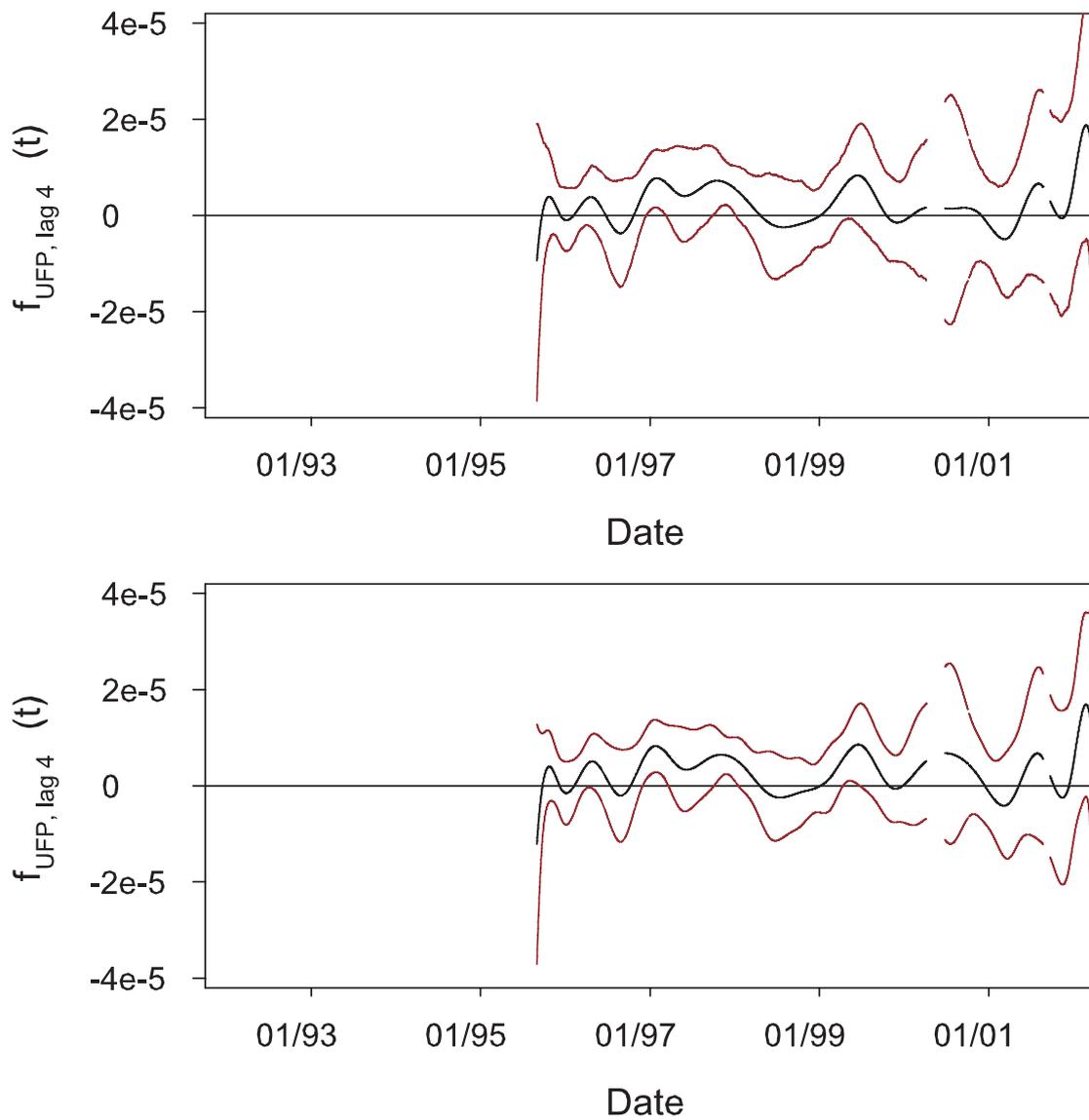


Figure 5.7: Time-varying effect of ultrafine particles (*UFP*) with a four-day lag estimated using the MCMC (top) and the mixed model based EB method (bottom).

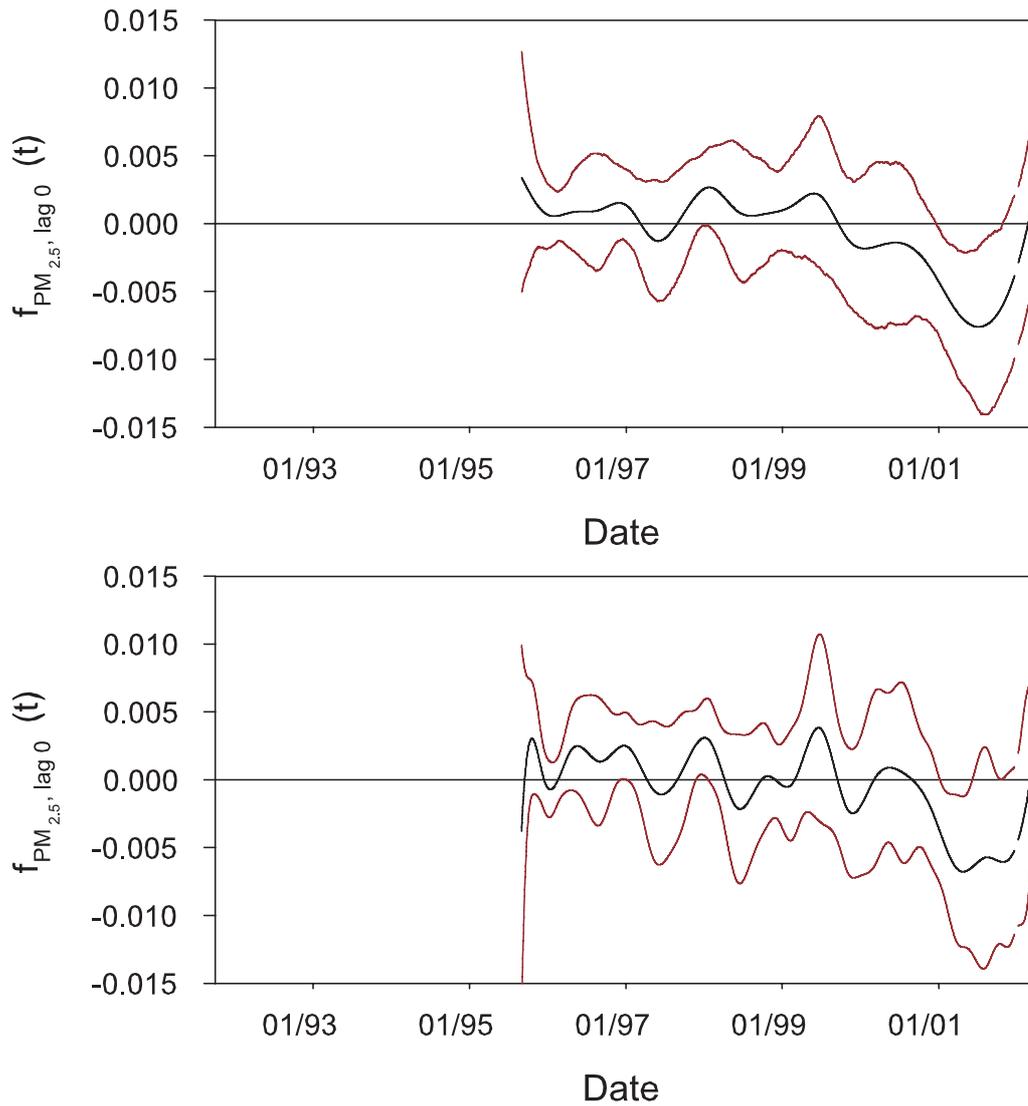


Figure 5.8: Time-varying effect of fine particle mass ($PM_{2.5}$) of the same day estimated using the MCMC (top) and the mixed model based EB method (bottom).

Table 5.3: Quantiles for the estimated time-varying coefficients of CO (lag 4), UFP number concentrations (lag 4) and $PM_{2.5}$ (lag 0), using the two methods that performed best in the simulation study.

Variable	Method	Mean/Mode	Min	25%	Median	75%	Max
CO	MCMC	0.032	-0.050	0.009	0.031	0.067	0.1133
	Empirical Bayes	0.039	-0.041	0.027	0.035	0.065	0.083
UFP	MCMC	2.39e-6	-9.38e-6	-7.93e-7	1.57e-6	5.30e-6	1.88e-5
	EB	2.75e-6	-1.21e-5	-6.01e-7	2.99e-6	5.70e-6	1.69e-5
$PM_{2.5}$	MCMC	-0.0007	-0.0076	-0.0017	0.0006	0.0012	0.0034
	EB	-0.0005	-0.0068	-0.0017	0.0000	0.0016	0.0038

We also calculated relative risks (RR) for mortality per increase in one interquartile range (IQR) of the respective pollutant to be able to compare the pollutant effects.

Relative risks of mortality per increase of CO in one IQR (IQR: $0.48 \text{ mg}/\text{m}^3$) ranged from 0.98 with a 95% credible interval of (0.89, 1.07) in 2001 to 1.06 (95% CI: 1.02, 1.10) in 1997 for the fully Bayesian approach. For the empirical Bayes approach, the RR varied from 0.98 (95% CI: 0.89, 1.08) to 1.04 (95% CI: 1.01, 1.08). Relative risks per IQR increase in the UFP number concentration (IQR: 9418) ranged between 0.92 with a 95% credible interval of (0.70, 1.20) and 1.19 (95% CI: 0.95, 1.50) for the fully Bayesian approach and between 0.90 (95% CI: 0.71, 1.13) and 1.17 (95% CI: 0.98, 1.40) for the empirical Bayes approach. Hence, the effects of the UFP number concentration showed a slightly greater range compared to CO . In contrast, for $PM_{2.5}$, relative risks per an interquartile range increase (IQR: $16.3 \text{ }\mu\text{g}/\text{m}^3$) varied from 0.88 with a 95% credible interval of 0.80 and 0.97 to 1.06 (95% CI: 0.92; 1.23) for the fully Bayesian method and from 0.90 (95% CI: 0.82; 0.98) to 1.07 (95% CI: 0.95; 1.19). Comparing all three pollutants, UFP number concentrations showed the largest effect estimates.

5.2.1 Sensitivity within MCMC estimation

The estimated nonlinear functions f_j may in some situations be very sensitive to the particular choice of hyperparameters, e.g. the parameters a_j and b_j defining the inverse gamma prior of the variances of nonparametric effects. This may be the case for very low signal to noise ratios and/or sparse data situations. Therefore, it is often recommended to re-estimate the models under consideration using a (small) number of different choices a_j and b_j to assess the dependence of results on minor changes in the model assumptions. We assessed the sensitivity of our inferences to the choice of hyperparameters by repeating the analyses with different choices for the parameters a and b for each effect in the model. The values of choice were: $a = 0.001, b = 0.001$ (standard choice);

$a = 0.00001, b = 0.00001; a = 1, b = 0.005$. As an example, Figure 5.9 shows the results for the time-varying effect of CO depending on different choices of hyperparameters. Obviously, the estimated function using $a = 0.00001, b = 0.00001$ was somewhat smoother, whereas for $a = 1, b = 0.005$, a slightly curvier time-varying coefficient was obtained; but it did not differ that much from the estimates with the standard choices.

5.2.2 Time-varying effect or dose-response?

In principle, two competing factors may explain a change in effect estimates. The first possibility is that there is a linear exposure response relationship, but because the measured pollutant serves as an indicator for changes in source composition or air pollution mixtures, associations with the outcome vary. The second possibility is that there is a non-linear exposure-response relationship which converts to changes in the effect estimates over time as concentrations in the exposure change. We therefore assessed the exposure-response functions of the three pollutants considered. We used model (5.1), but included nonlinear functions of the pollutants instead of a time-varying coefficient term

$$\log(\mu_t) = \log(pop) + confounder + f(poll).$$

Figure 5.10 shows the exposure-response-relationships of the three air pollutants and mortality. The dose-response curve of CO , lag 4 was clearly linear (Figure 5.10, a)), suggesting that there was indeed a time-varying association between CO and mortality. The dose-response curve of $PM_{2.5}$ of the same day and mortality also suggested a linear relationship (Figure 5.10, c)). This, however means, that changes in the effect estimates over time were not caused by a conversion of a non-linear exposure-response relationship. The dose-response curve of UFP , lag 4 reveals a slightly non-linear dose-response relationship of UFP and mortality, shown in Figure 5.10, b). We therefore assessed a number of sensitivity plots; a plot showing the time series of the lagged UFP concentrations and the time-varying association of UFP , lag 4, and mortality (Figure 5.11, middle row); and a scatterplot of the UFP concentrations versus the product of the time-varying coefficient times the UFP concentrations, which is shown in the bottom of Figure 5.11. Both plots suggest that the highest time-varying coefficient estimates were not obtained while concentration levels were highest.

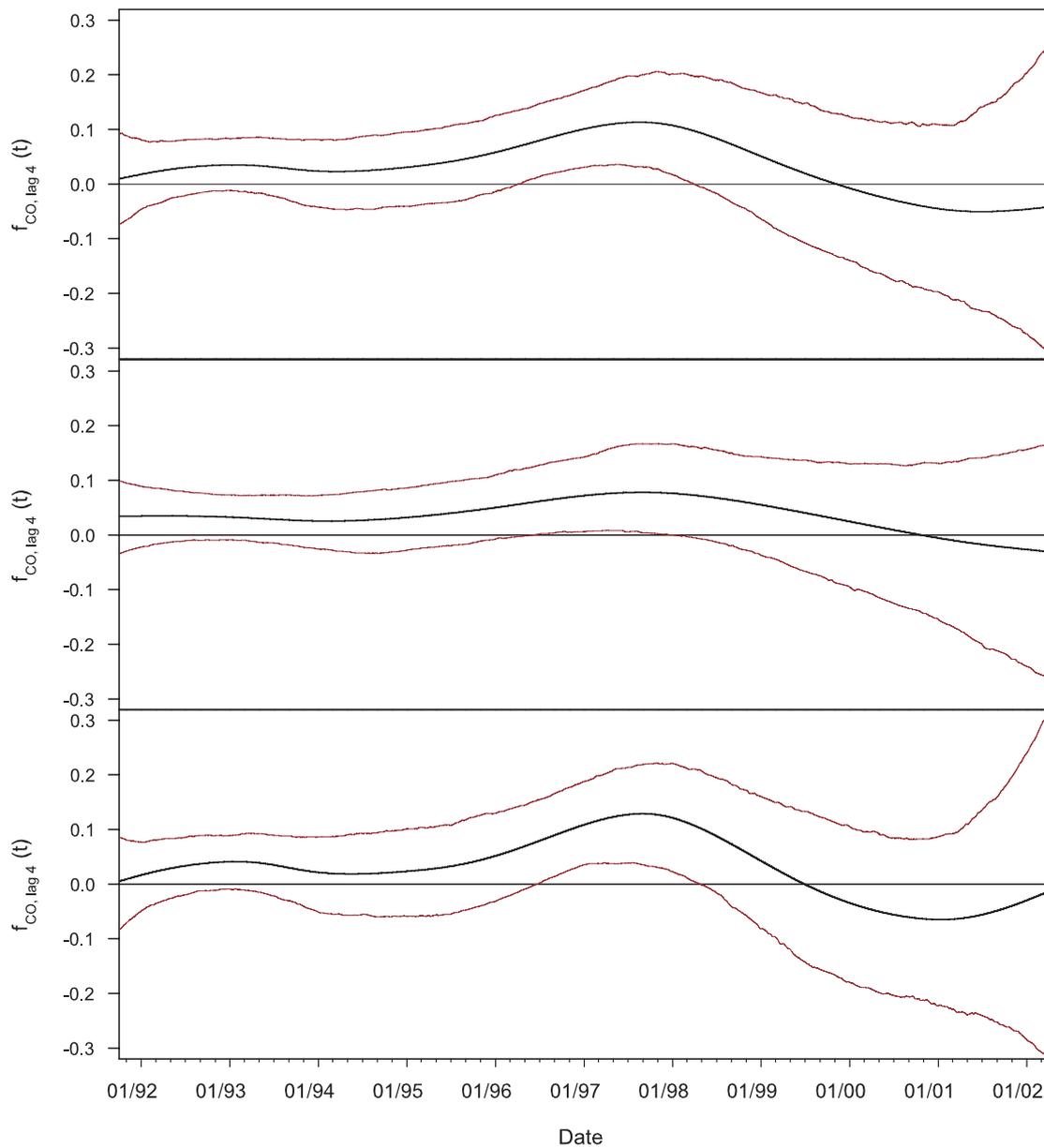


Figure 5.9: Time-varying association of carbon monoxide (CO) at lag 4 with different choices of the hyperparameters: $a = b = 0.001$ (top, standard choice); $a = 0.00001, b = 0.00001$ (middle) and $a = 1, b = 0.005$ (bottom).

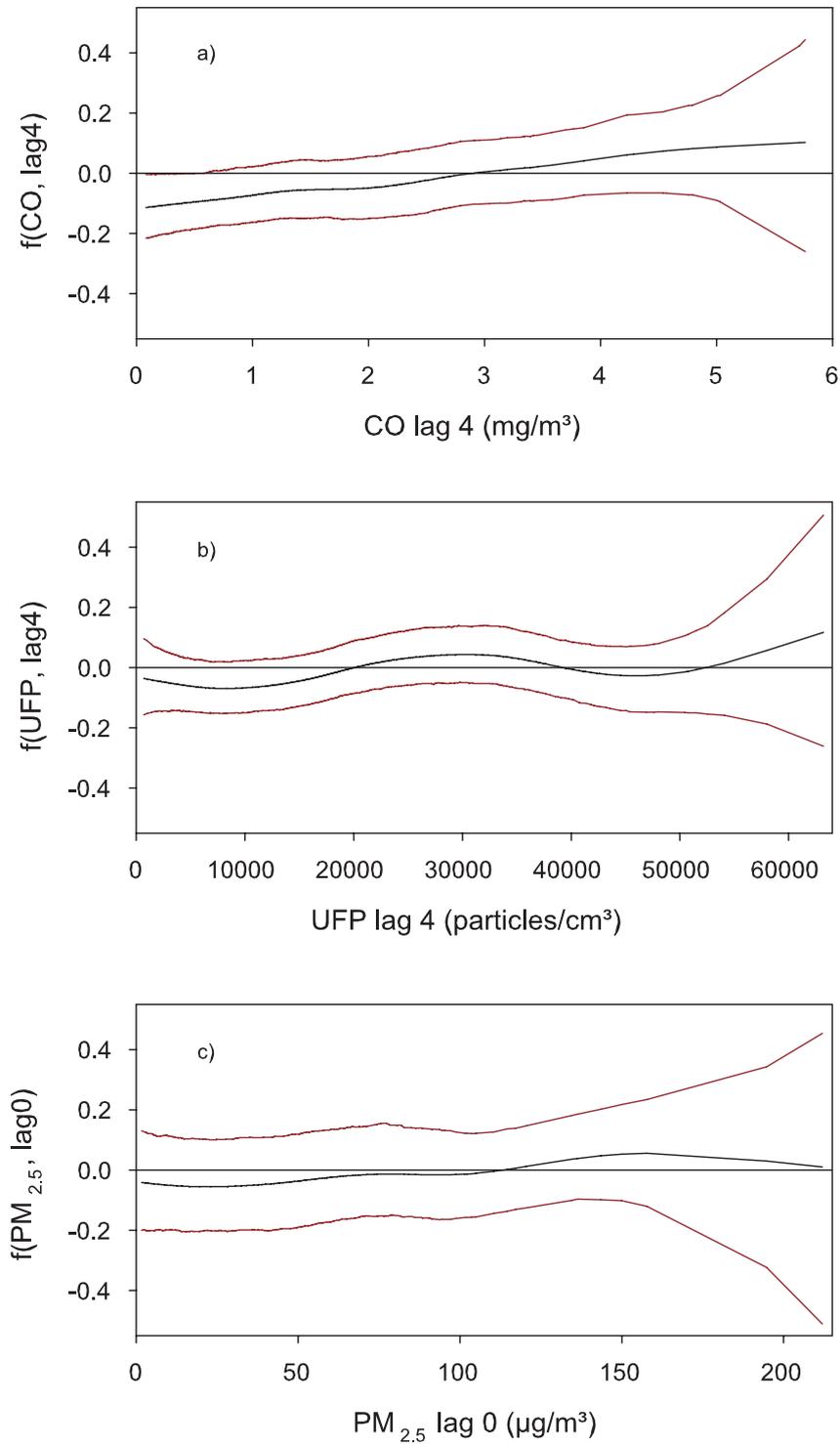


Figure 5.10: Exposure-response-relationships of CO concentrations with lag 4 (a), of UFP concentrations with lag 4 (b) and of $PM_{2.5}$ concentrations of the same day (c).

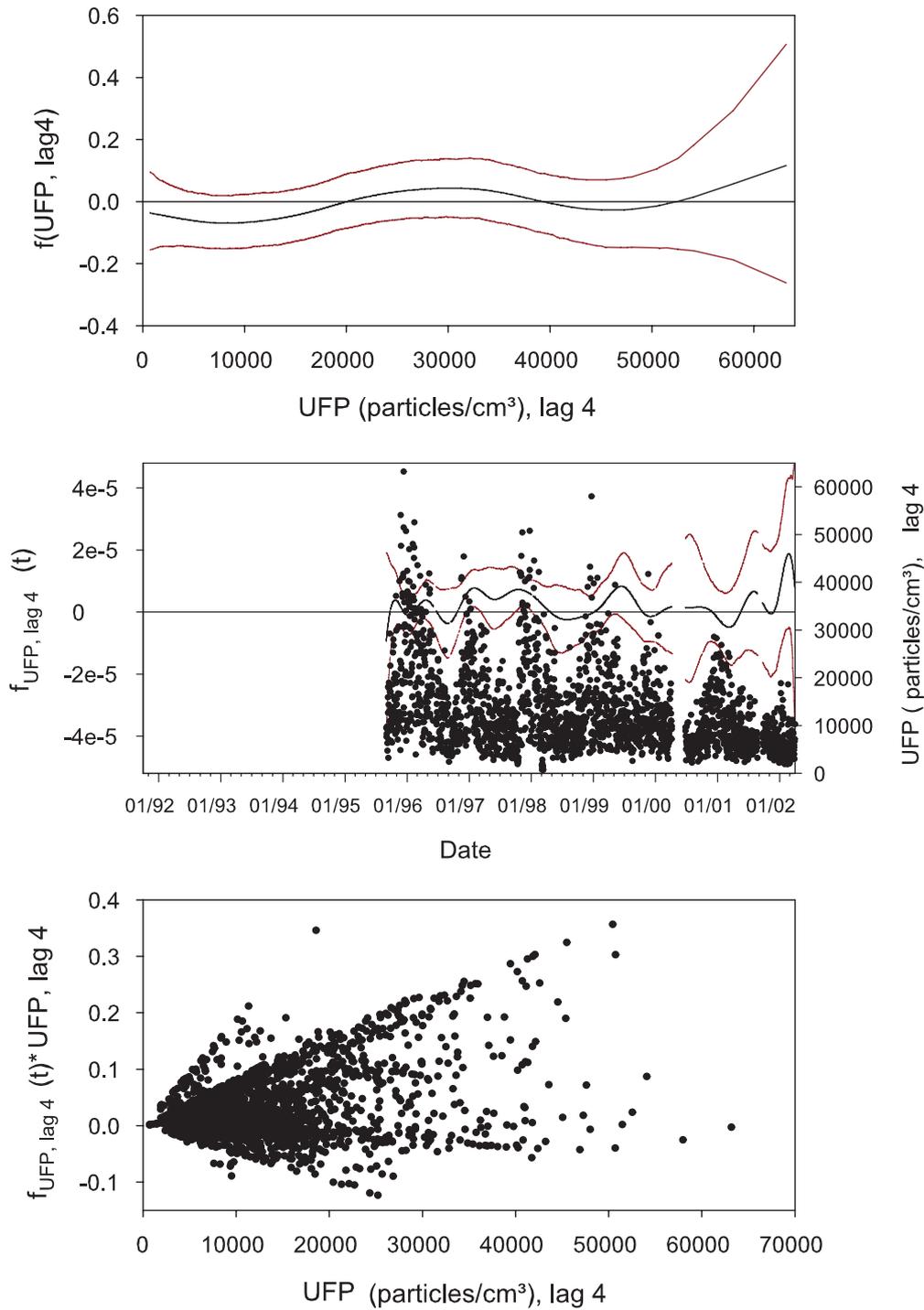


Figure 5.11: Exposure-response-relationship of UFP with lag 4 (top); Time series of UFP concentrations with lag 4 and time-varying association of UFP , lag 4 (middle); and scatterplot of UFP concentrations versus the product of the time-varying coefficient times the UFP concentrations (bottom).

5.2.3 Inference

To get a clear indication whether the time-varying coefficient approach in the applications is really necessary, a test of the fixed-parameter hypothesis against a time-varying coefficient alternative is called for. For the application considered here, this results in testing, for example, the hypotheses

$$H_0 : f_{CO_{lag4}}(time) = constant \leftrightarrow H_1 : f_{CO_{lag4}}(time) \neq constant. \quad (5.4)$$

There are only a few proposals for tests in the context of generalized varying coefficient models, mainly using local likelihood modeling and spline smoothing. Zhang (2004) proposes a scaled chi-squared test in the context of generalized longitudinal data to examine whether a varying coefficient is a polynomial of certain degree. Smoothing splines are used in this approach to estimate the smooth varying coefficient functions. Kauermann & Tutz (1999) and Kauermann & Tutz (2001) present tests of a general fixed-coefficient hypothesis against a varying-coefficient alternative, estimating the alternative model by local likelihood smoothing. Fan, Zhang & Zhang (2001) present a method focusing on the comparison of a nonparametric estimate of the whole regression function with its parametric counterpart. Based on their results, Cai et al. (2000) proposes a nonparametric maximum likelihood ratio test procedure for generalized varying coefficient models which can also be used for identifying whether each of the coefficient functions, but not the whole regression function, of a varying-coefficient model is constant. A corresponding test within the Bayesian framework has been provided by Fahrmeir & Mayer (2001), however only in the context of Kalman filtering and EM algorithms.

Similar to Cai et al. (2000) and Fahrmeir & Mayer (2001), we propose to use a likelihood ratio test that is based on the following test statistic

$$T = 2\{l(H_1) - l(H_0)\}, \quad (5.5)$$

where $l(H_1)$ and $l(H_0)$ are the corresponding log-likelihood functions under the null and the alternative hypothesis, respectively.

In contrast to parametric models, where the likelihood ratio statistic follows asymptotically a χ^2 -distribution, the analytic form of the null distribution of the above proposed test is hard to find. Therefore, we apply a bootstrapping procedure which makes it possible to mimic the distribution of a test statistic under the hypothetical model. The testing procedure can be summarized as follows:

1. Fit the observed data with the model under H_0 , that means for example

$$\hat{\eta}_t = \sum_{j=0}^l \beta_j x_j + \sum_{j=l+1}^{p-1} f_j(x_j) + \beta_p CO_{lag4},$$

using empirical Bayes (EB) inference methods.

2. Generate a bootstrap sample Y_t^* from the above model under H_0 .
3. Estimate the fixed as well as the time-varying coefficient model for this sample, again using empirical Bayes inference methods. In each bootstrap replicate, choose the smoothing parameter using a REML estimate to incorporate the variability of estimating the smoothing parameter in the bootstrap as well.
4. Calculate the corresponding test statistic (5.5) for each of the bootstrap samples.
5. Calculate the empirical p -value.

Given the covariates, Y_t^* is generated 500 times for CO and 1000 times for UFP and $PM_{2.5}$. Further the test statistic T^* in (5.5) is computed. The distribution of T^* can then be used as an approximation to the distribution of T . For power considerations of the proposed test, we refer to the empirical work conducted by Cai et al. (2000) and Fahrmeir & Mayer (2001). They show that this testing procedure may indeed be powerful.

An application of this testing procedure to the model estimated with empirical Bayes inference revealed that the test statistic based on 500 bootstrap samples is $T = 15.64$ and the p -value is 0.03, which suggests that the model with a time-varying term for CO is a better fit. For UFP and $PM_{2.5}$, the testing procedure revealed a p -value of 0.11 and 0.09, respectively. Hence, in both cases, the assumption of a constant effect of air pollution cannot be rejected.

5.3 Discussion of the (time-varying) air pollution effects

The time-varying models revealed variation in the effect estimates over time, which was reflected in seasonal variation as well as overall variation of the effect estimates over time. A test procedure further showed that the long-term trend shape in the time-varying coefficient of CO was even significant. CO might therefore be a marker of slowly-varying sources. Although not significant, the results of UFP and $PM_{2.5}$ showed some seasonal variation. Studies in a number of locations have shown a change in the characteristics of the particulate matter mixture throughout the year and that the relative and absolute contributions of particular components to particulate matter mass may be different during different times of the year (see for example Bell et al. (in press)).

The changing effects over time could have been induced by a non-linear dose-relationship.

However, an examination of the size of effect with relation to the magnitude of the pollution level showed this not to be the case.

A striking observation was that the effect estimates were largest for CO during the period 1995 to 1997. Also, $PM_{2.5}$ and UFP showed positive effect estimates during these years (see also Wichmann et al. (2000)). In this period, changes in source characteristics took place. The two main changes were: Energy production (by power plants and by local heating) and the number and type of vehicles. To be able to give more insight on these changes, additional data about the changes in the sources were collected.

For example, the proportion of stove heating in Erfurt decreased in the time period 1992 - 2001 from 95% to 5%, whereas during the same time period, the proportion of homes heated by central heating increased from 12% to 51%. Also the mobile sources have undergone modernization after the German unification. Since 1990 the old car fleet was gradually replaced by vehicles with comparatively modern engine technology including three-way catalysts. The number of cars in Erfurt equipped with a three-way catalyst increased from 50% to 96% during the study period. Although air pollution concentrations were already reduced (see Figure 5.2) during the years 1995 to 1997, it seems that the benefits of ambient air quality were not yet completely achieved. One might, therefore, speculate that the oxidative stress potential was markedly increased during this period leading to more toxic air pollution mixtures.

Evidence on smooth time-varying associations of air pollution with mortality is very limited and no direct evaluation of time-varying effects of CO , $PM_{2.5}$ and UFP is possible. The only evidence of time-varying associations with mortality has, so far, been given for PM_{10} data.

Chiogna & Gaetan (2002) investigate a potential time-varying effect of PM_{10} on mortality over a four-year period. They detect, as in other studies, "*a positive effect of particular matter on mortality, but find that the significance of this effect might change in time*". In an extended version of the Dynamic Generalized linear model approach, Lee & Shaddick (2005) analyze PM_{10} data from the Greater London area. In their study, the effects of PM_{10} show little change over the first two years of the study, and then a marked increase in 1997. However, the possibility of a constant effect could not be excluded, as the credible intervals were wide. Finally, Peng et al. (2005) find a seasonally-varying effect of PM_{10} throughout the U.S. with a higher effect on mortality during summer. However, it is important to note that in the U.S. maximum particle mass concentrations are observed in summer due to regionally transported secondary sulphates. In contrast, European ambient particle concentrations peak during winter time and this is caused by increased source emissions and a decrease of air mixture. Therefore, the apparent differences between the results of Peng et al. (2005) and our study might be due to different

seasonal variations in particle concentrations rather than a fundamental difference.

Chapter 6

A measurement error model for time-varying coefficient models

So far, we have assumed that all of the covariates are obtained without any error. However, a more realistic view is to allow, for example, for an error-prone air pollutant variable, where sources of error include, among others, imperfect monitoring devices. In this chapter, we first specify a measurement error model for time-varying coefficient models as they were considered in Chapter 5. Having described this measurement error model, we then turn to presenting methods to estimate parameters and measurement error variances of the model in the case of an autocorrelated latent exposure variable. These methods are based on a representation of the measured surrogate variables as autoregressive moving average processes.

6.1 Measurement error model formulation

Measurement error models generally have a common structure, consisting of three sub-models (Clayton 1992). The three parts are:

1. A model for the true, but latent exposure variable ξ . This is often also called *exposure model*;
2. An *outcome* or *disease model* relating the response y to the true latent covariate ξ ; and
3. A *measurement model* linking the unobservable true exposure ξ to an observable surrogate variable X .

6.1.1 Model for the latent exposure variable

In general, there are two approaches to a measurement error analysis, depending on how ξ is treated. In the so-called *functional variant* of an *error-in-variables model*, ξ is assumed to be an unknown parameter pertaining to the observation X . The second possibility is the so-called structural variant. In this model, ξ is considered as a random variable. Further, the structural variant of a regression model with measurement error is characterized by the assumption of an underlying (known) distribution of the latent covariate.

We use the structural variant in this work, as it can be assumed that the true air pollution values are random variables. Therefore, we have to specify a distribution of the latent exposure variable.

Past experience suggests that daily air pollution measurements are temporally correlated. Therefore, a potential distribution of ξ that will be investigated thoroughly is given by a multivariate normal distribution with an autoregressive covariance structure of order p (AR(p)):

$$\xi \sim N_T(\mu, \Omega), \quad (6.1)$$

where μ is a column vector of length T and Ω is a $T \times T$ covariance matrix which is symmetric and positive definite and has the following form

$$\Omega_{tt'} = \text{Cov}(\xi_t, \xi_{t'}) = \sigma_\xi^2 \rho_\xi^{|t-t'|}. \quad (6.2)$$

This formulation assumes the stationarity of the mean and variance over time, but in general it is more reasonable to assume that the mean of ξ follows a linear trend which can be expressed as

$$E(\xi) = \kappa_0 + \kappa_1 t, \quad (6.3)$$

where t denotes the calendar time, κ_0 is an intercept term and κ_1 is the slope of the long-term time trend. An extension of model (6.3) could also include, for example, a seasonal indicator or a harmonic component such as

$$E(\xi) = \beta_0 + \beta_1 t + \beta_2 \cos(2\pi t/365) + \beta_3 \sin(2\pi t/365) \quad (6.4)$$

for the seasonality effect.

6.1.2 Disease model

As in Chapter 5, we use a Poisson model as *outcome* or *disease model*, with daily mortality as response, that means we assume

$$y|\xi, z \sim Po(\mu) \quad (6.5)$$

and the predictor to have the following form

$$\log(\mu) = \eta = \beta_0 + \sum_{j=1}^l \beta z_j + \sum_{j=l+1}^{p-1} f_j(z_j) + f_p(t) \xi,$$

where z_j , $j = 1, \dots, p-1$ denote additional covariates. Instead of an observed surrogate measure (as in (5.1)), now the true exposure ξ is included in the predictor. We further assume that both response and additional covariates are measured or obtained without error and that there are no dependencies between the covariates z_j and the true exposure variable ξ - strong restrictions indeed. While a more flexible model is desirable, we use these restrictions to ease the illustration and simplify the estimation procedures.

6.1.3 Measurement model

The measurement model relates the observed error-prone surrogate X to the true, but unobservable variable ξ . Broadly, there are two different types of error structures (see, for example, Carroll et al. (2006) for more details).

The *classical* measurement model assumes that the observed variable X is the sum of the true variable ξ and an additive random error with mean zero, which is independent of ξ . A classical model is, for example, reasonable for the difference between the measured levels of an environmental exposure and the true level of that agent.

An alternative is the so-called Berkson model, where the true variable is assumed to be the result of the observed variable plus some independent random deviation. In this model the error is independent of X . This error type is reasonable, for example, when the individual exposure to a certain air pollutant is measured by a monitoring device, which is on a fixed location. The measured level is then assigned to all persons living in a certain radius of that monitor.

In this thesis, we consider the situation in which two environmental monitoring systems are observed which are abbreviated X_{t1} and X_{t2} , hereafter. The two devices collect data on the same variable ξ at regular time points t , $t = 1, \dots, T$. We further treat the measurements of the two monitoring systems as replicate measurements of ξ . Hence, the

assumption of a classical error model is appropriate, since we want to model the true ambient level of an air pollutant in a city or region, but the monitoring devices do not allow for a correct measurement.

Specifically, the relationship between the true and perturbed observed series is given by the following models

$$\begin{aligned} X_{t1} &= \xi_t + \alpha_1 + \epsilon_{t1} \\ X_{t2} &= \xi_t + \alpha_2 + \epsilon_{t2}, \quad i = 1, 2, \end{aligned} \tag{6.6}$$

where α_i are measurement biases associated with the two monitoring devices and

$$\begin{aligned} \epsilon_{ti} &\sim N(0, \sigma_{\epsilon_i}^2) \\ \text{Cov}(\xi_t, \epsilon_{t'i}) &= 0 \quad \forall t, t' \\ \text{Cov}(\epsilon_{t1}, \epsilon_{t'2}) &= 0 \quad \forall t, t' \end{aligned}$$

A further essential assumption is that, conditional on the latent covariate ξ , the surrogates X_i do not contain additional information on the response y . This is known as non-differential measurement error and can be expressed as

$$y_t \perp X_{ti} | \xi_t \quad \text{for all } t.$$

The measurement error may also be correlated over time. The dependence structure of measurement errors can be caused by common disturbing factors. Conceivable is an influence of season or other exogenous variables as temperature on the measurement error. A potential model then could be

$$\begin{aligned} \epsilon_{ti} &\sim N(\mu_t, \sigma_{\epsilon_i}^2) \\ \text{Cov}(\epsilon_{ti}, \epsilon_{t'i}) &= \sigma_{\epsilon_i}^2 \rho_i^{|t-t'|}. \end{aligned}$$

Another possibility could be to take logarithmic values into consideration. Therewith, one allows a multiplicative structure of the model; however, in this thesis, we only regard the case of an additive measurement error structure.

6.2 Identifiability of parameters

Before we can estimate parameters of models such as (6.6), there are some assumptions necessary for the identifiability of the parameters. This is due to the fact that ξ is latent.

One possible identifiability condition is given by

$$\alpha_1 + \alpha_2 = 0. \quad (6.7)$$

This means that none of both measurement series is favored concerning the systematic bias. Correspondingly, such considerations may be required for potential other effects in the measurement error model.

A further possibility to ensure identifiability is to set one of the α_i , $i = 1, 2$ to zero, which implies

$$\alpha_1 = 0 \quad \text{or} \quad \alpha_2 = 0. \quad (6.8)$$

Note, however, that adding a constant to one exposure variable only affects the intercept and, therefore, does not affect the regression results in any interesting way.

6.3 Estimating the measurement error model parameters

To adjust for the bias in the estimated coefficients in a measurement error analysis, an estimate of the measurement error variance is needed. In the following, we describe several techniques for obtaining these estimated variances. We first present a simple estimation procedure without accounting for autocorrelation. Then, we consider an approach for obtaining the autoregressive parameter and corrected variance estimates for the case of autocorrelated true values, but temporally independent measurement errors. Finally, we present a method when both, ξ and ϵ_i , follow autoregressive processes. Both methods are based on a representation of the surrogate variables as autoregressive moving average processes. We conclude with some remarks about the identifiability and the estimation of the intercept and linear trend coefficients of the exposure model.

6.3.1 Simple Estimation

A simple estimate of the parameters in models such as mentioned above can be obtained by applying assumption (6.7) or (6.8) and using the mean, x_m , of the two observed variables, with

$$x_m := \frac{x_1 + x_2}{2}, \quad (6.9)$$

where

$$\begin{aligned} E(x_m) &= \mu \\ \text{Var}(x_m) &= \sigma_\xi^2 + \frac{1}{4}(\sigma_{\epsilon_1}^2 + \sigma_{\epsilon_2}^2) \\ \text{Var}(x_1) &= \sigma_\xi^2 + \sigma_{\epsilon_1}^2 \\ \text{Var}(x_2) &= \sigma_\xi^2 + \sigma_{\epsilon_2}^2. \end{aligned}$$

Using (6.9) and the corresponding assumptions, the parameters can be estimated from the data using the appropriate moments.

6.3.2 Models allowing for an autoregressive structure of the observations

As we assume that the true concentrations are temporally correlated, the simple estimation procedure might not lead to 'correct' estimated parameters. To account for the dependence between successive observations, we assume that the autocorrelation structure of the ξ_t follows an autoregressive process of order p . This means that observations close in time are more correlated than those far away. Furthermore, the measurement errors might also be temporally correlated.

6.3.2.1 AR(p) + white noise case

In this section, we consider a model in which the observed surrogate variables are assumed to be the sum of an autoregressive series ξ_t plus a white noise error e_t . Such a model is also known as 'signal plus noise' model in engineering (see for example Pagano (1974) or Granger & Morris (1976)).

In general, an autoregressive model for ξ is specified as:

$$\xi_t = \mu_t + \rho_{\xi,1} \xi_{t-1} + \dots + \rho_{\xi,p} \xi_{t-p} + e_t, \quad (6.10)$$

where the $\rho_{\xi,1}, \dots, \rho_{\xi,p}$ denote the autoregressive parameters and μ_t is the mean of ξ_t . The e_t are iid distributed with mean 0 and variance σ_e^2 .

Instead of ξ_t which is not observable, we now take the observed measurement values $X_{ti} = \xi_t + \epsilon_{ti}$, $i = 1, 2$ with $\epsilon_{ti} \sim N(0, \sigma_{\epsilon_i}^2)$ and

$$\text{Cov}(\epsilon_{ti}, \epsilon_{t'i}) = 0 \quad \forall t \neq t'.$$

Following Pagano (1974) and Granger & Morris (1976), it is known that if ξ_t is an autoregressive process of order p with parameters $\rho_{\xi,1}, \dots, \rho_{\xi,p}$, and ϵ_{ti} is white noise

with (constant) variance $\sigma_{\epsilon_i}^2$, then the sum $X_{ti} = \xi_t + \epsilon_{ti}$, follows an autoregressive moving average (ARMA) process of order (p, p) . The corresponding proof is given in Appendix B.

In general, an stationary ARMA(p,p) process for X_{ti} is given by

$$\rho(B) X_{ti} = \theta(B) \gamma_{ti}.$$

The model consists of two parts, an autoregressive (AR) part $\rho(B) X_{ti}$, and a moving average (MA) part $\theta(B)\gamma_{ti}$. $\rho(B)$ and $\theta(B)$ are defined as p -th order lag polynomials, where $\rho(B) = 1 - \rho_1 B - \dots - \rho_p B^p$, $\theta(B) = 1 + \theta_1 B + \dots + \theta_p B^p$ and B is the so-called Backward-Shift operator, such that $B^j X_t = X_{t-j}$. γ_{ti} is a white noise series with mean 0 and variance $\sigma_{\gamma_i}^2$.

Applying the definition of X_{ti} as sum of ξ_t and ϵ_{ti} results in an ARMA(p,p) process of the form

$$\rho(B) X_{ti} = \rho(B) (\xi_t + \epsilon_{ti}) = e_t + \rho(B) \epsilon_{ti}.$$

Having specified the representation of X_{ti} as ARMA process, we now show how to obtain the estimates for the parameters of interest. As an example, a zero-mean AR(1) process for ξ_t is chosen.

In the case of a zero-mean AR(1) process for ξ_t , we are seeking a model

$$X_{ti} = \xi_t + \epsilon_{ti} \quad \xi_t = \rho_{\xi,1} \xi_{t-1} + e_t, \quad \text{and} \quad \epsilon_{ti} \sim WN(0, \sigma_{\epsilon_i}^2). \quad (6.11)$$

Using the definition of the lag polynomials, the ARMA (1,1) process for X_{ti} is obtained as

$$\begin{aligned} (1 - \rho_{\xi,1} B) X_{ti} &= (1 - \rho_{\xi,1} B) \xi_t + (1 - \rho_{\xi,1} B) \epsilon_{ti} \\ &= e_t + (1 - \rho_{\xi,1} B) \epsilon_{ti}, \end{aligned} \quad (6.12)$$

and applying the definition of the Backward-Shift operator then yields

$$X_{ti} - \rho_{\xi,1} X_{t-1i} = e_t + \epsilon_{ti} - \rho_{\xi,1} \epsilon_{t-1i}.$$

Suppose now that we estimate parameters for a given ARMA(1,1) process

$$(1 - \rho_{ARMA_i} B) X_{ti} = (1 + \theta_{1i} B) \gamma_{ti}, \quad (6.13)$$

where γ_{ti} is defined as above and is uncorrelated with e_t .

For the ARMA(1,1) processes (6.12) and (6.13) to be equivalent, the autoregressive parameter ρ_{ARMA_i} in (6.13) has to be the same as in model (6.12); however, this also means that the AR parameter is the same as $\rho_{\xi,1}$ in (6.11).

Further, the moving averages θ_{1i} as well as the variances $\sigma_{\gamma_i}^2$ obtained by (6.13) can be related to $\rho_{\xi,1}$, σ_e^2 and $\sigma_{\epsilon_i}^2$ by equating the auto-covariance functions of X_i and $\xi + \epsilon_i$

$$\frac{\sigma_{\gamma_i}^2 (1 + \theta_{1i}z)(1 + \theta_{1i}z^{-1})}{(1 - \rho_{\xi,1}z)(1 - \rho_{\xi,1}z^{-1})} = \frac{\sigma_e^2}{(1 - \rho_{\xi,1}z)(1 - \rho_{\xi,1}z^{-1})} + \sigma_{\epsilon_i}^2, \quad (6.14)$$

where z is a complex scalar. After some transformations, this results in

$$\sigma_{\gamma_i}^2 (1 + \theta_{1i}^2) + \sigma_{\gamma_i}^2 \theta_{1i}(z + z^{-1}) = \sigma_e^2 + \sigma_{\epsilon_i}^2 (1 + \rho_{\xi,1}^2) - \sigma_{\epsilon_i}^2 \rho_{\xi,1} (z + z^{-1}). \quad (6.15)$$

For an equivalence of the processes (6.12) and (6.13), one needs

$$\sigma_{\gamma_i}^2 (1 + \theta_{1i}^2) = \sigma_e^2 + \sigma_{\epsilon_i}^2 (1 + \rho_{\xi,1}^2), \quad (6.16)$$

that means equal variances. Using this equivalence in (6.15) finally leads to

$$\sigma_{\gamma_i}^2 \theta_{1i} = -\sigma_{\epsilon_i}^2 \rho_{\xi,1}. \quad (6.17)$$

Hence, in summary, for the two ARMA models (6.12) and (6.13) to be equivalent, one needs the conditions

$$\begin{aligned} \rho_{ARMA_i} &= \rho_{\xi,1} \\ \sigma_{\gamma_i}^2 (1 + \theta_{1i}^2) &= \sigma_e^2 + \sigma_{\epsilon_i}^2 (1 + \rho_{\xi,1}^2) \\ \sigma_{\gamma_i}^2 \theta_{1i} &= -\sigma_{\epsilon_i}^2 \rho_{\xi,1}. \end{aligned}$$

The last equation can be solved for $\sigma_{\gamma_i}^2$, yielding

$$\sigma_{\gamma_i}^2 = -\frac{\sigma_{\epsilon_i}^2 \rho_{\xi,1}}{\theta_{1i}}.$$

This can then be substituted into (6.16) to deduce

$$-\frac{\sigma_{\epsilon_i}^2 \rho_{\xi,1}}{\theta_{1i}} (1 + \theta_{1i}^2) = \sigma_e^2 + \sigma_{\epsilon_i}^2 (1 + \rho_{\xi,1}^2).$$

Under the condition $\sigma_e^2 > 0$, we get after some transformations realizability conditions in the form

$$\left| \frac{1}{1 + \rho_{\xi,1}^2} \right| > \left| \frac{\theta_{1i}}{(1 + \theta_{1i})\rho_{\xi,1}} \right| \geq 0.$$

Note that the processes that were added together in this special case were assumed to have zero mean. However, Hamilton (1994) states that "*adding constant terms to the processes will not change the results in any interesting way*".

As a result of this, an approach to estimating $\rho_{\xi,1}$ is to fit an ARMA(1,1) model to the observed time series of measurements X_i , $i = 1, 2$ and to use the resultant estimates of the autoregressive parameters $\hat{\rho}_{ARMA}$. Note that, as only one autoregressive parameter $\rho_{\xi,1}$ in the measurement error correction methods is needed, we always use the mean of the two estimated autoregressive parameters obtained for each observed series.

Having obtained estimates of $\hat{\rho}_{ARMA_i}$, $\hat{\theta}_{1i}$ and $\hat{\sigma}_{\gamma_i}^2$, one can then solve equation (6.17) to get an estimate of $\sigma_{\epsilon_i}^2$. Substitution of $\hat{\sigma}_{\epsilon_i}^2$ into equation (6.16) finally leads to an estimate of σ_e^2 .

6.3.2.2 AR(p) + AR(q) case

Suppose now, that not only the time series ξ_t of true values, but also the measurement errors are autocorrelated. We further assume that the autocorrelation structure of the ϵ_{ti} follows an autoregressive process of order q (AR(q)). This dependence structure of the measurement errors may be caused by common disturbing factors such as meteorological influences.

Again, the time series of observed measurement values $X_{ti} = \xi_t + \epsilon_{ti}$, $i = 1, 2$ is taken, which, however, is now the sum of two autoregressive processes of order p and q ,

$$\phi(B) \xi_t = e_t, \quad (6.18)$$

and

$$\pi(B) \epsilon_{ti} = u_{ti}, \quad (6.19)$$

where $e_t \sim N(0, \sigma_e^2)$ and $u_{ti} \sim N(0, \sigma_{u_i}^2)$.

In general, it is likely that the autoregressive parameters of the two processes are different. In this case, adding an AR(p) process (6.18) to an AR(q) process (6.19) under the assumption that the two processes are uncorrelated produces an ARMA($p+q, \max\{p, q\}$) process of the following form

$$\phi(B) \pi(B) X_{ti} = \pi(B) e_t + \phi(B) u_{ti}.$$

Corresponding to the AR(p) + white noise case, we again show exemplarily for AR(1) processes how to obtain the parameters of interest.

For the special case of AR(1) processes for ξ_t

$$\xi_t = \phi_1 \xi_{t-1} + e_t,$$

as well as for ϵ_{ti}

$$\epsilon_{ti} = \pi_1 \epsilon_{t-1} + u_{ti},$$

this implies that the sum of these two processes is an ARMA(2,1) model, given by

$$\begin{aligned} (1 - \phi_1 B)(1 - \pi_1 B)(X_{ti}) &= (1 - \phi_1 B)(1 - \pi_1 B)(\xi_t + \epsilon_{ti}) \\ &= (1 - (\phi_1 + \pi_1)B + \phi_1 \pi_1 B^2)(\xi_t + \epsilon_{ti}) \\ &= (1 - \pi_1 B)e_t + (1 - \phi_1 B)u_{ti}. \end{aligned} \quad (6.20)$$

Now assume that a given ARMA(2,1) process is

$$(1 - \rho_{1i}B - \rho_{2i}B^2)X_{ti} = (1 + \theta_{1i}B)\gamma_{ti}, \quad (6.21)$$

where γ_{ti} is a white noise process uncorrelated with e_t and u_{ti} . Note that an ARMA(2,1) model (6.21) is stationary and the roots of $1 - \rho_1 z - \rho_2 z^2 = 0$ are real, if the ρ_i satisfy the following conditions (see for example Brockwell & Davis (1998)):

$$\rho_2 + \rho_1 < 1, \quad \rho_2 - \rho_1 < 1, \quad \text{and} \quad |\rho_2| < 1.$$

For the two ARMA models (6.20) and (6.21) to be equivalent, the autoregressive parameters of the two models have to be related in the following way:

$$\rho_{1i} = \phi_1 + \pi_1 \quad \text{and} \quad \rho_{2i} = -\phi_1 \pi_1,$$

which implies

$$\begin{aligned} \phi_1^2 - \rho_{1i} \phi_1 - \rho_{2i} &= 0 \quad \text{and} \\ \pi_1^2 - \rho_{1i} \pi_1 - \rho_{2i} &= 0. \end{aligned}$$

Comparable to the *AR(1) + white noise* case, one further needs conditions based on all non-zero autocovariances

$$\sigma_{\gamma_i}^2 (1 + \theta_{1i}^2) = \sigma_e^2 (1 + \pi_1^2) + \sigma_{\epsilon_i}^2 (1 + \phi_1^2) \quad (6.22)$$

$$\sigma_{\gamma_i}^2 \theta_{1i} = -\pi_1 \sigma_e^2 - \phi_1 \sigma_{\epsilon_i}^2. \quad (6.23)$$

Using the estimates $\hat{\rho}_{1i}$, $\hat{\rho}_{2i}$, θ_{1i} and $\hat{\sigma}_{\gamma_i}^2$, one can solve equation (6.23) for σ_e^2 , for example, yielding

$$\sigma_e^2 = \frac{-\sigma_{\gamma_i}^2 \theta_{1i} - \phi_1 \sigma_{\epsilon_i}^2}{\pi_1}$$

Substitution of the corresponding equations into equation (6.22) finally leads to estimates $\hat{\sigma}_e^2$ and $\hat{\sigma}_{\epsilon_i}^2$.

6.3.2.3 Remarks

We conclude this Chapter with some remarks:

1. Uniqueness of the estimated parameters of an ARMA(2,1) model when the ARMA(2,1) process arises as a sum process.

Following Ku & Seneta (1998), the coefficients have a unique solution, that means they are identifiable, if the following condition holds

$$-(\theta_{1i} - \rho_{1i}\theta_{1i} - \rho_{2i}) \neq 0.$$

The case that this condition is equal to 0 effectively results in the reduction to an AR(1) process. See Ku & Seneta (1998) for a detailed discussion of this point.

2. Estimation of ARMA models.

There are several possibilities for the parameter estimation of an ARMA model. Examples are estimation using Yule-Walker equations, ordinary least-squares or maximum likelihood. For an exact maximum likelihood estimation of an ARMA model, it is convenient to use a state-space representation of the ARMA process (see for example Hamilton (1994) or Jones (1980)). This enables the calculation of the exact likelihood function by means of the Kalman filter. To get the maximum likelihood estimate of the parameters, one needs to maximize the obtained likelihood function. This can be done by optimization algorithms such as the Newton-Raphson procedure.

3. Estimation of the trend parameters κ_0 and κ_1

To obtain estimates for the exposure model parameters κ_0 and κ_1 , a linear regression is fitted with an ARMA model for the error term. Parameter estimates can be obtained by using maximum likelihood estimation as shown in Hamilton (1994).

4. The existence of measurement errors not only affects the estimation of model parameters, but also the lag order selection. Measurement errors may have two opposite effects on the lag order selection. Recently, Chong et al. (2006) show that for any given sample size, the estimated lag order obtained by the Akaike as well as the Bayes information criteria tends to be positively associated with the variance of measurement error if the variance is small, whereas they become negatively related when the variance is large. Therefore, order selection should be done with due care.

Chapter 7

Measurement error correction methods

In this chapter, we present two different approaches for analyzing temporally correlated time series data with measurement error. The first one is a Bayesian analysis including a measurement model with common time trend and measurement errors with system-specific biases. To estimate the required posterior distributions, Markov chain Monte Carlo (MCMC) simulation methods (see also Section 3.2) are used. The second approach is a variant of the well-known regression calibration approach.

7.1 A Bayesian measurement error correction model

Bayesian analysis of measurement error problems has been developed from the work of Clayton (1992), Richardson & Gilks (1993) and Dellaportas & Stephens (1995). It is based on structural specifications and functional specifications. The measurement error model based on structural specifications entails the formulation of three sub-models (see also Section 6.1). The submodels consist of:

- A *disease model* that relates the response y to the true latent covariate ξ ;
- A *measurement model* relating X and ξ ; and
- An *exposure model* for the prior distribution of ξ .

A necessary assumption is that we have a non-differential measurement error which can be expressed as

$$y_t \perp X_t | \xi_t \quad \text{for all } t.$$

7.1.1 Model formulation

At the second stage, the functional forms of the distributions involved in the sub-models have to be chosen. For the outcome and the measurement models, this choice is guided by epidemiological knowledge. Thus, the current measurement error problem can be formalized as follows:

1. Disease model

We consider a Poisson model with log-link and predictor given similar to Section 6.1.2 as disease model.

2. Measurement model

We specify a *classical measurement model* relating the unobservable ξ to the observable variables X_1 and X_2 . Specifically, we assume that the measurements are governed by the model

$$X_{ti} = \xi_t + \alpha_i + \epsilon_{ti}, \quad \text{for } t = 1, \dots, T, \text{ and } i = 1, 2, \quad (7.1)$$

where α_i are measurement biases associated with the two devices; the ϵ_{ti} are independent $N(0, \sigma_{\epsilon_i}^2)$, $i = 1, 2$. Further, we assume $\alpha_2 = 0$ for identification.

Let τ_1 and τ_2 denote the measurement error precisions which are defined as $\tau_1 \equiv 1/\sigma_{\epsilon_1}^2$ and $\tau_2 \equiv 1/\sigma_{\epsilon_2}^2$, respectively. We assume the conditional distribution of the system measurements given the true value being measured at time t , ξ_t , the device bias α_1 , and the precisions τ_1 and τ_2 , to be as follows:

$$\left(\begin{bmatrix} X_{t1} \\ X_{t2} \end{bmatrix} \mid \xi_t, \alpha_1, \tau_1, \tau_2 \right) \sim N_2 \left(\begin{pmatrix} \xi_t + \alpha_1 \\ \xi_t \end{pmatrix}, \begin{pmatrix} 1/\tau_1 & 0 \\ 0 & 1/\tau_2 \end{pmatrix} \right). \quad (7.2)$$

In the case of autocorrelated measurement errors, an appropriate modification is given by the assumption of a first order autoregressive process satisfying

$$\epsilon_{ti} = \rho_i \epsilon_{t-1i} + u_{ti}, \quad i = 1, 2, \quad (7.3)$$

where the u_{ti} are $N(0, \sigma_{u_i}^2)$ and independent of other errors over time. This results in a conditional distribution for the vectors of all observations, X_1 and X_2 , as follows:

$$\left(\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \mid \xi, \alpha_1, \tau_1, \tau_2, \rho_1, \rho_2 \right) \sim N_{2T} \left(\begin{pmatrix} \xi + \alpha_1 \\ \xi \end{pmatrix}, \begin{pmatrix} \frac{1}{\tau_1} V_1 & 0 \\ 0 & \frac{1}{\tau_2} V_2 \end{pmatrix} \right), \quad (7.4)$$

with

$$V_i = \frac{1}{(1 - \rho_i^2)} \begin{bmatrix} 1 & \rho_i & \cdots & \rho_i^{T-1} \\ \rho_i & 1 & \cdots & \rho_i^{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_i^{T-1} & \rho_i^{T-2} & \cdots & 1 \end{bmatrix}.$$

3. Exposure model

In the exposure model, we specify the distribution of the latent variable ξ . There might be non-stationarity in the time series. Following Isaacson & Zimmerman (2000), we assume that ξ has a mean that follows a linear trend. It will be further assumed that deviations from the mean follow an autoregressive process of order 1 (AR(1)). This implies a model of the following form:

$$\xi_t = \kappa_0 + \kappa_1 t + \rho_{\xi,1}[\xi_{t-1} - (\kappa_0 + (t-1)\kappa_1)] + e_t, \quad t = 1, \dots, T, \quad (7.5)$$

where e_1 is assumed to be $N(0, \sigma_e^2/(1 - \rho^2))$, independently of e_2, \dots, e_T . The e_2, \dots, e_T , on the other hand, are iid $N(0, \sigma_e^2)$.

Starting from this exposure model, the conditional distribution of ξ_1 given the parameters $\kappa_0, \kappa_1, \rho_{\xi,1}$ and τ_e is given by

$$(\xi_1 | \kappa_0, \kappa_1, \rho_{\xi,1}, \tau_e) \sim N \left(\kappa_0 + \kappa_1, \frac{1}{\tau_e(1 - \rho_{\xi,1}^2)} \right), \quad (7.6)$$

where $\tau_e \equiv 1/\sigma_e^2$. The conditional distribution of all other ξ_t given a previous value ξ_{t-1} and the parameters is assumed to be:

$$(\xi_t | \xi_{t-1}, \kappa_0, \kappa_1, \rho_{\xi,1}, \tau_e) \sim N(\kappa_0 + t\kappa_1 + \rho_{\xi,1}(\xi_{t-1} - (\kappa_0 + (t-1)\kappa_1)), 1/\tau_e). \quad (7.7)$$

It is easily seen that $\xi = (\xi_1, \dots, \xi_T)'$ is multivariate Gaussian with covariance matrix elements

$$\Omega_{tt'} = \text{Cov}(\xi_t, \xi_{t'}) = \frac{1}{\tau_e(1 - \rho_{\xi,1}^2)} \rho_{\xi,1}^{|t-t'|}, \quad t, t' \in 1, \dots, T, \quad (7.8)$$

defining the joint/full covariance matrix

$$\Omega = \frac{\sigma_e^2}{(1 - \rho_{\xi,1}^2)} \begin{bmatrix} 1 & \rho_{\xi,1} & \cdots & \rho_{\xi,1}^{T-1} \\ \rho_{\xi,1} & 1 & \cdots & \rho_{\xi,1}^{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{\xi,1}^{T-1} & \rho_{\xi,1}^{T-2} & \cdots & 1 \end{bmatrix}.$$

Its inverse, Ω^{-1} , is tri-diagonal with elements

$$\Omega^{-1} = \frac{1}{\sigma_e^2} \begin{bmatrix} 1 & -\rho_{\xi,1} & 0 & 0 & \dots & 0 \\ -\rho_{\xi,1} & 1 + \rho_{\xi,1}^2 & -\rho_{\xi,1} & 0 & \dots & 0 \\ 0 & -\rho_{\xi,1} & 1 + \rho_{\xi,1}^2 & -\rho_{\xi,1} & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & -\rho_{\xi,1} & 1 + \rho_{\xi,1}^2 & -\rho_{\xi,1} \\ 0 & \dots & 0 & 0 & -\rho_{\xi,1} & 1 \end{bmatrix},$$

which is very useful for computationally efficient MCMC updating schemes.

7.1.2 Prior specifications for model parameters and variances

The remaining random quantities, not yet specified, are the parameter vector β , as well as the unknown functions f_j , $j = l + 1, \dots, p - 1$ and $f_p(t)$ of the disease model and the measurement model parameter α_1 . Due to the complexity of the model, the parameters $\rho_{\xi,1}$, ρ_1 , ρ_2 as well as σ_1^2 , σ_2^2 and σ_e^2 and the two exposure model parameters κ_0, κ_1 are taken to be fixed, obtained by the methods described in Section 6.3.

7.1.2.1 Disease model parameters

For each of the fixed effects, we assume a diffuse prior

$$p(\beta_j) \propto \text{const},$$

whereas the unknown functions f_j and $f_j(t)$, $j = l + 1, \dots, p$ are modeled using Bayesian P-splines as well as first or second order random walks.

Bayesian P-splines The basic idea behind Bayesian P-splines is to approximate a function f_j or $f_j(t)$ by a linear combination of B-spline basis functions. A moderately large number of equally spaced knots (usually between 20 and 40) is used to ensure enough flexibility. To obtain sufficient smoothness of the fitted curves, it is assumed that adjacent B-spline coefficients follow first- or second-order random walks with Gaussian errors. For a more detailed description of Bayesian P-splines, we refer to Section 3.2.2 or Lang & Brezger (2004).

Random walks Instead of assuming a random walk prior for the parameters of a P-spline, a random walk can also be directly applied on the function evaluations $f_j = (f_{j1}, \dots, f_{jm}, \dots, f_{jT})'$, with $f_{jm} = f(x_{j(m)})$. Such random walk models are frequently

used in the analysis of time series but can also be applied in additive regression models (see Fahrmeir & Lang (2001), for example).

In case of equidistant knots, random walks of first or second order are defined as

$$f_m - f_{m-1} \sim N(0, \tau_f^2), \quad \text{for } m = 2, \dots, M, \quad (7.9)$$

and

$$f_m - 2f_{m-1} + f_{m-2} \sim N(0, \tau_f^2), \quad \text{for } m = 3, \dots, M,$$

respectively (see also Section 3.2.3). Diffuse priors

$$f_1 \propto \text{const},$$

as well as

$$f_1 \propto \overline{\text{const}} \quad \text{and} \quad f_2 \propto \text{const}$$

are chosen for initial values.

The joint distribution of f in the RW1 case (7.9) can be factorized as:

$$\begin{aligned} p(f) &= p(f_M|f_{M-1}) \cdots p(f_m|f_{m-1}) \cdots p(f_2|f_1)p(f_1) \\ &\propto \exp\left(-\frac{1}{2\tau_f^2} \sum_{m=2}^M (f_m - f_{m-1})^2\right) \\ &\propto \exp\left(-\frac{1}{2\tau_f^2} f' K_{RW1} f\right). \end{aligned} \quad (7.10)$$

The penalty matrix K_{RW1} is of the form $K_{RW1} = D_1' D_1$, where D_1 is a first order difference matrix such as

$$D_1 = \begin{pmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & & \ddots & \ddots \\ & & & & -1 & 1 \end{pmatrix}.$$

The same procedure applied to a second order random walk gives the joint density

$$\begin{aligned} p(f) &= p(f_M|f_{M-1}, f_{M-2}) \cdots p(f_m|f_{m-1}, f_{m-2}) \cdots p(f_3|f_2, f_1) p(f_2) p(f_1) \\ &\propto \exp\left(-\frac{1}{2\tau_f^2} \sum_{m=2}^M (f_m - 2f_{m-1} + f_{m-2})^2\right) \\ &\propto \exp\left(-\frac{1}{2\tau_f^2} f' K_{RW2} f\right), \end{aligned} \quad (7.11)$$

f_j have to be redefined by multiplying each element in row t of f_j with x_{tj} . Hence, the overall matrix formulation for the time-varying terms is given by

$$f_j(t, x_j) = f_j(t)x_j = \text{diag}(x_{1j}, \dots, x_{Tj}) f_j.$$

7.1.2.2 Measurement model

The measurement device bias α_1 is assumed to have a normal distribution

$$\alpha_1 \sim N(\mu_{\alpha_1}, 1/\tau_{\alpha_1}), \quad (7.13)$$

where $\tau_{\alpha_1} = 1/\sigma_{\alpha_1}^2$.

7.1.3 MCMC Estimation

Let $\gamma = (\gamma_{l+1}, \dots, \gamma_{p-1})'$ denote the vector of all regression coefficients of the B-spline bases, $f = f_p$ the vector of the random walk coefficients of function $f_p(t)$, β the vector of all fixed effects, and τ_γ^2 as well as τ_f^2 the vectors of the corresponding variance parameters.

We get the posterior of the unknown parameters as

$$\begin{aligned} \pi(\gamma, \tau_\gamma^2, f, \tau_f^2, \beta, \alpha_1, \xi | \text{data, other parameters}) &\propto L(y, x_1, x_2, \gamma, \tau_\gamma^2, f, \tau_f^2, \beta, \alpha_1, \xi) \\ &\quad p(\gamma, \tau_\gamma^2, f, \tau_f^2, \beta, \alpha_1, \xi). \end{aligned}$$

By assuming prior independence, as is usually done, the joint prior distribution of the unknown parameters can be expressed as the product of the marginal priors:

$$p(\gamma, \tau_\gamma^2, f, \tau_f^2, \beta, \alpha_1, \xi) = p(\gamma | \tau_\gamma^2) p(\tau_\gamma^2) p(f | \tau_f^2) p(\tau_f^2) p(\beta) p(\alpha_1) p(\xi).$$

By suppressing conditioning parameters as well as data notationally, the posterior is obtained in the following form

$$\begin{aligned} \pi(\gamma, \tau_\gamma^2, f, \tau_f^2, \beta, \alpha_1, \xi | \cdot) &\propto L(y, x_1, x_2, \gamma, \tau_\gamma^2, f, \tau_f^2, \beta, \alpha_1, \xi) \\ &\quad p(\gamma | \tau_\gamma^2) p(\tau_\gamma^2) p(f | \tau_f^2) p(\tau_f^2) p(\beta) p(\alpha_1) p(\xi). \end{aligned}$$

Bayesian inference is based on the analysis of the posterior distribution. The posterior distribution described above has no 'nice' closed form which could be used to draw samples for inference, but rather a complicated high dimensional density only known up to the proportionality constant. Estimation of the unknown parameters is, therefore, done by simulation of the posterior with Markov Chain Monte Carlo (MCMC) techniques.

The key idea of MCMC is to construct a Markov chain that has the posterior distribution of interest as its stationary distribution. Then samples are collected from that chain. This way, a sample of dependent random numbers from the posterior is obtained. For a detailed introduction and an overview of MCMC methods see, for example, Casella & George (1992), Chib & Greenberg (1995), Gilks, Richardson & Spiegelhalter (1996), Robert & Casella (2004) and the references therein.

7.1.3.1 Full Conditionals and updating of the full conditionals

MCMC simulation methods with proposal densities split the model parameters into subsets and then sample from the conditional distributions given the remaining parameters and data, which are called *full conditionals*. These full conditionals often have a much simpler structure than the posterior distribution itself. In the following, the full conditionals will be denoted, for example, by $\pi(\nu | \dots)$.

Consequently, we can now obtain the full conditionals and the corresponding updating schemes for each subset or block of parameters. We use the blocks $\gamma_{l+1}, \dots, \gamma_{p-1}, \tau_{\gamma_{l+1}}^2, \dots, \tau_{\gamma_{p-1}}^2, f_p, \tau_f^2, \beta, \alpha_1, \xi$.

7.1.3.2 Disease model parameters and their hyperparameters

Due to the unified form for the priors of γ and f given in (3.10) as well as in (7.10) and in (7.11), we can represent their full conditionals in a compact way as product of the joint likelihood and the prior,

$$\begin{aligned} \pi(\nu | \dots) &\propto L(y|\eta) p(\nu|\tau_\nu^2) \\ &\propto L(y|\eta) \exp\left(-\frac{1}{2\tau_\nu^2} \nu' K_\nu \nu\right), \end{aligned} \quad (7.14)$$

where $\nu \in \gamma, f$, the penalty matrix $K_\nu \in K_\gamma, K_f$ and the hyperparameter $\tau_\nu^2 \in \tau_\gamma^2, \tau_f^2$, respectively.

Sampling For updating the parts ν_j of the parameter vector ν corresponding to the time-independent functions f_j , a Metropolis Hastings (MH)-algorithm based on iteratively weighted least squares (IWLS) proposals is used. These IWLS proposals were originally proposed by Gamerman (1997) in the context of generalized linear mixed models and adopted to a more general setting by Brezger & Lang (2006).

Suppose now that we want to update the regression coefficients ν_j of the function f_j with current state ν_j^c of the chain. Then, following the concept of IWLS-MH, a new value,

denoted as ν_j^n , is proposed by drawing random numbers from a multivariate Gaussian proposal distribution

$$\nu_j^n \sim N(m_j, P_j^{-1}).$$

The mean m_j is obtained by one Fisher scoring step to maximize the full conditional $\pi(\nu|\dots)$, whereas the precision matrix P_j is the inverse expected Fisher information, evaluated at ν_j^c . Hence, the mean and the precision matrix are given as

$$m_j = P_j^{-1} X_j' W(\nu_j^c) (\tilde{y}(\nu_j^c) - \tilde{\eta}), \quad (7.15)$$

$$P_j = X_j' W(\nu_j^c) X_j + \frac{1}{\tau_j^2} K_j, \quad (7.16)$$

where X_j are the corresponding design matrices and $\tilde{\eta}$ corresponds to the part of the predictor associated with all remaining effects in the model. K_j is a penalty matrix as defined in (3.10). Furthermore, $W(\nu_j^c) = \text{diag}(w_1, \dots, w_T)$ is the usual weight matrix for IWLS (see for example Fahrmeir & Tutz (2001)), calculated from ν_j^c ; the working observations $\tilde{y} = (\tilde{y}_1(\nu_j^c), \dots, \tilde{y}_T(\nu_j^c))'$ are defined as

$$\tilde{y}_t(\nu_j^c) = \eta_t + (y_t - \mu_t) g'(\mu_t).$$

The proposed vector ν_j^n is accepted as the new state of the chain with probability

$$\alpha(\nu_j^c, \nu_j^n) = \min \left\{ 1, \frac{\pi(\nu_j^n | \cdot) q(\nu_j^c, \nu_j^n)}{\pi(\nu_j^c | \cdot) q(\nu_j^n, \nu_j^c)} \right\}$$

where $\pi(\nu_j^c | \cdot)$ and $\pi(\nu_j^n | \cdot)$ is the full conditional evaluated at either ν_j^c or ν_j^n , and $q(\nu_j^n, \nu_j^c)$ and $q(\nu_j^c, \nu_j^n)$ are the corresponding proposal densities, that means multivariate Gaussian densities corresponding to (7.15) and (7.16).

In principle, the IWLS-MH algorithm could also be used for updating the parameter vector ν_p , which corresponds to the time-dependent function f_p . However, especially in the context of latent variables, this would require considerably more computational effort. Therefore, we adopt a computationally faster MH-algorithm based on *conditional prior proposals*, although mixing properties of these proposals are inferior compared to the IWLS-MH algorithm (Brezger & Lang (2006)). The conditional prior proposal algorithm was first suggested by Knorr-Held (1999) for dynamic generalized linear models and extended for generalized additive mixed models in Fahrmeir & Lang (2001).

Suppose that $f_{[m:n]}$ denote sub-vectors $(f_m, f_{m+1}, \dots, f_n)'$ out of the function evaluations $f_p = (f_1, f_2, \dots, f_t, \dots, f_T)$. Full conditionals for these sub-vectors are given by the

product of all likelihood contributions of the posterior that depend on $f_{[m:n]}$, and the conditional distribution of $f_{[m:n]}$:

$$\pi(f_{mn} | \dots) \propto L(f_{mn}) p(f_{[m:n]} | f_{s, s \notin [m,n]}, \tau^2).$$

In order to obtain the conditional distribution of $f_{[m:n]}$, let $K_{[m:n]}$ be the corresponding sub-matrix out of the first- or second-order precision matrices K_{RW1} and K_{RW2} , given by the rows and columns m to n . Further, $K_{[1:m-1]}$ as well as $K_{[n+1:T]}$ denote the matrices to the left and right of $K_{[m:n]}$ such that

$$K = \begin{pmatrix} & K'_{[1:m-1]} & \\ K_{[1:m-1]} & K_{[m:n]} & K_{[n+1:T]} \\ & K'_{[n+1:T]} & \end{pmatrix}.$$

Then the conditional distribution of $f_{[m:n]}$ given $f_{[1:m-1]}$ and $f_{[n+1:T]}$ is defined by

$$f_{[m:n]} | f_{[1:m-1]}, f_{[n+1:T]} \sim N(\mu_{[m:n]}, \Sigma_{[m:n]}) \quad (7.17)$$

with

$$\mu_{[m:n]} = \begin{cases} -K_{[m:n]}^{-1} K_{[n+1:T]} f_{[n+1:T]} & m = 1 \\ -K_{[m:n]}^{-1} (K_{[1:m-1]} f_{[1:m-1]} + K_{[n+1:T]} f_{[n+1:T]}) & m > 1, n < T \\ -K_{[m:n]}^{-1} K_{[1:m-1]} f_{[1:m-1]} & n = T \end{cases}$$

and $\Sigma_{[m:n]} = K_{mn}^{-1}$. All elements in K outside the z off-diagonals are zero, where z is the order of the corresponding difference matrix. Therefore, only the f_{m-z}, \dots, f_{m-1} and f_{n+1}, \dots, f_{n+z} enter in $\mu_{[m:n]}$ (Knorr-Held (1999)).

One way to update f_p is to use a so-called *single move* conditional prior proposal, that means f_t is updated one at a time. Updates are obtained by generating a new proposed f_t^n from the conditional distribution $N(\mu_{tt}, \Sigma_{tt})$. The Hastings acceptance probability is then given by

$$\min \left\{ 1, \frac{L(y_t | f_t^n)}{L(y_t | f_t^c)} \right\}, \quad (7.18)$$

which is simply the likelihood ratio for observation y_t .

Since the single move scheme might converge very slowly, it is recommended to alternatively use so-called *block* moves based on updating one block $f_{[m:n]}$ at a time. This way, a conditional prior proposal can be implemented similarly as for the single move scheme. Block move updates for $f_{[m:n]}$ are obtained by generating $f_{[m:n]}^n$ from the conditional prior proposal $N(\mu_{[m:n]}, \Sigma_{[m:n]})$. The acceptance probability is then given by

$$\min \left\{ 1, \frac{\prod_{t=m}^n L(y_t | f_t^n)}{\prod_{t=m}^n L(y_t | f_t^c)} \right\}.$$

The full conditional of β is given by the product of the likelihood of the disease model and the prior

$$\begin{aligned}\pi(\beta|\dots) &\propto L(y|\eta) p(\beta) \\ &\propto L(y|\eta),\end{aligned}\tag{7.19}$$

where the last factor can be omitted, as we have chosen $p(\beta) \propto \text{const}$ for the fixed effects. Hence, the full conditional of β is proportional to the likelihood of the disease model.

Sampling A new value β_j^n is proposed by drawing from the Gaussian proposal density $q(\beta_j^c, \beta_j^n)$ with mean and precision matrix

$$\begin{aligned}m_\beta &= P_\beta^{-1} U' W(\beta^c) (\tilde{y}(\beta^c) - \tilde{\eta}), \\ P_\beta &= U' W(\beta^c) U\end{aligned}$$

where U is the design matrix of fixed effects. The working observations \tilde{y} and $\tilde{\eta}$ are defined as for the sampling of ν .

In accordance to the unknown functions f_j , we can also obtain a general form for the full conditional for all $\tau_\nu^2 = \tau_\gamma^2, \tau_f^2$. The full conditional is then given by

$$\begin{aligned}\pi(\tau_\nu^2|\dots) &\propto p(\nu|\tau_\nu^2) p(\tau_\nu^2) \\ &\propto IG\left(a + \frac{1}{2}\text{rank}(K_\nu), \frac{1}{2}\nu' K_\nu \nu + b\right).\end{aligned}\tag{7.20}$$

To update the τ_ν^2 , a new value can be directly sampled from the inverse gamma densities,

$$\tau_\nu^{2n} \sim IG(a', b'),\tag{7.21}$$

with $a' = a + \frac{1}{2}\text{rank}(K_\nu)$ and $b' = b + \frac{1}{2}\nu' K_\nu \nu$. The sampled value τ_ν^{2n} is then accepted as the next stage in the chain for τ_ν^2 .

7.1.3.3 Measurement error model parameter

In the measurement error model, the underlying distribution is a multivariate Normal. We further assume a normal prior for α_1 (see assumption (7.13)). The full conditional for α_1 is given by the product of the likelihood of the measurement error model and the

prior for α_1 , that means

$$\begin{aligned}
\pi(\alpha_1|\dots) &\propto L(X_1|\cdot) p(\alpha_1) \\
&\propto \exp\left[-\frac{1}{2\sigma_1^2} \sum_{t=1}^T (X_{t1} - \alpha_1 - \xi_t)^2\right] \exp\left[-\frac{1}{2\sigma_{\alpha_1}^2} (\alpha_1 - \mu_{\alpha_1})^2\right] \\
&\propto \exp\left[-\frac{1}{2} \left\{ \tau_1 \sum_t (X_{t1} - \alpha_1 - \xi_t)^2 + \tau_{\alpha_1} (\alpha_1 - \mu_{\alpha_1})^2 \right\}\right] \\
&\propto \exp\left[-\frac{1}{2} \left\{ \alpha_1 (T \tau_1 + \tau_{\alpha_1}) \alpha_1 - 2\alpha_1 (\tau_1 \sum_t (X_{t1} - \xi_t) + \tau_{\alpha_1} \mu_{\alpha_1}) \right\}\right] \\
&= N(m_{\alpha_1}, \tau_{\alpha_1} + T \tau_1), \tag{7.22}
\end{aligned}$$

with

$$m_{\alpha_1} = \frac{\mu_{\alpha_1} \tau_{\alpha_1} + \tau_1 \sum_T (X_{t1} - \xi_t)}{\tau_{\alpha_1} + T \tau_1}$$

where τ_1 is the measurement error precision of the corresponding device.

Hence, the full conditional of α_1 is given by a Normal distribution. Therefore, an update of α_1 can be obtained by using an additional Gibbs step.

7.1.3.4 True values ξ

Let X denote the combined vector of all measured observations, that means

$$X = \begin{bmatrix} X_{11} \\ \vdots \\ X_{T1} \\ X_{12} \\ \vdots \\ X_{T2} \end{bmatrix}.$$

Furthermore, let α be a vector of dimension $2T$, with $\alpha = (\alpha_1, \dots, \alpha_1, 0, \dots, 0)'$. The full conditional for ξ is given by the product of the disease model likelihood, the measurement error model likelihood and the prior for ξ :

$$\begin{aligned}
\pi(\xi|\dots) &\propto L(y|\eta) L(X|\cdot) p(\xi) \\
&\propto \exp(y \log(\mu) - \mu) \exp\left(-\frac{1}{2} (X - \alpha - (Z\xi))' \Sigma^{-1} (X - \alpha - (Z\xi))\right) \\
&\quad \exp\left(-\frac{1}{2} (\xi - \mu_\xi)' \Omega^{-1} (\xi - \mu_\xi)\right) \\
&\propto \exp(y \log(\mu) - \mu) \exp\left(-\frac{1}{2} [(X - \alpha - (Z\xi))' \Sigma^{-1} (X - \alpha - (Z\xi)) \right. \\
&\quad \left. + (\xi - \mu_\xi)' \Omega^{-1} (\xi - \mu_\xi)]\right), \tag{7.23}
\end{aligned}$$

where $Z = \begin{bmatrix} I_T \\ I_T \end{bmatrix}$, i. e. a matrix of dimension $2T \times T$, containing two identity matrices of dimension $T \times T$. One can see that it is impossible to find an appropriate known distribution which is proportional to this full conditional. This, however, means that an additional Metropolis-Hastings (MH) step for the update of ξ has to be implemented.

Sampling It could be considered to update all the ξ_t , $t = 1, \dots, T$, simultaneously. However, since the number of ξ_t can be large, it is hard to find proposal densities which approximate the full conditional of ξ well. Consequently, this approach suffers from slow mixing due to low acceptance probabilities. Therefore, we use again a conditional prior proposal for updating the latent true exposure ξ .

In the following, we specify the proposal for ξ to be

$$\begin{aligned} p(\xi|\cdot) &\propto \exp\left(-\frac{1}{2}[(X - \alpha - (Z\xi))'\Sigma^{-1}(X - \alpha - (Z\xi))] - \frac{1}{2}[(\xi - \mu_\xi)'\Omega^{-1}(\xi - \mu_\xi)]\right) \\ &\quad \exp\left(-\frac{1}{2}[(X - \alpha - (Z\xi))'\Sigma^{-1}(X - \alpha - (Z\xi)) + (\xi - \mu_\xi)'\Omega^{-1}(\xi - \mu_\xi)]\right) \\ &\propto \exp\left(-\frac{1}{2}[\xi'(Z'\Sigma^{-1}Z + \Omega^{-1})\xi - 2\xi'(Z'\Sigma^{-1}(X - \alpha) + \Omega^{-1}\mu_\xi)]\right), \end{aligned}$$

which is a multivariate normal distribution $N(m, \Sigma)$, where

$$m = (Z'\Sigma^{-1}Z + \Omega^{-1})^{-1} [Z'\Sigma^{-1}(X - \alpha) + \Omega^{-1}\mu_\xi].$$

For the assumption of autocorrelated true values, but independent measurement errors, Σ is given by

$$\begin{aligned} \Sigma &= (Z'\Sigma^{-1}Z + \Omega^{-1})^{-1} \\ &= \begin{bmatrix} \tau_1 + \tau_2 + \tau_e & -\rho_{\xi,1}\tau_e & & & \\ -\rho_{\xi,1}\tau_e & \tau_1 + \tau_2 + \tau_e(1 + \rho_{\xi,1}^2) & -\rho_{\xi,1}\tau_e & & \\ & \ddots & \ddots & \ddots & \\ & -\rho_{\xi,1}\tau_e & \tau_1 + \tau_2 + \tau_e(1 + \rho_{\xi,1}^2) & -\rho_{\xi,1}\tau_e & \\ & & -\rho_{\xi,1}\tau_e & \tau_1 + \tau_2 + \tau_e & \end{bmatrix}^{-1}. \end{aligned}$$

If we also assume autocorrelated measurement errors, the only difference in the proposal for ξ is given by a slightly modified covariance matrix such that

$$\begin{aligned} \Sigma &= (Z'\Sigma^{-1}Z + \Omega^{-1})^{-1} = \\ &= \begin{bmatrix} \tau_1 + \tau_2 + \tau_e & -\rho_1\tau_1 - \rho_2\tau_2 - \rho_{\xi,1}\tau_e & & & \\ -\rho_1\tau_1 - \rho_2\tau_2 - \rho_{\xi,1}\tau_e & \tau_1(1 + \rho_1^2) + \tau_2(1 + \rho_2^2) + \tau_e(1 + \rho_{\xi,1}^2) & -\rho_1\tau_1 - \rho_2\tau_2 - \rho_{\xi,1}\tau_e & & \\ & \ddots & \ddots & \ddots & \\ & & & & \end{bmatrix}^{-1}. \end{aligned}$$

The conditional distribution of $\xi_{[m:n]}$ given by $\xi_{[1:m-1]}$ and $\xi_{[n+1:T]}$ is

$$\xi_{[m:n]} | \xi_{[1:m-1]}, \xi_{[n+1:T]} \sim N(\mu_{[m:n]}, \Sigma_{[m:n]}). \quad (7.24)$$

In contrast to the case of random walks, which are assumed to have zero mean, the mean has now a slightly different form

$$\mu_{[m:n]} = \begin{cases} m_{[m:n]} - K_{[m:n]}^{-1} K_{[n+1:T]} (\xi_{[n+1:T]} - m_{[n+1:T]}) & m = 1 \\ m_{[m:n]} - K_{[m:n]}^{-1} K_{[1:m-1]} (\xi_{[1:m-1]} - m_{[1:m-1]}) & n = T \\ m_{[m:n]} - K_{[m:n]}^{-1} \{ K_{[1:m-1]} (\xi_{[1:m-1]} - m_{[1:m-1]}) \\ + K_{[n+1:T]} (\xi_{[1:m-1]} - m_{[1:m-1]}) \} & \text{else} \end{cases}$$

and $\Sigma_{[m:n]} = K_{[m:n]}^{-1}$.

Comparable to the conditional prior proposal for nonlinear functions, single move or block move updates can be chosen.

7.1.4 Sampling Algorithm

The steps of the sampling algorithm can be summarized as follows:

1. Initialize $\gamma_{l+1}^{(0)}, \dots, \gamma_{p-1}^{(0)}, \tau_{\gamma_{l+1}}^{2(0)}, \dots, \tau_{\gamma_{p-1}}^{2(0)}, f_p^{(0)}, \tau_f^{2(0)}, \beta^{(0)}, \alpha_1^{(0)}$ and $\xi^{(0)}$, and set the number of iterations $c = 0$
2. Set $c = c + 1$.
3. For $j = l + 1, \dots, p - 1$ update γ_j with a IWLS MH-step in the following way:
 - a) Compute the likelihood $L(y, \dots, \gamma_j^c, \dots, \xi^c)$.
 - b) Draw a proposed new value γ_j^n from the Gaussian proposal density $q(\gamma_j^n, \gamma_j^c)$, with mean m_j and precision matrix P_j as given in (7.15) and (7.16), respectively.
 - c) Compute the likelihood $L(y, \dots, \gamma_j^n, \dots, \xi^c)$ as well as the full conditionals $\pi(\gamma_j^c | \cdot)$, $\pi(\gamma_j^n | \cdot)$ and proposal densities $q(\gamma_j^c, \gamma_j^n)$ and $q(\gamma_j^n, \gamma_j^c)$.
 - d) Accept γ_j^n as the new state of the chain γ_j^c with the corresponding probability, otherwise keep γ_j^c as the current state.
4. Update f_p with a conditional prior proposal MH-step in the following way:
 - a) Choose the block size.
For every block $m : n$:

- b) Compute the likelihood $L(y, \dots, f_{[m:n]}^c, \dots, \xi^c)$.
 - c) Draw a proposed new value $f_{[m:n]}^n$ from the conditional prior proposal (7.17).
 - d) Compute the likelihood $L(y, \dots, f_{[m:n]}^n, \dots, \xi^c)$.
 - e) Accept $f_{[m:n]}^n$ as the new state of the chain $f_{[m:n]}^c$ with the corresponding probability, otherwise keep $f_{[m:n]}^c$ as the current state.
5. Update the fixed parameters by similar steps as for updating the γ_j .
 6. For $j = l + 1, \dots, p$ update τ_j^2 by drawing from inverse gamma full conditionals with parameters given in (7.21).
 7. Update ξ with a conditional prior proposal MH-step in the following way:
 - a) Choose the block size.
For every block $m : n$:
 - b) Compute the likelihood $L(y, \dots, \xi_{[m:n]}^c, \dots)$.
 - c) Draw a proposed new value $\xi_{[m:n]}^n$ from the conditional prior proposal (7.24).
 - d) Compute the likelihood $L(y, \dots, \xi_{[m:n]}^n, \dots)$.
 - e) Accept $\xi_{[m:n]}^n$ as the new state of the chain $\xi_{[m:n]}^c$ with probability

$$\min \left\{ 1, \frac{\prod_{t=m}^n L(y_t | \xi_t^n)}{\prod_{t=m}^n L(y_t | \xi_t^c)} \right\},$$
 otherwise keep $\xi_{[m:n]}^c$ as the current state.
 8. Update α_1 by drawing from the normal full conditionals given in (7.22).
 9. Go to 2. until $c = C$, that means the total number of iterations.

We conclude this section with some remarks:

- Starting values

For the MCMC algorithm, we have to specify starting values for all parameters. For the terms in the disease model predictor the starting values are the posterior mode estimates (Brezger & Lang (2006)). As starting values for the sampling of the ξ_t , the corresponding mean of the two observed series X_{t1} and X_{t2} is used, whereas for α_1 we take the estimates obtained by an ARMA model as described in Section 6.3.

- Optimal block size

The choice of the block sizes is an important point and we follow here Fahrmeir & Lang (2001) who suggest to specify a minimum and maximum block size. The block size is then chosen randomly in every iteration of the MCMC simulations. In our experience, the best results in terms of mixing and autocorrelations in the sampled parameters are obtained with a block size between 1 and 100.

- The Bayesian measurement error correction method has been implemented in BayesX (Brezger, Kneib & Lang (2005)). Further details about estimating such models can be found in Appendix A.

7.2 A measurement error calibration model

One of the most popular methods for dealing with measurement error in covariates is regression calibration. Armstrong (1985) originally introduced regression calibration as a method in linear models, whereas Gleser (1990) proposed regression calibration methods in the context of logistic regression. Another formulation of the same method was suggested by Carroll & Stefanski (1990).

The basic idea of regression calibration is to predict the unobservable variable ξ by means of regression, and then to include this predicted variable in the main regression model. After this approximation, one performs a standard analysis. As such, it is applicable to any regression modelling setting.

Again, we specify a *classical measurement model* for the observed series X_{ti} (see Section 7.1.1):

$$X_{ti} = \xi_t + \alpha_i + \epsilon_{ti}, \quad \text{for } t = 1, \dots, T, \text{ and } i = 1, 2. \quad (7.25)$$

We suppose that $\alpha_2 = 0$ for identification.

It will be further assumed that the latent variable ξ is given by

$$\xi_t = \kappa_0 + \kappa_1 t + \rho_{\xi,1} [\xi_{t-1} - (\kappa_0 + (t-1)\kappa_1)] + e_t, \quad t = 1, \dots, T, \quad (7.26)$$

with the specifications already defined in Section 7.1.1.

The complete measurement data can now be written in the vector $X = (X_{11}, X_{21}, \dots, X_{T1}, X_{12}, X_{22}, \dots, X_{T2})'$. Combining models (7.25) and (7.26), we get that

$$X \sim N_{2T}(U\theta, \Sigma), \quad (7.27)$$

where

$$U = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ \vdots & \vdots & \vdots \\ 1 & T & 1 \\ 1 & 1 & 0 \\ \vdots & \vdots & \vdots \\ 1 & T & 0 \end{pmatrix}, \quad \theta = \begin{pmatrix} \kappa_0 \\ \kappa_1 \\ \alpha_1 \end{pmatrix}, \quad (7.28)$$

and the variance-covariance matrix is given by

$$\begin{aligned} \Sigma &= \Sigma(\varphi) = \Sigma_\epsilon + \Omega \\ &= \left(\begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix} \otimes I_T \right) + \left(\frac{\sigma_\epsilon^2}{1-\rho^2} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \otimes \Phi_T \right) \end{aligned} \quad (7.29)$$

with

$$\Phi_T = \begin{bmatrix} 1 & \rho & \dots & \rho^{T-1} \\ \rho & 1 & \dots & \rho^{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \dots & 1 \end{bmatrix}, \quad \text{and} \quad \varphi = (\sigma_1^2, \sigma_2^2, \rho, \sigma_\epsilon^2)'. \quad (7.30)$$

Comparing this model formulation to that of Section 7.1, we make the same distributional assumptions that were previously described for the Bayesian correction model; however, we express them in a different way.

Note that for the case of autocorrelated measurement errors, an appropriate modification is given by the assumption of $\sigma_1^2 V_1$ and $\sigma_2^2 V_2$, where V_j has an equivalent structure to Φ_T . Hence, Σ_ϵ is replaced by

$$\Sigma_\epsilon = \begin{bmatrix} \sigma_1^2 & \sigma_1^2 \rho_1 & \dots & \sigma_1^2 \rho_1^{T-1} & 0 & \dots & \dots & 0 \\ \sigma_1^2 \rho_1 & \sigma_1^2 & \dots & \sigma_1^2 \rho_1^{T-2} & \vdots & & & \vdots \\ \vdots & \vdots & \ddots & \vdots & \vdots & & & \vdots \\ \sigma_1^2 \rho_1^{T-1} & \sigma_1^2 \rho_1^{T-2} & \dots & \sigma_1^2 & 0 & \dots & \dots & 0 \\ 0 & \dots & \dots & 0 & \sigma_2^2 & \sigma_2^2 \rho_2 & \dots & \sigma_2^2 \rho_2^{T-1} \\ \vdots & & & \vdots & \sigma_2^2 \rho_2 & \sigma_2^2 & \dots & \sigma_2^2 \rho_2^{T-2} \\ \vdots & & & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \dots & 0 & \sigma_2^2 \rho_2^{T-1} & \sigma_2^2 \rho_2^{T-2} & \dots & \sigma_2^2 \end{bmatrix}.$$

In the following, we assume that ξ is unrelated to other covariates - a strong restriction indeed. In general, the best linear approximation to ξ given the average of the observed

surrogates X_1 and X_2 , \bar{X} , is (see, for example, Carroll et al. (2006))

$$E(\xi|\bar{X}) \approx \mu_\xi + \Sigma_{\xi\xi} [\Sigma_{\xi\xi} + \Sigma_{\epsilon\epsilon}/2]^{-1} (\bar{X} - \mu_X), \quad (7.31)$$

where $\Sigma_{\xi\xi}$ denotes the $T \times T$ covariance matrix of ξ , $\Sigma_{\epsilon\epsilon}$ is a $T \times T$ measurement error covariance matrix, and μ_ξ and μ_X are the means of the variables ξ and \bar{X} , respectively. An estimated regression calibration function is then obtained by replacing the means and covariance matrices in (7.31) by the corresponding methods of moments calibration parameter estimators as describe, for example, in Carroll et al. (2006).

To be applicable in our case, we modify this general form, yielding

$$E(\xi|X) \approx \frac{1}{2} Z' U \theta + \Omega_{11} Z' \Sigma^{-1} Z \left[\frac{1}{2} Z' (X - U \theta) \right], \quad (7.32)$$

where X is the vector of all observed measurements, $Z = \begin{bmatrix} I_T \\ I_T \end{bmatrix}$, i. e. a matrix of dimension $2T \times T$, and Ω_{11} denotes a $T \times T$ sub-matrix of Ω . U and θ are defined in (7.28).

Estimates for the autoregressive parameters and variances involved in Σ and Ω_{11} are obtained by estimating the corresponding ARMA models for X_1 and X_2 , as described in the last chapter. Using these estimators, parameter estimates for the parameter vector θ , which consists of the trend parameters κ_0 and κ_1 as well as the device bias α_1 , are obtained by maximizing the log-likelihood function

$$l(\theta|X) = -(2T) \ln 2\pi - \frac{1}{2} \ln |\hat{\Sigma}| - \frac{1}{2} (X - U\theta)' \hat{\Sigma}^{-1} (X - U\theta),$$

which results in

$$\hat{\theta} = \left(U' \hat{\Sigma}^{-1} U \right)^{-1} U' \hat{\Sigma}^{-1} X.$$

The resulting estimated series $E(\xi|X)$ can then be used in the main model to evaluate the time-varying coefficients. Note, however, that this approach does not take into account the uncertainty due to estimation of the latent variable and, thus, confidence intervals for the exposure effects might not correctly reflect the true uncertainty in the data. To obtain adjusted standard errors, one would need asymptotic formulae for the standard errors or a bootstrap procedure; however, both methods are crucial in this case and will be topics of future research.

Chapter 8

Measurement error analysis of the Erfurt data

In this chapter, we again use the Erfurt data introduced in Section 5.1. In particular, we look at $\text{PM}_{2.5}$ data, for which we have measurements from two monitoring systems. We first describe the two measurement devices and present some descriptive statistics of the observed values. In the following, the estimated parameters and measurement error variances obtained by the simple estimation procedure as well as by using ARMA models are shown. Finally, we present the results of the application of a 'naive' correction. We further show the application results of the two proposed measurement error adjustment methods and compare the different approaches.

8.1 Description of the data sources

The mass of fine particles ($\text{PM}_{2.5}$) was measured using two different monitoring systems. Particle size distributions for fine particles were measured on a daily basis by an aerosol spectrometer (AS). As described elsewhere (Brand, Ruoff & Gebhart (1992), Wichmann et al. (2000)), it consists of two measurement devices, the differential mobility particle sizer (DMPS) and the optical laser aerosol spectrometer (LAS-X). The two instruments cover different size ranges.

Particles in the size range from 0.01 up to 0.5 μm are measured using the DMPS. This device consists of a differential mobility analyser (DMA) combined with a condensation particle counter (CPC). The DMA allows for a segregation of particle fractions of uniform electrical mobility from a polydisperse aerosol. The CPC counts the number of particles selected by the DMA in 13 discrete size ranges. Particles in the size range from 0.1 up to 2.5 μm are measured by LAS-X, which classifies particles according to their light

Table 8.1: Summary statistics of daily mortality, air pollution concentrations and key meteorology for the period March 10, 1996 to March 31, 2000.

Variable	Mean (SD)	Min	25%	Median	75 %	Max
Daily counts of death	4.89 (2.2)	0	3	5	6	13
PM _{2.5} AS ($\mu\text{g}/\text{m}^3$)	21.5 (17.2)	1.6	10.4	15.9	27.0	143.6
PM _{2.5} HI ($\mu\text{g}/\text{m}^3$)	23.3 (17.5)	2.6	11.8	18.2	29.0	197.0
Temperature (°C)	8.4 (7.4)	-16.9	2.9	8.5	14.3	26.5
Relative humidity (%)	80.5 (11.1)	34.3	73.8	81.7	88.7	99.3

scattering properties into 45 size-dependent channels.

Particle number distributions are connected to particle mass distributions by assuming spherical particles with an average density of 1.5. The integral of the particle mass distributions between 10 nm to 2.5 μm is a good measure for PM_{2.5} (Wichmann et al. (2000)) and is called PM_{2.5}AS.

The second measurement device was a PM_{2.5} Harvard Impactor (HI) (Marple et al. (1987)). The measurement principle of the HI is based on gravimetric analysis. The sampler consists of an inlet, an impaction plate, and filter mounted in a plastic holder. The concentration of particles is determined from the calculated mass change on a filter by weighing under controlled temperature and relative humidity conditions, taking the total volume of air sampled (at local temperature and pressure) into account. Before weighing, the filters are equilibrated in a temperature and relative humidity controlled weighing room for at least 24 hours.

In this example, regular measurements for both devices are available for the period of March 10, 1996 to March 31, 2000. Time series as well autocorrelation functions plots of PM_{2.5}AS_t and PM_{2.5}HI_t for the this time period are given in Figure 8.1.

Summary statistics of both PM_{2.5} time series as well as of daily counts of death and key meteorology for the respective time period are given in Table 8.1. It can be seen that the mean concentration obtained by the HI is slightly higher compared to those of the AS. Moreover, the range of the HI measurements is wider.

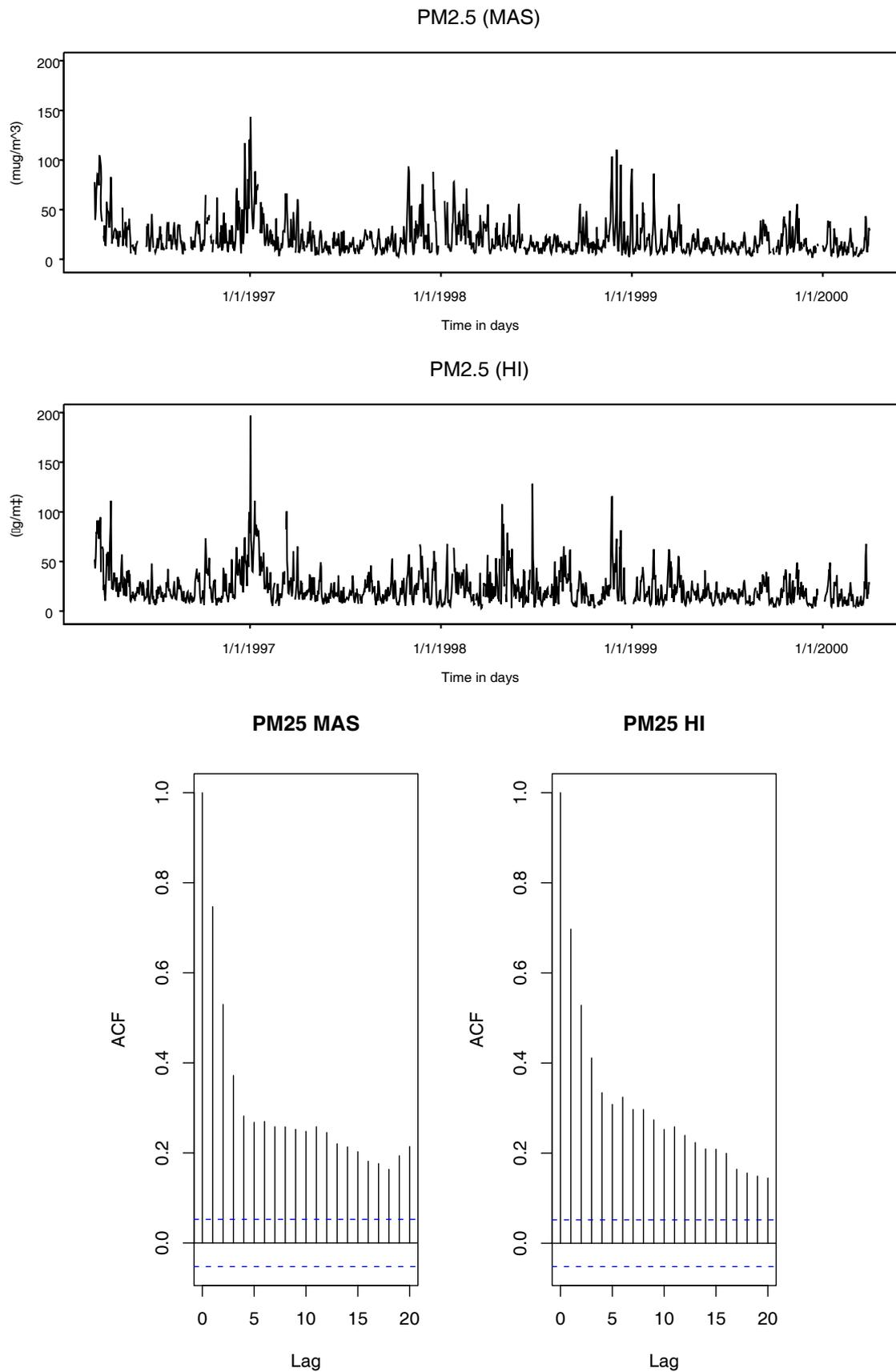


Figure 8.1: Time series of the two PM_{2.5} measurement devices (top) and the corresponding autocorrelation functions (bottom).

8.2 Application of the measurement error models

We assume that all measurements $PM_{2.5}AS_t$ and $PM_{2.5}HI_t$ are perturbed version of ξ_t , which represents the true $PM_{2.5}$ value at time point t .

The disease model used in the following applications is given by

$$\begin{aligned} y_t &\sim \text{Poisson}(\mu_t) \\ \log(\mu_t) &= \log(\text{pop}) + \text{confounder} + f_{PM_{2.5}}(t) PM_{2.5}, \end{aligned} \quad (8.1)$$

where $f_{PM_{2.5}}(t)$ is modeled using a second order random walk. Correspondingly to Chapter 5, we consider same-day $PM_{2.5}$ mass concentrations.

The confounder model used in this application is slightly different to that for the period 1995 to 2002 (see Section 5.1). It is given by

$$\begin{aligned} \log(\mu_t) &= \log(\text{pop}) + f(t) + pi9899_{l2} + pi9900_{l2} + so \\ &+ f(\text{temp}_{l0}) + \text{temp}_{l1} + f(\text{rh}_{l2}). \end{aligned} \quad (8.2)$$

Again, cubic P-splines are estimated for calendar time (t), and for the meteorological confounders temperature with lag 0 days (temp_{l0}) and relative humidity, lag 2 days (rh_{l2}). Adjustment for influenza is done by including the doctor's practice indices of the winters 1998/1999 and 1999/2000.

In the following, we present the results of estimating the measurement error model parameters and error variances. First, the simple estimation procedure based on the mean of the two measured series is used. Thereafter, the results of the application of the ARMA models are shown. To estimate the different ARMA models, we used the `ARIMA` function of the `ts` package in *R*, as suggested by Staudenmayer & Buonaccorsi (2005). The parameters were obtained using maximum likelihood estimation. More specifically, a Kalman filter was used to compute the Gaussian likelihood and its partial derivatives, and ML estimates were obtained using quasi-Newton methods.

Finally, we present the results of the application of the considered correcting methods in the context of Bayesian time-varying coefficient models. For all models, the first 50,000 iterations were discarded as burn-in. Further 200,000 iterations were then completed for the chain, which were thinned by 200 to reduce autocorrelation (see Fahrmeir & Lang (2001)).

Table 8.2: Estimated parameters using the simple estimation procedure.

Parameter	Estimate
$E(x_m)$	21.84
$V(x_m)$	240.56
$V(x_1)$	258.83
$V(x_2)$	287.82
σ_ξ^2	207.87
$\sigma_{\epsilon_1}^2$	50.95
$\sigma_{\epsilon_2}^2$	79.95

8.2.1 Estimating the model parameters and measurement error variances

Applying the method of moments using the mean, $X_m = (\text{PM}_{2.5} \text{ AS} + \text{PM}_{2.5} \text{ HI})/2$, of the two observed variables, and

$$\begin{aligned} E(X_m) &= \mu \\ \text{Var}(X_m) &= \sigma_\xi^2 + \frac{1}{4}(\sigma_{\epsilon_1}^2 + \sigma_{\epsilon_2}^2) \\ \text{Var}(X_1) &= \sigma_\xi^2 + \sigma_{\epsilon_1}^2 \\ \text{Var}(X_2) &= \sigma_\xi^2 + \sigma_{\epsilon_2}^2, \end{aligned}$$

as described in Section 6.3.1, we get parameter estimates as given in Table 8.2. However, an inspection of the correlation between consecutive observations of the averaged time series X_m reveals that there is a considerable amount of autocorrelation in the data ($\rho \approx 0.74$) which should be accounted for.

Next, we present results of the application of ARMA models of order (1,1) to both measured $\text{PM}_{2.5}$ series (see also Section 6.3.2.1). The models are given by

$$(1 - \rho_{ARMA_i} B) X_{ti} = (1 + \theta_{1i} B) \gamma_{ti},$$

where X_1 corresponds to the $\text{PM}_{2.5}$ AS series and X_2 to the $\text{PM}_{2.5}$ HI series. The results of these analyses are shown in Table 8.3. Both models show a considerable amount of autocorrelation, ranging between 0.69 and 0.75. Using the obtained parameters $\hat{\rho}_{ARMA_i}$, $\hat{\theta}_{1i}$, and $\hat{\sigma}_{\gamma_i}^2$, we can solve the equations

$$\hat{\sigma}_{\gamma_i}^2 \hat{\theta}_{1i} = -\sigma_{\epsilon_i}^2 \hat{\rho}_{ARMA_i}$$

to get an estimate for the measurement error variances $\sigma_{\epsilon_i}^2$, and

$$\sigma_{\gamma_i}^2 (1 + \theta_{1i}^2) = \sigma_{\epsilon_i}^2 + \sigma_{\epsilon_i}^2 (1 + \rho_{ARMA,i}^2)$$

for obtaining an estimated variance $\hat{\sigma}_{e_i}^2$ of the autoregressive process for ξ , shown in Table 8.3. A comparison of the estimated measurement error variances obtained by the ARMA(1,1) models to those of the simple approach shows that not accounting for autocorrelation yields much larger variances. Estimated intercept and linear trend coefficients of the true exposure process are nearly equal for the three models. In Figure 8.2, the autocorrelation functions of standardized residuals obtained by the ARMA(1,1) models are shown for $PM_{2.5}AS$ and $PM_{2.5}HI$. The ACF functions show that there is nearly no autocorrelation left. The small peaks at lags 6 and 11 for both $PM_{2.5}AS$ and $PM_{2.5}HI$ may be due to a day-of-week pattern.

As a sensitivity analysis, we further explore the difference of the two observed $PM_{2.5}$ series

$$PM_{2.5}AS - PM_{2.5}HI.$$

By using this difference, we are able to investigate whether the observed autocorrelation is only caused by the latent true values. Computing the ACF for the difference series, however, reveals that there is still correlation between consecutive observations, and investigating the pattern of autocorrelation indicates again an autoregressive structure with a value of 0.38 for the first-order autocorrelation. This means that there is autocorrelation within the measurement errors, and ARMA(2,1) representations might be more appropriate.

Such ARMA models for X_i , $i = 1, 2$ are given as

$$(1 - \rho_{1i}B - \rho_{2i}B^2)X_{ti} = (1 + \theta_{1i}B)\gamma_{ti},$$

where γ_{ti} is white noise. Further details can be found in Section 6.3.2.2.

The results of applying ARMA(2,1) models are given in Table 8.4. As can be shown, the maximum likelihood estimators ρ_{1i} and ρ_{2i} satisfy the conditions $\rho_{1i}^2 + 4\rho_{2i} > 0$, $\rho_{1i} + \rho_{2i} < 1$ and $\rho_{2i} - \rho_{1i} < 1$. These conditions are necessary for an ARMA(2,1) process to be stationary and the roots of $1 - \rho_{1i}z - \rho_{2i}z^2 = 0$ to be real. This, however, means that the corresponding solutions ϕ_i and π_i are real and maximum likelihood estimates. We can further show that the condition of uniqueness $-(\theta_{1i}^2 - \rho_{1i}\theta_{1i} - \rho_{2i}) \neq 0$ is satisfied. In contrast to the ARMA(1,1) models, the autoregressive parameters have a greater range, though, they show again a considerable amount of autocorrelation in the series.

Table 8.3: Estimated parameters using an ARMA(1,1) model.

	PM _{2.5} AS	PM _{2.5} HI
ρ_{ARMA}	0.69	0.75
θ_1	0.11	0.12
σ_γ^2	123.40	153.6
σ_ε^2	19.67	24.57
σ_e^2	95.86	117.42
κ_0	46.96	46.26
κ_1	-0.01	-0.01
AIC	10716.69	11270.14
BIC	10678.39	11296.65

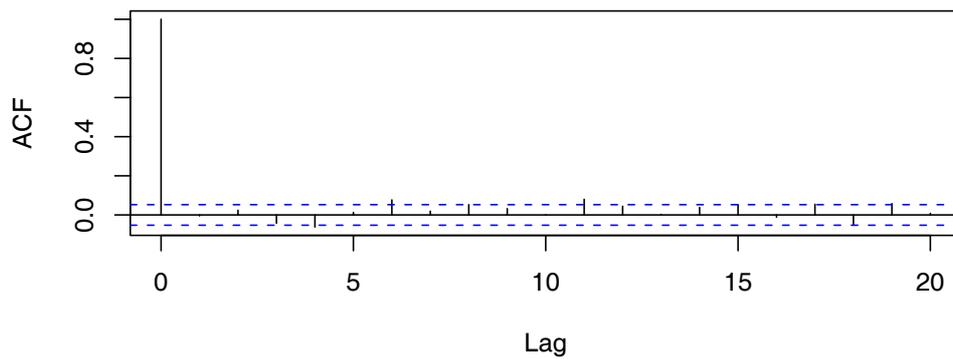
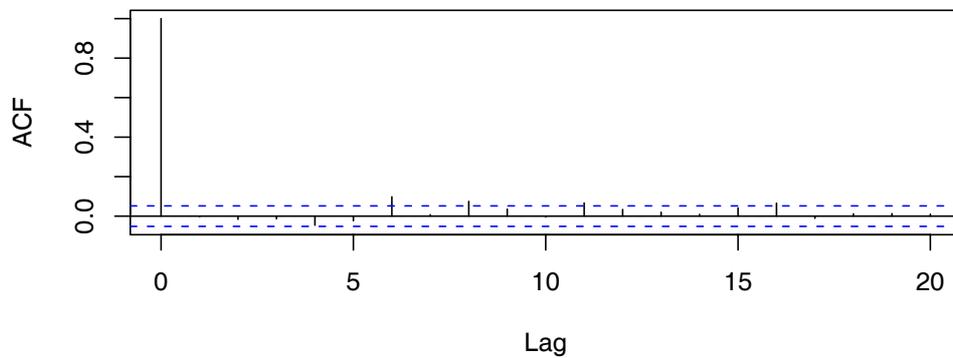
ACF of standardized residuals of an ARMA(1,1) model for PM25 MAS**ACF of standardized residuals of an ARMA(1,1) model for PM25 HI**

Figure 8.2: Autocorrelation functions (ACF) of the standardized residuals obtained by ARMA(1,1) models.

Using the obtained parameters $\hat{\rho}_{1i}$ and $\hat{\rho}_{2i}$, we can solve for the autoregressive parameters of the latent ξ process as well as the measurement error processes

$$\rho_{1i} = \phi_i + \pi_i \quad \text{and} \quad \rho_{2i} = -\phi_i \pi_i,$$

which implies

$$\begin{aligned} \phi_i^2 - \rho_{1i} \phi_i - \rho_{2i} &= 0 \quad \text{and} \\ \pi_i^2 - \rho_{1i} \pi_i - \rho_{2i} &= 0. \end{aligned}$$

Using the estimates $\hat{\phi}_i$, $\hat{\pi}_i$, θ_{1i} and $\hat{\sigma}_{\gamma_i}^2$, one can then solve

$$\sigma_{e_i}^2 = \frac{-\sigma_{\gamma_i}^2 \theta_{1i} - \phi_i \sigma_{\epsilon_i}^2}{\pi_i},$$

for $\sigma_{e_i}^2$. Substitution of the obtained equation into equation

$$\sigma_{\gamma_i}^2 (1 + \theta_{1i}^2) = \sigma_{e_i}^2 (1 + \pi_i^2) + \sigma_{\epsilon_i}^2 (1 + \phi_i^2)$$

finally leads to estimates $\hat{\sigma}_{e_i}^2$, as shown in Table 8.4. The estimated measurement error variances are now larger, whereas the estimated variances of the true values are smaller. Estimated intercept and linear trend coefficients of the true exposure process are comparable for all models. The autocorrelation functions of standardized residuals obtained by the ARMA(2,1) models are shown in Figure 8.3. Comparable to the ARMA(1,1) models, the ACF plots show some slight autocorrelation for lags 6 and 11. Overall, however, adjusting for autocorrelation by the ARMA(2,1) models seems to work very well. This is corroborated by the inspection of the model selection criteria AIC and BIC which reveals that the AIC and BIC values for the ARMA(2,1) models are smaller compared to those obtained for the ARMA(1,1) models. Further, t-tests are performed which show the statistical significance of the additional autoregressive parameters of the ARMA(2,1) models (with p-values of 0.05 and 0.001 for the $PM_{2.5}$ AS and the HI series, respectively).

8.2.2 Results of the correction methods - time-varying effect of $PM_{2.5}$

The top left panel of Figure 8.4 shows the time-varying coefficients of the imputed $PM_{2.5}$ series as described in Chapter 5. They correspond to 'naive' estimates that ignore measurement error. Further, the time-varying coefficients obtained by using the mean of the two observed $PM_{2.5}$ series are shown (top right panel).

Table 8.4: Estimated parameters using an ARMA(2,1) model.

	PM _{2.5} AS	PM _{2.5} HI
ρ_1	0.87	1.49
ρ_2	-0.14	-0.51
θ_1	0.06	0.87
ϕ	0.67	0.95
π	0.21	0.54
σ_γ^2	123.30	151.1
σ_ϵ^2	46.20	132.77
σ_e^2	52.11	50.15
κ_0	46.87	48.37
κ_1	-0.01	-0.01
AIC	10641.27	11249.25
BIC	10678.39	11281.07

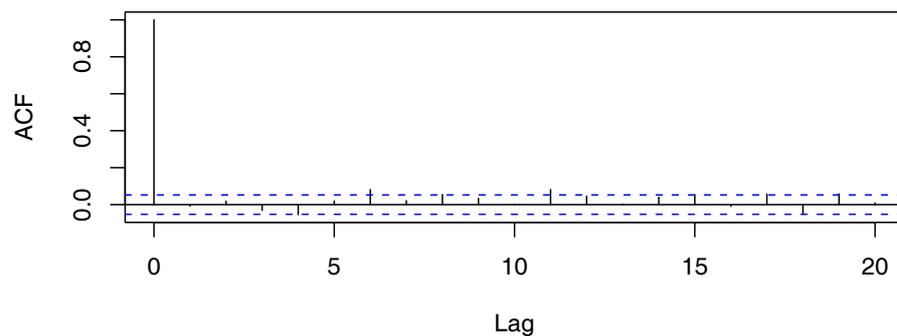
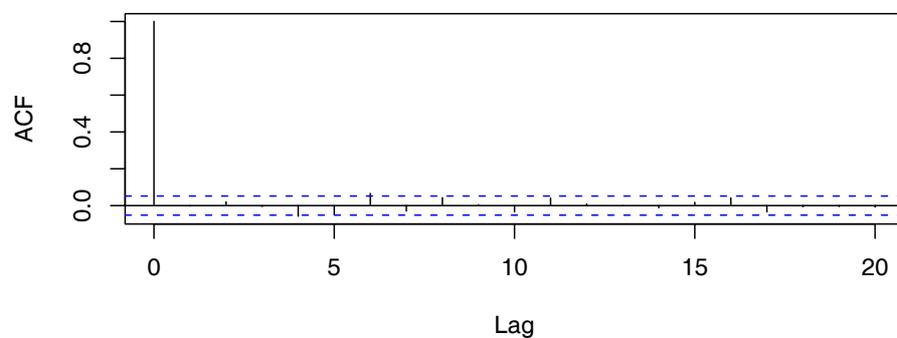
ACF of standardized residuals of an ARMA(2,1) model for PM25 MAS**ACF of standardized residuals of an ARMA(2,1) model for PM25 HI**

Figure 8.3: Autocorrelation functions (ACF) of the standardized residuals obtained by ARMA(2,1) models.

We see a slightly decreasing time-varying effect of $PM_{2.5}$ for the imputed series, whereas for the mean series a clear decreasing time-varying effect is obtained. Further, the estimated coefficients have a larger range and point to some significant effects at the beginning and the end of the specific period.

The results of the application of the two correction methods are shown in Figure 8.4, middle and bottom rows. Note that, as only one autoregressive parameter $\rho_{\xi,1}$ or ϕ_1 in the measurement error correction methods is needed, we always use the mean of the two estimated autoregressive parameters obtained for each observed series.

Figure 8.4, middle row, provides the results assuming an underlying ARMA(1,1) model for the observed series, that means the true values are supposed to follow an AR(1) process, whereas the measurement errors are white noise. Both methods provide a decreasing effect until the year 2000. Compared to the time-varying coefficients obtained by the naive analysis and using the mean, we now observe stronger effects, where those obtained based on the regression calibration approach even point to a significant association. The Bayesian correction method shows wider credible intervals. This is mainly because this approach takes into account uncertainty due to estimating the true exposure, whereas the credible intervals obtained for the estimates based on the regression calibration function do not and, hence, can only communicate a rough impression of significance for the coefficients. Furthermore, the time-varying curve of the Bayesian correction approach shows a wigglier shape.

Figure 8.4, bottom row, shows the results of the measurement error adjustment methods assuming underlying ARMA(2,1) models for the observed series. In this case, both the true values and the measurement error are assumed to follow AR(1) processes. Compared to the results of the ARMA(1,1) case, the same functional forms of the coefficients can be observed. Moreover, an even more pronounced effect can be seen for both correction methods. The estimated time-varying coefficients obtained by the Bayesian correction approach again show wider credible intervals compared to the regression calibration method.

8.2.3 Conclusion

Overall, we conclude from the results that not accounting for autocorrelation within the time series, leads to biased results. Based on the sensitivity analysis of the difference, we prefer the results obtained by assuming an underlying ARMA(2,1) model. Furthermore, the Bayesian correction model is a more conservative method as it takes key sources of

uncertainty in the time-varying estimate into account. Results of the correction methods point to an association between $PM_{2.5}$ and mortality for the period 1996 to 1998, which was also found in a previous analyses of the Erfurt data (Wichmann et al. (2000)). The results further indicate that the estimates obtained without measurement correction are biased towards the null, an observation already established for other applications in the context of air pollution time series data (see Dominici et al. (2000)).

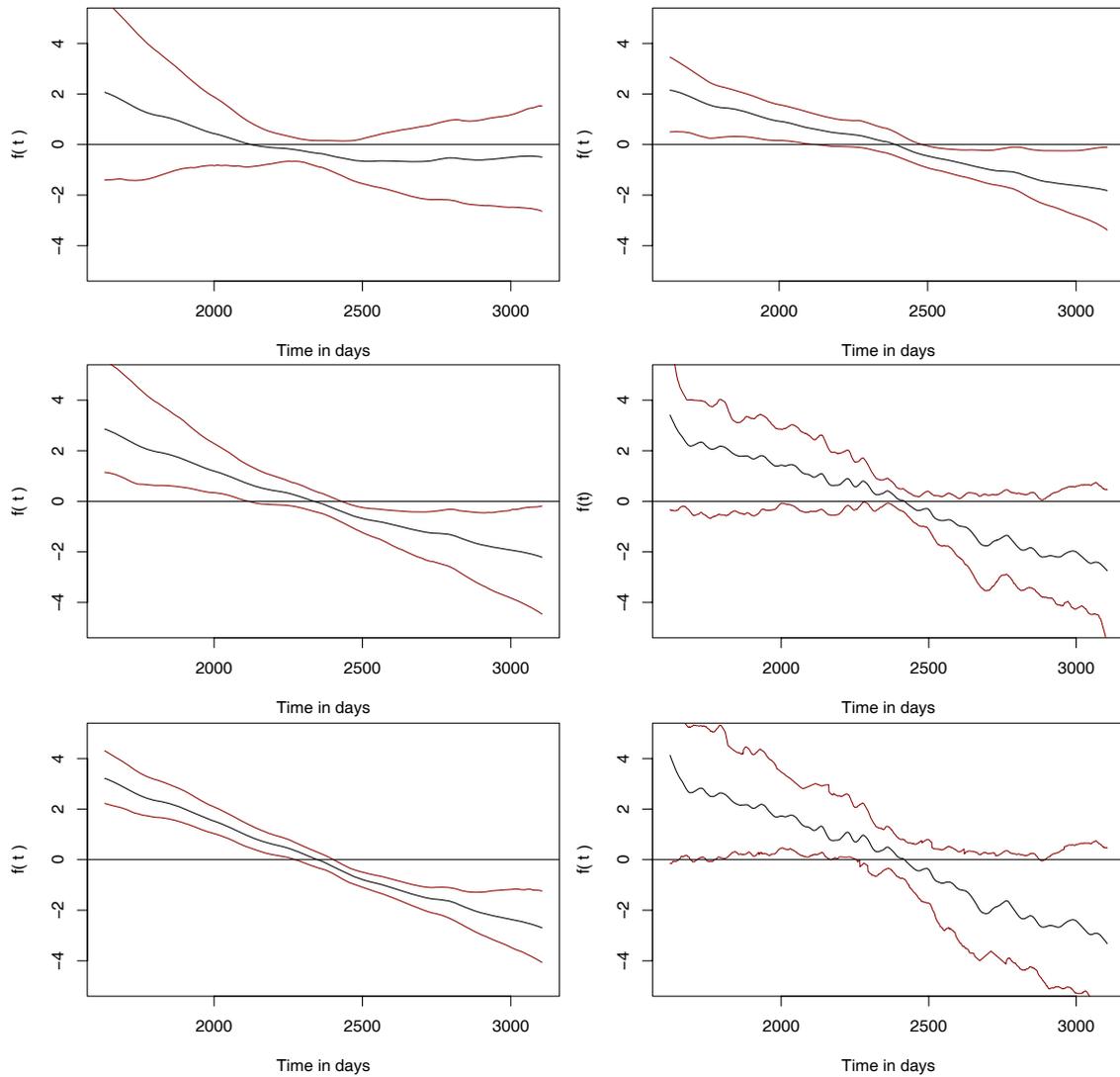


Figure 8.4: Time-varying coefficients of $PM_{2.5}$ series. Top row: 'Naive' time-varying coefficients of the imputed $PM_{2.5}$ series (left panel) and of the mean of the two observed $PM_{2.5}$ series (right panel). Middle row: Time-varying coefficients of $PM_{2.5}$ obtained by the (Bayesian) regression calibration model (left) and by the Bayesian measurement error model (right). Underlying assumption of ARMA(1,1) models. Bottom row: Time-varying coefficients of $PM_{2.5}$ obtained by the (Bayesian) regression calibration model (left) and by the Bayesian measurement error model (right). Underlying assumption of ARMA(2,1) models.

Chapter 9

Summary and Outlook

In this thesis, we first presented and developed five approaches for estimating time-varying coefficients. All methods allow for a smoothly time-varying effect. The five approaches were: Time-varying coefficient models using regression splines; Bayesian time-varying coefficient models with P-splines; Time-varying coefficient models with penalized linear splines based on a generalized linear mixed model framework; Time-varying coefficient models with P-splines based on empirical Bayes inference, and Adaptive generalized TVC models. In addition to the theoretical development of the models, a simulation study was conducted to assess the performance of the approaches. These comparisons revealed that the fully Bayesian time-varying coefficient models using P-splines and time-varying coefficient models with P-splines based on empirical Bayes inference performed best. These approaches performed quite well in terms of root mean squared errors, but also in terms of bias. They outperformed especially the approach using local likelihood smoothing, but also the two approaches based on truncated power series bases. In most of the simulation cases as well as in the analysis of the Erfurt data, both Bayesian approaches yielded estimates which were quite close to each other. The regression results pointed to a time-varying association between daily mortality and air pollution. We further presented a test procedure for investigating whether a time-varying coefficient term is indeed required. This test was based on a likelihood ratio test statistics and performed by using bootstrapping methods. The procedure can be seen as an extension of the test procedure proposed by Fahrmeir & Mayer (2001).

Regarding the analysis of the Erfurt data, we are aware of some severe limitations in interpreting trends and variation of the short-term effects of air pollutants. These include power limitations caused by the small population size in a town such as Erfurt. Due to only five deaths per day on average, power for statistical analyses is limited and therefore, the assessment of time-varying air pollution effects is a challenging task. Because of the

limited power of the study, we also did not extend the varying coefficient models to interactions with higher dimensions. However, a straightforward extension of the Bayesian P-splines to the two-dimensional case can be obtained by using bivariate P-splines as proposed by Lang & Brezger (2004) and Brezger & Lang (2006) for fully Bayesian models and by Kneib (2006) in the context of empirical Bayes inference. An unknown surface is then approximated by the tensor product of two (univariate) B-splines. In analogy to the univariate case, two-dimensional random walk priors are further assigned to enforce smoothness of the estimated surfaces.

In the second part of this thesis we approached the problem of estimating time-varying coefficient models in the presence of measurement errors in the interacting variable. As time series are usually assumed to be temporally correlated, particular attention was given to by an appropriate selection of the latent variable's distribution. In addition, we presented methods for estimating the measurement error model parameters in this special context. These methods are based on an ARMA representation of measurement error model. We further developed two methods adjusting for measurement errors in the context of time-varying coefficients. One is a hierarchical Bayesian model based on a MCMC correction. In this approach, we achieved a corrected time-varying estimate by including an additional sampling step, the sampling of the latent variable values. The second approach is an extension of the well-known regression calibration to the case of autocorrelated data. The obtained estimated true values can then be included into the main model to assess the effect of the variable of interest.

In the future, application of time-varying coefficient models could be of great importance, especially in case studies related to the evaluation of the public health impact of environmental regulations. However, there are still questions concerning model validation which will require additional attention. First of all, power and sample size considerations within the time-varying effects framework might be a beneficial addition. The sample size determination for studies intending the application of such models is challenging and has, to our knowledge, not been adequately addressed in the literature. In general, the calculation of the optimal sample size has to be based on simulations.

As only limited information can be obtained from a single city and it is difficult to clearly identify temporal confounders of risk, multiple city temporal risk models would be a valuable extension of the proposed approaches. Such a model would be based on larger amounts of data and, therefore, provide a wider basis of evidence. The extension could be done by adding a random effects term, which would fit naturally within Bayesian hierarchical model context. Bayesian multi-city models estimating time-constant effects are proposed by Samet et al. (2000a) and Dominici et al. (2002), for example. If there

was evidence of correlation between the effects from cities in close proximity, then the covariance structure could be formulated to incorporate a spatial structure.

Furthermore, a detailed evaluation of the properties and the power of the proposed test procedure is required. In particular, an extensive simulation study designed to assess the validity of the bootstrap-based test is needed. Additionally, the test procedure could be extended with respect to different robust adjustments for residual bootstrapping. Kauermann, Claeskens & Opsomer (2006), for example, present such adjustments for tests on generalized additive models. In the context of smoothing and mixed models, likelihood ratio tests on variance components have been developed throughout the last years. However, these tests either rely on the assumption of a Gaussian distributed response (Crainiceanu & Ruppert (2004), Crainiceanu et al. (2005)) or consider only one variance component (Crainiceanu & Ruppert (2004)), which might not be adequate in rather complex models as considered in this thesis. Within a fully Bayesian framework, the question whether nonparametric modeling is really necessary can be answered by the use of simultaneous contour probabilities based on highest posterior density regions, as proposed by Brezger (2005) for generalized additive models. An extension of these models might be a valuable enhancement in the context of time-varying coefficient models.

Besides power assessments of testing procedures, further developments could aim at the generalization of the Bayesian measurement error correction method proposed in this work. For example, a further extension could be given by replacing the assumption of a random walk for the unknown time-varying effect of the latent variable by Bayesian P-splines. As for the models without measurement error corrections, this should give smoother time-varying estimates.

A more challenging task is the question of the specification of the prior distribution of the unknown covariate. In structural measurement error problems, there is a general concern about the parametric specification of the unknown covariates. Richardson et al. (2002), therefore, present a flexible semiparametric model for this distribution based on a mixture of normal distributions with an unknown number of components. Using mixtures of normal distributions increases the robustness to model miss-specification. However, the application of such mixtures in our models is not straightforward, since computation in large sample cases might become vastly complicated and very time-consuming.

Further generalizations of the Bayesian measurement error correction methods could include models accounting for autocorrelation in the latent variable in combination with those accounting for exposure error resulting from using centrally measured data as a surrogate for personal exposure. One possibility would be an extension with an additional

model using external information about personal exposure as proposed by Dominici et al. (2000). A further correction approach is based on the use of spatio-temporal models to properly specify the relationship between observed and true underlying exposure levels. There is growing literature on modeling, for example, air pollution data in time and space. Because of the complexity in general spatio-temporal models, interest has focused on models which are separable over time and space (Gelfand, Zhu & Carlin (2001)). This subclass of spatial temporal processes has several advantages, including rapid fitting and simple extensions of many techniques developed and successfully used in time series and classical geostatistics. Among many others, Tonellato (2001) presents a Bayesian dynamic modeling approach for hourly measurements of CO obtained from a small number of sites. A random walk process measured with error is used. The spatial dependence is modeled by a heteroscedastic Gaussian spatial process. A similar approach is proposed by Shaddick & Wakefield (2002). They also use a hierarchical dynamic linear model with a slightly different spatial covariance structure to model data on four different pollutants measured at eight monitoring sites in London. A combination of their approaches with the Bayesian measurement error model considered in this work would allow for both the modeling of appropriate exposure data and assessing a corrected health-exposure relationship.

Chapter A

Time-varying coefficient models and measurement error analysis with BayesX

The focus of this chapter is on demonstrating how air pollution data with measurement errors can be analyzed in the program package *BayesX*.

For this purpose, a description of the general use of BayesX and some comments about its structure are given first. We then describe how some of the results presented in Chapter 5 can be obtained. Further, we show how to estimate time-varying coefficient models that allow for measurement error in the pollution variable. This chapter concludes with a description of methods that are implemented in *BayesX* to plot and further explore regression results.

BayesX is a free software package available at

<http://www.stat.uni-muenchen.de/~bayesx>.

It allows for performing complex full and empirical Bayesian inference to estimate generalized structured additive regression models. Functions for handling and manipulating data sets as well as for the visualization of results are added for convenient use. For an overview over the capabilities of *BayesX* and a detailed description of all available features we refer to Brezger et al. (2005) and the manuals provided in addition.

A.1 Getting started

After having started *BayesX*, a main window consisting of four sub-windows appears on the screen. The sub-windows are a *command window*, where commands can be entered

and executed, an *output window* for displaying results, a *review window* allowing for an easy access to past commands, and, finally, an object browser, where all objects currently available are displayed.

To estimate Bayesian regression models we need at least a dataset object which allows for incorporating, handling and manipulating of data, a bayesreg or a remlreg object to estimate the regression models, and a graph object to be able to visualize the estimated results.

The syntax for generating a new object in *BayesX* is

```
> objecttype objectname
```

where **objecttype** is the type of the object, for example `dataset`, and **objectname** is the name to be given to the new generated object.

A.2 Dataset objects

In a first step the available data set information is loaded into *BayesX*. This is done by creating a dataset object named *erfurt* by typing

```
> dataset erfurt
```

in the *command window* and then using the method `infile`

```
> erfurt.infile [, options] using path\ filename
```

Note, that this command supposes that the variable names are given in the first row of the corresponding external file. If this is not the case, we would have to supply them right after the keyword `infile`. For large data sets, that means with more than 10000 observations, it is recommended to use the option `maxobs` to speed up the execution time.

After having read in the data, we can inspect the data by double-clicking on the respective object in the object-browser. This can also be achieved by executing the `describe` command

```
> erfurt.describe
```

Further commands for handling and manipulating the data are described in Brezger et al. (2005). Examples are the `descriptive` command to obtain summary statistics for the variables or the `generate` command which allows the creation of new variables.

A.3 Bayesreg and remlreg objects

Time-varying coefficient models can be estimated based on MCMC inference using the `regress` command of `bayesreg` objects. These models can further be estimated by methods based on mixed model representations using the `regress` command of `remlreg` objects. The general syntax for both types of objects is

```
> objecttype objectname
> objectname.regress model [, options] using erfurt
```

Executing these commands first creates a `bayesreg` or `remlreg` object and then estimates the regression model specified in `model` using `erfurt`, which is the dataset object created previously. By default, estimation results are written to the sub-directory `output` of the installation directory. The default filenames are then composed of the name of the object and the type of the specific file. Usually, however, it is more convenient to specify a directory where the results are stored. This can be done by using the command

```
> objectname.outfile = path\\outfilename.
```

Execution of this command leads to the storage of the results in the directory specified in 'path' and all generated filenames start with the characters 'outfilename'.

Hence, for example, a fully Bayesian time-varying coefficient model for CO with a lag of four days, which was specified by (see Chapter 5)

$$\begin{aligned} \log(\mu_t) = & \log(pop) + f(t) + f(pi9394_{l0}) + f(pi9495_{l-3}) + f(pi9596_{l-1}) + f(pi9900_{l2}) \\ & + f(pi0001_{l3}) + so + f(temp_{l0}) + temp_{l1} + f(rh_{l2}) + f(t) CO_{l4}, \end{aligned} \quad (A.1)$$

can be estimated by executing

```
> bayesreg erf
> erf.outfile = c:\temp\co_mcmc
> erf.regress mortagnk = logpop(offset) + tag(psplinerw2)
+ pi9394(psplinerw2) + pi9495lm3(psplinerw2) + pi9596lm1(psplinerw2)
+ pi9900l2(psplinerw2) + pi0001l3(psplinerw2) + so + temp_dwd(psplinerw2)
+ temp_dwdl1 + rh_dwdl2(psplinerw2) + CO_Krl4*tag(psplinerw2),
iterations=60000 step=40 burnin=10000 family=poisson predict using erfurt
```

The effects of the unknown functions are modeled by P-splines with a second order random walk prior specified by `psplinerw2`. By default, the degree of a spline is 3 and the number of inner knots is 20. Options `iterations`, `burnin` and `step` define properties of the MCMC-algorithm. In particular, `iterations` is used to specify the total number

of MCMC iterations, while the number of burn-in iterations is defined by `burnin`. Using the above specifications hence leads to a sample of 50000 random numbers. However, as these random numbers are, in general, correlated, we thin out the Markov chain by using the thinning parameter `step`. Specifying `step = 40`, for example, forces *BayesX* to store only every 40-th sampled parameter leading to a random sample of length 1250 for every parameter in our example. If additionally, the option `predict` is specified, then samples of the unstandardized deviance, the effective number of parameters, and the deviance information criterion DIC of the model are computed, see Spiegelhalter et al. (2002). Furthermore, estimates for the linear predictor as well as the expectation of every observation are obtained.

The influence of different choices for hyperpriors can be investigated by specifying different parameters a and b for the $IG(a, b)$ priors. The following command may for example be used to specify $a = 1$ and $b = 0.005$:

```
> erf.regress mortagnk = logpop(offset)
+ tag(psplinerw2,a=1,b=0.005) + pi9394(psplinerw2,a=1,b=0.005)
+ pi9495lm3(psplinerw2,a=1,b=0.005) + pi9596lm1(psplinerw2,a=1,b=0.005)
+ pi9900l2(psplinerw2,a=1,b=0.005) + pi0001l3(psplinerw2,a=1,b=0.005)
+ so + temp_dwd(psplinerw2,a=1,b=0.005) + temp_dwdl1
+ rh_dwdl2(psplinerw2,a=1,b=0.005) + CO_Krl4*tag(psplinerw2,a=1,b=0.005),
iterations=60000 step=40 burnin=10000 family=poisson predict using erfurt
```

Note that in the case that no parameters for a and b are specified, these parameters are set to the default values $a = b = 0.001$.

For estimation of model (A.1) using the mixed model methodology, we enter

```
> remlreg erf1
> erf1.outfile = c:\temp\co_reml
> erf1.regress mortagnk = logpop(offset) + tag(psplinerw2)
+ pi9394(psplinerw2) + pi9495lm3(psplinerw2) + pi9596lm1(psplinerw2)
+ pi9900l2(psplinerw2) + pi0001l3(psplinerw2) + so + temp_dwd(psplinerw2)
+ temp_dwdl1 + rh_dwdl2(psplinerw2) + CO_Krl4*tag(psplinerw2),
family=poisson using erfurt
```

By default, *BayesX* produces external ASCII files containing the posterior means or modes obtained by `bayesreg` or `remlreg`, respectively. The names of these files are given in the *output window*. The files further contain the 80% and 95% credible intervals, and the corresponding 80% and 95% posterior probabilities of the estimated effects. The

levels of the credible intervals and posterior probabilities may be changed by the user using the options `level1` and `level2`. For example, specifying `level1=99` in the option list of the `regress` command leads to the computation of 99% credible intervals and posterior probabilities.

A.4 Time-varying coefficients and measurement error

To estimate the fully Bayesian time-varying coefficient models with a correction for measurement error in the $PM_{2.5}$ series presented in Section 8.2.2, we again start by creating a `bayesreg` object by typing

```
> bayesreg mem
```

We could also use the existing `bayesreg` object `erf`, however, for clarity, we prefer to create a new object named `mem`.

The predictor of the measurement error correction model based on an underlying assumption of an ARMA(1,1) representation of the observed surrogates is given by

$$\begin{aligned} \log(\mu_t) &= \log(\text{pop}) + f(t) + \text{pi9899}_{l2} + \text{pi9900}_{l2} + \text{so} \\ &+ f(\text{temp}_{l0}) + \text{temp}_{l1} + f(\text{rh}_{l2}) + f_{PM_{2.5}}(t) PM_{2.5}, \end{aligned}$$

To estimate this model, we execute the following commands

```
> mem.outfile = c:\temp\pm25
> mem.regress mortngnk = logpop(offset) + tag(psplinerw2)
+ pi9899l2 + pi9900l2 + so + temp_dwd(psplinerw2) + temp_dwdl1
+ rh_dwdl2(psplinerw2) + x_merror*tag(merrorrw2,proposal=cp,min=1,max=80,
  rhoxi = 0.72, sig1 = 19.76, sig2 = 24.57, sig3 = 138),
iterations=250000 step=200 burnin=50000 maxint=1500 family=poisson
predict using erfurt
```

Note that the only difference to the estimation of the fully Bayesian models presented in the previous section is the term

```
x_merror*tag(merrorrw2,proposal=cp,min=1,max=80,rhoxi=0.72,sig1=19.76,
sig2=24.57,sig3=138)
```

The term `x_merror` specifies the object which is used for the sampling of the true $PM_{2.5}$ values. Note that the two observed variables should be denoted as `x1` and `x2` in the

dataset. The unknown time-varying effect for the sampled $PM_{2.5}$ values is modeled as a second order random walk specified as `merrorrw2`. Updating is done by using the conditional prior proposal which is specified by `proposal=cp`. Minimum and maximum block sizes are defined by the `min` and `max` option. The options `rhoxi`, `sig1`, `sig2` and `sig` specify additional parameters needed for the sampling of the true $PM_{2.5}$ values. They correspond to the autoregressive parameter $\rho_{\xi,1}$, the measurement error variances σ_1^2 and σ_2^2 , and the variance of the autoregressive process σ_e^2 .

Again, additionally to the information being printed to the *output window*, results for each effect are written to external ASCII files, with the names of these files being given in the *output window*.

A.5 Post-estimation commands and visualization of the results

A.5.1 Post-estimation commands

In addition to the `regress` command, `bayesreg` objects provide some post estimation commands to get sampled parameters or to compute autocorrelation functions of sampled parameters. These may be obtained by

```
> mem.outfile = c:\temp\pm25sample  
> mem.getsample
```

and

```
> mem.plotautocor, maxlag = 150
```

respectively, where `maxlag` specifies the maximum lag number (which is by default 250).

A.5.2 Visualization of the results

BayesX automatically creates appropriate plots of some of the estimated effects and stores the graphs as postscript files. The file names are given in the *output window* for each effect. Moreover, a batch-file is created that contains all commands necessary to reproduce the plots. The advantage is that additional options may be added by the user to customize the graphs (for example to change the title or axis labels).

It is also possible to visualize effects on the screen by using post estimation plot commands. For the nonlinear effect of the trend as well as the time-varying effect obtained from model (A.1), for example, such plots can be obtained by executing the commands

```
> erf.plotnonp 1
> erf.plotnonp 17
```

where the numbers following the `plotnonp` command depend on the order in which the model terms have been specified. These numbers are supplied in the *output window* after estimation. The graphs produced for this commands then appear in an *Object-Viewer window*.

Finally, `graph` objects may be used to produce graphics using the obtained ASCII files containing the estimation results. Using these `graph` objects, we can, for example, visualize sampling paths for parameters.

To create a `graph` object, we execute

```
> graph g
```

Having created the `graph` object, we need a new `dataset` object to store the data to be plotted:

```
> dataset pmsample
```

After having executed the method `getsample` on a `bayesreg` object, we can now use the command `plotsample` to visualize the sampling paths for the parameters. For example, sampling paths for the sampled true $PM_{2.5}$ values can be obtained by the following code:

```
> pmsample.infile using c:\ temp\ pm25sample_x_f_tag_merror_sample.raw
> g.plotsample using pmsample
```

Some example sampling paths are shown in Figure A.1.

A.6 Convergence diagnostics

Convergence behavior of a chosen MCMC algorithm should be controlled for each particular analysis. This is typically done by an output analysis of the samples.

The first approach is the visual inspection of the samples of all parameters involved to determine the length of the burn-in phase. As a rule of thumb, it is recommended to discard the "*first one or two percent of a run long enough to give sufficient precision*" (Knorr-Held (1997)). Good indicators of convergence are plots of the sampling paths and an analysis of the autocorrelation functions. These can provide evidence whether

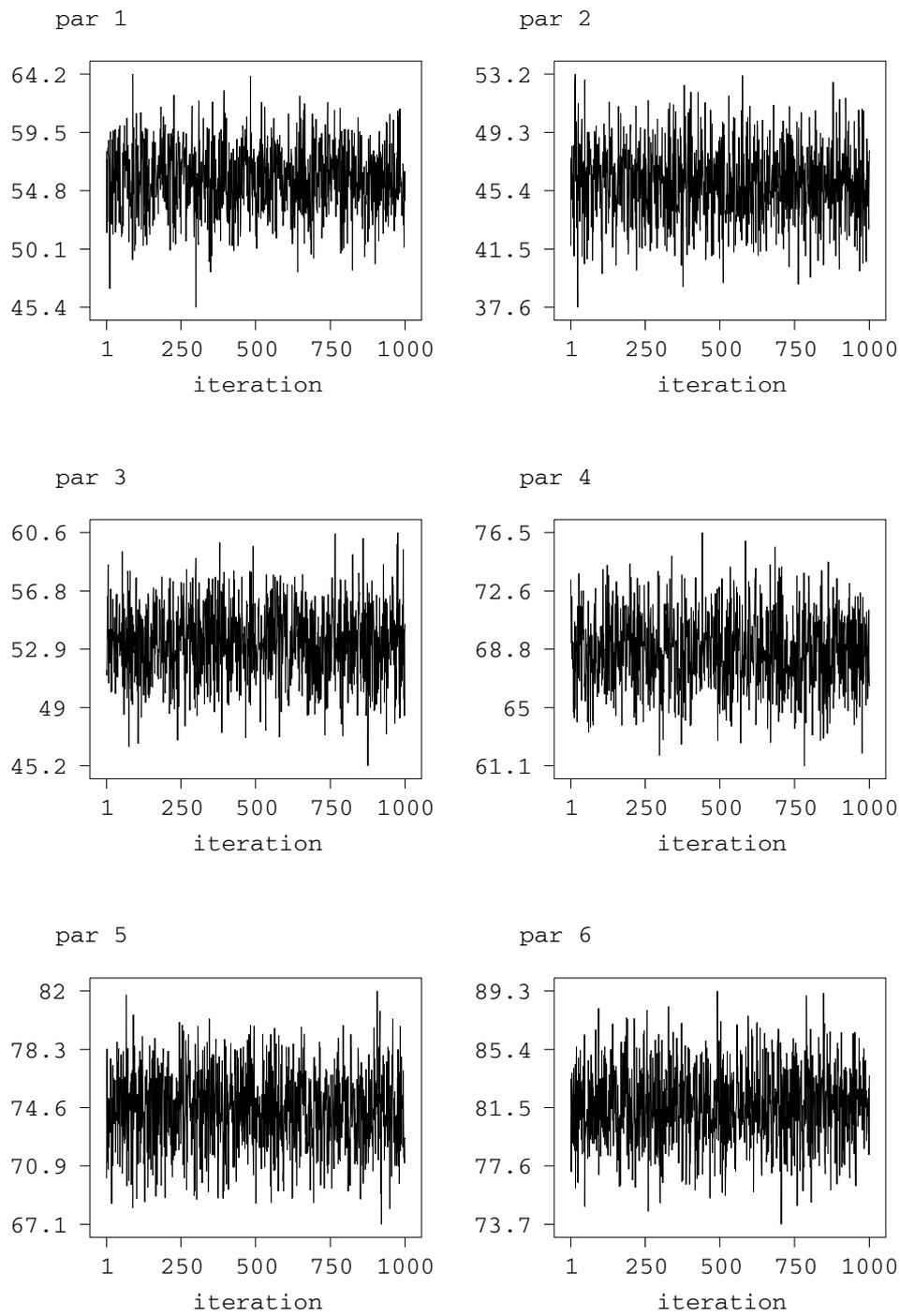


Figure A.1: Sampling paths for the first $PM_{2.5}$ values of the Bayesian measurement error model.

the chain shows good convergence behavior or not. If there is high autocorrelation between successive states of the generated chain, 'thinning' can be used. This means that only every k -th sampled parameter after the burn in phase will be taken as a stage in the chain. A further tool is the acceptance rate which is defined as the proportion of accepted changes of state in the generated chain. A high acceptance rate means that too many of the proposed values are accepted and the proposed values are very close to the current ones. Because of the small steps, the support of the target density is covered very slowly, which, consequently, results in a poor mixing behavior. On the other hand, a too low acceptance rate mean that the chain does not change often enough the states, since the proposed values may fall in low probability zones of the support of the target distribution, far away from the current value. As a consequence slow convergence and poor mixing are achieved. For an optimal mixing the acceptance rate should be as a rule of thumb between 30% and 80% (Fahrmeir & Lang (2001)). Another approach to check for convergence is given by comparing a set of chains based on different starting values.

Chapter B

Proof of the ARMA(p,p) representation of AR(p) models plus white noise

From Pagano (1974) and Granger & Morris (1976), it is known that: If ξ_t is an AR(p) process with autoregressive parameters ρ and ϵ_t is white noise with (constant) variance σ_ϵ^2 , then $X_t = \xi_t + \epsilon_t$ follows an autoregressive moving average (ARMA) process, ARMA(p,p).

Proof An AR process of order p for ξ_t can be represented by

$$\rho(B)\xi_t = e_t,$$

where $\rho(B)$ is defined as p -th order lag polynomial, $\rho(B) = 1 - \rho_1 B - \dots - \rho_p B^p$, with B is the so-called Backward-Shift operator $B^j \xi_t = \xi_{t-j}$, and e_t is white noise. Since the measurement error ϵ_t is a white noise process which is uncorrelated with ξ_{t-j} for all j , it can be inferred that ϵ_t is also uncorrelated with e_{t-j} for all j ; however, this implies:

$$\rho(B)X_t = \rho(B)\xi_t + \rho(B)\epsilon_t = e_t + \rho(B)\epsilon_t = \nu_t,$$

where $\nu_t = e_t + \rho(B)\epsilon_t$. ν_t then is the sum of a white noise process and an (uncorrelated) MA(p) process. However, this sum is, following a lemma of Granger & Morris (1976), itself an MA(p) process:

$$\nu_t = \theta(B)\gamma_t,$$

where $\theta(B)$ is defined as $\theta(B) = 1 + \theta_1 B + \dots + \theta_p B^p$ and γ_t is white noise, since all autocovariances of ν_t at lags greater than p are zero. Therefore, one obtains

$$\rho(B)X_t = \rho(B)\xi_t + \rho(B)\epsilon_t = \theta(B)\gamma_t,$$

and thus X_t follows an ARMA(p,p) process. \square

Bibliography

- Armstrong, B. (1985). Measurement error in the generalized linear model, *Communications in Statistics: Simulation and Computation* **14**: 529–544.
- Bateson, T. F. & Schwartz, J. (1999). Control for seasonal variation and time trend in case-crossover studies of acute effects of environmental exposures, *Epidemiology* **10**: 539–544.
- Bell, M. L., Dominici, F., Ebisu, K., Zeger, S. L. & Samet, J. M. (in press). Spatial and temporal variation in pm_{2.5} chemical composition in the united states for health effects studies, *Environmental Health Perspectives* .
- Bell, M. L., Samet, J. M. & Dominici, F. (2004). Time Series Studies of Particulate Matter, *Annual Review of Public Health* **25**: 247–280.
- Besag, J., Green, P., Higdon, D. & Mengersen, K. (1995). Bayesian Computation and Stochastic Systems (with discussion), *Statistical Science* **10**: 3–41.
- Brand, P., Ruoß, K. & Gebhart, J. (1992). Performance of a mobile aerosol spectrometer for an in situ characterization of environmental aerosols in Frankfurt city, *Atmospheric Environment A* **26**: 2451–2457.
- Breslow, N. E. & Clayton, D. G. (1993). Approximate Inference in Generalized Linear Mixed Models, *Journal of the American Statistical Association* **88**: 9–25.
- Brezger, A. (2005). *Bayesian P-Splines in Structured Additive Regression Models*, PhD thesis, Department of Statistics, University of Munich.
- Brezger, A., Kneib, T. & Lang, S. (2005). BayesX: Analysing Bayesian Structured Additive Regression Models, *Journal of Statistical Software* **14**(11).
- Brezger, A. & Lang, S. (2006). Generalized structured additive regression based on Bayesian P-splines, *Computational Statistics and Data Analysis* **50**: 967–991.

- Brockwell, P. J. & Davis, R. A. (1998). *Time Series: Theory and Methods*, 2 edn, Springer, New York.
- Brumback, B. A., Ruppert, D. & Wand, M. P. (1999). Comment on "variable Selection and Function Estimation in Additive Nonparametric Regression Using a Data-Based Prior" by Shively, T. S., Kohn, R., and Wood, S., *Journal of the American Statistical Association* **94**: 794–797.
- Burnett, R. T., Cakmak, S., Raizenne, M. E., Stieb, D., Vincent, R., Krewski, D., Brook, J. R., Philips, O. & Ozkaynak, H. (1998). The association between Ambient Carbon Monoxide Levels and Daily Mortality in Toronto, Canada, *Journal of the Air & Waste Management Association* **48**: 689–700.
- Cai, Z., Fan, J. & Li, R. (2000). Efficient Estimation and Inferences for Varying-Coefficient Models, *Journal of the American Statistical Association* **95**: 888–902.
- Carroll, R. J., Ruppert, D., Stefanski, L. A. & Crainiceanu, C. M. (2006). *Measurement Error in Nonlinear Models: A Modern Perspective*, 2 edn, Chapman & Hall, London.
- Carroll, R. J. & Stefanski, L. A. (1990). Approximate Quasi-likelihood Estimation in Models With Surrogate Predictors, *Journal of the American Statistical Association* **85**: 652–663.
- Casella, G. & George, E. (1992). Explaining Gibbs sampler, *Journal of the American Statistical Association* **46**: 167–174.
- Chib, S. & Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm, *Journal of the American Statistical Association* **49**: 327–335.
- Chiogna, M. & Gaetan, C. (2002). Dynamic generalized linear models with applications to environmental epidemiology, *Applied Statistics* **51**: 453–468.
- Chong, T. T.-L., Liew, V., Zhang, Y. & Wong, C.-L. (2006). Estimation of the Autoregressive Order in the Presence of Measurement Errors, *Economics Bulletin* **3**: 1–10.
- Ciocco, A. & Thompson, D. J. (1961). A Follow-Up of Donora Ten Years After: Methodology and Findings, *American Journal of Public Health* **51**: 155–164.
- Clayton, D. G. (1992). Models for the analysis of cohort and case-control studies with inaccurately measured exposures. In: Dwyer, J. H., Feinleib, M., Lippert, P., Hoffmeister, H. (eds): *Statistical Models for Longitudinal Studies of Health* *Monographs in Epidemiology and Biostatistics*, Oxford University Press, New York, pp. 301–331.

- Coull, B. A., Schwartz, J. & Wand, M. P. (2001). Respiratory health and air pollution: additive mixed model analyses, *Biostatistics* **2**: 337–349.
- Crainiceanu, C. M. & Ruppert, D. (2004). Likelihood ratio tests in linear mixed models with one variance component, *Journal of the Royal Statistical Society B* **66**: 165–185.
- Crainiceanu, C., Ruppert, D., Claeskens, G. & Wand, M. P. (2005). Exact likelihood ratio tests for penalised splines, *Biometrika* **92**: 91–103.
- Daniels, M. J., Dominici, F., Samet, J. M. & Zeger, S. L. (2000). Estimating Particulate Matter-Mortality Dose-Response Curves and Threshold Levels: An Analysis of Daily Time-Series for the 20 Largest US Cities, *American Journal of Epidemiology* **152**: 397–406.
- de Boor, C. (1978). *A Practical Guide to Splines*, Springer, Berlin.
- Dellaportas, P. & Stephens, D. A. (1995). Bayesian analysis of error-in-variables regression models, *Biometrics* **51**: 1085–1095.
- Dierckx, P. (1993). *Curve and Surface Fitting with Splines*, Clarendon Press, Oxford.
- Dockery, D. W., Schwartz, J. & Spengler, J. D. (1992). Air Pollution and Daily Mortality: Associations with Particulates and Acid Aerosols, *Environmental Research* **59**: 362–373.
- Dominici, F., Daniels, M., Zeger, S. L. & Samet, J. M. (2002). Air Pollution and Mortality: Estimating Regional and National Dose-Response Relationships, *Journal of the American Statistical Association* **97**: 100–111.
- Dominici, F., Zeger, S. L. & Samet, J. M. (2000). A measurement error model for time-series studies of air pollution and mortality, *Biostatistics* **1**: 157–175.
- Eilers, P. H. C. & Marx, B. D. (1996). Flexible Smoothing with B-splines and Penalties (with comments and rejoinder), *Statistical Science* **11**: 89–121.
- Eilers, P. H. C. & Marx, B. D. (2004). Splines, Knots, and Penalties, *submitted*.
- Eubank, R. L., Huang, C., Maldonado, Y. M., Wang, N., Wang, S. & Buchanan, R. J. (2004). Smoothing spline estimation in varying-coefficient models, *Journal of the Royal Statistical Society B* **66**: 653–667.
- Fahrmeir, L. & Kneib, T. (2006). Propriety of Posteriors in Structured Additive Regression Models: Theory and Empirical Evidence.

- Fahrmeir, L., Kneib, T. & Lang, S. (2004). Penalized structured additive regression for space-time data: a Bayesian perspective, *Statistica Sinica* **14**: 731–761.
- Fahrmeir, L. & Lang, S. (2001). Bayesian inference for generalized additive mixed models based on Markov random field priors, *Journal of the Royal Statistical Society C* **50**: 201–220.
- Fahrmeir, L. & Mayer, J. (2001). Bayesian-type count data models with varying coefficients: estimation and testing in the presence of overdispersion, *Applied Stochastic Models in Business and Industry* **17**: 165–179.
- Fahrmeir, L. & Tutz, G. (2001). *Multivariate Statistical Modelling Based on Generalized Linear Models*, 2 edn, Springer, New York.
- Fairly, D. (1990). The Relationship of Daily Mortality to Suspended Particulates in Santa Clara County, 1980-1986, *Environmental Health Perspectives* **89**: 159–168.
- Fan, J. & Chen, J. (1999). One-step local quasi-likelihood estimation, *Journal of the Royal Statistics Society, Series B* **61**: 927–943.
- Fan, J. & Gijbels, I. (1996). *Local Polynomial Modeling and Its Applications*, Chapman & Hall, London.
- Fan, J. & Yao, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods*, Springer, New York.
- Fan, J., Yao, Q. & Cai, Z. (2003). Adaptive varying-coefficient linear models, *Journal of the Royal Statistical Society B* **65**: 57–80.
- Fan, J., Zhang, C. & Zhang, J. (2001). Generalized Likelihood Ratio Statistics and Wilks Phenomenon, *Annals Of Statistics* **29**: 153–193.
- Fan, J. & Zhang, W. (1999). Statistical estimation in varying-coefficient models, *Annals of Statistics* **27**: 1491–1518.
- Fan, J. & Zhang, W. (2000). Simultaneous Confidence Bands and Hypothesis Testing in Varying-coefficient Models, *Scandinavian Journal of Statistics* **27**: 715–731.
- Forastiere, F., Stafoggia, M., Picciotto, S., Bellander, T., D'Ippoliti, D., Lanzi, T., von Klot, S., Nyberg, F., Paatero, P., Peters, A., Pekkanen, J., Sunyer, J. & Perucci, C. A. (2005). A Case-Crossover Analysis of Out-of-Hospital Coronary Deaths and Air Pollution in Rome, Italy, *American Journal of Respiratory and Critical Care Medicine* **172**: 1549–1555.

- Friket, J. (1931). The cause of the symptoms found in the Meuse Valley during the fog of December, 1930, *Bulletin et Memoires de l'Academie Royale de Medecine de Belgique* **11**: 683–741.
- Fuller, W. A. (1987). *Measurement Error Models*, Wiley, New York.
- Galindo, C. D., Liang, H., Kauermann, G. & Carroll, R. J. (2001). Bootstrap confidence intervals for local likelihood, local estimating equations and varying coefficient models, *Statistica Sinica* **11**: 121–134.
- Gamerman, D. (1997). Sampling from the posterior distribution in generalized linear mixed models, *Statistics and Computing* **7**: 57–68.
- Gamerman, D. (1998). Markov chain Monte Carlo for dynamic generalised linear models, *Biometrika* **85**: 215–227.
- Gelfand, A. E., Zhu, L. & Carlin, B. P. (2001). On the change of support problem for spatio-temporal data, *Biostatistics* **2**: 31–45.
- Gilks, W. R., Richardson, S. & Spiegelhalter, D. J. (1996). *Markov Chain Monte Carlo in Practice*, Chapman & Hall, London.
- Gleser, L. J. (1990). Improvements of the naive approach to estimation in nonlinear errors-in-variables regression models. In: P. J. Brown and W. A. Fuller (eds): *Statistical Analysis of Error Measurement Models and Application*, American Mathematics Society, Providence.
- Granger, C. W. J. & Morris, M. J. (1976). Time Series Modelling and Interpretation, *Journal of the Royal Statistical Society A* **139**: 246–257.
- Hamilton, J. D. (1994). *Time Series Analysis*, Princeton University Press, Princeton.
- Hastie, T. J. & Tibshirani, R. J. (1990). *Generalized Additive Models*, Chapman & Hall, London.
- Hastie, T. & Tibshirani, R. (1993). Varying-Coefficient Models, *Journal of the Royal Statistical Society B* **55**: 757–796.
- Hastie, T. & Tibshirani, R. (2000). Bayesian Backfitting (with comments and rejoinder), *Statistical Science* **15**: 196–223.
- Health Effects Institute (2003). Revised Analyses of Time-Series Studies of Air Pollution and Health, *Special report*, Health Effects Institute.

- Hoover, D. R., Rice, J. A., Wu, C. O. & Yang, L. P. (1998). Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data, *Biometrika* **85**: 809–822.
- Huang, J. Z., Wu, C. O. & Zhou, L. (2002). Varying-coefficient models and basis function approximations for the analysis of repeated measurements, *Biometrika* **89**: 111–128.
- Isaacson, J. D. & Zimmerman, D. L. (2000). Combining temporally correlated environmental data from two measurement systems, *Journal of Agricultural, Biological, and Environmental Statistics* **5**: 398–416.
- Janes, H., Sheppard, L. & Lumley, T. (2005). Case-Crossover Analyses of Air Pollution Exposure Data: Referent Selection Strategies and Their Implications for Bias, *Epidemiology* **16**: 717–726.
- Jerak, A. & Lang, S. (2005). Locally Adaptive Function Estimation for Binary Regression Models, *Biometrical Journal* **47**: 151–166.
- Jones, R. H. (1980). Maximum Likelihood Fitting of ARMA Models to Time Series with Missing Observations, *Technometrics* **22**: 389–395.
- Katsouyanni, K., Touloumi, G., Samoli, E., Gryparis, A., Tertre, A. L., Monopoli, Y., Rossi, G., Zmirou, D., Ballester, F., Boumghar, A., Anderson, H. R., Wojtyniak, B., Paldy, A., Braunstein, R., Pekkanen, J., Schindler, C. & Schwartz, J. (2001). Confounding and Effect Modification in the Short-Term Effects of Ambient Particles on Total Mortality: Results from 29 European Cities within the APHEA2 Project, *Epidemiology* **12**: 521–531.
- Katsouyanni, K., Touloumi, G., Samoli, E., Gryparis, A., Tertre, A. L., Monopoli, Y., Rossi, G., Zmirou, D., Ballester, F., Boumghar, A., Anderson, H. R., Wojtyniak, B., Paldy, A., Braunstein, R., Pekkanen, J., Schindler, C. & Schwartz, J. (2002). Different Convergence Parameters Applied to the S-Plus GAM function, *Epidemiology* **13**: 742.
- Kauermann, G., Claeskens, G. & Opsomer, J. D. (2006). Bootstrapping for Penalized Spline Regression, *Technical report*, University of Bielefeld.
- Kauermann, G. & Tutz, G. (1999). On model diagnostics using varying coefficient models, *Biometrika* **86**: 119–128.
- Kauermann, G. & Tutz, G. (2001). Testing generalized linear and semiparametric models against smooth alternatives, *Journal of the Royal Statistical Society B* **63**: 147–166.

- Kneib, T. (2006). *Mixed model based inference in structured additive regression*, PhD thesis, Department of Statistics, University of Munich.
- Knorr-Held, L. (1997). *Hierarchical Modelling of Discrete Longitudinal Data*, Herbert Utz Verlag, München.
- Knorr-Held, L. (1999). Conditional Prior Proposals in Dynamic Models, *Scandinavian Journal of Statistics* **26**: 129–144.
- Krivobokova, T., Crainiceanu, C. M. & Kauermann, G. (2006). Fast Adaptive Penalized Splines, *Working paper*, Johns Hopkins University, Department of Biostatistics.
- Ku, S. & Seneta, E. (1998). Practical estimation from the sum of AR(1) processes, *Communication in Statistics: Simulation and Computation* **27**: 981–998.
- Lang, S. & Brezger, A. (2004). Bayesian P-Splines, *Journal of Computational & Graphical Statistics* **13**: 183–212.
- Lang, S., Fronk, E. & Fahrmeir, L. (2002). Function estimation with locally adaptive dynamic models, *Computational Statistics* **17**: 479–499.
- Lee, D. & Shaddick, G. (2005). Modelling the effects of air pollution on health using Bayesian Dynamic Generalised Linear Models, *Technical report*, University of Bath.
- Liang, H., Wu, H. & Carroll, R. J. (2003). The relationship between virologic and immunologic responses in AIDS clinical research using mixed-effects varying-coefficient models with measurement error, *Biostatistics* **4**: 297–312.
- Lin, X. & Zhang, D. (1999). Inference in generalized additive mixed models using smoothing splines, *Journal of the Royal Statistical Society B* **61**: 381–400.
- Lindstrom, M. J. & Bates, D. M. (1990). Nonlinear Mixed Effects Models for Repeated Measures Data, *Biometrics* **46**: 673–687.
- Lu, Z., Tjøstheim, D. & Yao, Q. (2007). Adaptive Varying-Coefficient Linear Models for Stochastic Processes: Asymptotic Theory, *Statistica Sinica* **17**: 177–197.
- Marple, V., Rubow, K. L., Turner, W. & Spengler, J. D. (1987). Low flow rate sharp cut impactors for indoor sampling: design and calibration, *Journal of the Air Pollution Control Association* **37**: 1303–1307.
- McCullagh, P. & Nelder, J. A. (1989). *Generalized Linear Models*, 2 edn, Chapman & Hall, London.

- Ministry of Health (1954). Mortality and morbidity during the London fog of December 1952, *Report on Public Health and Medical Subjects No 95*, Her Majesty's Station Office, London.
- Neas, L. M., Schwartz, J. & Dockery, D. (1999). A Case-Crossover Analysis of Air Pollution and Mortality in Philadelphia, *Environmental Health Perspectives* **107**: 629–631.
- Nelder, J. A. & Wedderburn, R. W. M. (1972). Generalized Linear Models, *Journal of the Royal Statistical Society A* **135, Part 3**: 370–384.
- Ngo, L. & Wand, M. P. (2004). Smoothing with Mixed Model Software, *Journal of Statistical Software* **9**: Issue 1.
- Özkaynak, H. & Spengler, J. D. (1985). Analysis of Health Effects Resulting from Population Exposures to Acid Precipitation Precursors, *Environmental Health Perspectives* **63**: 45–55.
- Pagano, M. (1974). Estimation of Models of Autoregressive Signal Plus White Noise, *Annals of Statistics* **2**: 99–108.
- Peng, R. D., Dominici, F., Pastor-Barriuso, R., Zeger, S. L. & Samet, J. M. (2005). Seasonal Analyses of Air Pollution and Mortality in 100 U.S. Cities, *American Journal of Epidemiology* **161**: 585–594.
- Peters, A., Stölzel, M., Cyrys, J., Breitner, S., Pitz, M., Wölke, G., Heinrich, J., Kreyling, W., Küchenhoff, H. & Wichmann, H.-E. (2007). Improved Air Quality and its Influences on Short-Term Health Effects in Erfurt, Eastern Germany, *Research report*, Under review at the Health Effects Institute, Cambridge, MA.
- Pope, C. A., Dockery, D. W. & Schwartz, J. (1995). Review of epidemiologic evidence of health effects of particulate air pollution, *Inhalation Toxicology* **7**: 1–18.
- Pope, C. A., Schwartz, J. & Ransom, M. R. (1992). Daily mortality and PM₁₀ Pollution in Utah Valley - particulate pollution, *Archives of Environmental Health* **47**: 211–217.
- Ramsay, J. O. & Silverman, B. W. (1997). *Functional Data Analysis*, Springer, Berlin.
- Ramsay, J. O. & Silverman, B. W. (2002). *Applied Functional Data Analysis: Methods and Case Studies*, Springer, New York.
- Richardson, S. & Gilks, W. R. (1993). Conditional independence models for epidemiological studies with covariate measurement error, *Statistics in Medicine* **12**: 1703–1722.

- Richardson, S., Leblond, L., Jaussent, I. & Green, P. J. (2002). Mixture models in measurement error problems, with reference to epidemiological studies, *Journal of the Royal Statistical Society A* **165**: 549–566.
- Robert, C. P. & Casella, G. (2004). *Monte Carlo statistical methods*, 2 edn, Springer, New York.
- Ruppert, D. (2002). Selecting the Number of Knots for Penalized Splines, *Journal of Computational and Graphical Statistics* **11**: 735–757.
- Ruppert, D., Wand, M. P. & Carroll, R. J. (2003). *Semiparametric Regression*, Cambridge University Press, Cambridge, UK.
- Samet, J. M., Dominici, F., Zeger, S. L., Schwartz, J. & Dockery, D. W. (2000a). The National Morbidity, Mortality, and Air Pollution Study, Part I: Methods and Methodologic Issues, *Research Report No 94*, Health Effects Institute, Cambridge, MA.
- Samet, J. M., Zeger, S. L., Dominici, F., Curriero, F., Coursac, I., Dockery, D. W., Schwartz, J. & Zanobetti, A. (2000b). The National Morbidity, Mortality, and Air Pollution Study, Part II: Morbidity and mortality from air pollution in the United States, *Research Report No 94*, Health Effects Institute, Cambridge, MA.
- Samoli, E., Analitis, A., Touloumi, G., Schwartz, J., Anderson, H. R., Sunyer, J., Bisanti, L., Zmirou, D., Vonk, J. M., Pekkanen, J., Goodman, P., Paldy, A., Schindler, C. & Katsouyanni, K. (2005). Estimating the Exposure-Response Relationships between Particulate Matter and Mortality within the APHEA Multicity Project, *Environmental Health Perspectives* **113**: 88–95.
- Schrenk, H. H., Heimann, H., Clayton, G. D., Gafafer, G. D. & Wexler, H. (1949). Air pollution in Donora, Pennsylvania. Epidemiology of the unusual smog episode in October, 1948, *Public Health Bulletin No 306*, Public Health Service, Washington D.C.
- Schwartz, J. (1993). Air Pollution and Daily Mortality in Birmingham, Alabama, *American Journal of Epidemiology* **137**: 1136–1147.
- Schwartz, J. (1994). Total Suspended Particulate Patter and Daily Mortality in Cincinnati, Ohio, *Environmental Health Perspectives* **102**: 186–189.
- Schwartz, J., Ballester, F., Saez, M., Pérez-Hoyos, S., Bellido, J., Cambra, K., Arribas, F., Cañada, A., Pérez-Boillos, M. J. & Sunyer, J. (2001). The Concentration-Response relation between Air Pollution and Daily Deaths, *Environmental Health Perspectives* **109**: 1001–1006.

- Schwartz, J. & Coull, B. A. (2003). Control for confounding in the presence of measurement error in hierarchical models, *Biostatistics* **4**: 539–553.
- Schwartz, J. & Dockery, D. W. (1992). Particulate Air Pollution and Daily Mortality in Steubenville, Ohio, *American Journal of Epidemiology* **135**: 12–19.
- Schwartz, J., Dockery, D. W. & Neas, L. M. (1996). Is daily mortality associated specifically with fine particles?, *Journal of the Air & Waste Management Association* **46**: 927–939.
- Seaton, A. & Dennekamp, M. (2003). Hypothesis: Ill health associated with low concentrations of nitrogen dioxide—an effect of ultrafine particles?, *Thorax* **58**: 1012–1015.
- Seifert, B. & Gasser, T. (1996). Finite-Sample Variance of Local Polynomials: Analysis and Solutions, *Journal of the American Statistical Association* **91**: 267–275.
- Shaddick, G. & Wakefield, J. (2002). Modelling daily multivariate pollutant data at multiple sites, *Journal of the Royal Statistical Society C* **51**: 351–372.
- Sheppard, L. & Damian, D. (2000). Estimating short-term PM effects accounting for surrogate exposure measurements from ambient monitors, *Environmetrics* **11**: 675–687.
- Spix, C., Heinrich, J., Dockery, D., Schwartz, J., Völksch, G., Schwinkowski, K., Cöllen, C. & Wichmann, H.-E. (1993). Air Pollution and Daily Mortality in Erfurt, East Germany, 1980-1989, *Environmental Health Perspectives* **101**: 518–526.
- Staudenmayer, J. & Buonaccorsi, J. P. (2005). Measurement Error in Linear Autoregressive Models, *Journal of the American Statistical Association* **100**: 841–852.
- Stölzel, M., Breitner, S., Cyrys, J., Pitz, M., Wölke, G., Kreyling, W., Heinrich, J., Wichmann, H.-E. & Peters, A. (in press). Daily mortality and particulate matter in different size classes in Erfurt, Germany, *Journal of Exposure Science and Environmental Epidemiology* .
- Stölzel, M., Breitner, S., Cyrys, J., Pitz, M., Wölke, G., Kreyling, W., Küchenhoff, H., Heinrich, J., Wichmann, H.-E. & Peters, A. (submitted). Short-term health effects during a decade of improved air quality in Erfurt, Germany.
- Tonellato, S. F. (2001). A multivariate time series model for the analysis and prediction of carbon monoxide atmospheric concentrations, *Journal of the Royal Statistical Society C* **50**: 187–200.

- Touloumi, G., Atkinson, R., Tertre, A. L., Samoli, E., Schwartz, J., Schindler, C., Vonk, J. M., Rossi, G., Saez, M., Rabszenko, D. & Katsouyanni, K. (2004). Analysis of health outcome time series data in epidemiological studies, *Environmetrics* **15**: 101–117.
- Touloumi, G., Samoli, E., Pipikou, M., Tertre, A. L., Atkinson, R. & Katsouyanni, K. (2006). Seasonal confounding in air pollution and health time-series studies: effect on air pollution effect estimates, *Statistics in Medicine* **25**: 4164–4178.
- Tutz, G. & Kauermann, G. (2003). Generalized linear random effect models with varying coefficients, *Computational Statistics & Data Analysis* **43**: 13–28.
- Wand, M. P. (2003). Smoothing and Mixed Models, *Computational Statistics* **18**: 223–249.
- West, M., Harrison, J. P. & Migon, H. S. (1985). Dynamic generalized linear models and bayesian forecasting, *Journal of the American Statistical Association* **80**: 73–96.
- Wichmann, H.-E., Spix, C., Tuch, T., Wölke, G., Peters, A., Heinrich, J., Kreyling, W. G. & Heyder, J. (2000). Daily Mortality and Fine and Ultrafine Particles in Erfurt, Germany. Part I: Role of Particle Number and Particle Mass, *Research Report No 98*, Health Effects Institute, Cambridge, MA.
- Wolfinger, R. & O’Connell, M. (1993). Generalized Linear Mixed Models: A Pseudo-Likelihood Approach, *Journal of Statistical Computation and Simulation* **48**: 233–243.
- You, J., Zhou, X. & Chen, G. (2006). Corrected local polynomial estimation in varying-coefficient models with measurement errors, *Canadian Journal of Statistics* **34**: 391–410.
- Zeger, S. L., Thomas, D., Dominici, F., Samet, J. M., Schwartz, J., Dockery, D. & Cohen, A. (2000). Exposure Measurement Error in Time-Series Studies of Air Pollution: Concepts and Consequences, *Environmental Health Perspectives* **108**: 419–426.
- Zeka, A. & Schwartz, J. (2004). Estimating the Independent Effects of Multiple Pollutants in the Presence of Measurement Error: An Application of a Measurement-Error-Resistant Technique, *Environmental Health Perspectives* **112**: 1686–90.
- Zhang, D. (2004). Generalized Linear Mixed Models with Varying Coefficients for Longitudinal Data, *Biometrics* **60**: 8–15.

LEBENS LAUF

Susanne Breitner
geboren am 08.12.1977 in Wolnzach

Schulbildung:

1984 - 1988 Joseph-Maria-Lutz-Grundschule in Pfaffenhofen
1988 - 1997 Schyrenngymnasium Pfaffenhofen

Studium:

11/1997 - 02/2003 Studium der Statistik an der Ludwig-Maximilians-Universität München mit dem Anwendungsgebiet Medizinische Biometrie und der speziellen Ausrichtung mathematische Stochastik
05/2000 Diplom-Vorprüfung
02/2003 Abschluss: Diplom

Beruf:

01/2000 - 02/2003 Studentische Hilfskraft im Statistischen Beratungslabor (StaBLab) am Department für Statistik der LMU München
03/2003 - 01/2004 Wissenschaftliche Mitarbeiterin im Statistischen Beratungslabor (StaBLab)
02/2004 - 04/2004 Auslandsaufenthalt beim Environmental Epidemiology Program der Harvard School of Public Health, Boston
05/2004 - 12/2004 Wissenschaftliche Mitarbeiterin im Statistischen Beratungslabor (StaBLab)
seit 01/2005 wissenschaftliche Mitarbeiterin am Institut für Epidemiologie am GSF-Forschungszentrum für Umwelt und Gesundheit

Pfaffenhofen, den 23.02.2007