Exposure Modeling and Exposure Measurement Error Correction in Health Outcome Models with Longitudinal Data Structure: Exposure to Particulate Matter

Veronika Deffner

München 2016

Exposure Modeling and Exposure Measurement Error Correction in Health Outcome Models with Longitudinal Data Structure: Exposure to Particulate Matter

Veronika Deffner

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

vorgelegt von
Veronika Deffner
aus Hagau

München, den 06.06.2016

Erstgutachter: Prof. Dr. Helmut Küchenhoff
Zweitgutachter: Prof. Dr. Ciprian Crainiceanu
Tag der Disputation: 28. Juli 2016

Abstract

The human organism is permanently exposed to various environmental factors, which influence its performance, e.g. climatic conditions, nutrition intake behavior or air quality. This thesis focuses on the human exposure to particulate matter.

The "Augsburger Umweltstudie" (2007–2008) was conducted at the KORA study center in Augsburg and aimed at the investigation of the association between particulate matter and human health. An accompanying validation study was conducted in 2011 in order to collect information about the errors in the measurement of particulate matter.

The complexity of these errors arises from different data sources: 1.) The usage of population-specific exposure measurements, e.g. from one or several fixed site measurement stations, instead of personal measurements involves a Berkson type error. The deviations between personal and populations–specific measurements are driven by the microenvironment of the person and the climatic conditions. In the first part of the work, a two–level exposure model is developed for the association between the fixed–site and the personal exposure measurements of the validation study including the selection of relevant covariates and the appropriate consideration of categorical covariates in the analysis of longitudinal data. 2.) The mobile devices used for personal exposure measurements exhibit classical measurement error, which is partially device–specific and autocorrelated. 3.) In order to use as much information as possible missing personal exposure measurements are filled in with population–specific exposure measurements resulting in a mixture of Berkson and classical measurement error.

The second part of the work aims at the development and application of methods to include knowledge about Berkson, classical and mixture error into regression models of the health outcome. Therefore, the method–of–moments is extended to longitudinal data and to the different types of errors with individual–specific and autocorrelated structures.

Validation studies and expert knowledge provide information about the size of the measurement error, but prior knowledge is often afflicted with uncertainty. Approaches for the adequate inclusion of prior knowledge about the measurement errors in the Bayesian health outcome model are evaluated in the third part of the thesis. The role of prior knowledge in regression models with an error–prone covariate differs from conventional Bayesian regression models and is strongly affected by the interaction between the parameters in the model.

The thesis is closed with the application of the developed method–of–moments and Bayesian approach to the Augsburger Umweltstudie by integrating information from the validation studies.

Zusammenfassung

Der menschliche Organismus wird ständig durch verschiedene Umwelteinflüsse, wie klimatische Bedingungen, Gewohnheiten bei der Nahrungsaufnahme oder Luftqualität, belastet, die seine Leistungsfähigkeit beeinflussen. Die vorliegende Arbeit beschäftigt sich mit der Belastung des Menschen durch Feinstaub.

Die "Augsburger Umweltstudie" (2007–2008) wurde vom KORA Studienzentrum in Augsburg mit dem Ziel durchgeführt, den Zusammenhang zwischen Feinstaub und menschlicher Gesundheit zu untersuchen. Darauf aufbauend wurde 2011 eine Validierungsstudie durchgeführt, um Informationen über Fehler bei der Messung von Feinstaub zu erhalten.

Die Komplexität dieser Fehler ist in unterschiedlichen Arten der Datenerhebung begründet: 1.) Die Verwendung populations–spezifischer Expositionsmessungen, z.B. von einer oder mehreren Messstationen, anstatt persönlicher Messungen bringt einen Berkson–Fehler mit sich. Die Stärke der Abweichungen zwischen persönlichen und populations–spezifischen Messungen hängt von der Mikro–Umgebung der Person und den klimatischen Bedingungen ab. Im ersten Teil der Arbeit wird ein zweistufiges Expositionsmodell für den Zusammenhang zwischen stationären und persönlichen Feinstaubexpositionsmessungen der Validierungsstudie entwickelt, wobei relevante Variablen selektiert werden und kategoriale Variablen geeignet in der Analyse longitudinaler Daten berücksichtigt werden. 2.) Die mobilen Geräte zur Messung der persönlichen Feinstaubexposition weisen einen klassischen Messfehler auf, der teilweise vom Gerät abhängt und autokorreliert ist. 3.) Fehlende persönliche Feinstaub–Messungen werden durch populations–spezifische Expositionsmessungen ersetzt, um möglichst viele Informationen in die Analyse einzuschließen, was eine Mischung aus Berkson und klassischem Messfehler zur Folge hat.

Das Ziel im zweiten Teil der Arbeit ist die Entwicklung und Anwendung von Methoden zur Berücksichtigung von Wissen über Berkson, klassischen und gemischten Fehler in Regressionsmodellen mit Gesundheitsparametern als Zielgröße. Dazu wird die Momentenmethode für longitudinale Daten und für verschiedenen Messfehlerarten mit individuen–spezifischen und autokorrelierten Strukturen erweitert.

Zwar liefern Validierungsstudien und Expertenwissen Informationen über die Größe des Messfehlers, aber das Vorwissen ist oft mit Unsicherheit behaftet. Ansätze zur adäquaten Einbindung von Vorwissen über Messfehler in Bayesianische Gesundheitsmodelle werden im dritten Teil der Arbeit bewertet. Die Rolle des Vorwissens in Regressionsmodellen mit einer fehlerhaften Kovariable unterscheidet sich von konventionellen Bayesianischen Regressionsmodellen und wird stark von der Interaktion zwischen den Modellparametern beeinflusst.

Die Arbeit schließt mit der Anwendung der entwickelten Momentenmethode und dem Bayesiansichen Ansatz auf die Augsburger Umweltstudie unter Berücksichtigung der Informationen aus den Validierungsstudien.

Acknowledgments

I would like to sincerely thank

- Helmut Küchenhoff, who supervised and advised my work on measurement error as well as on various other interesting projects during my employment at the Statistical Consulting Unit of the LMU,
- Josef Cyrys, Alexandra Schneider, Susanne Breitner, Annette Peters, Regina Hampel, Jianwei Gu, Mike Pitz and Uta Geruschkat from the Helmholtz Zentrum München and the University of Augsburg for providing the data and background materials, for the cooperation within our joint projects and for many valuable discussions,
- Ciprian Crainiceanu for enabling a research stay at Johns Hopkins School of Public Health in Baltimore and pointing out new perspectives,
- my friends and colleagues
- and particularly my family.

Contents

1. Introduction

1.1. Errors in the measurements of personal exposure to air pollutants

Accurate and precise measurements of personal exposure, a surrogate for the biologically effective dose (Hatch and Thomas, 1993) of a pollutant, are a prerequisite for the evaluation of short–term associations between air quality and health. However, meteorological variability (Hudda et al., 2010), individual differences and laboratory errors inevitably involve exposure measurement error in epidemiological questions. In the direct measurement approach, personal samplers are used to determine the exposure levels on an individual basis (i.e. for each study participant). A number of studies has been undertaken to measure personal exposure directly using personal monitors often with the aim to determine the correlations of personal exposure with outdoor concentrations (e.g. Gu et al., 2015; Meng et al., 2005; Oglesby et al., 2000). Since personal exposure measurements are time–consuming, cumbersome and cost–intensive, indirect exposure measurements from one or several outdoor monitoring sites located in the study area or estimated exposure concentrations of the study participants are often used in health outcome models as a proxy for average personal exposure (Monn, 2001), as e.g. in McCracken et al. (2009). Indeed, personal exposure measurements reflect the personal exposure more precisely than fixed–site outdoor measurements, but the majority of epidemiological studies aims to investigate the effect of personal exposure to outdoor air pollutants (Mage et al., 1999). The usage of personal exposure measurements for this purpose may even confound the exposure–outcome association, because the personal measurements are influenced by sources other than outdoor sources (Monn, 2001), e.g. indoor penetration of outdoor-generated air pollutants, indoor sources of air pollutants under consideration, time activity patterns of the observed population (e.g. time spent indoors, close to traffic, at the work place) and the spatial variation in ambient air concentrations. Therefore, the correlations between personal exposure and ambient levels of air pollutants are in general low.

Exposure models are an alternative to the direct usage of fixed–site measurements. Long–term outdoor concentration is often estimated through a spatio–temporal model, e.g. a dispersion model, (e.g. Szpiro et al., 2010) whereas short–term outdoor concentration is estimated using measurements of fixed–site monitors, which may be combined with a spatio–temporal model (e.g. Gulliver and Briggs, 2011). Using predicted values of exposure from an exposure model based on temporally or spatially smoothing of observed exposure values results in Berkson error (Gryparis et al., 2009). Since people spend the majority of their time indoors, the validity of using ambient concentrations for approximating personal exposure has raised concerns, because exposure misclassification could bias epidemiological results. Indoor concentrations are usually estimated through mass balance models based on outdoor concentration data.

The approximation of personal exposure through monitoring data represents the population–averaged exposure and does not adequately reflect the small–scale temporal variation in personal exposure through frequent alternation between different microenvironments (e.g. high–exposure microenvironments like during commuting or low–exposure microenvironments like during being at home) and individual time–activity patterns. The

combination of estimated ambient concentrations with time–activity data (Beckx et al., 2009; Calder et al., 2008; Ebelt et al., 2005; Gerharz et al., 2009; Lanki et al., 2007) and optionally with spatial information (Calder et al., 2008; Gerharz et al., 2009) permits to estimate personal exposure with a so–called microenvironmental model (Berhane et al., 2004) providing adequate estimates for personal exposure (Gerharz et al., 2013). However, Spatial and temporal (Gryparis et al.,2007b, 2009) as well as microenvironment–specific variability of the true personal exposure cannot completely considered by these approximations and bias health effect estimates (Chang et al., 2011; Peng and Bell, 2010).

Exposure measurement error occurs by using population–averaged exposure as well as by using personally measured exposure (e.g. Berhane et al., 2004; Thomas, 2000)). Population–averaged exposure measurements is assumed to exhibit Berkson error, if the average difference between personal and ambient exposure is constant, which is usually violated when the data comprises also personal indoor exposure measurements (Zeger et al., 2000). Low precision of mobile exposure measurement devices induce classical measurement error to personal exposure measurements. Besides the measurement uncertainty caused by the measurement devices, the temporal and spatial variation of the exposure affects the precision of the estimation of the average population exposure (White et al., 2008). Regionally averaged measurements, like measurements of a central monitor or an estimated exposure surface, deliver unbiased personal health effect estimations, if the averaged exposure measurements correctly reflect the actual average exposure of the considered population (Monn, 2001).

Another issue is the fact, that the degree of air pollution cannot be gathered by a single measure, but by a multitude of measures each describing only a few or some aspects of air pollution (Wilson and Suh, 1997). This involves additional errors, which do not relate to the measurement itself. We do not go into detail regarding the adequacy of the measures for air pollution.

1.2. Covariate measurement error in regression models

A short introduction to the terminology in the context of measurement errors is given in this section on the basis of Thomas et al. (1993). The term measurement error is commonly used for continuous variables, whereas misclassification is used when referring to discrete variables. Measurement errors are described in different ways: In contrast to *random* errors, which randomly fluctuate around the true value, *systematic* error exhibits some non–random, systematic regularity or structure. The measurement error is called *non–differential*, if the error does not contain any information about the outcome variable beyond the information of the precise, latent value of the variable. The latent variable can be considered as random variable (*structural* approach) or as fixed parameter (*functional* approach). The discrimination between *Berkson* and *classical* measurement error is essential for data analyses and was already introduced for air pollution measurements in the previous section. Classical measurement error, U^{C}, and Berkson error, U^{B}, are defined as a normally distributed random variables ($U^{\mathrm{C}} \sim \mathrm{N}(0, \sigma^2_{U^{\mathrm{C}}})$ and $U^{\mathrm{B}} \sim \mathrm{N}(0, \sigma^2_{U^{\mathrm{B}}})$), which are independent from the true measurement X and from the Berkson error–prone measurement $X^{*\mathrm{B}}$, respectively.

Statistical analyses rely on the accuracy of the collected data. Measurement error possibly yields distorted results Greenland (e.g. 1980). Thomas et al. (1993) and Zeger et al. (2000) give an overview of the impact of exposure measurement error on the relationship between exposure and health outcome.

In a linear regression model with normally distributed X, classical measurement error only affects the estimates of the effect coefficients and the residual variance but neither the linearity of the model nor the form of the conditional distribution of the response (Chesher, 1991). Classical measurement error mostly causes an attenuation of the effect estimate in linear regression analyses, i.e. a bias towards zero, and only a slight increase in the uncertainty of the effect estimate (Fuller, 1987; Carroll et al., 2006).

Effect estimates in a regression analysis with Berkson error–prone measurements are unbiased, but more variable than the estimates based on the true measurements (e.g. Zeger et al., 2000). However, the usage of ambient exposure measurements as a surrogate for personal exposure measurements involves bias in the exposure effect estimate of the regression model due to the neglected indoor particle sources (Calder et al., 2008; Zeger et al., 2000). Goldman et al. (2011) observed attenuated estimates for measurements with multiplicative Berkson and classical measurement error in time series data; the degree of attenuation depends on the type and size of the error.

Mixtures of Berkson and classical measurement error are considered by Li et al. (2007) and Mallick et al. (2002): measurements with Berkson error, e.g. from a central measurement unit, are overlayed with an instrumental error of the classical error type.

In order to account for covariate measurement error in regression analyses additional information for the identification of the model parameters are required, which can be provided by data from a validation study, repeated measurements or instrumental variables (Carroll et al., 2006). In the context of air pollution exposure measurements data may be collected from varying sources, including questionnaires, direct and indirect measurements (Monn, 2001; Needham et al., 2005), as well as subsample data from individual exposure measurements and microenvironment measurements (Gilliland et al., 2005; Monn, 2001). More details on methods for the assessment of the size of measurement error can be found in Hatch and Thomas (1993). This information can be used to account for a possible bias and the inflation of the standard errors due to measurement error. Although quantifying and allowing for exposure measurement error amplifies the benefit of epidemiological data (Burns et al., 2014), its impact on the study results is often ignored (Jurek et al., 2006) and methods for the correction of the measurement error are rarely applied (Willett, 1989).

The easiest method for the measurement error correction is the method of moments (MOM) (Fuller, 1987; Carroll et al., 2006). If the bias of the effect estimate can be expressed through the moments of the response and the predictor variables, this bias is corrected using an estimate for the bias based upon empirical moments.

The replacement of the actual, but unknown exposure, in the health effects model with predictions from a model for the true exposure is one possibility for error correction towards correcting exposure measurement error and is known as “regression calibration”; this approach yields too narrow confidence intervals (Rosner et al., 1989). For example, Strand et al. (2006) modeled the personal ambient exposure to $PM_{2.5}$ (particulate matter with an aerodynamic diameter less than 2.5 μm) with measurements from fixed–site monitors.

Another regression calibration approach is the usage of validation data to examine the relationship between a surrogate and the actual, directly measured exposure. Based on that, the effect estimate of the surrogate on the health outcome can be corrected, as it is applied in Bateson and Wright (2010); Horick et al. (2006); Li et al. (2006); Freedman et al. (2008) or Smith et al. (2010) with internal and in Dominici et al. (2000) or Van Roosbroeck et al. (2008) with external validation data.

Refer to Zeger et al. (2000) or Spiegelman (2010) for an introduction and Sarnat et al. (2007) for an overview regarding regression calibration.

The idea of the SIMEX method (Cook and Stefanski, 1994), a functional approach for measurement error correction, is to describe the impact of artificially added measurement error to the data on the effect estimate and to extrapolate the calculated trend to the situation without measurement error. In the past, the approach was extended e.g. to nonparametric regression models (Carroll et al., 1999), to misclassification (Küchenhoff et al., 2006), to clustered survival data (Li and Lin, 2003), to longitudinal data (Wang et al., 1998; Yi, 2008) to spatial data (Alexeeff et al., 2016). The SIMEX approach is developed for a single measurement error parameter; therefore, SIMEX cannot simultaneously deal with several kinds of measurement error, e.g. an individual–specific and a random classical measurement error.

Likelihood methods (e.g. ML in Stefanski and Carroll (1985); Spiegelman et al. (2000), Quasi–likelihood in Whittemore and Keller (1988) and Wang et al. (1996) and approximate Quasi–likelihood in Carroll and Stefanski (1990)) are an alternative approach allowing for measurement error in regression models. The specification of an error model is required in addition to the specification of the likelihood involving additional distributional assumptions, which result in a more or less pronounced increase in efficiency compared to e.g. the method of moments or regression calibration (Carroll et al., 2006), especially for large measurement error (Messer and Natarajan, 2008).

In 1983, Suggs and Curran proposed a Bayesian approach to account for measurement error in air pollution models and Richardson and Gilks introduced the method for epidemiological studies. Beside the likelihood and the error model, as for the ML methods, the priors for the model parameters are the third component of a Bayesian regression model for a measurement error–prone covariate. In comparison to the method of moments and regression calibration, the strengths of likelihood methods and Bayesian modeling are, that the exposure model and the health outcome model are simultaneously calculated and, that the health outcome model accounts for uncertainty of the exposure model (Dominici et al., 2000). At the same time, missing exposure data is estimated (Molitor et al., 2006). Another appealing advantage of Bayesian measurement error models is that the components for accounting for measurement error can be easily integrated in the majority of more complex models, e.g. in hierarchical models (Gryparis et al., 2007a; Schwartz and Coull, 2003) and that the extension of these components e.g. regarding spatial autocorrelated error terms (Molitor et al., 2007) or regarding spatial misalignment (Peng and Bell, 2010) is often feasible.

Many other methods for measurement error correction in regression models are mentioned in the literature including robust techniques (Guolo, 2008), multiple imputation (Blackwell et al., 2015) and the EM algorithm (Schafer, 1987), but we will concentrate in the following on the previous developments concerning measurement error methods for longitudinal data.

Recent work has extended the existing knowledge about covariate measurement errors and the corresponding methods to models with random effects (e.g. Tosteson et al., 1998; Wang et al., 1998), correlated measurement error terms (Wang et al., 1996, 2000), autocorrelated model error (Moberg and Brattström, 2011) and correlated predictors with correlated measurement error (Schwartz and Coull, 2003). So far, the impact of measurement error in linear mixed models with autocorrelated error terms has not been studied.

Apart from the previously mentioned Bayesian methods, many approaches for measurement error correction in regression models for longitudinal data base upon unbiased score functions (as e.g. Pan et al., 2009), which have been introduced by Stefanski and Carroll (1987). The quasilikelihood approach of Wang et al. (1996) for dealing with correlated measurement error was extended by Wang and Sullivan Pepe (2000) to unbalanced and unequally spaced longitudinal data using an expected estimation equations estimator.

1.3. Background of the Augsburger Umweltstudie

Data from the Augsburger Umweltstudie (main data)

The Augsburger Umweltstudie, an observational study, which was part of the Rochester Particle Center investigations, (Hampel et al., 2012a; Hampel et al., 2012b; Kraus et al., 2015; Peters et al., 2015; Rückerl et al., 2014) was conducted between March 2007 and December 2008 in the city of Augsburg and two adjacent counties at the KORA (Cooperative Health Research in the Augsburg Region) study center in Augsburg. The aim of the study was to examine the association between fine and ultrafine particles and human health. Since short–term effects of particle concentrations were the focal point, their association with blood parameters was investigated. One hundred and twelve individuals were enrolled and measured up to four times for a 5–6 hour period their individual exposure to particle number concentration (PNC) in their daily life by carrying a portable particle counter. During the same time the individuals wore ECG devices recording their cardiac rhythm activity and documented their activities in a diary. Personally measured particle number concentration (PNC) as well as ambient $PM_{2.5}$ concentrations measured at a fixed–site measurement station are associated with changes in cardiac function in individuals with metabolic disorders (Peters et al., 2015).

The mobile exposure measurements revealed two major problems: Firstly, the precision of the devices is $\pm$ 20 % as quoted by the manufacture. Secondly, about 23 % of the outdoor measurements are missing due to breakdown or incorrect appliance of the device by the study participants, i.e. the handling of mobile devices is sometimes too difficult for elderly people without technical skills. The problem of missing values becomes even more severe, if lagged exposure effects are considered. Substituting missing personal data with data from the measurement station, which are available for the complete study period, induces a mixture of classical and Berkson error, i.e. about 77 % of the observations exhibit classical measurement error and the remaining part of the observations exhibits Berkson error. For the analyses, one minute resolved data are aggregated to five minutes.

Hatch and Thomas (1993) suggest four ways for handling exposure measurement error: a) validation studies, b) replicate measurements, c) multiple types of measurements and d) sensitivity analysis. The Augsburger Umweltstudie is accompanied by a validation study and replicate measurements in order to quantify the within–subject and between–subjects variability.

Measurement error in the Augsburger Umweltstudie consists of several parts and types. Deriving information about the measurement error from repeated measurements in a linear mixed model setting is elaborate and hardly realizable, because many repetitions and time points are necessary. Information from validation studies or comparison measurements seem to be more promising. The parameters for the correction of the measurement errors are estimated with validation data as well as with data from the Augsburger Umweltstudie.

The validation study accompanying the Augsburger Umweltstudie

A follow–up study of the Augsburger Umweltstudie was conducted at the KORA study center in Augsburg to quantify the size of the measurement errors. The description of personal microenvironments and the identification of factors determining personal exposure to PNC in everyday situations was a further aim of the validation study. These analyses provide information, which can be integrated in correction methods for the exposure measurement error in the analysis of the Augsburger Umweltstudie.

The study design is visualized in Figure 1.1. Ten non–smoking volunteers (not among

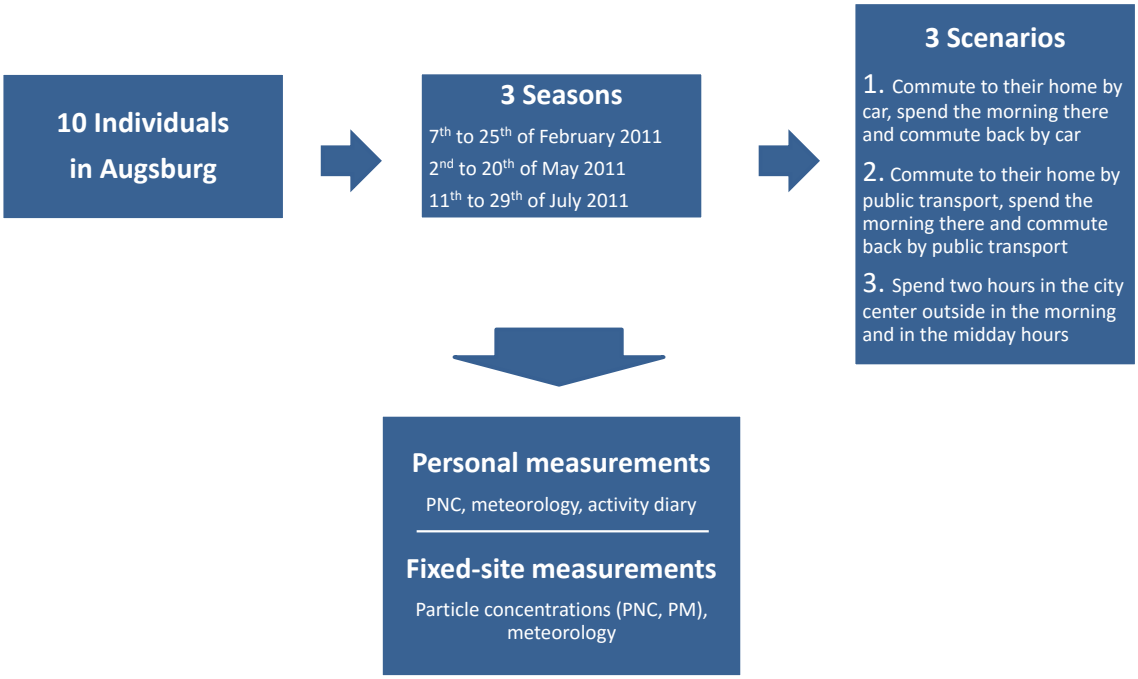


FIGURE 1.1.: *Study design of the validation study accompanying the Augsburger Umweltstudie.*

research staff or university employees) from Augsburg were recruited to accomplish personal exposure measurement campaigns in different seasons each lasting 3 weeks: from 7th to 25th of February 2011, from 2nd to 20th of May 2011 and from 11th to 29th of July 2011. On each day of the week (Monday to Friday), two volunteers conducted personal PNC measurements starting between 7 and 8 a.m. and ending between 1 and 3 p.m. at the KORA study center located in the city center of Augsburg. In each seasonal campaign, the volunteers had to follow three different scenarios for 5–6 h in optional order:

1. commute to their home by car, spend the morning there and commute back by car ("car")
2. commute to their home by public transport, spend the morning there and commute back by public transport ("public transport")
3. spend 2 h in the city center outside in the morning and in the midday hours ("city center")

Each volunteer conducted the measurements once per week and on the same weekday. The scenarios on the same day were different.

The volunteers kept activity diaries, in which they filled in their activities during the measurement phases. Every 5 min. they recorded, whether they stayed indoors or outdoors, at home or elsewhere. When being outdoors, the study participants documented, whether they were in traffic and which means of transportation they used. When staying indoors, the ventilation conditions of the room and domestic activities related to stirring up or producing particles including heating, cleaning, cooking, ironing, lighting of candles, using sprays and other "dust"–producing activities were noted. The term "dust" is casually used in the questionnaire and refers to activities, which may lift dust from ground or table to the air. Particular activities well–known to be associated with increased particulate matter (PM) exposure like passive smoking or staying in an environment with a

construction area, wood smoke, smell of food, steam, smoke or vehicle exhaust, and laser printer were also collected.

The personal PNC of each individual was measured every minute with a portable device; in addition, personal environmental conditions (temperature and relative humidity) were recorded. During the complete study phase, data on ambient air quality and environmental conditions were gathered at a site located near the University of Applied Sciences (UAS). The particles size distribution of ambient particles, total PNC, certain fractions of total PNC and concentrations of other air pollutants were measured at the UAS site as well as ambient environmental conditions (temperature, relative humidity, wind direction, wind speed, pressure radiation and rainfall).

Additional exposure measurements following a different study design were conducted within the scope of the validation study. We refer to the work of Gu et al. (2015) for further details and for results.

Comparison measurements

The three portable CPC devices were compared with each other as well as with the devices a the UAS site before and after each measurement campaign of the validation study. The comparison measurements lasted between one and nine days and reveal information about the measurement accuracy of the devices. During each period of comparison measurements, several battery changes were necessary. To conform to the temporal resolution of the main study one minute resolved data were aggregated to five minutes.

Exposure measurements

Since 2004 a monitoring station has recorded the urban background air pollution at the carefully chosen location (Cyrys et al., 2008) of the campus of the University of Applied Sciences in Augsburg. Various air pollution parameters are measured. The total particle number concentration (PNC) comprising particles with a diameter between 3 nm and 3 μm was measured with a condensation particle counter (CPC) with a resolution of 1 min. The manufacturer of the device adverts to imprecise measurements above a concentration of 10,000 particles/cm^3 and discourages from the usage of the device for higher concentrations. Moreover, certain fractions of total PNC were recorded (20–min. resolution). The particle size distribution of ambient particles (3 nm to 10 μm) was measured as well as particle mass concentrations of PM_{10} (particulate matter with an aerodynamic diameter less than 10 μm) and $PM_{2.5}$ (1–h. running means, which were updated every 5 min.) and black carbon (5–min. resolution). Quality control, quality assurance and analyses of seasonal and diurnal variations were accomplished by Pitz et al., 2008a and Pitz et al., 2008b.

Personal exposure to particulate matter is measured with portable CPCs (TSI Inc. USA, Model 3007), which cover particles ranging from 10 nm to 1 μm in their diameter. Three portable devices were used in the Augsburger Umweltudie, in the successive validation studies and for comparison measurements. The concentration accuracy is specified with $\pm$20% by the manufacturer.

Characterization of the breakdowns

In 23 % of the observations the portable devices did not record any measurement as a consequence of the breakdown of the device. Most frequently, the device broke down during staying indoors or in a car. In only 3 % of the observations the device broke down

and the person did not stay in a closed room at the same time. The breakdown of a device took 82 minutes on average.

1.4. Outline of the thesis

The key target of the thesis is the development and application of methods for taking into account exposure measurement errors inherent to ambient and personal exposure measurements in the analysis of health effects in the Augsburger Umweltstudie.

In Chapter 2, the notation is established and the models for the Augsburger Umweltstudie, the validation study and the comparison measurements are introduced. Thereby, the properties and size of Berkson error and classical measurement error in the Augsburger Umweltstudie are elucidated. A particular focus is on the development of a model for personal exposure to ultrafine particles during activities in everyday situations. The effects of individual–specific and autocorrelated classical, Berkson and mixture error of a covariate in a general simple linear mixed model are derived and illustrated by simulations in Chapter 3. The considerations are extended to regression models with several covariates and to unbalanced and non–equidistant observations. These insights enable the application of the method of moments in Chapter 4 as a simple, frequentist method for correcting classical and mixture measurement error. Especially the distributional properties of the corrected estimator are investigated. Bayesian methods for the consideration of covariate measurement error are focused in Chapter 5. Pitfalls regarding the specification of prior knowledge and of the model structure are demonstrated and guidelines for a successful composition of the model are given. The frequentist and the Bayesian approach for correcting covariate measurement error are applied to the Augsburger Umweltstudie in Chapter 6. The thesis closes with a summary of the principal findings and an outline of unsettled research questions in Chapter 7.

Parts of the Chapters 1, 2 and 7 have been published in advance in the article “Personal exposure to ultrafine particles: Two-level statistical modeling of background exposure and time-activity patterns during three seasons” (Deffner et al., 2016) in Journal of Exposure Science and Environmental Epidemiology. The ideas, the implementation and the interpretation of the analysis as well as the writing of the article is my own work. Helmut Küchenhoff was involved in the methodological considerations. The coauthors from the Helmholtz Zentrum München and the University Augsburg provided the preprocessed data, contributed to the discussion of the results and commented the manuscript.

2. Health, exposure and measurement models

Health (HM), exposure (EM) and measurement (MM) models are the essential parts of a measurement error problem (Thomas et al., 1993). In this chapter these models are defined for the data of the Augsburger Umweltstudie and the attendant validation studies. Person and time index are marked with M in the main study to distinguish the indices of the main study from the comparison measurements (C) and the validation study (V). The model definitions are related to several assumptions regarding the personal and fixed–site exposure measurements, which will be explained in the following.

2.1. Model definition

The health model is defined in Subsection 2.1.1 and describes the association between a latent covariate X (exposure to PM) and the response Y (health outcome). It is assumed that only deficient realizations of the random variable X can be measured. The assumptions regarding the associations between the unknown X and the different error–prone, but observable surrogates for X are presented in Subsection 2.1.2 including the exposure model and measurement model for the Augsburger Umweltstudie.

2.1.1. Health model

A linear mixed model with random person intercept τ_{i_M} is considered as the main outcome model for the association between the health outcome $Y_{i_M t_M}$ and the true exposure to particulate matter $X_{i_M t_M}$, confounded by the additional covariates $\boldsymbol{Z}_{i_M t_M}$ of i_M, $i_M = 1, \ldots, n$, independent individuals at time point $t_M, t_M = 1, \ldots, T_{i_M}$:

HMM: Health model for true measurements of main data

$$\begin{aligned} Y_{i_M t_M} =& \beta_0 + \beta_X X_{i_M t_M} + \boldsymbol{Z}_{i_M t_M} \boldsymbol{\beta}_Z + \tau_{i_M} + \varepsilon_{i_M t_M} \\ \boldsymbol{\varepsilon}_{i_M} \sim& N(\mathbf{0}, \boldsymbol{\Sigma}_{\varepsilon, i_M}), \boldsymbol{\Sigma}_{\varepsilon, i_M} = \sigma_\varepsilon^2 \boldsymbol{W}_{\rho, i_M} \\ \tau_{i_M} \sim& N(0, \sigma_\tau^2), \\ \boldsymbol{\theta}^H =& (\beta_0, \beta_X, \boldsymbol{\beta}_Z, \sigma_\tau^2, \sigma_\varepsilon^2, \rho)^\top \end{aligned} \tag{2.1}$$

β_0 is the intercept of the model and β_X the effect coefficient for the impact of exposure on health. $\boldsymbol{\beta}_Z$ denotes the effect coefficients of the confounder variables. In this chapter we focus only on the Augsburger Umweltstudie and therefore, subscript M is neglected for simplicity and is only used for definitions which are relevant for subsequent chapters. As a start, it is assumed that the number of observations of each individual is equal for the simplification of the considerations, i.e. $T_i = T$. Extensions to unbalanced designs are discussed in Subsection 3.5.5. A simple linear model without additional covariates $\boldsymbol{Z}$ is considered at the first glance:

$$Y_{it} = \beta_0 + \beta_X X_{it} + \tau_i + \varepsilon_{it}$$

Due to longitudinal data, the model errors $\boldsymbol{\varepsilon}_i$ are assumed to be independent and to follow an AR(1) process with autocorrelation coefficient ρ: $\boldsymbol{\varepsilon}_i \sim \mathrm{N}(0, \boldsymbol{\Sigma}_\varepsilon)$. In the following the correlation matrix of some first–order autocorrelated variable with autocorrelation coefficient ρ is denoted by $\boldsymbol{W}_\rho$:

$$\boldsymbol{W}_\rho = \begin{pmatrix} 1 & \rho & \rho^2 & \dots \\ \rho & 1 & \rho & \\ \rho^2 & \rho & 1 & \\ \vdots & & & \ddots \end{pmatrix}.$$

Thus, $\boldsymbol{\Sigma}_\varepsilon = \sigma_\varepsilon^2 \boldsymbol{W}_\rho$. For further calculations it is important, that the correlation matrix $\boldsymbol{W}_\rho$ can easily be inverted:

$$\boldsymbol{W}_\rho^{-1} = \frac{1}{1-\rho^2} \begin{pmatrix} 1 & -\rho & 0 & \cdots & 0 & 0 \\ -\rho & 1+\rho^2 & -\rho & \cdots & 0 & 0 \\ 0 & -\rho & 1+\rho^2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1+\rho^2 & -\rho \\ 0 & 0 & 0 & \cdots & -\rho & 1 \end{pmatrix}; \tag{2.2}$$

further, the following holds:

$$\sum_{i=1}^{T}\sum_{j=1}^{T} [\boldsymbol{W}_\rho^{-1}]_{ij} = \mathbf{1}_T^\top \boldsymbol{W}_\rho^{-1} \mathbf{1}_T = \frac{1}{1-\rho^2}\left[(T-2)(1-\rho)^2 + 2(1-\rho)\right]. \tag{2.3}$$

$\mathbf{1}_T$ denotes a vector of ones of length T.

2.1.2. Measurement error structures

Motivated by the measurement of the exposure to particulate matter, three types of measurement errors are evaluated:

(1) Berkson error–prone measurements $X^{*\mathrm{B}}$ are unbiased aggregated measurements of the truth; e.g., fixed–site exposure measurements do not account for the individual structure of personal exposure but the exposure of the population close to the site and represents a spatially averaged value.

(2) Classical error–prone measurements $X^{*\mathrm{C}}$ result from a noise process overlapping the true values of X; e.g., measurements of individual exposure to particulate matter recorded by portable devices exhibit a certain amount of classical measurement error according to the manufacturer of the device.

(3) Mixture error–prone measurements $X^{*\mathrm{M}}$ occur, when one part of the observations exhibits Berkson error, whereas the other part exhibits classical measurement error; e.g., missing values of individual exposure measurements with classical measurement error are substituted with the corresponding measurements from a monitoring station at a fixed site.

For our considerations, classical and Berkson errors are both allowed to comprise individual–specific effects (as it is e.g. also found in McShane et al., 2001). Autocorrelated and individual–specific Berkson error ("between person error", cf. Willett, 1989) may arise, because the exposure levels in the microenvironments of the individuals differ from the levels in the microenvironment of the monitoring site. Individual–specific handling of the portable measurement devices, as well as differing accuracy depending on the device and the environmental conditions and varying over time may yield individual–specific and

possibly autocorrelated classical measurement error. The classical measurement error of the fixed–site measurements is neglected.

The formal definition of the error structures is given by eq. 2.4–2.6.

BMM$_1$: One–stage EM for personal measurements of main data

$$\begin{aligned} X_{i_M t_M} =& X^{*\mathrm{B}}_{i_M t_M} + \nu^{\mathrm{B}}_{i_M} + U^{\mathrm{B}}_{i_M t_M} \\ \nu^{\mathrm{B}}_{i_M} \sim& N(0, \sigma^2_{\nu^{\mathrm{B}}}) \\ \boldsymbol{U}^{\mathrm{B}}_{i_M} \sim& N(\mathbf{0}, \boldsymbol{\Sigma}^{\mathrm{B}}_{i_M}), \boldsymbol{\Sigma}^{\mathrm{B}}_{i_M} = \sigma^2_{U^{\mathrm{B}}} \boldsymbol{W}_{\rho^{\mathrm{B}}, i_M} \\ \boldsymbol{\theta}^{E_{B1}} =& (\sigma^2_{U^{\mathrm{B}}}, \sigma^2_{\nu^{\mathrm{B}}}, \rho^{\mathrm{B}})^\top \end{aligned} \tag{2.4}$$

CMM: Classical measurement error model for main data

$$\begin{aligned} X^{*\mathrm{C}}_{i_M t_M} =& X_{i_M t_M} + \nu^{\mathrm{C}}_{i_M} + U^{\mathrm{C}}_{i_M t_M} \\ \nu^{\mathrm{C}}_{i_M} \sim& N(0, \sigma^2_{\nu^{\mathrm{C}}}) \\ \boldsymbol{U}^{\mathrm{C}}_{i_M} \sim& N(\mathbf{0}, \boldsymbol{\Sigma}^{\mathrm{C}}_{i_M}), \boldsymbol{\Sigma}^{\mathrm{C}}_{i_M} = \sigma^2_{U^{\mathrm{C}}} \boldsymbol{W}_{\rho^{\mathrm{C}}, i_M} \\ \boldsymbol{\theta}^{M} =& (\sigma^2_{U^{\mathrm{C}}}, \sigma^2_{\nu^{\mathrm{C}}}, \rho^{\mathrm{C}})^\top \end{aligned} \tag{2.5}$$

MMM: Mixture measurement error model for main data

$$\begin{aligned} X^{*\mathrm{M}}_{i_M t_M} &= X^{*\mathrm{B}}_{i_M t_M} G_{i_M t_M} + X^{*\mathrm{C}}_{i_M t_M} (1 - G_{i_M t_M}) \\ &= \begin{cases} X^{*\mathrm{B}}_{i_M t_M} & \text{for } p \cdot 100\% \text{ of the measurements} \\ X^{*\mathrm{C}}_{i_M t_M} & \text{for } (1-p) \cdot 100\% \text{ of the measurements} \end{cases} \end{aligned} \tag{2.6}$$

The random variable G_{it} indicates, whether the measurement at time point t of individual i exhibits Berkson error. G_{it} follows a Bernoulli distribution with success probability p and with values 0 (individual measurement available) and 1 (only fixed–site measurement available).

The Berkson error–prone measurements $\boldsymbol{X}^{*\mathrm{B}}_i = (X^{*\mathrm{B}}_{i1}, \ldots, X^{*\mathrm{B}}_{iT})^\top$ are assumed to follow a normal distribution with expectation $\mu^{X^{*\mathrm{B}}} \mathbf{1}_T$ and variance–covariance matrix $\boldsymbol{\Sigma}_{X^{*\mathrm{B}}} = \sigma^2_{X^{*\mathrm{B}}} \boldsymbol{W}_{\rho^{X^{*\mathrm{B}}}}$ (see Section 2.2 for the definition of $\boldsymbol{W}_{\rho^{X^{*\mathrm{B}}}}$). The measurement errors $\boldsymbol{U}^{\mathrm{B}}_i = (U^{\mathrm{B}}_{i1}, \ldots, U^{\mathrm{B}}_{iT})^\top$, $\boldsymbol{U}^{\mathrm{B}}_i \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}_{U^{\mathrm{B}}})$, and $\boldsymbol{U}^{\mathrm{C}}_i = (U^{\mathrm{C}}_{i1}, \ldots, U^{\mathrm{C}}_{iT})^\top, \boldsymbol{U}^{\mathrm{C}}_i \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}_{U^{\mathrm{C}}})$, denote the classical and Berkson measurement errors, which are assumed to follow autoregressive processes of order one for each independent individual, i.e. $\boldsymbol{\Sigma}_{U^{\mathrm{B}}} = \sigma^2_{U^{\mathrm{B}}} \boldsymbol{W}_{\rho^{\mathrm{B}}}$ and $\boldsymbol{\Sigma}_{U^{\mathrm{C}}} = \sigma^2_{U^{\mathrm{C}}} \boldsymbol{W}_{\rho^{\mathrm{C}}}$. The individual–specific measurement errors ν^{B}_i and ν^{C}_i are assumed to be i.i.d. normally distributed with expectation zero and constant variances $\sigma^2_{\nu^{\mathrm{B}}}$ and $\sigma^2_{\nu^{\mathrm{C}}}$. U^{B}_{it} and ν^{B}_i are independent from $X^{*\mathrm{B}}_{it}$; U^{C}_{it} and ν^{C}_i are independent from X_{it} and the measurement errors are independent of each other and of the model errors τ_i and ε_{it}.

Further, it is assumed, that the measurement errors, i.e. $\boldsymbol{\nu}^{\mathrm{B}} = (\nu^{\mathrm{B}}_1, \ldots, \nu^{\mathrm{B}}_n)^\top, \boldsymbol{\nu}^{\mathrm{C}} = (\nu^{\mathrm{C}}_1, \ldots, \nu^{\mathrm{C}}_n)^\top, \boldsymbol{U}^{\mathrm{B}} = (\boldsymbol{U}^{\mathrm{B}}_1, \ldots, \boldsymbol{U}^{\mathrm{B}}_n)^\top$ and $\boldsymbol{U}^{\mathrm{C}} = (\boldsymbol{U}^{\mathrm{B}}_1, \ldots, \boldsymbol{U}^{\mathrm{B}}_n)^\top$, are non–differential.

These assumptions involve

$$\boldsymbol{X}_i \sim \mathrm{N}\left((\mu^{X^{*\mathrm{B}}} + \nu^{\mathrm{B}}_i)\mathbf{1}_T, \boldsymbol{\Sigma}_{X^{*\mathrm{B}}} + \boldsymbol{\Sigma}_{U^{\mathrm{B}}}\right), \tag{2.7}$$

$$\boldsymbol{X}^{*\mathrm{C}}_i \sim \mathrm{N}\left((\mu^{X^{*\mathrm{B}}} + \nu^{\mathrm{B}}_i + \nu^{\mathrm{C}}_i)\mathbf{1}_T, \boldsymbol{\Sigma}_{X^{*\mathrm{B}}} + \boldsymbol{\Sigma}_{U^{\mathrm{B}}} + \boldsymbol{\Sigma}_{U^{\mathrm{C}}}\right) \tag{2.8}$$

and

$$\boldsymbol{X}_i^{*\mathrm{C}} \sim \mathrm{N}\left(\boldsymbol{X}_i^{*\mathrm{B}} + (\nu_i^{\mathrm{B}} + \nu_i^{\mathrm{C}})\mathbf{1}_T, \boldsymbol{\Sigma}_{U^{\mathrm{B}}} + \boldsymbol{\Sigma}_{U^{\mathrm{C}}}\right). \tag{2.9}$$

$\boldsymbol{\Sigma}_{X^{*\mathrm{B}}} + \boldsymbol{\Sigma}_{U^{\mathrm{B}}}$ and $\boldsymbol{\Sigma}_{X^{*\mathrm{B}}} + \boldsymbol{\Sigma}_{U^{\mathrm{B}}} + \boldsymbol{\Sigma}_{U^{\mathrm{C}}}$ are variance–covariance matrices of sums of AR(1) processes. The underlying processes are not AR(1) processes but ARMA processes (Lütkepohl, 1984).

2.2. PNC measurements of the fixed–site monitor

We characterize the distribution of the Berkson error–prone measurements $X^{*\mathrm{B}}$, i.e. the measurements of the fixed–site monitor, with the empirical distribution of the fixed–site measurements during the measurements within the Augsburger Umweltstudie usually between 7 a.m. and 3 p.m.:

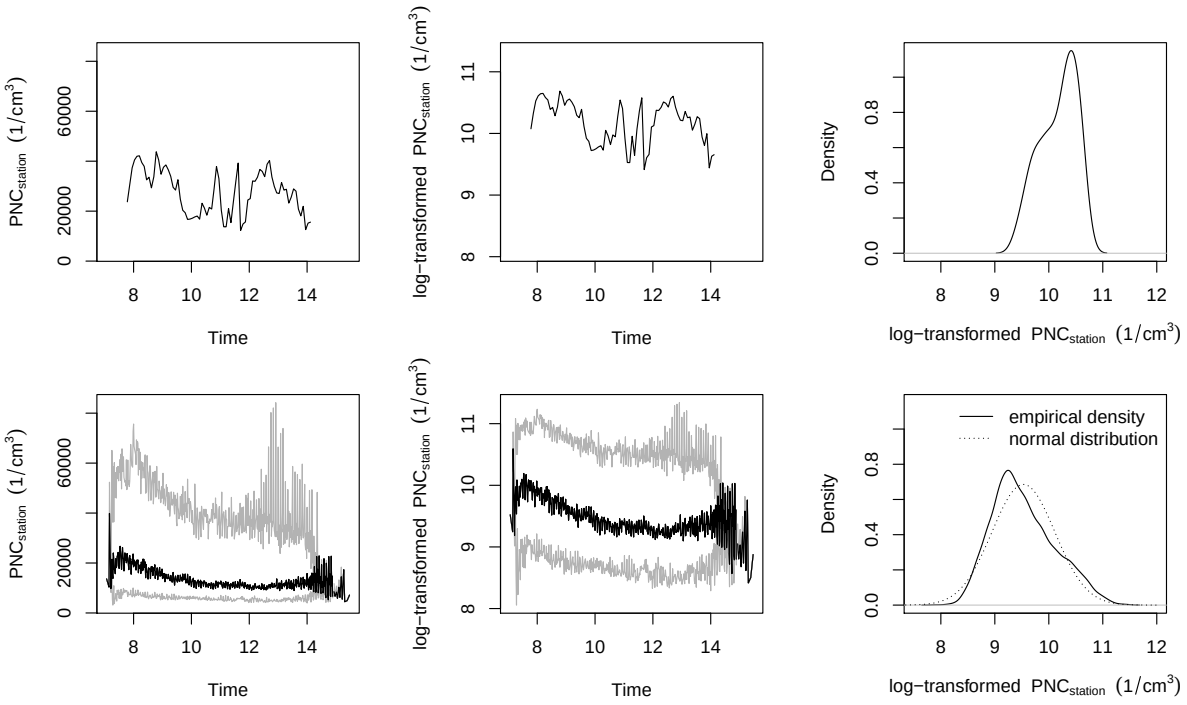


FIGURE 2.1.: *Fixed–site PNC levels during the Augsburger Umweltstudie. Upper row: exemplary raw and log–transformed time series of fixed–site PNC measurements and the density of the log–transformed fixed–site PNC levels; lower row: minute–wise median (black), 2.5 %– and 97.5 %-quantiles (gray) of fixed–site PNC measurements and density of all log–transformed fixed–site levels.*

Measurement error analysis requires assumptions concerning the distribution of the true and error–prone measurements as well as of the error itself. Since concentrations of air pollutants approximate a lognormal distribution, additive error models for the log–transformed exposure seem to be appropriate (Goldman et al., 2011), as also indicated by Figure 2.1.

Mean, variance and autocorrelation coefficient of $X^{*\mathrm{B}}$ are empirically estimated from the log–transformed fixed–site PNC measurements of the Augsburger Umweltstudie: $\widehat{\mu}^{X^{*\mathrm{B}}} = 9.53$, variance $\widehat{\sigma}^{X^{*\mathrm{B}}} = 0.34$ and autocorrelation coefficient $\widehat{\rho}^{X^{*\mathrm{B}}} = 0.93$.

BMM$^{*\mathrm{B}}$: EM for fixed–site measurements of main data

$$
\begin{aligned}
X^{*\mathrm{B}}_{i_M t_M} =& \mu^{X^{*\mathrm{B}}} + U^{*\mathrm{B}}_{i_M t_M} \\
U^{*\mathrm{B}}_{i_M} \sim& N(0, \boldsymbol{\Sigma}^{*\mathrm{B}}_{i_M}), \boldsymbol{\Sigma}^{*\mathrm{B}}_{i_M} = \sigma^2_{X^{*\mathrm{B}}} \boldsymbol{W}_{\rho^{X^{*\mathrm{B}}}, i_M} \\
\boldsymbol{\theta}^{E_{*B}} =& (\mu^{X^{*\mathrm{B}}}, \sigma^2_{X^{*\mathrm{B}}}, \rho^{X^{*\mathrm{B}}})^\top
\end{aligned}
\tag{2.10}
$$

Indeed the normality assumption for the fixed–site measurements is a strong simplification, because various factors like weather conditions, holidays or rush hours influence the PNC concentrations at the fixed–site monitor.

2.3. Classical measurement error of mobile PNC measurements

Comparison measurements of the mobile devices and the devices at the fixed–site characterize the classical measurement error of the mobile devices.

Classical measurement error of PNC measurements is heteroscedastic (Figure 2.2) and therefore multiplicative. The logarithmically transformed PNC measurements entail an additive measurement error. Since the measurement range of the fixed–site device for measuring PNC differs from the mobile devices (see Section 1.3), fixed–site PNC measurements are mostly constantly higher than PNC measured with the mobile devices; therefore, observations with high peaks in the ratio (above a ratio of 4) between fixed–site and portable devices are not involved in the measurement error calculation.

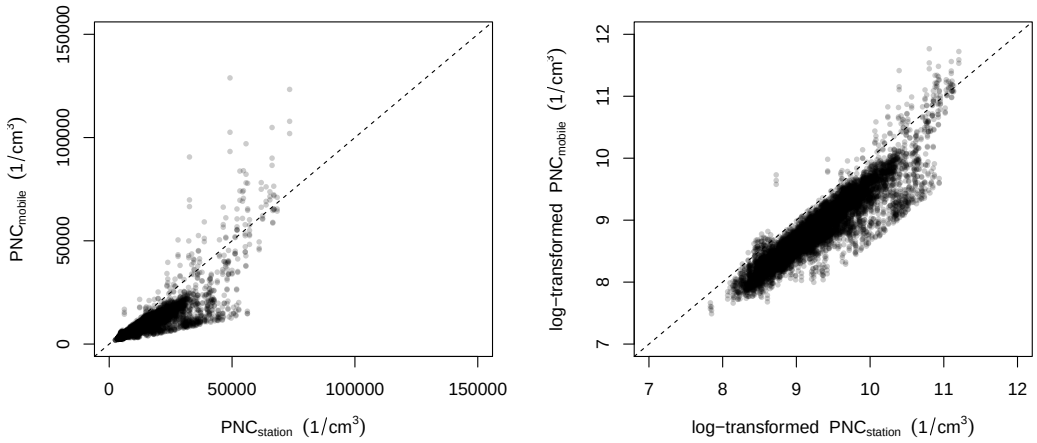


FIGURE 2.2.: *Comparison between (raw and log–transformed) PNC measurements with the portable and the fixed–site devices.*

The comparison measurements are used to estimate $\sigma^2_{U^{\mathrm{C}}}, \sigma^2_{\nu^{\mathrm{C}}}$ and ρ^{C} with a regression model for the differences between (log–transformed) values of the mobile (denoted by $\breve{X}^{*\mathrm{C}}_{jkt_C}$) and fixed–site measurement devices (denoted by $\breve{X}_{jkt_C}$):

CMC: Classical measurement error model for comparison measurements

$$
\begin{aligned}
(\breve{X}^{*\mathrm{C}}_{jkt_C} - \breve{X}_{kt_C}) =& \alpha_0^C + \tau_j^{dev} + \tau_{jk}^{devbat} + \breve{U}^C_{jkt_C} \\
\breve{U}^C_{jkt_C} \sim& \mathrm{N}(0, \sigma^2_{U^\mathrm{C}}) \\
\sigma^2_{\nu^\mathrm{C}} =& \mathrm{Var}(\breve{X}^{*\mathrm{C}}_{jkt_C} - \breve{X}_{kt_C}) - \sigma^2_{U^\mathrm{C}} \\
\widehat{\breve{U}}^C_{jkt_C} =& \alpha_1^C + \breve{V}^C_{jkt_C} \\
\breve{\boldsymbol{V}}^C_{jk} \sim& \mathrm{N}(0, \boldsymbol{\Sigma}^\mathrm{C}_{jk}), \boldsymbol{\Sigma}^\mathrm{C}_{jk} = \sigma^2_{V^\mathrm{C}} \boldsymbol{W}_{\rho^\mathrm{C},jk} \\
\boldsymbol{\omega}^\mathrm{C} =& (\alpha_0^C, \alpha_1^C, \sigma^2_{V^\mathrm{C}}, \sigma^2_{\nu^\mathrm{C}}, \sigma^2_{U^\mathrm{C}}, \rho^\mathrm{C})^\top
\end{aligned}
\tag{2.11}
$$

j is the index for three portable devices, which measure personal PNC levels, k the index for the periods between battery changes and t_C the index for the time. α_0^C and α_1^C are intercepts. The classical measurement error in this experimental setup consists of three components: 1) a random error $\breve{U}^C_{jkt_C}$, 2) an error defined by periods between the battery changes τ_{jk}^{devbat} and 3) a device error τ_j^{dev}. This error structure is modeled with a simple linear model including categorical effects for device and battery change period. ρ^C is estimated by the empirical autocorrelation of the residuals $\widehat{\breve{\boldsymbol{U}}}^\mathrm{C}$.

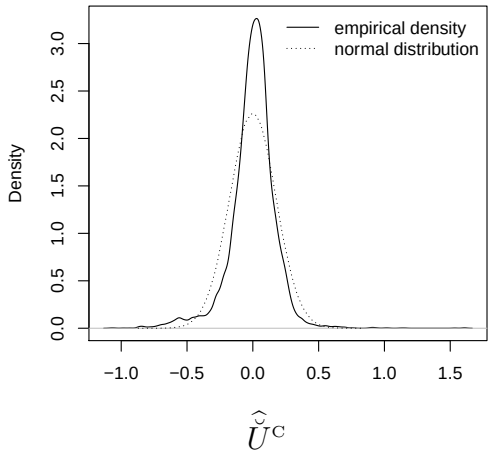


FIGURE 2.3.: *Density of $\widehat{\breve{U}}^C$ in comparison to the corresponding normal distribution.*

Estimations for parameters describing the Berkson error according to eq. 2.11 are: $\widehat{\sigma}^2_{U^\mathrm{C}} = 0.03, \widehat{\sigma}^2_{\nu^\mathrm{C}} = 0.03$ and $\widehat{\rho}^\mathrm{C} = 0.696$. 82.53 % of the residuals lie with in the boundaries announced by the manufacturer of the portable devices (±20 % for 1–min. measurements, TSI Incorporated (2016a)). We neglect the measurement error of the fixed–site PNC device (±10 % for 1–min. concentrations below 10,000 particles/cm^3, TSI Incorporated (2016b)). $\widehat{\breve{\boldsymbol{U}}}^\mathrm{C}$ is depicted in Figure 2.3.

Especially the change of batteries has a large effect on the discrepancy between the mobile and the fixed–site measurements. Considerable reasons are, that the calibration of the sensitive devices slightly changes through the change of batteries or that the exact position of the mobile devices plays a role. The variability of the measurements through the battery change and through the different measurement devices are used as a lower boundary for individual–specific classical measurement error, since this is the only way

to extract information about the individual–specific classical measurement error from our data.

2.4. Berkson error of fixed–site PNC measurements

The main objective of the analysis presented in this section is to understand the factors determining the personal exposure to PNC in everyday situations. This corresponds to a very detailed description of the Berkson error arising by using fixed–site exposure measurements instead of personal measurements. Parts of this section have been published in advance in the article "Personal exposure to ultrafine particles: Two-level statistical modeling of background exposure and time-activity patterns during three seasons" (Deffner et al., 2016) in Journal of Exposure Science and Environmental Epidemiology.

2.4.1. Current knowledge about the association between personal PNC and activities

Personal exposure to PM consists of two components according to Wilson and Brauer (2006): particles of ambient and particles of non–ambient origin. Ambient PNC levels are closely associated with meteorology (temporally) (Wehner and Wiedensohler, 2003) and locations/microenvironments (spatially). Concentrations and composition of outdoor air pollutants vary seasonally (Gómez-Moreno et al., 2011; Hussein et al., 2004; Krudysz et al., 2009; Lonati and Giugliano, 2006; Pitz et al., 2003) and diurnally (Krudysz et al., 2009; Pitz et al., 2003; Ruuskanen et al., 2001).

Hussein et al. (2004) described an inverse relationship between temperature and PNC levels. As increasing wind speed conduces to dilution, PNC decreases (Berghmans et al., 2009; Gómez-Moreno et al., 2011). In an urban area traffic intensity and traffic infrastructure strongly influence PNC levels (Berghmans et al., 2009; Cattaneo et al., 2009).

According to the observations of Sturm et al. (2003) and Zhu et al. (2002), the PNC levels at traffic or traffic–influenced sites are higher than at urban background sites, because of different physical and chemical transformations. However, PNC levels were found well correlated among urban background sites (Cyrys et al., 2008).

Indoor air quality substantially differs from outdoor air quality regarding particle composition and concentration. Indoor air quality may be particularly relevant because people spend the majority of time indoors (Brasche and Bischof, 2005; Cohen Hubal et al., 2000; Moschandreas, 1981). As the indoor PNC level is a mixture of particles from indoor as well as from outdoor sources, which may penetrate into indoors (Franck et al., 2003). Meteorological conditions, the building ventilation behavior and the air exchange rate are important factors determining indoor exposure to ambient particles. The majority of non–ambient ultrafine particles originate from indoor sources, for example from human activities (Morawska et al., 2003). In particular, food preparation, heating (particularly with wood), active/passive smoking or other domestic combustion processes like candle burning are known particle–producing activities (He et al., 2004) and are the major sources for residential exposure to PM (Bekö et al., 2013).

2.4.2. Two–stage statistical modeling

The shape as well as the size of the impact of ambient exposure to PNC on personal exposure varies temporarily as well as with the location/microenvironment of the individual. Also the predictor set may change, i.e. the relevance of the single size fractions of PM or

of the meteorological parameters for the personal PNC level may underlie temporal and spatial variations.

Therefore, the modeling procedure for the explorative analysis of the association of personal exposure to PNC with activites, fixed–site PM levels and meteorological conditions is divided into two steps. Firstly, season–specific personal background exposure to PNC is modeled with simultaneous variable selection separately for indoor and outdoor observations (as e.g. in Gerharz et al. (2013)) using all available quantities of the fixed–site monitor without any preselection ("background model"); in the second step, the association between personal exposure to PNC and the activities is evaluated adjusted for background exposure ("activity model"). This two–stage approach provides the advantage that the effects of the background pollution and the effects of the activities are clearly separated.

Fixed–site measurement with a resolution below 1 min., are carried forward to fit the 1–min. resolution of the personal measurement devices.

2.4.3. Background exposure model

Semi–parametric regression models with a random person intercept are applied for the background model using observations without any of the activities examined in the second stage; for the outdoor background models only observations during being in transit as pedestrian are used:

$$\mathrm{PNC}_{it} = \boldsymbol{X}_{it}\boldsymbol{\beta} + f_1(z_{it1}) + \cdots + f_q(z_{itq}) + \tau_i + \varepsilon_{it} \quad ,$$

with $t = 1, \ldots, T, i = 1, \ldots, n$. The random person intercept is denoted by τ_i. $f_1, \ldots, f_q$ represent smooth functions.

Air quality parameters from the fixed–site monitor, personally measured meteorology data and time of the day are smoothly modeled with P–splines; the models are further adjusted for deterministic time covariates including week, day of the week and time of the day, and categorical covariates describing the environment of the individuals (indoor models: being at home/not at home, window open/tilted/closed; outdoor models: none). The detailed model equations can be found in Appendix A.1.

Variable selection is conducted by means of model–based boosting (Bühlmann and Hothorn, 2007) using the total set of the collected covariates. Boosting algorithms estimate coefficients by iteratively approximating the solution with small steps and by using weak learners. Adding up the coefficients of a specific covariate of all iteration steps results in the common regression coefficients (if the number of coefficients is smaller than the number of observations). As usual, these coefficients can be interpreted as conditional effects adjusted for the other covariates included in the model. This approach offers three advantages: (I) only the relevant covariates are included in the model (variable selection), (II) the multicollinearity problem due to the highly correlated exposure quantities at the fixed–site monitor is avoided in contrast to the usual regression analysis and (III) the combination of weak learners, so–called base–learners, for linear, smooth and random effects allows the flexible modeling of the underlying structure in the data.

For each season observations with extreme personal PNC levels (above their 90 % quantile) or extreme particle concentration levels measured at the fixed monitoring site (above their 95 % quantile) are omitted from the calculation of the background models. The reason is that these extreme values may have been possibly evolved from particular situations, which had no bearing on the modeled association, for example, not recorded particle–producing activities or heavy traffic near the measurement station. Focusing on the scope of data with sufficient information ensures the reliability of the model within

this scope. Deficiently recorded particle concentrations (e.g., negative levels) are excluded from the analysis. Separate models for each season are calculated accounting for seasonal variation of the included predictors.

2.4.4. Relevance of variable groups in a linear regression model

Variable importance in linear regression was broadly discussed in the literature and many approaches were proposed (see e.g. Bi (2012) for an overview). The relevance of the following predictor groups in the background model is evaluated:

- Individual
- Time: season, week, day of the week, time of the day
- Meteorology (personally measured): temperature, relative humidity, dew point temperature
- Categorical surrounding variables (only for indoor models): being at home/not at home, status of the windows
- PM/NC/BC (UAS): particle mass and number concentrations of various size ranges and black carbon concentration measured at the fixed monitoring site

The relevance of a predictor group is assessed by the fraction of variance explained by the predictor group, which will be explained in the following.

Let

$$\boldsymbol{X} = (X_1, X_2, \ldots, X_p)$$

denote the predictors in a linear regression model for the outcome variable Y:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p + \varepsilon.$$

Further, it is assumed that the set of the predictors $\boldsymbol{X}$ can be divided into n uncorrelated subgroups

$$\boldsymbol{X}^{G_1} = (X_1^{G_1}, \ldots, X_{g_1}^{G_1}), \boldsymbol{X}^{G_2} = (X_1^{G_2}, \ldots, X_{g_2}^{G_2}), \ldots, \boldsymbol{X}^{G_n} = (X_1^{G_n}, \ldots, X_{g_n}^{G_n})$$

with group indices $G_1, G_2, \ldots, G_n$.

The marginal variance model, as in Grömping (2007), is given by

$$\begin{aligned}
\mathrm{Var}(Y) &= \mathrm{Var}(\boldsymbol{X}\boldsymbol{\beta}) + \mathrm{Var}(\varepsilon) \\
&= \mathrm{Var}((X_1^{G_1}, \ldots, X_{g_1}^{G_1}, X_1^{G_2}, \ldots, X_{g_2}^{G_2}, \ldots, X_1^{G_n}, \ldots, X_{g_n}^{G_n})^\top \boldsymbol{\beta}) + \mathrm{Var}(\varepsilon) \\
&= \sum_{i=1}^{n} \beta_i^2 \mathrm{Var}(\boldsymbol{X}^{G_i}) + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \beta_i \beta_j \mathrm{Cov}(\boldsymbol{X}^{G_i}, \boldsymbol{X}^{G_j}) + \mathrm{Var}(\varepsilon)
\end{aligned}$$

The fraction of variance explained by a specific predictor group $\boldsymbol{X}^{G_k}$ is determined through

$$R_{\mathrm{frac}}^{G_k\,2} = \frac{1}{\widehat{\mathrm{Var}}(Y)} \left[\sum_{i=1}^{g_k} \widehat{\beta}_i^2 \widehat{\mathrm{Var}}(X_i^{G_k}) + 2 \sum_{i=1}^{g_k - 1} \sum_{j=i+1}^{g_k} \widehat{\beta}_i \widehat{\beta}_j \widehat{\mathrm{Cov}}(X_i^{G_k}, X_j^{G_k}) \right].$$

Thus, the ${R^{G_k}_{\text{frac}}}^2$ is given by the sum of the corresponding block of the variance–covariance matrix of the $\boldsymbol{X\beta}$, divided by $\text{Var}(Y)$ and

$$R^2 = \sum_{i=1}^{n} {R^{G_i}_{\text{frac}}}^2$$

Usually, the variable groups are correlated. Thus, $\text{Var}(\boldsymbol{X\beta})$ is not block–diagonal and the entries of $\text{Var}(\boldsymbol{X\beta})$ not belonging to the blocks defined by the variable groups are not equal to zero. The sum of all these entries, divided by $\text{Var}(Y)$ is denoted by ${R^{\text{cov}}_{\text{frac}}}^2$ and contains that part of the explained variance, which is explained through predictors from different variable groups. For correlated predictor groups, R^2 can be fractionized as follows:

$$R^2 = \sum_{i=1}^{n} {R^{G_i}_{\text{frac}}}^2 + {R^{\text{cov}}_{\text{frac}}}^2$$

Many approaches to split ${R^{\text{cov}}_{\text{frac}}}^2$ and to assign the fractions to the single covariates (Bi, 2012; Grömping, 2007) or in this case, to single groups, are proposed in the literature. However, the properties, benefits and drawbacks of the methods are not completely understood and there exists no clear recommendation towards one of the methods. Therefore, ${R^{\text{cov}}_{\text{frac}}}^2$ is not further split in the presented analysis. Especially, in the case of uncorrelated or slightly correlated groups, ${R^{\text{cov}}_{\text{frac}}}^2$ should be small.

This relation is used to evaluate the relevance of the predictor groups in the background model. The application of the fractional importance values for smoothly modeled covariates using P–splines is straight forward.

2.4.5. Modeling categorical effects in time series

Particle producing activities yield a gradual instead of an abrupt change of PNC. The identification of categorical activity effects in time series data with two commonly used approaches and the attending drawbacks are examined in this subsection. The inclusion of a highly flexible smooth trend for the small–scaled temporal structure is considered as the first approach to allow for the autocorrelation in the data. A general linear model with explicit modeling of the correlation structure through an iterative estimation process is the second approach.

The following simple data example will be used for the theoretical considerations: Let $\boldsymbol{X}$ be the design matrix of a time series regression problem for response $\boldsymbol{Y} = (Y_1, \ldots, Y_T)^\top$ with only an intercept and one categorical, binary predictor, e.g. an activity of duration n_a:

$$\boldsymbol{X} = \begin{pmatrix} 1 & \cdots & 1 & 1 & \cdots & 1 & 1 & \cdots & 1 \\ 0 & \cdots & 0 & 1 & \cdots & 1 & 0 & \cdots & 0 \end{pmatrix}^\top$$

Suppose further for simplicity, that for the first and the last two observations of the time series, the reference category of the activity covariate is observed, that the activity is conducted without breaks (starting at time point t_{a_1} and ending at time point t_{a_S}) and that $n_a > 2$.

Smooth time trend

For this approach, temporal autocorrelation of $Y|X$ is modeled through a smooth function f, represented through P–Splines, depending on time t assuming $\varepsilon_t \sim \text{N}(0, 1)$:

$$Y_t = \beta_0 + \beta_1 X_t + f(t) + \varepsilon_t.$$

To prevent the modeling of the activity effect through $f(t)$, the knots are placed only at time points without any activity and first order differences of the coefficients for the B–Splines representing f are penalized. Dense knots at e.g. every third observation allow for small–scale autocorrelation. The weaknesses of this approach are, that the autocorrelation within the activities will be insufficiently removed and the gradual change of the PNC levels between different activities is not adequately considered.

General linear model

The general simple linear model assuming $\boldsymbol{\varepsilon} \sim \mathrm{N}(0, \boldsymbol{\Sigma})$ is defined as

$$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t, \tag{2.12}$$

with $t = 1 \ldots, T$ and $\boldsymbol{\Sigma} = \sigma_\varepsilon^2 \boldsymbol{W}_\rho$. Effect estimates for $\boldsymbol{\beta} = (\beta_0, \beta_1)^\top$ can be received by the weighted least squares method:

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \boldsymbol{W}_\rho^{-1} \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{W}_\rho^{-1} \boldsymbol{y}.$$

With

$$\boldsymbol{X}^\top \boldsymbol{W}_\rho^{-1} \boldsymbol{X} \overset{\text{eq. 2.3}}{=} \frac{1}{1-\rho^2} \begin{pmatrix} (T-2)(1-\rho)^2 + 2(1-\rho) & n_a(1-\rho)^2 \\ n_a(1-\rho)^2 & (n_a-2)(1-\rho)^2 + 2(1+\rho^2-\rho) \end{pmatrix}$$

and

$$\begin{aligned} |\boldsymbol{X}^\top \boldsymbol{W}_\rho^{-1} \boldsymbol{X}| =& \frac{1}{1-\rho^2} \Big\{ \Big[2(1-\rho) + (T-2)(1-\rho)^2 \Big] \cdot \\ & \Big[(n_a-2)(1-\rho)^2 + 2(1+\rho^2-\rho) \Big] - n_a^2 (1-\rho)^4 \Big\} =: \frac{1}{1-\rho^2} D^{-1} \end{aligned}$$

the effect estimate is given by

$$\begin{aligned} \widehat{\boldsymbol{\beta}} =& D \cdot \\ & \begin{pmatrix} (n_a-2)(1-\rho^2) + 2(1+\rho^2-\rho) & -n_a(1-\rho)^2 \\ -n_a(1-\rho)^2 & (T-2)(1-\rho)^2 + 2(1-\rho) \end{pmatrix} \cdot \\ & \begin{pmatrix} (1-\rho)(Y_1 + Y_T) + (1-\rho)^2 \sum_{t=2}^{T-1} Y_t \\ -\rho(Y_{t_{a_1}-1} + Y_{t_{a_S}+1}) + (1+\rho^2-\rho)(Y_{t_{a_1}} + Y_{t_{a_S}}) + (1-\rho)^2 \sum_{t=t_{a_1}+1}^{t_{a_S}-1} Y_t \end{pmatrix}. \end{aligned}$$

Thus, $\widehat{\beta}_0$ and $\widehat{\beta}_1$ are calculated as follows:

$$\begin{aligned} \widehat{\beta}_0 =& D \cdot (1-\rho) \cdot \\ & \Bigg[\Big((n_a-2)(1-\rho)^2 + 2(1+\rho^2-\rho) \Big) \left((Y_1 + Y_T) + (1-\rho) \sum_{t=2}^{T-1} Y_t \right) - \\ & n_a(1-\rho) \left(-\rho(Y_{t_{a_1}-1} + Y_{t_{a_S}+1}) + (1+\rho^2-\rho)(Y_{t_{a_1}} + Y_{t_{a_S}}) + (1-\rho)^2 \sum_{t=t_{a_1}+1}^{t_{a_S}-1} Y_t \right) \Bigg] \end{aligned}$$

and

$$\begin{aligned} \widehat{\beta}_1 =& D \cdot (1-\rho) \cdot \\ & \Bigg[-n_a(1-\rho) \left((1-\rho)(Y_1 + Y_T) + (1-\rho)^2 \sum_{t=2}^{T-1} Y_t \right) + \\ & (2 + (T-2)(1-\rho)) \\ & \left(-\rho(Y_{t_{a_1}-1} + Y_{t_{a_S}+1}) + (1+\rho^2-\rho)(Y_{t_{a_1}} + Y_{t_{a_S}}) + (1-\rho)^2 \sum_{t=t_{a_1}+1}^{t_{a_S}-1} Y_t \right) \Bigg] \end{aligned}$$

$\widehat{\beta}_0$ and $\widehat{\beta}_1$ are weighted sums of $\boldsymbol{Y}$. The weights for each observation are listed in Table 2.1:

	Weights for Y_t
Weights for $\widehat{\beta}_0$	
$t \in \{1, T\}$	$D(1-\rho)\left((n_a-2)(1-\rho)^2+ 2(1+\rho^2-\rho)\right)$
$t \in \{2, \ldots, t_{a_1}-2, t_{a_S}+2, \ldots, T-1\}$	$D(1-\rho)^2\left((n_a-2)(1-\rho)^2+ 2(1+\rho^2-\rho)\right)$
$t \in \{t_{a_1}-1, t_{a_S}+1\}$	$D(1-\rho)^2\left((n_a-2)(1-\rho)^2+ 2(1+\rho^2-\rho)+\rho n_a\right)$
$t \in \{t_{a_1}, t_{a_S}\}$	$D(1-\rho)^2\left((n_a-2)(1-\rho)^2+ 2(1+\rho^2-\rho)-(1+\rho^2-\rho)n_a\right)$
$t \in \{t_{a_1}+1, \ldots, t_{a_S}-1\}$	$D(1-\rho)^2\left((n_a-2)(1-\rho)^2+ 2(1+\rho^2-\rho)-(1-\rho)^2 n_a\right)$
Weights for $\widehat{\beta}_1$	
$t \in \{1, T\}$	$D\left(-n_a(1-\rho)^3\right)$
$t \in \{2, \ldots, t_{a_1}-2, t_{a_S}+2, \ldots, T-1\}$	$D\left(-n_a(1-\rho)^4\right)$
$t \in \{t_{a_1}-1, t_{a_S}+1\}$	$D\left(-n_a(1-\rho)^4- \rho(2(1-\rho)+(T-2)(1-\rho)^2)\right)$
$t \in \{t_{a_1}, t_{a_S}\}$	$D\left(-n_a(1-\rho)^4+ (1+\rho^2-\rho)(2(1-\rho)+(T-2)(1-\rho)^2)\right)$
$t \in \{t_{a_1}+1, \ldots, t_{a_S}-1\}$	$D\left(-n_a(1-\rho)^4+ (1-\rho)^2(2(1-\rho)+(T-2)(1-\rho)^2)\right)$

TABLE 2.1.: *Weights for Y_t, $t = 1, \ldots, T$ for the calculation of $\widehat{\beta}_0$ and $\widehat{\beta}_1$.*

For uncorrelated data ($\rho = 0$), each observation of the response variable has the same influence on the estimation, i.e. they are equally weighted. With increasing autocorrelation the observations before and after an activity change become more influential, whereas all other observations become less relevant (see first row of Figure 2.4). For $\rho \to 1$, the following holds:

$$\lim_{\rho \to 1} \widehat{\beta}_0 = \frac{1}{2}(Y_1 + Y_T)$$
$$\lim_{\rho \to 1} \widehat{\beta}_1 = \frac{1}{2}(Y_{t_{a_1}} + Y_{t_{a_S}}) - \frac{1}{2}(Y_{t_{a_1}-1} + Y_{t_{a_S}+1}).$$

The disadvantage thereof is, that conducting a certain activity does not yield to an abrupt change of PNC levels to another level, but the change is gradually. In addition, the start of an activity is often not precisely enough documented through the activity dictionary. High autocorrelation in a model with a single overall autocorrelation coefficient will result in an activity effect, which is mainly based on the observations before and after the activity and on the first and last observations of the activity. Thus, the effect estimate represents only the PNC change at the beginning and at the end of the activity, which does not represent the actual effect of the activity in the case of a gradual change.

Each temporal coherent period with the same activity (or non–activity) of an individual is called the realization of an activity. Assuming a block–diagonal structure for $\boldsymbol{\Sigma}$, i.e. Y_{t_i} and Y_{t_j} $(t_i \neq t_j)$ are assumed to be uncorrelated, if they belong to different activities or different realizations of an activity, remedies from biased effect coefficients for moderate autocorrelation as shown in the second row of Figure 2.4:

The attendant calculations of the effect coefficients can be found in Appendix B.1. The influence of the observations directly before and after the activity changes ($t \in \{t_{a_1} - 1, t_{a_1}, t_{a_S}, t_{a_S} + 1\}$) on $\widehat{\beta}_1$ strongly increases only with high autocorrelation, when a block–diagonal structure is assumed for $\mathbf{\Sigma}$. Independently of the size of the autocorrelation coefficient, the observations during the activity do not have any impact on the estimation of the intercept. Note, that the line for $t \in \{t_{a_1} - 1, t_{a_S} + 1\}$ is overplotted by the line for $t \in \{1, T\}$ in both subfigures and the line for $t \in \{t_{a_1} + 1, \ldots, t_{a_S} - 1\}$ is overplotted by the line for $t \in \{t_{a_1}, t_{a_S}\}$ in the subfigure for $\widehat{\beta}_0$. Nevertheless, $\widehat{\boldsymbol{\beta}}$ is mainly determined through the observations at the activity margins if autocorrelation is very high:

$$\lim_{\rho \to 1} \widehat{\beta}_0 = \frac{1}{4}(Y_1 + Y_{t_{a_1}-1} + Y_{t_{a_S}+1} + Y_T)$$
$$\lim_{\rho \to 1} \widehat{\beta}_1 = \frac{1}{2}(Y_{t_{a_1}} + Y_{t_{a_S}}) - \frac{1}{4}(Y_1 + Y_{t_{a_1}-1} + Y_{t_{a_S}+1} + Y_T)$$

In addition to the observations directly before and after the activity, also the first and the last observation of the time series are relevant.

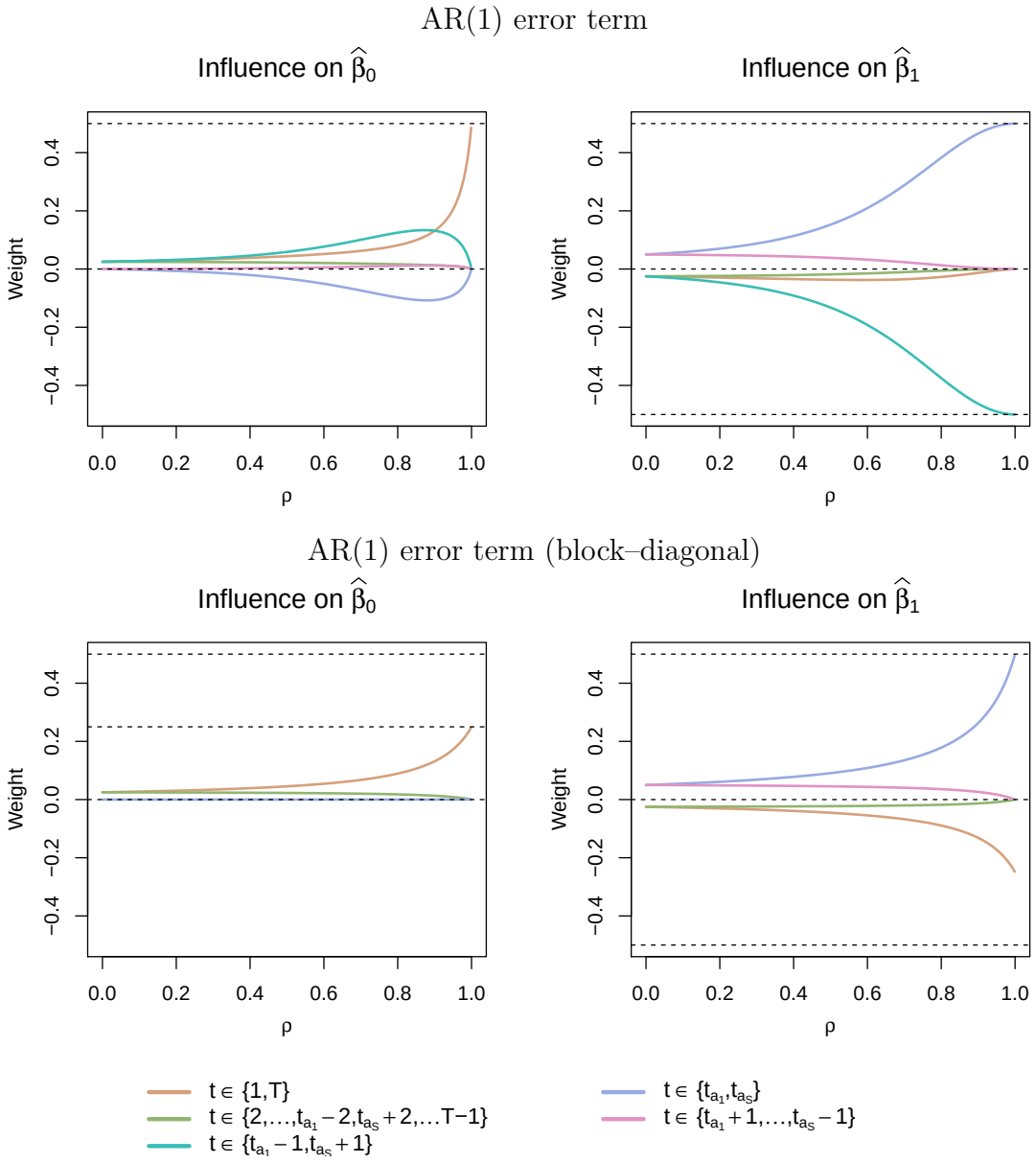


FIGURE 2.4.: *Relationship between the autocorrelation coefficient and the effect estimates in a linear model with a categorical covariate assuming an AR(1) process for the model errors. upper row:* $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{W}_\rho$*; lower row: a block–diagonal structure of the correlation matrix of the error term defined by the activity is additionally assumed:* $\boldsymbol{\Sigma} = \sigma^2 diag(\boldsymbol{W}_{\rho,1}, \boldsymbol{W}_{\rho,2}, \boldsymbol{W}_{\rho,3})$.

Further concepts for modeling the correlation structure

Other correlation structures are considered to overcome the shortcomings of the AR(1) process, discussed in the previous paragraphs. Despite the modeling of the underlying autocorrelation structure, the requirements for the correlation structure are: flexibility, few parameters and inclusion of all observations in the estimation of the effect. The latter requirement is essential for the modeling of the gradual change of the exposure levels in succession of an activity change. Several correlation functions are proposed in the literature including the spherical, exponential, Gaussian, linear, rational quadratic or Matérn function (see e.g. Pinheiro and Bates, 2000), pages 226-239 or Fahrmeir et al. (2007), pages 328-330, for a description). In contrast to higher order autoregressive processes, these stationary correlation functions get along with only a few parameters. The following correlation functions are taken into account Pinheiro et al. (2013):

Rational quadratic: $\rho(d,r,n) = \dfrac{1-n}{1+(\frac{d}{r})^2}$

Exponential: $\rho(d,r,n) = (1-n)\exp\left(-\dfrac{d}{r}\right)$

Gaussian: $\rho(d,r,n) = (1-n)\exp\left(-\left(\dfrac{d}{r}\right)^2\right)$

Linear: $\rho(d,r,n) = (1-n)\left(1-\dfrac{d}{r}\right)$ for $d<r$, 0 otherwise

Spherical: $\rho(d,r,n) = (1-n)\left(1-1.5\dfrac{d}{r}+0.5\left(\dfrac{d}{r}\right)^3\right)$ for $d<r$, 0 otherwise

The nugget is denoted by n, the distance between the time points by d and the range by $r(r>0)$. For an arbitrary variance–covariance matrix of the error term, the weights for β_0, $w_t^{\beta_0}$, and for β_1, $w_t^{\beta_1}$, $t=1,\ldots,T$, are defined by

$$w_t^{\beta_0} = [(\boldsymbol{X}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}(\boldsymbol{X}^\top\boldsymbol{\Sigma}^{-1})]_{1,t}$$

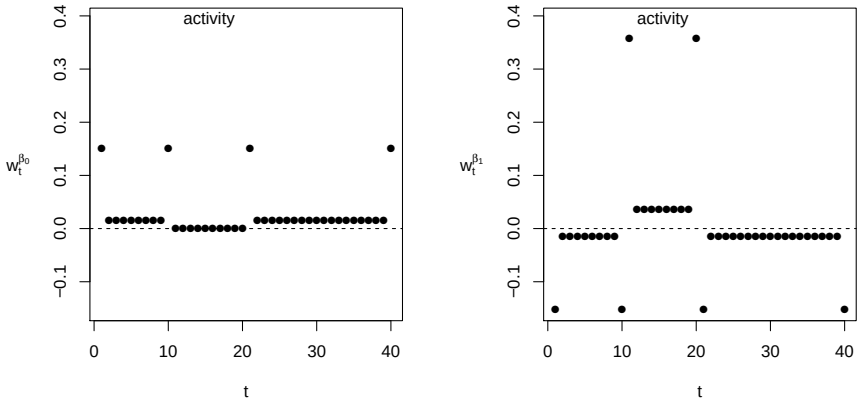
and

$$w_t^{\beta_1} = [(\boldsymbol{X}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}(\boldsymbol{X}^\top\boldsymbol{\Sigma}^{-1})]_{2,t}.$$

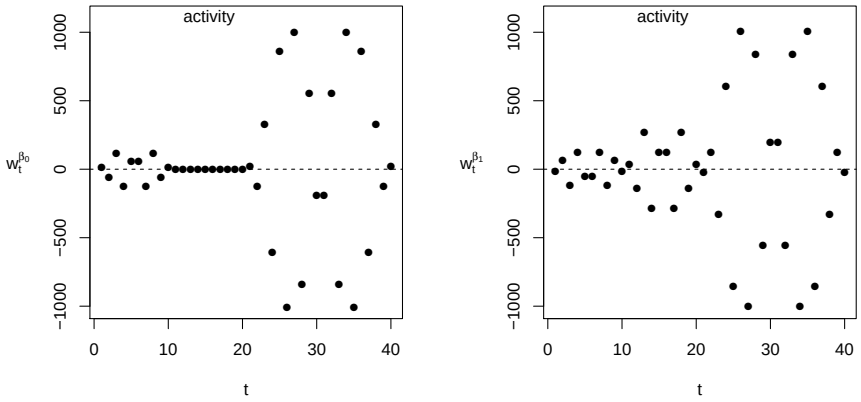
Obviously, the nugget n does not affect the weights. We examine $w_t^{\beta_0}$ and $w_t^{\beta_1}$ for the correlation functions mentioned above using the model defined in eq. 2.12 with a block–diagonal variance–covariance matrix for the error term, , $t=1,\ldots,T$, $T=40$, $t_{a_1}=11$, $t_{a_S}=20$, $n=0.1$ and r=10; $\rho=0.9$ is chosen for the AR(1) structure. The results are depicted in the Figures 2.5 and 2.6.

The correlation functions behave rather differently regarding the weighting of single observations in the weighted least squares estimation. In contrast to the usage of an autoregressive process of order one or one of the remaining correlation structures, all observations will be relevant for the estimation of the activity effect coefficient, when the rational quadratic correlation structure is used. Thus, gradual changes of the PNC levels after an activity change are modeled more adequately.

(A) AR(1)



(B) Rational quadratic



(C) Exponential

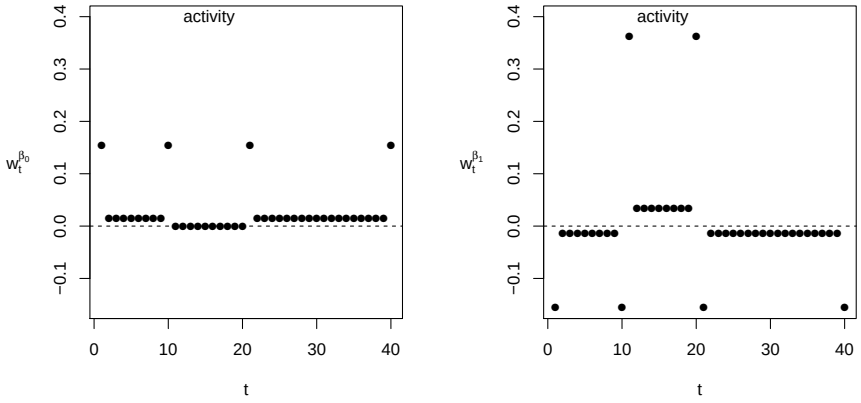
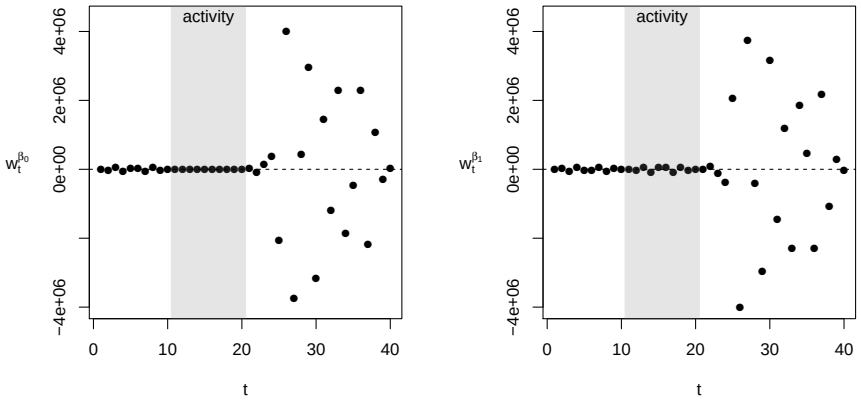
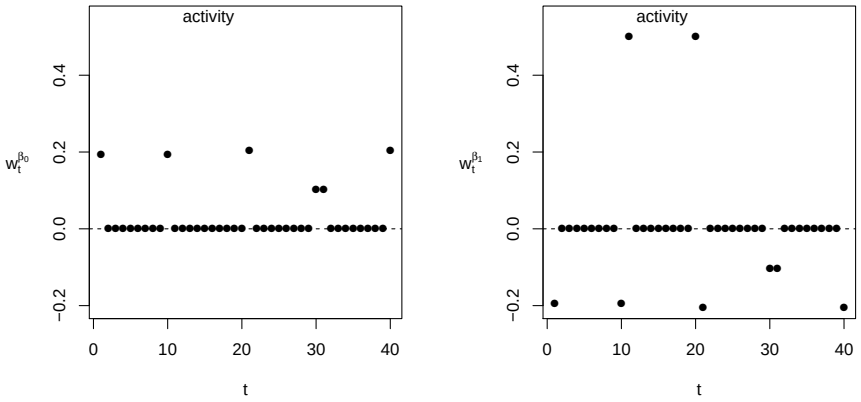


FIGURE 2.5.: *Weights for the observations in the estimation of the intercept and the effect coefficient in a simple general linear model with a binary covariate and different correlation structures for the error term: (A) AR(1), (B) Rational quadratic, (C) Exponential.*

(D) Gaussian



(E) Linear



(F) Spherical

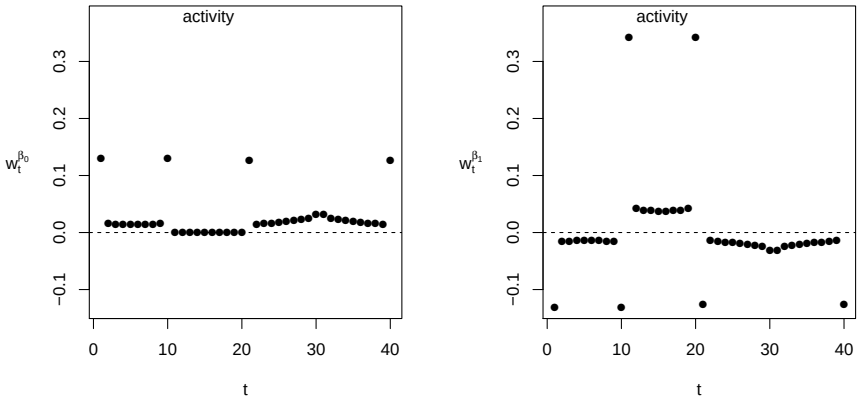


FIGURE 2.6.: *Weights for the observations in the estimation of the intercept and the effect coefficient in a simple general linear model with a binary covariate and different correlation structures for the error term: (D) Gaussian, (E) Linear, (F) Spherical.*

2.4.6. Modeling categorical effects in highly autocorrelated longitudinal data: a comparison with real data

Different modeling strategies for categorical effects in longitudinal data, which were theoretically outlined in the previous subsection, are evaluated in this subsection with an example data set for exposure and activity data. Since the data generating process is very complex, the modeling approaches are compared using real data instead of simulated data. A small data set is used in order to explain the characteristics of models in combination with a visual imagination of the data. Three exemplary time series from the validation study, depicted in Figure 2.7, are chosen, one without any activity and two further time series with cooking and candle lighting.

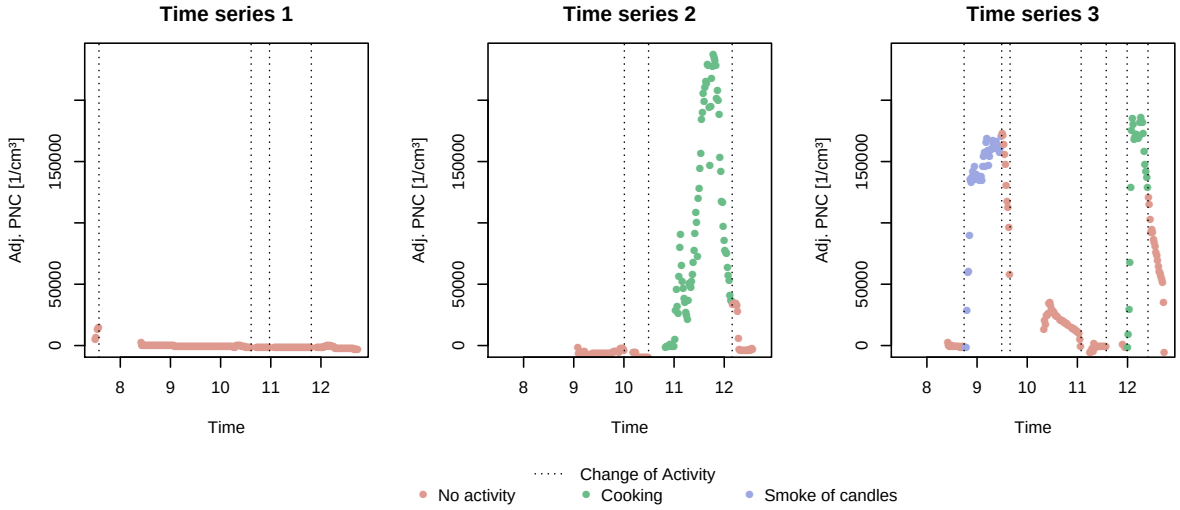


FIGURE 2.7.: *Three exemplary time series of personal exposure to 1–min. PNC (in* $1/cm^3$*), adjusted for background exposure. The vertical lines mark activity changes.*

Since the model contains only categorical covariates, Figure 2.7 leads to suspect heteroscedastic residuals, i.e. the variance of the residuals within a realization of an activity varies between the different realizations. Also realization–specific correlation structures are considerable. Furthermore, categorical effects in highly autocorrelated continuous time series may be biased, if the exposure levels gradually change after a change of the activity and if an AR(1) process is assumed for the error term, because mainly the marginal observations of the activities are used to calculate the effect (see Subsection 2.4.6). The following models are evaluated concerning their coefficient estimates, model fit and compliance of the assumptions using the example data:

- M1: Linear mixed model (M1a) and additional realization–specific variance (M1b)
- M2: General linear mixed model with AR(1) structure for the error term (M2a) and additional realization–specific variance (M2b)
- M3: General linear mixed model with AR(1) structure for the error term with realization–specific correlation coefficients (M3a) and additional realization–specific variance (M3b)
- M4: General linear mixed model with rational quadratic correlation structure for the error term (M4a) and additional realization–specific variance (M4b)
- M5: Additive mixed models with few (every 10–th observation), (M6a) and many (every third observation) (M6b) knots at time points without any activity

The exposure to PNC during single realizations of activities is assumed to be independent resulting in a block–diagonal correlation structure; the blocks are determined through the realizations of the activities. The autocorrelation coefficient as well as the nugget and the range of the correlation functions are simultaneously estimated with the effect coefficients. Also realization–specific correlation structures are considered, because the level of autocorrelation may vary depending on the way of conducting a certain activity. A realization–specific random intercept is included in all models.

If the model is correctly specified, the linear mixed model M1 provides consistent estimates of the effect coefficients. Therefore, models accounting for heteroscedastic and autocorrelated residuals should provide similar estimates for the coefficients. Further reference criteria for the evaluation of the different approaches refer to the homoscedasticity (Breusch–Pagan test, Breusch and Pagan (1979)) and independence of the residuals . In addition, the models are compared concerning their fit via the AIC. The results of the model comparison are depicted in Figure 2.8.

The linear mixed model (M1a) with random person intercept shows a bad model fit and strong violations of the model assumptions regarding highly autocorrelated and heteroscedastic residuals, which is only partially eliminated through the realization–specific residual variances (M1b). Including an overall as well as a realization–specific correlation structure (M2 and M3) fractionally alleviates the problem of autocorrelation and improves the model fit, but does not identify the visually apparent effects of the activities due to the properties of the model discussed in chapter 2.4.5. The models with an autoregressive process of order one for the error term exhibit a strong bias in one or even both activity effect estimates of about 40,000 particles. Instead, using the rational quadratic correlation structure (M4) provides adequate estimates for the coefficients, a comparably good model fit and meets the model assumptions. Smoothing splines (M5) for modeling the temporal dependencies are not able to remove the autocorrelation of the residuals; furthermore, the AIC of these models is high. Model M4b shows the best properties and is used for the further analyses.

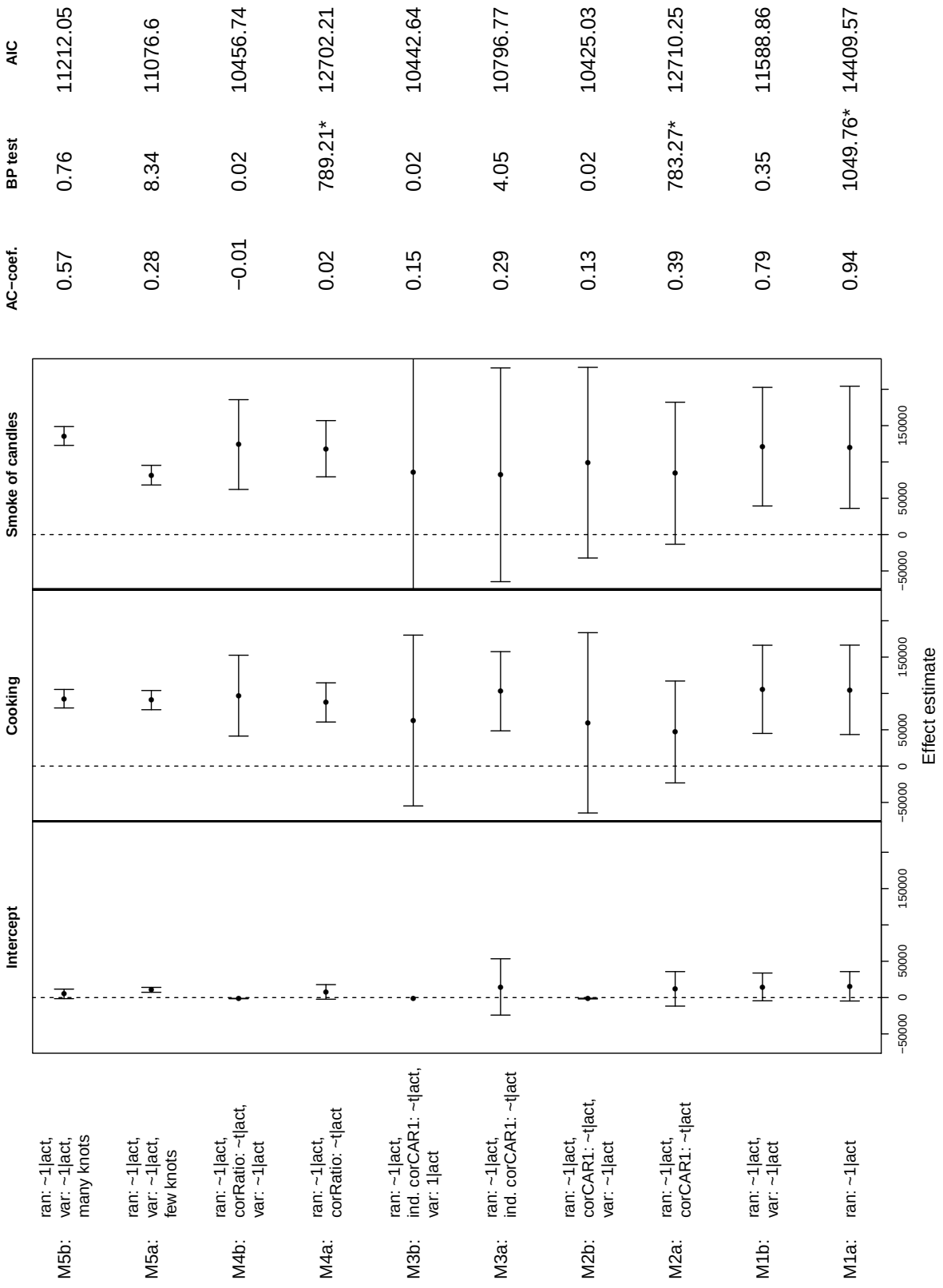


FIGURE 2.8.: *Model comparison regarding effect coefficients, residual autocorrelation (AC–coef.), heteroscedasticity (BP test, asterisk mark significance) and model fit (AIC).*

2.4.7. Activity model

In order to quantify the effect size of the collected activities (smell of smoke/food/vehicle exhaust, construction area, laser printer, steam, spray, ironing, cleaning, cooking, smoke of candles/wood, passive smoking, other recorded "dust"–producing activities) and the means of transport on personal PNC in the second modeling stage, regression models with random effects are calculated. The residuals from the first–stage models are used as outcome variable and the activities and seasons as predictors:

$$\text{residual}_{k_j t} = \mathbf{activities}_{k_j t}\boldsymbol{\beta}_1 + \mathbf{period}_{k_j t}\boldsymbol{\beta}_2 + \boldsymbol{b}_{k_j} + \varepsilon_{k_j t}$$

with $b_{k_j} \sim \text{N}(0, \sigma^2_{b_k})$ and b_k independent for all different activities; further, $\varepsilon_{k_j} \sim \text{N}(0, \boldsymbol{\Sigma}_\varepsilon)$, i.e. $\boldsymbol{\Sigma}_\varepsilon$ is block–diagonal with the blocks defined by the realizations of the activities. The detailed model equations can be found in Appendix A.2.

As the mean level of each realization of a certain activity varies around the corresponding population effect, a random slope is considered for each realization of an activity. Autocorrelation is accounted for with a rational quadratic correlation structure for the error term (Pinheiro and Bates, 2000), because the usual AR(1) process lack the inclusion of the complete information of a highly correlated time series for the estimation of categorical effects (Subsections 2.4.5 and 2.4.6). Activity–specific residual variances are assessed to cope with heteroscedasticity between observations of different activity realizations. The model formulation is based on the assumption that activities yield an additive change to the background exposure and that several simultaneous activities additively change the PNC levels.

Data with fixed–site PNC levels above the respective 95 % quantile are also omitted in the second–stage models. Furthermore, observations of activities, which were conducted less or equal than 3 times, are excluded; since these activities describe specific and uncommon microenvironments, their effect cannot be generalized as it is also indicated in Gerharz et al. (2013).

2.4.8. Results

Descriptive results

On average, the 1–min, resolved PNC records covered 358 min. (SD: 21) per individual during one measurement series. During one measurement series in spring, the device recorded only the concentration of the first 45 min and in another case, also in spring, no values at all.

Mean personal temperature was on average higher, when the individuals stayed indoors (winter: 17.8 °C, spring: 21.0 °C, summer: 22.1 °C), than outdoors (winter: 11.6 °C, spring: 19.5 °C, summer: 21.4 °C). Relative humidity was similar in winter and spring indoors as well as outdoors between 41 and 45 %; in summer, relative humidity was 57 % indoors and 56% outdoors. Dew point temperature increased from winter (indoors: 4.7 °C, outdoors: −1.1°C) to summer (indoors: 13.2 °C, outdoors: 12.0 °C). Personal temperature and humidity data loggers showed a delayed response to large variation in temperature and humidity levels. This mostly affected the short outdoor periods in the Scenarios "car" and "public transport" and yielded higher mean personal outdoor temperature levels and lower mean personal outdoor humidity levels in comparison to the measurements at the fixed–site monitor.

Table 2.2 lists descriptive statistics of the PNC levels measured by the individuals and at the fixed monitoring site. Median personal PNC levels were highest during all seasons while conducting the Scenario "city center". Extreme personal measurements

	Personal PNC levels				Fixed–site PNC levels				Ratio	
	Median	Mean	Q95	SD	Median	Mean	Q95	SD	Median	Mean
Scenario 1: Car										
Winter	8.6	21.7	112.7	38.1	10.4	11.6	20.8	6.9	0.7	2.2
Spring	8.5	20.4	89.0	34.3	6.6	8.2	19.1	5.2	1.2	3.6
Summer	8.9	18.7	77.0	29.5	4.6	7.0	20.4	5.6	1.3	3.4
Scenario 2: Public transport										
Winter	11.0	38.8	226.6	66.8	9.9	13.0	34.8	10.3	1.0	4.1
Spring	9.8	23.5	101.2	35.4	6.5	9.1	20.8	7.7	1.3	3.9
Summer	7.8	11.2	35.2	11.8	6.2	8.6	21.9	7.0	1.1	1.7
Scenario 3: City center										
Winter	14.4	22.9	64.9	27.9	10.0	11.9	27.1	8.0	1.4	2.1
Spring	12.0	14.5	33.5	10.9	7.1	9.7	22.0	7.8	1.4	1.9
Summer	10.4	12.8	30.6	10.0	5.5	7.5	21.2	6.7	1.7	2.3
Exposure to PNC while indoors										
Winter	7.7	30.1	170.8	55.9	10.3	12.5	30.9	8.6	0.7	3.1
Spring	8.5	21.7	98.5	35.7	6.3	7.9	17.2	5.6	1.2	3.8
Summer	7.6	14.4	48.6	22.9	5.2	7.5	21.2	5.9	1.1	2.5
Exposure to PNC while outdoors										
Winter	15.3	20.5	51.8	18.9	10.0	12.6	34.8	9.2	1.5	2.0
Spring	12.9	15.8	36.5	11.8	7.4	10.2	26.8	8.6	1.5	2.0
Summer	11.6	14.2	32.9	11.3	6.1	8.2	22.3	7.7	1.8	2.3

TABLE 2.2.: *Descriptive statistics of personal and fixed–site 1–min. PNC (in 1000/cm^3) and of the ratio between personal and fixed–site measurements (personal measurement/fixed–site measurement). Q95: 95 % quantile; SD: standard deviation.*

during the Scenarios "car" and "public transport" induced high mean personal PNC levels. Median personal indoor PNC did not any show strong seasonal variation. Mean personal indoor exposure to PNC was high in winter in comparison to spring and summer. Outdoors, median personal exposure to PNC was higher than indoors and higher in winter than in spring and summer. The mean personal indoor PNC level exceeded the mean personal outdoor level especially in winter, because more extreme personal PNC levels were recorded indoors. The fixed–site measurements showed indoors as well as outdoors lower mean and median PNC levels than the personal levels with the only exception that the fixed–site median concentration in winter was higher than the respective personal indoor concentration. The personally measured values showed larger variability than the fixed–site measurements; this was especially apparent while staying indoors. The discrepancy between the median and the mean fixed–site PNC levels was weaker than between the median and the mean personal PNC levels indicated by a positive skewness of the corresponding ratio. Personal exposure levels exceeded the values at the fixed–site monitor by a mean factor of 2.0 – 2.3 outdoors and of 2.5 – 3.8 indoors. The high outdoor personal PNC levels occurred because the volunteers mainly stayed in traffic environments while outdoors. Those environments have higher PNC levels than measured at the fixed–site monitor.

In Figure 2.9, the ratio between the median exposure to PNC of each volunteer during a particular activity and the median exposure to PNC of all individuals, stratified on the season, is shown separately for exposure to PNC occurring indoors and outdoors. This ratio quantifies the person–specific deviance from the average exposure during the respective activity. Increased exposure to indoor PNC was observed during cooking, which

was particularly pronounced in winter and spring; also, concentrations were increased during the smell of food. Most of the individuals either cooked or smelled food at about noon. Window opening may act as a source of outdoor particles as well as a sink for indoor particles. In the winter season, higher–than–average PNC levels were observed when the windows were tilted. As one can suppose, windows were tilted or opened during particle–producing activities. For example, during cooking, above one–half of the observation time, the windows were tilted or opened. Since in winter the windows were tilted in 25.4 % during cooking, the association between exposure and the opening of the windows was heavily confounded by cooking. Smoke originating from wood combustion and especially from candles resulted in higher exposure to PNC. Slightly higher PNC levels were observed in winter season during domestic activities including cleaning, ironing or other activities related to the production of "dust". The volunteers were mainly exposed to passive smoking while staying outdoors, but no clear effect was visible. Regarding outdoor exposure, PNC was higher in traffic than in situations when the individuals did not stay in traffic, particularly in summer. Median outdoor exposure mainly represented the exposure when participating as a pedestrian in traffic, because this was the most frequent outdoor activity. While using other means of transportation, the PNC levels changed only slightly. Atypical outdoor activities including staying in vehicles with opened windows or staying near an outdoor cash machine are not depicted in Figure 2.9. The association between a certain activity and personal exposure was affected by other co–occurring activities, background concentrations and meteorological conditions. Therefore, models comprising all potential influencing factors simultaneously yielded further insights.

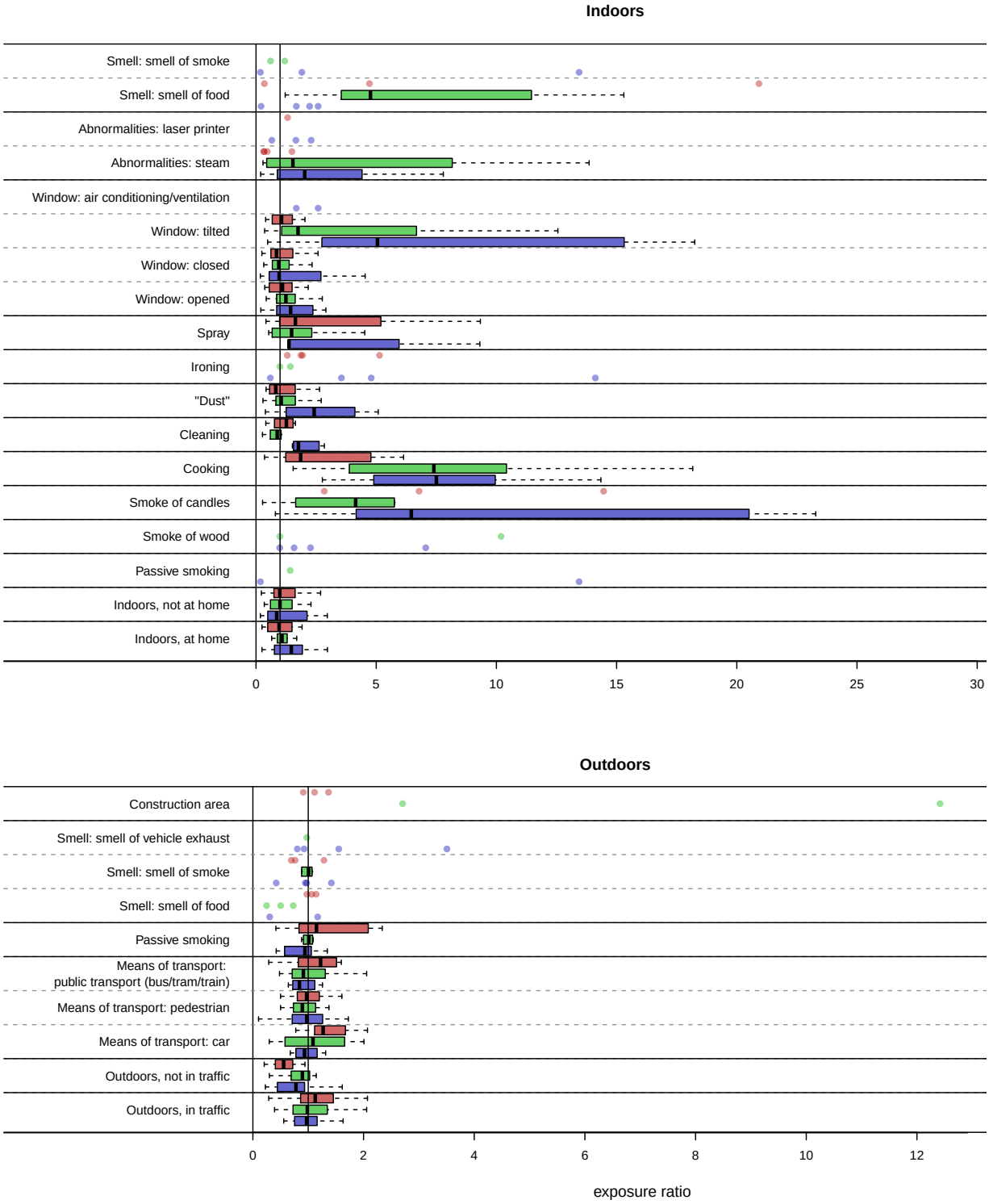


FIGURE 2.9.: *Ratio between the median 1–min. exposure to PNC of each volunteer during a particular activity and the median 1–min. exposure of PNC of all individuals, stratified on the season, separately for indoor and outdoor activities (dots; blue: winter, green: spring, red: summer). Exposure ratios of more than four persons are summarized by a boxplot.*

Two-Stage Exposure Models

Model fit and relevance of different variable groups in the background exposure models are summarized in Table 2.3.

	N	R^2	Individual	Time	Meteo-rology	Cat. Sur-rounding	PM/NC/BC (UAS)
Indoors							
Winter	3433	0.055	0.342	0.197	[a]	0.020	0.106
Spring	2682	0.135	0.004	[a]	0.117	[a]	0.717
Summer	3342	0.111	[a]	0.522	[a]	0.174	0.082
Outdoors							
Winter	1257	0.135	[a]	[a]	0.075	[a]	0.737
Spring	1472	0.137	0.006	0.076	0.022	[a]	0.686
Summer	1344	0.191	0.045	0.068	0.000	[a]	0.572

[a] None of the variables in the respective variable group was selected.

TABLE 2.3.: *First two columns: number of observations N and R^2 of the corresponding model; subsequent columns: contribution of specific variable groups (see Subsection 2.4.4 for the definitions of the groups) to the explained variance of 1–min. personal exposure in the background models.*

The smooth covariate effects of the indoor background models for the three different seasons are shown in Figure 2.10.

For the winter model, most of the explained variance is due to the variance between the volunteers. Thus, individual behavior and housing conditions other than the collected covariates provide the majority of variance explanation. PNC measured at the fixed monitoring site of particles between an aerodynamic diameter of 2.5 μm and 10 μm show a positive, linear association with personal PNC. Fixed–site PNC levels are the most important predictors for the indoor background model for spring. Fixed–site PNC levels of particles with size fractions between 50 nm and 100 nm and between 100 nm and 500 nm have a nearly linear positive effect on personally measured PNC and elevated temperatures above 20 °C are negatively associated with personal PNC. For the summer model, daily variation explains a major part of the variance in the background PNC levels. The number concentration of particles between an aerodynamic diameter of 1 μm and 2.5 μm at the fixed–site monitor is inversely associated with personal PNC levels. The time of day is chosen by the variable selection procedure, but shows only a marginal effect.

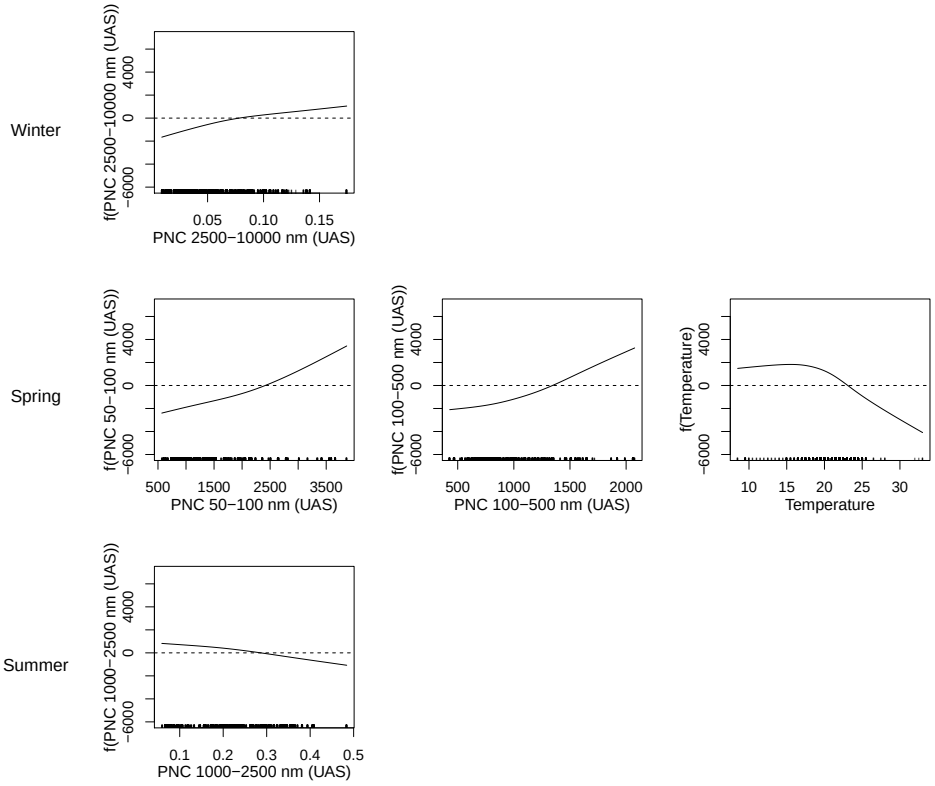


FIGURE 2.10.: *Smooth covariate effects (in $1/cm^3$) in the background model on personal indoor exposure to 1–min. PNC for each season.*

Estimated regression coefficients in the second–stage indoor model can be found in Table 2.4.

			CI		
	Effect	SE	Lower	Upper	P-value
Intercept	12.9	2.0	9.0	16.9	0.000
Smoke of candles	52.5	7.0	38.8	66.2	0.000
Smoke of wood	3.0	11.8	-20.1	26.2	0.796
Cooking	37.6	6.7	24.5	50.8	0.000
Smell of food	50.4	21.1	8.9	91.9	0.017
Cleaning	-0.9	3.8	-8.3	6.5	0.806
Steam	10.3	11.9	-13.1	33.7	0.387
"Dust"	-2.5	2.0	-6.4	1.4	0.210
Ironing	16.1	11.7	-7.0	39.1	0.172
Spray	4.8	4.7	-4.4	14.0	0.306
Period: spring	-6.9	2.4	-11.6	-2.2	0.004
Period: summer	-8.6	2.3	-13.1	-4.0	0.000

TABLE 2.4.: *Effects (in 1000/cm^3) of activities on personal indoor exposure to 1–min. PNC adjusted for the measurements period and background exposure.*

The intercept in the activity models quantifies the average exposure to PNC, when none of the activities were conducted in winter. The estimation for the intercept and the period effects differ from zero, although they are already included in the background model. The reason is that extreme personal measurements, which are excluded for the background model, are included in the activity model. The personal PNC levels are significantly increased during exposure to smoke of candles. Furthermore, cooking and smell of food are significantly associated with higher PNC levels by about 38,000 particles/cm^3 and 50,000 particles/cm^3, respectively. The occurrence of steam originating from, for example, boiling water or showering non–significantly increases the PNC levels by about 10,000 particles/ cm^3 on average. Non–significantly increased concentrations are also observed during ironing.

The smooth covariate effects of the outdoor background models for the three different seasons are depicted in Figure 2.11.

Outdoor exposure to PNC as a pedestrian is mainly driven by ambient PNC levels measured with the fixed–site monitor. In winter, PNC levels of particles with aerodynamic diameters between 3 nm and 10 nm, 50 nm and 100 nm, and 2.5 μm and 10 μm at the fixed–site monitor show a linear, positive effect. The personally measured dew point temperature have a strong negative impact on PNC. The number concentrations of particles with an aerodynamic diameter smaller than 1 μm and of particles in the size range of the CPC device at the fixed–site monitor are positively associated with personal outdoor exposure to PNC in spring. Elevated PNC levels are found in the morning hours. Dew point temperatures above 7 °C result in slightly decreasing background exposure to PNC. The summer model also showed a morning increase in personal PNC. The number concentration of PM exhibit a positive, nearly linear effect. The following variables are chosen by the variable selection procedure, but show only a slight effect: relative humidity in the winter model; PNC 50–100 nm and PNC 100–500 nm in the spring model; PNC CPC, PNC 500–1000 nm, PM and relative humidity in the summer model.

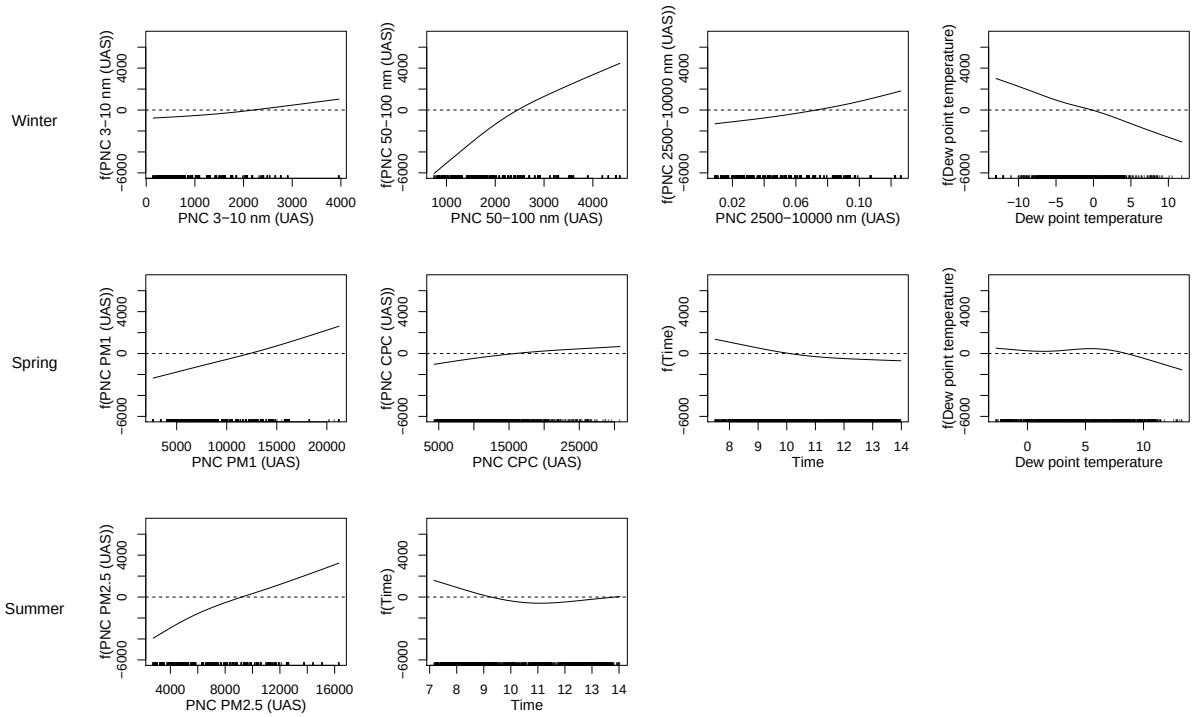


FIGURE 2.11.: *Smooth covariate effects (in 1/cm^3) in the background model on personal outdoor exposure to 1–min. PNC for each season.*

Table 2.5 illustrates the estimated regression coefficients in the second–stage outdoor model.

			CI		
	Effect	SE	Lower	Upper	P-value
Intercept	2.6	0.7	1.3	3.9	0.000
Car driving	2.0	2.1	-2.2	6.1	0.352
Using public transport (bus/tram/train)	-2.5	0.8	-4.0	-0.9	0.002
Not in traffic	-2.3	1.9	-5.9	1.4	0.224
Passive smoking	1.6	2.1	-2.4	5.6	0.434
Smell of food	-1.9	2.0	-5.8	2.0	0.340
Smell of smoke	0.5	2.9	-5.2	6.1	0.876
Smell of vehicle exhaust	4.8	4.2	-3.4	12.9	0.253
Construction area	31.7	28.6	-24.5	87.9	0.268
Period: spring	-0.2	0.8	-1.8	1.5	0.833
Period: summer	0.1	0.8	-1.5	1.7	0.914

TABLE 2.5.: *Effects (in 1,000/cm^3 of activities on personal outdoor exposure to 1–min. PNC adjusted for the measurements period and background exposure.*

The means of transport significantly (P=0.005) affected the exposure to PNC; lowest exposure was observed during staying in public transport or not in traffic, followed by commuting as a pedestrian, the reference category; car drivers were most severely exposed to PNC. PNC levels were (not significantly) increased while staying near a construction area.

2.4.9. Discussion and conclusion: exposure model for everyday situations

Whereas our study examines a wide range of everyday situations, recent studies on personal exposure to PNC considered specific aspects, activities or environments. Our findings are consistent with those of others. Mean fixed–site PNC is lower than in the outskirts of Leipzig (Germany) between February 1997 and February 2001 (Wehner and Wiedensohler, 2003) and in Erfurt between January 1999 and November 2000 (Pitz et al., 2003). The PNC concentrations in schools were found to be lower than or similar to that found in our study (Fromme et al., 2007) because extreme particle–generating activities as they occur at home are missing. (Morawska et al., 2003) observed similar PNC levels in residential houses in Australia during daytime activities in comparison with the mean indoor PNC levels deriving from our analyses in spring. The authors detected a high variability between the houses resulting presumably from differing activities. This suggestion is affirmed through our analyses, which show, that the random person effect adjusted for the activities is low. Temperature is also found in previous works to be negatively associated with outdoor PNC (Pekkanen et al., 1997; Penttinen et al., 2001; Pitz et al., 2003). (Wehner and Wiedensohler, 2003) suggested that lower temperatures reduce the vertical mixing in the lower atmosphere resulting in higher PNC. Our results back up these findings and further indicate that ambient humidity is impacting PNC concentrations as well. In the literature, ambiguous results are reported regarding the effect of relative humidity (Penttinen et al., 2001; Pitz et al., 2003; Wehner and Wiedensohler, 2003). In contrast to other findings (Morawska et al., 2003), particle concentration is not clearly increased, when the volunteers are exposed to passive smoking. A strong influence of the presence of lighted candles on the number concentration of PM is also reported by Bekö et al. (2013) and Wallace and Ott (2011). Cooking or food preparation in general is consistent with the results of other authors, detected as the main particle-generating activity (Bekö et al., 2013; Buonanno et al., 2012; Kearney et al., 2011; Wheeler et al., 2011). He et al. (2004) found varying effects depending on the type of food preparation, for example, PNC exposure was higher during cooking or frying than during using microwave or oven. We observe substantial variation in PNC levels during cooking, but data enabling a further differentiation of this activity regarding duration or the mode of cooking were not collected. This study also shows that outdoor exposure to PNC is higher in traffic, especially while driving with a car. Brauer et al. (1999) as well found a higher geometric mean of particle concentrations for being in a car than for walking, going by bus or bicycle (for particles in the range between 0.3 μm and 5.0 μm) whereas Kaur and Nieuwenhuijsen (2009) observed higher PNC in buses and cars than during cycling or walking in London.

The results confirm the findings of Gerharz et al. (2013) regarding the reduced performance of a microenvironment model for the indoor microenvironment due to its strong heterogeneity and regarding the need of data on more specific microenvironments for exposure models.

2.4.10. Internal characterization of Berkson error

Personal exposure to particulate matter depends on many parameters, which include ambient and non–ambient conditions and which vary between indoor and outdoor locations (Subsection 2.4.8; Gu et al. (2015)): Meteorological conditions and ambient PM concentrations have an impact on the background PM level of a person, whereas certain activities like cooking, lighting candles and car driving affect the personal exposure in a very indi-

vidual and situation–specific way. The transferability of the results from the validation study, particularly the transferability of properties of the residuals as properties of the Berkson error, is questionable. As mentioned in eq. 2.9, the model definition implicitly involves the assumption that

$$\boldsymbol{X}_i^{*\mathrm{C}} \sim \mathrm{N}\left(\boldsymbol{X}_i^{*\mathrm{C}} + (\nu_i^{\mathrm{B}} + \nu_i^{\mathrm{C}})\mathbf{1}_T, \boldsymbol{\Sigma}_{U^{\mathrm{B}}} + \boldsymbol{\Sigma}_{U^{\mathrm{C}}}\right).$$

For frequentist approaches for measurement error correction, this assumption is restrictive, because the uncertainty about the knowledge from external studies is not directly included in the estimation procedure of the regression analysis and knowledge from external validation studies is often not adequate, because this assumption is not met. Therefore, we consider additionally an internal characterization of the Berkson error. A regression model with the differences between (centered) mobile and (centered) fixed–site PNC measurement of the Augsburger Umweltstudie, while the individual was staying outdoors as response and the individual as covariate:

$\mathbf{BMM_{int}}$: Model for the internal estimation of the Berkson error

$$\begin{aligned}
(X^{*\mathrm{C}}_{i_M t_M} - X^{*\mathrm{B}}_{i_M t_M}) =& \alpha_0^M + \nu^{\mathrm{BC}}_{i_M} + U^{\mathrm{BC}}_{i_M t_M} \\
U^{\mathrm{BC}}_{i_M t_M} \sim& \mathrm{N}(0, \sigma^2_{U^{\mathrm{BC}}}) \\
\sigma^2_{U^{\mathrm{B}}} =& \sigma^2_{U^{\mathrm{BC}}} - \sigma^2_{U^{\mathrm{C}}} \\
\widehat{U}^{\mathrm{BC}}_{i_M t_M} =& \alpha_1^M + V^{\mathrm{BC}}_{i_M t_M} \\
\boldsymbol{V}^{\mathrm{BC}}_{i_M} \sim& \mathrm{N}(0, \boldsymbol{\Sigma}^{\mathrm{BC}}_{i_M}), \boldsymbol{\Sigma}^{\mathrm{BC}}_{i_M} = \sigma^2_{V^{\mathrm{BC}}} \boldsymbol{W}_{\rho^{\mathrm{B}}, i_M} \\
\sigma^2_{\nu^{\mathrm{B}}} =& \mathrm{Var}(X^{*\mathrm{C}}_{i_M t_M} - X^{*\mathrm{B}}_{i_M t_M}) - \sigma^2_{U^{\mathrm{BC}}} - \sigma^2_{\nu^{\mathrm{C}}}
\end{aligned} \tag{2.13}$$

The Berkson error is characterized using the relationship between mobile and fixed–site measurements of the Augsburger Umweltstudie and the information regarding the classical measurement error, derived in Section 2.4, in three steps: First, the residual variance of a regression model with the differences between mobile and fixed–site PNC measurements of the Augsburger Umweltstudie as response and the individual as covariate provided information about $\sigma^2_{U^{\mathrm{BC}}} = \sigma^2_{U^{\mathrm{C}}} + \sigma^2_{U^{\mathrm{B}}}$. Second, ρ^{B} is estimated by the empirical autocorrelation of the residuals $\widehat{\boldsymbol{U}}^{\mathrm{BC}}_{i_M}$ and is only an approximation, because actually, the sum of classical and Berkson error is considered. Third, $\sigma^2_{U^{\mathrm{B}}}$ and $\sigma^2_{\nu^{\mathrm{B}}}$ are determined through the differences between the total measurement error variance and the size of the classical measurement error.

Estimations for parameters describing the Berkson error according to eq. 2.13 are: $\widehat{\sigma}^2_{U^{\mathrm{B}}} = 0.3, \widehat{\sigma}^2_{\nu^{\mathrm{B}}} = 0.21$ and $\widehat{\rho}^{\mathrm{B}} = 0.582$. In comparison to the classical measurement error of the devices for measuring PNC, the size of Berkson error is considerably higher.

Indeed, the normality assumption for the measurement errors is violated regarding the kurtosis of the empirical distribution of $\widehat{U}^{\mathrm{BC}}$ (see Figure 2.12), but the assumption is sustained to facilitate the examination of the complex error structure in the following chapters. The Bayesian approach presented in Chapter 5 and Section 6.2 enables the easy consideration of non–normal distributional assumptions.

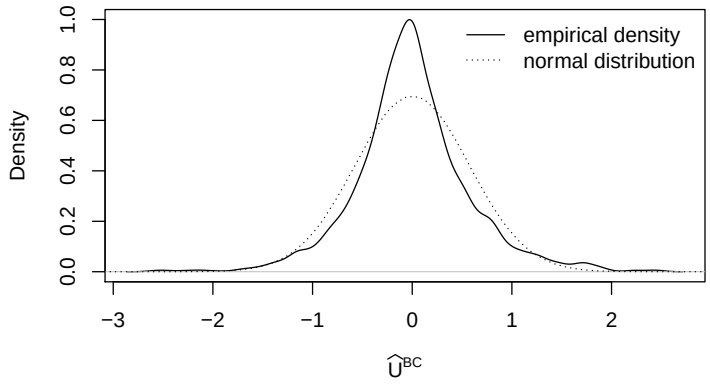


FIGURE 2.12.: *Density of $\widehat{U}^{BC}$ in comparison to the corresponding normal distribution.*

3. Bias through measurement errors in linear mixed models

The calculation of the effect of measurement error enables the easy evaluation of its relevance, transparently elucidates its impact and involves the calculation of correction formulas. The role of covariate measurement error in linear mixed models is theoretically examined in this chapter. After introducing three considered types of measurement error (classical, Berkson and mixture error) in Section 2.1, a general introduction to the naive estimation with error–prone measurements is given in Section 3.1. The resulting biases on the effect estimates and the error variances are derived in the Sections 3.2–3.4. Extensions of the considered models regarding alternative calculations of the attenuation factor, the utilization of the information of the breakdown times of the devices, the inclusion of further covariates, the consideration of classical measurement error of the fixed–site measurements and the handling of unbalanced observations are discussed in Section 3.5. A simulation study in the style of the Augsburger Umweltstudie and the attendant validation studies illustrates the theoretical findings (Section 3.6).

3.1. Naive estimation

The naive models are defined as:

HMM$^{*\mathrm{B}}$: HM for fixed–site measurements of main data

$$\begin{aligned} Y_{i_M t_M} =& \beta_0 + \beta_X^{*\mathrm{B}} X_{i_M t_M}^{*\mathrm{B}} + \boldsymbol{Z}_{i_M t_M} \boldsymbol{\beta}_Z + \tau_{i_M}^{*\mathrm{B}} + \varepsilon_{i_M t_M}^{*\mathrm{B}} \\ \boldsymbol{\varepsilon}_{i_M}^{*\mathrm{B}} \sim& N(\mathbf{0}, \boldsymbol{\Sigma}_{\varepsilon^{*\mathrm{B}}, i_M}), \boldsymbol{\Sigma}_{\varepsilon^{*\mathrm{B}}, i_M} = \sigma_{\varepsilon^{*\mathrm{B}}}^2 \boldsymbol{W}_{\rho^{*\mathrm{B}}, i_M} \\ \tau_{i_M}^{*\mathrm{B}} \sim& N(0, \sigma_{\tau^{*\mathrm{B}}}^2) \end{aligned} \tag{3.1}$$

HMM$^{*\mathrm{C}}$: HM for personal measurements of main data

$$\begin{aligned} Y_{i_M t_M} =& \beta_0 + \beta_X^{*\mathrm{C}} X_{i_M t_M}^{*\mathrm{C}} + \boldsymbol{Z}_{i_M t_M} \boldsymbol{\beta}_Z + \tau_{i_M}^{*\mathrm{C}} + \varepsilon_{i_M t_M}^{*\mathrm{C}} \\ \boldsymbol{\varepsilon}_{i_M}^{*\mathrm{C}} \sim& N(\mathbf{0}, \boldsymbol{\Sigma}_{\varepsilon^{*\mathrm{C}}, i_M}), \boldsymbol{\Sigma}_{\varepsilon^{*\mathrm{C}}, i_M} = \sigma_{\varepsilon^{*\mathrm{C}}}^2 \boldsymbol{W}_{\rho^{*\mathrm{C}}, i_M} \\ \tau_{i_M}^{*\mathrm{C}} \sim& N(0, \sigma_{\tau^{*\mathrm{C}}}^2) \end{aligned} \tag{3.2}$$

HMM$^{*\mathrm{M}}$: HM for mixed fixed and personal measurements of main data

$$\begin{aligned} Y_{i_M t_M} =& \beta_0 + \beta_X^{*\mathrm{M}} X_{i_M t_M}^{*\mathrm{M}} + \boldsymbol{Z}_{i_M t_M}\boldsymbol{\beta}_Z + \tau_{i_M}^{*\mathrm{M}} + \varepsilon_{i_M t_M}^{*\mathrm{M}} \\ \boldsymbol{\varepsilon}_{i_M}^{*\mathrm{M}} \sim& N(\mathbf{0}, \boldsymbol{\Sigma}_{\varepsilon^{*\mathrm{M}}, i_M}), \boldsymbol{\Sigma}_{\varepsilon^{*\mathrm{M}}, i_M} = \sigma^2_{\varepsilon^{*\mathrm{M}}} \boldsymbol{W}_{\rho^{*\mathrm{M}}, i_M} \\ \tau_{i_M}^{*\mathrm{M}} \sim& N(0, \sigma^2_{\tau^{*\mathrm{M}}}) \end{aligned} \tag{3.3}$$

The impact of an error–prone covariate $\boldsymbol{X}^* \in \{\boldsymbol{X}^{*\mathrm{B}}, \boldsymbol{X}^{*\mathrm{C}}, \boldsymbol{X}^{*\mathrm{M}}\}$ on a simple linear mixed model without the covariates $\boldsymbol{Z}$ is examined in the current and in the following three subsections.

3.1.1. Score equations

$\boldsymbol{\beta}^* = (\beta_0^*, \beta_X^*)^\top$ and $\boldsymbol{\phi}^* = (\sigma^2_{\varepsilon^*}, \rho^*, \sigma^2_{\tau^*})^\top = (\phi_1^*, \phi_2^*, \phi_3^*)^\top$ denoting the probability limits of the naive estimators for $n \to \infty$ are obtained as solutions from the linear mixed model score equations, if $n \to \infty$:

$$\mathbb{E}\left[\boldsymbol{\chi}^{*\top} \boldsymbol{V}^{*-1}(\boldsymbol{Y} - \boldsymbol{\chi}^*\boldsymbol{\beta}^*)\right] = \mathbf{0} \tag{3.4}$$

and

$$\frac{1}{2}\left\{\mathbb{E}\left[(\boldsymbol{Y} - \boldsymbol{\chi}^*\boldsymbol{\beta}^*)^\top \boldsymbol{V}^{*-1}\frac{\partial \boldsymbol{V}^*}{\partial \phi_j^*}\boldsymbol{V}^{*-1}(\boldsymbol{Y} - \boldsymbol{\chi}^*\boldsymbol{\beta}^*)\right] - \mathrm{tr}\left(\boldsymbol{V}^{*-1}\frac{\partial \boldsymbol{V}^*}{\partial \phi_j^*}\right)\right\} = 0 \tag{3.5}$$

with $\boldsymbol{\chi}^* = (\mathbf{1}_T, \boldsymbol{X}^*)$ and $j = 1, 2, 3$ (see Wang et al., 1998)). $\boldsymbol{V}^*$, $\boldsymbol{V}^* = \sigma^2_{\tau^*}\mathbf{J}_T + \boldsymbol{\Sigma}_{\varepsilon^*}$, denotes the probability limit for $n \to \infty$ of the variance–covariance matrix of the error term in the naive model and $\mathbf{J}_T$ denotes the $T \times T$ matrix of ones. $\tau_i^* \sim \mathrm{N}(0, \sigma^2_{\tau^*})$ is assumed and ε_{it}^* is assumed to follow an autoregressive process of order 1 with autocorrelation parameter ρ^* and variance matrix $\boldsymbol{\Sigma}_{\varepsilon^*} = \sigma^2_{\varepsilon^*}\boldsymbol{W}_{\rho^*}$. Due to clarity, the index i for the i.i.d. samples is neglected in eq. 3.4 and eq. 3.5 and in the following considerations.

3.1.2. Naive regression coefficients

The first score equation (eq. 3.4) for the effect coefficients can be transformed equivalently to Wang et al. (1998) using calculation rules for the expected value:

$$\begin{aligned} & \mathbb{E}\left[\boldsymbol{\chi}^{*\top} \boldsymbol{V}^{*-1}(\boldsymbol{Y} - \boldsymbol{\chi}^*\boldsymbol{\beta}^*)\right] = \mathbf{0} \\ \Leftrightarrow\ & \mathbb{E}\left[\boldsymbol{\chi}^{*\top} \boldsymbol{V}^{*-1}\boldsymbol{Y}\right] = \mathbb{E}\left[\boldsymbol{\chi}^{*\top} \boldsymbol{V}^{*-1}\boldsymbol{\chi}^*\right]\boldsymbol{\beta}^* \\ \Leftrightarrow\ & \mathbb{E}\left[\boldsymbol{\chi}^{*\top} \boldsymbol{V}^{*-1}\boldsymbol{\chi}\right]\boldsymbol{\beta} = \mathbb{E}\left[\boldsymbol{\chi}^{*\top} \boldsymbol{V}^{*-1}\boldsymbol{\chi}^*\right]\boldsymbol{\beta}^*, \end{aligned} \tag{3.6}$$

with $\boldsymbol{\beta} = (\beta_0, \beta_X)^\top$ and $\boldsymbol{\chi} = (\mathbf{1}_T, \boldsymbol{X})$. Similarly to the calculation of the expected value of a quadratic form with an arbitrary vector $\boldsymbol{X}, \boldsymbol{X} = (X_1, \ldots, X_T)^\top$, of random variables and a positive definite $T \times T$ matrix $\boldsymbol{A}$ through

$$\mathbb{E}(\boldsymbol{X}^\top \boldsymbol{A}\boldsymbol{X}) = \mathrm{tr}(\boldsymbol{A}\mathrm{Var}(\boldsymbol{X})) + \mathbb{E}(\boldsymbol{X})^\top \boldsymbol{A}\mathbb{E}(\boldsymbol{X}), \tag{3.7}$$

$\mathbb{E}(\boldsymbol{X_1}^\top \boldsymbol{A}\boldsymbol{X_2})$ with the arbitrary vectors $\boldsymbol{X_1}, \boldsymbol{X_1} = (X_{11}, \ldots, X_{1T})^\top$, and $\boldsymbol{X_2}$, $\boldsymbol{X_2} = (X_{21}, \ldots, X_{2T})^\top$, of random variables is given by

$$\begin{aligned}
\mathbb{E}(\boldsymbol{X}_1^\top \boldsymbol{A}\boldsymbol{X}_2) &= \mathbb{E}[\mathrm{tr}(\boldsymbol{X}_1^\top \boldsymbol{A}\boldsymbol{X}_2])\\
&= \mathbb{E}[\mathrm{tr}(\boldsymbol{A}\boldsymbol{X}_2\boldsymbol{X}_1^\top)]\\
&= \mathrm{tr}[\mathbb{E}(\boldsymbol{A}\boldsymbol{X}_2\boldsymbol{X}_1^\top)]\\
&= \mathrm{tr}[\boldsymbol{A}\mathrm{Cov}(\boldsymbol{X}_1, \boldsymbol{X}_2) + \boldsymbol{A}\mathbb{E}(\boldsymbol{X}_2)\mathbb{E}(\boldsymbol{X}_1)^\top]\\
&= \mathrm{tr}[\boldsymbol{A}\mathrm{Cov}(\boldsymbol{X}_1, \boldsymbol{X}_2)] + \mathbb{E}(\boldsymbol{X}_1)^\top \boldsymbol{A}\mathbb{E}(\boldsymbol{X}_2) \qquad (3.8)
\end{aligned}$$

Thus, eq. 3.6 can be written as

$$\begin{aligned}
&\left[\begin{pmatrix} 0 & 0 \\ 0 & \mathrm{tr}(\boldsymbol{V}^{*-1}\mathrm{Cov}(\boldsymbol{X}, \boldsymbol{X}^*)) \end{pmatrix} + \mathbb{E}(\boldsymbol{\chi}^*)^\top \boldsymbol{V}^{*-1}\mathbb{E}(\boldsymbol{\chi})\right]\boldsymbol{\beta}\\
&= \left[\begin{pmatrix} 0 & 0 \\ 0 & \mathrm{tr}(\boldsymbol{V}^{*-1}\mathrm{Var}(\boldsymbol{X}^*)) \end{pmatrix} + \mathbb{E}(\boldsymbol{\chi}^*)^\top \boldsymbol{V}^{*-1}\mathbb{E}(\boldsymbol{\chi}^*)\right]\boldsymbol{\beta}^* .
\end{aligned}$$

The probability limits for the naive coefficients result as

$$\begin{aligned}
\beta_0^* &= \beta_0 + \mathbb{E}(X)(\beta_X - \beta_X^*)\\
\beta_X^* &= \beta_X \frac{\mathrm{tr}(\boldsymbol{V}^{*-1}\mathrm{Cov}(\boldsymbol{X}, \boldsymbol{X}^*))}{\mathrm{tr}(\boldsymbol{V}^{*-1}\mathrm{Var}(\boldsymbol{X}^*))} =: \beta_X \lambda. \qquad (3.9)
\end{aligned}$$

Detailed expressions of λ depend on the type of measurement error and are given in the sections 3.2–3.4.

3.1.3. Naive covariance structure

Using a general linear mixed model with an AR(1) process for the error term as a naive model with the error–prone covariate $\boldsymbol{X}^*$, $\boldsymbol{X}^* \in \{\boldsymbol{X}^{*\mathrm{B}}, \boldsymbol{X}^{*\mathrm{C}}, \boldsymbol{X}^{*\mathrm{M}}\}$ results in a misspecified model because actually, the model error, as a sum of AR(1) processes, follows an ARMA(p, q) process with $q > 0$ (Lütkepohl (1984); see also Appendix B.2). Thus, the actual data generating process does not belong to the model class of the naive model.

Maximum Likelihood estimations of the coefficients in a linear regression model under misspecification of the covariance structure of the error term are unbiased. In contrast, the estimation of the covariance matrix of the error term itself is biased.

Since the aim is the quantification of the bias due to measurement error and not due to misspecification, the AR(1) error structure is further investigated in the naive model. Using eq. 3.8, the three score equations for $\boldsymbol{\phi}^*$ (eq. 3.5) can be written as

$$\begin{aligned}
&\mathrm{tr}\left(\boldsymbol{V}^{*-1}\boldsymbol{W}_{\rho^*}\boldsymbol{V}^{*-1}\boldsymbol{V}_T\right) = \mathrm{tr}\left(\boldsymbol{V}^{*-1}\boldsymbol{W}_{\rho^*}\right)\\
&\mathrm{tr}\left(\boldsymbol{V}^{*-1}\frac{\partial \boldsymbol{W}_{\rho^*}}{\partial \rho^*}\boldsymbol{V}^{*-1}\boldsymbol{V}_T\right) = \mathrm{tr}\left(\boldsymbol{V}^{*-1}\frac{\partial \boldsymbol{W}_{\rho^*}}{\partial \rho^*}\right)\\
&\mathrm{tr}\left(\boldsymbol{V}^{*-1}\mathbf{J}_T\boldsymbol{V}^{*-1}\boldsymbol{V}_T\right) = \mathrm{tr}\left(\boldsymbol{V}^{*-1}\mathbf{J}_T\right)
\end{aligned}$$

with $\boldsymbol{V}_T = \mathrm{Var}(\boldsymbol{Y} - \boldsymbol{\chi}^*\boldsymbol{\beta}^*)$.

The score equations (eq. 3.4 and 3.5) do not allow to write $\phi_j^*, j = 1, 2, 3$, in terms of other parameters. Two special cases where solutions are available are presented in the next paragraphs.

If $\boldsymbol{V}_T$ exhibits the same matrix structure as $\boldsymbol{V}^*$, i.e. $\boldsymbol{V}_T = \sigma_a^2 \boldsymbol{W}_{\rho_a} + \sigma_c^2 \mathbf{J}_T$ with arbitrary variances σ_a^2 and σ_c^2 and arbitrary autocorrelation parameter ρ_a, $\boldsymbol{V}_T = \boldsymbol{V}^*$. Thus, the probability limits of the variance parameters of the error term in the naive model are:

$$\sigma_{\varepsilon^*}^2 \boldsymbol{W}_{\rho^*} = \sigma_a^2 \boldsymbol{W}_{\rho_a} \tag{3.10}$$

$$\sigma_{\tau^*}^2 = \sigma_c^2 \tag{3.11}$$

For the second special case no random effects, neither in the main model nor in the measurement errors, are considered ($\sigma_\tau^2 = \sigma_{\nu^{\mathrm{B}}}^2 = \sigma_{\nu^{\mathrm{C}}}^2 = 0$), but $\boldsymbol{V}_T$ consists of a sum of, w.l.o.g two, AR(1) variance–covariance matrices: $\boldsymbol{V}_T = \boldsymbol{\Sigma}_a + \boldsymbol{\Sigma}_b = \sigma_a^2 \boldsymbol{W}_{\rho_a} + \sigma_b^2 \boldsymbol{W}_{\rho_b}$ with arbitrary variances σ_a^2 and σ_b^2 and arbitrary autocorrelation parameters ρ_a and ρ_b, and $\boldsymbol{V}^* = \boldsymbol{\Sigma}_{\varepsilon^*} = \sigma_{\varepsilon^*}^2 \boldsymbol{W}_{\rho^*}$.

The resulting score equations for $\sigma_{\varepsilon^*}^2$ and ρ^* are

$$\operatorname{tr}(\boldsymbol{W}_{\rho^*}^{-1} \boldsymbol{V}_T) = T \sigma_{\varepsilon^*}^2 \tag{3.12}$$

$$\operatorname{tr}\left(\boldsymbol{W}_{\rho^*}^{-1} \frac{\partial \boldsymbol{W}_{\rho^*}}{\partial \rho^*} \boldsymbol{W}_{\rho^*}^{-1} \boldsymbol{V}_T \right) = \sigma_{\varepsilon^*}^2 \operatorname{tr}\left(\boldsymbol{W}_{\rho^*}^{-1} \frac{\partial \boldsymbol{W}_{\rho^*}}{\partial \rho^*} \right) \tag{3.13}$$

$\sigma_{\varepsilon^*}^2$ is directly derived from eq. 3.12:

$$\sigma_{\varepsilon^*}^2 = \frac{1}{T} \operatorname{tr}(\boldsymbol{W}_{\rho^*}^{-1} \boldsymbol{V}_T) \tag{3.14}$$

This relation is introduced in eq. 3.13:

$$\operatorname{tr}\left(\boldsymbol{W}_{\rho^*}^{-1} \frac{\partial \boldsymbol{W}_{\rho^*}}{\partial \rho^*} \boldsymbol{W}_{\rho^*}^{-1} \boldsymbol{V}_T \right) = \frac{1}{T} \operatorname{tr}\left(\boldsymbol{W}_{\rho^*}^{-1} \frac{\partial \boldsymbol{W}_{\rho^*}}{\partial \rho^*} \right) \operatorname{tr}\left(\boldsymbol{W}_{\rho^*}^{-1} \boldsymbol{V}_T \right) \tag{3.15}$$

Dissolving eq. 3.14 and eq. 3.15 is described in Appendix B.3 and B.4 and results in

$$\rho^* = \frac{\rho_a \sigma_a^2 + \rho_b \sigma_b^2}{\sigma_a^2 + \sigma_b^2},$$
$$\sigma_{\varepsilon^*}^2 = \sigma_a^2 + \sigma_b^2.$$

3.2. Berkson error

The independence between model error and measurement error yields under the simple random intercept regression model

$$\mathbb{E}(\boldsymbol{Y}|\boldsymbol{X}^*) = \mathbb{E}(\mathbb{E}(\boldsymbol{Y}|\boldsymbol{X})|\boldsymbol{X}^*) = \beta_0 + \beta_X \mathbb{E}(\boldsymbol{X}|\boldsymbol{X}^*). \tag{3.16}$$

Based on the main outcome model eq. 3.1 (without confounder variables $\boldsymbol{Z}$) and the Berkson error model eq. 2.4, the effect estimate for a covariate with Berkson error is just as unbiased as with a conventional Berkson error. The reason therefore is, that modeling the conditional expectation of the outcome $\boldsymbol{Y}$ given the error–free predictor $\boldsymbol{X}$ results in the same slope β_X as with the Berkson error–prone covariate $\boldsymbol{X}^{*\mathrm{B}}$, since the theorem of the iterated expectation yields:

$$\begin{aligned} \mathbb{E}(\boldsymbol{Y}|\boldsymbol{X}^{*\mathrm{B}}) &\overset{\text{eq. 3.16}}{=} \beta_0 + \beta_X \mathbb{E}(\boldsymbol{X}|\boldsymbol{X}^{*\mathrm{B}}) \\ &= \beta_0 + \beta_X \boldsymbol{X}^{*\mathrm{B}}. \end{aligned}$$

In comparison to the error–free measurements $\boldsymbol{X}$, the observations $\boldsymbol{X}^{*\mathrm{B}}$ lack to explain personal variability described by the random person effect ν_i^{B} and the random error $\boldsymbol{U}_i^{\mathrm{B}}$.

Thus, the fraction of the variability of $\boldsymbol{X}$ exceeding the variability of the Berkson error–prone measurement $\boldsymbol{X}^{*\mathrm{B}}$ this variance is conferred on the variance of the error term in the main outcome model:

$$\begin{aligned}\mathrm{Var}(\boldsymbol{Y} - \boldsymbol{\chi}^{*\mathrm{B}}\boldsymbol{\beta}) &= \boldsymbol{\Sigma}_\varepsilon + \sigma_\tau^2\mathbf{J}_T + \beta_X^2(\boldsymbol{\Sigma}_{U^\mathrm{B}} + \sigma_{\nu^\mathrm{B}}^2\mathbf{J}_T) \\ &= (\sigma_\tau^2 + \beta_X^2\sigma_{\nu^\mathrm{B}}^2)\mathbf{J}_T + \boldsymbol{\Sigma}_\varepsilon + \beta_X^2\boldsymbol{\Sigma}_{U^\mathrm{B}}\end{aligned}$$

Data with Berkson error entail an overestimation of the variability of the error term and consequently also of the effect estimate $\widehat{\beta_1^*}$, since $\mathrm{Var}(X_{it}^{*\mathrm{B}}) \leq \mathrm{Var}(X_{it})$, analogically to a simple linear model as main outcome model. Estimating the residual variance by assuming a random intercept and an AR(1) process for the error term cannot identify the true underlying structure of the error term as it is discussed in Subsection 3.1.2.

3.3. Classical measurement error

To understand the complex impact of classical covariate measurement error in mixed models, the relations are stepwisely illuminated.

3.3.1. Homogeneous covariate and error structure

Usual classical measurement error (without random effects and autocorrelation: $\sigma_\tau^2 = \sigma_{\nu^\mathrm{B}}^2 = \sigma_{\nu^\mathrm{C}}^2 = 0, \rho^{X^{*\mathrm{B}}} = \rho = \rho^\mathrm{B} = \rho^\mathrm{C} = 0$) causes an attenuation, i.e. a bias towards 0, of the effect estimate β_X of the corresponding covariate in a simple linear regression model by the so–called attenuation factor or reliability ratio λ^C with $\lambda^\mathrm{C} \in [0, 1]$ (Carroll et al., 2006; Fuller, 1987):

$$\beta_X^{*\mathrm{C}} = \lambda^\mathrm{C}\beta_X.$$

If measurement error and model error are independent, this attenuation factor λ^C can be calculated based on $\mathbb{E}(X|X^{*\mathrm{C}})$ (eq. 3.16) as the ratio between the variance of the actual values X and the deficient values $X^{*\mathrm{C}}$ in linear regression models due to the normality of $\boldsymbol{X}$ and $\boldsymbol{U}^\mathrm{C}$ (Carroll et al., 2006):

$$\lambda^\mathrm{C} = \mathbb{E}(X|X^{*\mathrm{C}}) = \frac{\mathrm{Cov}(X, X^{*\mathrm{C}})}{\mathrm{Var}(X^{*\mathrm{C}})} = \frac{\mathrm{Var}(X)}{\mathrm{Var}(X^{*\mathrm{C}})} = \frac{\sigma_X^2}{\sigma_X^2 + \sigma_{U^\mathrm{C}}^2}. \tag{3.17}$$

This property holds also if a simple linear mixed model with random intercept, as defined in eq. 3.1, is considered. With increasing measurement error variance, the effect coefficient is stronger attenuated as depicted in Fig. 3.1.

The variance of the error term increases with the size of the classical measurement error:

$$\begin{aligned}&\mathrm{Var}(Y - (\beta_0^{*\mathrm{C}} + \beta_X^{*\mathrm{C}}X^{*\mathrm{C}})) \\ &\quad= \mathrm{Var}(Y - \beta_0^{*\mathrm{C}} - \beta_X^{*\mathrm{C}}(X + U^\mathrm{C})) \\ &\quad= \sigma_\varepsilon^2 + \beta_X^2(1 - \lambda^\mathrm{C})^2\sigma_X^2 + \beta_X^2{\lambda^\mathrm{C}}^2\sigma_{U^\mathrm{C}}^2 \\ &\quad\overset{\text{eq. } 3.17}{=} \sigma_\varepsilon^2 + \lambda^\mathrm{C}\beta_X^2\sigma_{U^\mathrm{C}}^2 \\ &\quad= \sigma_{\varepsilon^{*\mathrm{C}}}^2.\end{aligned}$$

Although $\mathrm{Var}(Y|X) \leq \mathrm{Var}(Y|X^{*\mathrm{C}})$, the uncertainty of the naive estimate $\widehat{\beta_X^{*\mathrm{C}}}$ may either increase or decrease, because $\mathrm{Var}(X) \leq \mathrm{Var}(X^{*\mathrm{C}})$. The linearity of $\mathbb{E}(Y_{it}|X^{*\mathrm{C}})$ and the

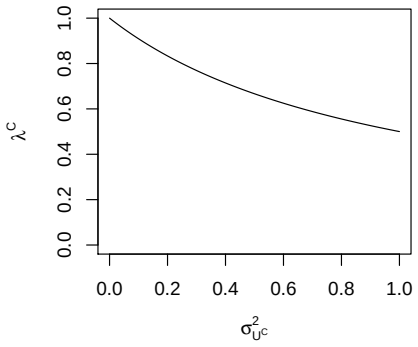


FIGURE 3.1.: *Attenuation of the effect coefficient of a classical error–prone covariate in a simple linear model with homogeneous covariate structure.* $\sigma_X^2 = 1$.

joint normality of the model errors, the true covariate and the classical measurement error implicate that the naive model of the response Y and the covariate $X^{*\mathrm{C}}$ is a linear regression model (Carroll et al., 2006).

3.3.2. Heterogeneous covariate and error structure

Wang et al. (1998) have studied the consequences of a heterogeneous covariate structure in the linear mixed model, i.e. $\boldsymbol{X}$ is not presumed to be i.i.d. itself but to originate from independent clusters with i.i.d. observations. In order to remain in the introduced model and parameter definitions, such a heterogeneous covariate X_{it} can be regarded as the sum of a global, Berkson error–prone mean $X_t^{*\mathrm{B}}$, an individual–specific component of the Berkson error, ν_i^{B}, and a random Berkson error U_{it}^{B}, as defined in Subsection 2.1.2. For this type of model $\sigma_{\nu^{\mathrm{C}}}^2 = 0, \rho^{X^{*\mathrm{B}}} = \rho = \rho^{\mathrm{B}} = \rho^{\mathrm{C}} = 0$ is assumed.

According to Wang et al. (1998), the effects of a heterogeneous covariate with classical measurement error is underestimated by the factor

$$\lambda^{\mathrm{C}} = \frac{\sigma_{\nu^{\mathrm{B}}}^2 + \sigma_X^2 \left[1 + (T-1)\frac{\sigma_{\tau^{*\mathrm{C}}}^2}{\sigma_{\varepsilon^{*\mathrm{C}}}^2}\right]}{\sigma_{\nu^{\mathrm{B}}}^2 + (\sigma_X^2 + \sigma_{U^{\mathrm{C}}}^2)\left[1 + (T-1)\frac{\sigma_{\tau^{*\mathrm{C}}}^2}{\sigma_{\varepsilon^{*\mathrm{C}}}^2}\right]}. \tag{3.18}$$

The variance of the random effect ν_i^{B}, $\sigma_{\nu^{\mathrm{B}}}^2$, contributes to the attenuation factor with additional terms in the numerator and the denominator of the attenuation factor, but in comparison to $\sigma_{U^{\mathrm{B}}}^2$, with an additional weight $\left[1 + (T-1)\frac{\sigma_{\tau^{*\mathrm{C}}}^2}{\sigma_{\varepsilon^{*\mathrm{C}}}^2}\right]^{-1}$ approximating zero for $T \to \infty$ (Figure 3.2). Thus, the attenuation factor is higher in a model with a heterogeneous covariate than in a model with a homogeneous covariate. Further, the weight depends on the ratio between the variance of the random effects and the random model error in the naive model.

The variance–covariance matrix of the error term is block–diagonal and (neglecting index i) each block is given by

$$\begin{aligned}
&\mathrm{Var}(\boldsymbol{Y} - (\beta_0^{*\mathrm{C}} + \beta_X^{*\mathrm{C}} \boldsymbol{X}^{*\mathrm{C}})) \\
&\quad = \mathrm{Var}(\boldsymbol{Y} - \beta_0^{*\mathrm{C}} - \beta_X^{*\mathrm{C}}(\boldsymbol{X}^{*\mathrm{B}} + \nu^{\mathrm{B}}\mathbf{1}_T + \boldsymbol{U}^{\mathrm{B}} + \boldsymbol{U}^{\mathrm{C}})) \\
&\quad = [\sigma_\varepsilon^2 + \beta_X^2(1-\lambda^{\mathrm{C}})^2\left(\sigma_{X^{*\mathrm{B}}}^2 + \sigma_{U^{\mathrm{B}}}^2\right) + \beta_X^2 {\lambda^{\mathrm{C}}}^2 \sigma_{U^{\mathrm{C}}}^2]\mathbf{I}_T + \\
&\qquad [\sigma_\tau^2 + \beta_X^2(1-\lambda^{\mathrm{C}})^2\sigma_{\nu^{\mathrm{B}}}^2]\mathbf{J}_T.
\end{aligned} \tag{3.19}$$

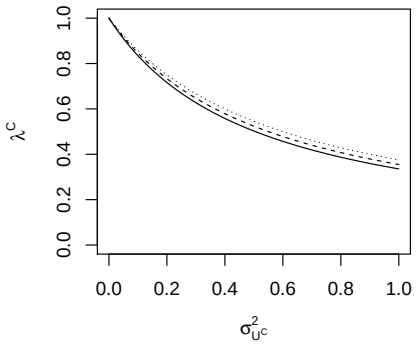


FIGURE 3.2.: *Attenuation of the effect coefficient of a classical error–prone covariate in a simple linear mixed model with heterogeneous covariate structure. T: 5 (dotted), 10 (dashed), 100 (solid);* $\sigma_X^2 = 0.5; \sigma_{\nu^B}^2 = 0.5; \sigma_{\tau^*}^2 = 1; \sigma_\varepsilon^2 = 1$.

The probability limit of the variance of the random intercept in the naive model is according to eq. 3.11 and eq. 3.19

$$\sigma_{\tau^*}^2 = \sigma_\tau^2 + (1 - \lambda^{\mathrm{C}})^2 \sigma_{\nu^{\mathrm{B}}}^2 \beta_X^2$$

and the probability limit of the variance of the error term in the naive model is according to eq. 3.10 and eq. 3.19

$$\sigma_{\varepsilon^*}^2 = \sigma_\varepsilon^2 + \left[(1 - \lambda^{\mathrm{C}})^2 (\sigma_{X^{*\mathrm{B}}}^2 + \sigma_{U^{\mathrm{B}}}^2) + {\lambda^{\mathrm{C}}}^2 \sigma_{U^{\mathrm{C}}}^2\right] \beta_X^2.$$

The heterogeneity of the deficient covariate affects also the variance of the random intercept in the main outcome model, $\sigma_{\tau^*}^2$, in a similar way as the variance of X affects the error variance.

The attenuation factor for a heterogeneous, classical error–prone covariate depends on the main outcome model. Let $\sigma_\tau^2 = \sigma_{\nu^{\mathrm{C}}}^2 = 0, \rho^{X^{*\mathrm{B}}} = \rho = \rho^{\mathrm{B}} = \rho^{\mathrm{C}} = 0$, i.e. the main outcome model is a simple linear model and the covariate is heterogeneous and exhibits classical measurement error. If a simple linear mixed model is used for the analysis, eq. 3.9 results in

$$\begin{aligned}
\lambda^{\mathrm{C}} &\overset{\text{eq. } 3.9}{=} \frac{\left[\mathrm{tr}(\boldsymbol{V}^{*\mathrm{C}^{-1}} \mathrm{Cov}(\boldsymbol{X}, \boldsymbol{X}^{*\mathrm{C}}))\right]}{\left[\mathrm{tr}(\boldsymbol{V}^{*\mathrm{C}^{-1}} \mathrm{Var}(\boldsymbol{X}^{*\mathrm{C}}))\right]} \\
&= \frac{\mathrm{tr}\left[\frac{1}{\sigma_{\varepsilon^{*\mathrm{C}}}^2}(\sigma_{X^{*\mathrm{B}}}^2 \mathbf{I}_T + \sigma_{U^{\mathrm{B}}}^2 \mathbf{I}_T + \sigma_{\nu^{\mathrm{B}}}^2 \mathbf{J}_T)\right]}{\mathrm{tr}\left[\frac{1}{\sigma_{\varepsilon^{*\mathrm{C}}}^2}(\sigma_{X^{*\mathrm{B}}}^2 \mathbf{I}_T + \sigma_{U^{\mathrm{B}}}^2 \mathbf{I}_T + \sigma_{\nu^{\mathrm{B}}}^2 \mathbf{J}_T + \sigma_{U^{\mathrm{C}}}^2 \mathbf{I}_T)\right]} \\
&= \frac{\sigma_{X^{*\mathrm{B}}}^2 + \sigma_{U^{\mathrm{B}}}^2 + \sigma_{\nu^{\mathrm{B}}}^2}{\sigma_{X^{*\mathrm{B}}}^2 + \sigma_{U^{\mathrm{B}}}^2 + \sigma_{\nu^{\mathrm{B}}}^2 + \sigma_{U^{\mathrm{C}}}^2},
\end{aligned} \tag{3.20}$$

with $\boldsymbol{V}^{*\mathrm{C}} = \sigma_{\varepsilon^{*\mathrm{C}}}^2 \mathbf{I}_T$; a similar result is found by Rosner et al. (1989).

If $\left[1 + (T-1)\frac{\sigma_{\tau^*\mathrm{C}}^2}{\sigma_{\varepsilon^{*\mathrm{C}}}^2}\right] > 1$ as in the most cases, the attenuation factor in eq. 3.20 resulting from a simple linear model as main outcome model, is higher than in eq. 3.18 resulting from a simple linear mixed model as main outcome model. Thus, a heterogeneous, error–prone covariate, attenuates the effect coefficients less if a simple linear mixed main outcome model is used in comparison to a simple linear model.

Let additionally admit subgroup–specific classical measurement error (without autocorrelation in the errors of the main model and the measurement error, i.e. $\rho^{X^{*\mathrm{B}}} = \rho = \rho^{\mathrm{B}} = \rho^{\mathrm{C}} = 0$). The variance–covariance matrix of the heterogeneous covariate $\boldsymbol{X}^{*\mathrm{C}}$ with heterogeneous error structure, i.e. $X_{it}^{*\mathrm{C}} = X_t^{*\mathrm{B}} + \nu_i^{\mathrm{B}} + U_{it}^{\mathrm{B}} + \nu_i^{\mathrm{C}} + U_{it}^{\mathrm{C}}$,

$$\mathrm{Var}(\boldsymbol{X}^{*\mathrm{C}}) = (\sigma^2_{X^{*\mathrm{B}}} + \sigma^2_{U^{\mathrm{B}}} + \sigma^2_{U^{\mathrm{C}}})\mathbf{I}_T + (\sigma^2_{\nu^{\mathrm{B}}} + \sigma^2_{\nu^{\mathrm{C}}})\mathbf{J}_T$$

has the same structure as the variance–covariance matrix of the heterogeneous covariate, i.e. $X_{it}^{*\mathrm{C}} = X_t^{*\mathrm{B}} + \nu_i^{\mathrm{B}} + U_{it}^{\mathrm{B}} + U_{it}^{\mathrm{C}}$, considered in the previous paragraphs:

$$\mathrm{Var}(\boldsymbol{X}^{*\mathrm{C}}) = (\sigma^2_{X^{*\mathrm{B}}} + \sigma^2_{U^{\mathrm{B}}} + \sigma^2_{U^{\mathrm{C}}})\mathbf{I}_T + \sigma^2_{\nu^{\mathrm{B}}}\mathbf{J}_T.$$

Therefore, the derivation of the attenuation factor in the case of a heterogeneous, error–prone covariate can easily be adapted to a heterogeneous form of the classical measurement error. The attenuation factor λ^{C} is calculated with

$$\lambda^{\mathrm{C}} = \frac{\sigma^2_{\nu^{\mathrm{B}}} + (\sigma^2_{X^{*\mathrm{B}}} + \sigma^2_{U^{\mathrm{B}}})\left[1 + (T-1)\frac{\sigma^2_{\tau^{*\mathrm{C}}}}{\sigma^2_{\varepsilon^{*\mathrm{C}}}}\right]}{\sigma^2_{\nu^{\mathrm{B}}} + \sigma^2_{\nu^{\mathrm{C}}} + (\sigma^2_{X^{*\mathrm{B}}} + \sigma^2_{U^{\mathrm{B}}} + \sigma^2_{U^{\mathrm{C}}})\left[1 + (T-1)\frac{\sigma^2_{\tau^{*\mathrm{C}}}}{\sigma^2_{\varepsilon^{*\mathrm{C}}}}\right]}. \tag{3.21}$$

Increasing $\sigma^2_{\nu^{\mathrm{C}}}$ while fixing the other parameters enlarges the attenuation of the effect. However, this impact approximates zero if $T \to \infty$. The effect of individual–specific classical measurement error on the regression coefficient in a simple linear regression model with random intercept is in most cases (if $1 + (T-1)\frac{\sigma^2_{\tau^*}}{\sigma^2_{\varepsilon^*}} > 1$) less severe than the usual classical measurement error. If the main model does not contain a random intercept, additional individual–specific classical measurement error will strengthen the attenuation of the effect.

The variance of the individual–specific classical measurement error is transferred to the variance of the error term:

$$\begin{aligned}
&\mathrm{Var}(\boldsymbol{Y} - (\beta_0^{*\mathrm{C}} + \beta_X^{*\mathrm{C}}\boldsymbol{X}^{*\mathrm{C}})) \\
&\quad = \mathrm{Var}(\boldsymbol{Y} - \beta_0^{*\mathrm{C}} - \beta_X^{*\mathrm{C}}(\boldsymbol{X}^{*\mathrm{B}} + \nu^{\mathrm{B}}\mathbf{1}_T + \boldsymbol{U}^{\mathrm{B}} + \nu^{\mathrm{C}}\mathbf{1}_T + \boldsymbol{U}^{\mathrm{C}})) \\
&\quad = [\sigma^2_\varepsilon + \beta^2_X(1-\lambda^{\mathrm{C}})^2\sigma^2_X + \beta^2_X\lambda^{\mathrm{C}^2}\sigma^2_{U^{\mathrm{C}}}]\mathbf{I}_T + \\
&\qquad [\sigma^2_\tau + \beta^2_X(1-\lambda^{\mathrm{C}})^2\sigma^2_{\nu^{\mathrm{B}}} + \beta^2_X\lambda^{\mathrm{C}^2}\sigma^2_{\nu^{\mathrm{C}}}]\mathbf{J}_T.
\end{aligned} \tag{3.22}$$

The probability limits for the variance parameters in the naive model are, according to eq. 3.10, eq. 3.11 and eq. 3.22,

$$\begin{aligned}
\sigma^2_{\tau^*} &= \sigma^2_\tau + \lambda^{\mathrm{C}^2}\beta^2_X\sigma^2_{\nu^{\mathrm{C}}} + (1-\lambda^{\mathrm{C}})^2\beta^2_X\sigma^2_{\nu^{\mathrm{B}}}, \\
\sigma^2_{\varepsilon^*} &= \sigma^2_\varepsilon + \left[(1-\lambda^{\mathrm{C}})^2\left(\sigma^2_{X^{*\mathrm{B}}} + \sigma^2_{U^{\mathrm{B}}}\right) + \lambda^{\mathrm{C}^2}\sigma^2_{U^{\mathrm{C}}}\right]\beta^2_X.
\end{aligned}$$

3.3.3. Heterogeneous and autocorrelated error structure

In this subsection, the models of the previous sections are generalized to the framework described in the Section 2.1: The covariate as well as the measurement error structure are considered heterogeneous. In addition, the error terms of the main model, of the measurement error model and of the covariate itself are assumed to follow an autoregressive process of order one.

The attenuation factor λ^{C} can be written as

$$\begin{aligned}\lambda^{\mathrm{C}} &\overset{\text{eq. 3.9}}{=} \frac{\operatorname{tr}\left[\boldsymbol{V}^{*\mathrm{C}\,-1}\operatorname{Cov}(\boldsymbol{X},\boldsymbol{X}^{*\mathrm{C}})\right]}{\operatorname{tr}\left[\boldsymbol{V}^{*\mathrm{C}\,-1}\operatorname{Var}(\boldsymbol{X}^{*\mathrm{C}})\right]} \\ &= \frac{\operatorname{tr}\left[(\sigma^2_{\tau^{*\mathrm{C}}}\mathbf{J}_T + \boldsymbol{\Sigma}_{\varepsilon^{*\mathrm{C}}})^{-1}(\boldsymbol{\Sigma}_{X^{*\mathrm{B}}} + \boldsymbol{\Sigma}_{U^{\mathrm{B}}} + \sigma^2_{\nu^{\mathrm{B}}}\mathbf{J}_T)\right]}{\operatorname{tr}\left[(\sigma^2_{\tau^{*\mathrm{C}}}\mathbf{J}_T + \boldsymbol{\Sigma}_{\varepsilon^{*\mathrm{C}}})^{-1}(\boldsymbol{\Sigma}_{X^{*\mathrm{B}}} + \boldsymbol{\Sigma}_{U^{\mathrm{B}}} + \boldsymbol{\Sigma}_{U^{\mathrm{C}}} + (\sigma^2_{\nu^{\mathrm{B}}} + \sigma^2_{\nu^{\mathrm{C}}})\mathbf{J}_T)\right]}.\end{aligned} \tag{3.23}$$

The Sherman–Morrison formula (Sherman and Morrison, 1950) is used to invert $\sigma^2_{\tau^{*\mathrm{C}}}\mathbf{J} + \boldsymbol{\Sigma}_{\varepsilon^{*\mathrm{C}}}$. In the notation of Bartlett (1951), the Sherman–Morrison formula is given as

$$(\boldsymbol{A} + \boldsymbol{u}\boldsymbol{v}^\top)^{-1} = \boldsymbol{A}^{-1} - \frac{\boldsymbol{A}^{-1}\boldsymbol{u}\boldsymbol{v}^\top\boldsymbol{A}^{-1}}{1 + \boldsymbol{v}^\top\boldsymbol{A}^{-1}\boldsymbol{u}} \tag{3.24}$$

for an invertible square matrix $\boldsymbol{A}$ of dimension $n \times n$, vectors $\boldsymbol{u}$ and $\boldsymbol{v}$, both of length n and $1 + \boldsymbol{v}^\top\boldsymbol{A}^{-1}\boldsymbol{u} \neq 0$. Because

$$d := 1 + \sigma^2_{\tau^{*\mathrm{C}}}\mathbf{1}_T^\top\boldsymbol{\Sigma}^{-1}_{\varepsilon^{*\mathrm{C}}}\mathbf{1}_T = 1 + \frac{\sigma^2_{\tau^{*\mathrm{C}}}}{\sigma^2_{\varepsilon^{*\mathrm{C}}}(1-\rho^{*\mathrm{C}2})}\left[(T-2)(1-\rho^{*\mathrm{C}})^2 + 2(1-\rho^{*\mathrm{C}})\right] \neq 0, \tag{3.25}$$

the Sherman–Morrison formula yields:

$$\begin{aligned}(\sigma^2_{\tau^{*\mathrm{C}}}\mathbf{J}_T + \boldsymbol{\Sigma}_{\varepsilon^{*\mathrm{C}}})^{-1} &= (\boldsymbol{\Sigma}_{\varepsilon^{*\mathrm{C}}} + \sigma^2_{\tau^{*\mathrm{C}}}\mathbf{1}_T\mathbf{1}_T^\top)^{-1} \\ &\overset{\text{eq. 3.24}}{=} \boldsymbol{\Sigma}^{-1}_{\varepsilon^{*\mathrm{C}}} - \sigma^2_{\tau^{*\mathrm{C}}}(1 + \sigma^2_{\tau^{*\mathrm{C}}}\mathbf{1}_T^\top\boldsymbol{\Sigma}^{-1}_{\varepsilon^{*\mathrm{C}}}\mathbf{1}_T)^{-1}\boldsymbol{\Sigma}^{-1}_{\varepsilon^{*\mathrm{C}}}\mathbf{J}_T\boldsymbol{\Sigma}^{-1}_{\varepsilon^{*\mathrm{C}}} \\ &\overset{\text{eq. 3.25}}{=} \boldsymbol{\Sigma}^{-1}_{\varepsilon^{*\mathrm{C}}} - \sigma^2_{\tau^{*\mathrm{C}}}d^{-1}\boldsymbol{\Sigma}^{-1}_{\varepsilon^{*\mathrm{C}}}\mathbf{J}_T\boldsymbol{\Sigma}^{-1}_{\varepsilon^{*\mathrm{C}}}.\end{aligned}$$

The numerator of the attenuation factor eq. 3.23 is therefore given by

$$\begin{aligned}&\operatorname{tr}[\ (\sigma^2_{\tau^{*\mathrm{C}}}\mathbf{J}_T + \boldsymbol{\Sigma}_{\varepsilon^{*\mathrm{C}}})^{-1}(\boldsymbol{\Sigma}_{X^{*\mathrm{B}}} + \boldsymbol{\Sigma}_{U^{\mathrm{B}}} + \sigma^2_{\nu^{\mathrm{B}}}\mathbf{J}_T)] \\ =\ &\operatorname{tr}[\ \underbrace{\boldsymbol{\Sigma}^{-1}_{\varepsilon^{*\mathrm{C}}}\boldsymbol{\Sigma}_{X^{*\mathrm{B}}}}_{\text{(A)}} - \sigma^2_{\tau^{*\mathrm{C}}}d^{-1}\underbrace{\boldsymbol{\Sigma}^{-1}_{\varepsilon^{*\mathrm{C}}}\mathbf{J}_T\boldsymbol{\Sigma}^{-1}_{\varepsilon^{*\mathrm{C}}}\boldsymbol{\Sigma}_{X^{*\mathrm{B}}}}_{\text{(B)}} + \\ &\underbrace{\boldsymbol{\Sigma}^{-1}_{\varepsilon^{*\mathrm{C}}}\boldsymbol{\Sigma}_{U^{\mathrm{B}}}}_{\text{(C)}} - \sigma^2_{\tau^{*\mathrm{C}}}d^{-1}\underbrace{\boldsymbol{\Sigma}^{-1}_{\varepsilon^{*\mathrm{C}}}\mathbf{J}_T\boldsymbol{\Sigma}^{-1}_{\varepsilon^{*\mathrm{C}}}\boldsymbol{\Sigma}_{U^{\mathrm{B}}}}_{\text{(D)}} + \\ &\sigma^2_{\nu^{\mathrm{B}}}\underbrace{\boldsymbol{\Sigma}^{-1}_{\varepsilon^{*\mathrm{C}}}\mathbf{J}_T}_{\text{(E)}} - \sigma^2_{\tau^{*\mathrm{C}}}\sigma^2_{\nu^{\mathrm{B}}}d^{-1}\underbrace{\boldsymbol{\Sigma}^{-1}_{\varepsilon^{*\mathrm{C}}}\mathbf{J}_T\boldsymbol{\Sigma}^{-1}_{\varepsilon^{*\mathrm{C}}}\mathbf{J}_T}_{\text{(F)}}\ ]\end{aligned} \tag{3.26}$$

The single components of eq. 3.26 are calculated in the following:

(A)

$$\begin{aligned}&\operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}_{\varepsilon^{*\mathrm{C}}}\boldsymbol{\Sigma}_{X^{*\mathrm{B}}}\right) \\ =\ &\operatorname{tr}\left(\frac{\sigma^2_{X^{*\mathrm{B}}}}{\sigma^2_{\varepsilon^{*\mathrm{C}}}}\boldsymbol{W}^{-1}_{\rho^{*\mathrm{C}}}\boldsymbol{W}_{\rho^{X^{*\mathrm{B}}}}\right) \\ \overset{\text{Sec. B.3}}{=}\ &\frac{\sigma^2_{X^{*\mathrm{B}}}}{\sigma^2_{\varepsilon^{*\mathrm{C}}}\left(1-\rho^{*\mathrm{C}2}\right)}\left(T - 2(T-1)\rho^{*\mathrm{C}}\rho^{X^{*\mathrm{B}}} + (T-2)\rho^{*\mathrm{C}2}\right) \\ =\ &\frac{\sigma^2_{X^{*\mathrm{B}}}}{\sigma^2_{\varepsilon^{*\mathrm{C}}}\left(1-\rho^{*\mathrm{C}2}\right)}g^{*\mathrm{C}}_{1,T}(\rho^{X^{*\mathrm{B}}})\end{aligned}$$

with $g^{*\mathrm{C}}_{1,T}(\rho) := T - 2(T-1)\rho\rho^{*\mathrm{C}} + (T-2)\rho^{*\mathrm{C}2}$.

Ⓑ

$$
\begin{aligned}
&\boldsymbol{\Sigma}_{\varepsilon^{*\mathrm{C}}}^{-1}\mathbf{J}_T\boldsymbol{\Sigma}_{\varepsilon^{*\mathrm{C}}}^{-1}\boldsymbol{\Sigma}_{X^{*\mathrm{B}}} \\
&= \frac{\sigma^2_{X^{*\mathrm{B}}}}{(\sigma^2_{\varepsilon^{*\mathrm{C}}})^2}\boldsymbol{W}_{\rho^{*\mathrm{C}}}^{-1}\mathbf{J}_T\boldsymbol{W}_{\rho^{*\mathrm{C}}}^{-1}\boldsymbol{W}_{\rho^{X^{*\mathrm{B}}}} \\
&\overset{\text{eq. 3.3}}{=} \frac{\sigma^2_{X^{*\mathrm{B}}}}{\left(\sigma^2_{\varepsilon^{*\mathrm{C}}}\left(1-\rho^{*\mathrm{C}^2}\right)\right)^2}\cdot \\
&\underbrace{\begin{pmatrix}
(1-\rho^{*\mathrm{C}})^2 & (1-\rho^{*\mathrm{C}})^3 & (1-\rho^{*\mathrm{C}})^3 & \cdots & (1-\rho^{*\mathrm{C}})^2 \\
(1-\rho^{*\mathrm{C}})^3 & (1-\rho^{*\mathrm{C}})^4 & (1-\rho^{*\mathrm{C}})^4 & \cdots & (1-\rho^{*\mathrm{C}})^3 \\
(1-\rho^{*\mathrm{C}})^3 & (1-\rho^{*\mathrm{C}})^4 & (1-\rho^{*\mathrm{C}})^4 & \cdots & (1-\rho^{*\mathrm{C}})^3 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
(1-\rho^{*\mathrm{C}})^3 & (1-\rho^{*\mathrm{C}})^4 & (1-\rho^{*\mathrm{C}})^4 & \cdots & (1-\rho^{*\mathrm{C}})^3 \\
(1-\rho^{*\mathrm{C}})^2 & (1-\rho^{*\mathrm{C}})^3 & (1-\rho^{*\mathrm{C}})^3 & \cdots & (1-\rho^{*\mathrm{C}})^2
\end{pmatrix}}_{\boldsymbol{D}}\boldsymbol{W}_{\rho^{X^{*\mathrm{B}}}}
\end{aligned}
$$

$$
\begin{aligned}
\Rightarrow \quad & \operatorname{tr}(\boldsymbol{\Sigma}_{\varepsilon^{*\mathrm{C}}}^{-1}\mathbf{J}_T\boldsymbol{\Sigma}_{\varepsilon^{*\mathrm{C}}}^{-1}\boldsymbol{\Sigma}_{X^{*\mathrm{B}}}) \\
&= \frac{\sigma^2_{X^{*\mathrm{B}}}}{\left(\sigma^2_{\varepsilon^{*\mathrm{C}}}\left(1-\rho^{*\mathrm{C}^2}\right)\right)^2}\left(\sum_{i=1}^{T}\sum_{j=1}^{T}[\boldsymbol{D}]_{ij}[\boldsymbol{W}_{\rho^{X^{*\mathrm{B}}}}]_{ji}\right) \\
&= \frac{\sigma^2_{X^{*\mathrm{B}}}}{\left(\sigma^2_{\varepsilon^{*\mathrm{C}}}\left(1-\rho^{*\mathrm{C}^2}\right)\right)^2} \\
&\left[2\left(1-\rho^{*\mathrm{C}}\right)^2\left(1+\left(\rho^{X^{*\mathrm{B}}}\right)^{T-1}\right)+4\left(1-\rho^{*\mathrm{C}}\right)^3\sum_{t=1}^{T-2}\left(\rho^{X^{*\mathrm{B}}}\right)^t+\right. \\
&\left.\left(1-\rho^{*\mathrm{C}}\right)^4\left(T-2+2\sum_{k=1}^{T-3}\sum_{t=1}^{k}\left(\rho^{X^{*\mathrm{B}}}\right)^t\right)\right] \\
&= \frac{\sigma^2_{X^{*\mathrm{B}}}}{\left(\sigma^2_{\varepsilon^{*\mathrm{C}}}\left(1-\rho^{*\mathrm{C}^2}\right)\right)^2}g^{*\mathrm{C}}_{2,T}(\rho^{X^{*\mathrm{B}}})
\end{aligned}
$$

with

$$
\begin{aligned}
g^{*\mathrm{C}}_{2,T}(\rho) :=& 2(1-\rho^{*\mathrm{C}})^2(1+\rho^{T-1})+4\left(1-\rho^{*\mathrm{C}}\right)^3\sum_{t=1}^{T-2}\rho^t+ \\
&\left(1-\rho^{*\mathrm{C}}\right)^4\left(T-2+2\sum_{k=1}^{T-3}\sum_{t=1}^{k}\rho^t\right).
\end{aligned}
$$

Ⓒ results from Ⓐ :

$$
\operatorname{tr}(\boldsymbol{\Sigma}_{\varepsilon^{*\mathrm{C}}}^{-1}\boldsymbol{\Sigma}_{U^{\mathrm{B}}}) = \frac{\sigma^2_{U^{\mathrm{B}}}}{\sigma^2_{\varepsilon^{*\mathrm{C}}}\left(1-\rho^{*\mathrm{C}^2}\right)}g^{*\mathrm{C}}_{1,T}\left(\rho^{\mathrm{B}}\right)
$$

Ⓓ results from Ⓑ :

$$
\operatorname{tr}(\boldsymbol{\Sigma}_{\varepsilon^{*\mathrm{C}}}^{-1}\mathbf{J}_T\boldsymbol{\Sigma}_{\varepsilon^{*\mathrm{C}}}^{-1}\boldsymbol{\Sigma}_{U^{\mathrm{B}}}) = \frac{\sigma^2_{U^{\mathrm{B}}}}{\left(\sigma^2_{\varepsilon^{*\mathrm{C}}}\left(1-\rho^{*\mathrm{C}^2}\right)\right)^2}g^{*\mathrm{C}}_{2,T}\left(\rho^{\mathrm{B}}\right)
$$

Ⓔ results from Ⓐ with 1 instead of $\rho^{X^{*\mathrm{B}}}$ and $\sigma^2_{X^{*\mathrm{B}}}$:

$$\mathrm{tr}(\mathbf{\Sigma}^{-1}_{\varepsilon^{*\mathrm{C}}}\mathbf{J}_T) = \frac{1}{\sigma^2_{\varepsilon^{*\mathrm{C}}}\left(1-\rho^{*\mathrm{C}2}\right)} g^{*\mathrm{C}}_{1,T}(1)$$

Ⓕ results from Ⓑ with 1 instead of $\rho^{X^{*\mathrm{B}}}$ and $\sigma^2_{X^{*\mathrm{B}}}$:

$$\mathrm{tr}(\mathbf{\Sigma}^{-1}_{\varepsilon^{*\mathrm{C}}}\mathbf{J}_T\mathbf{\Sigma}^{-1}_{\varepsilon^{*\mathrm{C}}}\mathbf{J}_T) = \frac{1}{\left(\sigma^2_{\varepsilon^{*\mathrm{C}}}\left(1-\rho^{*\mathrm{C}2}\right)\right)^2} g^{*\mathrm{C}}_{2,T}(1)$$

Using the auxiliary calculations Ⓐ – Ⓕ, eq. 3.26 can be transformed to

$$\begin{aligned}
&\Bigg[\sigma^2_{X^{*\mathrm{B}}} g^{*\mathrm{C}}_{1,T}\left(\rho^{X^{*\mathrm{B}}}\right) - \frac{\sigma^2_{\tau^{*\mathrm{C}}}\sigma^2_{X^{*\mathrm{B}}}}{\sigma^2_{\varepsilon^{*\mathrm{C}}}\left(1-\rho^{*\mathrm{C}2}\right)d} g^{*\mathrm{C}}_{2,T}\left(\rho^{X^{*\mathrm{B}}}\right) + \\
&\quad \sigma^2_{U^{\mathrm{B}}} g^{*\mathrm{C}}_{1,T}\left(\rho^{\mathrm{B}}\right) - \frac{\sigma^2_{\tau^{*\mathrm{C}}}\sigma^2_{U^{\mathrm{B}}}}{\sigma^2_{\varepsilon^{*\mathrm{C}}}\left(1-\rho^{*\mathrm{C}2}\right)d} g^{*\mathrm{C}}_{2,T}\left(\rho^{\mathrm{B}}\right) + \\
&\quad \sigma^2_{\nu^{\mathrm{B}}} g^{*\mathrm{C}}_{1,T}(1) - \frac{\sigma^2_{\tau^{*\mathrm{C}}}\sigma^2_{\nu^{\mathrm{B}}}}{\sigma^2_{\varepsilon^{*\mathrm{C}}}\left(1-\rho^{*\mathrm{C}2}\right)d} g^{*\mathrm{C}}_{2,T}(1) \Bigg] \frac{1}{\sigma^2_{\varepsilon^{*\mathrm{C}}}\left(1-\rho^{*\mathrm{C}2}\right)} \\
= &\Bigg[\sigma^2_{X^{*\mathrm{B}}} g^{*\mathrm{C}}_{T}\left(\rho^{X^{*\mathrm{B}}}\right) + \sigma^2_{U^{\mathrm{B}}} g^{*\mathrm{C}}_{T}\left(\rho^{\mathrm{B}}\right) + \sigma^2_{\nu^{\mathrm{B}}} g^{*\mathrm{C}}_{T}(1) \Bigg] \frac{T}{\sigma^2_{\varepsilon^{*\mathrm{C}}}\left(1-\rho^{*\mathrm{C}2}\right)}
\end{aligned}$$

with

$$\begin{aligned}
g^{*\mathrm{C}}_T(\rho) :=& \left[g^{*\mathrm{C}}_{1,T}(\rho) - \frac{\sigma^2_{\tau^{*\mathrm{C}}}}{\sigma^2_{\varepsilon^{*\mathrm{C}}}\left(1-\rho^{*\mathrm{C}2}\right)d} g^{*\mathrm{C}}_{2,T}(\rho) \right] /T \\
=& 1 - 2\frac{T-1}{T}\rho\rho^{*\mathrm{C}} + \frac{T-2}{T}\rho^{*\mathrm{C}2} - \\
&\Bigg\{ (T-2)\left(1-\rho^{*\mathrm{C}}\right)^4 + 2\left(1-\rho^{*\mathrm{C}}\right)^2\left(1+\rho^{T-1}\right) + 4\left(1-\rho^{*\mathrm{C}}\right)^3 \sum_{t=1}^{T-2}\rho^t + \\
&2\left(1-\rho^{*\mathrm{C}}\right)^4 \sum_{k=1}^{T-3}\sum_{t=1}^{k}\rho^t \Bigg\} \\
&\left\{ T\left[T - 2(T-1)\rho^{*\mathrm{C}} + (T-2)\rho^{*\mathrm{C}2} + \frac{\sigma^2_{\varepsilon^{*\mathrm{C}}}\left(1-\rho^{*\mathrm{C}2}\right)}{\sigma^2_{\tau^{*\mathrm{C}}}} \right] \right\}^{-1} .
\end{aligned}$$

Since the denominator of eq. 3.23 is calculated analogically to its numerator, the attenuation factor for the heterogeneous and autocorrelated error structure results as

$$\begin{aligned}
\lambda^{\mathrm{C}} =& \frac{\beta^{*\mathrm{C}}_X}{\beta_X} \\
=& \frac{\sigma^2_{X^{*\mathrm{B}}} g^{*\mathrm{C}}_T\left(\rho^{X^{*\mathrm{B}}}\right) + \sigma^2_{\nu^{\mathrm{B}}} g^{*\mathrm{C}}_T(1) + \sigma^2_{U^{\mathrm{B}}} g^{*\mathrm{C}}_T\left(\rho^{\mathrm{B}}\right)}{\sigma^2_{X^{*\mathrm{B}}} g^{*\mathrm{C}}_T\left(\rho^{X^{*\mathrm{B}}}\right) + \sigma^2_{\nu^{\mathrm{B}}} g^{*\mathrm{C}}_T(1) + \sigma^2_{U^{\mathrm{B}}} g^{*\mathrm{C}}_T\left(\rho^{\mathrm{B}}\right) + \sigma^2_{\nu^{\mathrm{C}}} g^{*\mathrm{C}}_T(1) + \sigma^2_{U^{\mathrm{C}}} g^{*\mathrm{C}}_T\left(\rho^{\mathrm{C}}\right)} .
\end{aligned} \tag{3.27}$$

Eq. 3.27 reduces to eq. 3.21 if $\rho^{X^{*\mathrm{B}}} = \rho^{\mathrm{B}} = \rho^{\mathrm{C}} = 0$, as shown in Appendix B.5. The limit of $g_T^{*\mathrm{C}}(\rho)$ for $T \to \infty$ is $g_\infty^{*\mathrm{C}}(\rho) = 1 - 2\rho\rho^{*\mathrm{C}} + \rho^{*\mathrm{C}^2}$ for $|\rho| < 1$ and $\rho^{*\mathrm{C}} > 0$ and $g_\infty^{*\mathrm{C}}(1) = 0$ for $\rho^{*\mathrm{C}} \in (0, 1)$ as shown in Appendix B.6. Thus, the attenuation factor λ^{C} approximates

$$\lambda^{\mathrm{C}} = \frac{\sigma^2_{X^{*\mathrm{B}}}\left(1 - 2\rho^{*\mathrm{C}}\rho^{X^{*\mathrm{B}}} + \rho^{*\mathrm{C}^2}\right) + \sigma^2_{U^{\mathrm{B}}}\left(1 - 2\rho^{*\mathrm{C}}\rho^{\mathrm{B}} + \rho^{*\mathrm{C}^2}\right)}{\sigma^2_{X^{*\mathrm{B}}}\left(1 - 2\rho^{*\mathrm{C}}\rho^{X^{*\mathrm{B}}} + \rho^{*\mathrm{C}^2}\right) + \sigma^2_{U^{\mathrm{B}}}\left(1 - 2\rho^{*\mathrm{C}}\rho^{\mathrm{B}} + \rho^{*\mathrm{C}^2}\right) + \sigma^2_{U^{\mathrm{C}}}\left(1 - 2\rho^{*\mathrm{C}}\rho^{\mathrm{C}} + \rho^{*\mathrm{C}^2}\right)} \tag{3.28}$$

for $T \to \infty$. Especially high values of ρ in comparison to ρ^* reduce the size of the weight $g_T(\rho)$ and vice–versa.

The variance of the error term is:

$$\begin{aligned}
\mathrm{Var}(\boldsymbol{Y} - \boldsymbol{\chi}^{*\mathrm{C}}\boldsymbol{\beta}^{*\mathrm{C}}) &= \mathrm{Var}(\boldsymbol{Y} - \beta_0 - \beta_X\boldsymbol{X}) + \mathrm{Var}(\beta_0 + \beta_X\boldsymbol{X} - \beta_X^{*\mathrm{C}}\boldsymbol{X}^{*\mathrm{C}}) \\
&= \mathrm{Var}(\boldsymbol{Y} - \beta_0 - \beta_X\boldsymbol{X}) + \mathrm{Var}(\beta_X\boldsymbol{X}) + \mathrm{Var}(\beta_X^{*\mathrm{C}}\boldsymbol{X}^{*\mathrm{C}}) - \\
&\quad 2\mathrm{Cov}(\beta_X\boldsymbol{X}, \beta_X^{*\mathrm{C}}\boldsymbol{X}^{*\mathrm{C}}) \\
&= \boldsymbol{\Sigma}_\varepsilon + \sigma^2_\tau\mathbf{J}_T + \\
&\quad \beta_X^2(\boldsymbol{\Sigma}_{X^{*\mathrm{B}}} + \boldsymbol{\Sigma}_{U^{\mathrm{B}}} + \sigma^2_{\nu^{\mathrm{B}}}\mathbf{J}_T) + \\
&\quad \beta_X^{*\mathrm{C}^2}(\boldsymbol{\Sigma}_{X^{*\mathrm{B}}} + \boldsymbol{\Sigma}_{U^{\mathrm{B}}} + \sigma^2_{\nu^{\mathrm{B}}}\mathbf{J}_T + \boldsymbol{\Sigma}_{U^{\mathrm{C}}} + \sigma^2_{\nu^{\mathrm{C}}}\mathbf{J}_T) - \\
&\quad 2\beta_X\beta_X^{*\mathrm{C}}(\boldsymbol{\Sigma}_{X^{*\mathrm{B}}} + \boldsymbol{\Sigma}_{U^{\mathrm{B}}} + \sigma^2_{\nu^{\mathrm{B}}}\mathbf{J}_T) \\
&= \boldsymbol{\Sigma}_\varepsilon + \sigma^2_\tau\mathbf{J}_T + \\
&\quad \beta_X^2\left[(1 - \lambda^{\mathrm{C}})^2\left(\boldsymbol{\Sigma}_{X^{*\mathrm{B}}} + \boldsymbol{\Sigma}_{U^{\mathrm{B}}} + \sigma^2_{\nu^{\mathrm{B}}}\mathbf{J}_T\right) + \lambda^{\mathrm{C}^2}\left(\boldsymbol{\Sigma}_{U^{\mathrm{C}}} + \sigma^2_{\nu^{\mathrm{C}}}\mathbf{J}_T\right)\right]
\end{aligned}$$

As mentioned in Subsection 3.1.3, a simple expression for the variance parameters $\sigma^2_{\varepsilon^{*\mathrm{C}}}, \sigma^2_{\tau^{*\mathrm{C}}}$ and $\rho^{*\mathrm{C}}$ in the naive model is not available, because the correlation structure of the model error is misspecified.

3.4. Mixture measurement error

A random variable with mixture measurement error is composed by a certain percentage p of observations with Berkson error and the percentage $1-p$ of observations with classical measurement error (see eq. 2.6).

3.4.1. Homogeneous covariate and error structure

At first, the simple linear regression case with usual classical and Berkson errors ($\sigma^2_\tau = \sigma^2_{\nu^{\mathrm{B}}} = \sigma^2_{\nu^{\mathrm{C}}} = 0, \rho^{X^{*\mathrm{B}}} = \rho = \rho^{\mathrm{B}} = \rho^{\mathrm{C}} = 0$) is considered. $\mathrm{Var}(X^{*\mathrm{M}})$ and $\mathrm{Cov}(X, X^{*\mathrm{M}})$ are calculated by:

$$\begin{aligned}
\mathrm{Var}(X^{*\mathrm{M}}) =& \mathrm{Var}(X^{*\mathrm{B}} + (1-G)U^{\mathrm{B}} + (1-G)U^{\mathrm{C}}) \\
=& \mathrm{Var}(X^{*\mathrm{B}}) + \mathrm{Var}((1-G)U^{\mathrm{B}}) + \mathrm{Var}((1-G)U^{\mathrm{C}}) \\
=& \sigma^2_{X^{*\mathrm{B}}} + \mathbb{E}(\mathrm{Var}((1-G)U^{\mathrm{B}}|G)) + \mathrm{Var}(\mathbb{E}((1-G)U^{\mathrm{B}}|G)) + \\
& \mathbb{E}(\mathrm{Var}((1-G)U^{\mathrm{C}}|G)) + \mathrm{Var}(\mathbb{E}((1-G)U^{\mathrm{C}}|G)) \\
=& \sigma^2_{X^{*\mathrm{B}}} + (1-p)(\sigma^2_{U^{\mathrm{B}}} + \sigma^2_{U^{\mathrm{C}}}) \\
\mathrm{Cov}(X, X^{*\mathrm{M}}) =& \mathrm{Cov}(X^{*\mathrm{B}} + U^{\mathrm{B}}, X^{*\mathrm{B}} + (1-G)U^{\mathrm{B}} + (1-G)U^{\mathrm{C}}) \\
=& \mathrm{Var}(X^{*\mathrm{B}}) + \mathrm{Var}((1-G)U^{\mathrm{B}}) \\
=& \sigma^2_{X^{*\mathrm{B}}} + (1-p)\sigma^2_{U^{\mathrm{B}}}.
\end{aligned}$$

Thus, the reliability ratio for mixture measurement error with a homogeneous covariate and error structure is calculated through

$$\lambda^{\mathrm{M}} = \mathbb{E}(X|X^{*\mathrm{M}}) = \frac{\mathrm{Cov}(X, X^{*\mathrm{M}})}{\mathrm{Var}(X^{*\mathrm{M}})} = \frac{\sigma^2_{X^{*\mathrm{B}}} + (1-p)\sigma^2_{U^{\mathrm{B}}}}{\sigma^2_{X^{*\mathrm{B}}} + (1-p)(\sigma^2_{U^{\mathrm{B}}} + \sigma^2_{U^{\mathrm{C}}})}.$$

Since $\lambda^{\mathrm{M}} \leq 1$, mixed measurement error entails an attenuation of the true effect estimate, similar to classical measurement error, but the attenuation of the effect estimate due to mixture measurement error is less severe if $\sigma^2_{X^{*\mathrm{B}}} > 0, \sigma^2_{U^{\mathrm{C}}} > 0$ and $p < 1$.

Equivalent calculations for the variance of the error term as in Subsection 3.3.1 result in the following variance of the error term:

$$\begin{aligned}
&\mathrm{Var}(Y - (\beta_0^{*\mathrm{M}} + \beta_X^{*\mathrm{M}} X^{*\mathrm{M}})) \\
&\quad = \mathrm{Var}(Y - \beta_0 - \beta_X X) + \mathrm{Var}(\beta_X X) + \mathrm{Var}(\beta_X^{*\mathrm{M}} X^{*\mathrm{M}}) - \\
&\quad\quad 2\mathrm{Cov}(\beta_X X, \beta_X^{*\mathrm{M}} X^{*\mathrm{M}}) \\
&\quad = \sigma^2_\varepsilon + \beta_X^2(\sigma^2_{X^{*\mathrm{B}}} + \sigma^2_{U^{\mathrm{B}}}) + {\beta_X^{*\mathrm{M}}}^2(\sigma^2_{X^{*\mathrm{B}}} + (1-p)(\sigma^2_{U^{\mathrm{B}}} + \sigma^2_{U^{\mathrm{C}}})) - \\
&\quad\quad 2\beta_X \beta_X^{*\mathrm{M}}(\sigma^2_{X^{*\mathrm{B}}} + (1-p)\sigma^2_{U^{\mathrm{B}}}) \\
&\quad = \sigma^2_\varepsilon + \beta_X^2(1-\lambda^{\mathrm{M}})^2\sigma^2_{X^{*\mathrm{B}}} + \\
&\quad\quad \beta_X^2(1 + (1-p)\lambda^{\mathrm{M}}(\lambda^{\mathrm{M}} - 2))\sigma^2_{U^{\mathrm{B}}} + \\
&\quad\quad \beta_X^2 {\lambda^{\mathrm{M}}}^2(1-p)\sigma^2_{U^{\mathrm{C}}} \\
&\quad = \sigma^2_{\varepsilon^*}.
\end{aligned}$$

3.4.2. Heterogeneous covariate and error structure

For a heterogeneous error–prone covariate and error term $\rho^{X^{*\mathrm{B}}} = \rho = \rho^{\mathrm{B}} = \rho^{\mathrm{C}} = 0$ is assumed. $\mathrm{Cov}(\boldsymbol{X}, \boldsymbol{X}^{*\mathrm{M}})$ denotes the cross–covariance between the measurements with mixture error $\boldsymbol{X}^{*\mathrm{M}}$ and the precise measurements $\boldsymbol{X}$ and is derived by

$$\mathrm{Cov}(\boldsymbol{X}, \boldsymbol{X}^{*\mathrm{M}}) = \mathrm{Cov}(\boldsymbol{X}^{*\mathrm{B}} + (\mathbf{1}_T - \boldsymbol{G}) \circ (\nu^{\mathrm{B}}\mathbf{1}_T) + (\mathbf{1}_T - \boldsymbol{G}) \circ \boldsymbol{U}^{\mathrm{B}}, \boldsymbol{X}^{*\mathrm{B}} + \nu^{\mathrm{B}}\mathbf{1}_T + \boldsymbol{U}^{\mathrm{B}}).$$

$\mathrm{Var}(\boldsymbol{X}^{*\mathrm{M}})$ results from

$$\begin{aligned}
\mathrm{Var}(\boldsymbol{X}^{*\mathrm{M}}) = &\mathrm{Var}(\boldsymbol{X}^{*\mathrm{B}} + (\mathbf{1}_T - \boldsymbol{G}) \circ (\nu^{\mathrm{B}}\mathbf{1}_T) + (\mathbf{1}_T - \boldsymbol{G}) \circ \boldsymbol{U}^{\mathrm{B}} + \\
&(\mathbf{1}_T - \boldsymbol{G}) \circ (\nu^{\mathrm{C}}\mathbf{1}_T) + (\mathbf{1}_T - \boldsymbol{G}) \circ \boldsymbol{U}^{\mathrm{C}}).
\end{aligned}$$

For an arbitrary T–dimensional random variable $\boldsymbol{U}$ with expected value $\mathbf{0}$ $\mathrm{Cov}(\boldsymbol{G} \circ \boldsymbol{U}, \boldsymbol{U})$ and $\mathrm{Var}(\boldsymbol{G} \circ \boldsymbol{U})$ are calculated as follows:

$$\begin{aligned}
\mathrm{Cov}(\boldsymbol{G} \circ \boldsymbol{U}, \boldsymbol{U}) &= \mathbb{E}\left[(\boldsymbol{G} \circ \boldsymbol{U})\boldsymbol{U}^\top\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[(\boldsymbol{G} \circ \boldsymbol{U})\boldsymbol{U}^\top | \boldsymbol{G}\right]\right] \\
&= \mathbb{E}\left[(\boldsymbol{G}\mathbf{1}_T^\top) \circ \mathrm{Var}(\boldsymbol{U})\right] \\
&= (1-p)\mathrm{Var}(\boldsymbol{U}) \\
\mathrm{Var}(\boldsymbol{G} \circ \boldsymbol{U}) &= \mathbb{E}\left[\mathrm{Var}\left[\boldsymbol{G} \circ \boldsymbol{U} | \boldsymbol{G}\right]\right] + \mathrm{Var}\left[\mathbb{E}\left[\boldsymbol{G} \circ \boldsymbol{U} | \boldsymbol{G}\right]\right] \\
&= \mathbb{E}\left[(\boldsymbol{G}\boldsymbol{G}^\top) \circ \mathrm{Var}(\boldsymbol{U})\right] \\
&= ((1-p)^2\mathbf{J}_T + p(1-p)\mathbf{I}_T)\mathrm{Var}(\boldsymbol{U}).
\end{aligned}$$

These relationships are used for the derivation of $\mathrm{Cov}(\boldsymbol{X}, \boldsymbol{X}^{*\mathrm{M}})$ and $\mathrm{Var}(\boldsymbol{X}^{*\mathrm{M}})$:

$$\begin{aligned}
\mathrm{Cov}(\boldsymbol{X}, \boldsymbol{X}^{*\mathrm{M}}) =& \sigma^2_{X^{*\mathrm{B}}}\mathbf{I}_T + \mathbb{E}\left[((\mathbf{1}_T - \boldsymbol{G})\mathbf{1}_T^\top) \circ (\sigma^2_{\nu^{\mathrm{B}}}\mathbf{J}_T + \sigma^2_{U^{\mathrm{B}}}\mathbf{I}_T)\right] \\
=& \sigma^2_{X^{*\mathrm{B}}}\mathbf{I}_T + (1-p)\sigma^2_{\nu^{\mathrm{B}}}\mathbf{J}_T + (1-p)\sigma^2_{U^{\mathrm{B}}}\mathbf{I}_T \qquad (3.29) \\
\mathrm{Var}(\boldsymbol{X}^{*\mathrm{M}}) =& \sigma^2_{X^{*\mathrm{B}}}\mathbf{I}_T + \mathbb{E}\left[((\mathbf{1}_T - \boldsymbol{G})(\mathbf{1}_T - \boldsymbol{G})^\top) \circ \left(\sigma^2_{U^{\mathrm{B}}}\mathbf{I}_T + \sigma^2_{U^{\mathrm{C}}}\mathbf{I}_T + (\sigma^2_{\nu^{\mathrm{B}}} + \sigma^2_{\nu^{\mathrm{C}}})\mathbf{J}_T\right)\right] \\
=& \sigma^2_{X^{*\mathrm{B}}}\mathbf{I}_T + (\sigma^2_{\nu^{\mathrm{B}}} + \sigma^2_{\nu^{\mathrm{C}}})\left((1-p)^2\mathbf{J}_T + p(1-p)\mathbf{I}_T\right) + \\
& (1-p)(\sigma^2_{U^{\mathrm{B}}} + \sigma^2_{U^{\mathrm{C}}})\mathbf{I}_T. \qquad (3.30)
\end{aligned}$$

Calculations along the lines of Wang et al. (1998) result in the following attenuation factor:

$$\lambda^{\mathrm{M}} = \frac{\sigma^2_{\nu^{\mathrm{B}}} + (\sigma^2_{X^{*\mathrm{B}}} + (1-p)\sigma^2_{U^{\mathrm{B}}})\left[1 + (T-1)\frac{\sigma^2_{\tau^{*\mathrm{C}}}}{\sigma^2_{\varepsilon^{*\mathrm{C}}}}\right]}{\sigma^2_{\nu^{\mathrm{B}}} + \sigma^2_{\nu^{\mathrm{C}}} + \left(\sigma^2_{X^{*\mathrm{B}}} + (1-p)(\sigma^2_{U^{\mathrm{B}}} + \sigma^2_{U^{\mathrm{C}}}) + p(1-p)(\sigma^2_{\nu^{\mathrm{B}}} + \sigma^2_{\nu^{\mathrm{C}}})\right)\left[1 + (T-1)\frac{\sigma^2_{\tau^{*\mathrm{C}}}}{\sigma^2_{\varepsilon^{*\mathrm{C}}}}\right]}.$$

The individual–specific measurement errors and the random effects of the main model, i.e. $\boldsymbol{\nu}^{\mathrm{B}}$, $\boldsymbol{\nu}^{\mathrm{C}}$ and $\boldsymbol{\tau}$, cannot be identified with the main model. The mixture of the two error types implicates, that also the two individual–specific measurement errors are mixed, i.e. they cannot be separately calculated. As a result, the variance of the individual–specific measurement errors is not entirely transferred to the variance of the error term and the variance between the individual mean of the Berkson error–prone measurements and the individual mean of the measurements with classical measurement error affects the denominator of the attenuation factor.

In contrast to the classical measurement error with heterogeneous covariate and error structure, the attenuation factor for mixture measurement error is also for $T \to \infty$ affected by the individual–specific measurement errors $\boldsymbol{\nu}^{\mathrm{B}}$ and $\boldsymbol{\nu}^{\mathrm{C}}$. The additional term $p(1-p)(\sigma^2_{\nu^{\mathrm{B}}} + \sigma^2_{\nu^{\mathrm{C}}})$ intensifies the attenuation of the effect coefficient.

The variance of the error term is

$$\begin{aligned}
&\mathrm{Var}(\boldsymbol{Y} - (\beta_0^{*\mathrm{M}} + \beta_X^{*\mathrm{M}}\boldsymbol{X}^{*\mathrm{M}})) \\
&\quad = \mathrm{Var}(\boldsymbol{Y} - \beta_0 - \beta_X\boldsymbol{X}) + \mathrm{Var}(\beta_X\boldsymbol{X}) + \mathrm{Var}(\beta_X^{*\mathrm{M}}\boldsymbol{X}^{*\mathrm{M}}) - \\
&\qquad 2\mathrm{Cov}(\beta_X\boldsymbol{X}, \beta_X^{*\mathrm{M}}\boldsymbol{X}^{*\mathrm{M}}) \\
&\quad = \sigma^2_\varepsilon\mathbf{I}_T + \sigma^2_\tau\mathbf{J}_T + \beta_X^2(\sigma^2_{X^{*\mathrm{B}}}\mathbf{I}_T + \sigma^2_{\nu^{\mathrm{B}}}\mathbf{J}_T + \sigma^2_{U^{\mathrm{B}}}\mathbf{I}_T) + \\
&\qquad \beta_X^{*\mathrm{M}2}\Big[\sigma^2_{X^{*\mathrm{B}}}\mathbf{I}_T + (\sigma^2_{\nu^{\mathrm{B}}} + \sigma^2_{\nu^{\mathrm{C}}})\left((1-p)^2\mathbf{J}_T + p(1-p)\mathbf{I}_T\right) + \\
&\qquad\qquad (1-p)\left(\sigma^2_{U^{\mathrm{B}}} + \sigma^2_{U^{\mathrm{C}}}\right)\mathbf{I}_T\Big] - \\
&\qquad 2\beta_X^2\lambda^{\mathrm{M}}\Big[\sigma^2_{X^{*\mathrm{B}}}\mathbf{I}_T + (1-p)\sigma^2_{\nu^{\mathrm{B}}}\mathbf{J}_T + (1-p)\sigma^2_{U^{\mathrm{B}}}\mathbf{I}_T\Big] \\
&\quad = \sigma^2_\varepsilon\mathbf{I}_T + \sigma^2_\tau\mathbf{J}_T + \\
&\qquad \beta_X^2\Big[(1-\lambda^{\mathrm{M}})^2\sigma^2_{X^{*\mathrm{B}}}\mathbf{I}_T + \\
&\qquad\qquad (1 - 2\lambda^{\mathrm{M}}(1-p) + \lambda^{\mathrm{M}2}(1-p))\sigma^2_{U^{\mathrm{B}}}\mathbf{I}_T + \\
&\qquad\qquad (1-p)\lambda^{\mathrm{M}2}\sigma^2_{U^{\mathrm{C}}}\mathbf{I}_T + \\
&\qquad\qquad (1-(1-p)\lambda^{\mathrm{M}})^2\sigma^2_{\nu^{\mathrm{B}}}\mathbf{J}_T + p(1-p)\lambda^{\mathrm{M}2}\sigma^2_{\nu^{\mathrm{B}}}\mathbf{I}_T + \\
&\qquad\qquad (1-p)^2\lambda^{\mathrm{M}2}\sigma^2_{\nu^{\mathrm{C}}}\mathbf{J}_T + p(1-p)\lambda^{\mathrm{M}2}\sigma^2_{\nu^{\mathrm{C}}}\mathbf{I}_T\Big] \qquad (3.31)
\end{aligned}$$

The probability limits for the variance parameters in the naive model are, according to eq. 3.10, eq. 3.11 and eq. 3.31,

$$\begin{aligned}\sigma^2_{\tau^*} &= \sigma^2_\tau + {\lambda^{\mathrm{M}}}^2(1-p)^2\beta_X^2\sigma^2_{\nu^{\mathrm{C}}} + (1-(1-p)\lambda^{\mathrm{M}})^2\beta_X^2\sigma^2_{\nu^{\mathrm{B}}},\\ \sigma^2_{\varepsilon^*} &= \sigma^2_\varepsilon + \beta_X^2\Big[(1-\lambda^{\mathrm{M}})^2\sigma^2_{X^{*\mathrm{B}}} + (1-2\lambda^{\mathrm{M}}(1-p)+{\lambda^{\mathrm{M}}}^2(1-p))\sigma^2_{U^{\mathrm{B}}} + (1-p){\lambda^{\mathrm{M}}}^2\sigma^2_{U^{\mathrm{C}}} + \\ &\quad p(1-p){\lambda^{\mathrm{M}}}^2\sigma^2_{\nu^{\mathrm{B}}} + p(1-p){\lambda^{\mathrm{M}}}^2\sigma^2_{\nu^{\mathrm{C}}}\Big].\end{aligned}$$

3.4.3. Heterogeneous and autocorrelated error structure

The derivation of the attenuation factor λ^{M} in a simple linear mixed model with autocorrelated errors and individual–specific and autocorrelated mixture covariate measurement error is similar to the classical measurement error (Subsection 3.3.3). Therefore, only the changing steps are mentioned in this subsection. The attenuation factor λ^{M} for the effect coefficient of a covariate with mixture measurement error is

$$\lambda^{\mathrm{M}} \overset{\text{eq. } 3.9}{=} \frac{\mathrm{tr}\left[\boldsymbol{V}^{*\mathrm{M}\,-1}\mathrm{Cov}(\boldsymbol{X},\boldsymbol{X}^{*\mathrm{M}})\right]}{\mathrm{tr}\left[\boldsymbol{V}^{*\mathrm{M}\,-1}\mathrm{Var}(\boldsymbol{X}^{*\mathrm{M}})\right]}. \tag{3.32}$$

$\mathrm{Cov}(\boldsymbol{X},\boldsymbol{X}^{*\mathrm{M}})$ is, analogically to eq. 3.29, given through

$$\mathrm{Cov}(\boldsymbol{X},\boldsymbol{X}^{*\mathrm{M}}) = \boldsymbol{\Sigma}_{X^{*\mathrm{B}}} + (1-p)\sigma^2_{\nu^{\mathrm{B}}}\mathbf{J}_T + (1-p)\boldsymbol{\Sigma}_{U^{\mathrm{B}}}, \tag{3.33}$$

The variance of $\boldsymbol{X}^{*\mathrm{M}}$ is, analogously to eq. 3.30, determined by

$$\begin{aligned}\mathrm{Var}(\boldsymbol{X}^{*\mathrm{M}}) =& \boldsymbol{\Sigma}_{X^{*\mathrm{B}}} + (\sigma^2_{\nu^{\mathrm{B}}} + \sigma^2_{\nu^{\mathrm{C}}})\left((1-p)^2\mathbf{J}_T + p(1-p)\mathbf{I}_T\right) + \\ & (1-p)^2\left(\boldsymbol{\Sigma}_{U^{\mathrm{B}}} + \boldsymbol{\Sigma}_{U^{\mathrm{C}}}\right) + p(1-p)(\sigma^2_{U^{\mathrm{B}}} + \sigma^2_{U^{\mathrm{C}}})\mathbf{I}_T.\end{aligned}$$

Eq. 3.32 is continued with

$$\begin{aligned}\lambda^{\mathrm{M}} =& \left\{\mathrm{tr}\left[(\sigma^2_{\tau^{*\mathrm{M}}}\mathbf{J}_T + \boldsymbol{\Sigma}_{\varepsilon^{*\mathrm{M}}})^{-1}(\boldsymbol{\Sigma}_{X^{*\mathrm{B}}} + (1-p)\sigma^2_{\nu^{\mathrm{B}}}\mathbf{J}_T + (1-p)\boldsymbol{\Sigma}_{U^{\mathrm{B}}})\right]\right\} \\ & \left\{\mathrm{tr}\left[(\sigma^2_{\tau^{*\mathrm{M}}}\mathbf{J}_T + \boldsymbol{\Sigma}_{\varepsilon^{*\mathrm{M}}})^{-1}(\boldsymbol{\Sigma}_{X^{*\mathrm{B}}} + (\sigma^2_{\nu^{\mathrm{B}}} + \sigma^2_{\nu^{\mathrm{C}}})((1-p)^2\mathbf{J}_T + p(1-p)\mathbf{I}_T) + \right.\right. \\ & \left.\left. (1-p)^2\left(\boldsymbol{\Sigma}_{U^{\mathrm{B}}} + \boldsymbol{\Sigma}_{U^{\mathrm{C}}}\right) + p(1-p)(\sigma^2_{U^{\mathrm{B}}} + \sigma^2_{U^{\mathrm{C}}})\mathbf{I}_T)\right]\right\}.\end{aligned} \tag{3.34}$$

Calculations similar to the classical measurement error in Subsection 3.3.3 yield the attenuation factor

$$\begin{aligned}\lambda^{\mathrm{M}} =& \left\{\sigma^2_{X^{*\mathrm{B}}}g_T^{*\mathrm{M}}\left(\rho^{X^{*\mathrm{B}}}\right) + (1-p)\sigma^2_{\nu^{\mathrm{B}}}g_T^{*\mathrm{M}}(1) + (1-p)\sigma^2_{U^{\mathrm{B}}}g_T^{*\mathrm{M}}\left(\rho^{\mathrm{B}}\right)\right\} \\ & \left\{\sigma^2_{X^{*\mathrm{B}}}g_T^{*\mathrm{M}}\left(\rho^{X^{*\mathrm{B}}}\right) + (\sigma^2_{\nu^{\mathrm{B}}} + \sigma^2_{\nu^{\mathrm{C}}})[(1-p)^2g_T^{*\mathrm{M}}(1) + p(1-p)g_T^{*\mathrm{M}}(0)] + \right. \\ & \sigma^2_{U^{\mathrm{B}}}[(1-p)^2g_T^{*\mathrm{M}}\left(\rho^{\mathrm{B}}\right) + p(1-p)g_T^{*\mathrm{M}}(0)] + \\ & \left.\sigma^2_{U^{\mathrm{C}}}[(1-p)^2g_T^{*\mathrm{M}}\left(\rho^{\mathrm{C}}\right) + p(1-p)g_T^{*\mathrm{M}}(0)]\right\}^{-1}\end{aligned} \tag{3.35}$$

with the same weighting function $g_T^{*\mathrm{M}}$ as used for the bias calculation of classical measurement error, but using $\rho^{*\mathrm{M}}$ instead of $\rho^{*\mathrm{C}}$.

The error variance of the naive model results as

$$\begin{aligned}
\mathrm{Var}(\boldsymbol{Y}-\boldsymbol{\chi}^{*\mathrm{M}}\boldsymbol{\beta}^{*\mathrm{M}}) =& \mathrm{Var}(\boldsymbol{Y}-\beta_0-\beta_X\boldsymbol{X})+\mathrm{Var}(\beta_X\boldsymbol{X})+\mathrm{Var}(\beta_X^{*\mathrm{M}}\boldsymbol{X}^{*\mathrm{M}})- \\
& 2\mathrm{Cov}(\beta_X\boldsymbol{X},\beta_X^{*\mathrm{M}}\boldsymbol{X}^{*\mathrm{M}}) \\
=& \boldsymbol{\Sigma}_\varepsilon+\sigma_\tau^2\mathbf{J}_T+\beta_X^2(\boldsymbol{\Sigma}_{X^{*\mathrm{B}}}+\boldsymbol{\Sigma}_{U^{\mathrm{B}}}+\sigma_{\nu^{\mathrm{B}}}^2\mathbf{J}_T)+ \\
& \beta_X^{*\mathrm{M}2}\Big[\boldsymbol{\Sigma}_{X^{*\mathrm{B}}}+(\sigma_{\nu^{\mathrm{B}}}^2+\sigma_{\nu^{\mathrm{C}}}^2)\Big((1-p)^2\mathbf{J}_T+p(1-p)\mathbf{I}_T\Big)+ \\
& \qquad (1-p)^2(\boldsymbol{\Sigma}_{U^{\mathrm{B}}}+\boldsymbol{\Sigma}_{U^{\mathrm{C}}})+p(1-p)(\sigma_{U^{\mathrm{B}}}^2+\sigma_{U^{\mathrm{C}}}^2)\mathbf{I}_T\Big]- \\
& 2\beta_X\beta_X^{*\mathrm{M}}\Big[\boldsymbol{\Sigma}_{X^{*\mathrm{B}}}+(1-p)\boldsymbol{\Sigma}_{U^{\mathrm{B}}}+(1-p)\sigma_{\nu^{\mathrm{B}}}^2\mathbf{J}_T\Big] \\
=& \boldsymbol{\Sigma}_\varepsilon+\sigma_\tau^2\mathbf{J}_T+ \\
& \beta_X^2\Big[(1-\lambda^{\mathrm{M}})^2\boldsymbol{\Sigma}_{X^{*\mathrm{B}}}+ \\
& \qquad (1-(1-p)\lambda^{\mathrm{M}})^2(\boldsymbol{\Sigma}_{U^{\mathrm{B}}}+\sigma_{\nu^{\mathrm{B}}}^2\mathbf{J}_T)+p(1-p)\lambda^{\mathrm{M}2}(\sigma_{U^{\mathrm{B}}}^2+\sigma_{\nu^{\mathrm{B}}}^2)\mathbf{I}_T+ \\
& \qquad \lambda^{\mathrm{M}2}(1-p)^2(\boldsymbol{\Sigma}_{U^{\mathrm{C}}}+\sigma_{\nu^{\mathrm{C}}}^2\mathbf{J}_T)+p(1-p)\lambda^{\mathrm{M}2}(\sigma_{U^{\mathrm{C}}}^2+\sigma_{\nu^{\mathrm{C}}}^2)\mathbf{I}_T\Big]
\end{aligned}$$

Again, as mentioned in Subsection 3.1.3, simple expressions for the variance parameters $\sigma_{\varepsilon^{*\mathrm{M}}}^2, \sigma_{\tau^{*\mathrm{M}}}^2$ and $\rho^{*\mathrm{M}}$ in the naive model are not available.

3.5. Extensions of the models

3.5.1. Alternatives for the calculation of the attenuation factor

Since the true values of $\boldsymbol{X}$ are unknown, $\mathrm{Cov}(\boldsymbol{X},\boldsymbol{X}^*)$ has to be calculated with additional information, e.g. through a validation study. This can be accomplished either based on $\boldsymbol{X}^* \in \{\boldsymbol{X}^{*\mathrm{C}},\boldsymbol{X}^{*\mathrm{M}}\}$ or based on $\boldsymbol{X}^{*\mathrm{B}}$:

- Classical measurement error:

  1. $$\mathrm{Cov}\left(\boldsymbol{X},\boldsymbol{X}^{*\mathrm{C}}\right)=\mathrm{Var}\left(\boldsymbol{X}^{*\mathrm{C}}\right)-\sigma_{\nu^{\mathrm{C}}}^2\mathbf{J}_T-\boldsymbol{\Sigma}_{U^{\mathrm{C}}} \tag{3.36}$$
  2. $$\mathrm{Cov}\left(\boldsymbol{X},\boldsymbol{X}^{*\mathrm{C}}\right)=\boldsymbol{\Sigma}_{X^{*\mathrm{B}}}+\sigma_{\nu^{\mathrm{B}}}^2\mathbf{J}_T+\boldsymbol{\Sigma}_{U^{\mathrm{B}}} \tag{3.37}$$

- Mixture measurement error:

  1. $$\begin{aligned}\mathrm{Cov}\left(\boldsymbol{X},\boldsymbol{X}^{*\mathrm{M}}\right) =& \mathrm{Var}\left(\boldsymbol{X}^{*\mathrm{M}}\right)+p(1-p)\sigma_{\nu^{\mathrm{B}}}^2(\mathbf{J}_T-\mathbf{I}_T)+p(1-p)\boldsymbol{\Sigma}_{U^{\mathrm{B}}}- \\ & p(1-p)\sigma_{U^{\mathrm{B}}}^2\mathbf{I}_T-(1-p)^2\sigma_{\nu^{\mathrm{C}}}^2\mathbf{J}_T-p(1-p)\sigma_{\nu^{\mathrm{C}}}^2\mathbf{I}_T- \\ & (1-p)^2\boldsymbol{\Sigma}_{U^{\mathrm{C}}}-p(1-p)\sigma_{U^{\mathrm{C}}}^2\mathbf{I}_T\end{aligned} \tag{3.38}$$
  2. $$\mathrm{Cov}\left(\boldsymbol{X},\boldsymbol{X}^{*\mathrm{M}}\right)=\boldsymbol{\Sigma}_{X^{*\mathrm{B}}}+(1-p)\boldsymbol{\Sigma}_{U^{\mathrm{B}}}+(1-p)\sigma_{\nu^{\mathrm{B}}}^2\mathbf{J}_T \tag{3.39}$$

The relations in eq. 3.36 and eq. 3.37 are evident. Eq. 3.39 results from eq. 3.33. The derivation of eq. 3.38 is given in the following.

$$\begin{aligned}
\mathrm{Cov}\left(\boldsymbol{X}^{*\mathrm{M}}, \boldsymbol{X}\right) =& \mathrm{Cov}\left(\boldsymbol{X}^{*\mathrm{M}}, \boldsymbol{X}^{*\mathrm{M}} + \nu^{\mathrm{B}}\boldsymbol{G} + \boldsymbol{G} \circ \boldsymbol{U}^{\mathrm{B}} - \nu^{\mathrm{C}}(\mathbf{1}_T - \boldsymbol{G}) - (\mathbf{1}_T - \boldsymbol{G}) \circ \boldsymbol{U}^{\mathrm{C}}\right) \\
=& \mathrm{Cov}\left(\boldsymbol{X}^{*\mathrm{M}}, \boldsymbol{X}^{*\mathrm{M}}\right) + \mathrm{Cov}\left(\nu^{\mathrm{B}}\left(\mathbf{1}_T - \boldsymbol{G}\right), \nu^{\mathrm{B}}\boldsymbol{G}\right) + \\
& \mathrm{Cov}\left((\mathbf{1}_T - \boldsymbol{G}) \circ \boldsymbol{U}^{\mathrm{B}}, \boldsymbol{G} \circ \boldsymbol{U}^{\mathrm{B}}\right) - \mathrm{Cov}\left(\nu^{\mathrm{C}}\left(\mathbf{1}_T - \boldsymbol{G}\right), \nu^{\mathrm{C}}\left(\mathbf{1}_T - \boldsymbol{G}\right)\right) - \\
& \mathrm{Cov}\left((\mathbf{1}_T - \boldsymbol{G}) \circ \boldsymbol{U}^{\mathrm{C}}, (\mathbf{1}_T - \boldsymbol{G}) \circ \boldsymbol{U}^{\mathrm{C}}\right) \\
=& \mathrm{Var}\left(\boldsymbol{X}^{*\mathrm{M}}\right) + p(1-p)\sigma^2_{\nu^{\mathrm{B}}}\mathbf{J}_T - p(1-p)\sigma^2_{\nu^{\mathrm{B}}}\mathbf{I}_T + p(1-p)\boldsymbol{\Sigma}_{U^{\mathrm{B}}} - \\
& p(1-p)\sigma^2_{U^{\mathrm{B}}}\mathbf{I}_T - (1-p)^2\sigma^2_{\nu^{\mathrm{C}}}\mathbf{J}_T - p(1-p)\sigma^2_{\nu^{\mathrm{C}}}\mathbf{I}_T - (1-p)^2\boldsymbol{\Sigma}_{U^{\mathrm{C}}} - \\
& p(1-p)\sigma^2_{U^{\mathrm{C}}}\mathbf{I}_T
\end{aligned}$$

The first mentioned alternatives use the available personal measurements and therefore, more detailed information.

3.5.2. Including knowledge about the exact breakdown times

Previous considerations (eq. 3.34) are based on the stochasticity and on the "missing completely at random" assumption of the failure process of the portable measurement device. However, known failure times occur in the Augsburger Umweltstudie, i.e. the entries of $\boldsymbol{G}$ are known. The introduction of this knowledge into the bias calculations for the mixture error is discussed in this subsection.

Known breakdown times of the personal measurement device result in non–identically distributed data for the individuals. If equal lengths of the measurement series T are presumed, eq. 3.23 changes to

$$\lambda^{\mathrm{M}} = \frac{\sum_{i=1}^{n} \mathrm{tr}\left[\boldsymbol{V}^{*-1}\mathrm{Cov}(\boldsymbol{X}_i, \boldsymbol{X}_i^*)\right]}{\sum_{i=1}^{n} \mathrm{tr}\left[\boldsymbol{V}^{*-1}\mathrm{Var}(\boldsymbol{X}_i^*)\right]}. \tag{3.40}$$

because $\boldsymbol{V}^*$ is identical for all individuals. The attenuation factor λ^{M} results as

$$\lambda^{\mathrm{M}} = \frac{\sum_{i=1}^{n} \mathrm{tr}\left[\boldsymbol{V}^{*\mathrm{M}-1}(\boldsymbol{\Sigma}_{X^{*\mathrm{B}}} + ((\mathbf{1}_T - \boldsymbol{G}_i)\mathbf{1}_T^\top) \circ (\sigma^2_{\nu^{\mathrm{B}}}\mathbf{J}_T + \boldsymbol{\Sigma}_{U^{\mathrm{B}}}))\right]}{\sum_{i=1}^{n} \mathrm{tr}\left[\boldsymbol{V}^{*\mathrm{M}-1}(\boldsymbol{\Sigma}_{X^{*\mathrm{B}}} + ((\mathbf{1}_T - \boldsymbol{G}_i)(\mathbf{1}_T - \boldsymbol{G}_i)^\top) \circ (\sigma^2_{\nu^{\mathrm{B}}}\mathbf{J}_T + \sigma^2_{\nu^{\mathrm{C}}}\mathbf{J}_T + \boldsymbol{\Sigma}_{U^{\mathrm{B}}} + \boldsymbol{\Sigma}_{U^{\mathrm{C}}}))\right]} \tag{3.41}$$

Alternatively,

$$\begin{aligned}
\mathrm{Cov}\left(\boldsymbol{X}_i^{*\mathrm{M}}, \boldsymbol{X}_i\right) =& \mathrm{Var}\left(\boldsymbol{X}^{*\mathrm{M}}\right) + \sigma^2_{\nu^{\mathrm{B}}}(\mathbf{1}_T - \boldsymbol{G}_i)\boldsymbol{G}_i^\top \circ \mathbf{J}_T + (\mathbf{1}_T - \boldsymbol{G}_i)\boldsymbol{G}_i^\top \circ \boldsymbol{\Sigma}_{U^{\mathrm{B}}} - \\
& \sigma^2_{\nu^{\mathrm{C}}}(\mathbf{1}_T - \boldsymbol{G}_i)(\mathbf{1}_T - \boldsymbol{G}_i)^\top \circ \mathbf{J}_T - (\mathbf{1}_T - \boldsymbol{G}_i)(\mathbf{1}_T - \boldsymbol{G}_i)^\top \circ \boldsymbol{\Sigma}_{U^{\mathrm{C}}}
\end{aligned}$$

can be used in the numerator of eq. 3.40.

In other applications distributional assumptions of the failure process may be known. For example, the breakdown of the device of a certain time point depends on the situation in the past or on the measured quantity.

3.5.3. Further covariates

The previous considerations were restricted to a single covariate X. The inclusion of q additional precisely measured covariates $\boldsymbol{Z}$ does not affect the bias of the effect coefficient

induced by the deficient covariate X^*, if $\boldsymbol{Z}$ and X^* are independent. However, the attenuation of the effect may change and biased effect estimates may possibly occur if X^* and $\boldsymbol{Z}$ are dependent. Attenuation could even be inverted if the measurement errors of two covariates are correlated (Carroll et al., 2006) or if several covariates are measured with error (Buzas et al., 2005).

The following considerations are restricted to a single error–prone covariate. The first score equation of the linear mixed model with additional covariates $\boldsymbol{Z}$ correlating with the classical error–prone covariate is (see Carroll et al., 2006), when neglecting index i,

$$\mathbb{E}\left[\begin{pmatrix}\mathbf{1}_T & \boldsymbol{X}^* & \boldsymbol{Z}\end{pmatrix}^\top \boldsymbol{V}^{*-1}\left(\boldsymbol{Y} - \begin{pmatrix}\mathbf{1}_T & \boldsymbol{X}^* & \boldsymbol{Z}\end{pmatrix}\begin{pmatrix}\beta_0^* \\ \beta_X^* \\ \boldsymbol{\beta}_Z^*\end{pmatrix}\right)\right] = 0$$

$$\Leftrightarrow \mathbb{E}\left[\begin{pmatrix}\mathbf{1}_T & \boldsymbol{X}^* & \boldsymbol{Z}\end{pmatrix}^\top \boldsymbol{V}^{*-1}\boldsymbol{Y}\right] = \mathbb{E}\left[\begin{pmatrix}\mathbf{1}_T & \boldsymbol{X}^* & \boldsymbol{Z}\end{pmatrix}^\top \boldsymbol{V}^{*-1}\begin{pmatrix}\mathbf{1}_T & \boldsymbol{X}^* & \boldsymbol{Z}\end{pmatrix}\begin{pmatrix}\beta_0^* \\ \beta_X^* \\ \boldsymbol{\beta}_Z^*\end{pmatrix}\right]$$

$$\Leftrightarrow \mathbb{E}\left[\begin{pmatrix} \mathbf{1}_T^\top\boldsymbol{V}^{*-1}\mathbf{1}_T & \mathbf{1}_T^\top\boldsymbol{V}^{*-1}\boldsymbol{X} & \mathbf{1}_T^\top\boldsymbol{V}^{*-1}\boldsymbol{Z}_1 & \dots & \mathbf{1}_T^\top\boldsymbol{V}^{*-1}\boldsymbol{Z}_q \\ \boldsymbol{X}^{*\top}\boldsymbol{V}^{*-1}\mathbf{1}_T & \boldsymbol{X}^{*\top}\boldsymbol{V}^{*-1}\boldsymbol{X} & \boldsymbol{X}^{*\top}\boldsymbol{V}^{*-1}\boldsymbol{Z}_1 & \dots & \boldsymbol{X}^{*\top}\boldsymbol{V}^{*-1}\boldsymbol{Z}_q \\ \boldsymbol{Z}_1^\top\boldsymbol{V}^{*-1}\mathbf{1}_T & \boldsymbol{Z}_1^\top\boldsymbol{V}^{*-1}\boldsymbol{X} & \boldsymbol{Z}_1^\top\boldsymbol{V}^{*-1}\boldsymbol{Z}_1 & \dots & \boldsymbol{Z}_1^\top\boldsymbol{V}^{*-1}\boldsymbol{Z}_q \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{Z}_q^\top\boldsymbol{V}^{*-1}\mathbf{1}_T & \boldsymbol{Z}_q^\top\boldsymbol{V}^{*-1}\boldsymbol{X} & \boldsymbol{Z}_q^\top\boldsymbol{V}^{*-1}\boldsymbol{Z}_1 & \dots & \boldsymbol{Z}_q^\top\boldsymbol{V}^{*-1}\boldsymbol{Z}_q \end{pmatrix}\right]\begin{pmatrix}\beta_0 \\ \beta_X \\ \boldsymbol{\beta}_Z\end{pmatrix} =$$

$$\mathbb{E}\left[\begin{pmatrix} \mathbf{1}_T^\top\boldsymbol{V}^{*-1}\mathbf{1}_T & \mathbf{1}_T^\top\boldsymbol{V}^{*-1}\boldsymbol{X}^* & \mathbf{1}_T^\top\boldsymbol{V}^{*-1}\boldsymbol{Z}_1 & \dots & \mathbf{1}_T^\top\boldsymbol{V}^{*-1}\boldsymbol{Z}_q \\ \boldsymbol{X}^{*\top}\boldsymbol{V}^{*-1}\mathbf{1}_T & \boldsymbol{X}^{*\top}\boldsymbol{V}^{*-1}\boldsymbol{X}^* & \boldsymbol{X}^{*\top}\boldsymbol{V}^{*-1}\boldsymbol{Z}_1 & \dots & \boldsymbol{X}^{*\top}\boldsymbol{V}^{*-1}\boldsymbol{Z}_q \\ \boldsymbol{Z}_1^\top\boldsymbol{V}^{*-1}\mathbf{1}_T & \boldsymbol{Z}_1^\top\boldsymbol{V}^{*-1}\boldsymbol{X}^* & \boldsymbol{Z}_1^\top\boldsymbol{V}^{*-1}\boldsymbol{Z}_1 & \dots & \boldsymbol{Z}_1^\top\boldsymbol{V}^{*-1}\boldsymbol{Z}_q \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{Z}_q^\top\boldsymbol{V}^{*-1}\mathbf{1}_T & \boldsymbol{Z}_q^\top\boldsymbol{V}^{*-1}\boldsymbol{X}^* & \boldsymbol{Z}_q^\top\boldsymbol{V}^{*-1}\boldsymbol{Z}_1 & \dots & \boldsymbol{Z}_q^\top\boldsymbol{V}^{*-1}\boldsymbol{Z}_q \end{pmatrix}\right]\begin{pmatrix}\beta_0^* \\ \beta_X^* \\ \boldsymbol{\beta}_Z^*\end{pmatrix},$$

which can be further simplified with eq. 3.7 to

$$\begin{pmatrix}\beta_0 \\ \beta_X \\ \boldsymbol{\beta}_Z\end{pmatrix} =$$

$$\mathbb{E}\left[\begin{pmatrix} \operatorname{tr}(\boldsymbol{V}^{*-1}\mathbf{J}_T) & \operatorname{tr}(\boldsymbol{V}^{*-1}\boldsymbol{X}\mathbf{1}_T^\top) & \operatorname{tr}(\boldsymbol{V}^{*-1}\boldsymbol{Z}_1\mathbf{1}_T^\top) & \dots & \operatorname{tr}(\boldsymbol{V}^{*-1}\boldsymbol{Z}_q\mathbf{1}_T^\top) \\ \operatorname{tr}(\boldsymbol{V}^{*-1}\mathbf{1}_T\boldsymbol{X}^{*\top}) & \operatorname{tr}(\boldsymbol{V}^{*-1}\boldsymbol{X}\boldsymbol{X}^{*\top}) & \operatorname{tr}(\boldsymbol{V}^{*-1}\boldsymbol{Z}_1\boldsymbol{X}^{*\top}) & \dots & \operatorname{tr}(\boldsymbol{V}^{*-1}\boldsymbol{Z}_q\boldsymbol{X}^{*\top}) \\ \operatorname{tr}(\boldsymbol{V}^{*-1}\mathbf{1}_T\boldsymbol{Z}_1^\top) & \operatorname{tr}(\boldsymbol{V}^{*-1}\boldsymbol{X}\boldsymbol{Z}_1^\top) & \operatorname{tr}(\boldsymbol{V}^{*-1}\boldsymbol{Z}_1\boldsymbol{Z}_1^\top) & \dots & \operatorname{tr}(\boldsymbol{V}^{*-1}\boldsymbol{Z}_q\boldsymbol{Z}_1^\top) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \operatorname{tr}(\boldsymbol{V}^{*-1}\mathbf{1}_T\boldsymbol{Z}_q^\top) & \operatorname{tr}(\boldsymbol{V}^{*-1}\boldsymbol{X}\boldsymbol{Z}_q^\top) & \operatorname{tr}(\boldsymbol{V}^{*-1}\boldsymbol{Z}_1\boldsymbol{Z}_q^\top) & \dots & \operatorname{tr}(\boldsymbol{V}^{*-1}\boldsymbol{Z}_q\boldsymbol{Z}_q^\top) \end{pmatrix}\right]^{-1} \cdot$$

$$\mathbb{E}\left[\begin{pmatrix} \operatorname{tr}(\boldsymbol{V}^{*-1}\mathbf{J}_T) & \operatorname{tr}(\boldsymbol{V}^{*-1}\boldsymbol{X}^*\mathbf{1}_T^\top) & \operatorname{tr}(\boldsymbol{V}^{*-1}\boldsymbol{Z}_1\mathbf{1}_T^\top) & \dots & \operatorname{tr}(\boldsymbol{V}^{*-1}\boldsymbol{Z}_q\mathbf{1}_T^\top) \\ \operatorname{tr}(\boldsymbol{V}^{*-1}\mathbf{1}_T\boldsymbol{X}^{*\top}) & \operatorname{tr}(\boldsymbol{V}^{*-1}\boldsymbol{X}^*\boldsymbol{X}^{*\top}) & \operatorname{tr}(\boldsymbol{V}^{*-1}\boldsymbol{Z}_1\boldsymbol{X}^{*\top}) & \dots & \operatorname{tr}(\boldsymbol{V}^{*-1}\boldsymbol{Z}_q\boldsymbol{X}^{*\top}) \\ \operatorname{tr}(\boldsymbol{V}^{*-1}\mathbf{1}_T\boldsymbol{Z}_1^\top) & \operatorname{tr}(\boldsymbol{V}^{*-1}\boldsymbol{X}^*\boldsymbol{Z}_1^\top) & \operatorname{tr}(\boldsymbol{V}^{*-1}\boldsymbol{Z}_1\boldsymbol{Z}_1^\top) & \dots & \operatorname{tr}(\boldsymbol{V}^{*-1}\boldsymbol{Z}_q\boldsymbol{Z}_1^\top) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \operatorname{tr}(\boldsymbol{V}^{*-1}\mathbf{1}_T\boldsymbol{Z}_q^\top) & \operatorname{tr}(\boldsymbol{V}^{*-1}\boldsymbol{X}^*\boldsymbol{Z}_q^\top) & \operatorname{tr}(\boldsymbol{V}^{*-1}\boldsymbol{Z}_1\boldsymbol{Z}_q^\top) & \dots & \operatorname{tr}(\boldsymbol{V}^{*-1}\boldsymbol{Z}_q\boldsymbol{Z}_q^\top) \end{pmatrix}\right]\begin{pmatrix}\beta_0^* \\ \beta_X^* \\ \boldsymbol{\beta}_Z^*\end{pmatrix}.$$

Since the measurement errors are assumed to be independent from the confounder variables $\boldsymbol{Z}$,

$$\mathbb{E}\left[\text{tr}(\boldsymbol{V}^{*-1}\boldsymbol{X}\boldsymbol{Z}^\top)\right] = \mathbb{E}\left[\text{tr}(\boldsymbol{V}^{*-1}\boldsymbol{X}^*\boldsymbol{Z}^\top)\right]$$

and thus,

$$\begin{pmatrix}\beta_0\\ \beta_X\\ \boldsymbol{\beta}_Z\end{pmatrix} =$$

$$\begin{pmatrix}
\text{tr}(\boldsymbol{V}^{*-1}\mathbf{J}_T) & \text{tr}(\boldsymbol{V}^{*-1}\mathbb{E}(X)\mathbf{J}_T) & \text{tr}(\boldsymbol{V}^{*-1}\mathbb{E}(Z_1)\mathbf{J}_T) & \cdots & \text{tr}(\boldsymbol{V}^{*-1}\mathbb{E}(Z_q)\mathbf{J}_T)\\
\text{tr}(\boldsymbol{V}^{*-1}\mathbb{E}(X^*)\mathbf{J}_T) & \text{tr}(\boldsymbol{V}^{*-1}\text{Cov}(\boldsymbol{X}^*,\boldsymbol{X})) & \text{tr}(\boldsymbol{V}^{*-1}\text{Cov}(\boldsymbol{X}^*,\boldsymbol{Z_1})) & \cdots & \text{tr}(\boldsymbol{V}^{*-1}\text{Cov}(\boldsymbol{X}^*,\boldsymbol{Z_q}))\\
\text{tr}(\boldsymbol{V}^{*-1}\mathbb{E}(Z_1)\mathbf{J}_T) & \text{tr}(\boldsymbol{V}^{*-1}\text{Cov}(\boldsymbol{X}^*,\boldsymbol{Z_1})) & \text{tr}(\boldsymbol{V}^{*-1}\text{Var}(\boldsymbol{Z_1})) & & \\
\vdots & & \ddots & \cdots & \vdots\\
\text{tr}(\boldsymbol{V}^{*-1}\mathbb{E}(Z_q)\mathbf{J}_T) & \text{tr}(\boldsymbol{V}^{*-1}\text{Cov}(\boldsymbol{X}^*,\boldsymbol{Z_q})) & \cdots & & \text{tr}(\boldsymbol{V}^{*-1}\text{Var}(\boldsymbol{Z_q}))
\end{pmatrix}^{-1}$$

$$\begin{pmatrix}
\text{tr}(\boldsymbol{V}^{*-1}\mathbf{J}_T) & \text{tr}(\boldsymbol{V}^{*-1}\mathbb{E}(X^*)\mathbf{J}_T) & \text{tr}(\boldsymbol{V}^{*-1}\mathbb{E}(Z_1)\mathbf{J}_T) & \cdots & \text{tr}(\boldsymbol{V}^{*-1}\mathbb{E}(Z_q)\mathbf{J}_T)\\
\text{tr}(\boldsymbol{V}^{*-1}\mathbb{E}(X^*)\mathbf{J}_T) & \text{tr}(\boldsymbol{V}^{*-1}\text{Var}(\boldsymbol{X}^*)) & \text{tr}(\boldsymbol{V}^{*-1}\text{Cov}(\boldsymbol{X}^*,\boldsymbol{Z_1})) & \cdots & \text{tr}(\boldsymbol{V}^{*-1}\text{Cov}(\boldsymbol{X}^*,\boldsymbol{Z_q}))\\
\text{tr}(\boldsymbol{V}^{*-1}\mathbb{E}(Z_1)\mathbf{J}_T) & \text{tr}(\boldsymbol{V}^{*-1}\text{Cov}(\boldsymbol{X}^*,\boldsymbol{Z_1})) & \text{tr}(\boldsymbol{V}^{*-1}\text{Var}(\boldsymbol{Z_1})) & & \\
\vdots & & \ddots & \cdots & \vdots\\
\text{tr}(\boldsymbol{V}^{*-1}\mathbb{E}(Z_q)\mathbf{J}_T) & \text{tr}(\boldsymbol{V}^{*-1}\text{Cov}(\boldsymbol{X}^*,\boldsymbol{Z_q})) & \cdots & & \text{tr}(\boldsymbol{V}^{*-1}\text{Var}(\boldsymbol{Z_q}))
\end{pmatrix}$$

$$\begin{pmatrix}\beta_0^*\\ \beta_X^*\\ \boldsymbol{\beta}_Z^*\end{pmatrix} \tag{3.42}$$

is obtained for centered covariates X^* and $\boldsymbol{Z}$. Two terms cause the differences to the simple case: (I) The probability limit of the variance–covariance matrix of the naive model errors $\boldsymbol{V}^*$ changes due to additional explanatory power of the additional covariates and (II) $\boldsymbol{Z}$ directly influences the estimation of β_X^* through $\text{Cov}(\boldsymbol{X}^*,\boldsymbol{Z})$, $\text{Var}(\boldsymbol{Z})$ and $\boldsymbol{\beta}_Z$. Indeed, the explicit calculation of the bias is possible, but would be complex and would require knowledge about the cross–covariances between the covariates, which may be difficult to estimate in real situations: in the Augsburger Umweltstudie, these parameters strongly vary between the individuals. Therefore, empirical, individual–specific equivalents of the expectations, variances and cross–covariances are used to estimate the coefficients β_0, β_X and β_Z by numerically solving the following equation system:

$$
\begin{pmatrix} \widehat{\beta_0} \\ \widehat{\beta_X} \\ \widehat{\boldsymbol{\beta}_Z} \end{pmatrix} =
$$

$$
\sum \left[\begin{pmatrix}
\mathrm{tr}(\widehat{\boldsymbol{V}^*}^{-1}\mathbf{J}_T) & \mathrm{tr}(\widehat{\boldsymbol{V}^*}^{-1}\boldsymbol{X}\mathbf{1}_T^\top) & \mathrm{tr}(\widehat{\boldsymbol{V}^*}^{-1}\boldsymbol{Z}_1\mathbf{1}_T^\top) & \cdots & \mathrm{tr}(\widehat{\boldsymbol{V}^*}^{-1}\boldsymbol{Z}_q\mathbf{1}_T^\top) \\
\mathrm{tr}(\widehat{\boldsymbol{V}^*}^{-1}\mathbf{1}_T\boldsymbol{X}^{*\top}) & \mathrm{tr}(\widehat{\boldsymbol{V}^*}^{-1}\boldsymbol{X}\boldsymbol{X}^{*\top}) & \mathrm{tr}(\widehat{\boldsymbol{V}^*}^{-1}\boldsymbol{Z}_1\boldsymbol{X}^{*\top}) & \cdots & \mathrm{tr}(\widehat{\boldsymbol{V}^*}^{-1}\boldsymbol{Z_q}\boldsymbol{X}^{*\top}) \\
\mathrm{tr}(\widehat{\boldsymbol{V}^*}^{-1}\mathbf{1}_T\boldsymbol{Z}_1^\top) & \mathrm{tr}(\widehat{\boldsymbol{V}^*}^{-1}\boldsymbol{X}^*\boldsymbol{Z}_1^\top) & \mathrm{tr}(\widehat{\boldsymbol{V}^*}^{-1}\boldsymbol{Z}_1\boldsymbol{Z}_1^\top) & & \\
\vdots & & \ddots & \cdots & \vdots \\
\mathrm{tr}(\widehat{\boldsymbol{V}^*}^{-1}\mathbf{1}_T\boldsymbol{Z}_q^\top) & \mathrm{tr}(\widehat{\boldsymbol{V}^*}^{-1}\boldsymbol{X}^*\boldsymbol{Z}_q^\top) & \cdots & & \mathrm{tr}(\widehat{\boldsymbol{V}^*}^{-1}\boldsymbol{Z_q}\boldsymbol{Z_q}^\top)
\end{pmatrix} \right]^{-1}
$$

$$
\sum \begin{pmatrix}
\mathrm{tr}(\widehat{\boldsymbol{V}^*}^{-1}\mathbf{J}_T) & \mathrm{tr}(\widehat{\boldsymbol{V}^*}^{-1}\boldsymbol{X}^*\mathbf{1}_T^\top) & \mathrm{tr}(\widehat{\boldsymbol{V}^*}^{-1}\boldsymbol{Z}_1\mathbf{1}_T^\top) & \cdots & \mathrm{tr}(\widehat{\boldsymbol{V}^*}^{-1}\boldsymbol{Z}_q\mathbf{1}_T^\top) \\
\mathrm{tr}(\widehat{\boldsymbol{V}^*}^{-1}\mathbf{1}_T\boldsymbol{X}^{*\top}) & \mathrm{tr}(\widehat{\boldsymbol{V}^*}^{-1}\boldsymbol{X}^*\boldsymbol{X}^{*\top}) & \mathrm{tr}(\widehat{\boldsymbol{V}^*}^{-1}\boldsymbol{Z}_1\boldsymbol{X}^{*\top}) & \cdots & \mathrm{tr}(\widehat{\boldsymbol{V}^*}^{-1}\boldsymbol{Z}_q\boldsymbol{X}^{*\top}) \\
\mathrm{tr}(\widehat{\boldsymbol{V}^*}^{-1}\mathbf{1}_T\boldsymbol{Z}_1^\top) & \mathrm{tr}(\widehat{\boldsymbol{V}^*}^{-1}\boldsymbol{X}^*\boldsymbol{Z}_1^\top) & \mathrm{tr}(\widehat{\boldsymbol{V}^*}^{-1}\boldsymbol{Z}_1\boldsymbol{Z}_1^\top) & & \\
\vdots & & \ddots & \cdots & \vdots \\
\mathrm{tr}(\widehat{\boldsymbol{V}^*}^{-1}\mathbf{1}_T\boldsymbol{Z}_q^\top) & \mathrm{tr}(\widehat{\boldsymbol{V}^*}^{-1}\boldsymbol{X}^*\boldsymbol{Z}_q^\top) & \cdots & & \mathrm{tr}(\widehat{\boldsymbol{V}^*}^{-1}\boldsymbol{Z}_q\boldsymbol{Z}_q^\top)
\end{pmatrix} \begin{pmatrix} \widehat{\beta_0^*} \\ \widehat{\beta_X^*} \\ \widehat{\boldsymbol{\beta}_Z^*}. \end{pmatrix}
\tag{3.43}
$$

Please remember, index i for the person is neglected; therefore, no summation limits are specified. The only unknown component in eq. 3.43 is $\mathrm{tr}(\widehat{\boldsymbol{V}^*}^{-1}\boldsymbol{X}\boldsymbol{X}^{*\top})$. Since the true values of $\boldsymbol{X}$ are unknown, $\mathrm{tr}(\widehat{\boldsymbol{V}^*}^{-1}\boldsymbol{X}\boldsymbol{X}^{*\top})$ has to be calculated with additional information, e.g. through a validation study (see Subsection 3.4.2).

3.5.4. Berkson error–prone measurements with classical measurement error

Classical measurement error of the fixed–site measurements was neglected in the previous considerations, but does actually occur. This data situation results in a mixture of Berkson and classical measurement error, which is different from the mixture measurement error defined in eq. 2.6 and is defined by Mallick et al. (2002) as follows:

$$
\begin{aligned}
X &= X^{*\mathrm{B}} + U^{\mathrm{B}} \\
X^{*\mathrm{BC}} &= X^{*\mathrm{B}} + U^{\mathrm{BC}}
\end{aligned}
$$

The true values of the Berkson error–prone measurements $X^{*\mathrm{B}}$ are not observable. Only $X^{*\mathrm{BC}}$ can be observed, which represents the latent fixed–site exposure with an additive measurement error.

Mixture error

The mixture error structure with classical error–prone fixed–site and individual measurements is defined as:

$$
\begin{aligned}
X_{it}^{*\mathrm{M}_2} &= X_{it}^{*\mathrm{BC}} G_{it} + X_{it}^{*\mathrm{C}}(1 - G_{it}) \\
&= \begin{cases} X_{it}^{*\mathrm{BC}} & \text{for } p \cdot 100\% \text{ of the measurements} \\ X_{it}^{*\mathrm{C}} & \text{for } (1-p) \cdot 100\% \text{ of the measurements} \end{cases} .
\end{aligned}
$$

The classical measurement error overlaying $\boldsymbol{X}^{*\mathrm{B}}$, $\boldsymbol{U}^{\mathrm{BC}}$, is presumed to be independent from the other errors, i.e. $\boldsymbol{\tau}, \boldsymbol{\nu}^{\mathrm{B}}, \boldsymbol{\nu}^{\mathrm{C}}, \boldsymbol{\varepsilon}, \boldsymbol{U}^{\mathrm{B}}$ and $\boldsymbol{U}^{\mathrm{C}}$ and to be normally distributed with an autoregressive error structure of order 1: $\boldsymbol{U}^{\mathrm{BC}} \sim \mathrm{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_{U^{\mathrm{BC}}}), \boldsymbol{\Sigma}_{U^{\mathrm{BC}}} = \sigma^2_{U^{\mathrm{BC}}} \boldsymbol{W}_{\rho^{\mathrm{BC}}}$. In order to derive the attenuation factor λ^{M_2}, $\mathrm{Cov}(\boldsymbol{X}, \boldsymbol{X}^{*\mathrm{M}_2})$ and $\mathrm{Var}(\boldsymbol{X}^{*\mathrm{M}_2})$ are calculated:

$$\begin{aligned}
\mathrm{Cov}(\boldsymbol{X}, \boldsymbol{X}^{*\mathrm{M}_2}) =& \mathrm{Cov}(\boldsymbol{G} \circ \boldsymbol{X}^{*\mathrm{BC}} + (\boldsymbol{1}_T - \boldsymbol{G}) \circ \boldsymbol{X}^{*\mathrm{B}} + (1 - \boldsymbol{G}) \circ \nu^{\mathrm{B}} \boldsymbol{1}_T + (1 - \boldsymbol{G}) \circ \boldsymbol{U}^{\mathrm{B}}, \\
& \boldsymbol{X}^{*\mathrm{B}} + \boldsymbol{\nu}^{\mathrm{B}} + \boldsymbol{U}^{\mathrm{B}}) \\
=& \boldsymbol{\Sigma}_{X^{*\mathrm{B}}} + (1-p)\sigma^2_{\nu^{\mathrm{B}}} \mathbf{J}_T + (1-p)\boldsymbol{\Sigma}_{U^{\mathrm{B}}} \\
\mathrm{Var}(\boldsymbol{X}^{*\mathrm{M}_2}) =& \mathrm{Var}(\boldsymbol{G} \circ \boldsymbol{X}^{*\mathrm{B}} + \boldsymbol{G} \circ \boldsymbol{U}^{\mathrm{BC}} + (\boldsymbol{1}_T - \boldsymbol{G}) \circ \boldsymbol{X}^{*\mathrm{B}} + (\boldsymbol{1}_T - \boldsymbol{G}) \circ \nu^{\mathrm{B}} \boldsymbol{1}_T + \\
& (\boldsymbol{1}_T - \boldsymbol{G}) \circ \boldsymbol{U}^{\mathrm{B}} + (\boldsymbol{1}_T - \boldsymbol{G}) \circ \nu^{\mathrm{C}} \boldsymbol{1}_T + (\boldsymbol{1}_T - \boldsymbol{G}) \circ \boldsymbol{U}^{\mathrm{C}}) \\
=& \boldsymbol{\Sigma}_{X^{*\mathrm{B}}} + (\sigma^2_{\nu^{\mathrm{B}}} + \sigma^2_{\nu^{\mathrm{C}}}) \left((1-p)^2 \mathbf{J} + p(1-p)\mathbf{I}_T \right) + \\
& p^2 \boldsymbol{\Sigma}_{U^{\mathrm{BC}}} + (1-p)^2 \left(\boldsymbol{\Sigma}_{U^{\mathrm{B}}} + \boldsymbol{\Sigma}_{U^{\mathrm{C}}} \right) + p(1-p)(\sigma^2_{U^{\mathrm{B}}} + \sigma^2_{U^{\mathrm{C}}} + \sigma^2_{U^{\mathrm{BC}}})\mathbf{I}_T.
\end{aligned}$$

$\mathrm{Cov}(\boldsymbol{X}, \boldsymbol{X}^{*\mathrm{M}_2})$ does not change in comparison to the case when $\sigma^2_{U^{\mathrm{BC}}} = 0$, whereas $\mathrm{Var}(\boldsymbol{X}^{*\mathrm{M}_2})$ increases by $p^2 \boldsymbol{\Sigma}_{U^{\mathrm{BC}}} + p(1-p)\sigma^2_{U^{\mathrm{BC}}} \mathbf{I}_T$. The attenuation factor is deduced similar to Section 3.6 and results in:

$$\begin{aligned}
\lambda^{\mathrm{M}_2} =& [\sigma^2_{X^{*\mathrm{B}}} g_T^{*\mathrm{M}_2}(\rho^{X^{*\mathrm{B}}}) + (1-p)\sigma^2_{\nu^{\mathrm{B}}} g_T^{*\mathrm{M}_2}(1) + (1-p)\sigma^2_{U^{\mathrm{B}}} g_T^{*\mathrm{M}_2}(\rho^{\mathrm{B}})] \\
& [\sigma^2_{X^{*\mathrm{B}}} g_T^{*\mathrm{M}_2}(\rho^{X^{*\mathrm{B}}}) + (\sigma^2_{\nu^{\mathrm{B}}} + \sigma^2_{\nu^{\mathrm{C}}})[(1-p)^2 g_T^{*\mathrm{M}_2}(1) + p(1-p) g_T^{*\mathrm{M}_2}(0)] + \\
& \sigma^2_{U^{\mathrm{B}}}[(1-p)^2 g_T^{*\mathrm{M}_2}(\rho^{\mathrm{B}}) + p(1-p) g_T^{*\mathrm{M}_2}(0)] + \\
& \sigma^2_{U^{\mathrm{C}}}[(1-p)^2 g_T^{*\mathrm{M}_2}(\rho^{\mathrm{C}}) + p(1-p) g_T^{*\mathrm{M}_2}(0)] + \\
& \sigma^2_{U^{\mathrm{BC}}}[p^2 g_T^{*\mathrm{M}_2}(\rho^{\mathrm{BC}}) + p(1-p) g_T^{*\mathrm{M}_2}(0)]]^{-1}. \qquad (3.44)
\end{aligned}$$

λ^{M_2} exhibits the additional term $\sigma^2_{U^{\mathrm{BC}}}[p^2 g_T^{*\mathrm{M}_2}(\rho^{\mathrm{BC}}) + p(1-p) g_T^{*\mathrm{M}_2}(0)]$ in the denominator in comparison to λ^{M}. This additional term may alleviate or strengthen the attenuation depending on the weighting functions $g_T^{*\mathrm{M}_2}(\rho^{\mathrm{BC}})$ and $g_T^{*\mathrm{M}_2}(0)$. If $\rho^{X^{*\mathrm{B}}} = \rho^{\mathrm{B}} = \rho^{\mathrm{BC}} = \rho^{\mathrm{C}} = 0$, $g_T^{*\mathrm{M}_2}(0) = 1 + (T-1)\sigma^2_{\tau^{*\mathrm{M}}} 2/\sigma^2_{\varepsilon^{*\mathrm{M}_2}}$ yields a stronger attenuation of the effect coefficient.

Mixture error as defined in Mallick et al. (2002)

For $p = 1$, the mixture error equals the mixture error defined by Mallick et al. (2002). The attenuation factor in eq. 3.44 changes to:

$$\lambda^{\mathrm{M}_2} = \frac{\sigma^2_{X^{*\mathrm{B}}} g_T^{*\mathrm{M}_2}(\rho^{X^{*\mathrm{B}}})}{\sigma^2_{X^{*\mathrm{B}}} g_T^{*\mathrm{M}_2}(\rho^{X^{*\mathrm{B}}}) + \sigma^2_{U^{\mathrm{BC}}} g_T^{*\mathrm{M}_2}(\rho^{\mathrm{BC}})}.$$

The Berkson error does not affect the attenuation of the regression coefficient. Only the classical measurement error of the fixed–site measurements, $\boldsymbol{U}^{\mathrm{BC}}$, attenuates the effect coefficient.

3.5.5. Unbalanced and non–equidistant observations

So far, the developed bias correction assumes a balanced sample, i.e. the observations for each individual are equidistant and have the same number T. An extension of the correction approach to a sample with missing values or with non–equidistant observation times is considered in this subsection. In these situations $\boldsymbol{X}_i$ and $\boldsymbol{Y}_i$, $i = 1, \ldots, n$ are not identically distributed and the attenuation factor is given by:

$$\lambda = \frac{\sum_{i=1}^n \mathrm{tr}\left[\boldsymbol{V}_i^{*-1} \mathrm{Cov}(\boldsymbol{X}_i, \boldsymbol{X}_i^*) \right]}{\sum_{i=1}^n \mathrm{tr}\left[\boldsymbol{V}_i^{*-1} \mathrm{Var}(\boldsymbol{X}_i^*) \right]}.$$

Individual lengths of measurement intervals

The derivation of the attenuation factor in the case of individually varying measurement durations but without missing values during a measurement series differs in comparison to the balanced case (Sections 3.5 and 3.6) only in the calculation of the weights $g^*_{T_i}(\rho)$, which are now individual–specific:

$$\begin{aligned}
\lambda^{\mathrm{C}} = &\left\{\sum_{i=1}^{n}\left[\sigma^2_{X^{*\mathrm{B}}} g^{*\mathrm{C}}_{T_i}(\rho^{X^{*\mathrm{B}}}) + \sigma^2_{\nu^{\mathrm{B}}} g^{*\mathrm{C}}_{T_i}(1) + \sigma^2_{U^{\mathrm{B}}} g^{*\mathrm{C}}_{T_i}(\rho^{\mathrm{B}})\right]\right\} \\
&\left\{\sum_{i=1}^{n}\left[\sigma^2_{X^{*\mathrm{B}}} g^{*\mathrm{C}}_{T_i}(\rho^{X^{*\mathrm{B}}}) + \sigma^2_{\nu^{\mathrm{B}}} g^{*\mathrm{C}}_{T_i}(1) + \sigma^2_{U^{\mathrm{B}}} g^{*\mathrm{C}}_{T_i}(\rho^{\mathrm{B}}) + \right.\right. \\
&\qquad \left.\left.\sigma^2_{\nu^{\mathrm{C}}} g^{*\mathrm{C}}_{T_i}(1) + \sigma^2_{U^{\mathrm{C}}} g^{*\mathrm{C}}_{T_i}(\rho^{\mathrm{C}})\right]\right\}^{-1}, \\
\lambda^{\mathrm{M}} = &\left\{\sum_{i=1}^{n}\left[\sigma^2_{X^{*\mathrm{B}}} g^{*\mathrm{M}}_{T_i}(\rho^{X^{*\mathrm{B}}}) + (1-p)\sigma^2_{\nu^{\mathrm{B}}} g^{*\mathrm{M}}_{T_i}(1) + (1-p)\sigma^2_{U^{\mathrm{B}}} g^{*\mathrm{M}}_{T_i}(\rho^{\mathrm{B}})\right]\right\} \\
&\left\{\sum_{i=1}^{n}\left[\sigma^2_{X^{*\mathrm{B}}} g^{*\mathrm{M}}_{T_i}(\rho^{X^{*\mathrm{B}}}) + (\sigma^2_{\nu^{\mathrm{B}}} + \sigma^2_{\nu^{\mathrm{C}}})[(1-p)^2 g^{*\mathrm{M}}_{T_i}(1) + p(1-p) g^{*\mathrm{M}}_{T_i}(0)] + \right.\right. \\
&\qquad \sigma^2_{U^{\mathrm{B}}}[(1-p)^2 g^{*\mathrm{M}}_{T_i}(\rho^{\mathrm{B}}) + p(1-p) g^{*\mathrm{M}}_{T_i}(0)] + \\
&\qquad \left.\left.\sigma^2_{U^{\mathrm{C}}}[(1-p)^2 g^{*\mathrm{M}}_{T_i}(\rho^{\mathrm{C}}) + p(1-p) g^{*\mathrm{M}}_{T_i}(0)]\right]\right\}^{-1}
\end{aligned}$$

Missing values

If missing values occur in $\boldsymbol{X}^{*\mathrm{M}}$, which cannot be imputed through fixed–site measurements, or in $\boldsymbol{X}^{*\mathrm{C}}$, the calculation of the attenuation factor changes. According to Wansbeek and Kapteyn (1985), the main outcome model can be rewritten as

$$\boldsymbol{D}_i\boldsymbol{Y}_i = \beta_0 + \beta_X \boldsymbol{D}_i\boldsymbol{X}_i + \tau_i \boldsymbol{D}_i\mathbf{1}_T + \boldsymbol{D}_i\boldsymbol{\varepsilon}_i$$

with the $m_i \times T$ deletion matrix $\boldsymbol{D}_i$; m_i denotes the number of missing values during the measurements of individual i, which corresponds to $\sum_{t=1}^{T} G_{it}$ in the case of classical measurement error. $\boldsymbol{D}_i$ is defined as $\mathbf{I}_T$ where the rows of observations with missing values are deleted. Since $\boldsymbol{\varepsilon}_i \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}_\varepsilon)$, $\boldsymbol{D}_i\boldsymbol{\varepsilon}_i \sim \mathrm{N}(0, \sigma^2_\varepsilon \boldsymbol{D}_i\boldsymbol{\Sigma}_\varepsilon\boldsymbol{D}_i^\top)$. $\boldsymbol{D}_i\boldsymbol{W}_\rho\boldsymbol{D}_i^\top$ represents the correlation matrix of a continuous AR(1) process:

$$\boldsymbol{D}_i\boldsymbol{W}_\rho\boldsymbol{D}_i^\top = \begin{pmatrix}
1 & \rho^{t_{2i}-t_{1i}} & \rho^{t_{3i}-t_{1i}} & \dots & \rho^{t_{Ti}-t_{1i}} \\
\rho^{t_{2i}-t_{1i}} & 1 & \rho^{t_{3i}-t_{2i}} & \dots & \rho^{t_{Ti}-t_{2i}} \\
\rho^{t_{3i}-t_{1i}} & \rho^{t_{3i}-t_{2i}} & 1 & \dots & \rho^{t_{Ti}-t_{3i}} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\rho^{t_{Ti}-t_{1i}} & \rho^{t_{Ti}-t_{2i}} & \rho^{t_{Ti}-t_{3i}} & \dots & 1
\end{pmatrix},$$

with $t_{1i}, t_{2i}, \dots, t_{Ti}$ denoting the observation times of the non–missing measurements of individual i. The attenuation factor changes to

$$\begin{aligned}
\lambda &\overset{\text{eq. 3.9}}{=} \frac{\sum_{i=1}^{n} \mathrm{tr}\left[(\boldsymbol{D}_i\boldsymbol{V}_i^*\boldsymbol{D}_i^\top)^{-1}\mathrm{Cov}(\boldsymbol{D}_i\boldsymbol{X}_i, \boldsymbol{D}_i\boldsymbol{X}_i^*)\right]}{\sum_{i=1}^{n} \mathrm{tr}\left[(\boldsymbol{D}_i\boldsymbol{V}_i^*\boldsymbol{D}_i^\top)^{-1}\mathrm{Var}(\boldsymbol{D}_i\boldsymbol{X}_i^*)\right]} \\
&= \frac{\sum_{i=1}^{n} \mathrm{tr}\left[(\boldsymbol{D}_i\boldsymbol{V}_i^*\boldsymbol{D}_i^\top)^{-1}\boldsymbol{D}_i\mathrm{Cov}(\boldsymbol{X}_i, \boldsymbol{X}_i^*)\boldsymbol{D}_i^\top\right]}{\sum_{i=1}^{n} \mathrm{tr}\left[(\boldsymbol{D}_i\boldsymbol{V}_i^*\boldsymbol{D}_i^\top)^{-1}\boldsymbol{D}_i\mathrm{Var}(\boldsymbol{X}_i^*)\boldsymbol{D}_i^\top\right]}.
\end{aligned} \tag{3.45}$$

Index i is neglected in the following for better readability. Wansbeek and Kapteyn (1985) showed that

$$(\boldsymbol{D}\boldsymbol{W}_{\rho^*}\boldsymbol{D}^\top)^{-1} =$$

$$\begin{pmatrix} \frac{1}{1-\rho^{*2(t_2-t_1)}} & -\frac{\rho^{*t_2-t_1}}{1-\rho^{*2(t_2-t_1)}} & 0 & 0 & 0 \\ -\frac{\rho^{*t_2-t_1}}{1-\rho^{*2(t_2-t_1)}} & \frac{1}{1-\rho^{*2(t_2-t_1)}} + \frac{\rho^{*2(t_3-t_2)}}{1-\rho^{*2(t_3-t_2)}} & -\frac{\rho^{*t_3-t_2}}{1-\rho^{*2(t_3-t_2)}} & 0 & 0 \\ 0 & -\frac{\rho^{*t_3-t_2}}{1-\rho^{*2(t_3-t_2)}} & \frac{1}{1-\rho^{*2(t_3-t_2)}} + \frac{\rho^{*2(t_4-t_3)}}{1-\rho^{*2(t_4-t_3)}} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & -\frac{\rho^{*t_T-t_{T-1}}}{1-\rho^{*2(t_T-t_{T-1})}} & \frac{1}{1-\rho^{*2(t_T-t_{T-1})}} \end{pmatrix}. \tag{3.46}$$

The Sherman–Morrison formula (eq. 3.24) is used to invert $\boldsymbol{D}\boldsymbol{V}^*\boldsymbol{D}^\top$:

$$\begin{aligned} (\boldsymbol{D}\boldsymbol{V}^*\boldsymbol{D}^\top)^{-1} =& (\sigma^2_{\tau^*}\boldsymbol{D}\mathbf{J}_T\boldsymbol{D}^\top + \boldsymbol{D}\boldsymbol{\Sigma}_{\varepsilon^*}\boldsymbol{D}^\top)^{-1} \\ =& (\boldsymbol{D}\boldsymbol{\Sigma}_{\varepsilon^*}\boldsymbol{D}^\top)^{-1} - \\ & \sigma^2_{\tau^*}\left[\underbrace{1+\sigma^2_{\tau^*}\mathbf{1}_T^\top\boldsymbol{D}^\top(\boldsymbol{D}\boldsymbol{\Sigma}_{\varepsilon^*}\boldsymbol{D}^\top)^{-1}\boldsymbol{D}\mathbf{1}_T}_{:=d}\right]^{-1} \\ & (\boldsymbol{D}\boldsymbol{\Sigma}_{\varepsilon^*}\boldsymbol{D}^\top)^{-1}(\boldsymbol{D}\mathbf{J}_T\boldsymbol{D}^\top)(\boldsymbol{D}\boldsymbol{\Sigma}_{\varepsilon^*}\boldsymbol{D}^\top)^{-1}. \end{aligned}$$

d is simplified to

$$d \overset{\text{eq. } 3.46}{=} 1 + \frac{\sigma^2_{\tau^*}}{\sigma^2_{\varepsilon^*}}\left(2\frac{1}{1+\rho^{*t_2-t_1}} + \sum_{j=2}^{T-1}\frac{1-\rho^{*t_{j+1}-t_j}}{1+\rho^{*t_{j+1}-t_j}}\right).$$

The weights are then

$$\begin{aligned} g^*_{1,T}(\rho) :=& (1-\rho^{*2})\text{tr}\left[(\boldsymbol{D}\boldsymbol{W}_{\rho^*}\boldsymbol{D}^\top)^{-1}\boldsymbol{D}\boldsymbol{W}_{\rho^*}\boldsymbol{D}^\top\right] \\ g^*_{2,T}(\rho) :=& (1-\rho^{*2})^2\text{tr}\left[(\boldsymbol{D}\boldsymbol{W}_{\rho^*}\boldsymbol{D}^\top)^{-1}(\boldsymbol{D}_T\mathbf{J}\boldsymbol{D}^\top)(\boldsymbol{D}\boldsymbol{W}_{\rho^*}\boldsymbol{D}^\top)^{-1}\boldsymbol{D}\boldsymbol{W}_{\rho}\boldsymbol{D}^\top\right] \\ g^*_T(\rho) :=& g^*_{1,T}(\rho) - \frac{\sigma^2_{\tau^*}}{\sigma^2_{\varepsilon^*}(1-\rho^{*2})d}g^*_{2,T}(\rho). \end{aligned}$$

3.6. Interpretation and discussion of the theoretical results via simulations

A simulation study is performed to visualize the theoretical results and to describe the properties of the estimators. The impact of measurement error on the effect estimates with and without confounding variables is examined as well as the properties of the estimated attenuation factor regarding T and missing values and the influence of autocorrelation of the measurement error and the model error.

3.6.1. Simulation setup

The data was generated according to the main model (eq. 3.1 without confounding variables $\boldsymbol{Z}$) and the measurement error models (eq. 2.4–2.6) defined in Section 2.1. The parameters are specified as described in Table 3.1, inspired by the Augsburger Umweltstudie:

	Mean	Variance	Autocorrelation
$X_{it}^{*\mathrm{B}}$	9.5	0.3	0.99
ν_i^{B}	0	0.04	
U_{it}^{B}	0	0.2	0.7
ν_i^{C}	0	0.03	
U_{it}^{C}	0	0.06	0.2
τ_i	0	140	
ε_{it}	0	70	0.7

TABLE 3.1.: *Parameter choices for simulation.*

Artificial measurement series of length $T = 75$ are generated for $n = 100$ individuals by conducting the following steps:

1. Draw fixed–site measurements $X_{it}^{*\mathrm{B}}$ from $N(9.5, 0.3 \cdot \boldsymbol{W}_{0.99})$.
2. Draw the true exposure measurements X_{it} according to eq. 2.4 with $\sigma^2_{\nu^{\mathrm{B}}} = 0.03, \sigma^2_{U^{\mathrm{B}}} = 0.06$ and $\rho^{\mathrm{B}} = 0.7$.
3. Draw the observed exposure measurements $X_{it}^{*\mathrm{C}}$ according to eq. 2.5 with $\sigma^2_{\nu^{\mathrm{C}}} = 0.04, \sigma^2_{U^{\mathrm{C}}} = 0.2$ and $\rho^{\mathrm{C}} = 0.2$.
4. Draw the health outcome Y_{it} according to eq. 3.1 (without confounding variables $\boldsymbol{Z}$) with β_0 =73 , β_X=1 , $\sigma^2_\tau = 140, \sigma^2_\varepsilon = 70$ and $\rho = 0.7$.

One out of the $N = 100$ artificial data sets is visualized in Figure 3.3.

Three scenarios are considered for the structure of the measurement errors: (I) uncorrelated errors without an individual–specific structure ($\boldsymbol{\nu}^{\mathrm{B}} = \mathbf{0}, \boldsymbol{\nu}^{\mathrm{C}} = \mathbf{0}, \rho^{\mathrm{B}} = 0, \rho^{\mathrm{C}} = 0, \rho = 0$), (II) uncorrelated random errors and individual–specific measurement error components ($\rho^{\mathrm{B}} = 0, \rho^{\mathrm{C}} = 0, \rho = 0$), (III) autocorrelated errors and individual–specific measurement error components.

These scenarios are evaluated for data with the three examined types of measurement error (Berkson, classical, mixture) as well as for the proper data without measurement error using linear models with a random intercept and an error term, which is assumed to be autocorrelated of order 1.

The size of measurement error is varied by scaling the variances of the random errors ($\sigma^2_{U^{\mathrm{B}}}, \sigma^2_{U^{\mathrm{C}}}$) with the scaling factors (SF) 0, 0.5, 1, 2 and 5; the other parameters remain equal within each scenario. The percentage of missing values in $X^{*\mathrm{C}}$ and the percentage of Berkson error–prone measurements in $X^{*\mathrm{M}}$ is, accordingly to the percentage of missing values in the Augsburger Umweltstudie, chosen to be 23 %.

3.6.2. Main results

The simulation results for the empirical attenuation factors depending on varying measurement error sizes are presented in Figure 3.4. The empirical attenuation factors, which

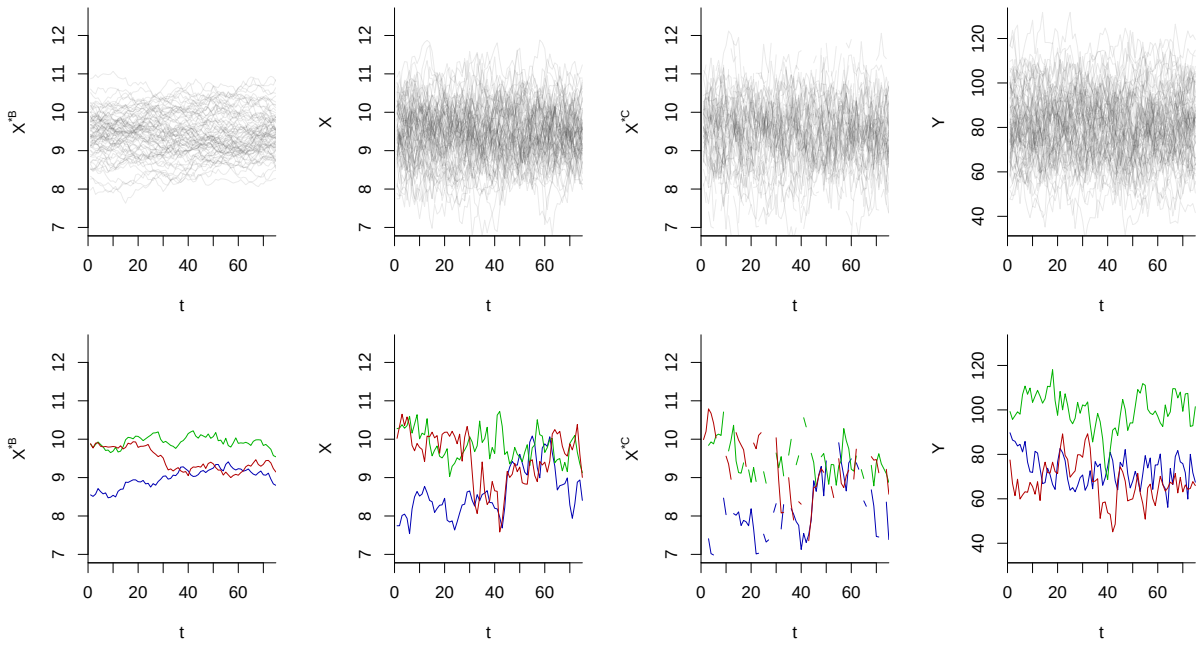


FIGURE 3.3.: *Individual time series for X^{*B}, X, X^{*C} and Y of an exemplary artificial data set of Scenario (III). First row: all individuals; second row: three randomly chosen individuals.*

are depicted with boxplots, are calculated as the ratio between the estimated effect using the error–prone data $\boldsymbol{X}^*$, $\widehat{\beta_X^*}$, and the estimated effects using the proper data $\boldsymbol{X}$, $\widehat{\beta_X}$.

The theoretical estimations of λ according to eq. 3.45 and eq. 3.41 are marked with circles and triangles. Note, that the theoretical estimations of the attenuation factor allow for the presence of missing values in the measurements with classical measurement error. The estimation of the attenuation factors for classical and mixture measurement error with the known values of $\boldsymbol{G}$ and $\boldsymbol{D}$ vary depending on the missing structure of the data set. Therefore, averaged estimations of the estimated attenuation factors according to eq. 3.45 and eq. 3.41 are depicted in Figure 3.4, since the variation of the estimations is only marginal.

The true parameter settings for the parameters describing the measurement error defined in the simulation setup are used for the estimation of $\widehat{\lambda}$; the parameters of the naive models, $\sigma^2_{\varepsilon^*}, \sigma^2_{\tau^*}$ and ρ^*, which are also required for the calculation of $\widehat{\lambda}$, are estimated from the corresponding naive models.

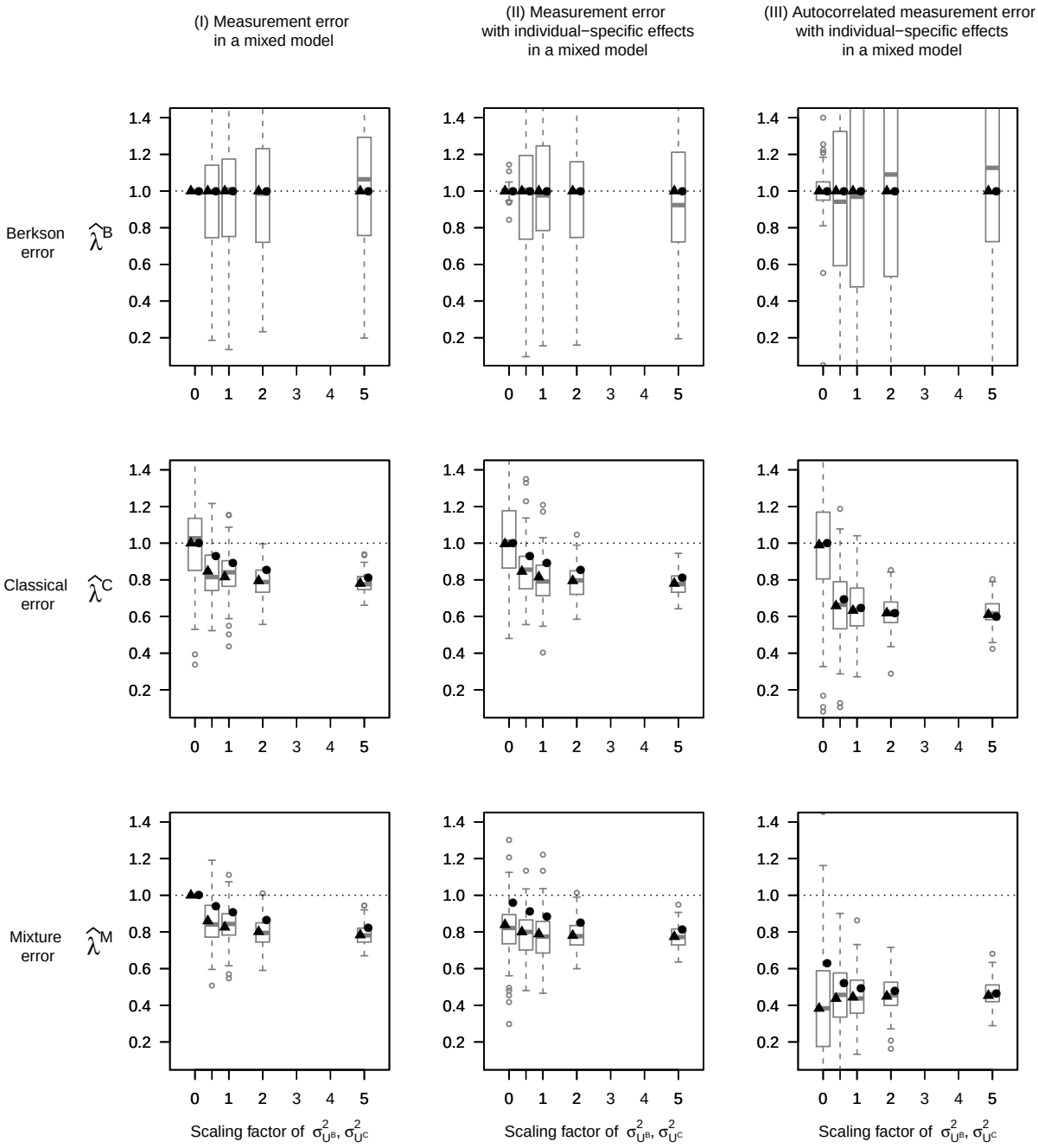


FIGURE 3.4.: *Simulation results for the empirical attenuation factor depicted in boxplots for increasing size of measurement errors; the theoretical estimations of λ according to eq. 3.45 and eq. 3.41 are marked with triangles for the exact values of T and with circles for calculations according to eq. 3.28 and eq. 3.35 assuming $T \to \infty$.*

On average, measurements with Berkson error provide unbiased effect estimates, also in the presence of individual specific or autocorrelated measurement error. The accuracy of the estimator diminishes with increasing measurement error variance. Individual–specific Berkson errors have not any visible impact on the accuracy of the estimator, whereas AR(1) Berkson errors (Scenario (III)) intensify the known accuracy reduction of the effect estimates.

The simulation results show, that heterogeneous classical measurement errors without autocorrelation have also an attenuating influence on the effect estimate as it is known for homogeneous errors. The theoretical examinations in Subsection 3.4.2 revealed, that individual–specific classical measurement error does not affect the attenuation for $T \to \infty$, which is also visible in the simulation results. In contrast, autocorrelated errors change the degree of attenuation; in our simulation, the degree of attenuation in Scenario (III) is stronger compared to Scenario (I) and Scenario (II), in particular regarding small measurement error variances.

If $\sigma^2_{\varepsilon^*}$ is substantially larger than the measurement error variances $\sigma^2_{U^{\mathrm{B}}}$ and $\sigma^2_{U^{\mathrm{C}}}$, as in our case, $\rho^{*\mathrm{C}}$ only marginally depends on ρ^{B} and ρ^{C}. In general, attenuation may increase or decrease with changing autocorrelation in the measurement error (see Subsection 3.6.4); however attenuation is intensified with increasing autocorrelation in the Berkson error and with decreasing autocorrelation in the classical measurement error for $T \to \infty$ (Figure 3.5).

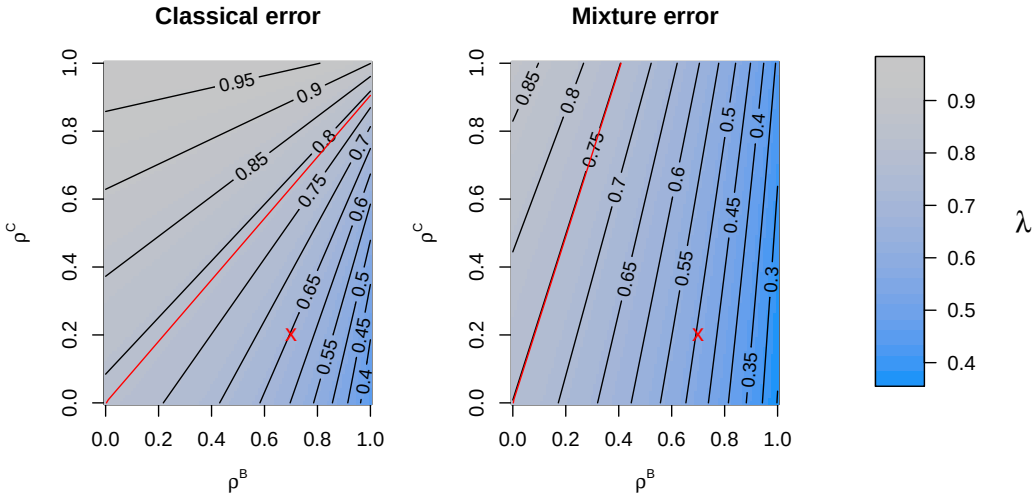


FIGURE 3.5.: *Attenuation factor for varying correlation of the measurement errors and* $T \to \infty$. *Classical measurement error (left) and mixture measurement error (right); red crosses: parameter settings for the simulations; red lines:* λ *with* $\rho^B = \rho^C = 0$; $\rho^* = 0.7$.

Mixture error attenuates the effect estimates as it is known for classical error–prone data, but to a different extent. Individual–specific components in the classical and Berkson error parts of the mixture error induce bias in the effect estimates even in situations without random measurement errors $(U^{\mathrm{B}}, U^{\mathrm{C}})$. The bias resulting from mixtures of autocorrelated measurement errors is high and strongly depends on both, the size of the measurement errors and the size of the autocorrelation coefficients. Increasing mixture measurement error, i.e. increasing Berkson and increasing classical measurement error, reduce the attenuation of the effect in Scenario (III); this is possible, because the numerator of the attenuation factor is not completely contained in its denominator. As with classical measurement error, correlation in the classical measurement error lowers

	SF of	Truth		Berkson		Classical		Mixture	
Scen.	$\sigma^2_{U^{\mathrm{B}}}, \sigma^2_{U^{\mathrm{C}}}$	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
(I)	0	0.015	0.360	**0.015**	**0.360**	0.020	0.421	**0.015**	**0.360**
	0.5	0.024	0.242	**-0.014**	0.407	0.016	0.295	0.019	**0.275**
	1	0.002	0.187	**-0.008**	0.378	0.022	0.246	0.014	**0.234**
	2	0.003	0.137	0.003	0.370	**-0.002**	0.182	-0.003	**0.177**
	5	-0.012	0.095	0.013	0.390	**-0.009**	**0.118**	-0.011	0.119
(II)	0	0.062	0.387	0.062	**0.389**	0.074	0.409	**0.033**	0.413
	0.5	-0.039	0.237	-0.067	0.364	**-0.035**	0.282	-0.050	**0.280**
	1	0.015	0.179	0.015	0.397	-0.010	0.243	**0.000**	**0.238**
	2	0.031	0.151	**0.000**	0.370	0.027	0.185	0.029	**0.180**
	5	-0.006	0.092	-0.034	0.387	-0.011	0.117	**-0.008**	**0.117**
(III)	0	-0.015	0.658	-0.018	**0.670**	-0.048	0.759	**0.005**	0.953
	0.5	0.036	0.272	**-0.007**	0.606	0.053	**0.402**	0.088	0.533
	1	-0.012	0.198	-0.002	0.672	0.017	**0.299**	**-0.001**	0.348
	2	-0.009	0.165	0.013	0.677	**-0.002**	**0.204**	0.013	0.255
	5	0.009	0.097	0.095	0.644	**0.031**	**0.150**	0.038	0.188

TABLE 3.2.: *Bias and RMSE of MOM–corrected effect estimations for Scenarios (I)-(III); bold: lowest absolute bias and lowest RMSE of each row (disregarding the results based on the true data).*

attenuation and correlation in the Berkson error intensifies attenuation for $T \to \infty$ (see Figure 3.5).

Comparing the correction methods in terms of bias and RMSE (Table 3.2) shows that the corrected estimations for classical or mixture error–prone data exceeded the accuracy of the estimations for Berkson error–prone data. In Scenarios (I) and (II) the correction based on mixture data is superior or comparable to the correction based on incomplete data with classical error, whereas in Scenario (III) the correction of classical error–prone data is more precise.

The RMSE values in Scenario (III) of corrected effect estimations for data with mixture error are not generally higher than for incomplete data with classical measurement error. For example, for lower $\rho^{X^{*\mathrm{B}}}$, the RMSE is better or comparably as good as for the estimated coefficient based on data with mixture error (see Appendix C.1). Hence, using incomplete data with classical error out–performed the usage of data with mixture error in Scenario (III) only under certain parameter specifications.

3.6.3. Convergence

As shown in Subsections 3.3.3 and 3.4.3, each component of the attenuation factor consists of a variance and a weight. The attenuation factor λ depends on the number of observations per individual through these weights $g^*_T(\cdot)$. The impact of T on λ^{C} and λ^{M} is depicted in Figure 3.6.

The dependence on T occurs in all scenarios: apart from small values of T, attenuation decreases with increasing T approximating a certain level for $T \to \infty$. As already seen in previous simulations, the attenuation of the effect coefficient is most pronounced in Scenario (III) for the used parameter specifications.

Since $\boldsymbol{X}^{*\mathrm{B}}$ is assumed to be autocorrelated for all three scenarios, the attenuation factor varies with T, even in Scenario (I). In Scenario (II) for the classical measurement error,

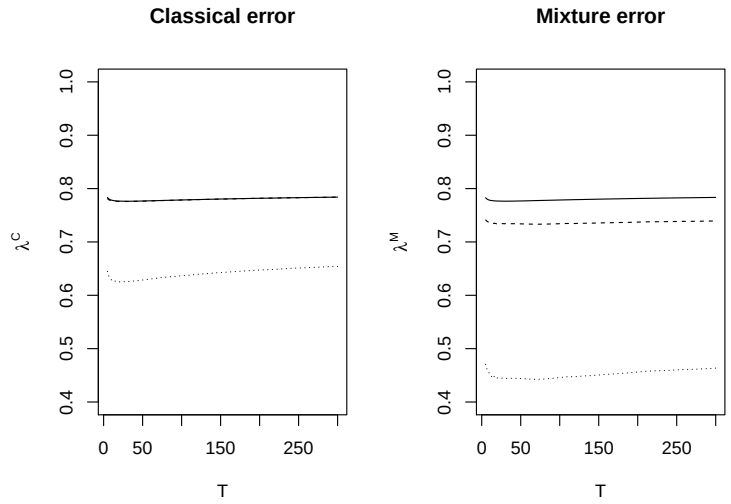


FIGURE 3.6.: *Probability limits for $n \to \infty$ of the empirical attenuation factor $\widehat{\lambda}$ of the three scenarios depending on the number of observations per individual T: solid: Scenario (I); dashed: Scenario (II); dotted: Scenario (III); $SF = 1$, $\rho^* = 0.7, \sigma^2_{\tau^*} = 140, \sigma^2_{\varepsilon^*} = 70$, $p = 0.23, \boldsymbol{D}/\boldsymbol{G}$ known.*

the run of the curve converges for increasing T more and more towards the curve from scenario (I). This effect is hardly visible in Figure 3.6, because $\sigma^2_{\nu^C}$ is chosen to be rather small. The curves of Scenario (I) and (II) for mixture measurement error are nearly parallelly shifted due to the additional component in the denominator of λ in Scenario (II).

Figures 3.4 and 3.6 indicate, that the finite sample correction is indeed relevant, because the weights $g^*_T(\cdot)$, and thus also the attenuation factor, strongly depend on the number of observations per individual T, especially in Scenario (III).

3.6.4. Weight $g^*_T(\rho)$

Autocorrelated errors change the attenuation factor by weighting the single components. Weights $g^*_T(\cdot)$ depending on T, as described in the previous sections are used for the estimation of the attenuation factors to account for the finite number of observations for each individual and unbalanced or non–equidistant data. Beside the number of observations per individual T, the weight $g^*_T(\rho)$ depends on the autocorrelation coefficient of the component itself, ρ, and of the naive model error, ρ^*. The weights are positive for the majority of possible parameter specifications. For $T \to \infty$, $g^*_T(\rho)$ equals one, i.e. autocorrelated structures do not affect the size of attenuation, if $\rho^* = 2\rho$; the weight is low for high values of ρ and high for low ρ and high ρ^*. Exemplary, $g^*_\infty(\rho)$ is visualized in Figure 3.7 as well as $g^*_T(\rho)$ for varying ρ, ρ^* and T.

The weights depend for small ρ^* only on T (except for very high values of ρ). For moderate and high values of ρ^*, the weights increase with decreasing values of ρ and increasing values of T. The product of variance and weight $g^*_T(\cdot)$ can be interpreted as a modified version of the sample variance, as it is shown in the following.

Let $\boldsymbol{Z}$, $\boldsymbol{Z} = (Z_1, Z_2, \ldots, Z_T)^\top$, denote a T–dimensional, first–order autocorrelated random variable with variance σ^2_Z and autocorrelation coefficient ρ_Z. Zięba (2010) shows, that the following holds for the biased sample variance s^2_Z of the realizations of $\boldsymbol{Z}$:

$$\mathbb{E}(s^2_Z) = \left(1 - \frac{1}{T_{\text{eff}}}\right) \sigma^2_Z$$

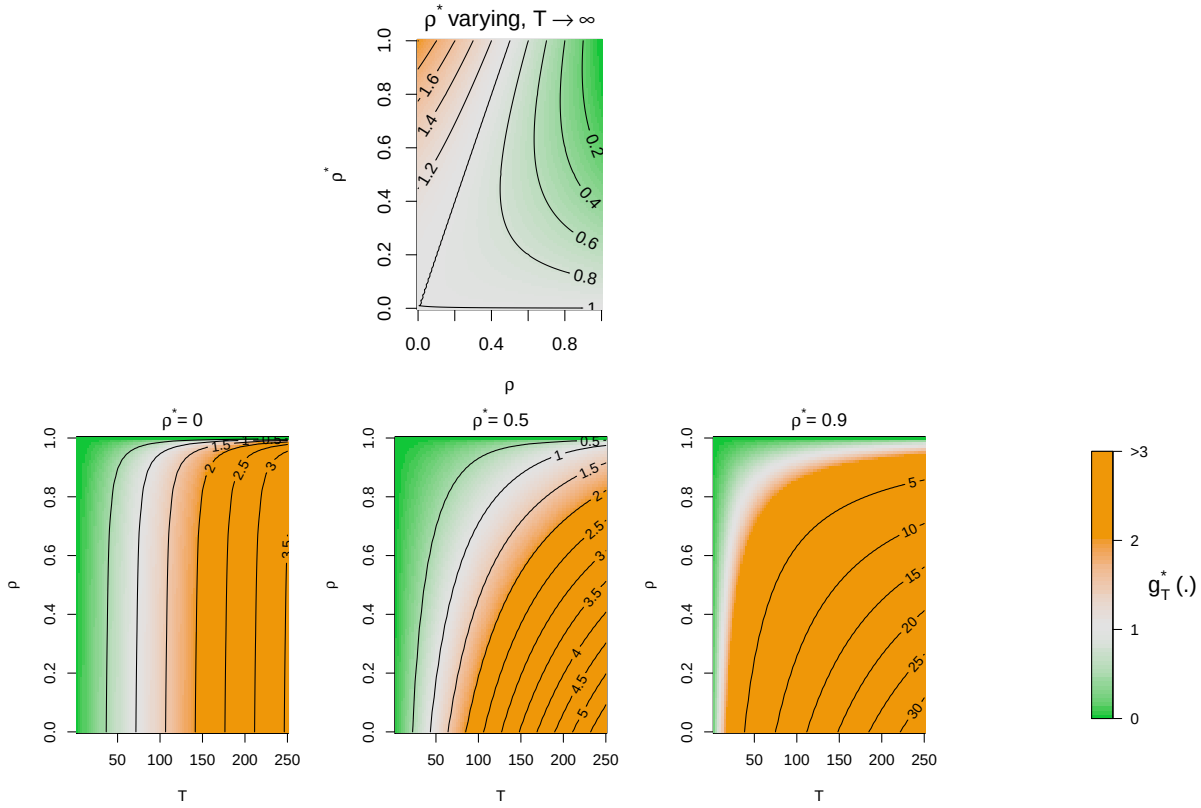


FIGURE 3.7.: *Above: Weights $g_T^*(\rho)$ for $T \to \infty$ and varying ρ and ρ^*; below: Weights $g_T^*(\rho)$ for varying ρ, ρ^* and T; SF=1, $\rho^* = 0.7, \sigma^2_{\tau^*} = 140, \sigma^2_{\varepsilon^*} = 70$.*

with

$$T_{\text{eff}} = \frac{T^2}{T + 2\sum_{k=1}^{T-1}\sum_{t=1}^{k}\rho_Z^k}.$$

If $\rho^* = 0$ and $\frac{\sigma^2_{\varepsilon^*}}{\sigma^2_{\tau^*}} + T \approx T$, $g_T^*(\rho_Z) \approx (1 - \frac{1}{T_{\text{eff}}})$:

$$\begin{aligned} g_T^*(\rho_Z) &= 1 - \frac{T + 2\rho_Z^{T-1} + 4\sum_{t=1}^{T-2}\rho_Z^t + 2\sum_{k=1}^{T-3}\sum_{t=1}^{k}\rho_Z^k}{T\left(T + \frac{\sigma^2_{\varepsilon^*}}{\sigma^2_{\tau^*}}\right)} \\ &= 1 - \frac{T + 2\sum_{k=1}^{T-1}\sum_{t=1}^{k}\rho_Z^k}{T\left(T + \frac{\sigma^2_{\varepsilon^*}}{\sigma^2_{\tau^*}}\right)} \\ &\approx 1 - \frac{1}{T^2}\left\{T + 2\sum_{k=1}^{T-1}\sum_{t=1}^{k}\rho_Z^t\right\} \\ &= 1 - \frac{1}{T_{\text{eff}}}. \end{aligned}$$

Thus, $\sigma_Z^2 g_T^*(\rho_Z) \approx \mathbb{E}(s_Z^2)$ for $\rho^* = 0$ and $\frac{\sigma^2_{\varepsilon^*}}{\sigma^2_{\tau^*}} + T \approx T$, i.e. for any first–order autocorrelated random variable $\boldsymbol{Z}$, with variance σ_Z^2 and correlation coefficient ρ_Z, $\sigma_Z^2 g_T^*(\rho_Z)$ is approximately equal to the expected biased sample variance of $\boldsymbol{Z}$, if $\rho^* = 0$ and $\frac{\sigma^2_{\varepsilon^*}}{\sigma^2_{\tau^*}} + T \approx T$.

3.6.5. Impact of missing value adjustment

The influence of the structure of missing values on the bias calculation of the effect coefficient was theoretically considered in Subsection 3.5.5. In the following, the impact

of neglecting the presence of missing values for bias calculation is explored. Therefore, the calculation of the attenuation factor for classical measurement error allowing for missing values according to eq. 3.45 is compared to the calculation according to eq. 3.27, which neglects the missing value structure.

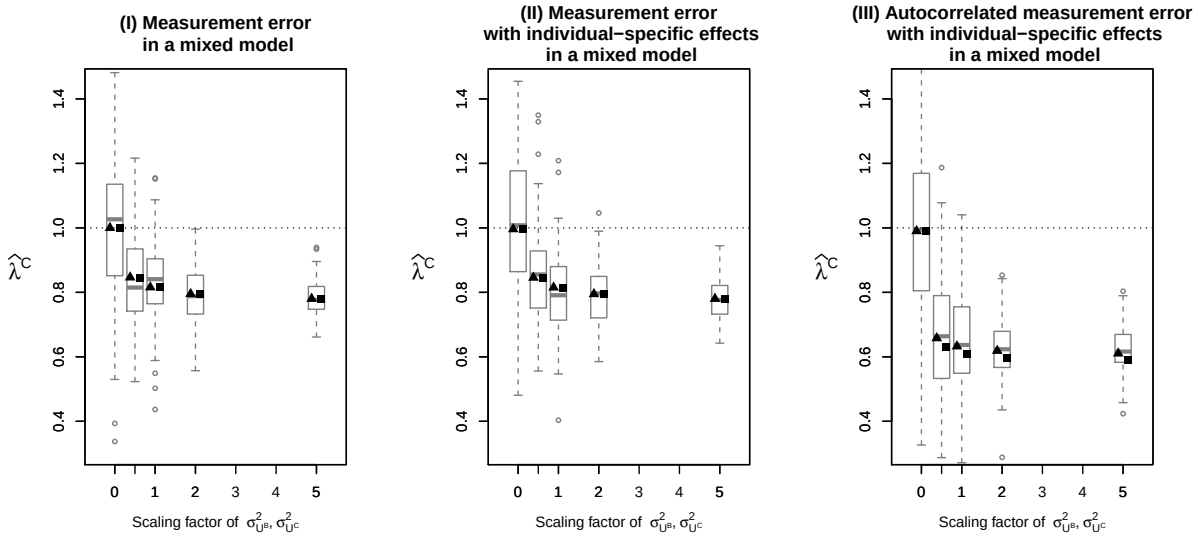


FIGURE 3.8.: *Simulation results for the empirical attenuation factor depicted in boxplots for increasing size of measurement errors; the theoretical estimations of λ according to 3.45 and 3.41 are marked with triangles for the exact values of $\boldsymbol{G}$ and with squares for $\boldsymbol{G} = \mathbf{0}_{75}$.*

The estimations for λ^{C} neglecting the missing values (filled squares in Figure 3.8) are similar to the exactly estimated values for λ^{C} in Scenario (I) and (II). In contrast, biased estimations are obtained in Scenario (III): the degree of attenuation is overestimated. Especially in Scenario (III), the attenuation factor depends on the percentage of missing values as displayed in Figure 3.9.

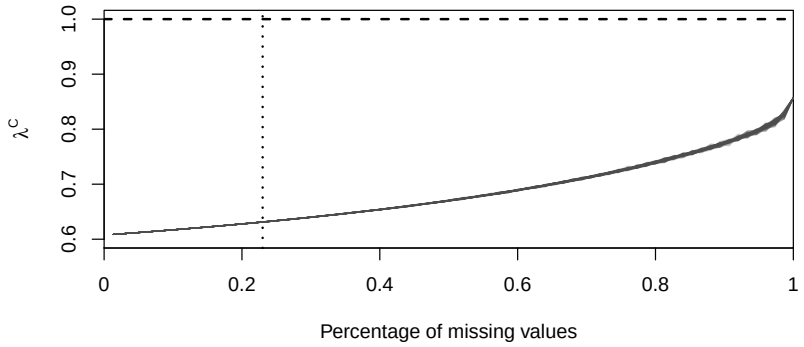


FIGURE 3.9.: *Theoretical attenuation factors for classical measurement error in Scenario (III) for varying percentages of missing values; vertical line: percentage of missing values used for simulations; $T = 75$, SF=1, $\rho^* = 0.7, \sigma^2_{\tau^*} = 140, \sigma^2_{\varepsilon^*} = 70$, 100 missing structures for each percentage of missing values.*

The attenuation of the effect and the variability of λ^{C} are increasing with the percentage of missing values.

3.6.6. Further covariates

For the Augsburger Umweltstudie we assume that the individual PNC level is the only parameter measured with error. Besides exposure to PNC, variable selection within the

main study identified some further covariates with a significant concurrent or lagged impact on the heart rate of the individuals (Hampel et al., 2012b). The following additional covariates affect the effect estimate of PNC, because they are correlated (according to the correlation coefficient of Spearman) with the fixed–site and thus also with the individual PNC measurements: temperature at lag 2 (r = -0.399), quadratic temperature at lag 2 (r = -0.376) and relative humidity at lag 1 (r = 0.17).

We examine the influence of correlated covariates on the attenuation of the effect coefficient in the framework of classical and mixture measurement error using $N = 100$ bootstrap samples (of the individuals) from the original data set. Realizations of $\boldsymbol{X}, \boldsymbol{X}^{*\mathrm{C}}$ and $\boldsymbol{X}^{*\mathrm{M}}$ are generated according to the settings described in Subsection 3.6.1 based on the log–transformed samples of the fixed–site measurements as realizations of $\boldsymbol{X}^{*\mathrm{B}}$. The main health outcome is generated with the simulated individual PNC measurements and the corresponding measurements of temperature, relative humidity, time of a day and the trend variable of the bootstrap sample using the naive parameter estimates of the precisely measured parameters in the main model of the Augsburger Umweltstudie. The attenuation effect is estimated with $\widehat{\beta_X^*}/\widehat{\beta_X}$ after the calculation of $\widehat{\beta_X}$ using eq. 3.42.

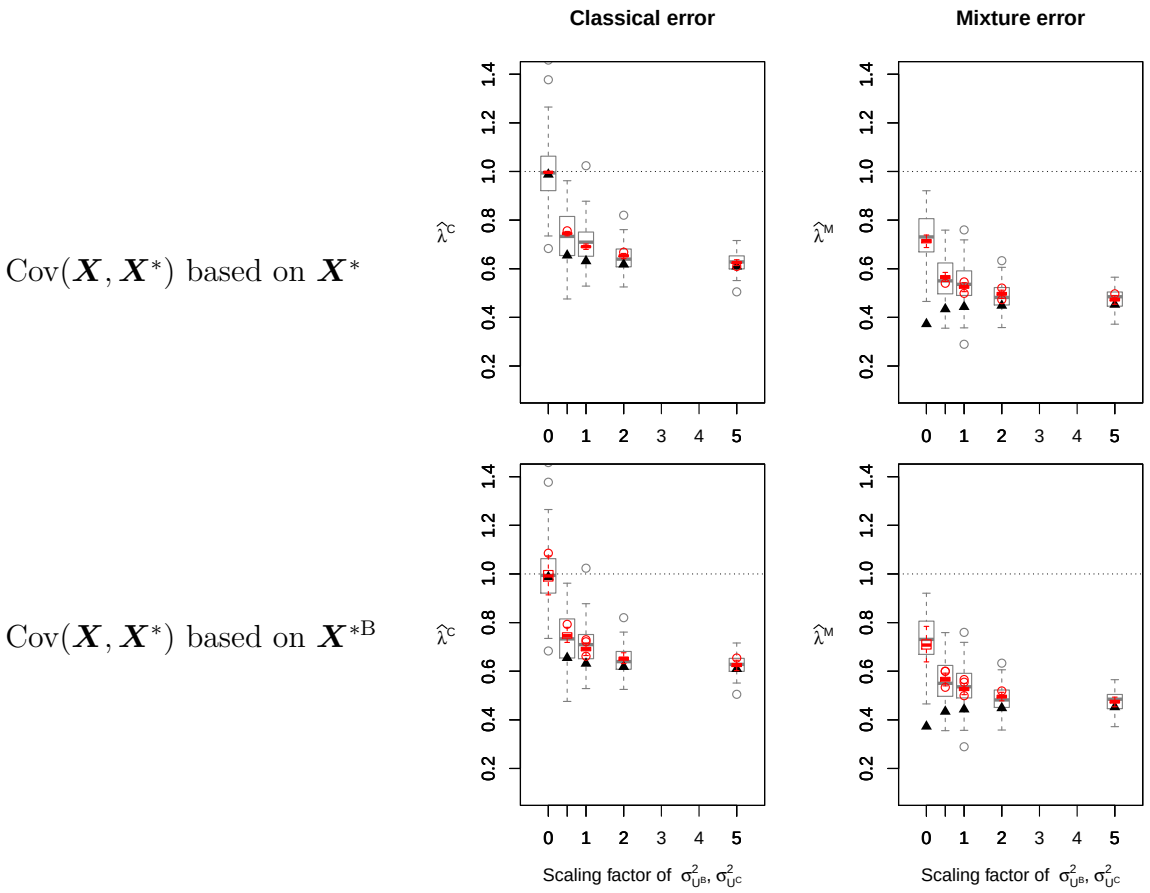


FIGURE 3.10.: *Simulation results of Scenario (III) for the empirical attenuation factor for increasing size of classical and Berkson measurement error in the presence of further covariates. Gray boxplots: empirical attenuation factor; red boxplots: estimated attenuation factors according to eq. 3.42 and eq. 3.36–3.39; triangles: estimated attenuation factors according to eq. 3.45 and eq. 3.35 assuming independence between the deficient and precisely measured covariates.*

The results in Figure 3.10 show that the degree of attenuation due to classical and mixture measurement error is reduced when further confounder variables (which correlate with the error–prone measurements) are included into the model. The attenuation factor

SF of	Truth		Classical error		Mixture error	
$\sigma^2_{U^{\mathrm{B}}}, \sigma^2_{U^{\mathrm{C}}}$	Bias	RMSE	Bias	RMSE	Bias	RMSE
0	0.014	0.219	**0.015**	**0.248**	0.038	0.275
0.5	0.010	0.132	**-0.010**	**0.169**	-0.010	0.176
1	0.002	0.123	**0.022**	**0.170**	0.025	0.188
2	0.007	0.090	**-0.005**	**0.126**	-0.008	0.144
5	0.008	0.063	**0.011**	**0.089**	0.022	0.104

TABLE 3.3.: *Bias and RMSE of MOM–corrected effect estimations for Scenarios (I)-(III) in multiple regression; bold: lowest absolute bias and lowest RMSE of each row (disregarding the results based on the true data).*

of the effect in the multiple regression model would be underestimated, if the correlation between the covariates is neglected and the correction formulas for the simple regression model are applied. Correcting the naive estimations according to eq. 3.42 yields preciser estimations. Bias and RMSE of the estimations are listed in Table 3.3. Measurement error correction based on classical measurement error–prone covariates provides better estimations regarding RMSE than measurement error correction based on mixture measurement error with the used parameter choice.

4. Bias correction with the method of moments

As seen in Chapter 3, measurements with classical or mixture error deliver biased estimates of the regression coefficient $\beta_X^* = \lambda\beta_X$ in a simple linear mixed model. Knowing the formula for λ enables us to correct bias due to classical or mixture measurement error; bias correction is not necessary in the case of Berkson error. For bias correction with the method of moments the components of the attenuation factor $\lambda \in \{\lambda^{\mathrm{C}}, \lambda^{\mathrm{M}}\}$ are directly estimated through their empirical equivalents resulting e.g. from validation data or repeated measurements. Subsequently, the estimated naive effect coefficient is corrected with the estimated attenuation factor $\widehat{\lambda}$:

$$\widehat{\beta_X} = \frac{1}{\widehat{\lambda}}\widehat{\beta_X^*}$$

This approach is known as the method of moments (e.g. Carroll et al., 2006).

4.1. Distributional properties of the attenuation factor

The variability of the estimated attenuation factor $\widehat{\lambda}$ influences the variability of the bias corrected effect coefficient $\widehat{\beta_X}$. Therefore, distributional properties of $\widehat{\lambda}$ are examined in the following.

4.1.1. Distribution of $\widehat{\lambda}$ (simple form)

For first considerations, we assume, that $X_i \sim \mathrm{N}(0, \sigma_X^2)$, $U_i \sim \mathrm{N}(0, \sigma_U^2)$, $X_i^* = X_i + U_i$, $i = 1, \ldots, n$ and X_i and U_i are independent. The attenuation factor according to the method of moments is:

$$\widehat{\lambda} = \frac{S_X^2}{S_{X^*}^2} = \frac{S_X^2}{S_X^2 + S_U^2},$$

with S_X^2, S_U^2 and $S_{X^*}^2$ denoting the respective unbiased sample variances. Thus, $\widehat{\lambda}$ is the ratio of two dependent χ^2–distributed, and for n large enough normally distributed, random variables S_X^2 and $S_{X^*}^2$.

Pham-Gia et al. (2006) found an analytical expression for the distribution of the ratio of two dependent bivariate normally distributed random variables using Kummer's classical confluent hypergeometric function of the first kind. Also recent works of Kim (2015) and Wang et al. (2015) examine this class of distributions.

A visual comparison of the empirical density of $\widehat{\lambda}$ based on simulated data ($\sigma_X^2 = 1; \sigma_U^2 = 0.25$), the density according to Pham-Gia et al. (2006), the density of the F distribution – the distribution of the ratio between two independent χ^2–distributed random variables – and the Beta distribution is presented in Figure 4.1.

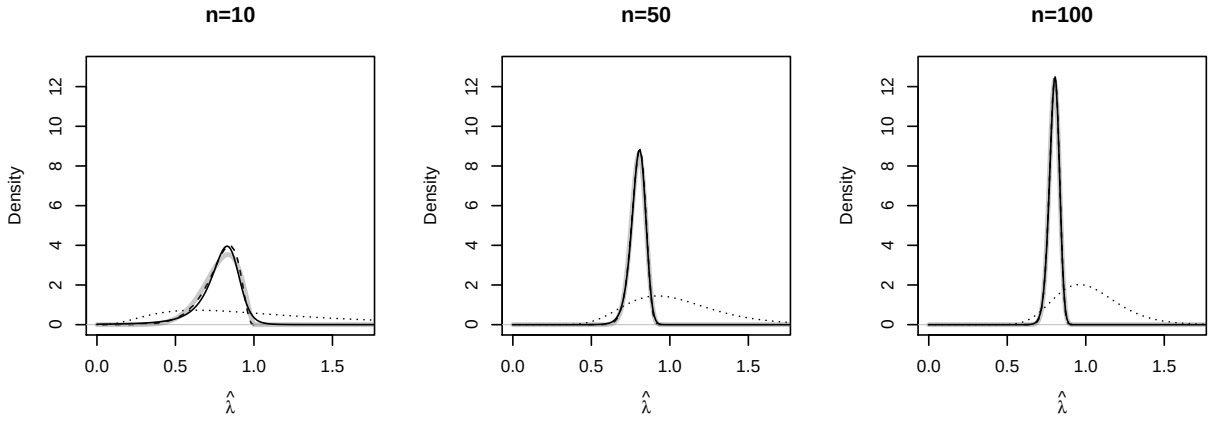


FIGURE 4.1.: *Comparison of the empirical density of $\widehat{\lambda}$ (dashed lines) based on simulated data, the closed form expression of the density according to Pham-Gia et al. (2006) (solid lines), the appropriate density assuming a F distribution (dotted lines) and the appropriate density of a Beta distribution (gray) for varying sample size n.*

Especially for a large number of observations, the complex density of Pham-Gia et al. (2006) seems to be well approximated through the Beta distribution with expectation $\mathbb{E}(\widehat{\lambda})$ and variance $\mathrm{Var}(\widehat{\lambda})$: the gray line indicating the Beta distributions in Figure 4.1 is nearly not visible due to overlaps of the density according to Pham-Gia et al. (2006) and the empirical density. In contrast, the F distribution does not provide appropriate results as expected.

4.1.2. Approximation of the first three central moments of $\widehat{\lambda}$ (simple form)

Instead of using the lengthy formula of the density, the expectation value, the variance and the skewness of $\widehat{\lambda}$ are examined in detail using the second–order Taylor series expansion assuming independence between S_X^2 and S_U^2.

Since $X_i \sim \mathrm{N}(0, \sigma_X^2)$ and $U_i \sim \mathrm{N}(0, \sigma_U^2), i = 1, \ldots, n,$, $(n-1)\frac{S_X^2}{\sigma_X^2} \sim \chi^2_{n-1}$ and $(n-1)\frac{S_U^2}{\sigma_U^2} \sim \chi^2_{n-1}$. The k–th raw moment of a χ^2–distributed random variable Y with m degrees of freedom can be derived by

$$\begin{aligned}\mathbb{E}\left(Y^k\right) &= 2^k \frac{\Gamma(k+0.5m)}{\Gamma(0.5m)} \\ &= m(m+2)\cdots(m+2k-2) \\ &= \prod_{i=1}^{k}(m+2i-2)\end{aligned}$$

according to Walck (2007), which is used to calculate $\mathbb{E}\left[\left(S_X^2\right)^k\right]$:

$$\mathbb{E}\left[\left(S_X^2\right)^k\right] = \prod_{i=1}^{k}\left[\frac{\sigma_X^2}{n-1}(n+2i-2)\right] .$$

By means of

$$\mathbb{E}\left[(Y-\mathbb{E}(Y))^k\right] = \sum_{i=0}^{k}\binom{k}{i}(-1)^{k-i}\mathbb{E}\left(Y^i\right)\left[\mathbb{E}(Y)\right]^{k-i}$$

with $\mathbb{E}\left(Y^0\right) = 1$, central moments of an arbitrary random variable Z can be deviated from raw moments (Papoulis, 1991). Hence, the first raw and the second to the sixth central moments of S_X^2 are

$$\begin{aligned}
\mathbb{E}\left(S_X^2\right) &= \sigma_X^2 \\
\mathbb{E}\left[(S_X^2 - \mathbb{E}(S_X^2))^2\right] &= \frac{2}{n-1}\sigma_X^4 \\
\mathbb{E}\left[(S_X^2 - \mathbb{E}(S_X^2))^3\right] &= \frac{8}{(n-1)^2}\sigma_X^6 \\
\mathbb{E}\left[(S_X^2 - \mathbb{E}(S_X^2))^4\right] &= \frac{12(n+3)}{(n-1)^3}\sigma_X^8 \\
\mathbb{E}\left[(S_X^2 - \mathbb{E}(S_X^2))^5\right] &= \frac{32(12+5(n-1))}{(n-1)^4}\sigma_X^{10} \\
\mathbb{E}\left[(S_X^2 - \mathbb{E}(S_X^2))^6\right] &= \frac{40(3(n-1)^2+52(n-1)+96)}{(n-1)^5}\sigma_X^{12}
\end{aligned}$$

The moments of S_U^2 are obtained analogously.

Let the function $h : \mathbb{R}^2 \to \mathbb{R}$ be defined by $h(\gamma_1, \gamma_2) = \frac{\gamma_1}{\gamma_1+\gamma_2}$ with the Jacobian matrix

$$J_h(\gamma_1, \gamma_2) = \begin{pmatrix} \frac{\gamma_2}{(\gamma_1+\gamma_2)^2} & -\frac{\gamma_1}{(\gamma_1+\gamma_2)^2} \end{pmatrix}$$

and the Hessian matrix

$$H_h(\gamma_1, \gamma_2) = \begin{pmatrix} -2\frac{\gamma_2}{(\gamma_1+\gamma_2)^3} & \frac{\gamma_1-\gamma_2}{(\gamma_1+\gamma_2)^3} \\ \frac{\gamma_1-\gamma_2}{(\gamma_1+\gamma_2)^3} & 2\frac{\gamma_1}{(\gamma_1+\gamma_2)^3} \end{pmatrix}.$$

The second–order Taylor series expansion of $h(S_X^2, S_U^2)$ around $\begin{pmatrix}\mathbb{E}(S_X^2) & \mathbb{E}(S_U^2)\end{pmatrix}^\top = \begin{pmatrix}\sigma_X^2 & \sigma_U^2\end{pmatrix}^\top$ is given by

$$\begin{aligned}
\widehat{\lambda} = h\left(S_X^2, S_U^2\right) \approx & \frac{\sigma_X^2}{\sigma_X^2+\sigma_U^2} + \frac{\sigma_U^2}{\left(\sigma_X^2+\sigma_U^2\right)^2}\left(S_X^2-\sigma_X^2\right) - \frac{\sigma_X^2}{\left(\sigma_X^2+\sigma_U^2\right)^2}\left(S_U^2-\sigma_U^2\right) - \\
& \frac{\sigma_U^2}{\left(\sigma_X^2+\sigma_U^2\right)^3}\left(S_X^2-\sigma_X^2\right)^2 + \frac{\sigma_X^2}{\left(\sigma_X^2+\sigma_U^2\right)^3}\left(S_U^2-\sigma_U^2\right)^2 + \\
& \frac{\sigma_X^2-\sigma_U^2}{\left(\sigma_X^2+\sigma_U^2\right)^3}\left(S_X^2-\sigma_X^2\right)\left(S_U^2-\sigma_U^2\right).
\end{aligned}$$

Thus, the first moment and the second and third central moments of $\widehat{\lambda}$ are approximated by

$$\begin{aligned}
\mathbb{E}\left(\widehat{\lambda}\right) \approx & \frac{\sigma_X^2}{\sigma_X^2+\sigma_U^2} - \frac{\sigma_U^2}{(\sigma_X^2+\sigma_U^2)^3}\mathrm{Var}\left(S_X^2\right) + \frac{\sigma_X^2}{(\sigma_X^2+\sigma_U^2)^3}\mathrm{Var}\left(S_U^2\right) + \\
& \frac{\sigma_X^2-\sigma_U^2}{(\sigma_X^2+\sigma_U^2)^3}\mathrm{Cov}\left(S_X^2, S_U^2\right) \\
= & \frac{\sigma_X^2}{\sigma_X^2+\sigma_U^2} = \lambda
\end{aligned}$$

$$
\begin{aligned}
\mathrm{Var}\left(\widehat{\lambda}\right) \approx & \frac{\sigma_U^4}{(\sigma_X^2+\sigma_U^2)^4}\mathrm{Var}\left(S_X^2\right)+\frac{\sigma_X^4}{(\sigma_X^2+\sigma_U^2)^4}\mathrm{Var}\left(S_U^2\right)- \\
& 2\frac{\sigma_U^4}{(\sigma_X^2+\sigma_U^2)^5}\mathbb{E}\left[\left(S_X^2-\mathbb{E}\left(S_X^2\right)\right)^3\right]-2\frac{\sigma_X^4}{(\sigma_X^2+\sigma_U^2)^5}\mathbb{E}\left[\left(S_U^2-\mathbb{E}\left(S_U^2\right)\right)^3\right]+ \\
& \frac{\sigma_U^4}{(\sigma_X^2+\sigma_U^2)^6}\mathbb{E}\left[\left(S_X^2-\mathbb{E}\left(S_X^2\right)\right)^4\right]+\frac{\sigma_X^4}{(\sigma_X^2+\sigma_U^2)^6}\mathbb{E}\left[\left(S_U^2-\mathbb{E}\left(S_U^2\right)\right)^4\right]+ \\
& \frac{(\sigma_X^2-\sigma_U^2)^2}{(\sigma_X^2+\sigma_U^2)^6}\mathrm{Var}\left(S_X^2\right)\mathrm{Var}\left(S_U^2\right)-2\frac{\sigma_X^2\sigma_U^2}{(\sigma_X^2+\sigma_U^2)^6}\mathrm{Var}\left(S_X^2\right)\mathrm{Var}\left(S_U^2\right) \\
= & \frac{\sigma_X^4\sigma_U^4}{(\sigma_X^2+\sigma_U^2)^4}\frac{4}{n-1}-\frac{\sigma_X^4\sigma_U^4}{(\sigma_X^2+\sigma_U^2)^4}\frac{16}{(n-1)^2}+ \\
& \frac{\sigma_X^4\sigma_U^4}{(\sigma_X^2+\sigma_U^2)^6}\frac{12(n+3)}{(n-1)^3}(\sigma_X^4+\sigma_U^4)+\frac{\sigma_X^4\sigma_U^4}{(\sigma_X^2+\sigma_U^2)^6}\frac{4}{(n-1)^2}(\sigma_X^4-4\sigma_X^2\sigma_U^2+\sigma_U^4) \\
= & \lambda^2(1-\lambda)^2\left[\frac{4}{n-1}-\frac{16}{(n-1)^2}\right]+ \\
& \lambda^2(1-\lambda)^2(\lambda^2+(1-\lambda)^2)\left(\frac{4}{(n-1)^2}+\frac{12(n+3)}{(n-1)^3}\right)- \\
& \lambda^3(1-\lambda)^3\frac{16}{(n-1)^2}
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}[(\widehat{\lambda}-\mathbb{E}(\widehat{\lambda}))^3] \approx & \mathbb{E}\Big[\frac{\sigma_U^2}{(\sigma_X^2+\sigma_U^2)^2}(S_X^2-\sigma_X^2)-\frac{\sigma_X^2}{(\sigma_X^2+\sigma_U^2)^2}(S_U^2-\sigma_U^2)- \\
& \frac{\sigma_U^2}{(\sigma_X^2+\sigma_U^2)^3}(S_X^2-\sigma_X^2)^2+\frac{\sigma_X^2}{(\sigma_X^2+\sigma_U^2)^3}(S_U^2-\sigma_U^2)^2+ \\
& \frac{\sigma_X^2-\sigma_U^2}{(\sigma_X^2+\sigma_U^2)^3}(S_X^2-\sigma_X^2)(S_U^2-\sigma_U^2)\Big]^3 \\
= & \frac{\sigma_X^6\sigma_U^6}{8(\sigma_X^2+\sigma_U^2)^6}\left\{36\frac{\sigma_X^2-\sigma_U^2}{\sigma_X^2+\sigma_U^2}\left[-\frac{1}{(n-1)^2}-\frac{n+3}{(n-1)^3}\right]+\right. \\
& 576\frac{\sigma_X^4-\sigma_U^4}{(\sigma_X^2+\sigma_U^2)^2}\left[\frac{1}{(n-1)^3}+\frac{2}{(n-1)^4}\right]+ \\
& 8\frac{\sigma_X^6-\sigma_U^6}{(\sigma_X^2+\sigma_U^2)^3}\left[-\frac{15}{(n-1)^3}-\frac{9(n+3)}{(n-1)^4}-\frac{252}{(n-1)^4}-\frac{480}{(n-1)^5}\right]+ \\
& \left.288\frac{\sigma_X^2\sigma_U^2(\sigma_X^2-\sigma_U^2)}{(\sigma_X^2+\sigma_U^2)^3}\left[\frac{n+3}{(n-1)^4}-\frac{2}{(n-1)^4}\right]\right\} \qquad (4.1)
\end{aligned}
$$

The estimator $\widehat{\lambda}$ is unbiased for λ and weakly consistent. The second and the third central moments of $\widehat{\lambda}$ are visualized in Figure 4.2 depending on the sample size n and the attenuation factor λ.

The variance of $\widehat{\lambda}$ is maximal for λ=0.5, i.e. for $\sigma_X^2=\sigma_U^2$. The third central moment indicates the symmetry of $\widehat{\lambda}$ as it is simply deduced from eq. 4.1. $\widehat{\lambda}$ is negatively skewed for $\sigma_X^2>\sigma_U^2$, positively skewed for $\sigma_X^2<\sigma_U^2$ and symmetric for $n\to\infty$ as well as for $\lambda=0.5$.

All the results of this subsection can easily be extended to the following structures of the estimated attenuation factor $\widehat{\lambda}$:

$$\widehat{\lambda}=\frac{\sum_{j=1}^m f_j(S_j^2)}{\sum_{k=1}^l g_k(S_k^2)}.$$

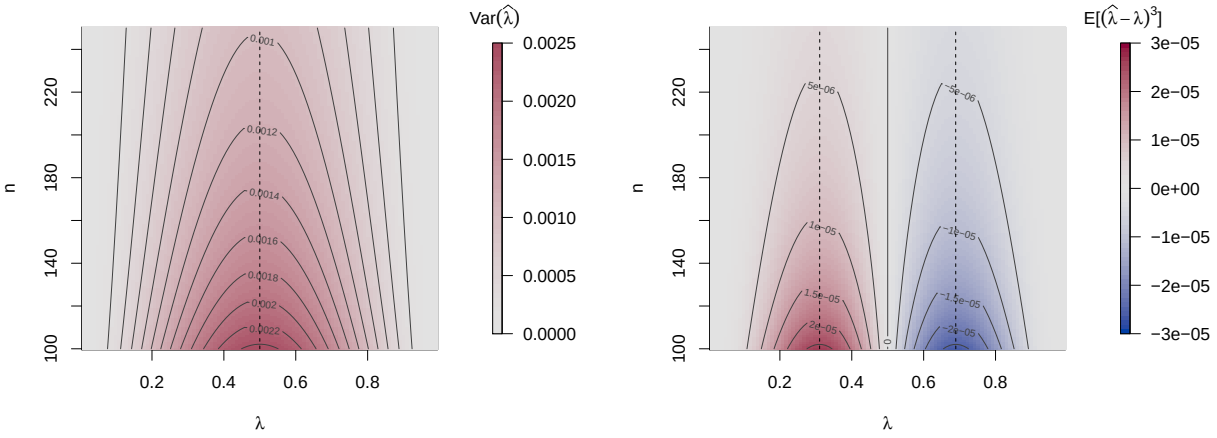


FIGURE 4.2.: *Approximate second (left) and third (right) central moment of $\widehat{\lambda}$ for varying values of the attenuation factor λ and the sample size n. The dashed lines mark the extreme values of the moments for different sample sizes.*

S_j^2 and S_k^2 denote sample variances of arbitrary random variables and f_j and g_j arbitrary deterministic linear transformations; $m, l \in \mathbb{N}$.

4.1.3. Approximation of the first three central moments of $\widehat{\lambda}$ (complex form)

The attenuation factor for autocorrelated data or autocorrelated measurement error has a more complex structure as considered in Subsection 4.1.2. The estimation of the autocorrelation coefficients involves an additional uncertainty for the estimation of the attenuation factor.

For the considerations in this subsection, we assume for the order–one autocorrelated random variables X_i and U_i with autocorrelation coefficients ρ_X and ρ_U, that $X_i \sim \mathrm{N}(0, \sigma_X^2)$ and $U_i \sim \mathrm{N}(0, \sigma_U^2), i = 1, \ldots, n,$. Further, let $X_i^* = X_i + U_i$, and X_i and U_i are independent for $i = 1, \ldots, n$. The empirical attenuation factor for the error–prone covariate X^* in a linear mixed model with order–one autocorrelated error term has the following general structure:

$$\widehat{\lambda} = h(\boldsymbol{\gamma}) = \frac{g_\infty^*(r_X) S_X^2}{g_\infty^*(r_X) S_X^2 + g_\infty^*(r_U) S_U^2} \ ,$$

with $\boldsymbol{\gamma} = (S_X^2, S_U^2, r_X, r_U, r^*)^\top$, $g_\infty^*(r) := 1 - 2rr^* + r^{*2}$ and r^*, r_X and r_U denoting sample autocorrelations of lag 1 for the autocorrelation coefficient of X (ρ_X), U (ρ_U) and ε^* (ρ^*). In addition to the assumption that S_X^2 and S_U^2 are independent, the independence with and between r_X, r_U and r is supposed for simplicity. The variance of $\widehat{\lambda}$ is approximated with the multivariate delta rule:

$$\begin{aligned}\operatorname{Var}\left(\widehat{\lambda}\right) \approx {} & [J_h(\mathbb{E}\boldsymbol{\gamma})]_{1,1} \operatorname{Var}\left(S_X^2\right) + [J_h(\mathbb{E}\boldsymbol{\gamma})]_{1,2} \operatorname{Var}\left(S_U^2\right) + [J_h(\mathbb{E}\boldsymbol{\gamma})]_{1,3} \operatorname{Var}(r_X) + \\ & [J_h(\mathbb{E}\boldsymbol{\gamma})]_{1,4} \operatorname{Var}(r_U) + [J_h(\mathbb{E}\boldsymbol{\gamma})]_{1,5} \operatorname{Var}(r^*) \ .\end{aligned}$$

$J_h(\cdot)$ denotes the Jacobian matrix for function h.

If ε^*, X, and U are approximately normally distributed, the variance of the corresponding empirical correlation coefficients r_U, r_X and r^* can be approximately estimated with $\frac{1-r^2}{m}$; $r \in \{r_U, r_X, r^*\}$ and m denoting the sample size (Bartlett, 1946). The approximation is imprecise for low sample size and high autocorrelation. Since the explicit formula for

$\text{Var}\left(\widehat{\lambda}\right)$ is elongate, approximations are computationally calculated. Figure 4.3 depicts the approximate variance of $\widehat{\lambda}$ depending on λ for different empirical correlation coefficients. The parameters are defined as: $m = 1000, S_X^2 = 3, r^* = 0.7$. Varying values of $\widehat{\lambda}$ are generated by varying values of S_U^2.

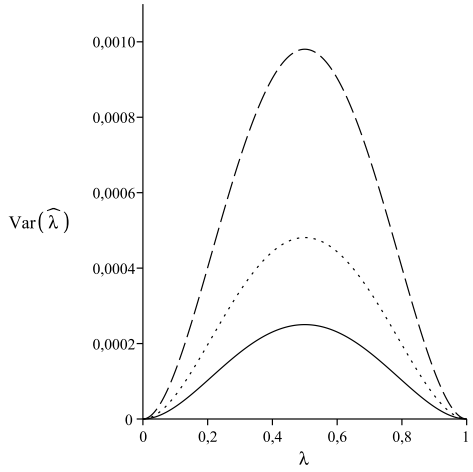


FIGURE 4.3.: *Approximate Var*$(\widehat{\lambda})$ *depending on* λ *and the empirical correlation coefficients* $r_X = 0$, $r_U = 0$ *(solid line),* $r_X = 0.5$, $r_U = 0.2$ *(dotted line) and* $r_X = 0.9$, $r_U = 0.2$ *(dashed line).*

Calculations for the more general form of the attenuation factor reveal the same shape of $\text{Var}\left(\widehat{\lambda}\right)$ regarding to varying λ as in the simple case, but the amplitude of $\text{Var}\left(\widehat{\lambda}\right)$ strongly depends on the size of the correlation coefficients.

The approximation of $\text{Var}\left(\widehat{\lambda}\right)$ can easily be adapted to a differing number of repetitions for each component of $\widehat{\lambda}$ (e.g. n_1 replications for the estimation of the size of the Berkson error and n_2 replications for the estimation of the size of the classical measurement error). In the case of a validation study without repeated measurements but with panel data, as in the Augsburger Umweltstudie, the specification of n is not trivial, but the above formulas give rough evidence about the variation of the estimated attenuation factor.

4.2. Confidence intervals for the bias corrected effect coefficient

The variability of the bias corrected effect coefficient $\widehat{\beta_X}$ depends on the variability of the estimated attenuation factor $\widehat{\lambda}$, which was discussed in the previous section. Numerous methods for estimating the variability of the measurement error corrected effect estimate are proposed and applied in the literature, e.g. jackknife (McShane et al., 2001). Two approaches for obtaining confidence intervals for $\widehat{\beta_X}$ are examined in the following: the delta method and bootstrapping.

4.2.1. Confidence intervals using the delta method

Let $\widehat{\boldsymbol{\theta}} = \left(\widehat{\beta_X^*} \quad \widehat{\lambda}\right)^\top$ with mean $\boldsymbol{\mu}_{\widehat{\boldsymbol{\theta}}}$ and variance–covariance matrix $\boldsymbol{\Sigma}_{\widehat{\boldsymbol{\theta}}}$. $\widehat{\beta_X^*}$ and $\widehat{\lambda}$ are assumed to be independent. According to the multivariate delta method the variance of

the corrected effect $\widehat{\beta_X} = h\left(\widehat{\boldsymbol{\theta}}\right) := \widehat{\beta_X^*}/\widehat{\lambda}$ with Jacobian matrix $J_h\left(\widehat{\boldsymbol{\theta}}\right) = \left(\frac{1}{\widehat{\lambda}} \quad -\frac{\widehat{\beta_X^*}}{\widehat{\lambda}^2}\right)$ is approximated through

$$\begin{aligned}
\mathrm{Var}\left(\widehat{\beta_X}\right) &\approx J_h(\boldsymbol{\mu}_{\widehat{\boldsymbol{\theta}}})\boldsymbol{\Sigma}_{\widehat{\boldsymbol{\theta}}}J_h(\boldsymbol{\mu}_{\widehat{\boldsymbol{\theta}}})^\top \\
&= \left(\frac{1}{\mathbb{E}\left(\widehat{\lambda}\right)} \quad -\frac{\mathbb{E}\left(\widehat{\beta_X^*}\right)}{\mathbb{E}\left(\widehat{\lambda}\right)^2}\right)\begin{pmatrix}\mathrm{Var}\left(\widehat{\beta_X^*}\right) & 0\\ 0 & \mathrm{Var}\left(\widehat{\lambda}\right)\end{pmatrix}\begin{pmatrix}\frac{1}{\mathbb{E}\left(\widehat{\lambda}\right)}\\ -\frac{\mathbb{E}\left(\widehat{\beta_X^*}\right)}{\mathbb{E}\left(\widehat{\lambda}\right)^2}\end{pmatrix} \\
&= \frac{1}{\mathbb{E}\left(\widehat{\lambda}\right)^2}\mathrm{Var}\left(\widehat{\beta_X^*}\right) + \frac{\mathbb{E}\left(\widehat{\beta_X^*}\right)^2}{E\left(\widehat{\lambda}\right)^4}\mathrm{Var}\left(\widehat{\lambda}\right)\ , \qquad (4.2)
\end{aligned}$$

as e.g. in Rosner et al. (1989). Using the second–order Taylor series expansion of $h\left(\widehat{\boldsymbol{\theta}}\right)$ with the Hessian matrix

$$H_h\left(\widehat{\boldsymbol{\theta}}\right) = \begin{pmatrix} 0 & -\frac{1}{\widehat{\lambda}^2} \\ -\frac{1}{\widehat{\lambda}^2} & 2\frac{\widehat{\beta_X^*}}{\widehat{\lambda}^3}\end{pmatrix}$$

of h for the approximation of $\mathrm{Var}\left(\widehat{\beta_X}\right)$ leads to

$$\begin{aligned}
\mathrm{Var}\left(\widehat{\beta_X}\right) &= \left[h\left(\widehat{\boldsymbol{\theta}}\right) - h\left(\boldsymbol{\mu}_{\widehat{\boldsymbol{\theta}}}\right)\right]^2 \\
&\approx \mathbb{E}\left[J_h\left(\boldsymbol{\mu}_{\widehat{\boldsymbol{\theta}}}\right)\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\mu}_{\widehat{\boldsymbol{\theta}}}\right) + \frac{1}{2}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\mu}_{\widehat{\boldsymbol{\theta}}}\right)^\top H_h\left(\boldsymbol{\mu}_{\widehat{\boldsymbol{\theta}}}\right)\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\mu}_{\widehat{\boldsymbol{\theta}}}\right)\right]^2 \\
&= \underbrace{J_h(\boldsymbol{\mu}_{\widehat{\boldsymbol{\theta}}})\boldsymbol{\Sigma}_{\widehat{\boldsymbol{\theta}}}J_h(\boldsymbol{\mu}_{\widehat{\boldsymbol{\theta}}})^\top}_{\text{(I)}} + \\
&\quad \underbrace{\mathbb{E}\left[J_h(\boldsymbol{\mu}_{\widehat{\boldsymbol{\theta}}})\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\mu}_{\widehat{\boldsymbol{\theta}}}\right)\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\mu}_{\widehat{\boldsymbol{\theta}}}\right)^\top H_h(\boldsymbol{\mu}_{\widehat{\boldsymbol{\theta}}})\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\mu}_{\widehat{\boldsymbol{\theta}}}\right)\right]}_{\text{(II)}} + \\
&\quad \underbrace{\mathbb{E}\left[\frac{1}{4}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\mu}_{\widehat{\boldsymbol{\theta}}}\right)^\top H_h(\boldsymbol{\mu}_{\widehat{\boldsymbol{\theta}}})\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\mu}_{\widehat{\boldsymbol{\theta}}}\right)\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\mu}_{\widehat{\boldsymbol{\theta}}}\right)^\top H_h(\boldsymbol{\mu}_{\widehat{\boldsymbol{\theta}}})\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\mu}_{\widehat{\boldsymbol{\theta}}}\right)\right]}_{\text{(III)}}.
\end{aligned}$$

The summands (I)-(III) are calculated in the following:

(I)

$$J_h\left(\boldsymbol{\mu}_{\widehat{\boldsymbol{\theta}}}\right)\boldsymbol{\Sigma}_{\widehat{\boldsymbol{\theta}}}J_h\left(\boldsymbol{\mu}_{\widehat{\boldsymbol{\theta}}}\right)^\top \overset{\text{eq. 4.2}}{=} \frac{1}{\mathbb{E}\left(\widehat{\lambda}\right)^2}\mathrm{Var}\left(\widehat{\beta_X^*}\right) + \frac{\mathbb{E}\left(\widehat{\beta_X^*}\right)^2}{\mathbb{E}\left(\widehat{\lambda}\right)^4}\mathrm{Var}\left(\widehat{\lambda}\right)$$

(II)

$$
\begin{aligned}
&\mathbb{E}\left[J_h\left(\boldsymbol{\mu}_{\widehat{\boldsymbol{\theta}}}\right)\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\mu}_{\widehat{\boldsymbol{\theta}}}\right)\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\mu}_{\widehat{\boldsymbol{\theta}}}\right)^{\top} H_h\left(\boldsymbol{\mu}_{\widehat{\boldsymbol{\theta}}}\right)\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\mu}_{\widehat{\boldsymbol{\theta}}}\right)\right] \\
=&\mathbb{E}\left\{-2\frac{1}{\mathbb{E}\left(\widehat{\lambda}\right)^3}\left(\widehat{\beta_X^*}-\mathbb{E}\left(\widehat{\beta_X^*}\right)\right)^2\left(\widehat{\lambda}-\mathbb{E}\left(\widehat{\lambda}\right)\right)+\right. \\
&\quad 4\frac{\mathbb{E}\left(\widehat{\beta_X^*}\right)}{\mathbb{E}\left(\widehat{\lambda}\right)^4}\left(\widehat{\beta_X^*}-\mathbb{E}\left(\widehat{\beta_X^*}\right)\right)\left(\widehat{\lambda}-\mathbb{E}\left(\widehat{\lambda}\right)\right)^2- \\
&\quad \left. 2\frac{\mathbb{E}\left(\widehat{\beta_X^*}\right)^2}{\mathbb{E}\left(\widehat{\lambda}\right)^5}\left(\widehat{\lambda}-\mathbb{E}\left(\widehat{\lambda}\right)\right)^3\right\} \\
=&-2\frac{\mathbb{E}\left(\widehat{\beta_X^*}\right)^2}{\mathbb{E}\left(\widehat{\lambda}\right)^5}\mathbb{E}\left(\widehat{\lambda}-\mathbb{E}\left(\widehat{\lambda}\right)\right)^3
\end{aligned}
\tag{4.3}
$$

(III)

$$
\begin{aligned}
&\mathbb{E}\left[\frac{1}{4}\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\mu}_{\widehat{\boldsymbol{\theta}}}\right)^{\top} H_h(\boldsymbol{\mu}_{\widehat{\boldsymbol{\theta}}})\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\mu}_{\widehat{\boldsymbol{\theta}}}\right)\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\mu}_{\widehat{\boldsymbol{\theta}}}\right)^{\top} H_h(\boldsymbol{\mu}_{\widehat{\boldsymbol{\theta}}})\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\mu}_{\widehat{\boldsymbol{\theta}}}\right)\right] \\
=&\frac{1}{4}\mathbb{E}\left[-2\frac{1}{\mathbb{E}\left(\widehat{\lambda}\right)^2}\left(\widehat{\beta_X^*}-\mathbb{E}\left(\widehat{\beta_X^*}\right)\right)\left(\widehat{\lambda}-\mathbb{E}\left(\widehat{\lambda}\right)\right)+2\frac{\mathbb{E}\left(\widehat{\beta_X^*}\right)}{\mathbb{E}\left(\widehat{\lambda}\right)^3}\left(\widehat{\lambda}-\mathbb{E}\left(\widehat{\lambda}\right)\right)^2\right]^2 \\
=&\mathbb{E}\left\{\frac{1}{\mathbb{E}\left(\widehat{\lambda}\right)^6}\left(\widehat{\lambda}-\mathbb{E}\left(\widehat{\lambda}\right)\right)^2\left[-\mathbb{E}\left(\widehat{\lambda}\right)\left(\widehat{\beta_X^*}-\mathbb{E}\left(\widehat{\beta_X^*}\right)\right)+\mathbb{E}\left(\widehat{\beta_X^*}\right)\left(\widehat{\lambda}-\mathbb{E}\left(\widehat{\lambda}\right)\right)\right]^2\right\} \\
=&\frac{1}{\mathbb{E}\left(\widehat{\lambda}\right)^4}\mathbb{E}\left[\left(\widehat{\beta_X^*}-\mathbb{E}\left(\widehat{\beta_X^*}\right)\right)^2\right]\mathbb{E}\left[\left(\widehat{\lambda}-\mathbb{E}\left(\widehat{\lambda}\right)\right)^2\right]+\frac{\mathbb{E}\left(\widehat{\beta_X^*}\right)^2}{\mathbb{E}\left(\widehat{\lambda}\right)^6}\mathbb{E}\left[\left(\widehat{\lambda}-\mathbb{E}\left(\widehat{\lambda}\right)\right)^4\right]
\end{aligned}
\tag{4.4}
$$

Eq. 4.4 is always positive or equals zero. Eq. 4.3 will equal zero if $\widehat{\lambda}$ exhibits a symmetric distribution. The distribution of $\widehat{\lambda}$ is symmetric for $n \to \infty$ as it is demonstrated in Subsection 4.1.2 and Subsection 4.1.3. For a finite sample size, a negatively skewed distribution of $\widehat{\lambda}$ assures the positiveness of eq. 4.3; $\widehat{\lambda}$ is negatively skewed if $\sigma_X^2 > \sigma_U^2$, which holds for the most practical applications

Assuming asymptotic normality for $\widehat{\beta_X}$, $\widehat{\beta_X} \pm z_{0.025} \cdot \sqrt{\mathrm{Var}(\widehat{\beta_X})}$ is an approximate 95% confidence interval; $z_{0.025}$ denotes the 2.5 % quantile of the Standard Normal distribution. However, CIs based on the Taylor series expansion of the variance may exhibit unbalanced coverage properties for small and moderate sample sizes (Lyles and Kupper, 1999).

4.2.2. Bootstrap percentile intervals

Bootstrap techniques provide confidence intervals with the advantage of relaxing the assumption of approximate normality of the estimator. For a longitudinal setting this can be realized by drawing $b, b = 1, \ldots, B$, samples with replacement from the independent individuals of the main study and of the independent units of the validation study resulting in B bootstrap replications of $\widehat{\lambda}$. Bootstrap percentile intervals can be derived through the appropriate percentiles of the B bootstrap replications.

4.3. Simulation study: Distributional properties of $\widehat{\lambda}$

The distributional properties of the attenuation factor $\widehat{\lambda}$ and the bias corrected estimate $\widehat{\beta_X}$ are further investigated with simulations.

4.3.1. Simulation setup

Exemplary, λ^{M} is examined for varying λ using N replication measurements for each time point and 100 simulation iterations. The simulation setup is chosen analogously to Subsection 3.6.1 with the following changes:

- N repeated classical error–prone measurements $X^{*\mathrm{C}}$ with varying missing value pattern $\boldsymbol{G}$ for each time point are generated in order to involve the variability of $\widehat{\lambda}$.
- $\rho^{X^{*\mathrm{B}}} = 0$, because the sample (least–squares) autocorrelation coefficient is biased in the case of high coefficients and a moderate sample size of 100 observations (Mudelsee, 2001) and hence, also $\widehat{\lambda^{\mathrm{M}}}$ would be biased.
- The size of the random classical measurement error $\sigma^2_{U^{\mathrm{C}}}$ is varied to obtain attenuation factors ranging over the complete domain of λ^{M}:

  $$\lambda^{\mathrm{M}} \in \{0.05, 0.10, 0.20, 0.30, 0.40, 0.50, 0.60, 0.70, 0.80, 0.90\}.$$

  The upper limit of λ^{M} cannot be chosen to be 1, because only $\sigma^2_{U^{\mathrm{C}}}$ is varying, whereas $\sigma^2_{U^{\mathrm{B}}}, \sigma^2_{\nu^{\mathrm{B}}}$ and $\sigma^2_{\nu^{\mathrm{C}}}$ are > 0 and remain fixed.

The components for the estimation of the attenuation factor are estimated as follows: $\sigma^2_{X^{*\mathrm{B}}}$ and $\rho^{X^{*\mathrm{B}}}$ are directly calculated as sample variance and sample autocorrelation coefficient of the realizations of $X^{*\mathrm{B}}$. Estimations for the true values of X are approximated through averaging over the repeated realizations of $X^{*\mathrm{C}}$. Information about the error models are gained regressing $\widehat{X}$ on $X^{*\mathrm{B}}$ and $X^{*\mathrm{C}}$ on $\widehat{X}$.

4.3.2. Results

The simulations with $\lambda^{\mathrm{M}} = 0.6$ in Figure 4.4 show, that the bias of the estimated attenuation factor decreases with increasing number of repetitions:

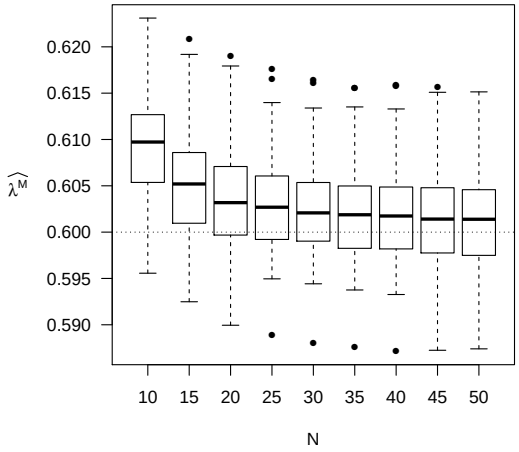


FIGURE 4.4.: *Dependency of* $\widehat{\lambda^{M}}$ *on the number of repeated measurements* N*;* $\lambda^{M} = 0.6$ *(dotted line).*

Figure 4.5 depicts the results of the simulation for $N = 100$ repeated measurements. The variability of $\widehat{\lambda^{\mathrm{M}}}$ is higher when the classical measurement error is small. The distribution of $\widehat{\lambda^{\mathrm{M}}}$ is symmetrical, also at the boundaries of its domain, as it is theoretically derived in Subsection 4.1.3.

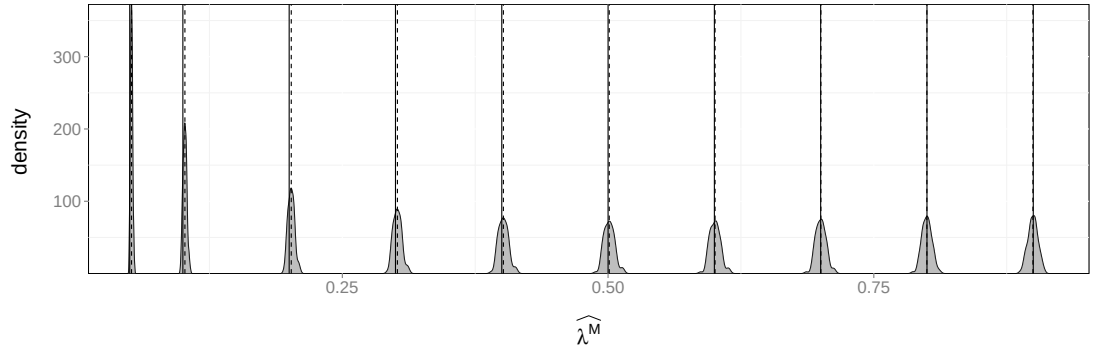


FIGURE 4.5.: *Empirical density estimation of* $\widehat{\lambda^{M}}$ *based on 100 simulation iterations and* $N = 100$ *repeated measurements; dashed lines: mean* $\widehat{\lambda^{M}}$*; solid lines: predefined values of* λ^{M}*.*

4.4. Simulation study: Comparison of confidence intervals

4.4.1. Simulation setup

Three types of confidence intervals (CIs) for the corrected effects $\widehat{\beta_X^*}$ are compared regarding their empirical coverage probabilities and their interval lengths:

1. "Naive": CIs assuming normality of $\widehat{\beta_X}$ are calculated with the variance of the naive estimate, $\widehat{\beta_X^*}$:

$$\widehat{\beta_X} \pm 1.96\sqrt{\text{Var}(\widehat{\beta_X^*})}.$$

2. "Delta method": CIs assuming normality of $\widehat{\beta_X}$ are calculated with an approximation of the variance using the delta method (eq. 4.2):

$$\widehat{\beta_X} \pm 1.96\sqrt{\text{Var}(\widehat{\beta_X})}.$$

3. "Bootstrap": Bootstrap percentile intervals using 100 bootstrap iterations are calculated as described in Subsection 4.2.2.

The comparison is based on 1000 CIs. Simulation data was generated as described in Subsection 3.6.1 for given values of λ: $\lambda^{\text{C}} \in \{0.05, 0.2, 0.4, 0.6, 0.8, 0.95\}$ and $\lambda^{\text{M}} \in \{0.05, 0.2, 0.4, 0.6\}$. λ^{C} and λ^{M} are generated through varying sizes of classical measurement error $\sigma^2_{U^{\text{C}}}$; the other components of λ^{C} and λ^{M} are fixed according to Table 3.1 and are assumed to be known.

Variability of $\widehat{\lambda}$ was generated through the usage of samples from a Beta distribution, as described in Subsection 4.1.1, around the true, but in reality unknown, λ, instead of $\widehat{\lambda}$. The variance of λ is chosen in the order of the variance of 100 Bootstrap replications of $\widehat{\lambda}$ in the application to the Augsburger Umweltstudie as 0.001.

4.4.2. Results

Table 4.1 presents the simulation results. Indeed, the naive estimations of the confidence intervals exhibit the most narrow intervals, but their coverage probability is in nearly all scenarios beyond 95 %. Only for strong attenuation, bootstrap percentile intervals provide inadequate coverage probabilities. In comparison to confidence intervals based on the normality assumption of the estimator and on variance calculations using the delta method, bootstrap confidence intervals are slightly narrower. The coverage probabilities of the delta method CIs are also for large measurement errors high. Uncertainty in the estimation of λ is adequately incorporated in both CIs, the CIs based on the delta method and the bootstrap percentile intervals.

Theoretical considerations for the case without autocorrelation in Subsection 4.2.1 lead to the conclusion that the variances of the corrected estimators are underestimated by the delta method since approximating $\text{Var}(\widehat{\beta_X})$ with the second order Taylor series yields larger values for $\sigma^2_X > \sigma^2_U$ compared to the delta method. The simulations for the more complex situation with autocorrelation in this Subsection show the opposite: confidence intervals based on the delta method and also on the bootstrap method tend to be too conservative, especially for small measurement error.

			Empirical CP			**Mean interval width**		
Error type	$\mathrm{Var}(\widehat{\lambda})$	λ	Naive	Bootstrap	Delta method	Naive	Bootstrap	Delta method
Classical error	0	0.050	0.085	0.920	0.941	0.205	3.899	4.099
		0.200	0.329	**0.954**	**0.970**	0.410	1.949	2.049
		0.400	0.634	**0.955**	**0.979**	0.579	1.377	1.448
		0.600	0.895	**0.987**	**0.992**	0.709	1.126	1.182
		0.800	**0.988**	**0.995**	**0.999**	0.818	0.975	1.023
		0.950	**1.000**	**0.998**	**1.000**	0.892	0.895	0.938
	0.0004	0.200	0.344	0.933	**0.968**	0.410	1.998	2.119
		0.400	0.642	**0.950**	**0.980**	0.579	1.382	1.464
		0.600	0.882	**0.986**	**0.994**	0.709	1.129	1.192
		0.800	**0.986**	**0.992**	**0.998**	0.818	0.977	1.030
Mixture error	0	0.050	0.098	0.928	**0.953**	0.203	3.872	4.062
		0.200	0.343	**0.958**	**0.974**	0.406	1.929	2.031
		0.400	0.683	**0.968**	**0.988**	0.574	1.361	1.436
		0.600	0.891	**0.988**	**0.994**	0.703	1.118	1.172
	0.0004	0.200	0.343	**0.950**	**0.974**	0.406	1.968	2.091
		0.400	0.656	**0.969**	**0.987**	0.574	1.368	1.456
		0.600	0.890	**0.986**	**0.994**	0.703	1.120	1.181

TABLE 4.1.: *Comparison of naive CIs, bootstrap percentile intervals and CIs based on the delta method regarding their empirical coverage probability (CP) and interval width; bold: empirical coverage probabilities greater than or equal to 0.95.*

5. The contribution of prior knowledge to Bayesian measurement error correction

5.1. Bayesian measurement error models

5.1.1. Background

Bayesian methods for measurement error adjustment allow the direct inclusion of knowledge and the underlying uncertainty regarding the size of the measurement error in the analysis through prior distributions in contrast to frequentist methods. Consequently, the estimated parameters and their variability are affected by the choice of the prior distributions.

Bayesian models have been successfully applied in the analysis of the effects of air pollution for a long time, see e.g. Best et al. (2000) as an example for Bayesian models in general and e.g. Chang et al. (2011), Gryparis et al. (2009) and Lee and Shaddick (2010) for models with measurement error. Moreover, Bayesian methods are used to correct measurement errors or misclassification in case–control studies (e.g. Espino-Hernandez et al., 2011; Liu et al., 2009; Prescott and Garthwaite, 2005; Rice, 2003). A further useful feature of Bayesian measurement error modeling is the possibility of simultaneously impute missing values as it is conducted e.g. by Chang et al. (2011).

Markov chain Monte Carlo (MCMC) sampling is widely used for the implementation of Bayesian measurement error models. We use JAGS (Plummer, 2003), a software for MCMC sampling using Gibbs sampling (Geman and Geman, 1984), for our analyses. Muff et al. (2015) proposed to use INLA (integrated nested Laplace approximation) as an alternative to MCMC sampling for Bayesian measurement error models.

Since linear regression models with a deficient covariate are known to lack identifiability, prior knowledge is necessary to overcome this problem within the Bayesian analysis of the data. The choice of the prior distribution for the unknown parameters describing the knowledge about the measurement errors is essential to get valid results. Additionally, the informativeness of prior knowledge influences the resulting effect estimations, as it is described by Gustafson (2005): even apparently non–informative prior distributions may affect the posterior distributions of the parameters describing the measurement errors and thus also the posterior distributions of the parameters of interest by indirect learning. These findings indicate that the direct usage of an exposure model as prior distribution in the health outcome model, as conducted in Chang et al. (2011), Gryparis et al. (2009) and Prescott and Garthwaite (2005), should be scrutinized.

We concentrate our considerations about Bayesian measurement error models on situations where prior knowledge about the measurement error is only available through expert knowledge or external validation data and not through repeated or simultaneous measurements within the main study. In these cases, the choice of proper prior distributions guarantees identifiability (Dellaportas and Stephens, 1995).

This chapter intends to examine the relation between the specification of prior knowledge and the Bayesian measurement error model with regard to varying uncertainty about the prior knowledge. More precisely, it will be evaluated 1) how much prior information is necessary to achieve adequate estimations, 2) if it is reasonable to prefer the simple

frequentist method of moments correction in situations with only vague prior information, 3a) whether the simultaneous inclusion of fixed–site and mobile measurements in a single model provides benefits and 3b) to what extent the specification of the relationship between fixed–site and personal measurements influences the results. Besides these questions we want to investigate the role of missing values and the presence of further covariates, which are correlated with the deficient covariate as it is conducted for frequentist models in Chapter 3. The findings will be applied to the Augsburger Umweltstudie in Section 6.2.

5.1.2. Model formulation and properties

In comparison to the frequentist procedure considered in Chapter 4, the Bayesian measurement error model formulation allows to incorporate personal measurements, exposure maps or ambient measurements as well as prior information about the measurement error into a single model. The knowledge about covariate measurement error is directly integrated into the Bayesian regression analysis.

According to Carroll et al. (2006) and Richardson and Gilks (1993), who give an introduction in Bayesian measurement error models, the usual Bayesian model formulation for a regression problem has to be extended through the specification of an exposure model and, in the case of classical measurement error, additionally through the specification of a measurement model (assuming non–differential measurement error):

Health model:	$f(\boldsymbol{Y}\|\boldsymbol{X},\boldsymbol{Z},\boldsymbol{\theta}^H)$
Measurement model:	$f(\boldsymbol{X}^{*\mathrm{C}}\|\boldsymbol{X},\boldsymbol{Z},\boldsymbol{\theta}^M)$
Exposure model:	$f(\boldsymbol{X}\|\boldsymbol{Z},\boldsymbol{\theta}^{E_C})$ or $f(\boldsymbol{X}\|\boldsymbol{X}^{*\mathrm{B}},\boldsymbol{Z},\boldsymbol{\theta}^{E_B})$

with additional covariates $\boldsymbol{Z}$. $\boldsymbol{\Theta}^{\mathrm{B}} = \{\boldsymbol{\theta}^H,\boldsymbol{\theta}^{E_B},\boldsymbol{\Omega}^{\mathrm{B}}\}, \boldsymbol{\Theta}^{\mathrm{C}} = \{\boldsymbol{\theta}^H,\boldsymbol{\theta}^M,\boldsymbol{\theta}^{E_C},\boldsymbol{\Omega}^{\mathrm{C}}\}$ and $\boldsymbol{\Theta}^{\mathrm{M}} = \{\boldsymbol{\theta}^H,\boldsymbol{\theta}^M,\boldsymbol{\theta}^{E_B},\boldsymbol{\Omega}^{\mathrm{M}}\}$ denote the vectors of the unknown parameters in the models with only fixed–site exposure measurements, $\boldsymbol{X}^{*\mathrm{B}}$, with only personal exposure measurements, $\boldsymbol{X}^{*\mathrm{C}}$, and with both types of exposure measurements ($\boldsymbol{X}^{*\mathrm{B}}$ and $\boldsymbol{X}^{*\mathrm{C}}$), respectively with additional hyper–parameters $\boldsymbol{\Omega}^{\mathrm{B}},\boldsymbol{\Omega}^{\mathrm{C}}$ and $\boldsymbol{\Omega}^{\mathrm{M}}$. The deficient measurements $\boldsymbol{X}^*$ are related to the true values $\boldsymbol{X}$. The inclusion of a prior distribution on $\boldsymbol{X}$ or an exposure model avoids the calculation of the expectation regarding $\boldsymbol{X}$ (two–stage Bayesian approach) (Richardson, 1996). We concentrate the Bayesian analysis on two–stage modeling approaches, where the health model and the exposure model are separately defined and connected through the latent exposure; this strategy is preferred to a fully Bayesian model, because the design of a well–mixing MCMC algorithm is easier in the case of complex exposure models (Gryparis et al., 2009) and the full conditional distributions of the parameters have a known form when conditionally conjugate priors are chosen.

Since $\boldsymbol{X}$ is unknown, as other model parameters like the effect coefficients, a conditional posterior distribution for $\boldsymbol{X}$ is estimated within the MCMC algorithm. The posterior of $\boldsymbol{X}$ depends among other parameters also on the outcome $\boldsymbol{Y}$, as it is shown in detail in Section 5.1.3 and also in Carroll et al. (2006) for the linear regression model.

The posterior distribution is given by

$$f(\boldsymbol{\Theta}^{\mathrm{C}},\boldsymbol{X}|\boldsymbol{Y},\boldsymbol{X}^{*\mathrm{C}},\boldsymbol{Z}) \propto f(\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{Z},\boldsymbol{\theta}^H)f(\boldsymbol{X}^{*\mathrm{C}}|\boldsymbol{X},\boldsymbol{\theta}^M)f(\boldsymbol{X}|\boldsymbol{Z},\boldsymbol{\theta}^{E_C})f(\boldsymbol{\Theta}^{\mathrm{C}}) \quad (5.1)$$

in the case of classical measurement error and by

$$f(\boldsymbol{\Theta}^{\mathrm{B}},\boldsymbol{X}|\boldsymbol{Y},\boldsymbol{X}^{*\mathrm{B}},\boldsymbol{Z}) \propto f(\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{Z},\boldsymbol{\theta}^H)f(\boldsymbol{X}|\boldsymbol{X}^{*\mathrm{B}},\boldsymbol{Z},\boldsymbol{\theta}^{E_B})f(\boldsymbol{\Theta}^{\mathrm{B}}) \quad (5.2)$$

in the case of Berkson error. If both data sources are combined in one model, the posterior distribution becomes:

$$f(\boldsymbol{\Theta}^{\mathrm{M}}, \boldsymbol{X}|\boldsymbol{Y}, \boldsymbol{X}^{*\mathrm{C}}, \boldsymbol{X}^{*\mathrm{B}}, \boldsymbol{Z}) \propto f(\boldsymbol{Y}|\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\theta}^{H}) f(\boldsymbol{X}^{*\mathrm{C}}|\boldsymbol{X}, \boldsymbol{\theta}^{M}) f(\boldsymbol{X}|\boldsymbol{X}^{*\mathrm{B}}, \boldsymbol{Z}, \boldsymbol{\theta}^{E_B}) f(\boldsymbol{\Theta}^{\mathrm{M}}) \,. \tag{5.3}$$

$f(\boldsymbol{\Theta})$ denotes the prior distribution of $\boldsymbol{\Theta}$, $\boldsymbol{\Theta} \in \{\boldsymbol{\Theta}^{\mathrm{B}}, \boldsymbol{\Theta}^{\mathrm{C}}, \boldsymbol{\Theta}^{\mathrm{M}}\}$.

The Bayesian model for measurement error lacks identifiability. Gustafson (2005) proposes three strategies to handle non–identifiability: 1) Fix unknown parameters at the "best guess", 2) expand the model adequately and 3) specify prior distributions for the unknown parameters and ignore the non–identifiability. Although model expansion yields identifiability, the examples in Gustafson (2005) did not show any superiority of model expansion in comparison to the specification of prior distributions. For this reason and due to lacking justifications for expanding the model, we concentrate on strategies 1 and 3.

5.1.3. Conditionally conjugate priors: role of measurement error characteristics

The theoretical considerations in this subsection are focused on the usage of both data sources, $\boldsymbol{X}^{*\mathrm{B}}$ and $\boldsymbol{X}^{*\mathrm{C}}$, the attendant posterior distribution eq. 5.3 and on the role of classical measurement error in the conditionally conjugate case.

Assuming model 2.1 without confounding variables $\boldsymbol{Z}$ and with classical and Berkson errors as defined in eq. 2.4 and eq. 2.5, the full conditional distributions of the model for both data sources and conditionally conjugate prior distributions are given in Appendix D. Since prior parameters of a single parameter only appear in the full conditional distribution of this parameter and not in the full conditional distribution of other parameters, prior knowledge does not have any direct influence on the estimation of other parameters.

The full conditional distribution of the classical measurement error variance is

$$\begin{aligned}\sigma^2_{U^{\mathrm{C}}}|\cdot \sim \mathrm{IG}\,(&\delta_{U^{\mathrm{C}},1} + \frac{nT}{2}, \\ &\delta_{U^{\mathrm{C}},2} + \frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{X}_i^{*\mathrm{C}} - \boldsymbol{X}_i - \nu_i^{\mathrm{C}}\mathbf{1}_T)^\top \boldsymbol{W}_{\rho^{\mathrm{C}}}^{-1}(\boldsymbol{X}_i^{*\mathrm{C}} - \boldsymbol{X}_i - \nu_i^{\mathrm{C}}\mathbf{1}_T)\Big)\,.\end{aligned} \tag{5.4}$$

The posterior mean of $\sigma^2_{U^{\mathrm{C}}}$ results as a mean between the prior expectation and the estimate from the data. The latter strongly depends on $\boldsymbol{X}$. If the estimation of $\boldsymbol{X}$ solely depends on $\sigma^2_{U^{\mathrm{C}}}$, the prior expectation of $\sigma^2_{U^{\mathrm{C}}}$ will equal the estimate based on the data.

The full conditional of $\boldsymbol{X}_i$ is a normal distribution:

$$\begin{aligned}\boldsymbol{X}_i|\cdot \sim \mathrm{N}\Bigg(&\left[\beta_X^2\boldsymbol{\Sigma}_\varepsilon^{-1} + \boldsymbol{\Sigma}_{U^{\mathrm{C}}}^{-1} + \boldsymbol{\Sigma}_{U^{\mathrm{B}}}^{-1}\right]^{-1} \\ &\left[\beta_X\boldsymbol{\Sigma}_\varepsilon^{-1}(\boldsymbol{Y}_i - \beta_0\mathbf{1} - \tau_i\mathbf{1}_T) + \boldsymbol{\Sigma}_{U^{\mathrm{C}}}^{-1}(\boldsymbol{X}_i^{*\mathrm{C}} - \nu_i^{\mathrm{C}}\mathbf{1}_T) + \boldsymbol{\Sigma}_{U^{\mathrm{B}}}^{-1}(\boldsymbol{X}_i^{*\mathrm{B}} + \nu_i^{\mathrm{B}}\mathbf{1}_T)\right], \\ &\left[\beta_X^2\boldsymbol{\Sigma}_\varepsilon^{-1} + \boldsymbol{\Sigma}_{U^{\mathrm{C}}}^{-1} + \boldsymbol{\Sigma}_{U^{\mathrm{B}}}^{-1}\right]^{-1}\Bigg)\end{aligned} \tag{5.5}$$

The expectation of this distribution is a weighted mean of $(\boldsymbol{Y}_i - \beta_0\mathbf{1}_T - \tau_i\mathbf{1}_T)/\beta_X$, $\boldsymbol{X}_i^{*\mathrm{C}} - \nu_i^{\mathrm{C}}\mathbf{1}_T$ and $\boldsymbol{X}_i^{*\mathrm{B}} + \nu_i^{\mathrm{B}}\mathbf{1}$ with the weights $\beta_X^2\boldsymbol{\Sigma}_\varepsilon^{-1}$, $\boldsymbol{\Sigma}_{U^{\mathrm{C}}}^{-1}$ and $\boldsymbol{\Sigma}_{U^{\mathrm{B}}}^{-1}$ and the random effects play

a major role in the estimation of $\boldsymbol{X}_i$. The variance of the distribution is given by the inverse of the sum of the weights.

Since

$$\begin{aligned}\mathrm{Var}(\boldsymbol{X}_i|\boldsymbol{Y}_i, \boldsymbol{X}_i^{*\mathrm{C}}, \boldsymbol{X}_i^{*\mathrm{B}}) =& \mathrm{Var}(\mathbb{E}(\boldsymbol{X}_i|\boldsymbol{Y}_i, \boldsymbol{X}_i^{*\mathrm{C}}, \boldsymbol{X}_i^{*\mathrm{B}}, \boldsymbol{\Theta}^{\mathrm{M}}))+ \\ & \mathbb{E}(\mathrm{Var}(\boldsymbol{X}_i|\boldsymbol{Y}_i, \boldsymbol{X}_i^{*\mathrm{C}}, \boldsymbol{X}_i^{*\mathrm{B}}, \boldsymbol{\Theta}^{\mathrm{M}})),\end{aligned} \tag{5.6}$$

the marginal posterior variance of $\boldsymbol{X}_i$, $\mathrm{Var}(\boldsymbol{X}_i|\boldsymbol{Y}_i, \boldsymbol{X}_i^{*\mathrm{C}}, \boldsymbol{X}_i^{*\mathrm{B}})$, exceeds or equals the expected variance of the full conditional distribution of $\boldsymbol{X}_i|\cdot$.

The full conditional distribution for $\boldsymbol{\beta}$ is directly related to the latent data $\boldsymbol{X}$:

$$\begin{aligned}\boldsymbol{\beta}|\cdot \sim \mathrm{N}\Bigg(& \left[\sum_{i=1}^{n} \boldsymbol{\chi}_i^\top \boldsymbol{W}_\rho^{-1} \boldsymbol{\chi}_i + \begin{pmatrix} \sigma_\varepsilon^2/\sigma_{\beta_0}^2 & 0 \\ 0 & \sigma_\varepsilon^2/\sigma_{\beta_X}^2 \end{pmatrix}\right]^{-1} \cdot \\ & \left[\sum_{i=1}^{n} (\boldsymbol{Y}_i - \tau_i \mathbf{1}_T)^\top \boldsymbol{W}_\rho^{-1} \boldsymbol{\chi}_i + \boldsymbol{\beta}^{(0)\top} \begin{pmatrix} \sigma_\varepsilon^2/\sigma_{\beta_0}^2 & 0 \\ 0 & \sigma_\varepsilon^2/\sigma_{\beta_X}^2 \end{pmatrix}\right], \\ & \sigma_\varepsilon^2 \left[\sum_{i=1}^{n} \boldsymbol{\chi}_i^\top \boldsymbol{W}_\rho^{-1} \boldsymbol{\chi}_i + \begin{pmatrix} \sigma_\varepsilon^2/\sigma_{\beta_0}^2 & 0 \\ 0 & \sigma_\varepsilon^2/\sigma_{\beta_X}^2 \end{pmatrix}\right]^{-1}\Bigg)\end{aligned} \tag{5.7}$$

The prior mean and variances of $\boldsymbol{\beta}$ are denoted with $\boldsymbol{\beta}^{(0)}$, $\sigma_{\beta_0}^2$ and $\sigma_{\beta_X}^2$; $\boldsymbol{\chi}_i^* = (\mathbf{1}_T, \boldsymbol{X}_i^*)$ as defined in Subsection 3.1.1.

Prior knowledge about the size of the measurement error only indirectly affects the full conditional distribution of $\boldsymbol{\beta}$ through the dependency of other unknown parameters with direct influence on the full conditional distribution of $\boldsymbol{\beta}$. The influence of prior knowledge about $\boldsymbol{\beta}$ is regulated by the ratio between σ_ε^2 and the prior variances of $\boldsymbol{\beta}$, $\sigma_{\beta_0}^2$ and $\sigma_{\beta_X}^2$.

If the prior distribution for β_X is chosen to be non–informative, i.e. $\sigma_{\beta_X}^2$ is high, the full conditional distribution of β_X will approximately equal the well–known normal distribution of the usual regression analysis with the same parameters as if $\boldsymbol{X}$ is known. In this case, the only impact of the prior knowledge about the measurement error on the estimation of the effect coefficient is through the estimation of $\boldsymbol{X}$.

If $\sigma_{U^{\mathrm{C}}}^2$ is substantially smaller than σ_ε^2 and $\sigma_{U^{\mathrm{B}}}^2$, $\sigma_{U^{\mathrm{C}}}^2$ will be the only unknown component, which bears information for $\boldsymbol{X}_i$ and the full conditional of $\boldsymbol{X}_i$ approximates a normal distribution with mean $\boldsymbol{X}^{*\mathrm{C}}$ and variance $\sigma_{U^{\mathrm{C}}}^2$. In this case, the data $(\boldsymbol{Y}, \boldsymbol{X}^{*\mathrm{C}}, \boldsymbol{X}^{*\mathrm{B}})$ only marginally introduces new knowledge to the full conditional distributions of $\sigma_{U^{\mathrm{C}}}^2$ and $\boldsymbol{X}_i$ involving the following consequences for expectations:

1. The expectation of the full conditional and of the marginal posterior distribution of $\sigma_{U^{\mathrm{C}}}^2$ equal the prior expectation denoted by $\sigma_{U^{\mathrm{C}}}^{2\,(0)}$.

2. The expectation of the full conditional and of the marginal posterior distribution of $\boldsymbol{X}_i$ equals $\boldsymbol{X}_i^{*\mathrm{C}}$.

3. The expectation of the full conditional and of the marginal posterior distribution of β_X depends only on the prior expectation of the size of the classical measurement error $\sigma_{U^{\mathrm{C}}}^{2\,(0)}$, and not on its variability.

Even precise prior knowledge about the measurement error results in a higher variance of the effect coefficient than in a model with the true data. The reason is that the

following holds for the marginal posterior distribution of β_X provided that the prior on β_X is non–informative ($\rho = 0$ and $\boldsymbol{\tau} = \mathbf{0}$ for simplicity):

$$\begin{aligned}\mathrm{Var}(\beta_X|\boldsymbol{Y}, \boldsymbol{X}^{*\mathrm{C}}, \boldsymbol{X}^{*\mathrm{B}}) =& \mathbb{E}(\mathrm{Var}(\beta_X|\boldsymbol{Y}, \boldsymbol{X}^{*\mathrm{C}}, \boldsymbol{X}^{*\mathrm{B}}, \boldsymbol{\Theta}_{-\beta_X}))+ \\ & \mathrm{Var}(\mathbb{E}(\beta_X|\boldsymbol{Y}, \boldsymbol{X}^{*\mathrm{C}}, \boldsymbol{X}^{*\mathrm{B}}, \boldsymbol{\Theta}_{-\beta_X})) \\ \approx& \mathbb{E}\left(\sigma_\varepsilon^2 \left(\sum_{i=1}^{n} \boldsymbol{X}_i^\top \boldsymbol{X}_i\right)^{-1}\right) + \mathrm{Var}\left(\sum_{i=1}^{n} (\boldsymbol{X}_i^\top \boldsymbol{X}_i)^{-1}(\boldsymbol{X}_i^\top \boldsymbol{Y}_i)\right) \quad (5.8)\end{aligned}$$

Only the second term of eq. 5.8 depends on the variability of $\boldsymbol{X}$ and thus on the uncertainty regarding the prior knowledge about the measurement error. The higher the variability of $\boldsymbol{X}$, the higher is the variability of the effect estimate.

The usage of an informative prior for $\boldsymbol{\beta}$ results in a conditional posterior expectation, which is the weighted mean between the conditional posterior expectation for the non–informative prior and the prior expectation.

The impact of prior knowledge on full conditional distributions of the variance parameters is only local in the first instance. But samples from these distributions are directly incorporated in the full conditional distribution of the latent variable $\boldsymbol{X}$, which affects the calculation of the effect β_X. The shape of the prior distribution for the size of the measurement error affects the conditional posterior distribution of β_X, but the expectation and the variance of the conditional posterior distribution only depend on prior expectation and variance regarding the size of the measurement error. Therefore, the adequate choice of prior distributions for the model parameters is essential to prevent undesired influences on the estimation of the relevant parameters.

5.2. Informativeness of prior distributions

5.2.1. Impact of the prior distribution in Bayesian models

This subsection aims to illustrate the impact of different amounts and shapes of prior knowledge on posterior distributions including general considerations, priors for variance and correlation parameters and the transferability of the concepts to Bayesian regression models for measurement error–prone covariates. Therefore, two meanings of the term "informativeness" regarding the prior knowledge of a parameter have to be differentiated. High information is generally understood as precise and correct knowledge with low bias and variance. This perception differs from the Bayesian idea of informativeness where a prior distribution, which affects the estimation of the parameters of interest to a relevant extent, is called informative. Informative priors in the Bayesian sense exhibit not necessarily low bias and variance.

Non–informative prior distributions are chosen if prior knowledge about a parameter is not available and only the information from the data should be involved for parameter estimation. We use in the following the term "non–informative" prior distribution for a prior distribution, which only has a negligible impact on the posterior distributions of the relevant parameters. Other commonly used terms in this context are "weakly informative", "vague", "flat" or "diffuse" prior distribution.

The choice of adequate prior distributions is crucial in Bayesian analyses. Gelman (2006) and Kass and Wasserman (1996) summarize criteria for an appropriate choice of non–informative prior distributions. A selection of the commonly used criteria is:

- Principle of insufficient reason (Laplace (1820), as cited in Kass and Wasserman (1996)): Applying this principle for a finite parameter space results in Uniform priors.

- Conditionally conjugate priors with appropriate parameters: These priors allow to interpret the posterior distribution as an updated version of the prior distribution within the same class of distributions.
- Invariance concerning bijective transformations: e.g. Jeffreys' prior $p(\theta) \propto \det(I(\theta))^{\frac{1}{2}}$
- Maximum entropy: The uncertainty of a prior distribution is measured with its entropy. This approach was mainly developed by Jaynes (e.g. Jaynes (1982)).
- Agreement with classical estimators (Box and Tiao, 1973; Meng and Zaslavsky, 2002)
- Reference priors: The reference prior is that prior in a class of admissible, proper priors, which maximizes the expected information regarding the data and the reference posterior (Bernardo, 1979).

Potential difficulties arise in the search of a non–informative prior distribution (Gelman, 2006): improper posterior distributions may occur, parameter transformation may change the degree of non–informativeness and non–informativeness depends on the data. Moreover, apparently non–informative prior distribution may have a relevant impact on the posterior inference (Spiegelhalter, 2001) and non–informativeness may involve unrealistic parameter values (Greenland, 1998).

The posterior distribution results from the product of the distribution of the data, the likelihood, and the prior distribution. The impact of likelihood and prior distribution on the posterior distribution depends on the number of observations, the shape of the prior distribution and the relative position of the probability masses of the two distributions. For instance, if the observed data is a priori known to be implausible, i.e. the data has a low probability to be observed under the prior distribution, an apparently flat prior distribution will be informative in the Bayesian sense.

Considering the two situations depicted in Figure 5.1a) and b): the likelihood is equal for both situations and the prior distribution has the same shape. However, the prior distribution will be informative for the situation depicted in the Figure 5.1a) and non–informative for the situation in Figure 5.1b) as it is seen with the corresponding posterior distributions (Figure 5.1c) and Figure 5.1d)).

If the probability masses of prior distribution and likelihood are in the same scope, the prior distribution has to exhibit a higher density in order to add information than in a situation with different scopes.

In general, the posterior distribution of a parameter can be considered as likelihood which is weighted by the prior distribution; e.g., using a Uniform prior distribution on the parameter results in a posterior distribution, which is proportional to the likelihood restricted to the domain where the prior distribution is positive. The posterior distribution resulting from a Uniform, improper prior with infinite limits is equal to the likelihood. Other prior densities change the shape of the likelihood through weighting due to non–Uniform weights and yield a mixture between data and prior knowledge. Informative prior distributions weight certain parameter areas of the observed likelihood stronger than others. Therefore, a non–informative prior distribution is defined as a prior distribution, which has no bearing on the posterior distribution of the parameters of interest. The distance between the likelihood and the posterior distribution may be an indicator for the informativeness of the prior distribution, e.g. the Kullback–Leibler distance can be used.

Thus, a rule of thumb for the choice of informative or non–informative prior distributions cannot be given; instead, the prior distribution has to be chosen with regard to the likelihood, especially when a non–informative prior is required. Indeed, the prior

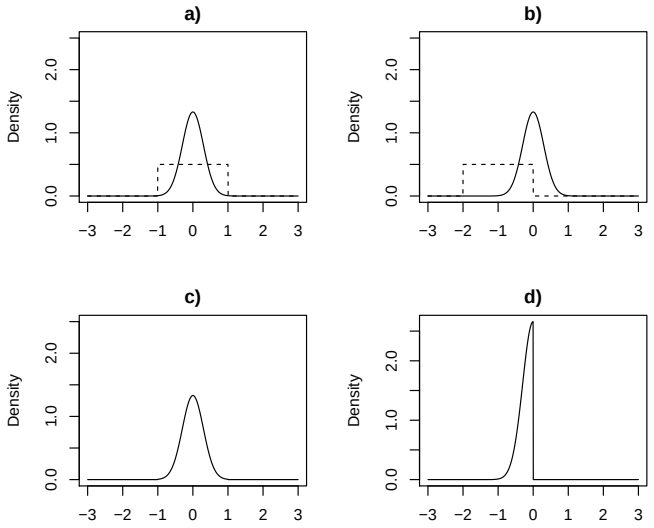


FIGURE 5.1.: *a) and b): Exemplary likelihood (solid) and prior distributions (dashed); c) and d): corresponding posterior distributions.*

distribution is characterized by concrete quantities, like expectation or variance, even if the prior distribution is flat, but these measures do not have any direct impact on the informativeness of the prior distribution.

Bayesian measurement error models rely on the informativeness of the prior distributions for the measurement error. Simple univariate examples are used in the following to illustrate the role and the impact of commonly used possibilities to specify prior knowledge. Prior distributions for the parameters describing the measurement errors, i.e. prior distributions on error variances and correlation coefficients, are most important in the context of the modeling strategy and the measurement error structure of the Augsburger Umweltstudie; therefore, the further discussions will be focused on these parameters.

5.2.2. Prior distributions for the variance

Common variance priors

Many distributions are suggested in the literature for non–informative prior distributions of variances (Table 5.1). The restricted parameter space and commonly used transformations of the variance provide a variety of possibilities to specify prior knowledge.

The Inverse Gamma distribution is the conjugate prior for modeling normally distributed data with known mean and unknown variance. Within this context, the Right–invariant Haar density ($p(\sigma) \propto 1/\sigma$) and the Location–scale prior ($p(\sigma^2) \propto 1/\sigma^2$) are Jeffreys' priors and represent the limiting distribution of the Inverse Gamma distribution with parameters δ_1 and δ_2 approximating 0. Further variance priors are the Beta, the (skewed) Truncated Normal and the Rayleigh distribution.

Informativeness of variance priors in a single–parameter model for normally distributed data

In the following, these suggestions will be investigated exemplary for normally distributed data with known mean and unknown variance regarding their properties with varying prior parameters: $x_1, \ldots, x_n | \mu, \sigma^2 \overset{\text{i.i.d.}}{\sim} \mathrm{N}(\mu, \sigma^2)$.

Distribution	Parameter	References
Inverse Gamma	σ^2	Gelman (2006); Lambert et al. (2005); Spiegelhalter (2001)
Uniform	$\log \sigma$	Gelman (2006)
Uniform	$\log \sigma^2$	Lambert et al. (2005); Spiegelhalter (2001)
Uniform	σ	Gelman (2006); Lambert et al. (2005)
Uniform	σ^2	Lambert et al. (2005); Spiegelhalter (2001)
Folded–noncentral–t	σ	Gelman (2006)
Location–scale	$\frac{1}{\sigma^2}$	Daniels (1999)
Right–invariant Haar density	$\frac{1}{\sigma}$	Daniels (1999)
Pareto	$\frac{1}{\sigma^2}$	Lambert et al. (2005)

TABLE 5.1.: *Commonly used distributions for the construction of a non–informative distribution for a variance parameter.*

The type of the prior distribution is only of secondary relevance for the search of a non–informative prior distribution as long as the parameters of the distribution are appropriately chosen. The situation is different when informative prior distributions are considered as we demonstrate in the next paragraphs.

The informativeness of the prior distribution depends amongst others on the likelihood. For the choice of an adequate prior, profound knowledge about the impact of the data and the prior distribution on the posterior distribution is necessary. To this end, a "confidence ratio" is defined for $\tilde{s}^2 \neq \sigma_0^2$ to quantify the confidence regarding the prior knowledge independently from the prior distribution, which relates the absolute difference between the posterior expectation $\mathbb{E}(\sigma^2|x)$ and the prior expectation σ_0^2 for σ^2 to the absolute difference between the frequentist estimation $\tilde{s}^2$, i.e. the posterior expectation without prior knowledge, and the prior expectation σ_0^2:

$$CR = 1 - \frac{|\mathbb{E}(\sigma^2|x) - \sigma_0^2|}{|\tilde{s}^2 - \sigma_0^2|} \tag{5.9}$$

Note, CR is only used for illustrating theoretical relationships and is not intended as a general measure for the quantification of the informativeness of prior knowledge since $\tilde{s}^2 \neq \sigma_0^2$ has to be assumed and prior distributions may be supposable, which result in $CR < 0$.

Using the Inverse Gamma distribution as conjugate prior distribution,

$$\sigma^2 \sim \text{IG}\left(\frac{\nu_0}{2}, \frac{\nu_0 - 2}{2}\sigma_0^2\right),$$

which is equivalent to the Scaled–Inverse–χ^2 distribution, $\sigma^2 \sim \text{Inv-}\chi^2(\nu_0, \frac{\nu_0-2}{\nu_0}\sigma_0^2)$, yields the following Inverse Gamma posterior distribution:

$$\sigma^2|x \sim \text{IG}\left(\delta_1 + \frac{n}{2}, \delta_2 + \frac{n}{2}\tilde{s}^2\right)$$

with $\tilde{s}^2 = \frac{1}{n}\sum_{i=1}^n (x_i - \mu)^2, \delta_1 = \frac{\nu_0}{2}$ and $\delta_2 = \frac{\nu_0-2}{2}\sigma_0^2$. The posterior distribution depends on the parameters of the prior distribution, the sample size and the sample variance $\tilde{s}^2$. The informativeness of a prior distribution is not a quality of the prior itself but can only

be evaluated with regard to the likelihood. Posterior expectation and variance are given by

$$\mathbb{E}(\sigma^2|x) = \frac{\delta_1 - 1}{\delta_1 + \frac{n}{2} - 1}\sigma_0^2 + \frac{\frac{n}{2}}{\delta_1 + \frac{n}{2} - 1}\tilde{s}^2,$$

$$\mathrm{Var}(\sigma^2|x) = \frac{\left(\delta_1\sigma_0^2 + \frac{n}{2}\tilde{s}^2\right)^2}{\left(\delta_1 + \frac{n}{2} - 1\right)^2\left(\delta_1 + \frac{n}{2} - 2\right)}.$$

The posterior expectation results from a weighted sum of the prior parameter σ_0^2 and the sample variance. Therefore, the confidence towards the prior knowledge, i.e. σ_0^2, is determined through the ratio between $\delta_1 - 1$ and $\delta_1 - 1 + n/2$, which is also the result of calculating the confidence ratio (eq. 5.9):

$$CR = \frac{\delta_1 - 1}{\delta_1 - 1 + \frac{n}{2}}$$

$\mathrm{Var}(\sigma^2|x)$ depends, aside from σ_0^2 and $\tilde{s}^2$, on the confidence ratio:

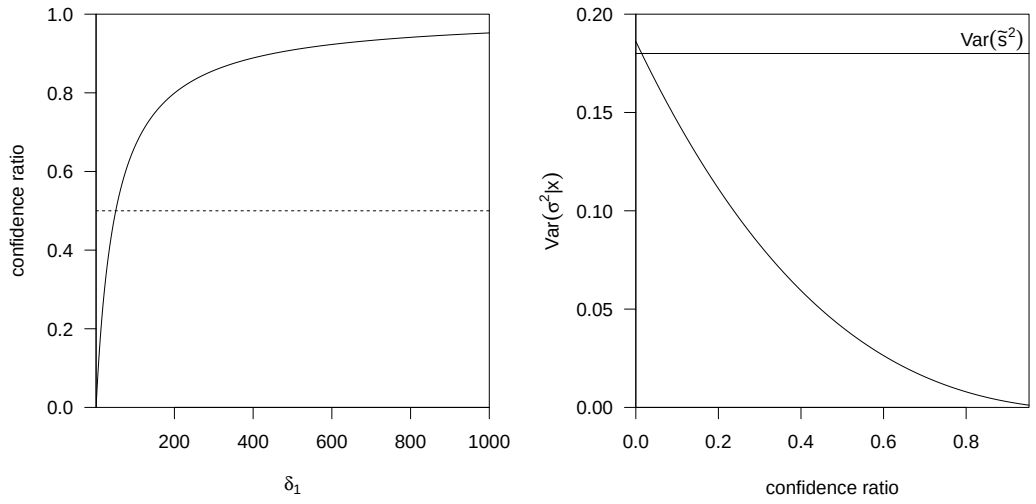


FIGURE 5.2.: *Dependencies between the parameters of the prior distribution, the confidence ratio and the posterior variance using an Inverse Gamma prior for* σ^2. $\sigma_0^2 = 1; \tilde{s}^2 = 3; n = 100; \delta_1$ *varying.*

If the confidence in σ_0^2 is small, i.e. δ_1 is small, the posterior variance will be high and will be about the variance of the frequentist estimation:

$$\mathrm{Var}(\tilde{s}^2) = \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2\right) = \frac{1}{n^2}\sigma^4\mathrm{Var}\left(\sum_{i=1}^{n}\frac{(x_i - \mu)^2}{\sigma^2}\right) = \frac{2}{n}\sigma^4$$

Assuming a Uniform prior distribution for σ^2 on the interval $[\gamma_1, \gamma_2]$, every value of the interval exhibits the same probability. The posterior distribution is given by

$$f(\sigma^2|x) = \frac{f(x|\sigma^2)\mathbf{I}_{[\gamma_1,\gamma_2]}(\sigma^2)}{\int_{-\infty}^{\infty} f(x|\sigma^2)\mathbf{I}_{[\gamma_1,\gamma_2]}(\sigma^2)\mathrm{d}\sigma^2}.$$

If a value $\sigma_0^2 \in [\gamma_1, \gamma_2]$ is chosen as the a priori best guess for σ^2, the posterior expectation should approximate σ_0^2 with increasing knowledge about σ^2, i.e. with $\gamma_1 \to \sigma_0^2$ and $\gamma_2 \to \sigma_0^2$. The posterior expectation of σ^2 is:

$$\mathbb{E}(\sigma^2|x) = \int_{-\infty}^{\infty}\sigma^2\frac{f(x|\sigma^2)\mathbf{I}_{[\gamma_1,\gamma_2]}(\sigma^2)}{\int_{-\infty}^{\infty} f(x|\sigma^2)\mathbf{I}_{[\gamma_1,\gamma_2]}(\sigma^2)\mathrm{d}\sigma^2}\mathrm{d}\sigma^2 = \int_{\gamma_1}^{\gamma_2}\sigma^2\frac{f(x|\sigma^2)}{\int_{\gamma_1}^{\gamma_2} f(x|\sigma^2)\mathrm{d}\sigma^2}\mathrm{d}\sigma^2.$$

Thus, the confidence ratio results as

$$CR = 1 - \frac{\left| \frac{\int_{\gamma_1}^{\gamma_2} \sigma^2 f(x|\sigma^2) \mathrm{d}\sigma^2}{\int_{\gamma_1}^{\gamma_2} f(x|\sigma^2) \mathrm{d}\sigma^2} - \sigma_0^2 \right|}{|\tilde{s}^2 - \sigma_0^2|} \approx 1 - \frac{\left| \frac{\int_{\gamma_1}^{\gamma_2} \sigma^2 f(x|\sigma^2) \mathrm{d}\sigma^2}{\int_{\gamma_1}^{\gamma_2} f(x|\sigma^2) \mathrm{d}\sigma^2} - \sigma_0^2 \right|}{\left| \frac{\int_{-\infty}^{\infty} \sigma^2 f(\sigma^2|x) \mathrm{d}\sigma^2}{\int_{-\infty}^{\infty} f(\sigma^2|x) \mathrm{d}\sigma^2} - \sigma_0^2 \right|}.$$

with $\gamma_1, \gamma_2 \geq 0$. The relation between the interval limits, the confidence ratio and the posterior variance is visualized in Figure 5.3 with $\gamma_1 = \gamma_2$:

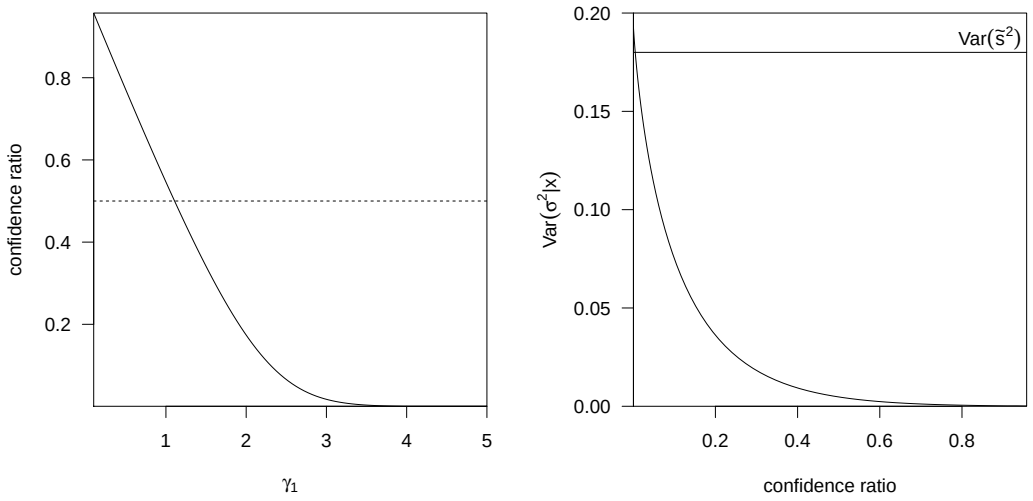


FIGURE 5.3.: *Dependencies between the parameters of the prior distribution, the confidence ratio and the posterior variance using a Uniform prior distribution for* σ^2. $\sigma_0^2 = 1; \tilde{s}^2 = 3; n = 100; \gamma_1 = \gamma_2$ *varying.*

Again, the posterior variance is about the variance of the frequentist estimate for high prior uncertainty. The formulation of informative prior knowledge using the Uniform prior distribution is often inappropriate, since the firm borders of prior distribution density strictly limits the domain with positive posterior density.

Choosing a Uniform prior distribution for $\log(\sigma^2)$ (Figure 5.4) yields a more rapid decrease of the confidence ratio with increasing interval width. The influence of the confidence ratio on the posterior variance is equal to the case with a Uniform prior distribution for σ^2.

By means of these examples we illuminate the association between the parameter choice for the prior distribution, the confidence ratio and the posterior variance. The confidence ratio allows to quantify the amount of prior knowledge about a parameter and to specify the prior knowledge about σ_0 independently from the prior distribution. Thus, the confidence ratio alleviates the determination of the parameters of the chosen prior distribution in practice, but quantifies the deviation between the prior distribution and the posterior distribution only regarding their expectations and not regarding the whole distributions.

Lower boundary of variance prior distributions

Situations are supposable where the occurrence of measurement error is not sure, i.e. the prior distribution has positive probability mass at $\sigma_U^2 = 0$. Lambert et al. (2005) observed biased variance estimates and high sensitivity towards the choice of the prior distribution in a Bayesian random effects model when the actual variance is very small.

If measurement error is probably absent, the usage of the commonly chosen conjugate Inverse Gamma prior distribution will be not reasonable because this prior distribution

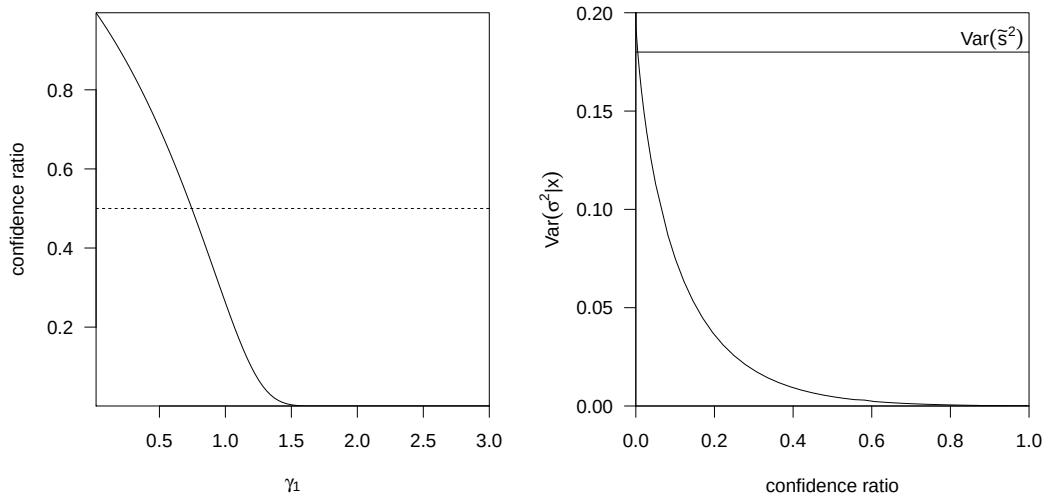


FIGURE 5.4.: *Dependencies between the parameters of the prior distribution, the confidence ratio and the posterior variance using a Uniform prior for* $\log \sigma^2$. $\sigma_0^2 = 1; \tilde{s}^2 = 3; n = 100; \gamma_1 = \gamma_2$ *varying.*

does not exhibit a positive probability mass at zero. We do not dwell on the problems arising with a measurement error size of zero, because measurement error exists in the data of the Augsburger Umweltstudie.

5.2.3. Prior distributions for the correlation coefficient

The Inverse Wishart distribution is the conjugate prior distribution for the variance–covariance matrix in a Bayesian model for multivariate normally distributed data with known mean and unknown variance–covariance matrix. However, distributional assumptions regarding the variance and the correlation matrix are not separately considered when the Inverse Wishart distribution is used as prior distribution. Furthermore, the uncertainty of the single entries of the variance–covariance matrix are assumed to be equal and the approach is limited to a balanced sample. Separation strategies (Barnard et al., 2000) for the variance–covariance matrix overcome the former drawback. Therefore, a prior distribution on the autocorrelation parameter ρ is preferred, even though there is not any conjugate prior distribution available and the effects off different priors are hard to conceive from a theoretical point of view.

We assume that the prior knowledge about the correlation matrix is independent from the prior knowledge about the variance, as it is proposed by Barnard et al. (2000). Possible prior distributions for the correlation coefficient are listed in Table 5.2. The last five prior distributions in Table 5.2 are suitable for the specification of informative prior distributions for correlation coefficients.

Prior distribution	References
$p(\rho) \propto 1$	Schisterman et al. (2003); Spiegelhalter (2001)
$p(\rho) \propto (1-\rho^2)^{-\frac{3}{2}}$	Schisterman et al. (2003)
$p(\rho) \propto \frac{\sqrt{1+\rho^2}}{1-\rho^2}$	Jeffreys (1961) as cited in Fosdick and Raftery (2012)
$p(\rho) = \frac{1}{\pi}\frac{1}{\sqrt{1-\rho^2}}$	Jeffreys (1961) as cited in Fosdick and Raftery (2012)
triangular shaped densities	Gokhale and Press (1982)
(Transformed) Beta	Gokhale and Press (1982); Spiegelhalter (2001)
Restricted Normal	Liechty (2004)
$\rho \sim \mathrm{N}(\mathrm{atanh}(\rho), \infty)$	Schisterman et al. (2003)
$0.5 \log \frac{1+\rho}{1-\rho} \sim \mathrm{N}(0, \sigma_\rho^2)$	Daniels and Kass (1999); Gryparis et al. (2007b)
$\rho \sim \mathrm{Unif}(-1, 1)$	Fosdick and Raftery (2012)

TABLE 5.2.: *Commonly used distributions for the construction of non–informative prior distributions for a correlation coefficient.*

5.2.4. Prior informativeness in Bayesian regression models for measurement errors

The considerations in Subsection 5.1.3 with regard to conditionally conjugate prior distributions for the sizes of the measurement errors in a Bayesian regression model for measurement error–prone data revealed that the expected value for the effect estimate depends solely on the prior expectation for the size of the measurement error but does not depend on its prior uncertainty. Thus, a property of Bayesian measurement error models is that the uncertainty about prior knowledge for a parameter describing the measurement error does not regulate the influence of the prior expectation and the estimate from the data on the posterior expectation. This property finally results from the fact that the effect estimate cannot be identified without any prior knowledge.

This property is illustrated by a small simulation. Data for a simple linear regression model with a classical error–prone covariate $X^{*\mathrm{C}}$ are generated assuming X and U^{C} normally distributed with $\mathbb{E}(X) = 9.5$, $\mathrm{Var}(X) = 0.3$, $\mathbb{E}(U^{\mathrm{C}}) = 0$ and $\mathrm{Var}(U^{\mathrm{C}}) = \sigma^2_{U^{\mathrm{C}}} = 0.06$ and the model errors normally distributed with $\mathbb{E}(\varepsilon) = 0$ and $\mathrm{Var}(\varepsilon) = \sigma^2_\varepsilon = 0.7$; further, $\beta_0 = 73, \beta_X = 1, n = 100$ and $T = 20$. Two models are considered: 1) A single–parameter model for estimating $\sigma^2_{U^{\mathrm{C}}}$ from $X^{*\mathrm{C}}$ with known $\mathbb{E}(X), \mathrm{Var}(X)$ and $\mathbb{E}(U^{\mathrm{C}})$ and 2) a Bayesian simple linear model accounting for a covariate with classical measurement error. The prior mean for $\sigma^2_{U^{\mathrm{C}}}$, $\sigma^{2\,(0)}_{U^{\mathrm{C}}}$, as well as prior uncertainty about the size of the classical measurement error are varied; the latter is varied according to the confidence ratio defined in Subsection 5.2.2. The Inverse Gamma distribution is used as a prior distribution for $\sigma^2_{U^{\mathrm{C}}}$. Non–informative prior distributions are chosen for the remaining parameters: Uniform prior distributions for σ^2_X and σ^2_ε and Normal prior distributions for β_0 and β_X. 2000 MCMC iterations are conducted and the first 1000 iterations are discarded as burnin–iterations. The results are visualized in Figure 5.5.

The estimations of $\sigma^2_{U^{\mathrm{C}}}$ with the single–parameter model behave as expected: the absolute bias of the posterior mean increases and the variance of the posterior mean decreases with increasing confidence in the prior knowledge.

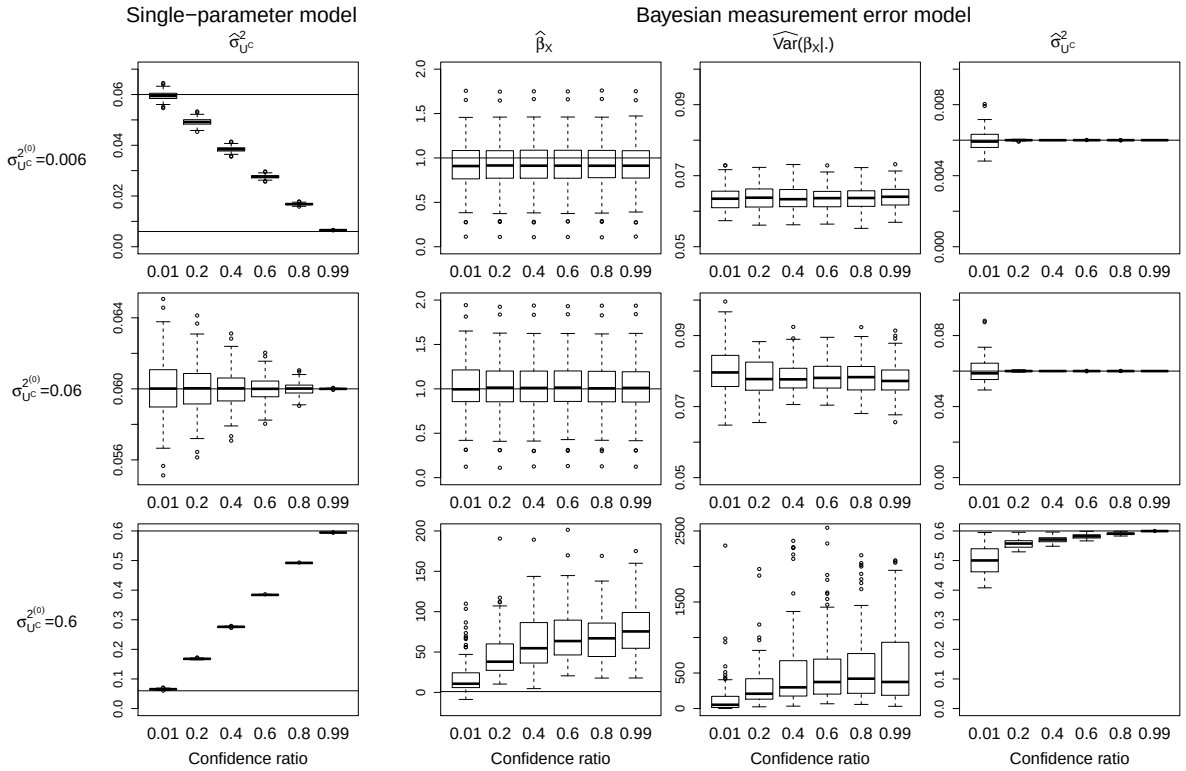


FIGURE 5.5.: *Posterior means ($\widehat{\sigma}^2_{UC}$) estimating σ^2_{UC} in a single–parameter model (left column). Posterior means ($\widehat{\beta_X}$) and variances ($\widehat{Var}(\beta_X|\cdot)$) for the effect coefficient β_X and posterior means ($\widehat{\sigma}^2_{UC}$) for the size of the measurement error σ^2_{UC} in a Bayesian model for a classical error–prone covariate (columns 2–4) for varying location of $\sigma^{2\,(0)}_{UC} \in \{0.006, 0.06, 0.6\}$ and confidence in the prior knowledge (confidence ratio).*

The Bayesian model for a classical error–prone covariate properly corrects the attenuation of effect estimate, if the prior expectation regarding the measurement error variance equals the actual parameter ($\sigma^{2\,(0)}_{UC} = 0.06$). However, the confidence regarding the prior knowledge has not any visible impact on the posterior mean of the effect coefficient and the variability thereof. For $\sigma^{2\,(0)}_{UC} = 0.6$, the Markov chains of the Bayesian measurement error models do not converge and the parameter estimations are highly biased since the prior distribution disagrees with the data ($\mathrm{Var}(X^{*\mathrm{C}}) < \sigma^{2\,(0)}_{UC}$). In contrast, the Markov chains converged for $\sigma^{2\,(0)}_{UC} = 0.006$, but the attenuation of the effect estimate is only insufficiently corrected, also in models with low confidence ratio.

Overestimation of σ^2_{UC} in the Bayesian model yields an overcorrection of the effect estimate attended by an increase of the variance of the effect estimate. Underestimation of σ^2_{UC} results in a too small correction of the attenuation attended by a reduction of the variance of the effect estimate. Prior expectations of σ^2_{UC} seem to equal posterior means for $\sigma^{2\,(0)}_{UC} = 0.006$ and for $\sigma^{2\,(0)}_{UC} = 0.06$, the cases, for which the MCMC chains converged. Even the use of strongly non–informative priors in the sense of a low confidence ratio does not yield adequate estimations for the slope parameter in the case of a misspecified prior expectation. The impact of the variability of the prior distribution for the measurement error parameters on the posterior distribution of the slope parameter in a simple linear regression model is theoretically considered in Section 5.3.

The choice of the prior expectation for σ^2_{UC} is essential for the adequate consideration of the measurement error. The parameter δ_1 has not any impact on the estimation of the

posterior expectation for $\sigma^2_{U^{\mathrm{C}}}$ (see Subsection 5.1.3). In fact, the conditional posterior distribution of β_X depends on the variation parameter δ_1, but δ_1 has only a marginal impact on $\widehat{\mathrm{Var}}(\beta_X)$ in the simulation.

The traditional way to specify the uncertainty of prior knowledge as a trade–off between data and prior knowledge is not adequate in Bayesian measurement error models with the prior distribution for the measurement error parameters as the single information source for the measurement error. This becomes obvious through the considerations in Subsection 5.1.3 and through the simulation presented in this subsection.

In addition, the simulations show, that the conditional posterior means for $\boldsymbol{X}$ are highly correlated with $\boldsymbol{X}^{*\mathrm{C}}$ (correlation coefficient about 1) and only slightly correlated with $\boldsymbol{Y}$ for $\sigma^{2\ (0)}_{U^{\mathrm{C}}} = 0.06$ and $\sigma^{2\ (0)}_{U^{\mathrm{C}}} = 0.006$ indicating that the response variable has not any impact on the estimations of $\boldsymbol{X}$ and β_X (eq. 5.5 and eq. 5.7). For $\sigma^{2\ (0)}_{U^{\mathrm{C}}} = 0.6$, the prior information is not plausible for the data; in this case the correlation between the estimations of $\boldsymbol{X}$ and $\boldsymbol{X}^{*\mathrm{C}}$ is lower and the correlation between the estimations of $\boldsymbol{X}$ and $\boldsymbol{Y}$ is higher.

5.3. Properties of posterior distributions

The notation and the basic ideas in Section 5.3 originate from the work of Gustafson (2005) and Gustafson (2015).

5.3.1. Theory

According to Gustafson (2005), some approximate properties of posterior distributions can be achieved by integrating over the model parameters under the assumption of proper joint prior distributions. This approach bases on the so–called transparent reparametrization of the model parameters $\boldsymbol{\Theta}$ to $\boldsymbol{\phi} = (\boldsymbol{\phi}_I, \boldsymbol{\phi}_N)$ with the aim, that the model for the data does not depend on the latent (non–identifiable) model parameters $\boldsymbol{\phi}_N$, i.e. $f(\boldsymbol{\phi}_N|\boldsymbol{\phi}_I, \text{data}) = f(\boldsymbol{\phi}_N|\boldsymbol{\phi}_I)$, but only on the so–called identifiable parameters $\boldsymbol{\phi}_I$: $f(\text{data}|\boldsymbol{\phi}) = f(\text{data}|\boldsymbol{\phi}_I)$. These identifiable parameters are determined so that they can be directly estimated from the data. For example, the transparent parametrization in a simple linear model with classical measurement error–prone covariate is given by $\boldsymbol{\phi}_I = (\beta_0^*, \beta_X^*, \mu_{X^*}, \sigma^2_{\varepsilon^*}, \sigma^2_{X^*})$ and $\boldsymbol{\phi}_N = r$; this example is used in the subsequent subsections to which we refer for further details. The posterior expectation of the parameter of interest $\psi = g_1(\boldsymbol{\phi})$ results through

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{\phi}}(\psi|\text{data}) &= \int\int g_1(\boldsymbol{\phi}_I, \boldsymbol{\phi}_N) f(\boldsymbol{\phi}_N|\boldsymbol{\phi}_I) d\boldsymbol{\phi}_N f(\boldsymbol{\phi}_I|\text{data}) d\boldsymbol{\phi}_I \\
&= \mathbb{E}_{\boldsymbol{\phi}_I}(\tilde{g}_1(\boldsymbol{\phi}_I)|\text{data}) \\
&= \mathbb{E}_{\boldsymbol{\phi}_I}(\mathbb{E}_{\boldsymbol{\phi}_N|\boldsymbol{\phi}_I}(g_1(\boldsymbol{\phi}))|\text{data}) \qquad (5.10)
\end{aligned}$$

with $\tilde{g}_1(\boldsymbol{\phi}_I) = \int g_1(\boldsymbol{\phi}_I, \boldsymbol{\phi}_N) f(\boldsymbol{\phi}_N|\boldsymbol{\phi}_I) d\boldsymbol{\phi}_N = \mathbb{E}_{\boldsymbol{\phi}_N|\boldsymbol{\phi}_I}(g_1(\boldsymbol{\phi}))$. The expectation of $\psi^2 = g_2(\boldsymbol{\phi})$ is derived as

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{\phi}}(\psi^2|\text{data}) &= \mathbb{E}_{\boldsymbol{\phi}_I}(\mathrm{Var}_{\boldsymbol{\phi}_N|\boldsymbol{\phi}_I}(\psi)|\text{data}) + \mathbb{E}_{\boldsymbol{\phi}_I}(\mathbb{E}_{\boldsymbol{\phi}_N|\boldsymbol{\phi}_I}(\psi)^2|\text{data}) \\
&= \mathbb{E}_{\boldsymbol{\phi}_I}(\tilde{g}_2(\boldsymbol{\phi}_I)|\text{data}) \\
&= \mathbb{E}_{\boldsymbol{\phi}_I}(\mathbb{E}_{\boldsymbol{\phi}_N|\boldsymbol{\phi}_I}(g_2(\boldsymbol{\phi}))|\text{data}),
\end{aligned}$$

with $\tilde{g}_2(\boldsymbol{\phi}_I) = \int g_2(\boldsymbol{\phi}_I, \boldsymbol{\phi}_N) f(\boldsymbol{\phi}_N|\boldsymbol{\phi}_I) d\boldsymbol{\phi}_N = \mathbb{E}_{\boldsymbol{\phi}_N|\boldsymbol{\phi}_I}(g_2(\boldsymbol{\phi}))$ resulting in the posterior variance of ψ:

$$\begin{aligned}
\mathrm{Var}_{\boldsymbol{\phi}}(\psi|\text{data}) =& \mathbb{E}_{\boldsymbol{\phi}}(\psi^2|\text{data}) - \mathbb{E}_{\boldsymbol{\phi}}(\psi|\text{data})^2 \\
=& \mathbb{E}_{\boldsymbol{\phi}_I}(\mathrm{Var}_{\boldsymbol{\phi}_N|\boldsymbol{\phi}_I}(\psi)|\text{data}) + \mathbb{E}_{\boldsymbol{\phi}_I}(\mathbb{E}_{\boldsymbol{\phi}_N|\boldsymbol{\phi}_I}(\psi)^2|\text{data}) - \mathbb{E}_{\boldsymbol{\phi}_I}(\mathbb{E}_{\boldsymbol{\phi}_N|\boldsymbol{\phi}_I}(\psi)|\text{data})^2 \\
=& \mathbb{E}_{\boldsymbol{\phi}_I}(\mathrm{Var}_{\boldsymbol{\phi}_N|\boldsymbol{\phi}_I}(\psi)|\text{data}) + \mathrm{Var}_{\boldsymbol{\phi}_I}(\mathbb{E}_{\boldsymbol{\phi}_N|\boldsymbol{\phi}_I}(\psi)|\text{data}) \\
=& \mathbb{E}_{\boldsymbol{\phi}_I}(\tilde{g}_2(\boldsymbol{\phi}_I)|\text{data}) - \mathbb{E}_{\boldsymbol{\phi}_I}(\tilde{g}_1(\boldsymbol{\phi}_I)^2|\text{data}) + \mathrm{Var}_{\boldsymbol{\phi}_I}(\tilde{g}_1(\boldsymbol{\phi}_I)|\text{data}). \qquad (5.11)
\end{aligned}$$

The asymptotic distribution of $\widehat{\psi}^{(n)} = \mathbb{E}(\psi|\text{data})$ as an estimate of ψ using n i.i.d. observations is derived by standard asymptotic theory if the model is correct (Gustafson, 2005):

$$n^{\frac{1}{2}} \left[\widehat{\psi}^{(n)} - \tilde{g}_1(\boldsymbol{\phi}_I)\right] \xrightarrow{d} N\left(0, [\tilde{g}_1'(\boldsymbol{\phi}_I)]^\top \boldsymbol{I}(\boldsymbol{\phi}_I)^{-1} [\tilde{g}_1'(\boldsymbol{\phi}_I)]\right)$$

for $n \to \infty$; $\boldsymbol{I}$ denotes the expected Fisher information. Thus, the following approximations hold:

$$\mathbb{E}\left(\widehat{\psi}^{(n)}\right) \approx \tilde{g}_1(\boldsymbol{\phi}_I) \qquad (5.12)$$

$$\mathrm{Var}\left(\widehat{\psi}^{(n)}\right) \approx \mathrm{Var}_{\boldsymbol{\phi}_I}(\tilde{g}_1(\boldsymbol{\phi}_I)|\text{data}) = n^{-1}[\tilde{g}_1'(\boldsymbol{\phi}_I)]^\top \boldsymbol{I}(\boldsymbol{\phi}_I)^{-1}[\tilde{g}_1'(\boldsymbol{\phi}_I)] \qquad (5.13)$$

$$\mathrm{Var}(\psi|\text{data}) \approx \tilde{g}_2(\boldsymbol{\phi}_I) - \tilde{g}_1(\boldsymbol{\phi}_I)^2 + n^{-1}[\tilde{g}_1'(\boldsymbol{\phi}_I)]^\top \boldsymbol{I}(\boldsymbol{\phi}_I)^{-1}[\tilde{g}_1'(\boldsymbol{\phi}_I)] \qquad (5.14)$$

$\mathrm{Var}\left(\widehat{\psi}^{(n)}\right)$ asymptotically does not depend on the variability of $\boldsymbol{\phi}_N$, if $\tilde{g}_1'(\boldsymbol{\phi}_I)$ does not depend on $\boldsymbol{\phi}_N$, e.g. if $g_1(\boldsymbol{\phi})$ is a linear function of $\boldsymbol{\phi}_N$.

This approach allows to examine quantities characterizing the posterior distribution of ψ without conducting the Bayesian analysis. The key issue for the asymptotic properties is the conditional prior distribution of $\boldsymbol{\phi}_N|\boldsymbol{\phi}_I$. The phenomenon of a priori dependencies between $\boldsymbol{\phi}_N$ and $\boldsymbol{\phi}_I$ calls Gustafson (2005) "indirect learning" (IL) about $\boldsymbol{\phi}_N$. If there is not any indirect learning (NIL) about the non–identifiable parameter $\boldsymbol{\phi}_N$, $f(\boldsymbol{\phi}_N|\boldsymbol{\phi}_I) = f(\boldsymbol{\phi}_N)$. When a proper joint prior distribution is chosen, preventing IL is mostly only approximately achieved. Then,

$$\tilde{g}(\boldsymbol{\phi}_I) = \mathbb{E}_{\boldsymbol{\phi}_N|\boldsymbol{\phi}_I}(g(\boldsymbol{\phi})) \overset{\text{NIL}}{\approx} \mathbb{E}_{\boldsymbol{\phi}_N}(g(\boldsymbol{\phi})),$$

with $g \in \{g_1, g_2\}$ and $\tilde{g} \in \{\tilde{g}_1, \tilde{g}_2\}$. If prior knowledge about $\boldsymbol{\Theta}_{-\boldsymbol{\phi}_N}$ is not available, indirect learning from these parameters should be prevented by an adequate parameter choice.

Since IL is visible through the discrepancy between the prior distribution of $\boldsymbol{\phi}_N$ and its conditional posterior distribution, the KL–distance D_{KL} between these densities quantifies the amount of IL:

$$D_{KL}(f(\boldsymbol{\phi}_N|\boldsymbol{\phi}_I), f(\boldsymbol{\phi}_N)).$$

D_{KL} depends only on the identifiable parameters and therefore,

$$\mathbb{E}(D_{KL}(f(\boldsymbol{\phi}_N|\boldsymbol{\phi}_I), f(\boldsymbol{\phi}_N))|\text{data}) = \mathbb{E}\left(\widehat{D}_{KL}^{(n)}(f(\boldsymbol{\phi}_N|\boldsymbol{\phi}_I), f(\boldsymbol{\phi}_N))\right) \approx D_{KL}(f(\boldsymbol{\phi_N}|\boldsymbol{\phi}_I), f(\boldsymbol{\phi}_N)) \qquad (5.15)$$

for $n \to \infty$.

Indirect learning has several aspects and implications. These will be discussed within the next paragraphs by means of a simple linear regression model with classical measurement error introduced in Subsection 5.3.2, after a short overview of properties in the absence of IL (Subsection 5.3.3). We will focus on IL from I) prior distributions on the

parameters of the main outcome model (Subsection 5.3.4), II) constraints on parameters in the main outcome model (Subsection 5.3.5) and III) constraints on parameters in the exposure model (Subsection 5.3.6). Since the transparent parametrization of the model is difficult to accomplish for more complex models, e.g. hierarchical models, we focus the considerations regarding the properties of posterior distributions on rather simple linear regression models and analyze more complex models in Section 5.4 via simulations.

5.3.2. Example setup: simple regression with classical covariate measurement error

Gustafson (2005) exemplifies his approach using a data situation with classical measurement error:

$$\begin{aligned} X^*|X,Y &\sim \mathrm{N}(X, r\sigma_X^2) \\ Y|X &\sim \mathrm{N}(\beta_0 + \beta_X X, \sigma_\varepsilon^2) \\ X &\sim \mathrm{N}(\mu_X, \sigma_X^2) \end{aligned} \tag{5.16}$$

X^* denotes the classical error–prone covariate throughout the Subsections 5.3.2 to 5.3.5. This parametrization will be used in the following to relate to the work of Gustafson (2005) and differs from our parametrization used in the former chapters:

$$X^*|X,Y \sim \mathrm{N}(X, \sigma_U^2) \tag{5.17}$$

In general, parametrization 5.16 describes a multiplicative association between σ_X^2 and $\sigma_{X^*}^2$, whereas parametrization 5.17 bases on an additive association. For parametrization 5.16, the transparent parametrization with $\boldsymbol{\phi}_N = r$ and $\boldsymbol{\phi}_I = (\beta_0^*, \beta_X^*, \mu_{X^*}, \sigma_{\varepsilon^*}^2, \sigma_{X^*}^2)$ of the original parameter vector $\boldsymbol{\Theta} = (\beta_0, \beta_X, \mu_X, r, \sigma_\varepsilon^2, \sigma_X^2)$ is obtained by the following transformations:

$$\begin{aligned} \beta_0^* &= \beta_0 + \mu_X\beta_X/(1+r) \\ \beta_X^* &= \beta_X/(1+r) \\ \mu_{X^*} &= \mu_X \\ \sigma_{\varepsilon^*}^2 &= \sigma_\varepsilon^2 + \beta_X^2\sigma_X^2 r/(1+r) \\ \sigma_{X^*}^2 &= \sigma_X^2(1+r). \end{aligned}$$

Thus, $\psi = g(\boldsymbol{\phi}) = \beta_X^*(1+r)$.

Constraining the domain of r to an interval between zero and one is realistic. Additionally, all variances have to be positive or zero; this will be the case, if

$$r \in (0, \min\{\sigma_{\varepsilon^*}^2/({\beta_X^*}^2\sigma_{X^*}^2), 1\}).$$

Some properties of the limiting posterior distribution of ψ are illustrated in the next paragraphs using the parameter settings

$$\beta_0 = 0, \beta_X = 1, \sigma_\varepsilon^2 = 0.25, \mu_X = 0, \sigma_X^2 = 1, r = 0.25$$

and the priors

$$\begin{aligned} &\beta_0 \sim \mathrm{N}(\beta_0^{(0)}, \sigma_{\beta_0}^2), \beta_X \sim \mathrm{N}(\beta_X^{(0)}, \sigma_{\beta_X}^2), \mu_X \sim \mathrm{N}(\mu_X^{(0)}, \sigma_{\mu_X}^2), \\ &\sigma_\varepsilon^2 \sim \mathrm{IG}(\delta_{\varepsilon,1}, \delta_{\varepsilon,2}), \sigma_X^2 \sim \mathrm{IG}(\delta_{X,1}, \delta_{X,2}), \\ &r: \text{ varying priors with expectation } r^{(0)} \end{aligned}$$

as in Gustafson (2005); $\beta_0^{(0)}, \beta_X^{(0)}$ and $\mu_X^{(0)}$ denote the prior expectations of the corresponding parameters.

The conditional prior distribution of $\boldsymbol{\phi}_N|\boldsymbol{\phi}_I$ is according to Gustafson (2005)

$$\begin{aligned} f(r|\beta_0^*, \beta_X^*, \mu_{X^*}, \sigma_{\varepsilon^*}^2, \sigma_{X^*}^2) \propto & f_{\mathrm{N}}(\beta_X^*(1+r)) f_{\mathrm{IG}}(\sigma_{\varepsilon^*}^2 - r(\beta_X^*)^2\sigma_{X^*}^2) \\ & f_{\mathrm{IG}}(\sigma_{X^*}^2/(1+r)) f_r(r) \mathbf{I}_{(0,\min\{\sigma_{\varepsilon^*}^2/(\beta_X^{*\,2}\sigma_{X^*}^2),1\})}(r) \end{aligned} \tag{5.18}$$

with $f_{\mathrm{N}}(\cdot)$ denoting the density function of $\mathrm{N}(0,1)$, $f_{\mathrm{IG}}(\cdot)$ the density function of $\mathrm{IG}(0.5, 0.5)$ and with an arbitrary proper prior density $f_r(\cdot)$ for the non–identifiable parameter r.

The conditional prior distribution of r (eq. 5.18) does not only depend on the prior distribution of r but also on the prior distributions of $\boldsymbol{\Theta}_{-r}$ and the restriction regarding the domain of the latent parameter, $\mathbf{1}_{(0,\min\{\sigma_{\varepsilon^*}^2/(\beta_X^{*\,2}\sigma_{X^*}^2),1\})}(r)$ resulting in IL. Hence, the asymptotic posterior expectation of r does not necessarily equal the prior expectation of r, even if the data does not contain any information about r. This is illustrated in Figure 5.6 using a Uniform prior distribution for r around 0.25 and gradually increasing γ:

$$r \sim \mathrm{Unif}(\gamma_1, \gamma_2) = \mathrm{Unif}(0.25 - \gamma, 0.25 + \gamma)$$

Figure 5.6 a)–c) present the asymptotic properties (Bias, standard deviation, RMSE) of $\widehat{\beta}_X$ for increasing prior uncertainty regarding r. Exemplary, three prior distributions for r with varying prior uncertainty are depicted in Figure 5.6 d).

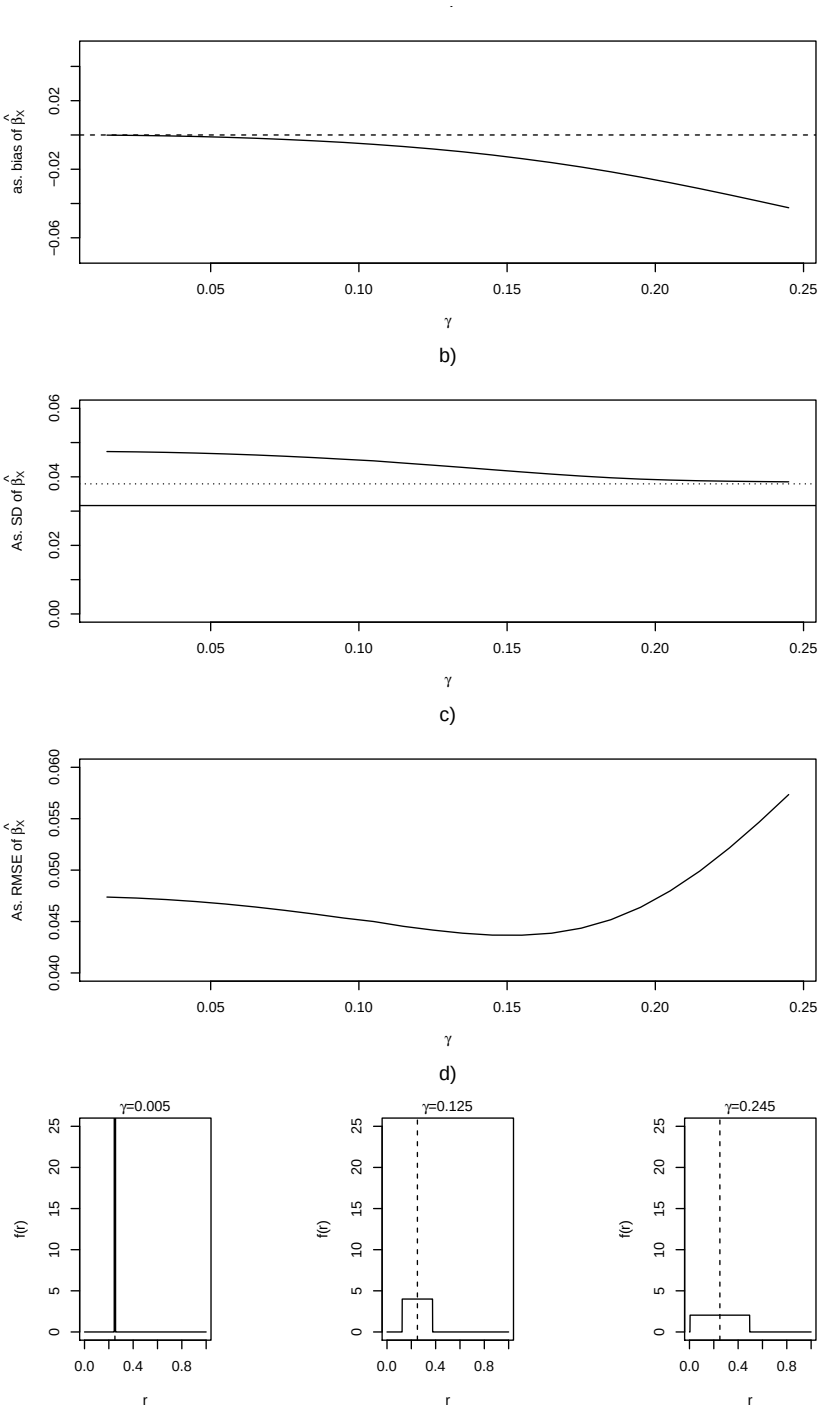


FIGURE 5.6.: *Asymptotic bias, standard deviation (SD) and RMSE of* $\widehat{\beta}_X = \mathbb{E}(\beta_X|data)$ *using Unif*$(0.25-\gamma, 0.25+\gamma)$ *for the marginal prior distribution of* r *with gradually decreasing values of* γ*; dashed vertical line:* $r = 0.25$*; dotted line: Var*$(\widehat{\beta_X^*}^{(n)})$*; solid, horizontal line: Var*$(\widehat{\beta_X}^{(n)})$*, if* X *is observed.* $\beta_0^{(0)} = 0, \sigma^2_{\beta_0} = 1, \beta_X^{(0)} = 0, \sigma^2_{\beta_X} = 1, \mu_X^{(0)} = 0, \sigma^2_{\mu_X} = 1, \delta_{\varepsilon,1} = 0.5, \delta_{\varepsilon,2} = 0.5, \delta_{X,1} = 0.5, \delta_{X,2} = 0.5, n = 250.$

Figure 5.6 shows a stronger bias for higher prior uncertainty and increased variability of $\widehat{\beta_X}$ for decreasing prior uncertainty; similar observations are made in Gustafson (2005). Thus, the shape of the prior distribution of r does not seem to be directly related to its informativeness. These are undesirable properties because actually, the prior uncertainty should not affect the expectation of the posterior distribution of β_X and increasing prior uncertainty should involve a higher variability of the parameter estimate.

5.3.3. Properties in the absence of indirect learning

In the absence of IL through using non–informative prior distributions for $\boldsymbol{\Theta}_{-r}$ and not imposing any constraints on the domain of r, eq. 5.18 simplifies to

$$f(r|\beta_0^*, \beta_X^*, \mu_{X^*}, \sigma_{\varepsilon^*}^2, \sigma_{X^*}^2) \propto f_r(r) \; . \tag{5.19}$$

Choosing proper joint prior distributions for $\boldsymbol{\phi}_I$ implicates the informativeness of the joint prior distribution of $\boldsymbol{\phi}_I$. Therefore, the simplification in eq. 5.19 is a limiting case, which is only asymptotically reached in realistic examples.

The properties of the posterior distribution of ψ are

$$\begin{aligned}
\mathbb{E}(\psi|\text{data}) &\overset{\text{eq. 5.10}}{=} & \mathbb{E}_{\boldsymbol{\phi}_I}(\mathbb{E}_{\boldsymbol{\phi}_N|\boldsymbol{\phi}_I}(g_1(\boldsymbol{\phi}))|\text{data}) \\
&\overset{\text{NIL}}{\approx} & \mathbb{E}_{\boldsymbol{\phi}_I}(\mathbb{E}_{\boldsymbol{\phi}_N}(g_1(\boldsymbol{\phi}))|\text{data}) = \mathbb{E}_{\beta_X^*}(\beta_X^*|\text{data})(1+\mathbb{E}_r(r)) \\
\text{Var}(\mathbb{E}(\psi|\text{data})) &\overset{\text{eq. 5.10}}{=} & \text{Var}_{\boldsymbol{\phi}_I}(\mathbb{E}_{\boldsymbol{\phi}_N|\boldsymbol{\phi}_I}(g_1(\boldsymbol{\phi}))|\text{data}) \\
&\overset{\text{NIL}}{\approx} & \text{Var}_{\boldsymbol{\phi}_I}(\mathbb{E}_{\boldsymbol{\phi}_N}(g_1(\boldsymbol{\phi}))|\text{data}) = (1+\mathbb{E}_r(r))^2\text{Var}_{\beta_X^*}(\beta_X^*|\text{data}) \\
\text{Var}(\psi|\text{data}) &\overset{\text{eq. 5.11}}{=} & \mathbb{E}_{\boldsymbol{\phi}_I}(\text{Var}_{\boldsymbol{\phi}_N|\boldsymbol{\phi}_I}(g_1(\boldsymbol{\phi}))|\text{data}) + \text{Var}_{\boldsymbol{\phi}_I}(\mathbb{E}_{\boldsymbol{\phi}_N|\boldsymbol{\phi}_I}(g_1(\boldsymbol{\phi}))|\text{data}) \\
&\overset{\text{NIL}}{\approx} & \mathbb{E}_{\boldsymbol{\phi}_I}(\text{Var}_{\boldsymbol{\phi}_N}(g_1(\boldsymbol{\phi}))|\text{data}) + \text{Var}_{\boldsymbol{\phi}_I}(\mathbb{E}_{\boldsymbol{\phi}_N}(g_1(\boldsymbol{\phi}))|\text{data}) \\
&= & \mathbb{E}_{\beta_X^*}({\beta_X^*}^2|\text{data})\text{Var}_r(r) + \text{Var}_{\beta_X^*}(\beta_X^*|\text{data})(1+\mathbb{E}_r(r))^2 \\
&= & \text{Var}_{\beta_X^*}(\beta_X^*|\text{data})\text{Var}_r(r) + \mathbb{E}_{\beta_X^*}(\beta_X^*|\text{data})^2\text{Var}_r(r) + \\
& & \text{Var}_{\beta_X^*}(\beta_X^*|\text{data})(1+\mathbb{E}_r(r))^2
\end{aligned}$$

$\mathbb{E}(\widehat{\beta}_X^{(n)})$ and $\text{Var}(\widehat{\beta}_X^{(n)})$ are not affected by the variability of $\boldsymbol{\phi}_N$; this relation holds for an arbitrary choice of prior distributions as long as the NIL assumption is valid and was already detected while using conditionally conjugate priors (Subsection 5.1.3).

The approximate variance of the posterior mean of β_X strongly depends on the prior mean of r, but not on the prior variance of r. Even in the case of exact prior knowledge about r, $\text{Var}_{\beta_X^*}(\beta_X^*|\text{data}) < \text{Var}(\widehat{\beta}_X^{(n)})$, if $r > 0$. The reason for this is, that higher measurement error results in increased uncertainty about the true measurements X.

$\text{Var}(\psi|\text{data})$ is affected by the prior variance of r, and thus by the prior uncertainty about the size of the measurement error.

The asymptotic properties of the posterior distribution can easily be calculated and are useful as a reference for evaluating the strength of IL.

5.3.4. Indirect learning from prior distributions on the parameters of the main outcome model

The role of the prior distribution of β_X (here a normal prior with density f_{N}) regarding the posterior distribution of β_X is exemplified in the following; the prior distributions for the remaining parameters $\boldsymbol{\Theta}_{-r,\beta_X}$ are taken to be non–informative. Furthermore, constraints resulting from limited parameter domains are neglected and discussed later (Subsection 5.3.5). Thus,

$$f(r|\beta_0^*, \beta_X^*, \mu_{X^*}, \sigma_{\varepsilon^*}^2, \sigma_{X^*}^2) \propto f(r|\beta_X^*) \propto f_{\text{N}}(\beta_X^*(1+r))f_r(r)$$

approximates $f_r(r)$ (aside from multiplicative constants), if $\sigma_{\beta_X}^2 \to \infty$, i.e. if the prior for β_X is non–informative. As long as the prior of β_X is non–informative, IL is prevented and the posterior distribution of the parameter of interest is characterized by the properties derived in Subsection 5.3.3.

A 3–point prior for r is used to facilitate integration regarding r:

$$\mathrm{P}(r = r^{(0)} - q) = \mathrm{P}(r = r^{(0)}) = \mathrm{P}(r = r^{(0)} + q) = \frac{1}{3}$$

with $r^{(0)} = 0.25$. For the present example, indications for IL are investigated for (i) $\beta_X^{(0)} = 1, \sigma_{\beta_X} = 100$, (ii) $\beta_X^{(0)} = 1, \sigma_{\beta_X} = 1$ and (iii) $\beta_X^{(0)} = 0, \sigma_{\beta_X} = 1$ using the approximate KL–distance (eq. 5.15).

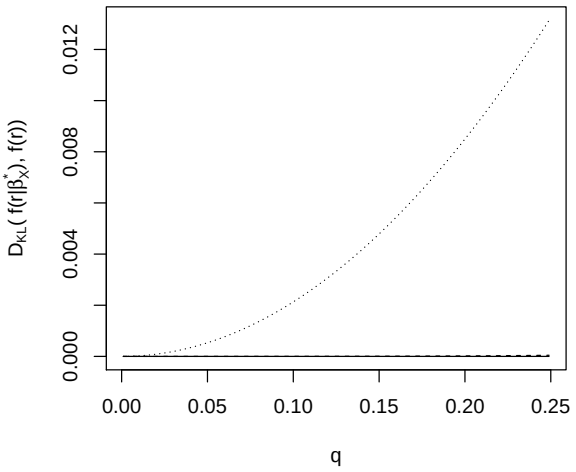


FIGURE 5.7.: $D_{KL}(f(r|\beta_X^*), f(r))$ *for varying prior uncertainty* q *and assuming normal prior distributions for* β_X *with (i)* $\beta_X^{(0)} = 1, \sigma_{\beta_X} = 100$ *(solid line), (ii)* $\beta_X^{(0)} = 1, \sigma_{\beta_X} = 1$ *(dashed line) and (iii)* $\beta_X^{(0)} = 0, \sigma_{\beta_X} = 1$ *(dotted line).*

The approximate NIL case occurs for the situations in (i) and (ii), because the approximate KL–distance between the conditional posterior distribution and the marginal prior distribution of r approximately equals zero (Figure 5.7). In contrast, IL of varying intensity occurs with a wider prior distribution for r in situation (iii). Figure 5.7 backs up the observation of Gustafson (2005), that wider priors for r increase the bias of the estimate for β_X; the reason for this behavior is the prevention of IL.

Since we focus on situations with knowledge regarding the measurement error originating solely from the prior distribution of $\boldsymbol{\phi}_N$, the prior distributions for $\boldsymbol{\Theta}_{-\boldsymbol{\phi}_N}$ should be chosen in a way to prevent IL. The amount of IL becomes also visible through the conditional distribution of r for different values of β_X^*:

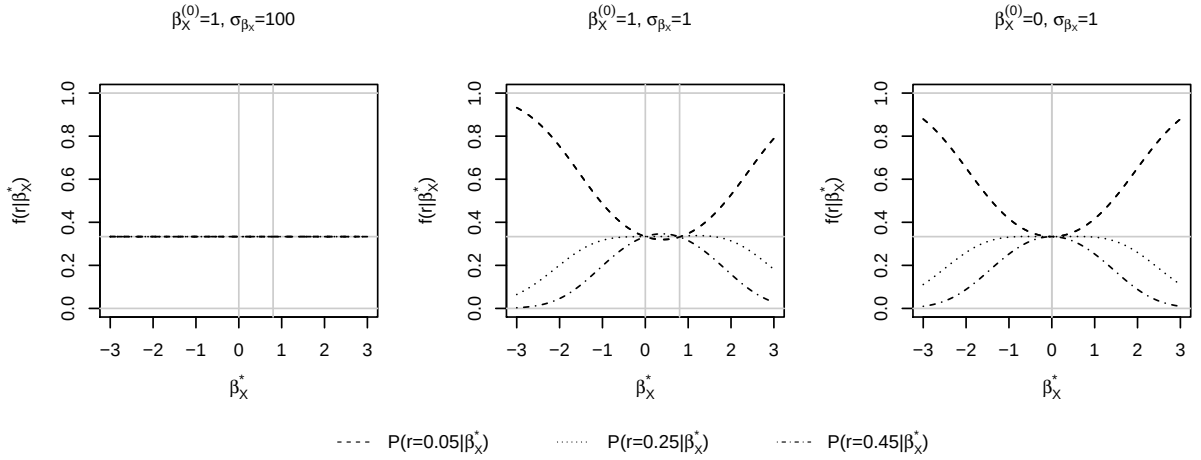


FIGURE 5.8.: *Limiting posterior probabilities of r for varying prior distributions of β_X and a 3–point prior for r with q=0.2. $P(r = 0.05|\beta_X^*)$: dashed lines, $P(r = 0.25|\beta_X^*)$: dotted lines, $P(r = 0.45|\beta_X^*)$: dotdashed lines. Horizontal gray lines: 0, $\frac{1}{3}$, 1; vertical lines: 0, $\beta_X^{(0)}/(1+0.25)$.*

The discrete Uniform prior distribution for r is only preserved for the prior distribution of β_X with $\beta_X^{(0)} = 1$ and $\sigma_{\beta_X} = 100$, whereas the conditional prior distribution for r differs from the marginal prior distribution for narrower prior distributions of β_X depending on the considered values of β_X^*. For the informative prior distributions of β_X with $\sigma_{\beta_X} = 1$, the discrete Uniform distribution is reached for $\beta_X^* = 0$ and approximately for $\beta_X^* = \beta_X^{(0)}/(1 - r^{(0)})$. Figure 5.8 shows that the effect of IL varies not only with $\sigma_{\beta_X}^2$, but also with $\beta_X^{(0)}$.

In the following, the effects of IL from the prior distribution of β_X on the properties of the posterior distribution of the parameter of interest β_X, will be discussed.

Approximate posterior mean

Figure 5.9 depicts the approximate posterior mean for varying prior information for β_X and r. Choosing informative priors for β_X results in an approximate posterior mean of β_X between $\beta_X^{(0)}$ and $\widehat{\beta}_X^*(1 + r^{(0)})$.

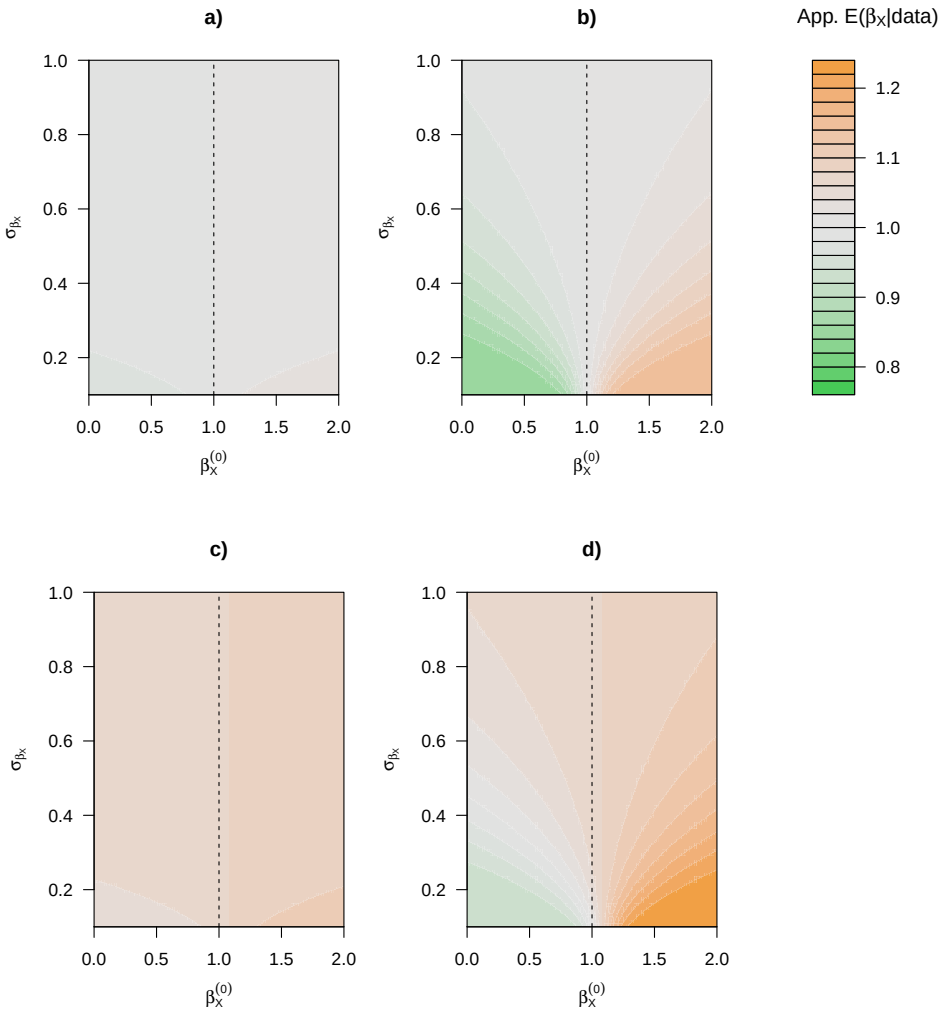


FIGURE 5.9.: *Approximate values of* $E(\beta_X|data)$ *for varying prior distributions of* β_X *and* r*; a):* $r^{(0)} = 0.25, q = 0.05$*, b):* $r^{(0)} = 0.25, q = 0.2$*, c):* $r^{(0)} = 0.35, q = 0.05$*, d):* $r^{(0)} = 0.35, q = 0.2$*; n=250.*

The informative priors for r affect the posterior means almost independently from the prior parameters of β_X (Figure 5.9 a) and c)) and are therefore the only source of prior knowledge for the posterior mean of β_X. In contrast, IL is relevant if the prior for r is less informative (Figure 5.9 b) and d)). The trade–off between the sources of prior knowledge, which depends strongly on the chosen prior distributions, becomes obvious.

Approximate variance of the posterior mean

If the prior distribution of β_X is non–informative and thus, IL is prevented, $\text{Var}(\widehat{\beta}_X^{(n)})$ will only be affected by the prior mean $r^{(0)}$, but not by the prior uncertainty about r (Subsection 5.3.3 and Figure 5.10 a) and c)) and will approximate $\text{Var}_{\beta_X^*}(\beta_X^*|\text{data})\mathbb{E}_r(1+r)^2$; in contrast, IL is present for the less informative prior distributions for r used in Figure 5.10 b) and d) and influences the variance of the posterior mean of β_X. The

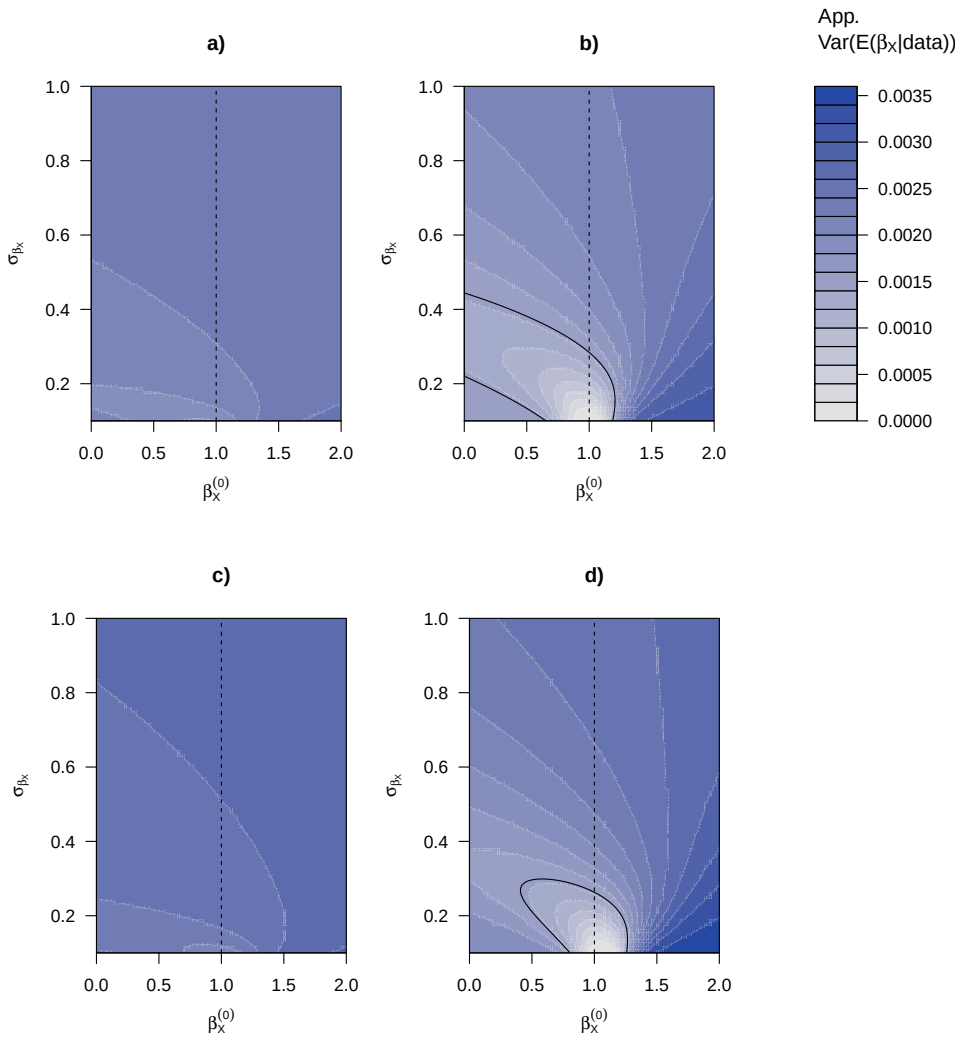


FIGURE 5.10.: *Approximate values of Var*$(\mathbb{E}(\beta_X|data))$ *for varying prior distributions of* β_X *and* r*; a):* $r^{(0)} = 0.25, q = 0.05$*, b):* $r^{(0)} = 0.25, q = 0.2$*, c):* $r^{(0)} = 0.35, q = 0.05$*, d):* $r^{(0)} = 0.35, q = 0.2; n = 250$*; solid line: app. Var*$(\mathbb{E}(\beta_X|data))$ *equals Var*$(\beta_X^*|data)$*.*

variance of the posterior mean may even increase through IL. This occurs in our example, if $\beta_X^{(0)}$ strongly deviates from the true value of β_X. IL reduces the variance of the posterior mean when the prior knowledge about β_X and the prior knowledge about r indicate the same value of β_X.

Approximate posterior variance

The approximate posterior variance is strongly affected by IL:

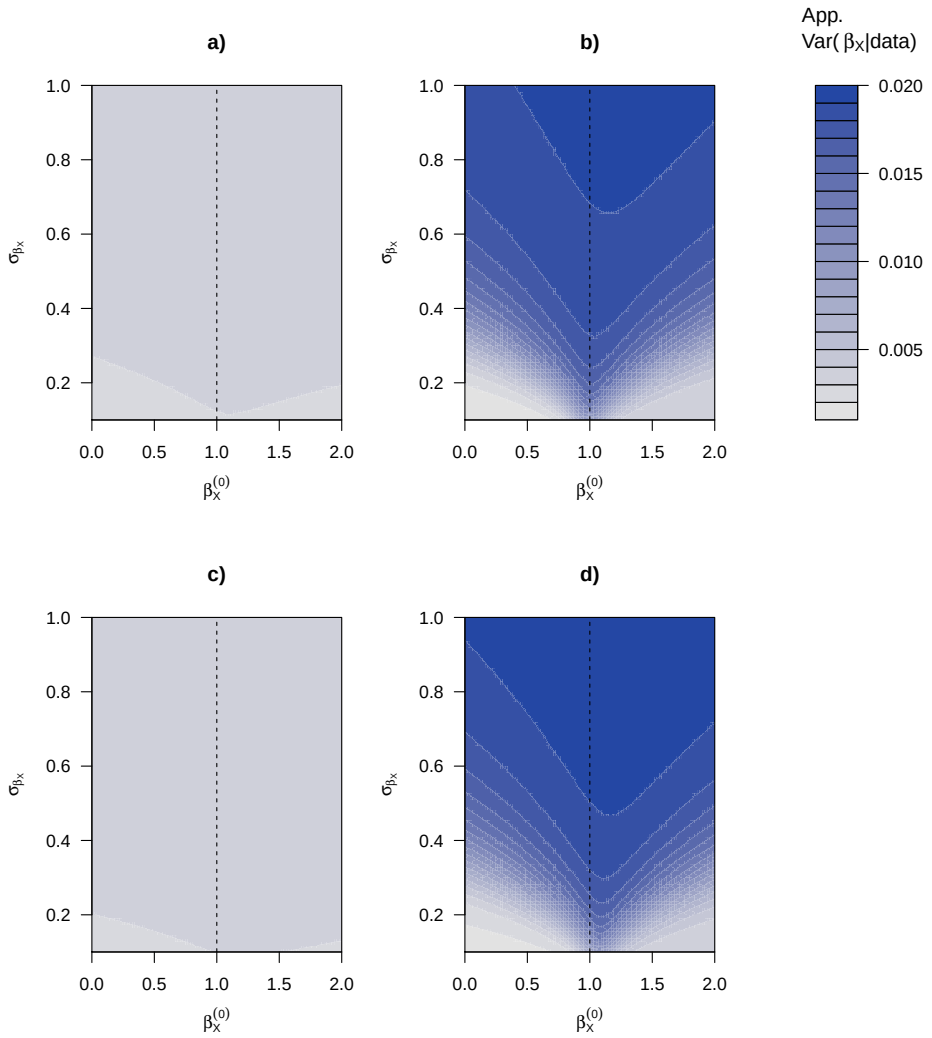


FIGURE 5.11.: *Approximate values of* $Var(\beta_X|data)$ *for varying prior distributions of* β_X *and* r*; a):* $r^{(0)} = 0.25, q = 0.05$*, b):* $r^{(0)} = 0.25, q = 0.2$*, c):* $r^{(0)} = 0.35, q = 0.05$*, d):* $r^{(0)} = 0.35, q = 0.2$*; n=250.*

Increasing σ_{β_X} involves a lower approximate posterior variance of β_X in the case of a less informative prior of r (Figure 5.11 b) and d)), whereas the posterior variance of β_X is nearly constant for an informative prior of r (Figure 5.11 a) and c)). Thus, the prior variability of r strongly influences the approximate posterior variance of β_X. IL reduces the posterior variance and the impact of the variability of r on the approximate posterior variance of β_X is smaller.

Conclusion

The influence of the informativeness of the prior distribution for r differs with the prior distributions of other parameters, which seem to modify the impact of the prior of r on the posterior distribution of β_X. Thus, the amount of IL does not only depend on the prior distribution of r, as it is also visible in Gustafson (2005), but also on the prior distributions of $\boldsymbol{\Theta}_{-r}$.

5.3.5. Indirect learning induced by constraints on parameters in the main outcome model

In the absence of IL from prior distributions (NILP), the limiting posterior of r (eq. 5.18) simplifies to

$$f(r|\beta_0^*, \beta_X^*, \mu_{X^*}, \sigma_{\varepsilon^*}^2, \sigma_{X^*}^2) \overset{\text{NILP}}{\approx} f_r(r)\mathbf{I}_{(0,\min\{\sigma_{\varepsilon^*}^2/({\beta_X^*}^2\sigma_{X^*}^2),1\})}(r) \cdot \text{const.}$$

The constraints on r directly affect $f(r|\boldsymbol{\Theta}_{-r})$ and the normalizing constant depends also on β_X^*; this implies IL from the constraints. For the simple examples in that subsection we take $\sigma_{\varepsilon^*}^2$ and $\sigma_{X^*}^2$ as fixed to the true values.

IL exists for the prior parameter choices (i)–(iii) for β_X, especially, if the prior distribution for r is less informative: The approximate KL–distance is strongly affected by

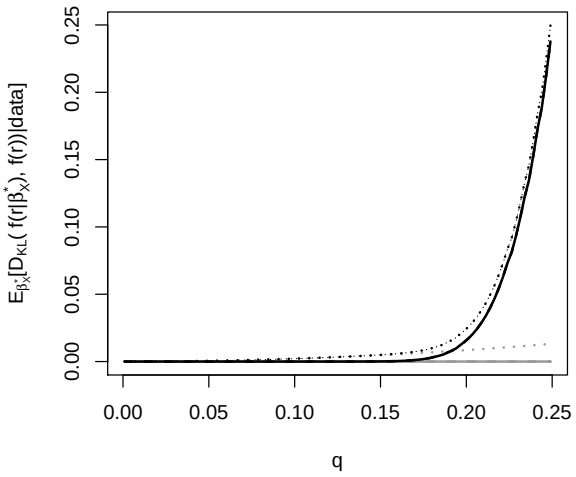


FIGURE 5.12.: *Estimation of* $\mathbb{E}_{\beta_X^*}(D_{KL}(f(r|\beta_X^*), f(r)))$ *based on* $n = 250$ *observations for varying* q *and assuming a normal prior distribution for* β_X *with (i)* $\beta_X^{(0)} = 1, \sigma_{\beta_X} = 100$ *(solid line), (ii)* $\beta_X^{(0)} = 1, \sigma_{\beta_X} = 1$ *(dashed line) and (iii)* $\beta_X^{(0)} = 0, \sigma_{\beta_X} = 1$ *(dotted line); gray: approximate* $\mathbb{E}_{\beta_X^*}(D_{KL}(f(r|\beta_X^*), f(r)))$ *for* $n \to \infty$.

the constraints for higher values of q, also for non–informative prior distributions of β_X. The reason is, that the conditional prior distribution of the non–identifiable parameter changes after including the constraints, as depicted in Figure 5.13.

Figure 5.12 and Figure 5.13 indicate that the usage of non–informative priors on β_X only mitigates IL through constraints but cannot avoid IL. IL through constraints will be prevented if the prior distribution of r is informative and n is large involving small variability of $\beta_X^*|$data in order to meet the limits of the constraints.

IL through constraints influences the posterior distribution of β_X in two aspects both originating from changes of the conditional prior distribution of r in comparison to the marginal prior distribution: 1) The prior mean of r changes and 2) the relationship between the information sources for the unobserved X changes.

In the NILP case, the magnitude of the effect arising from the constraints on the posterior distribution of β_X is driven by the distribution of β_X^*, and thus, also by the sample size n, and by the informativeness of the prior distribution for $\boldsymbol{\Theta}_{-r}$ as illustrated with Figure 5.14.

The constraint affects the properties of the posterior distribution of β_X, even if the probability mass of the non–identifiable parameter lies within the constraints and is of higher relevance, when the prior distribution for r is less informative. The reason for this is, that β_X^* is random and therefore, β_X^* delivers additional variability. The constraint

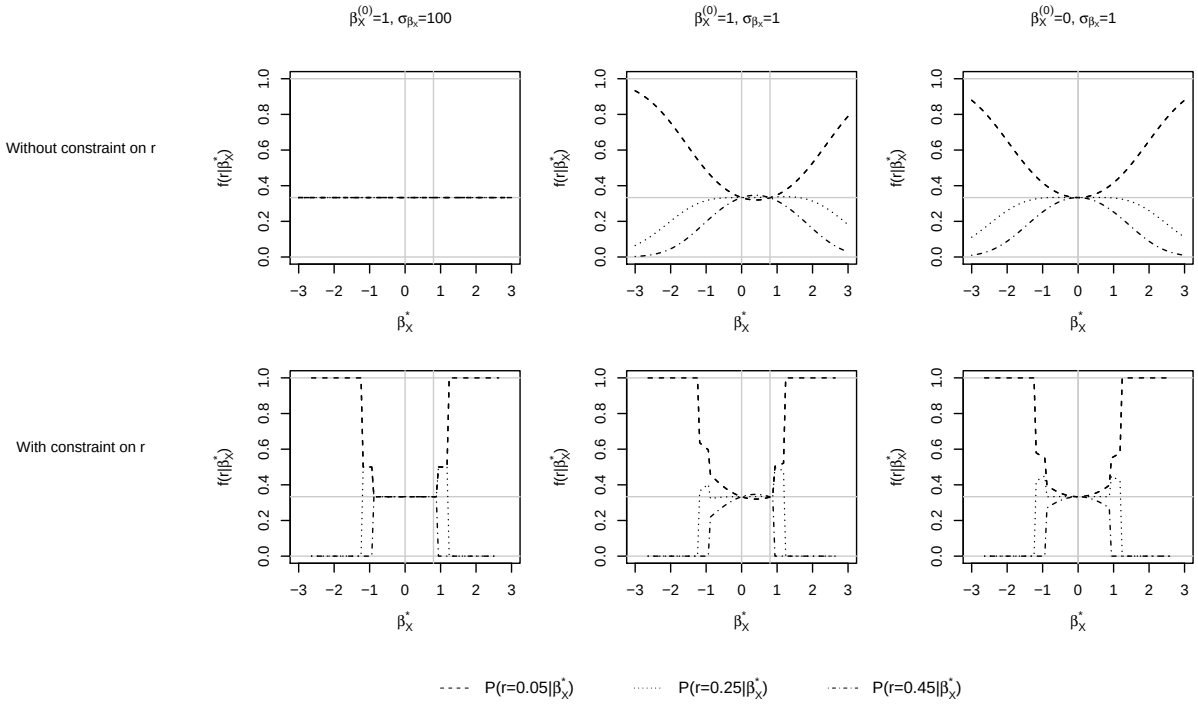


FIGURE 5.13.: *Limiting posterior probabilities for r for varying prior information of β_X contrasting the case with constraints on r and the case without constraints. $P(r = 0.05|\beta_X^*)$: dashed lines, $P(r = 0.25|\beta_X^*)$: dotted lines, $P(r = 0.45|\beta_X^*)$: dotdashed lines. Horizontal gray lines: 0, $\frac{1}{3}$, 1; vertical lines: 0, $(1 + 0.25)\beta_X^{(0)}$.*

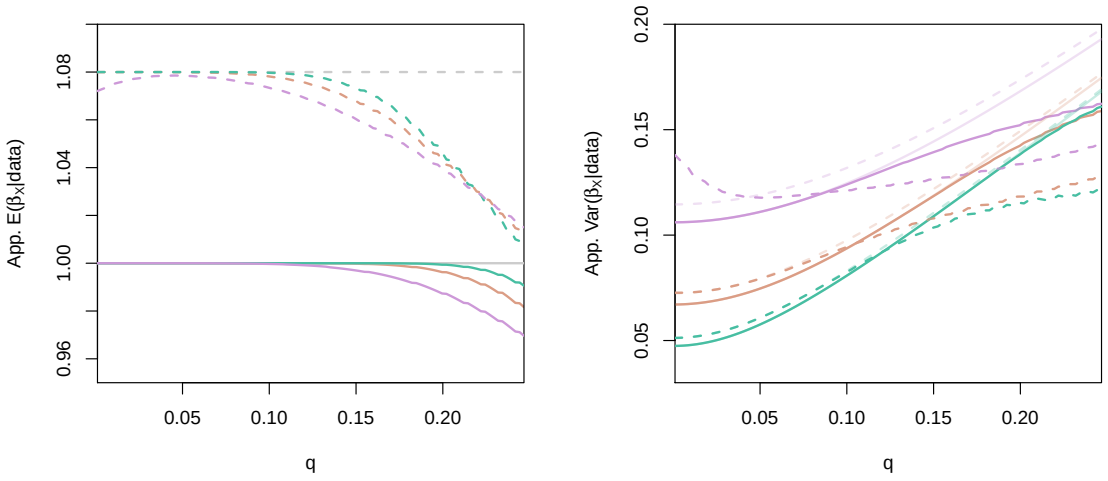


FIGURE 5.14.: *Approximate posterior expectation (left) and variance (right) of β_X for varying prior means of r (solid: $r^{(0)} = 0.25$, dashed: $r^{(0)} = 0.35$) and varying informativeness of the prior distribution of r based on 50 (purple), 125 (green) and 250 (brown) observations and approximate NILP; gray and lightened colors: approximate posterior mean and variance in the NIL case without constraints.*

negatively influences the posterior expectation of β_X and may yield reduced posterior variances of β_X if less informative prior distributions for r are chosen.

Under the chosen constraints, extreme values in the upper domain of r are a posteriori less probable (Figure 5.13), yielding $\mathbb{E}(r|\beta_X^*) < \mathbb{E}(r)$ and $\mathbb{E}(\beta_X|r, \beta_X^*) < \beta_X^{(0)}$; further, $r|\beta_X^*$ and thus, also $\beta_X|$data are less variable.

5.3.6. Indirect learning induced by constraints on parameters in the exposure model

In the previous subsections we focused our considerations on scalar $\boldsymbol{\phi}_N$. However, the measurement error in the Augsburger Umweltstudie is more complex and is described with several parameters. Indeed, the prior knowledge about these non–identifiable parameters may be independently specified, but the samples from the conditional posterior distributions may depend on each other. The reason is that the non–identifiable parameters may underlay constraints, as e.g. $\sigma^2_{U^{\mathrm{B}}}$ and $\sigma^2_{U^{\mathrm{C}}}$, because $\sigma^2_{X^{*\mathrm{B}}} + \sigma^2_{U^{\mathrm{B}}} + \sigma^2_{U^{\mathrm{C}}} = \sigma^2_{X^{*\mathrm{C}}}$ (in a situation without individual–specific measurement errors).

In eq. 5.17, an alternative parametrization of the measurement error is presented, which is used for this subsection to illustrate IL within the non–identifiable parameters. In addition, we consider a Berkson–type error:

$$\begin{aligned} X^{*\mathrm{C}}|X, Y &\sim \mathrm{N}(X, \sigma^2_{U^{\mathrm{C}}}) \\ Y|X &\sim \mathrm{N}(\beta_0 + \beta_X X, \sigma^2_\varepsilon) \\ X|X^{*\mathrm{B}} &\sim \mathrm{N}(X^{*\mathrm{B}}, \sigma^2_{U^{\mathrm{B}}}) \\ X^{*\mathrm{B}} &\sim \mathrm{N}(\mu^{X^{*\mathrm{B}}}, \sigma^2_{X^{*\mathrm{B}}}). \end{aligned}$$

In this case, the identifiable parameters are calculated as

$$\begin{aligned} \beta_0^* &= \beta_0 + \mu^{X^{*\mathrm{B}}}(1 - \lambda^{\mathrm{C}})\beta_X \\ \beta_X^* &= \lambda^{\mathrm{C}}\beta_X \\ \sigma^2_{\varepsilon^*} &= \sigma^2_\varepsilon + \beta_X^2 \sigma^2_{U^{\mathrm{C}}} \lambda^{\mathrm{C}} \\ \mu^{X^{*\mathrm{B}*}} &= \mu^{X^{*\mathrm{B}}} \\ \sigma^{2\ *}_{X^{*\mathrm{B}}} &= \sigma^2_{X^{*\mathrm{B}}} \\ \sigma^{2\ *}_{X^{*\mathrm{C}}} &= \sigma^2_{X^{*\mathrm{C}}} \end{aligned}$$

with $\lambda^{\mathrm{C}} = \frac{\sigma^2_{X^{*\mathrm{B}}} + \sigma^2_{U^{\mathrm{B}}}}{\sigma^2_{X^{*\mathrm{B}}} + \sigma^2_{U^{\mathrm{B}}} + \sigma^2_{U^{\mathrm{C}}}}$ and $\boldsymbol{\phi}_N = (\sigma^2_{U^{\mathrm{B}}}, \sigma^2_{U^{\mathrm{C}}})$. The conditional prior distribution of $\boldsymbol{\phi}_N|\boldsymbol{\phi}_I$ is

$$\begin{aligned} f(\sigma^2_{U^{\mathrm{B}}}, \sigma^2_{U^{\mathrm{C}}}|\beta_0^*, \beta_X^*, \sigma^2_{\varepsilon^*}, \mu^{X^{*\mathrm{B}*}}, \sigma^{2\ *}_{X^{*\mathrm{B}}}, \sigma^{2\ *}_{X^{*\mathrm{C}}}) \\ \overset{\text{NILP}}{\approx} f_{\sigma^2_{U^{\mathrm{B}}}}(\sigma^2_{U^{\mathrm{B}}}) f_{\sigma^2_{U^{\mathrm{C}}}}(\sigma^2_{U^{\mathrm{C}}}) \cdot \\ \mathbf{I}_{\{\sigma^2_{U^{\mathrm{B}}}, \sigma^2_{U^{\mathrm{C}}} | \sigma^{2\ *}_{X^{*\mathrm{C}}} - \sigma^{2\ *}_{X^{*\mathrm{B}}} = \sigma^2_{U^{\mathrm{B}}} + \sigma^2_{U^{\mathrm{C}}}\}}(\sigma^2_{U^{\mathrm{B}}}, \sigma^2_{U^{\mathrm{C}}}) \cdot \\ \mathbf{I}_{\{\sigma^2_{U^{\mathrm{B}}} | 0 \le \sigma^2_{U^{\mathrm{B}}}\}}(\sigma^2_{U^{\mathrm{B}}}) \mathbf{I}_{\{\sigma^2_{U^{\mathrm{C}}} | 0 \le \sigma^2_{U^{\mathrm{C}}}\}}(\sigma^2_{U^{\mathrm{C}}}) \cdot \text{const.} \end{aligned}$$

if IL from the prior distributions of $\boldsymbol{\Theta}_{-\boldsymbol{\phi}_N}$ is prevented, which can only approximately be achieved. As an example, suppose $\sigma^2_{X^{*\mathrm{C}}} = 0.56$, $\sigma^2_{X^{*\mathrm{B}}} = 0.2$, $\sigma^2_{U^{\mathrm{C}}} = 0.06$ and $\sigma^2_{U^{\mathrm{B}}} = 0.2$. We use truncated normal priors for $\sigma^2_{U^{\mathrm{B}}}$ and $\sigma^2_{U^{\mathrm{C}}}$

$$\begin{aligned} f_{\sigma^2_{U^{\mathrm{B}}}}(\sigma^2_{U^{\mathrm{B}}}) &\propto f_N(\sigma^2_{U^{\mathrm{B}}}) \mathbf{I}_{\{\sigma^2_{U^{\mathrm{B}}} | 0 \le \sigma^2_{U^{\mathrm{B}}} \le \sigma^2_{X^{*\mathrm{C}}} - \sigma^2_{X^{*\mathrm{B}}}\}}(\sigma^2_{U^{\mathrm{B}}}) \\ f_{\sigma^2_{U^{\mathrm{C}}}}(\sigma^2_{U^{\mathrm{C}}}) &\propto f_N(\sigma^2_{U^{\mathrm{C}}}) \mathbf{I}_{\{\sigma^2_{U^{\mathrm{C}}} | 0 \le \sigma^2_{U^{\mathrm{C}}} \le \sigma^2_{X^{*\mathrm{C}}} - \sigma^2_{X^{*\mathrm{B}}}\}}(\sigma^2_{U^{\mathrm{C}}}) \end{aligned}$$

with expectations 0.06 and 0.2 and standard deviations 0.03 and 0.1. In order to meet the condition

$$\sigma^2_{X^{*\mathrm{C}}} - \sigma^2_{X^{*\mathrm{B}}} = \sigma^2_{U^{\mathrm{B}}} + \sigma^2_{U^{\mathrm{C}}}$$

marginal and conditional prior distributions differ, possibly involving modified conditional prior expectations (see Figure 5.15).

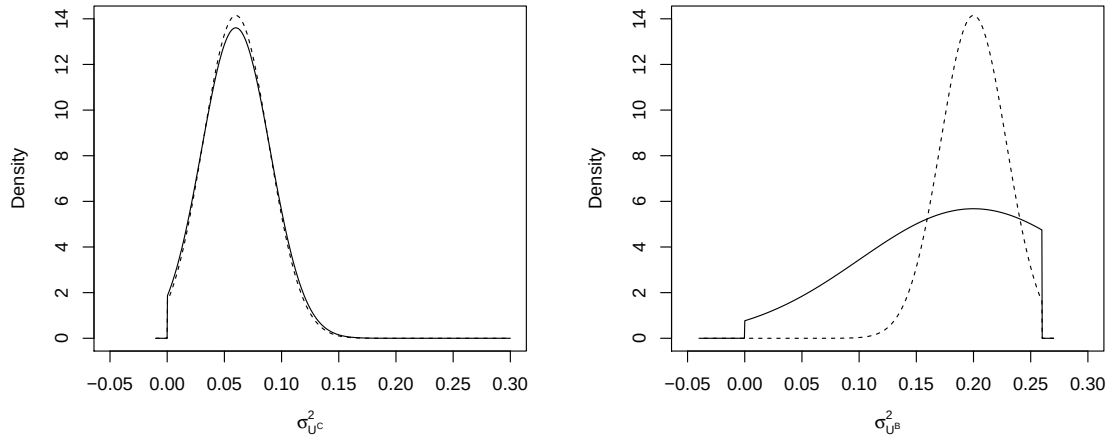


FIGURE 5.15.: *Marginal (solid lines) and conditional (dashed lines) prior densities of* $\sigma^2_{U^C}$ *and* $\sigma^2_{U^B}$.

The information described by the conditional prior distributions of $\boldsymbol{\phi}_N$ is used for the Bayesian analysis instead of the initially intended knowledge specified by the marginal prior distributions. This learning effect can only be prevented through a deliberate choice of the prior distributions on $\boldsymbol{\phi}_N$ including also the constraints since the dependency of the prior distributions originates in the design of the models. In comparison to Subsection 5.3.5, the impact of constraints on $\boldsymbol{\phi}_N$ does not depend on the data or the prior distributions regarding $\boldsymbol{\Theta}_{-\boldsymbol{\phi}_N}$.

5.4. Simulation study: Properties of Bayesian measurement error models

In Chapter 4 we derive method–of–moment correction formulas for the measurement error correction in linear mixed models with AR(1) errors. Bayesian methods on the same simulation example as in Section 3.6 are examined in this section. Again, 100 data sets are generated for each size of measurement error; the data are centralized for the Bayesian analyses.

Besides classical (C) and Berkson (B) measurement error, a mixture (M) of both error types is considered: in addition to the incomplete values of the personal exposure measurements fixed–site measurements are included in the analysis. Note, that the settings for mixture error for Bayesian analyses comprise more data than for the frequentist analyses: both data sources are used for Bayesian analyses. In the following simulations different aspects of Bayesian regression models for measurement error are considered and compared to the frequentist models. Results are also given for $SF = 0$ for the sake of completeness, but bear in mind that the model specification is different for computational reasons.

Non–informative normal prior distributions are used for the regression coefficients (β_0, β_X) of the main model and non–informative Uniform prior distributions are chosen for σ_τ^2, σ_ε^2 and ρ and, if necessary, for the variance of X, σ_X^2, and the autocorrelation parameter $\rho^{X^{*\mathrm{B}}}$:

- Effect coefficients: $\beta_0 \sim \mathrm{N}(0, 10^6), \beta_X \sim \mathrm{N}(0, 10^6)$
- Variances: $\sigma_\varepsilon^2 \sim \mathrm{Unif}(0, 100)$, $\sigma_\tau^2 \sim \mathrm{Unif}(0, 100)$, $\sigma_X^2 \sim \mathrm{Unif}(0, 100)$
- Autocorrelation coefficients: $\rho^{X^{*\mathrm{B}}} \sim \mathrm{Unif}(0, 1)$, $\rho \sim \mathrm{Unif}(0, 1)$

The comparison of Bayesian with MOM–corrected effect estimates is accomplished with fixed prior knowledge regarding the parameters characterizing the measurement error; the results are presented in Subsection 5.4.1. In the subsequent Subsections 5.4.2–5.4.5 the lack of knowledge about the exposure model in the analysis of classical error–prone data, of a varying fraction of missing personal measurements, of simultaneous estimation of missing values and of the sensitivity to the choice of the prior distribution for the parameters characterizing the measurement error on the effect estimations will be examined.

The Bayesian analyses in the present section are based on 5000 MCMC iterations with additional 1000 discarded burnin–iterations.

5.4.1. Comparison to frequentist estimation

Parameters characterizing the measurement error are assumed to be known and observations with missing values of $\boldsymbol{X}^{*\mathrm{C}}$ are excluded in models for classical measurement error–prone data. Further, the exposure model for the analysis of $\boldsymbol{X}^{*\mathrm{C}}$ is correctly specified, i.e. $\boldsymbol{X}$ is specified as a sum of the autocorrelated Berkson error prone measurements $\boldsymbol{X}^{*\mathrm{B}}$, (if reasonable) $\boldsymbol{\nu}^{\mathrm{B}}$ and (if reasonable, the autocorrelated) Berkson errors $\boldsymbol{U}^{\mathrm{B}}$.

The results for the Scenarios (I)–(III) are depicted in Appendix, Figure C.1 and the corresponding values for bias and RMSE are listed in Table 5.3 for Bayesian analyses and in Table 3.2 for the frequentist case.

As expected, the joint usage of the Berkson and classical error–prone measurements yields better estimates with regard to RMSE than using only a single data source, if $\sigma_{U^{\mathrm{B}}}^2, \sigma_{U^{\mathrm{C}}}^2 > 0$. Therefore, including both data sources always should be preferred, especially considering the high bias occurring with a misspecified exposure model in the model for classical error–prone data (Subsection 5.4.2).

Bayesian and frequentist error correction yield comparable results for the estimates of the effect coefficients in Scenarios (I) and (II); a favored method cannot be identified by the means of RMSE values. While the results are also similar for Berkson and classical error–prone data in Scenario (III), the Bayesian method for mixture error provides better RMSE values than the frequentist method for mixture error.

	SF of	Truth		Berkson		Classical		Mixture	
Scen.	$\sigma^2_{U\mathrm{B}}, \sigma^2_{U\mathrm{C}}$	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
(I)	0	0.016	0.363	**0.016**	**0.363**	0.019	0.422	0.016	**0.363**
	0.5	0.024	0.242	-0.016	0.406	**0.012**	0.295	0.014	**0.276**
	1	0.004	0.186	**-0.002**	0.380	0.024	0.248	0.013	**0.234**
	2	0.003	0.138	0.003	0.374	**-0.002**	0.183	-0.003	**0.177**
	5	-0.012	0.096	0.015	0.393	**-0.008**	**0.117**	-0.010	0.117
(II)	0	0.069	0.385	**0.071**	**0.392**	0.078	0.408	**0.071**	0.392
	0.5	-0.038	0.237	-0.063	0.366	**-0.034**	0.285	-0.048	**0.261**
	1	0.015	0.178	0.015	0.403	-0.006	0.239	**-0.001**	**0.225**
	2	0.031	0.150	**0.007**	0.376	0.026	0.185	0.028	**0.180**
	5	-0.006	0.092	-0.037	0.396	-0.010	0.116	**-0.009**	**0.115**
(III)	0	-0.011	0.654	**-0.006**	0.674	-0.055	0.749	-0.007	**0.671**
	0.5	0.033	0.273	**-0.005**	0.618	0.101	0.424	0.046	**0.374**
	1	-0.013	0.199	0.005	0.673	0.034	0.296	**0.004**	**0.272**
	2	-0.009	0.166	0.018	0.675	0.012	0.202	**0.002**	**0.196**
	5	0.009	0.097	0.109	0.680	0.033	0.147	**0.025**	**0.145**

TABLE 5.3.: *Bias and RMSE of the Bayesian effect estimations for Scenarios (I)-(III); bold: lowest absolute bias and lowest RMSE of each row (disregarding the results based on the true data).*

5.4.2. Lacking knowledge about the exposure model

In the previous subsection the exposure model in the Bayesian model for the classical measurement error was correctly specified, i.e. just as in the data generating process. However, the exposure model is in practice rather complex and often unknown. In particular when personal measurements are used, the exposure model depends on individual–specific parameters whereof the knowledge from validation studies cannot be transferred. Therefore, we examine the Bayesian models for classical error–prone data without prior knowledge about the structure and the parameters of the exposure model. To this end, we use a flat prior for the variance of X: $\sigma_X^2 \sim \text{Unif}(0, 100)$. Thus, the true exposure, which actually shows an individual–specific structure and autocorrelation, is misspecified.

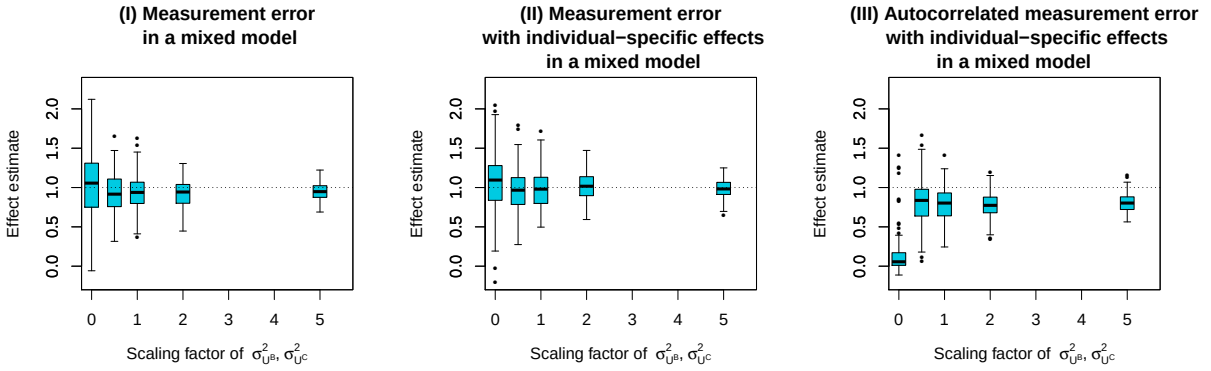


FIGURE 5.16.: *Bayesian estimates of corrected regression coefficients for increasing size of measurement errors based on measurements with classical measurement error and with lacking knowledge of the exposure model.*

The estimates for the classical error–prone data are biased, especially in Scenario (III). Of course, also the lack of knowledge regarding the Berkson error model in the analysis of both data sources would result in biased estimates due to the same reasons; in this case, the specification of a model for the fixed–site measurements is not needed.

Two aspects are relevant for the explanation of these results. First, IL about σ_X^2 occurs through the constraint $\sigma_X^2 + \sigma_{U^C}^2 = \sigma_{X^{*C}}^2$ (see Subsection 5.3.6) yielding an informative limiting posterior distribution of σ_X^2. Second, the attenuation of the effect estimate is insufficiently corrected if the autocorrelation structure of $\boldsymbol{X}^{*B}$ is neglected, as it becomes obvious for Scenario (I) using the formulas for calculating the attenuation factor derived in Subsection 3.3.3:

$$\frac{\sigma_{X^{*B}}^2 g_T^{*C}(\rho^{X^{*B}}) + \sigma_{U^B}^2}{\sigma_{X^{*B}}^2 g_T^{*C}(\rho^{X^{*B}}) + \sigma_{U^B}^2 + \sigma_{U^C}^2} \neq \frac{\sigma_{X^{*B}}^2 + \sigma_{U^B}^2}{\sigma_{X^{*B}}^2 + \sigma_{U^B}^2 + \sigma_{U^C}^2}$$

for $\rho^{X^{*B}} > 0$.

5.4.3. Varying fractions of missing personal measurements

Previously (Subsection 5.4.1), we discovered benefits of using both data sources in all scenarios in comparison to using only one data source. Since this benefit strongly depends on the fraction of missing personal measurements we will further evaluate this relation in the present subsection for scaling factor 1. The parameters characterizing the measurement error are again assumed to be known. Observations with missing values of $\boldsymbol{X}^{*C}$ are excluded and the exposure model is correctly specified in models for classical measurement

	SF of	Truth		Berkson		Classical		Mixture	
Scen.	$\sigma^2_{U\mathrm{B}}, \sigma^2_{U\mathrm{C}}$	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
(I)	0	0.012	0.181	0.035	0.369	0.019	0.200	**0.018**	**0.198**
	0.2	0.020	0.196	0.041	0.375	**0.033**	0.226	0.036	**0.220**
	0.4	-0.022	0.180	**-0.012**	0.368	-0.018	0.252	-0.022	**0.231**
	0.6	0.009	0.189	**0.003**	0.325	0.018	0.343	0.016	**0.268**
	0.8	0.019	0.191	0.040	0.391	**0.000**	0.458	0.012	**0.294**
	1	-0.047	0.214	**-0.015**	**0.320**			-0.015	0.320
(II)	0	-0.005	0.223	0.011	0.388	**0.002**	0.244	0.004	**0.242**
	0.2	0.020	0.189	**0.003**	0.382	0.030	0.230	0.026	**0.220**
	0.4	-0.002	0.196	0.065	0.406	-0.018	0.276	**-0.005**	**0.271**
	0.6	-0.003	0.179	0.038	0.350	**-0.001**	0.312	0.011	**0.244**
	0.8	-0.004	0.216	-0.055	0.401	**0.007**	0.443	-0.025	**0.322**
	1	0.012	0.169	0.025	0.384			**0.023**	**0.381**
(III)	0	-0.019	0.181	-0.095	0.648	**-0.007**	0.240	-0.035	**0.230**
	0.2	-0.023	0.208	-0.035	0.667	**-0.001**	0.276	-0.028	**0.258**
	0.4	-0.014	0.192	-0.039	0.651	0.019	0.317	**-0.012**	**0.298**
	0.6	-0.025	0.231	-0.039	0.628	**0.008**	0.408	-0.031	**0.384**
	0.8	0.010	0.204	0.072	0.700	0.112	0.506	**0.062**	**0.431**
	1	-0.019	0.223	**0.134**	**0.650**			**0.134**	**0.650**

TABLE 5.4.: *Bias and RMSE of the Bayesian effect estimations for Scenarios (I)–(III) with varying percentage of missing personal measurements (p); bold: lowest absolute bias or lowest RMSE of each row (disregarding the results based on the true data).*

error–prone data. The results for the Scenarios (I)–(III) are depicted in Appendix, Figure C.2 and the corresponding values for bias and RMSE are listed in Table 5.4.

The joint usage of Berkson and classical error–prone measurements out–performs the usage of only classical error–prone, incomplete measurements for all analyzed fractions of missing values, even for $p = 0$. Against the background that the underlying exposure model is implicitly estimated by the Bayesian regression model for classical measurement error, the benefit of any additional information for the exposure model can easily be recognized.

Further, the results indicate, that personal measurements with a high percentage of missing values provide worse results than complete Berkson error–prone measurements for Scenarios (I) and (II).

5.4.4. Impact of simultaneous estimation of missing personal measurements

In the Subsections 5.4.1–5.4.3 we did not estimate the missing values of $\boldsymbol{X}^{*\mathrm{C}}$ within the Bayesian analysis. We will analyze in the present subsection to what extent the simultaneous estimation of missing values will affect the results in Scenario (I). The parameters characterizing the measurement error are again assumed to be known and the exposure model is correctly specified in models with classical error–prone data.

The simultaneous prediction of missing values in the models for classical error–prone data provides a slight improvement of the effect estimates (Figure 5.5). The result is only valid for a correctly specified exposure model and cannot be transferred to situations with misspecified or unknown exposure model.

	SF of	Without prediction		With prediction	
	$\sigma^2_{U^{\mathrm{B}}}, \sigma^2_{U^{\mathrm{C}}}$	Bias	RMSE	Bias	RMSE
Classical	0	0.019	0.422	**0.015**	**0.361**
	0.5	0.012	0.295	**0.010**	**0.279**
	1	0.024	0.248	**0.022**	**0.235**
	2	**-0.002**	0.183	-0.002	**0.182**
	5	**-0.008**	**0.117**	-0.009	0.117
Mixture	0	**0.016**	**0.363**	**0.016**	**0.363**
	0.5	**0.014**	**0.276**	**0.014**	**0.276**
	1	**0.013**	**0.234**	**0.013**	**0.234**
	2	**-0.003**	**0.177**	**-0.003**	**0.177**
	5	**-0.010**	**0.117**	**-0.010**	**0.117**

TABLE 5.5.: *Bias and RMSE of the Bayesian effect estimations for Scenario (I) with and without prediction of missing values in $\boldsymbol{X}^{*C}$; bold: lowest absolute bias or lowest RMSE of each row.*

The results are equal for the models with and without predicting the missing values of $\boldsymbol{X}^{*\mathrm{C}}$ when both measurement sources are included in the analysis.

5.4.5. Sensitivity to prior parameter choices for the non–identifiable parameters

Setup

In regression models with a measurement error depending only on a single parameter Gustafson (2005) found out that less variable prior knowledge results in a reduced precision of the estimation. In contrast, we demonstrated in Section 5.3 that these observations may originate from indirect learning from the prior distribution on β_X; further, considerations with conditionally conjugate prior distributions in Subsection 5.1.3 reveal, that the variability of the prior knowledge is positively associated with the marginal posterior variability of the effect coefficient. Simulations with the following priors back up these results:

(I) $\sigma^2_{U^{\mathrm{C}}} \sim \mathrm{Unif}(0.9 \cdot \sigma^{2\,(0)}_{U^{\mathrm{C}}}, 1.1 \cdot \sigma^{2\,(0)}_{U^{\mathrm{C}}})$

(II) $\sigma^2_{U^{\mathrm{C}}} \sim \mathrm{Unif}(0.1 \cdot \sigma^{2\,(0)}_{U^{\mathrm{C}}}, 1.9 \cdot \sigma^{2\,(0)}_{U^{\mathrm{C}}})$

(III) $\sigma^2_{U^{\mathrm{C}}} \sim \mathrm{IG}(302.0, 60.2)$

(IV) $\sigma^2_{U^{\mathrm{C}}} \sim \mathrm{IG}(5.70, 0.94)$

The parameters in (III) and (IV) are chosen in that way, that the prior expectation equals $\sigma^{2\,(0)}_{U^{\mathrm{C}}}$ and the prior variance is equal to case (I) and (II), respectively, in order to analyze the effect of the shape of the measurement error prior distribution. The prior distributions (I)–(IV) are depicted in Figure 5.18 and Figure 5.19.

Data without autocorrelation within the measurement errors and the model error term is considered and the sensitivity to the prior choice in a simple linear mixed model is examined (Scenario (I)); the scaling factor for the measurement error variances $\sigma^2_{U^{\mathrm{B}}}$ and $\sigma^2_{U^{\mathrm{C}}}$ is fixed to 1. The latent parameters despite $\sigma^2_{U^{\mathrm{B}}}$ in the models including $\boldsymbol{X}^{*\mathrm{B}}$ and $\sigma^2_{U^{\mathrm{C}}}$ in the models including $\boldsymbol{X}^{*\mathrm{C}}$ are assumed to be known and the exposure model is correctly specified.

Bias, variance and RMSE

As already deviated from theoretical considerations in the previous sections, shape and variability of the prior distribution will not have any impact on $\mathbb{E}(\widehat{\beta_X})$ and $\mathrm{Var}(\widehat{\beta_X})$, if IL is prevented, as it is also visible in Table 5.6.

	interval $\cdot\sigma^2_{U\mathrm{B}}/\sigma^2_{U\mathrm{C}}$	Berkson		Classical		Mixture	
Prior	with 95 % prob. mass	Bias	RMSE	Bias	RMSE	Bias	RMSE
Point		-0.002	**0.380**	0.024	0.248	**0.013**	**0.234**
Gamma	[0.9,1.1]	-0.003	0.386	0.024	0.247	0.013	0.234
Gamma	[0.7,1.3]	-0.003	0.381	**0.023**	0.246	0.016	0.235
Unif	[0.9,1.1]	-0.002	0.381	0.024	0.246	0.014	0.234
Unif	[0.7,1.3]	**0.001**	0.382	0.025	**0.246**	0.015	0.236

TABLE 5.6.: *Bias and RMSE of the Bayesian effect estimations for Scenario (I) with varying shape and variability of the prior distributions for the sizes of the measurement errors; bold: lowest absolute bias or lowest RMSE of each column.*

Indeed, $\mathrm{Var}(\beta_X|\mathrm{data})$ depends on the type and the variability of the prior distributions, but the impact of different prior distributions for the measurement error on the effect estimations in the simulations is diminutive. Therefore, we demonstrate the dependency between $\mathrm{Var}(\beta_X|\mathrm{data})$ and the prior parameters with the approximate properties deviated in Section 5.3.

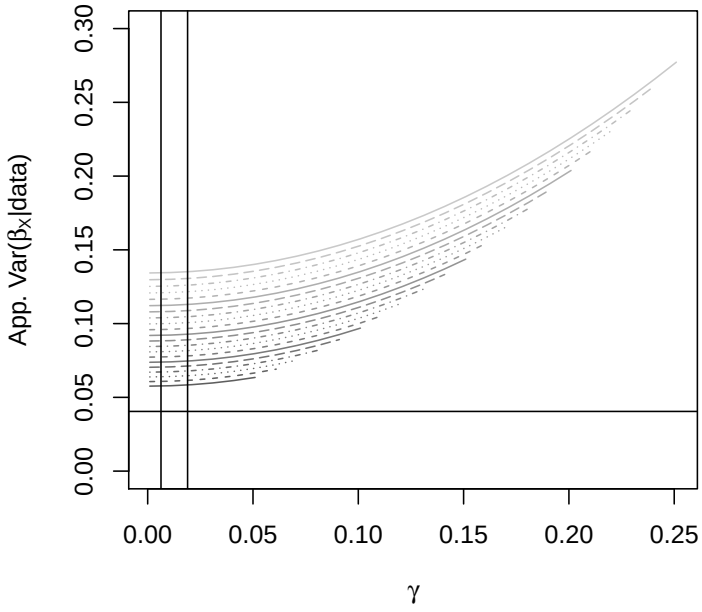


FIGURE 5.17.: *Approximate values for $Var(\beta_X|data)$ according to eq. 5.14 for a regression analysis with covariate $\boldsymbol{X}^{*C}$ with the latent parameters despite σ^2_{UC} assumed to be known and with Unif$(\sigma^2_{UC}{}^{(0)}-\gamma, \sigma^2_{UC}{}^{(0)}+\gamma)$ as the prior distribution for σ^2_{UC}; $\sigma^2_{UC}{}^{(0)}$ varies from 0.001 to 0.26 and is depicted with decreasing intensity; horizontal line: variance of the naive effect estimate; vertical lines: values for γ chosen for simulation.*

Figure 5.17 shows, that the approximate posterior variance of β_X changes only slightly for low values of γ when a Unif($\sigma^{2\,(0)}_{U^{\mathrm{C}}} - \gamma, \sigma^{2\,(0)}_{U^{\mathrm{C}}} + \gamma$) prior distribution is used for $\sigma^2_{U^{\mathrm{C}}}$ and increases with increasing γ.

Marginal posterior distributions

The prior choice in the simulation setup prevents IL from the prior distributions of $\boldsymbol{\Theta}_{-\phi_N}$ in the models for Berkson and classical error–prone measurements. For the models with Berkson error–prone covariates, this becomes obvious in the following figures: the posterior distributions for $\sigma^2_{U^{\mathrm{C}}}$ and $\sigma^2_{U^{\mathrm{B}}}$ are very similar to the respective prior distributions.

For the models with classical error–prone covariates and with measurements of both types, IL induced by constraints on the parameters in the exposure model exists, respectively. In the models for both measurement types, IL is stronger for $\sigma^2_{U^{\mathrm{B}}}$ in comparison to $\sigma^2_{U^{\mathrm{C}}}$ and the posterior distributions of $\sigma^2_{U^{\mathrm{C}}}$ and $\sigma^2_{U^{\mathrm{B}}}$ are correlated: the mean absolute Pearson correlation coefficients are higher for less informative priors (-0.59 for the more informative Uniform prior distribution, -0.90 for the less informative Uniform prior distribution, -0.49 for the more informative Inverse Gamma prior distribution, -0.86 for the less informative Inverse Gamma prior distribution).

Sensitivity to prior parameter choices for the autocorrelation coefficient of the measurement error

Whereas the variance of individual–specific measurement error, $\sigma^2_{\nu^{\mathrm{B}}}$ and $\sigma^2_{\nu^{\mathrm{C}}}$, may exhibit similar properties than $\sigma^2_{U^{\mathrm{B}}}$ and $\sigma^2_{U^{\mathrm{C}}}$, we examined the sensitivity to prior parameter choices for the autocorrelation coefficient of the measurement error using simulations. Therefore, Uniform prior distributions are imposed on ρ^{B} and ρ^{C}. Again, bias and RMSE change only slightly; the posterior distribution of ρ^{B} and ρ^{C} highly resemble the respective prior distribution in models for Berkson and classical error–prone measurements and are affected by IL if both data sources are included.

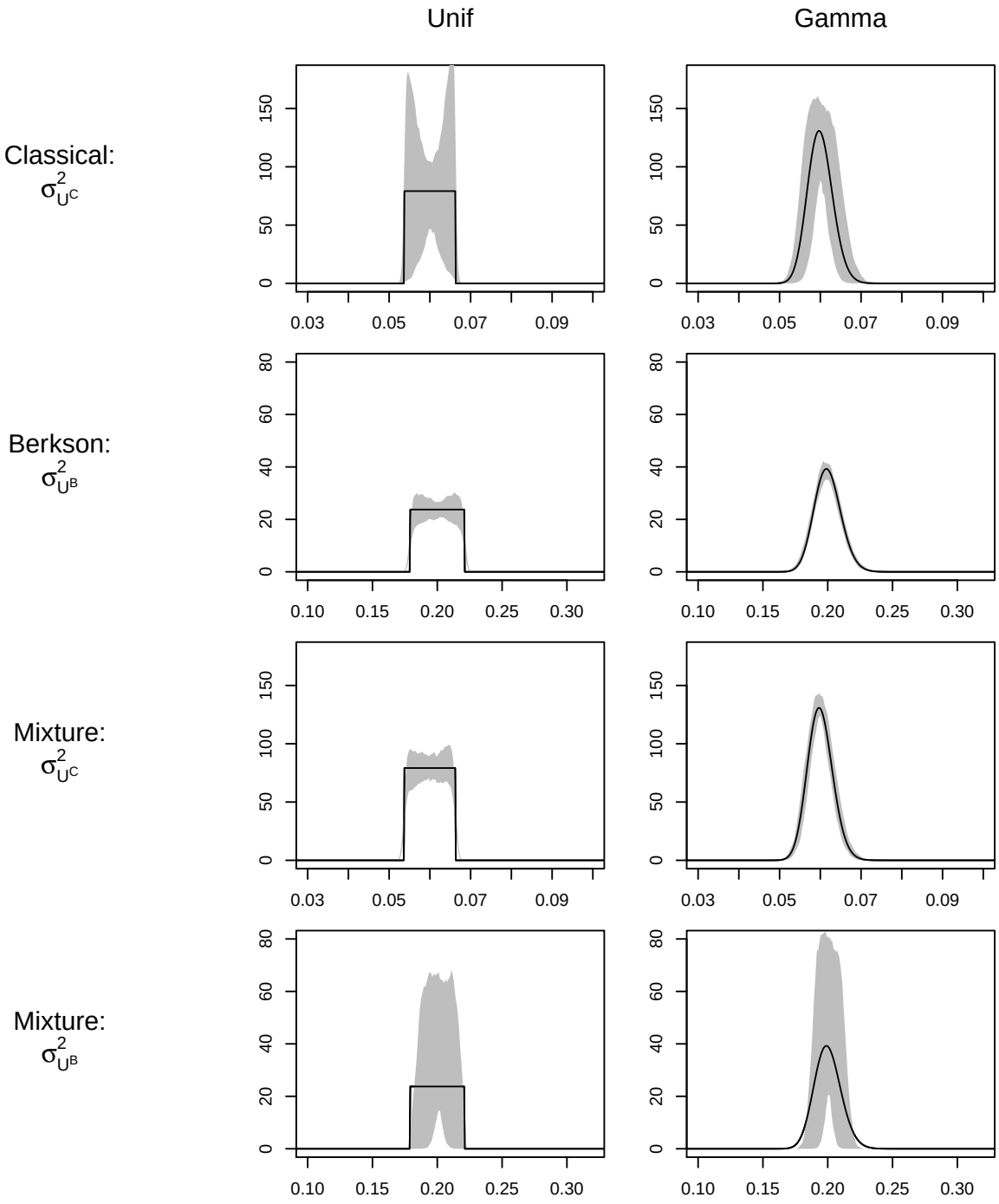


FIGURE 5.18.: *Prior distributions for* σ^2_{UB} *and* σ^2_{UC} *with 95 % of the probability mass between* $0.9\sigma^2_{UB}$ *and* $1.1\sigma^2_{UB}$ *or/and* $0.9\sigma^2_{UC}$ *and* $1.1\sigma^2_{UC}$ *(black); pointwise, empirical 95 % interval of the 100 posterior distributions of* σ^2_{UB} *and* σ^2_{UC} *(gray).*

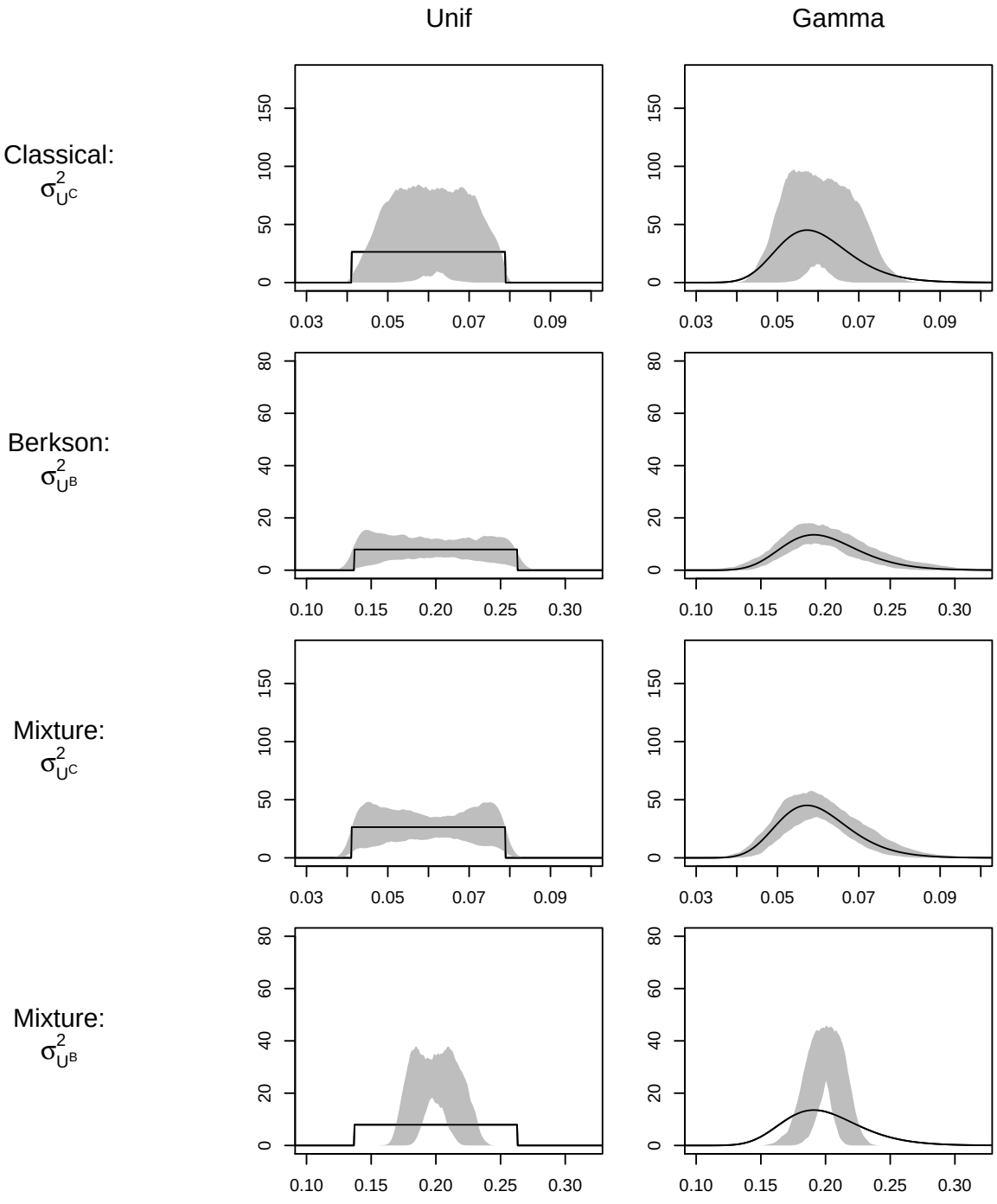


FIGURE 5.19.: *Prior distributions for σ^2_{UB} and σ^2_{UC} with 95 % of the probability mass between $0.7\sigma^2_{UB}$ and $1.3\sigma^2_{UB}$ or/and $0.7\sigma^2_{UC}$ and $1.3\sigma^2_{UC}$ (black); pointwise, empirical 95 % interval of the 100 posterior distributions of σ^2_{UB} and σ^2_{UC} (gray).*

5.5. Conclusions

Considerations with conditionally conjugate prior distributions and the transparent reparametrization according to Gustafson (2005) reveal, that the prior expectations regarding the non–identifiable parameters equal the respective posterior expectations, as seen also in Figure 5.5, if information about the measurement error solely originates from the informative prior distributions for the non–identifiable parameters (i.e. IL is prevented). Therefore, the expectations of the prior distributions for the non–identifiable parameters are the decisive parameters for the posterior expectation of the effect estimate. Even the marginal and conditional prior distributions and the conditional posterior distribution of the non–identifiable parameters are equal in that case. In many situations, this property is desirable, because the prior distribution for the non–identifiable parameters provides the only source of information about the measurement error. The considerations with conditionally conjugate prior distributions demonstrate, that the main model and the exposure model deliver additional information about the measurement error. We do not dwell on the former issue because the influence of these information sources are negligible in the data situation of the Augsburger Umweltstudie; we implicitly discussed the latter issue in the simultaneous analysis of personal and fixed–site measurements. The properties of the prior distribution for the non–identifiable parameters, e.g. mean and variability, are directly incorporated in the estimation of the regression coefficients in the absence of indirect learning. The shape of the prior distributions of the non–identifiable parameters only affects the shape of the posterior distribution of the effect estimate and thus also confidence intervals of the effect estimates, but not its expectation (Section 5.3). The variability of the estimated true data $\boldsymbol{X}$ is higher than its actual variability through the additional variability due to prior uncertainty regarding the size of the measurement error if the posterior mean of the size of the measurement error is positive.

Three types of indirect learning are to be differentiated: (I) IL from prior distributions on the parameters of the main outcome model, (II) IL from constraints on parameters of the main outcome model and (III) IL from constraints on parameters of the exposure model. The first type of IL is often undesirable and can be prevented through an adequate prior choice for the parameters of the main model. The second type of IL will affect the posterior distribution of the effect estimates if few prior knowledge about the non–identifiable parameters is available and at the same time, the variability of the identifiable parameters is high, e.g. through a small sample size. The third type of IL occurs in Bayesian regression models for measurement error, which include an exposure model or more than one non–identifiable parameters. IL type (III) may be volitional, but the analyst should be aware of the mutual interaction between the prior assumptions within the Bayesian model calculation; for instance, nescience about the exposure model in a Bayesian regression model with a measurement error–prone covariate results in an informative conditional prior distribution due to IL learning type (III) which may raise biases due to the erroneous specification of the exposure model (see the simulations results in Subsection 5.4.2). Further, IL type (III) cannot be alleviated by increasing the sample size of the main study in contrast to IL type (II). In general, indirect learning is highly relevant in situations with little knowledge about the size of the measurement error. A major difference to identifiable Bayesian regression models is, that it is not possible to specify a Bayesian regression model for error–prone covariates through non–informative prior distributions where the data speaks for itself (except further data is included describing the measurement error), because IL occurs. Since Bayesian regression models for error–prone covariates always comprise prior knowledge about the parameters describing

the measurement error, albeit through constraints, inference with the non–identifiable model is possible.

On the basis of these findings two steps should be conducted for the prior specification of a Bayesian regression model with measurement error–prone covariates:

1. Specify the prior distributions for the non–identifiable parameters on the basis of the prior knowledge regarding the shape and location.

2. Choose the prior parameters of $\boldsymbol{\Theta}_{-\phi_N}$ in a way that prevents undesired indirect learning effects

   a) through non–informative priors for the parameters of the main model to prevent IL of type (I)

   b) through enough information about the measurement error and through an adequate sample size to prevent IL of type (II)

   c) through the consistent specification of the prior distributions for all parameters describing the exposure model and the measurement error to prevent IL of type (III).

   The similarity of the prior and the posterior distribution of the parameters describing the exposure model and the measurement error indicates the success of preventing IL, e.g. the KL distance between the distributions may be used.

The focus of this chapter was on the adequate specification of a prior knowledge on the basis of expert knowledge. However, the prior knowledge about the measurement error may also originate from a preceding analysis, as e.g. in Chang et al. (2011), Gryparis et al. (2009) and Prescott and Garthwaite (2005). We showed, that this approach is adequate, if IL is prevented. The argumentation can also be extended to the usage of prior information gained from the simultaneous modeling of validation data. Several sources for prior information within a single model result in IL of type (III).

6. Measurement error correction for the Augsburger Umweltstudie

The method of moments correction and the Bayesian approach for measurements with Berkson, classical and mixture measurement error in a mixed model framework is applied to the models of the Augsburger Umwelstudie by using the data of the validation and comparison studies described in Section 1.3. For the main model of the Augsburger Umweltstudie, the 5–min. resolved heart rate is the health outcome and the corresponding (log–transformed) personal PNC measurements are regarded as the deficient covariate $X^{*\mathrm{C}}$; the fixed–site PNC measurements are considered as the Berkson error–prone covariate. The confounder model established for the analysis of the Augsburger Umweltstudie (Hampel et al., 2012b) is used, which includes a linear time trend, linear and quadratic effects of 2–h. lagged temperature at the measurement station, a linear effect of 1–h. lagged relative humidity and a binary variable for time of the day (before/after noon).

Due to differences in toxicity and thus also in health effects Wilson and Brauer (2006) argue to separately analyze personal ambient and personal non–ambient exposure. Further, the indoor measurements differ in the size, complexity and structure of Berkson error in comparison to outdoor measurements. Therefore, the main analyses are based on outdoor measurements; indoor measurements are only examined with a sensitivity analysis using the Bayesian approach.

The parameters for the (two–stage) correction of the measurement errors are estimated with validation data as well as with data from the Augsburger Umwelstudie as described in Chapter 2 and are listed in Table 6.1.

	Variance	Autocorrelation
$X^{*\mathrm{B}}$	0.34	0.932
U_{it}^{B}	0.3	0.582
ν_i^{B}	0.21	
U_{it}^{C}	0.03	0.696
ν_i^{C}	0.03	

TABLE 6.1.: *Parameter specifications for application (concurrent mobile and fixed–site measurements).*

6.1. Correction with the method of moments

The log–transformed fixed site measurements are adapted to the mean of the log–transformed individual measurements by adding a constant to fulfill the assumptions for the bias correction defined in Subsection 2.1.2. Varying degrees of lagged effects of the log–transformed PNC concentrations are examined up to a lag of 60 minutes.

Three models are considered: Fixed–site PNC levels are used as covariate in the first model (HMM$^{*\mathrm{B}}$, eq. 3.1). For the second model, only observations with individual measurements are used (HMM$^{*\mathrm{C}}$, eq. 3.2) and for the third model, missing individual PNC

observations are substituted with the values from the fixed–site measurement station (HMM$^{*\mathrm{M}}$, eq. 3.3).

Confidence intervals for the effect coefficients are based on the delta method due to higher coverage probabilities compared to confidence intervals based on bootstrap (Section 4.4). The variance of the attenuation factor, $\mathrm{Var}(\widehat{\lambda})$, is calculated using 100 bootstrap samples of the appropriate data sets: the different visits of the individuals are resampled from the data of the main study and the different periods defined by battery changes are resampled from the data of the comparison measurements.

The correlation structure of the model errors is individual–specific due to individual–specific patterns of missing values and varying measurement duration.

6.1.1. Multivariate analysis

For measurement error correction in the multivariate case, equation system 3.43 is adapted for unbalanced longitudinal data. $\mathrm{tr}(\widehat{\boldsymbol{V}}_i^{*^{-1}}\boldsymbol{X}_i\boldsymbol{X}_i^{*\top})$ is approximated as follows:

$$
\begin{aligned}
\mathrm{tr}(\widehat{\boldsymbol{V}}_i^{*\mathrm{C}^{-1}}\boldsymbol{X}_i\boldsymbol{X}_i^{*\mathrm{C}^\top}) \approx & \mathrm{tr}(\widehat{\boldsymbol{V}}_i^{*\mathrm{C}^{-1}}\boldsymbol{X}_i^{*\mathrm{C}}\boldsymbol{X}_i^{*\mathrm{C}^\top}) - \\
& \mathrm{tr}(\sigma^2_{U^\mathrm{C}}\widehat{\boldsymbol{V}}_i^{*\mathrm{C}^{-1}}\boldsymbol{D}_i\boldsymbol{W}_{\rho^\mathrm{C},i}\boldsymbol{D}_i^\top) - \mathrm{tr}(\sigma^2_{\nu^\mathrm{C}}\widehat{\boldsymbol{V}}_i^{*\mathrm{C}^{-1}}\boldsymbol{D}_i\mathbf{J}_{T_i}\boldsymbol{D}_i^\top) \\
\mathrm{tr}(\widehat{\boldsymbol{V}}_i^{*\mathrm{M}^{-1}}\boldsymbol{X}_i\boldsymbol{X}_i^{*\mathrm{M}^\top}) \approx & \mathrm{tr}(\widehat{\boldsymbol{V}}_i^{*\mathrm{M}^{-1}}\boldsymbol{X}_i^{*\mathrm{M}}\boldsymbol{X}_i^{*\mathrm{M}^\top}) + \\
& \mathrm{tr}(\sigma^2_{U^\mathrm{B}}\widehat{\boldsymbol{V}}_i^{*\mathrm{M}^{-1}}\boldsymbol{D}_i(\mathbf{1}_{T_i} - \boldsymbol{G}_i)\boldsymbol{G}_i^\top \circ \boldsymbol{W}_{\rho^\mathrm{B},i}\boldsymbol{D}_i^\top) + \\
& \mathrm{tr}(\sigma^2_{\nu^\mathrm{B}}\widehat{\boldsymbol{V}}_i^{*\mathrm{M}^{-1}}\boldsymbol{D}_i(\mathbf{1}_{T_i} - \boldsymbol{G}_i)\boldsymbol{G}_i^\top \circ \mathbf{J}_{T_i}\boldsymbol{D}_i^\top) - \\
& \mathrm{tr}(\sigma^2_{U^\mathrm{C}}\widehat{\boldsymbol{V}}_i^{*\mathrm{M}^{-1}}\boldsymbol{D}_i(\mathbf{1}_{T_i} - \boldsymbol{G}_i)(\mathbf{1}_{T_i} - \boldsymbol{G}_i)^\top \circ \boldsymbol{W}_{\rho^\mathrm{C},i}\boldsymbol{D}_i^\top) - \\
& \mathrm{tr}(\sigma^2_{\nu^\mathrm{C}}\widehat{\boldsymbol{V}}_i^{*\mathrm{M}^{-1}}\boldsymbol{D}_i(\mathbf{1}_{T_i} - \boldsymbol{G}_i)(\mathbf{1}_{T_i} - \boldsymbol{G}_i)^\top \circ \mathbf{J}_{T_i}\boldsymbol{D}_i^\top)
\end{aligned}
$$

The deletion matrices $\boldsymbol{D}_i$ are different in the two equations and indicate the missing values in $\boldsymbol{X}_i^{*\mathrm{C}}$ and $\boldsymbol{X}_i^{*\mathrm{M}}$ respectively. The results of the measurement error correction with the method of moments for the multiple regression analysis of the Augsburger Umwelstudie are depicted in Figure 6.1.

The application of the developed bias correction on the data of the Augsburger Umweltstudie reveals a slight increase of the estimated effect coefficient considering the models for concurrent measurements. Delayed effects are not significantly different from zero. The preferred strategy is the model based on mixture data, because more information is used.

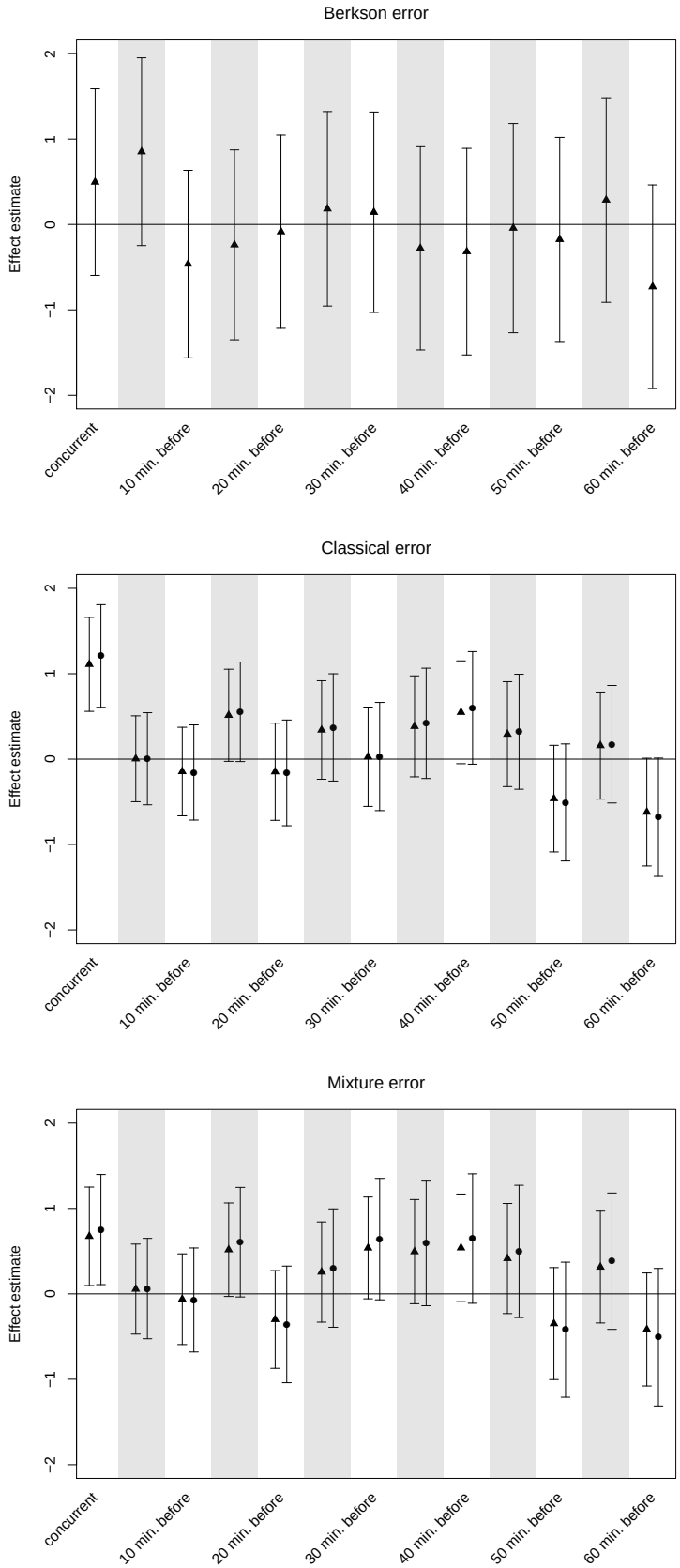


FIGURE 6.1.: *Naive (triangles) and MOM–corrected (circles) effect estimates for the association between heart rate and concurrent and lagged individual PNC levels with Berkson, classical and mixture measurement error in the multiple regression analysis.*

The attenuation of the effect estimate is stronger for the mixture error (Figure 6.2). The estimated attenuation factors are slightly left skewed. Distributional properties of the estimates of the parameters describing the measurement errors are presented in Section 6.2 and are depicted in Figure 6.4.

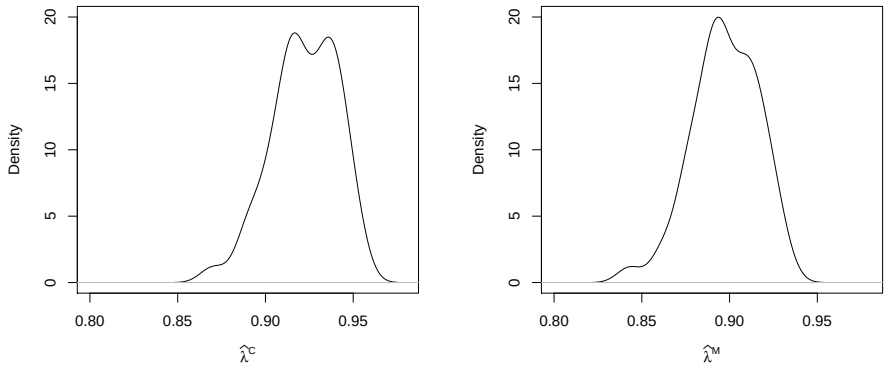


FIGURE 6.2.: *Density estimation for the estimated attenuation factors based on bootstrapped data for the multiple regression analysis.*

6.1.2. Bias correction in a simple regression model

To further illustrate the theoretical results, also simple regression models are calculated. The attenuation factors for the analyses using measurements with classical measurement error and with mixture error are estimated using

$$\widehat{\lambda}^{*\mathrm{C}} = \frac{\sum_{i=1}^{n} \operatorname{tr}\left[\widehat{\boldsymbol{V}}_i^{*\mathrm{C}^{-1}} \boldsymbol{D}_i \left(\widehat{\boldsymbol{\Sigma}}_{X^{*\mathrm{B}}} + \widehat{\sigma}^2_{\nu^{\mathrm{B}}} \mathbf{J}_T + \widehat{\boldsymbol{\Sigma}}_{U^{\mathrm{B}}}\right) \boldsymbol{D}_i^{\top}\right]}{\sum_{i=1}^{n} \operatorname{tr}\left[\widehat{\boldsymbol{V}}_i^{*\mathrm{C}^{-1}} \boldsymbol{D}_i \left(\widehat{\boldsymbol{\Sigma}}_{X^{*\mathrm{B}}} + \widehat{\sigma}^2_{\nu^{\mathrm{B}}} \mathbf{J}_T + \widehat{\sigma}^2_{\nu^{\mathrm{C}}} \mathbf{J}_T + \widehat{\boldsymbol{\Sigma}}_{U^{\mathrm{B}}} + \widehat{\boldsymbol{\Sigma}}_{U^{\mathrm{C}}}\right) \boldsymbol{D}_i^{\top}\right]} \tag{6.1}$$

and

$$\begin{aligned}\widehat{\lambda}^{*\mathrm{M}} = &\left\{\sum_{i=1}^{n} \operatorname{tr}\left[\widehat{\boldsymbol{V}}_i^{*\mathrm{M}^{-1}} \boldsymbol{D}_i \left(\widehat{\boldsymbol{\Sigma}}_{X^{*\mathrm{B}}} + ((\mathbf{1}_T - \boldsymbol{G}_i)\mathbf{1}_T^{\top}) \circ (\widehat{\sigma}^2_{\nu^{\mathrm{B}}} \mathbf{J}_T + \widehat{\boldsymbol{\Sigma}}_{U^{\mathrm{B}}})\right) \boldsymbol{D}_i^{\top}\right]\right\} \\ &\left\{\sum_{i=1}^{n} \operatorname{tr}\left[\widehat{\boldsymbol{V}}_i^{*\mathrm{M}^{-1}} \boldsymbol{D}_i \left(\widehat{\boldsymbol{\Sigma}}_{X^{*\mathrm{B}}} + ((\mathbf{1}_T - \boldsymbol{G}_i)(\mathbf{1}_T - \boldsymbol{G}_i)^{\top}) \circ \right.\right.\right. \\ &\left.\left.\left. (\widehat{\sigma}^2_{\nu^{\mathrm{B}}} \mathbf{J}_T + \widehat{\sigma}^2_{\nu^{\mathrm{C}}} \mathbf{J}_T + \widehat{\boldsymbol{\Sigma}}_{U^{\mathrm{B}}} + \widehat{\boldsymbol{\Sigma}}_{U^{\mathrm{C}}})\right) \boldsymbol{D}_i^{\top}\right]\right\}^{-1},\end{aligned} \tag{6.2}$$

based on the considerations in the Subsections 3.3.3, 3.4.3, 3.3.3 and 3.5.5. The values of the parameters of the variance–covariance matrices in eq. 6.1 and eq. 6.2 are described in Table 6.1. The regression coefficients and their MOM–corrected equivalents up to a lag of 60 minutes are displayed in Figure 6.3. The results only slightly differ from the results of the multiple regression model.

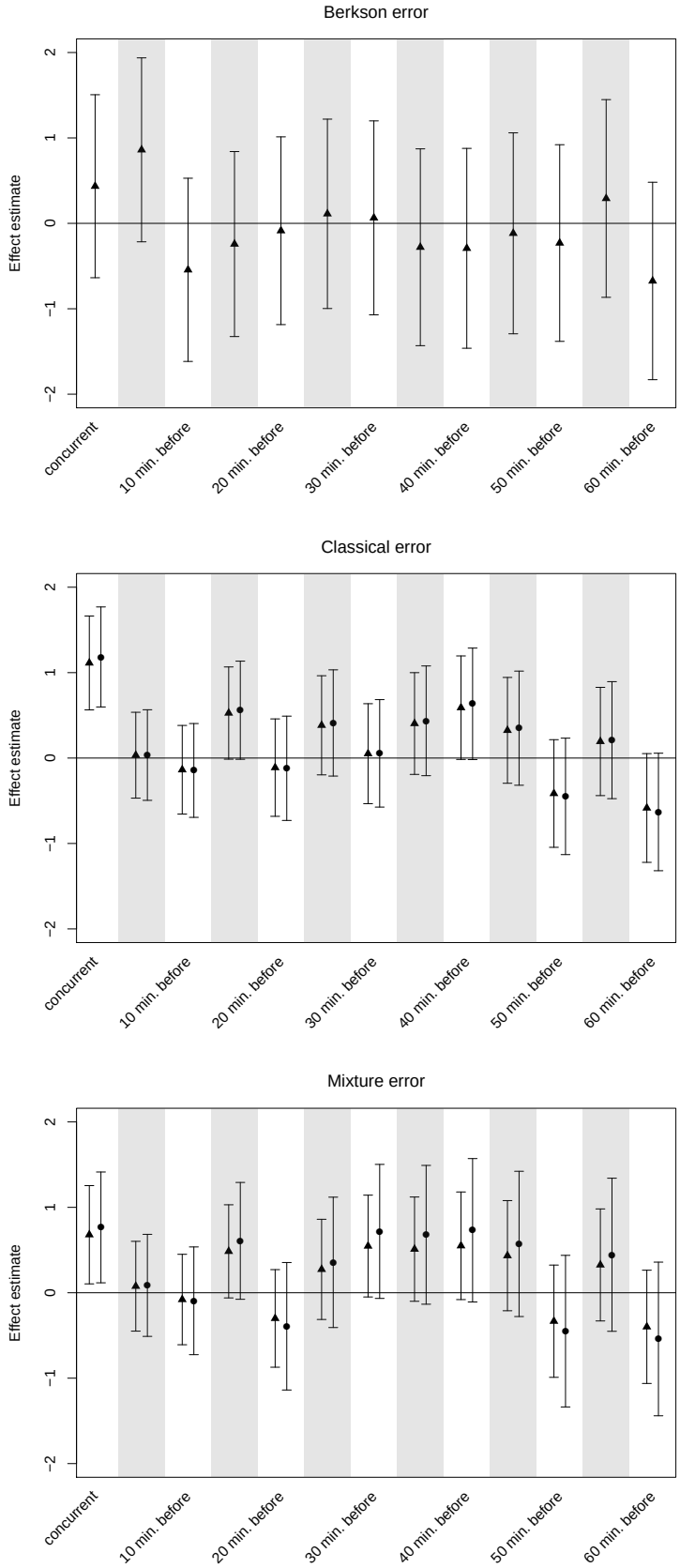


FIGURE 6.3.: *Naive (triangles) and MOM–corrected (circles) effect estimates for the association between heart rate and concurrent and lagged individual PNC levels with Berkson, classical and mixture measurement error in the simple regression analysis.*

6.2. Bayesian measurement error models

The application of Bayesian measurement error models is focused on varying approaches to include prior knowledge; delayed effects are not considered. Three Bayesian modeling approaches are compared: 1) the usage of fixed external information about the measurement errors, 2) the usage of uncertain external information and 3) the simultaneous modeling approach including all data sources (Augsburger Umweltstudie, the validation study and the comparison measurements).

The latter approach takes advantage of the flexible formulation of dependencies between different data sources in the Bayesian model. Thus, several information sources for the description of measurement errors can be simultaneously considered. Moreover, Bayesian modeling enables a more differentiated consideration of the Berkson error, e.g. separate Berkson error components for different seasons. The specification of the relations between the data sets is demanding. Model specifications are adopted from previous chapters.

6.2.1. Model formulation

Posterior distributions

Some abbreviations for models are introduced throughout previous chapters; these are strongly utilized for the following application of Bayesian modeling to the Augsburger Umweltstudie. The posterior distributions defined in eq. 5.1–5.3 are composed of several sub–models:

$$f(\boldsymbol{\Theta}^{\mathrm{C}}, \boldsymbol{X}|\boldsymbol{Y}, \boldsymbol{X}^{*\mathrm{C}}, \boldsymbol{Z}) \propto \underbrace{f(\boldsymbol{Y}|\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\theta}^{H})}_{\text{HMM}} \underbrace{f(\boldsymbol{X}^{*\mathrm{C}}|\boldsymbol{X}, \boldsymbol{\theta}^{M})}_{\text{CMM}} \underbrace{f(\boldsymbol{X}|\boldsymbol{Z}, \boldsymbol{\theta}^{E_C})}_{\text{BMM}_0} f(\boldsymbol{\Theta}^{\mathrm{C}}) \quad (6.3)$$

$$f(\boldsymbol{\Theta}^{\mathrm{B}}, \boldsymbol{X}|\boldsymbol{Y}, \boldsymbol{X}^{*\mathrm{B}}, \boldsymbol{Z}) \propto \underbrace{f(\boldsymbol{Y}|\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\theta}^{H})}_{\text{HMM}} \underbrace{f(\boldsymbol{X}|\boldsymbol{X}^{*\mathrm{B}}, \boldsymbol{\theta}^{E_B})}_{\text{BMM}_1 \text{ or } \text{BMM}_2} f(\boldsymbol{\Theta}^{\mathrm{B}}) \quad (6.4)$$

$$f(\boldsymbol{\Theta}^{\mathrm{M}}, \boldsymbol{X}|\boldsymbol{Y}, \boldsymbol{X}^{*\mathrm{C}}, \boldsymbol{X}^{*\mathrm{B}}, \boldsymbol{Z}) \propto \underbrace{f(\boldsymbol{Y}|\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\theta}^{H})}_{\text{HMM}} \underbrace{f(\boldsymbol{X}^{*\mathrm{C}}|\boldsymbol{X}, \boldsymbol{\theta}^{M})}_{\text{CMM}} \underbrace{f(\boldsymbol{X}|\boldsymbol{X}^{*\mathrm{B}}, \boldsymbol{Z}, \boldsymbol{\theta}^{E_B})}_{\text{BMM}_1 \text{ or } \text{BMM}_2} f(\boldsymbol{\Theta}^{\mathrm{M}}) \quad (6.5)$$

Likelihoods

Some of the sub–models, which are incorporated in the likelihoods included in eq. 6.3–6.5 are already defined earlier; the remaining sub–models are introduced in the following. An overview about the abbreviations used for sub–models included in the likelihoods of the Bayesian analyses is given in Table 6.2.

Abbr.	Description	Parameters	Equation
HMM	HM for true measurements of main data	$\boldsymbol{\theta}^{H} = (\beta_0, \beta_X, \boldsymbol{\beta}_Z, \sigma^2_\tau, \sigma^2_\varepsilon, \rho)^\top$	Eq. 2.1
HMM$^{*\mathrm{B}}$	HM for fixed–site measurements of main data	$\boldsymbol{\theta}^{H} = (\beta_0, \beta_X, \boldsymbol{\beta}_Z, \sigma^2_\tau, \sigma^2_\varepsilon, \rho)^\top$ [a]	Eq. 3.1
HMM$^{*\mathrm{C}}$	HM for personal measurements of main data	$\boldsymbol{\theta}^{H} = (\beta_0, \beta_X, \boldsymbol{\beta}_Z, \sigma^2_\tau, \sigma^2_\varepsilon, \rho)^\top$ [a]	Eq. 3.2
BMM$^{*\mathrm{B}}$	EM for fixed–site measurements of main data	$\boldsymbol{\theta}^{E_{*B}} = (\mu^{X^{*\mathrm{B}}}, \sigma^2_{X^{*\mathrm{B}}}, \rho^{X^{*\mathrm{B}}})^\top$	Eq. 2.10
BMM$_0$	EM for personal measurements of main data without $\boldsymbol{X}^{*\mathrm{B}}$	$\boldsymbol{\theta}^{E_{B0}} = (\mu^{X}, \sigma^2_{U^{\mathrm{B}}}, \sigma^2_{\nu^{\mathrm{B}}}, \rho^{\mathrm{B}})^\top$	Eq. 6.6
BMM$_1$	One–stage EM for personal measurements of main data	$\boldsymbol{\theta}^{E_{B1}} = (\sigma^2_{U^{\mathrm{B}}}, \sigma^2_{\nu^{\mathrm{B}}}, \rho^{\mathrm{B}})^\top$	Eq. 6.7
BMM$_2$	Two–stage EM for personal measurements of main data	$\boldsymbol{\theta}^{E_{B2}} = (\boldsymbol{\beta}_{Z^{\mathrm{B}}}, \sigma^2_{U^{\mathrm{BAct}}k}, \beta_0^{\mathrm{BBack}}, \beta^{*\mathrm{B}}, \boldsymbol{\beta}_{Z^{\mathrm{BBack}}}, \sigma^2_{\nu^{\mathrm{B}}}, \sigma^2_{U^{\mathrm{B}}}, \rho^{\mathrm{B}})^\top$	Eq. 6.8
CMM	Classical measurement error model for main data	$\boldsymbol{\theta}^{M} = (\sigma^2_{\nu^{\mathrm{C}}}, \sigma^2_{U^{\mathrm{C}}}, \rho^{\mathrm{C}})^\top$	Eq. 2.5

[a] Note, that the parameters in the naive health models HMM$^{*\mathrm{B}}$ and HMM$^{*\mathrm{C}}$ are no longer marked with a "*" to simplify and unify the referencing to the results with and without measurement error correction.

TABLE 6.2.: *Abbreviations for sub–models included in the likelihoods of the Bayesian analyses.*

The specification of an exposure model for fixed–site measurements (BMM$^{*\mathrm{B}}$) is optional for all models and necessary for models intended to simultaneously predict missing values of fixed–site exposure.

BMM$_1$ is used as exposure model for personal measurements in the frequentist approach. Two alternative exposure models are additionally considered: BMM$_0$, which is applied when fixed–site measurements are not available, and BMM$_2$, which allows for further parameters monitored at the fixed measurement site as well as for information about the activity and the microenvironment of the individuals.

The three types of exposure models are defined as follows:

BMM$_0$: EM for personal measurements of main data without $\boldsymbol{X}^{*\mathrm{B}}$

$$\begin{aligned} X_{i_M t_M} =& \mu^X + \nu^{\mathrm{B}}_{i_M} + U^{\mathrm{B}}_{i_M t_M} \\ \nu^{\mathrm{B}}_{i_M} \sim& N(0, \sigma^2_{\nu^{\mathrm{B}}}) \\ \boldsymbol{U}^{\mathrm{B}}_{i_M} \sim& N(\boldsymbol{0}, \boldsymbol{\Sigma}^{\mathrm{B}}_{i_M}), \boldsymbol{\Sigma}^{\mathrm{B}}_{i_M} = \sigma^2_{U^{\mathrm{B}}} \boldsymbol{W}_{\rho^{\mathrm{B}}, i_M} \\ \boldsymbol{\theta}^{E_{B0}} =& (\mu^X, \sigma^2_{U^{\mathrm{B}}}, \sigma^2_{\nu^{\mathrm{B}}}, \rho^{\mathrm{B}})^\top \end{aligned} \tag{6.6}$$

BMM$_1$: One–stage EM for personal measurements of main data

$$\begin{aligned} X_{i_M t_M} =& X^{*\mathrm{B}}_{i_M t_M} + \nu^{\mathrm{B}}_{i_M} + U^{\mathrm{B}}_{i_M t_M} \\ \nu^{\mathrm{B}}_{i_M} \sim& N(0, \sigma^2_{\nu^{\mathrm{B}}}) \\ \boldsymbol{U}^{\mathrm{B}}_{i_M} \sim& N(\boldsymbol{0}, \boldsymbol{\Sigma}^{\mathrm{B}}_{i_M}), \boldsymbol{\Sigma}^{\mathrm{B}}_{i_M} = \sigma^2_{U^{\mathrm{B}}} \boldsymbol{W}_{\rho^{\mathrm{B}}, i_M} \\ \boldsymbol{\theta}^{E_{B1}} =& (\sigma^2_{U^{\mathrm{B}}}, \sigma^2_{\nu^{\mathrm{B}}}, \rho^{\mathrm{B}})^\top \end{aligned} \tag{6.7}$$

BMM$_2$: Two–stage EM for personal measurements of main data

Activity model:

$$\begin{aligned} X_{i_M t_M} =& X^{*\mathrm{BBack}}_{i_M t_M} + \boldsymbol{Z}^{\mathrm{B}}_{i_M t_M} \boldsymbol{\beta}_{Z^{\mathrm{B}}} + U^{\mathrm{B}}_{i_M t_M} \\ U^{\mathrm{B}}_{i_M t_M} \sim& N(0, \sigma^2_{U^{\mathrm{BAct}}, k_{i_M t_M}}) \end{aligned} \tag{6.8}$$

Background model (based on observations without PM related activities):

$$\begin{aligned} X^{*\mathrm{BBack}}_{i_M t_M} =& \beta^{\mathrm{BBack}}_0 + \beta^{*\mathrm{B}} X^{*\mathrm{B}}_{i_M t_M} + \boldsymbol{Z}^{\mathrm{BBack}}_{i_M t_M} \boldsymbol{\beta}_{Z^{\mathrm{BBack}}} + \nu^{\mathrm{B}}_{i_M} + U^{\mathrm{BBack}}_{i_M t_M} \\ \nu^{\mathrm{B}}_{i_M} \sim& N(0, \sigma^2_{\nu^{\mathrm{B}}}) \\ \boldsymbol{U}^{\mathrm{BBack}}_{i_M} \sim& N(\boldsymbol{0}, \boldsymbol{\Sigma}^{\mathrm{BBack}}_{i_M}), \boldsymbol{\Sigma}^{\mathrm{BBack}}_{i_M} = \sigma^2_{U^{\mathrm{B}}} \boldsymbol{W}_{\rho^{\mathrm{B}}, i_M} \\ \boldsymbol{\theta}^{E_{B2}} =& (\boldsymbol{\beta}_{Z^{\mathrm{B}}}, \sigma^2_{U^{\mathrm{BAct}}_k}, \beta^{\mathrm{BBack}}_0, \beta^{*\mathrm{B}}, \boldsymbol{\beta}_{Z^{\mathrm{BBack}}}, \sigma^2_{\nu^{\mathrm{B}}}, \sigma^2_{U^{\mathrm{B}}}, \rho^{\mathrm{B}})^\top \end{aligned}$$

Time–activity and temporal data deliver additional information, which can be used to explain parts of the deviations of personal measurements from fixed–site measurements. The two–stage exposure model for the Augsburger Umweltstudie is based upon the exposure model developed for the validation study in Section 2.4. The background models are calculated separately for seasons (as applied e.g. by Gryparis et al., 2007b) and separately for indoor and outdoor observations (as applied e.g. by Bliznyuk et al., 2014); $\sigma^2_{U^{\mathrm{B}}}, \sigma^2_{\nu^{\mathrm{B}}}$ and ρ^{B} are assumed to be equal for spring and fall. The indoor and the outdoor models are not related through an infiltration model since appropriate data or expert knowledge about an infiltration model are not available. Only observations without PM related activities are used for the background model for indoor observations; i.e. only observations

during not staying in traffic are included in the analysis of outdoor data. For simplicity, the indices in eq. 6.8 are not adapted.

The confounding variables for the background model $\boldsymbol{Z}^{\text{BBack}}$ include personally measured temperature in the models for indoor observations and personally measured dew point temperature and time in the models for outdoor observations. The effects of the continuous covariates time and temperature in the background model are modeled with natural cubic splines with 6 degrees of freedom; the effect of dew point temperature is linearly modeled due to convergency issues. Modeling the smooth effects of the background model with natural cubic splines and the choice of the degrees of freedom are along the lines of Chang et al. (2011). The model errors in the background models are assumed to follow an AR(1) process (in contrast to Section 2.4, where we allowed for autocorrelated in the activity model instead of the background model).

The activity models are calculated separately for indoor and outdoor observations. $\boldsymbol{Z}^{\text{B}}$ includes for indoor observations indicators for cooking and smoke, and for outdoor observations indicators for traffic and smoke. Activity–specific variances $\sigma^2_{U^{\text{BAct}},k}$ are allowed.

Prior knowledge about non–identifiable parameters

We differentiate between three types of prior knowledge about the non–identifiable parameters $\boldsymbol{\phi}_N = (\boldsymbol{\theta}^M, \boldsymbol{\theta}^E)^\top$, with $\boldsymbol{\theta}^E \in \{\boldsymbol{\theta}^{E_{B0}}, \boldsymbol{\theta}^{E_{B1}}, \boldsymbol{\theta}^{E_{B2}}\}$:

1. *Fix prior knowledge*: $\boldsymbol{\phi}_N$ is assumed to be known (e.g. from external analyses) and to be non–stochastic.

2. *Uncertain prior knowledge*: $\boldsymbol{\phi}_N$ is stochastic and information about the uncertainty regarding $\boldsymbol{\phi}_N$ is available.

3. *Prior knowledge from validation data*: Validation data contain information about $\boldsymbol{\phi}_N$.

The detailed specification of the prior distributions is listed in Table 6.3 and visualized in Figure 6.4. For the analyses with *fix prior knowledge*, the same parameter values as for the frequentist analyses are used, which are assessed in Chapter 2 and depicted in Table 6.1. Further these parameter values serve as expectations of Uniform prior distributions in the analyses with *uncertain prior knowledge*. The specification of non–informative prior distributions for the non–identifiable parameters is required in the simultaneous modeling approach, i.e. if the actual prior information is deduced from the simultaneous modeling of *validation data*; the resulting prior distributions for $\sigma^2_{U^{\text{B}}}, \sigma^2_{\nu^{\text{B}}}$ and ρ^{B} differ from the prior knowledge gained from internal calculations (Figure 6.4). Non–informative Uniform priors (Unif$(0, 100)$) are chosen for variance parameters describing the classical and the Berkson error. Unif$(0, 1)$ distributions are chosen for autocorrelation parameters. The prior for the mean of the personal measurements (μ^X) is set to N$(0, 10^6)$.

Type	Prior knowledge	$\boldsymbol{\theta}^M$	$\boldsymbol{\theta}^E$			
			$\boldsymbol{\theta}^{E_{B0}}$	$\boldsymbol{\theta}^{E_{B1}}$	$\boldsymbol{\theta}^{E_{B2}}$	$\boldsymbol{\theta}^{E_{*B}}$
1	fix	$\sigma^2_{U^{\mathrm{C}}} = 0.03$	$\mu^X \sim \mathrm{N}(0, 10^6)$	$\sigma^2_{U^{\mathrm{B}}} = 0.3$	$\sigma^2_{U^{\mathrm{B}}} = 0.3$	$\mu^{X^{*\mathrm{B}}} = 0$
		$\sigma^2_{\nu^{\mathrm{C}}} = 0.03$	$\sigma^2_{U^{\mathrm{B}}} \sim \mathrm{Unif}(0, 100)$	$\sigma^2_{\nu^{\mathrm{B}}} = 0.21$	$\sigma^2_{\nu^{\mathrm{B}}} = 0.21$	$\sigma^2_{X^{*\mathrm{B}}} = 0.34$
		$\rho^{\mathrm{C}} = 0.696$	$\sigma^2_{\nu^{\mathrm{B}}} \sim \mathrm{Unif}(0, 100)$	$\rho^{\mathrm{B}} = 0.582$	$\rho^{\mathrm{B}} = 0.582$	$\rho^{X^{*\mathrm{B}}} = 0.932$
			$\rho^{\mathrm{B}} \sim \mathrm{Unif}(0, 1)$			
2	uncertain	$\sigma^2_{U^{\mathrm{C}}} \sim \mathrm{Unif}(0.02, 0.04)$	$\mu^X \sim \mathrm{N}(0, 10^6)$	$\sigma^2_{U^{\mathrm{B}}} \sim \mathrm{Unif}(0.2, 0.4)$	$\sigma^2_{U^{\mathrm{B}}} \sim \mathrm{Unif}(0.2, 0.4)$	$\mu^{X^{*\mathrm{B}}} \sim \mathrm{Unif}(-0.2, 0.2)$
		$\sigma^2_{\nu^{\mathrm{C}}} \sim \mathrm{Unif}(0.02, 0.04)$	$\sigma^2_{U^{\mathrm{B}}} \sim \mathrm{Unif}(0, 100)$	$\sigma^2_{\nu^{\mathrm{B}}} \sim \mathrm{Unif}(0.16, 0.26)$	$\sigma^2_{\nu^{\mathrm{B}}} \sim \mathrm{Unif}(0.16, 0.26)$	$\sigma^2_{X^{*\mathrm{B}}} \sim \mathrm{Unif}(0.3, 0.38)$
		$\rho^{\mathrm{C}} \sim \mathrm{Unif}(0.60, 0.80)$	$\sigma^2_{\nu^{\mathrm{B}}} \sim \mathrm{Unif}(0, 100)$	$\rho^{\mathrm{B}} \sim \mathrm{Unif}(0.53, 0.63)$	$\rho^{\mathrm{B}} \sim \mathrm{Unif}(0.53, 0.63)$	$\rho^{X^{*\mathrm{B}}} \sim \mathrm{Unif}(0.9, 0.96)$
			$\rho^{\mathrm{B}} \sim \mathrm{Unif}(0, 1)$			
3	validation	CMC,	$\mu^X \sim \mathrm{N}(0, 10^6)$	BMV_1	BMV_2	$\mathrm{BMV}^{*\mathrm{B}}$
	data	CMV (optionally)	$\sigma^2_{U^{\mathrm{B}}} \sim \mathrm{Unif}(0, 100)$			
			$\sigma^2_{\nu^{\mathrm{B}}} \sim \mathrm{Unif}(0, 100)$			
			$\rho^{\mathrm{B}} \sim \mathrm{Unif}(0, 1)$			

TABLE 6.3.: *Prior specifications for non–identifiable parameters $\boldsymbol{\phi}_N$ and for parameters describing the partially missing fixed–site PNC measurements, $\boldsymbol{\theta}^{E_{*B}}$.*

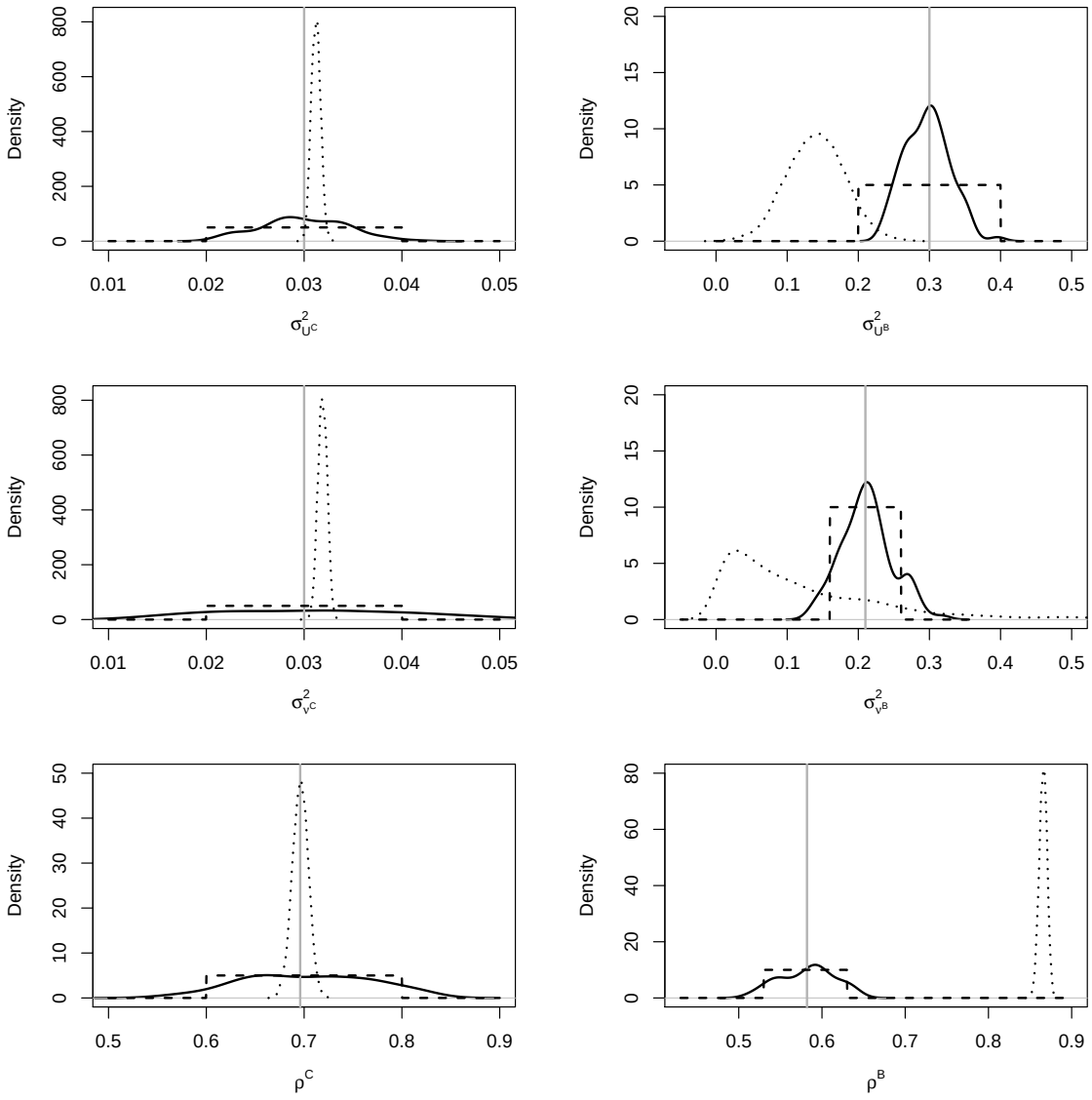


FIGURE 6.4.: *Prior densities for the three types of prior knowledge: fixed (vertical solid line), uncertain (dashed) and prior knowledge from validation data (dotted) in contrast to the density of the bootstrap replicates used for the frequentist analysis (solid).*

Models describing measurement errors using the validation data and the comparison measurements are established to include *prior knowledge from validation data* into the Bayesian health outcome analyses. Some parameters in the models for the comparison measurements and the validation data equal non–identifiable parameters in the models for the main data in order to connect the data sources, whereas some other parameters, like the effect coefficients of the activities, are separately defined for each data source. Table 6.4 gives an overview about these models.

CMC is described in Section 2.3; the remaining models will be introduced in the following. The specification of an exposure model for fixed–site measurements $\widetilde{X}^{*\mathrm{B}}_{i_V t_V}$ of individual $i_V, i_V = 1 \ldots, n_V$, at time point $t_V, t_V = 1 \ldots, T_V$, of the validation study (BMV$^{*\mathrm{B}}$) is optional:

Abbr.	Description	Parameters	Equation
$\text{BMV}^{*\text{B}}$	EM for fixed–site measurements measurements of validation data	$\boldsymbol{\omega}^{E_{*B}} = (\mu^{X^{*\text{B}}}, \sigma^2_{X^{*\text{B}}}, \rho^{X^{*\text{B}}})^\top$	Eq. 6.9
BMV_1	One–stage EM for personal measurements of validation data	$\boldsymbol{\omega}^{E_{B1}} = (\sigma^2_{U^{\text{B}}}, \sigma^2_{\nu^{\text{B}}}, \rho^{\text{B}})^\top$	Eq. 6.10
BMV_2	Two–stage EM for personal measurements of validation data	$\boldsymbol{\omega}^{E_{B2}} = (\boldsymbol{\beta}_{\widetilde{Z}^{\text{B}}}, \sigma^2_{U^{\text{Bact}}k}, \widetilde{f}^{*\text{B}}, \boldsymbol{\beta}_{\widetilde{Z}^{\text{BBack}}}, \sigma^2_{\nu^{\text{B}}}, \sigma^2_{U^{\text{B}}}, \rho^{\text{B}})^\top$	Eq. 6.11
CMV	Classical measurement error model for validation data	$\boldsymbol{\omega}^{M_V} = (\sigma^2_{\nu^{\text{C}}}, \sigma^2_{U^{\text{C}}}, \rho^{\text{C}})^\top$	Eq. 6.12
CMC	Classical measurement error model for comparison measurements	$\boldsymbol{\omega}^{M_C} = (\alpha^C_0, \alpha^C_1, \sigma^2_{V^{\text{C}}}, \sigma^2_{\nu^{\text{C}}}, \sigma^2_{U^{\text{C}}}, \rho^{\text{C}})$	Eq. 2.11

TABLE 6.4.: *Abbreviations for sub–models included in the prior distributions for non–identifiable parameters of the Bayesian analyses based on validation data.*

$\text{BMV}^{*\text{B}}$: EM for fixed–site measurements of validation data

$$\begin{aligned} \widetilde{X}^{*\text{B}}_{i_V t_V} =& \mu^{X^{*\text{B}}} + \widetilde{U}^{*\text{B}}_{i_V t_V} \\ \widetilde{U}^{*\text{B}}_{i_V} \sim& \text{N}(0, \widetilde{\boldsymbol{\Sigma}}^{*\text{B}}_{i_V}); \widetilde{\boldsymbol{\Sigma}}^{*\text{B}}_{i_V} = \sigma^2_{X^{*\text{B}}} \boldsymbol{W}_{\rho^{X^{*\text{B}}}, i_V} \\ \boldsymbol{\omega}^{E_{*B}} =& (\mu^{X^{*\text{B}}}, \sigma^2_{X^{*\text{B}}}, \rho^{X^{*\text{B}}})^\top \end{aligned} \tag{6.9}$$

Two alternative exposure models are considered for the personal measurements $\widetilde{X}_{i_V t_V}$ of the validation data. Firstly, the one–stage exposure model is based only on the fixed–site measurements. Secondly, the two–stage exposure model is separated in an activity model and a background model:

BMV$_1$: One–stage EM for personal measurements of validation data

$$
\begin{aligned}
\widetilde{X}_{i_V t_V} =& \widetilde{X}^{*\mathrm{B}}_{i_V t_V} + \widetilde{\nu}^{\mathrm{B}}_{i_V} + \widetilde{U}^{\mathrm{B}}_{i_V t_V} \\
\widetilde{\nu}^{\mathrm{B}}_{i_V} \sim& N(0, \sigma^2_{\nu^{\mathrm{B}}}) \\
\widetilde{\boldsymbol{U}}^{\mathrm{B}}_{i_V} \sim& N(\boldsymbol{0}, \widetilde{\boldsymbol{\Sigma}}^{\mathrm{B}}_{i_V}), \widetilde{\boldsymbol{\Sigma}}^{\mathrm{B}}_{i_V} = \sigma^2_{U^{\mathrm{B}}} \boldsymbol{W}_{\rho^{\mathrm{B}}, i_V} \\
\boldsymbol{\omega}^{E_{B1}} =& (\sigma^2_{U^{\mathrm{B}}}, \sigma^2_{\nu^{\mathrm{B}}}, \rho^{\mathrm{B}})^\top
\end{aligned}
\tag{6.10}
$$

BMV$_2$: Two–stage EM for personal measurements of validation data

Activity model:

$$
\begin{aligned}
\widetilde{X}_{i_V t_V} =& \widetilde{X}^{*\mathrm{BBack}}_{i_V t_V} + \widetilde{\boldsymbol{Z}}^{\mathrm{B}}_{i_V t_V} \boldsymbol{\beta}_{\widetilde{Z}^{\mathrm{B}}} + \widetilde{U}^{\mathrm{B}}_{i_V t_V} \\
\widetilde{U}^{\mathrm{B}}_{i_V t_V} \sim& N(0, \sigma^2_{U^{\mathrm{BAct}}, k_{i_V t_V}})
\end{aligned}
\tag{6.11}
$$

Background model (based on observations without PM related activities):

$$
\begin{aligned}
\widetilde{X}^{*\mathrm{BBack}}_{i_V t_V} =& \widetilde{\beta}^{\mathrm{BBack}}_0 + \widetilde{\boldsymbol{X}}^{*\mathrm{B}} \widetilde{\boldsymbol{\beta}}^{*\mathrm{B}} + \widetilde{\boldsymbol{Z}}^{\mathrm{BBack}}_{i_V t_V} \boldsymbol{\beta}_{\widetilde{Z}^{\mathrm{BBack}}} + \widetilde{\nu}^{\mathrm{B}}_{i_V} + \widetilde{U}^{\mathrm{BBack}}_{i_V t_V} \\
\widetilde{\nu}^{\mathrm{B}}_{i_V} \sim& N(0, \sigma^2_{\nu^{\mathrm{B}}}) \\
\widetilde{\boldsymbol{U}}^{\mathrm{BBack}}_{i_V} \sim& N(\boldsymbol{0}, \widetilde{\boldsymbol{\Sigma}}^{\mathrm{BBack}}_{i_V}), \widetilde{\boldsymbol{\Sigma}}^{\mathrm{BBack}}_{i_V} = \sigma^2_{U^{\mathrm{B}}} \boldsymbol{W}_{\rho^{\mathrm{B}}, i_V} \\
\boldsymbol{\omega}^{E_{B2}} =& (\boldsymbol{\beta}_{\widetilde{Z}^{\mathrm{B}}}, \sigma^2_{U^{\mathrm{BAct}}k}, \widetilde{\beta}^{\mathrm{BBack}}_0, \widetilde{\boldsymbol{\beta}}^{*\mathrm{B}}, \boldsymbol{\beta}_{\widetilde{Z}^{\mathrm{BBack}}}, \sigma^2_{\nu^{\mathrm{B}}}, \sigma^2_{U^{\mathrm{B}}}, \rho^{\mathrm{B}})^\top
\end{aligned}
$$

The stratified model structure and the included confounding variables in BMV$_2$ are deduced from the exposure model developed in Section 2.4 by variable selection. Separate background models for different seasons and indoors/outdoors and separate activity models for indoors/outdoors are specified. Only observations without PM related activities are used for the background model; for simplicity, the indices in eq. 6.11 are again not adapted. The included confounding variables are listed in Figure 6.5.

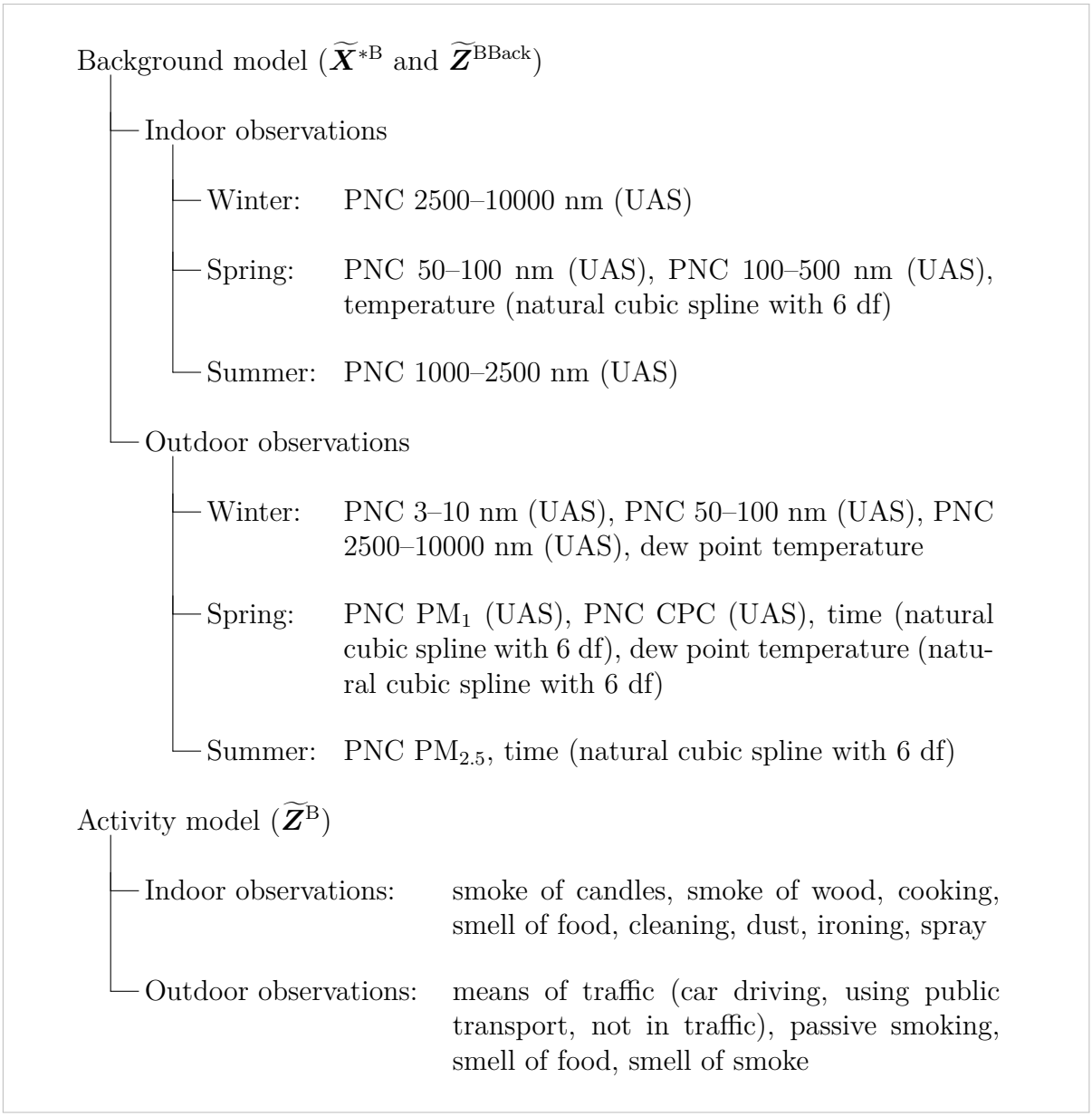


FIGURE 6.5.: *Variables in the two–stage EM for personal measurements of validation data (BMV_2).*

In the Bayesian analysis, we also allow for classical measurement error of the personal PNC measurements in the validation study assuming the following structure:

CMV: Classical measurement error model for validation data

$$\begin{aligned}
\widetilde{X}^{*\mathrm{C}}_{i_V t_V} =& \widetilde{X}^{\mathrm{C}}_{i_V t_V} + \widetilde{\nu}^{\mathrm{C}}_{i_V} + \widetilde{U}^{\mathrm{C}}_{i_V t_V} \\
\widetilde{\nu}^{\mathrm{C}}_{i_V} \sim& \mathrm{N}(0, \sigma^2_{\nu^{\mathrm{C}}}) \\
\widetilde{\boldsymbol{U}}^{\mathrm{C}}_{i_V} \sim& \mathrm{N}(\boldsymbol{0}, \widetilde{\boldsymbol{\Sigma}}^{\mathrm{C}}_{i_V}), \widetilde{\boldsymbol{\Sigma}}^{\mathrm{C}}_{i_V} = \sigma^2_{U^{\mathrm{C}}} \boldsymbol{W}_{\rho^{\mathrm{C}}, i_V} \\
\boldsymbol{\omega}^{M_V} =& (\sigma^2_{\nu^{\mathrm{C}}}, \sigma^2_{U^{\mathrm{C}}}, \rho^{\mathrm{C}})^\top
\end{aligned} \tag{6.12}$$

Non–informative prior distributions for the remaining parameters

Non–informative normal prior distributions ($\mathrm{N}(0, 10^6)$) are used for all regression coefficients of the health, exposure and measurement models. Non–informative Uniform priors are chosen for the remaining parameters, which do not describe the measurement errors: $\mathrm{Unif}(0, 1000)$ for σ^2_ε and σ^2_τ, $\mathrm{Unif}(0, 1000)$ for the remaining variance parameters) and $\mathrm{Unif}(0, 1)$ distributions for autocorrelation parameters. All parameters are assumed to be a priori independent of each other.

6.2.2. Implementation

Preprocessing of the data aims at a data basis for the Bayesian analyses, which is comparable to the data used for the frequentist analysis conducted in the previous subsection. Therefore, only outdoor measurements are used for the main analyses. Observations during activities or combinations of activities, which occur only less than ten times, are excluded from the data of the Augsburger Umweltstudie. The 1–min. resolution of the validation data requires either the transformation of the parameters describing the classical measurement error or the aggregation of the latent, true exposure data to a 5–min. resolution. The latter possibility is applied, because the transformation of variances of log–transformed data with different original resolution is challenging and the additivity assumption concerning the classical measurement error for a certain data resolution does not hold for a different temporal resolution. Again, observations during rare activities or combinations of activities (less than four times) are excluded from the validation study.

All PM concentrations are logarithmically transformed. The data are standardized for the calculations of the sub–models; parameters, which enter several sub–models are adequately re–transformed if necessary. Observations of the main study with missing data in the health outcome or some covariate are excluded from the main analysis.

The first 1000 MCMC iterations are discarded and the subsequent 10000 iterations are used for statistical inference.

6.2.3. Results

The comparison of different Bayesian modeling strategies for the main study is subject of this subsection. Beside three types of prior distributions, five options to build the likelihood are considered (the single model components for each option are listed in brackets):

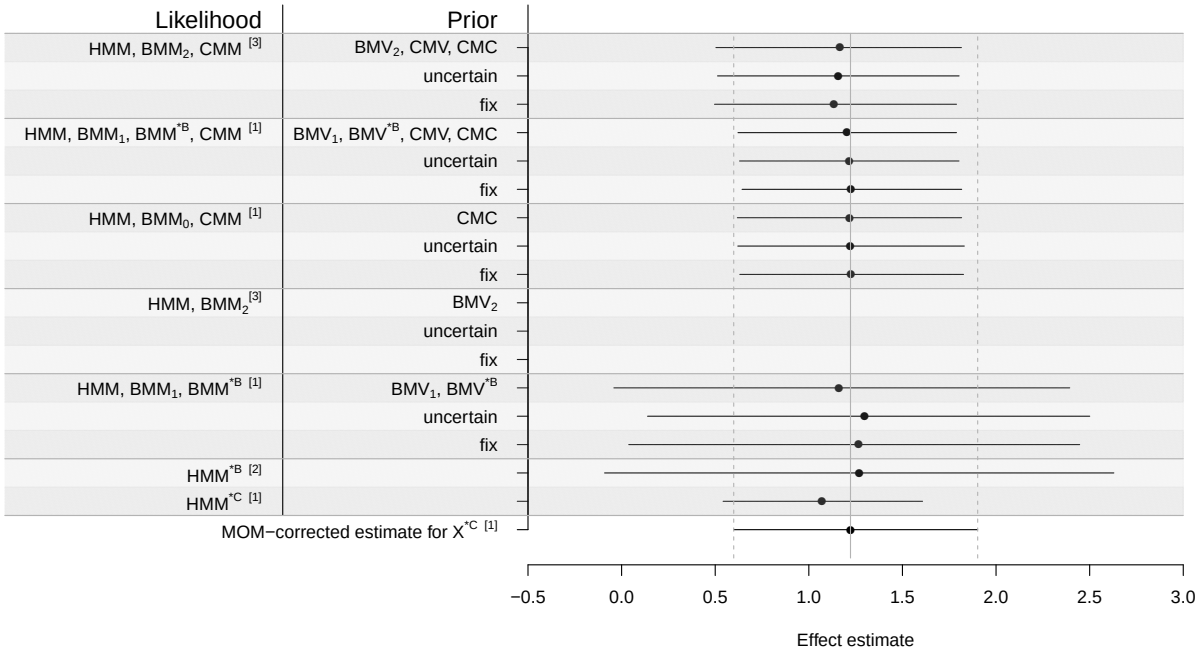
1. Using only Berkson error–prone data with a one–stage exposure model ($\mathrm{HMM}/\mathrm{BMM}_1/\mathrm{BMM}^{*\mathrm{B}}$) and

2. with a two–stage exposure model (HMM/BMM_2),
3. using only classical measurement error–prone data (HMM/BMM_0/CMM) and
4. simultaneously using Berkson and classical error–prone data with a one–stage exposure model (HMM/BMM_1/BMM^{*B}/CMM) and
5. with a two–stage exposure model (HMM/BMM_2/CMM).

We concentrate on models based on observations with available personal PNC measurements and on outdoor observations, because the Berkson error during staying indoors highly depends on the individuals and therefore, the external knowledge about the Berkson error of indoor observations may not be adequate.

Effects of personal PNC

The effect estimates for the effect of personal PNC on the heart rate for different model alternatives are depicted in Figure 6.6:



Included data:
[1]: Personal PNC measurements available
[2]: Personal and fixed–site PNC measurements available
[3]: Personal and fixed–site PNC measurements and values of confounding variables available

FIGURE 6.6.: *Posterior means for the effect of personal PNC on the heart rate and their 95 % credible intervals in comparison to frequentist estimates (solid, gray line: MOM–corrected effect estimate; dashed, gray lines: 95 % confidence interval for MOM–corrected effect estimate); effect estimates from models, which did not converge are not shown.*

The results of HMM^{*B} and HMM/BMM_1/BMM^{*B} differ from the results for Berkson error–prone measurements shown in Section 6.1 due to a different data basis. As expected, the Bayesian estimate using classical error–prone data (HMM^{*C}) is attenuated and the Bayesian estimate using Berkson error–prone data (HMM^{*B}) shows wider credible intervals in comparison to the MOM–corrected effect. Bayesian accounting for Berkson error in the models HMM/BMM_1/BMM^{*B} only slightly modifies the effect estimate and the

credible intervals. The Markov Chains of the effect estimates for HMM/BMM$_2$ do not converge.

The Bayesian effect estimates for HMM/BMM$_0$/CMM, HMM/BMM$_1$/BMM$^{*\mathrm{B}}$ and HMM/BMM$_2$/CMM converge and tend to equal or go below the MOM–corrected estimate; the effect estimates of the former two models exhibit credible intervals, which are slightly narrower than the confidence intervals of the MOM–corrected estimate. The results will be explained in detail with respect to various aspects in the subsequent paragraphs.

Convergence

Convergence problems only occur for the Bayesian health outcome models for Berkson error–prone data with a two–stage exposure model, i.e. HMM/BMM$_2$. The Markov chains of the exposure effect estimate strongly depend on the initial conditions. Data and prior information seem to provide an insufficient amount of information for the aim of getting a stable state of each chain. An improvement of these models may be possible through additional prior information, as e.g. by including knowledge about the activity model. The results demonstrate, that sufficient information about the exposure model is essential for the calculation of Bayesian regression models with a Berkson error–prone covariate beyond adequate knowledge about the size of the measurement error itself.

Indirect learning

Unintended indirect learning may severely affect the results of a Bayesian regression model for measurement error–prone data as explained in Chapter 5. IL is diagnosed by comparing the estimated conditional posterior distributions of the non–identifiable parameters in the models with the corresponding prior distributions. This comparison is depicted in the Figures 6.7 and 6.8 for the examined models except HMM/BMM$_2$ due to failure of convergence.

The prior distributions in the left columns of the Figures 6.7 and 6.8 differ from the prior distributions in the right columns: in the left columns, the estimated conditional posterior distributions of the parameters using uncertain prior distributions are depicted, whereas the right columns show the results from using Bayesian models for the validation data and the comparison measurements. Several prior distributions appear for Berkson error parameters in the case of prior knowledge from validation data, because different prior distributions are used depending on the exposure model. For the two–stage exposure model, season–specific prior distributions are allowed assuming equal prior distribution for spring and fall.

The resemblance of the estimated conditional posterior distributions for the non–identifiable parameters with the respective prior distribution of the models for either Berkson or classical measurement error (HMM/BMM$_1$ and HMM/BMM$_0$/CMM) confirms that the chosen prior distributions for $\boldsymbol{\Theta}_{-\phi_N}$ prevent IL type (I).

IL type (II) can be ruled out because $\sigma^2_{\varepsilon^*}$ and $\sigma^2_{\tau^*}$ are by far larger than $\sigma^2_{U\mathrm{C}}$ and $\sigma^2_{\nu^{\mathrm{C}}}$, even if the variability of the quantities is considered. Therefore, the Figures 6.7 and 6.8 indicate IL type (III). The occurrence of IL type (III) becomes clear because the prior distributions for the non–identifiable parameters are specified independently from each other.

In the models HMM/BMM$_0$/CMM and HMM/BMM$_1$/BMM$^{*\mathrm{B}}$/CMM, the parameters describing the Berkson error indirectly learn from the parameters describing the classical error, but not the other way round; indirect learning about the Berkson error strongly

depends on the model. In contrast, Berkson error and classical measurement error parameters are mutually affected in the models HMM/BMM$_2$/CMM, i.e. the models with a two–stage exposure model including season–specific Berkson errors. Subsequently, IL type (III) is examined in detail for each model.

For the model HMM/BMM$_0$/CMM, non–informative prior distributions for $\sigma^2_{U^{\mathrm{B}}}, \sigma^2_{\nu^{\mathrm{B}}}$ and ρ^{B} are specified for the model with uncertain expert prior knowledge as well as for the model with prior knowledge from validation data; note, that these prior distributions are not depicted in Figure 6.8. The estimated conditional posterior distributions for $\sigma^2_{U^{\mathrm{B}}}, \sigma^2_{\nu^{\mathrm{B}}}$ and ρ^{B} are informative due to IL type (III) and are very similar for the two types of prior knowledge for the classical measurement error (uncertain expert prior knowledge and prior knowledge from validation data).

The estimations of the conditional posterior distributions for $\sigma^2_{U^{\mathrm{B}}}, \sigma^2_{\nu^{\mathrm{B}}}$ and ρ^{B} differ from the corresponding prior distributions in the models HMM/BMM$_1$/BMM$^{*\mathrm{B}}$/CMM for both types of prior knowledge. Uncertain expert prior knowledge about $\sigma^2_{U^{\mathrm{B}}}$ and $\sigma^2_{\nu^{\mathrm{B}}}$ is moderately and prior knowledge about ρ^{B} is strongly adapted through IL. External prior knowledge for $\sigma^2_{U^{\mathrm{B}}}, \sigma^2_{\nu^{\mathrm{B}}}$ and ρ^{B} from a one–stage exposure model for the validation data substantially differs from the internal estimation of the parameters, e.g. prior values for $\sigma^2_{U^{\mathrm{B}}}$ are about two times higher with internal estimation compared to the external estimation. IL strongly modifies the prior knowledge about $\sigma^2_{U^{\mathrm{B}}}$ gained from the validation data.

Expert prior knowledge about the Berkson error is specified as not season–specific and is not accounted for activity effects; the estimated conditional posterior distributions for $\sigma^2_{U^{\mathrm{B}}}, \sigma^2_{\nu^{\mathrm{B}}}$ and ρ^{B} from model HMM/BMM$_2$/CMM strongly differ from the corresponding prior distributions hinting at a misspecified structure. Prior knowledge from validation data is moderately adapted through IL. Only the two–stage exposure model, which is based on prior knowledge from the validation data, provides prior knowledge about the Berkson error, which affects the prior knowledge about the classical measurement error.

In general, the results reveal, that modeling uncertain expert prior knowledge with uniform distributions restricts in some cases the adaption through IL type (III).

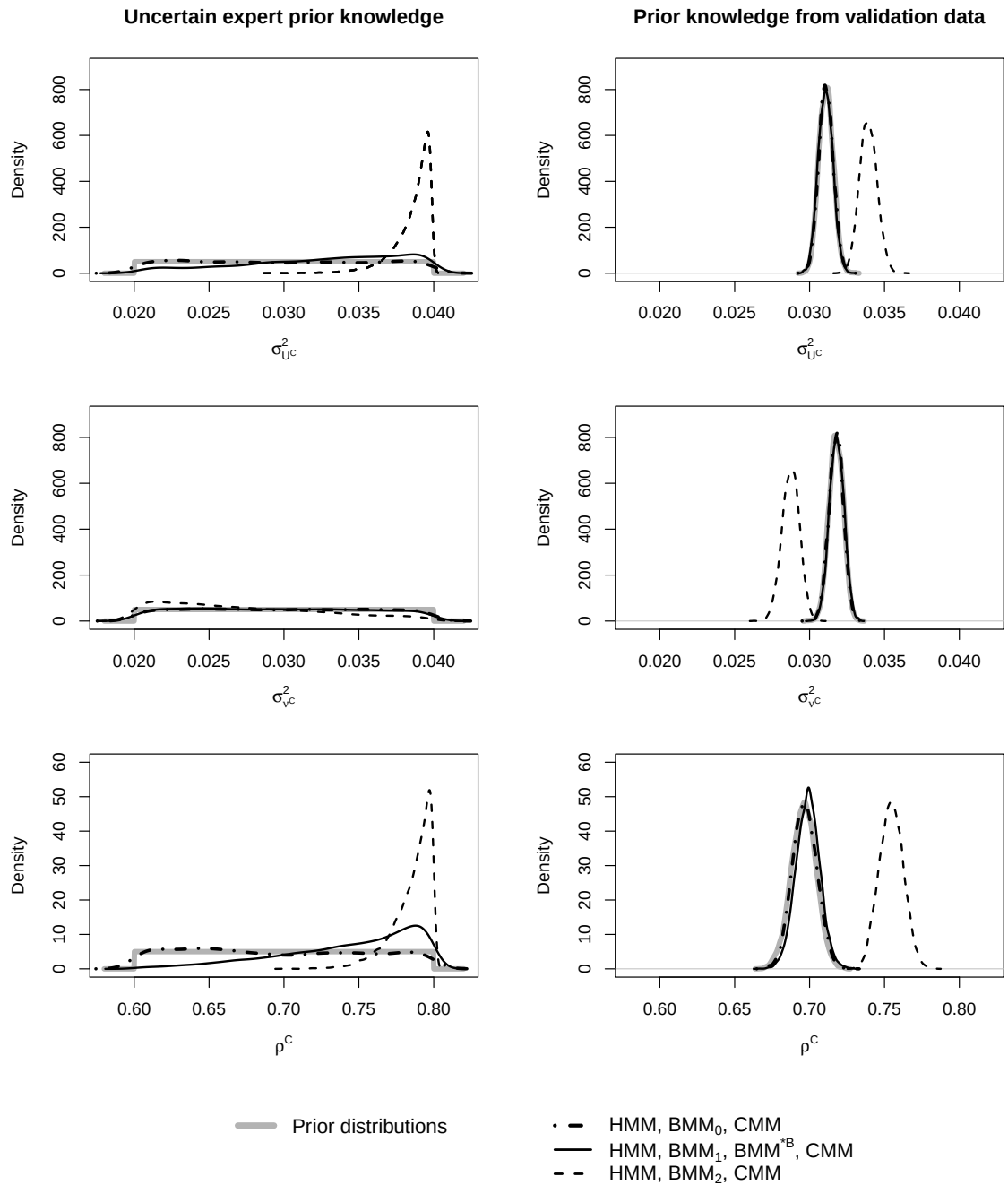


FIGURE 6.7.: *Estimated conditional posterior distributions of parameters characterizing the classical measurement error for uncertain expert prior knowledge (left) and prior knowledge gained from validation data (right). The gray lines, marking the prior distributions, represent in the left column the specified prior knowledge for* $\boldsymbol{\theta}^M$ *and in the right column the estimated conditional posterior distribution resulting from the model CMC.*

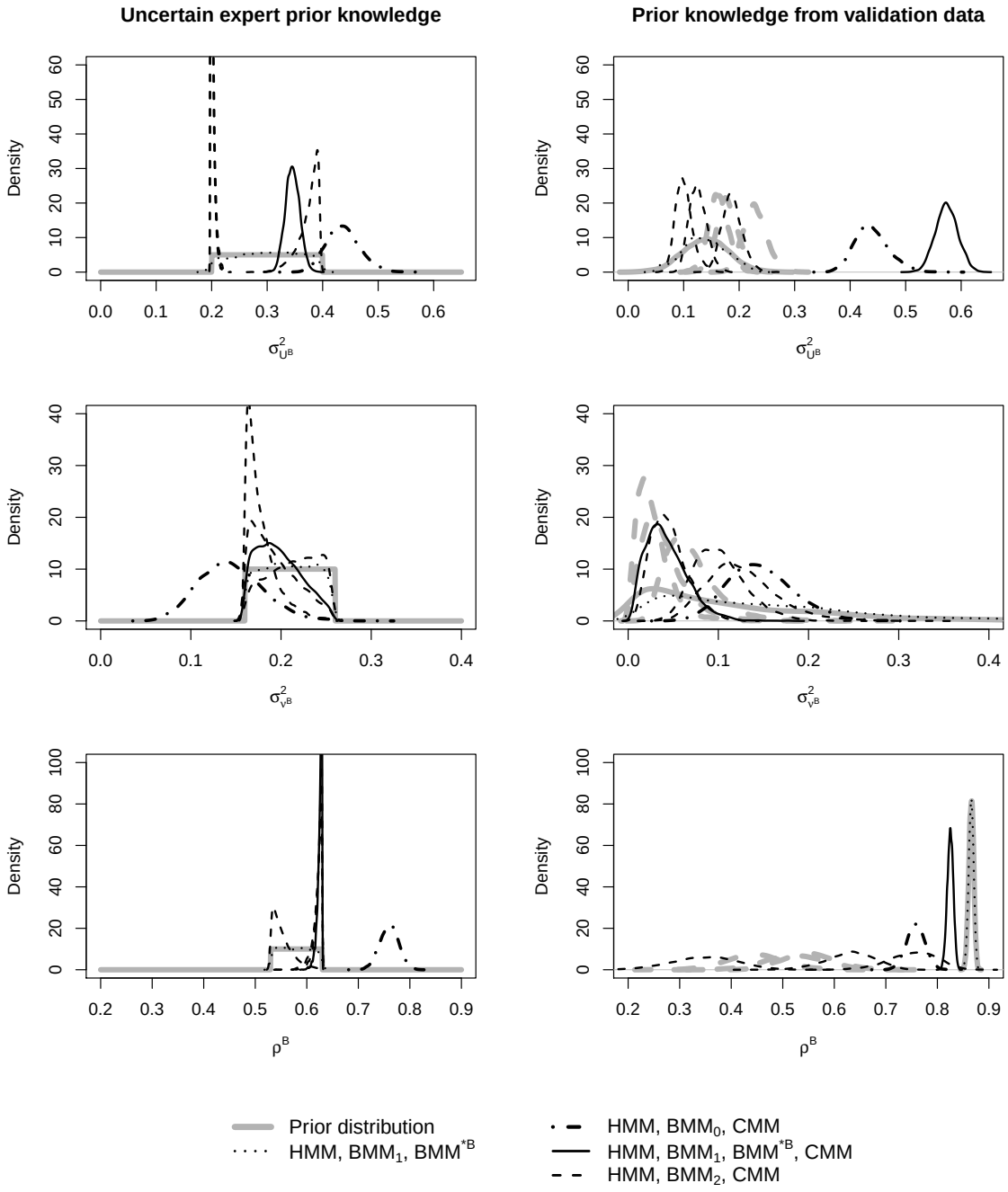


FIGURE 6.8.: *Estimated conditional posterior distributions of parameters characterizing the Berkson error for uncertain expert prior knowledge (left) and prior knowledge gained from validation data (right). The gray lines, marking the prior distributions, represent in the left column the specified prior knowledge for* $\boldsymbol{\theta}^{E_{B1}}/\boldsymbol{\theta}^{E_{B2}}$ *and in the right column the estimated conditional posterior distribution resulting from the models* BMV_1 *(solid line) and* BMV_2 *(dashed lines). Note the differing scale of the x–axis in the first row.*

Simultaneous modeling of several data sources

The model for the description of the classical measurement error is equal for all analyses – frequentist as well as Bayesian analyses. The properties of the measurement error are extracted in advance in the MOM approach and the Bayesian models with fixed and uncertain expert knowledge, whereas the measurement model is simultaneously estimated with the measurement error correction of the health outcome model. Simultaneous modeling of several data sources, i.e. the simultaneous modeling of the health outcome data and the validation and/or the comparison data, yields similar conditional posterior distributions for the non–identifiable parameters as separate modeling, if IL does not occur (see model HMM/BMM_0/CMM in Figure 6.7 and model HMM/BMM_1 in Figure 6.8). This indicates, that information about the measurement error can be adequately introduced into the health model by simultaneous modeling (see also Section 5.5). Moreover, simultaneous modeling enables to use the information about a non–identifiable parameter gained from one data source for the information extraction about another parameter from another data source; e.g., we adjusted the simultaneous models for classical measurement error in PNC measurements of the validation data.

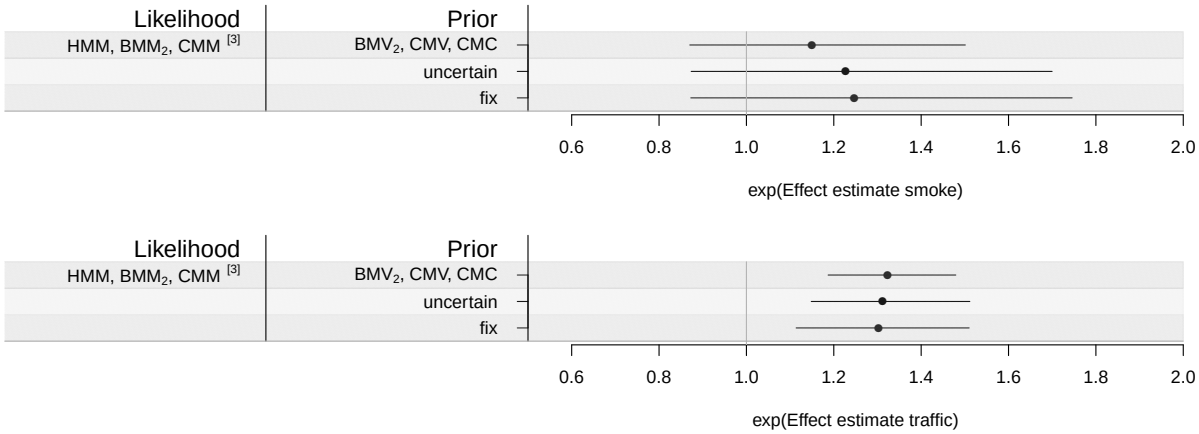
Transferability

Transferability of prior knowledge from external data sources is crucial for Bayesian regression analyses for measurement error. Indeed, possibly occurring IL type (III) may modify knowledge from external data and may perhaps correct inadequate external prior knowledge, but the extent of the modification depends on the uncertainty of the prior knowledge and on the degree of the compliance with the main data. In the present example, the adequacy of transferring prior knowledge about the Berkson error from exposure models for the validation data is doubted, because of the large discrepancy between internal and external information.

Activities

Activity effects, which are estimated within the two–stage exposure models for the Augsburger Umweltstudie, BMM_2, are depicted in Figure 6.9.

Smoke and traffic positively affect personal PNC, but only the multiplicative transformed effect estimate for traffic is significantly different from 1. The realization–specific variability of personal exposure to PNC is negligible small during not staying in traffic, whereas the realization–specific variability during staying in traffic or during being exposed to smoke is larger than 0.



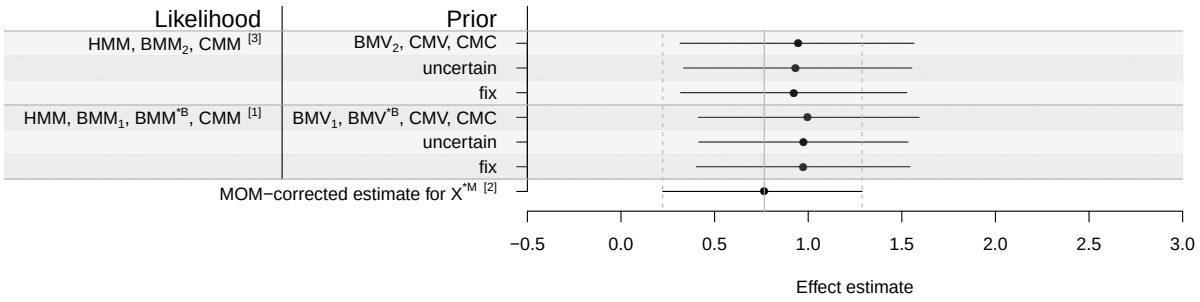
Included data:
[1]: Personal PNC measurements available
[2]: Personal and fixed–site PNC measurements available
[3]: Personal and fixed–site PNC measurements and values of confounding variables available

FIGURE 6.9.: *Posterior means for the effects of smoke and traffic on personal PNC in the exposure model for the main study and their 95 % credible intervals.*

6.2.4. Additional analyses

Bayesian mixture

The analyses presented in Subsection 6.2.3 are restricted on data with available personal PNC measurements for a better comparability between the model types. Effect estimates on the basis of the complete available outdoor data are presented in Figure 6.10.



Included data:
[1]: Complete data
[2]: Personal or fixed–site PNC measurements available
[3]: Fixed–site PNC measurements and values of confounding variables available

FIGURE 6.10.: *Posterior means for the effect of personal PNC on the heart rate based on the complete available data set and their 95 % credible intervals (solid, gray line: MOM–corrected effect estimate; dashed, gray lines: 95 % confidence interval for MOM–corrected effect estimate).*

The effect estimates lay between the MOM–corrected estimate for $X^{*\mathrm{M}}$ and the Bayesian–corrected estimates depicted in Figure 6.6. The lower effect estimate in comparison to the Bayesian–corrected estimates originates from the fact, that the Berkson error–prone measurements indicate a reduced effect, when the extended data basis is used as visible in Figure 6.1. The higher effect estimate in comparison to the MOM–corrected effect estimate for $X^{*\mathrm{M}}$ occurs due to partial indirect learning: IL about the Berkson error exists only for observations with fixed–site and personal measurements, whereas for the remaining observations the specified prior distributions provide the information about the Berkson error. In the present case, this trade–off yields a reduction in $\sigma^2_{U^{\mathrm{B}}}$, a stronger attenuation of the effect and hence a stronger measurement error correction (Figure 6.11).

Smooth effect of PNC

So far, a linear effect of personal PNC levels on the heart rate was presumed. Polynomial and semiparametric modeling approaches relax this assumption and can easily be combined with Bayesian regression models for an error–prone covariate. The health model with a forth degree polynomial for the effect of personal PNC is

$$\begin{aligned} Y_{i_M t_M} =& \beta_0 + \beta_{X,1} X_{i_M t_M} + \beta_{X,2} X^2_{i_M t_M} + \beta_{X,3} X^3_{i_M t_M} + \beta_{X,4} X^4_{i_M t_M} + \\ & \boldsymbol{Z}_{i_M t_M} \boldsymbol{\beta}_Z + \tau_{i_M} + \varepsilon_{i_M t_M}. \end{aligned}$$

Indeed, orthogonal polynomials are known to overcome the multicollinearity involved by raw polynomials, but orthogonalization of the data has to be accomplished before the calculation of the Bayesian model, when JAGS or WinBUGS is used. However, $\boldsymbol{X}$ is latent in our Bayesian regression model for an error–prone covariate and the values of $\boldsymbol{X}$ change in each iteration. Therefore, raw polynomials are used.

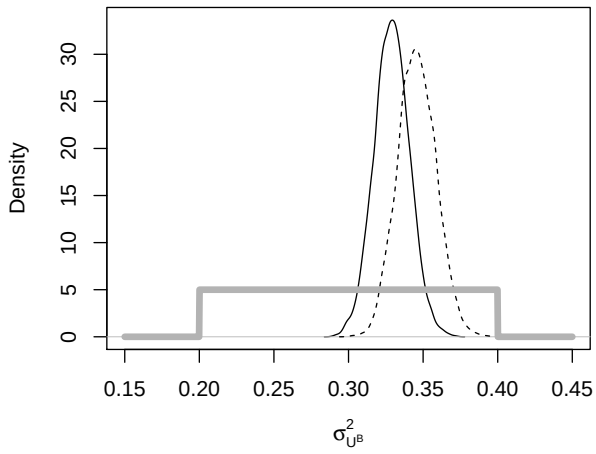


FIGURE 6.11.: *Example for partial indirect learning. Gray solid line: uncertain prior distribution; estimated conditional posterior distribution for HMM/BMM$_1$/BMM^{*B}/CMM based on the complete outdoor data (black solid line) and based on data restricted to observations with available personal PNC measurements (black dashed line).*

Crainiceanu et al. (2005) propose the usage of low–rank thin–plate splines for Bayesian semiparametric regression models due to favorable mixing properties in comparison to other approaches. The health model using low–rank thin–plate splines for the effect of personal PNC is

$$Y_{i_M t_M} = \beta_0 + \beta_X X_{i_M t_M} + \sum_{k=1}^{K} u_k |X_{i_M t_M} - \kappa_k|^3 + \boldsymbol{Z}_{i_M t_M} \boldsymbol{\beta}_Z + \tau_{i_M} + \varepsilon_{i_M t_M}.$$

with fixed knots $\kappa_1 < \kappa_2 < \ldots < \kappa_K$. Penalization of the regression coefficients $u_k, k = 1, \ldots, K$ is conducted through the equivalent representation of the penalized regression model as a linear mixed model (Crainiceanu et al., 2005):

$$\boldsymbol{Y} = \beta_0 + \boldsymbol{X}\beta_X + \boldsymbol{R}\boldsymbol{b} + \boldsymbol{Z}\boldsymbol{\beta}_Z + \boldsymbol{S}\boldsymbol{\tau} + \boldsymbol{\varepsilon}.$$

with $\boldsymbol{R} = \boldsymbol{R}_K \boldsymbol{\Omega}_K^{-1/2}$, $[\boldsymbol{R}_K]_{i,\cdot} = (|X_{i_M t_M} - \kappa_1|^3, |X_{i_M t_M} - \kappa_2|^3 \ldots, |X_{i_M t_M} - \kappa_K|^3)$ as the i–th row of $\boldsymbol{R}_K$, $\boldsymbol{b} = \boldsymbol{\Omega}_K^{-1/2}\boldsymbol{u}$ as random effects with $\mathbb{E}(\boldsymbol{b}) = \boldsymbol{0}$ and $\mathrm{Var}(\boldsymbol{b}) = \sigma_b^2$ and

$$\boldsymbol{S} = \begin{pmatrix} \boldsymbol{1}_{T_1} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{1}_{T_2} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \ldots \\ \boldsymbol{0} & \boldsymbol{0} & \vdots & \boldsymbol{1}_{T_n} \end{pmatrix}.$$

$\boldsymbol{b}, \boldsymbol{\tau}$ and $\boldsymbol{\varepsilon}$ are assumed to be mutually independent. $\boldsymbol{\Omega}_K$ denotes a penalty matrix with $[\boldsymbol{\Omega}_K]_{l,k} = |\kappa_l - \kappa_k|$. $K = 13$ knots and the prior $\sigma_b^2 \sim \mathrm{IG}(10^{-6}, 10^{-6})$ are used for the analysis. The linear and non–linear effect estimates are compared in Figure 6.12 for model HMM/BMM$_1$/BMM$^{*\mathrm{B}}$/CMM and fixed prior knowledge.

The polynomial and the semiparametric model show a nearly linear increasing effect in the lower and medium domain of the data, which is flattened involving high uncertainty in the upper domain of the data. Since the mixing of the Markov chains in the semiparameteric model is not adequate, these results, in particular the credible intervals, should

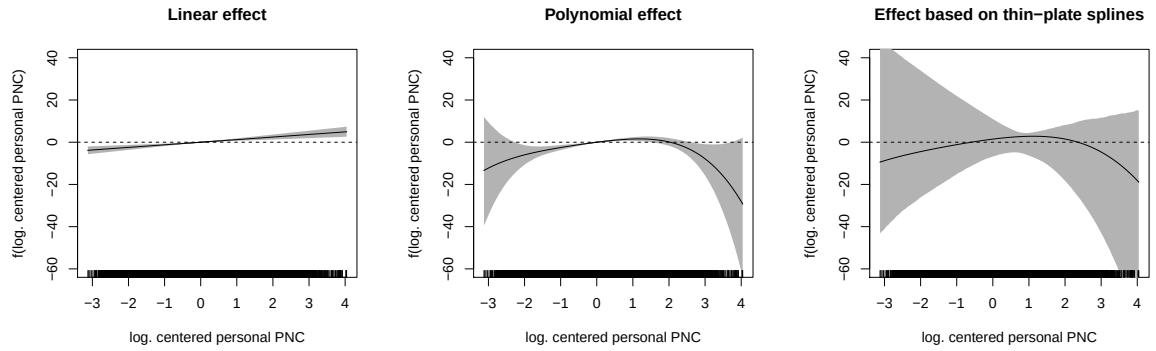


FIGURE 6.12.: *Estimates for the linear, polynomial and semiparametric effect of the log–transformed, centered personal PNC on heart rate based on model HMM/BMM$_1$/BMM^{*B}/CMM with fixed prior knowledge. Solid line: mean effect; gray regions: pointwise 95 % credible intervals of the effect.*

be interpreted with caution. However, the mean effect strongly resembles the mean polynomial effect indicating that the polynomial effect sufficiently describes the non–linear effect.

Final models

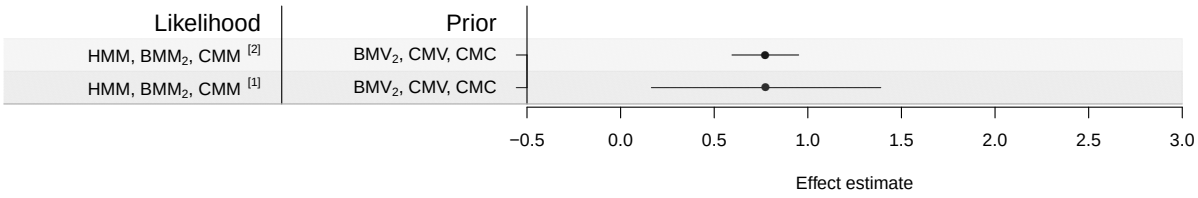
The previous examinations of Bayesian models for the Augsburg Umweltstudie are focused on the comparison to the frequentist approaches with regard to certain aspects, like different ways of specifying the likelihood (especially the exposure model) and of including prior knowledge, the simultaneous modeling of fixed–site and personal PNC measurements and a smooth effect of personal PNC on the heart rate. The developed MOM approach cannot deal with non–constant Berkson error. The Bayesian approach is more flexible and the extension of the model regarding indoor/outdoor–specific Berkson error can easily be accomplished. All these aspects are integrated in two final Bayesian models: one model for the outdoor observations and one model for the indoor and outdoor observations of the Augsburger Umweltstudie:

HMM$_{\text{fin}}$: Final health model for true measurements of main data

$$
\begin{aligned}
Y_{i_M t_M} =& \beta_0 + \beta_X X_{i_M t_M} + \boldsymbol{Z}_{i_M t_M}\boldsymbol{\beta}_Z + \tau_{\text{Pat}_{i_M}} + \tau_{i_M} + \varepsilon_{i_M t_M} \\
\boldsymbol{\varepsilon}_{i_M} \sim& N(\mathbf{0}, \boldsymbol{\Sigma}_{\varepsilon,i_M}), \boldsymbol{\Sigma}_{i_M} = \sigma^2_\varepsilon \boldsymbol{W}_{\rho,i_M} \\
\tau_{\text{Pat}_{i_M}} \sim& N(0, \sigma^2_{\tau_{\text{Pat}}}), \\
\tau_{i_M} \sim& N(0, \sigma^2_\tau), \\
\boldsymbol{\theta}^{H^{\text{fin}}} =& (\beta_0, \beta_X, \boldsymbol{\beta}_Z, \sigma^2_{\tau_{\text{Pat}}}, \sigma^2_\tau, \sigma^2_\varepsilon, \rho)^\top
\end{aligned}
\tag{6.13}
$$

For clarity, we again abstain from using separate indices for the models based on outdoor and on indoor/outdoor observations. In addition, the random intercept is changed for the final model, since up to four measurement series of a single patient are available for the analysis: to account for dependencies of measurement series of a single person a random patient intercept, $\tau_{\text{Pat}_{i_M}}$, $\text{Pat}_{i_M} = 1, \ldots, n_{\text{Pat}}$ (n_{Pat}: number of patients in the Augsburger Umweltstudie), is included in addition to the random intercept for the measurement series, τ_{i_M}.

We used the model HMM/BMM_2/CMM for the likelihood and the model BMV_2/CMV/CMC for the prior distribution to include all available data in the final models.



Included data:
[1]: Complete outdoor data
[2]: Complete indoor and outdoor data

FIGURE 6.13.: *Posterior means for the effect of personal PNC on the heart rate and their 95 % credible intervals for the final models.*

Indoor/outdoor and outdoor data yield the same posterior mean for the effect of personal PNC on the heart rate (Figure 6.13). The effect is smaller than in the previous results, which is attributed to the additional random effect for the patient.

7. Conclusion

7.1. Summary and discussion

An exposure model and methods for covariate measurement error correction in longitudinal health outcome analyses were developed in this thesis with the aim to combine several data sources motivated by the Augsburger Umweltstudie. The bias analyses regarding the impact of covariate measurement error in the longitudinal setting were the basis for the frequentist correction approach using the method of moments. Adequate inclusion of prior knowledge about a single or several parameters describing classical and Berkson error in Bayesian analyses for the correction of measurement error were examined and discussed. The developed methods and the obtained insights were applied to the Augsburger Umweltstudie.

Exposure model

This paragraph has been published in advance within the article "Personal exposure to ultrafine particles: Two-level statistical modeling of background exposure and time-activity patterns during three seasons" (Deffner et al., 2016) in Journal of Exposure Science and Environmental Epidemiology.

A two–stage modeling approach for the validation study was necessary in order to separate the effect of non–ambient sources from the background exposure. An activity model was established which accounts for autocorrelation in the longitudinal data without masking the categorical effects. The study design and the small sample size aimed at describing the exposure to PNC during certain activities and not at the general description of personal exposure to PNC. As the background model describes the data without many distributional assumptions, the model is of a rather explorative nature and requires further evaluation. Therefore, we refrained from using the model in a regression calibration approach for measurement error correction. Especially considering the limited sample size regarding the included individuals as well as the restricted number of scenarios and of days per season, larger studies are needed to further back up our findings of the activity model. Indeed, the volunteers and the frequency of the examined activities were not representative of the general population, but in fact, this requirement is not necessary. Instead, representativity of the measurements during a certain activity suffices for the applied models and interpretations. Our two–stage model is a first attempt to describe personal exposure to PNC within a single modeling approach. The flexible modeling approach allowed us to identify the relevant predictors.

The developed exposure model helped identifying and describing the determining factors for personal exposure to PNC in everyday situations. Personal PNC levels and their determinants varied depending on the location (indoor/outdoor) of the individual and on the season. Also, the influence of particles from outdoor sources differed seasonally. Personal outdoor exposure on background indoor concentrations was mainly driven by ambient particle concentrations and dew point temperature. Cooking, smell of food and smoke of candles were significantly associated with increased personal indoor PNC levels. Exposure during car driving was higher than as a pedestrian. Lowest outdoor levels were found during commuting with public transport or while not staying in traffic. The know-

ledge about the main predictors for personal PNC provides a basis for the development of strategies to reduce the PM concentration in the direct environment of human beings.

Model definition

Using surrogates, e.g. ambient exposure measurements, instead of personal exposure measurements entails measurement error. We examined Berkson error, classical measurement error and a special mixture error of a covariate as well as the simultaneous analysis of classical and Berkson error–prone data in linear mixed models with autocorrelated errors and permitted a wide range of error structures including individual–specific and autocorrelated measurement errors. Zeger et al. (2000) argue that exposure measurement error cannot be clearly attributed to either Berkson or classical measurement error. Three types of mixtures between Berkson and classical measurement error were considered throughout this work: 1) the occurrence of either Berkson or classical measurement error, 2) the occurrence of classical measurement error of Berkson error–prone data and 3) the simultaneous consideration of measurements with classical and Berkson error.

The adequate definition of an exposure model within error correction methods is essential. However, this step is exacerbated with longitudinal data due to often limited transferability of the results from other studies and due to complex, individual–specific structures and is especially challenging for indoor exposure.

The emphasis was on general linear mixed models, i.e. normally distributed model errors following an AR(1) process were assumed, to receive an impression for the impact of measurement errors on that type of model. Moreover, the measurement errors were supposed to be non–differential, which is fulfilled in the Augsburger Umweltstudie, and to exhibit a certain structure, which indeed extends the usual structure, but still may be too simplistic. The complex model structure enabled valuable insights into the role of exposure measurement error during the analysis of the association between exposure and human health.

In general, the AR(1) assumptions for model errors, measurement errors and Berkson error–prone measurements bears two unattractive consequences: 1) the true exposure measurements and the naive model errors do not follow an AR(1) process and 2) the AR(1) assumption does not hold for temporally aggregated data or for analyses on a different temporal resolution.

Assessing the sizes of measurement errors

Deriving information about the measurement error from repeated measurements in a linear mixed model setting is elaborate and hardly realizable, because many repetitions and time points are necessary. Information from validation studies or comparison measurements seem to be more promising as shown with the practical example of the Augsburger Umweltstudie.

Simultaneous modeling of mobile and fixed–site exposure measurements involves the restriction that assumptions regarding the size of Berkson and classical measurement error have to add up to the measurement error underlying the observed data. Thus, either information about Berkson or information about classical measurement error gained from external validation studies can be used. We decided to use the information from the comparison measurements for determining the classical measurement error in the application of the developed methods to the Augsburger Umweltstudie. This proceeding involves two questionable issues:

1. The individual–specific classical measurement error is assessed through the variability between battery changes, which may be too conservative.

2. There is not any gold standard method for the measurement of PM available; all used measurement devices exhibited some measurement error.

Method of moments

The effect of Berkson and classical measurement error is well known for the linear model but not for complex models with AR(1)-errors and random effects. Expressions for the attenuation factor in general simple and multiple linear mixed models for longitudinal unbalanced and unequally spaced data with covariate measurement error were derived. Analogously to the linear model, Berkson error–prone measurements yield unbiased effect estimations but the error variance of the model and the variance of the random effects are overestimated involving wider confidence intervals. The size of attenuation of the effect estimation resulting for a covariate with classical measurement error may change with autocorrelated errors in both directions. Each variance included in the attenuation factor is therefore modified with a multiplicative factor depending on the correlation coefficients. Individual–specific components in the error structure can be neglected in the bias calculation for classical measurement error, if the number of observations for each cluster is high enough. Expression for a small number of observations per individual were also derived. An explicit formula for the probability limit of the bias in the presence of additional covariates would be very complex, but the information about the measurement error was directly incorporated in the estimations equations, which were computationally solved. Mixture error has similar characteristics as classical measurement error. However, the main difference is, that the individual–specific components of the error strengthen attenuation. Furthermore, autocorrelated or individual–specific Berkson error has a complex impact on the estimation of the effect using data with mixture error. Our simulations demonstrated strong differences regarding the attenuation of the effect estimates depending on the presence of random effects and autocorrelation within the data and within the measurement error.

Error correction in the considered data situation through the method–of–moments provides only an approximation, because the estimation of the correlation structure of the model error term cannot be adapted. The presented approach is extensible to more complex correlation structures.

Although longitudinal data with covariate measurement error have already been examined in the past mostly by numerically solving possibly modified score equations, our considerations reveal valuable insights into the components and properties triggering the attenuation of the effect estimate. We presented basic relationships between covariate measurement error and the results of linear regression analyses for longitudinal data amongst strong assumptions regarding the health model and the measurement errors. These assumptions permitted us to comprehend the effects of measurement error a linear mixed model in detail within a defined model class. The knowledge about the attenuation of the effect estimate and its exact calculation even in our complex but practically relevant data situation and even under several restrictive assumptions enables the easy evaluation of the relevance of the measurement error. Our approach elucidates the complex effect of the measurement error in a transparent way in contrast to other methods, e.g. (RE)ML or Bayes estimation. Furthermore, the extensions of the approach, e.g. to several covariates, unbalanced observations and missing values, and the adequate allowance for the measurement error in the calculation of the variability of the effect estimate are realistic problems and of high practical relevance in the analysis of longitudinal data with measurement error.

The usage of a more complex exposure model, like a two–stage exposure model consisting of a background exposure model and an activity model, and the usage of a hierarchical multi–level model, as e.g. obtained through the inclusion of a random patient effect in addition to the random effect for each measurement series, requires further, possibly challenging methodological work.

In general, the method of moments is a simple, effective and practically established method for the correction of measurement error (Carroll et al., 2006) requiring only little computational demand.

Introducing prior knowledge in Bayesian models for measurement error correction

We examined the two–stage Bayesian linear mixed regression models with autocorrelated error terms with regard to covariate measurement error theoretically as well as with simulations. The main focus was the adequate introduction of prior knowledge about the measurement error.

The informativeness of a prior distribution of a certain parameter strongly depends on the data situation, the model and the prior distribution chosen for the other parameters. Broad literature is available for the adequate choice of non–informative priors, but only few research works are dedicated to the issue of informative prior choice especially in the field of covariate measurement error where prior knowledge about the measurement error is essential for estimating the non–identifiable model. Usually, the uncertainty of the prior distribution reflects the confidence in the prior knowledge in comparison to the information gained through the data. The transfer of this concept to Bayesian regression models for measurement error fails since the data lack information about the latent parameters as our theoretical considerations and simulations showed. Instead, the prior distributions of the non–identifiable parameters include the complete information about these parameters and the data is not relevant; this information can only be influenced by indirect learning, i.e. by the prior distributions of the remaining parameters or by the mutual interaction between the parameters through constraints. We defined three types of indirect learning: (I) IL from prior distributions on the parameters of the main outcome model, (II) IL induced by constraints on parameters in the main outcome model and (III) IL induced by constraints on parameters in the exposure model. In general, indirect learning changes the originally specified marginal prior distributions of the non–identifiable parameters. The concept applies also for prior knowledge from internal or external validation studies.

The uncertainty about the parameters describing the measurement error had only a diminutive effect on the variance of β_X in our simulations as well as the shape of the prior distribution; similar results are found by (Dellaportas and Stephens, 1995). The performance of Bayesian methods was similar to the frequentist approach for Berkson and classical error–prone data and yielded better results when both data sources are included. Using classical error–prone measurements with missing values provided a better performance than using Berkson error–prone measurements in Scenario (III), for less or equal than 60 % of missing values in Scenario (II) and for less or equal than 40 % of missing values in Scenario (I).

Allowing for mismeasured covariates by a Bayesian regression carries several advantages. Measurement error correction is accomplished simultaneously with model calculation including the implicit adaption of the variance of the estimated effect coefficient. Knowledge about the measurement error as well as the attendant uncertainty is directly included into the estimation procedure; therefore, the specification of the prior distribution regarding the size of the measurement error will be simple in practice, if IL is adequately consi-

dered. Compared to the elaborate or often impossible derivation of the attenuation factor for the method of moments in complex models (e.g. hierarchical models, non–linear effects of the error–prone covariate, multiple correlated, error–prone predictors), extending the Bayesian method is simple. The Bayesian approach allows the simultaneous modeling of several exposure measurements involving an improvement of the estimates also in comparison to the frequentist approach. Estimates for true measurements are provided and missing values of the personal measurements can be simultaneously estimated. Moreover, different types of exposure measurements may be incorporated into the model and improve the results. In comparison to frequentist approaches, like the method–of–moments or ML, the application of a Bayesian model for the correction of covariate measurement error has two major advantages: 1) increased flexibility through the possibility of specifying prior distributions on the non–identifiable parameters and 2) the possibility to combine various sources of information (Carroll et al., 2006).

The Bayesian approach has also some disadvantages. Undesired effects of IL may occur, if the model is specified in an inapt way. Furthermore, the flexibility of Bayesian models may entice the analyst to apply this approach in situations with little prior knowledge about the size and the structure of the measurement error by using less informative prior distributions for the parameters describing the exposure model and the classical measurement error. However, the misspecification of the underlying structure of the measurements may cause severely biased results. This is also observed by Richardson et al. (2002) and Wang and Sullivan Pepe (2000), who propose a mixture of distributions as prior distribution for the unknown, true measurements to allow a higher flexibility (see also Carroll et al., 1999a,b). In contrast, the frequentist approach directly urges the analyst to give thoughts to the underlying structure. Convergence problems appear in situations with implausible prior knowledge, which is not corrected through IL from prior distributions of other non–identifiable parameters. Long computing times for complex models and no detailed understanding of the impact of measurement error on the regression model are other weaknesses of the Bayesian method.

Application

Applying the correction methods to the Augsburger Umweltstudie revealed slight differences in the effect coefficients compared to the naive estimations. The examined complex error structure had little impact on the investigated effect estimations in the Augsburger Umweltstudie. Thus, the conclusions deduced from the original naive analyses persist.

The combination of the information from the main study and the validation studies were accomplished with a two–stage procedure: (I) modeling with deficient data and (II) error correction with validation data. Bayesian analyses presented in Section 5 and Subsection 6.2 combine these steps in one model.

In order to combine the data from the Augsburger Umweltstudie with the data from the validation study and the comparison measurements a Bayesian model was established. The Bayesian effect estimates were similar to the MOM–estimates for comparable models. Indirect learning type (III) occurs in simultaneous models for Berkson and classical error–prone data, but only slightly affected the estimated effect.

The very flexible, two–stage exposure model for the Bayesian model based on the previously developed exposure model for the validation study. However, logarithmically transformed exposure measurements were used in the Bayesian model to obtain additive measurement error, which is easier to handle than multiplicative error, but this approach required to presume multiplicative effects of the covariates in the exposure model. Indeed, methods for the inclusion of variable selection and the choice of the complexity of smooth

effects into the Bayesian model are available, but their application to each data source would enormously magnify the complexity of the model. Moreover, we abstained from using a rational quadratic correlation structure in favor of a sparse model. We only considered separate Berkson errors for different seasons and for staying indoors or outdoors, but not activity–specific or more detailed microenvironment–specific Berkson errors.

The outdoor observations in the Augsburger Umwelstudie were categorized in "staying in traffic" and "staying not in traffic", whereas the outdoor observations in the validation study can be differentiated into the categories "staying not in traffic", "staying in traffic as pedestrian", "staying in traffic with a car" and "staying in traffic with public transport". The background model for the Augsburger Umweltstudie was based on the observations during "staying in traffic", whereas the observations during "staying in traffic" as a pedestrian were used in the background model for the validation study, because only few observations during "staying not in traffic" were collected in the validation study.

The different data sources are solely connected through the non–identifiable parameters. However, also assuming equal distributions for other parameters, like activity effects, may be reasonable.

We have shown with our data example, that the specification of all parameters, which are necessary for the error correction, is possible through estimations from validation studies or from comparison measurements.

7.2. Outlook

The developed bias correction formulas are generally applicable to the considered types of covariate measurement errors in linear mixed models with a first–order autocorrelated error term. The theoretical, general considerations are transferable to other longitudinal studies. The method–of–moments for multiple regression can easily be extended to several error–prone covariates, like e.g. in multi–pollutant models, as proposed by McShane et al. (2001) for the analysis of matched case–control studies.

Moreover, the concept of the mixture error can be generalized to observations with differing measurement errors, like indoor and outdoor observations or observations from different microenvironments. Thus, $\boldsymbol{G}$ can be seen as an arbitrary categorical covariate; the size and structure of the measurement error has to be known for each category. Even temporally or spatially varying Berkson error are supposable through an adequate definition of $\boldsymbol{\Sigma}_{U^{\mathrm{B}}}$ depending on time and space.

Our considerations regarding the introduction of prior knowledge about latent parameters in not–identifiable models are of a generic nature and can be applied to other problems with latent parameters, e.g. to sensitivity analyses regarding unmeasured confounder variables (de Vocht et al., 2009; Steenland and Greenland, 2004). However, the transfer to more complex models or situations like e.g. to categorical covariates measured with error or to response variables, which do not follow a conditional Normal distribution requires further research. Especially, indirect learning from constraints regarding the domains of some parameters might pose a crucial issue in that situations.

We motivated our work through the Augsburger Umwelstudie for which only measurements from a single fixed–site monitor are available. Therefore, the spatial heterogeneity of ambient exposure is only insufficiently covered and the Berkson error is large. For example, Bliznyuk et al. (2014) and Gryparis et al. (2007b) propose a Bayesian approach for including data from several sources for the prediction of ambient exposure. Another possibility to cover spatial and temporal variation is the usage of predicted exposure form an adequate exposure model. The prediction uncertainty represents the size of the Berkson error and can directly be used for calculating the variance of a MOM–corrected

estimate in a model with mixture error–prone data or as prior uncertainty in a Bayesian model. Since we did not detect any lagged effects of personal exposure to PNC on the heart rate, we did not pursue more advanced methods for modeling lagged effects. The extension of our Bayesian model by distributed lagged effects may be easy to accomplish, but the effect of indirect learning of the imposed structure has to be clarified.

The theoretical achievements for correcting covariate measurement error in linear mixed models, which are exemplary applied to a single health outcome (heart rate) of the Augsburger Umweltstudie, provide the methodological background for correcting various other health outcomes analyzed within this study as well as for correcting the biases in analyses with error–prone mobile temperature and relative humidity as covariates. Personal exposure to air temperature or relative humidity, on which data was also collected within the Augsburger Umweltstudie and the validation study, is of growing interest, but its association with human health was only rarely examined in the past (Keatinge et al., 1986; Lanzinger et al., 2014; Nguyen et al., 2015).

A. Two–stage personal exposure model: model equations

A.1. Stage 1: Background model (before variable selection)

The regression equations for modeling PNC of individual i at time point t with the maximum set of predictors are presented in the following.

Model for indoor observations:

$$\begin{aligned}\text{PNC}_{it} =& \beta_0 + \mathbf{week}_{it}\boldsymbol{\beta}_1 + \mathbf{day\ of\ the\ week}_{it}\boldsymbol{\beta}_2 + \text{at home}_{it}\beta_3 + \\ & \text{not at home}_{it}\beta_4 + \mathbf{window}_{it}\boldsymbol{\beta}_5 + f_1(\text{time}_{it}) + f_2(\text{temperature}_{it}) + \\ & f_3(\text{rel. humidity}_{it}) + f_4(\text{dew point temperature}_{it}) + f_5(\text{ambient PM}_{2.5it}) + \\ & f_6(\text{ambient PM}_{10it}) + f_7(\text{ambient PNC}_{CPCit}) + f_8(\text{ambient BC}_{it}) + \\ & f_9(\text{ambient PNC}_{3-10nmit}) + f_{10}(\text{ambient PNC}_{10-30nmit}) + \\ & f_{11}(\text{ambient PNC}_{30-50nmit}) + f_{12}(\text{ambient PNC}_{50-100nmit}) + \\ & f_{13}(\text{ambient PNC}_{100-500nmit}) + f_{14}(\text{ambient PNC}_{500-1000nmit}) + \\ & f_{15}(\text{ambient PNC}_{1000-2500nmit}) + f_{16}(\text{ambient PNC}_{2500-10000nmit}) + \\ & f_{17}(\text{ambient PNC}_{10-300nmit}) + f_{18}(\text{ambient PNC}_{10-800nmit}) + \\ & f_{19}(\text{ambient PNC}_{PM_1it}) + f_{20}(\text{ambient PNC}_{PM_{2.5}it}) + \\ & f_{21}(\text{ambient PNC}_{PM_{10}it}) + \tau_i + \varepsilon_{it}\end{aligned}$$

Model for outdoor observations:

$$\begin{aligned}\text{PNC}_{it} =& \beta_0 + \mathbf{week}_{it}\boldsymbol{\beta}_1 + \mathbf{day\ of\ the\ week}_{it}\boldsymbol{\beta}_2 + f_1(\text{time}_{it}) + f_2(\text{temperature}_{it}) + \\ & f_3(\text{rel. humidity}_{it}) + f_4(\text{dew point temperature}_{it}) + f_5(\text{ambient PM}_{2.5it}) + \\ & f_6(\text{ambient PM}_{10it}) + f_7(\text{ambient PNC}_{CPCit}) + f_8(\text{ambient BC}_{it}) + \\ & f_9(\text{ambient PNC}_{3-10nmit}) + f_{10}(\text{ambient PNC}_{10-30nmit}) + \\ & f_{11}(\text{ambient PNC}_{30-50nmit}) + f_{12}(\text{ambient PNC}_{50-100nmit}) + \\ & f_{13}(\text{ambient PNC}_{100-500nmit}) + f_{14}(\text{ambient PNC}_{500-1000nmit}) + \\ & f_{15}(\text{ambient PNC}_{1000-2500nmit}) + f_{16}(\text{ambient PNC}_{2500-10000nmit}) + \\ & f_{17}(\text{ambient PNC}_{10-300nmit}) + f_{18}(\text{ambient PNC}_{10-800nmit}) + \\ & f_{19}(\text{ambient PNC}_{PM_1it}) + f_{20}(\text{ambient PNC}_{PM_{2.5}it}) + \\ & f_{21}(\text{ambient PNC}_{PM_{10}it}) + \tau_i + \varepsilon_{it}\end{aligned}$$

A.2. Stage 2: Activity model

The regression equations for modeling the residual (resulting from the background model) of realization j of activity k at time point t are presented in the following: Model for indoor observations:

$$\begin{aligned}\text{residual}_{k_j t} =& \beta_0 + \beta_1 \text{smoke of wood}_{k_j t} + \beta_2 \text{smoke of candles}_{k_j t} + \beta_3 \text{cooking}_{k_j t} + \\ & \beta_4 \text{cleaning}_{k_j t} + \beta_5 \text{ironing}_{k_j t} + \beta_6 \text{spray}_{k_j t} + \beta_7 \text{dust}_{k_j t} + \beta_8 \text{smell of food}_{k_j t} + \\ & \beta_9 \text{steam}_{k_j t} + \mathbf{season}\boldsymbol{\beta}_{10} + \boldsymbol{b}_{k_j} + \varepsilon_{k_j t}\end{aligned}$$

Model for outdoor observations:

$$\begin{aligned}\text{residual}_{k_j t} =& \beta_0 + \mathbf{means\ of\ traffic}_{k_j t}\boldsymbol{\beta}_1 + \beta_2 \text{construction area}_{k_j t} + \\ & \beta_3 \text{smell of food}_{k_j t} + \beta_4 \text{smell of smoke}_{k_j t} + \beta_5 \text{smell of vehicle exhaust}_{k_j t} + \\ & \mathbf{season}\boldsymbol{\beta}_6 + \boldsymbol{b}_{k_j} + \varepsilon_{k_j t}\end{aligned}$$

B. Derivations

B.1. Coefficient estimates for a single activity effect assuming a block–diagonal AR(1) correlation matrix for the error term

A block–diagonal correlation matrix for $\boldsymbol{\varepsilon}$ is considered:

$$\boldsymbol{\Sigma} = \text{diag}(\boldsymbol{W}_{\rho,1}, \boldsymbol{W}_{\rho,2}, \boldsymbol{W}_{\rho,3}).$$

With

$$\boldsymbol{X}^\top \boldsymbol{W}_\rho^{-1} \boldsymbol{X} = \frac{1}{1-\rho^2} \begin{pmatrix} 6(1-\rho) + (T-6)(1-\rho)^2 & 2(1-\rho) + (n_a-2)(1-\rho)^2 \\ 2(1-\rho) + (n_a-2)(1-\rho)^2 & 2(1-\rho) + (n_a-2)(1-\rho)^2 \end{pmatrix}$$

and

$$\begin{aligned} |\boldsymbol{X}^\top \boldsymbol{W}_\rho^{-1} \boldsymbol{X}| =& \frac{1}{(1-\rho^2)^2} \left[2(1-\rho) + (n_a-2)(1-\rho)^2\right] \left[4(1-\rho) + (T-n_a-4)(1-\rho)^2\right] \\ =:& \frac{1}{(1-\rho^2)^2} D_2^{-1} \end{aligned}$$

the effect estimate is given by

$$\begin{aligned} \widehat{\boldsymbol{\beta}} =& D \cdot \\ & \begin{pmatrix} 2(1-\rho) + (n_a-2)(1-\rho)^2 & -2(1-\rho) - (n_a-2)(1-\rho)^2 \\ -2(1-\rho) - (n_a-2)(1-\rho)^2 & 6(1-\rho) + (T-6)(1-\rho)^2 \end{pmatrix} \cdot \\ & \begin{pmatrix} (1-\rho)(Y_1 + Y_{t_{a_1}-1} + Y_{t_{a_1}} + Y_{t_{a_S}} + Y_{t_{a_S}+1} + Y_T) \\ \quad +(1-\rho)^2(\sum_{t=2}^{t_{a_1}-2} Y_t + \sum_{t=t_{a_1}+1}^{t_{a_S}-1} Y_t + \sum_{t=t_{a_1}+2}^{T-1} Y_t) \\ \quad (1-\rho)(Y_{t_{a_1}} + Y_{t_{a_S}}) + (1-\rho)^2 \sum_{t=t_{a_1}+1}^{t_{a_S}-1} Y_t \end{pmatrix} \\ \widehat{\beta}_0 =& D \cdot \\ & \left[2(1-\rho) + (n_a-2)(1-\rho)^2\right] \\ & \left[(1-\rho)(Y_1 + Y_{t_{a_1}-1} + Y_{t_{a_S}+1} + Y_T) + (1-\rho)^2 \left(\sum_{t=2}^{t_{a_1}-2} Y_t + \sum_{t=t_{a_S}+2}^{T-1} Y_t \right) \right] \\ \widehat{\beta}_1 =& D \cdot \\ & \Big[(Y_1 + Y_{t_{a_1}-1} + Y_{t_{a_S}+1} + Y_T)(1-\rho)^2 \left(-(1-\rho)(n_a-2) - 2\right) + \\ & \left(\sum_{t=2}^{t_{a_1}-2} Y_t + \sum_{t=t_{a_S}+2}^{T-1} Y_t \right) (1-\rho)^3 \left(-(1-\rho)(n_a-2) - 2\right) + \\ & (Y_{t_{a_1}} + Y_{t_{a_S}})(1-\rho)^2 \left(4 + (T-n_a-4)(1-\rho)\right) - \\ & \left(\sum_{t=t_{a_1}+1}^{t_{a_S}-1} Y_t \right) (1-\rho)^3 \left(4 + (T-n_a-4)(1-\rho)\right) \Big] \end{aligned}$$

B.2. Covariance matrix of two autocorrelated processes of order 1

Let $\boldsymbol{\varepsilon}^{\mathrm{A}} \sim \mathrm{N}(0, \boldsymbol{\Sigma}_{\mathrm{A}})$ with $\boldsymbol{\Sigma}_{\mathrm{A}} = \sigma_{\mathrm{A}}^2 \boldsymbol{W}_{\rho_{\mathrm{A}}}$, $\boldsymbol{\varepsilon}^{\mathrm{B}} \sim \mathrm{N}(0, \boldsymbol{\Sigma}_{\mathrm{B}})$ with $\boldsymbol{\Sigma}_{\mathrm{B}} = \sigma_{\mathrm{B}}^2 \boldsymbol{W}_{\rho_{\mathrm{B}}}$, $\boldsymbol{\varepsilon}^{\mathrm{A}}$ and $\boldsymbol{\varepsilon}^{\mathrm{B}}$ independent and $\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}^{\mathrm{A}} + \boldsymbol{\varepsilon}^{\mathrm{B}}$. For $\rho_{\mathrm{A}}, \rho_{\mathrm{B}} \neq 0$,

$$\begin{aligned} \mathrm{Var}(\boldsymbol{\varepsilon}) &= \sigma_{\mathrm{A}}^2 \boldsymbol{W}_{\rho_{\mathrm{A}}} + \sigma_{\mathrm{B}}^2 \boldsymbol{W}_{\rho_{\mathrm{B}}} \\ &\neq (\sigma_{\mathrm{A}}^2 + \sigma_{\mathrm{A}}^2) \boldsymbol{W}_{\rho_{\mathrm{AB}}} \end{aligned}$$

for any arbitrary choice of $\rho_{\mathrm{AB}} > 0$.

B.3. Auxiliary calculations for correlation matrices

$$\boldsymbol{W}_{\rho^*}^{-1} \overset{\text{eq. 2.2}}{=} \frac{1}{1-\rho^{*2}} \begin{pmatrix} 1 & -\rho^* & 0 & \cdots & 0 & 0 \\ -\rho^* & 1+\rho^{*2} & -\rho^* & \cdots & 0 & 0 \\ 0 & -\rho^* & 1+\rho^{*2} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1+\rho^{*2} & -\rho^* \\ 0 & 0 & 0 & \cdots & -\rho^* & 1 \end{pmatrix}$$

$$\frac{\partial \boldsymbol{W}_{\rho^*}}{\partial \rho^*} = \begin{pmatrix} 0 & 1 & 2\rho^* & \cdots & (T-2)\rho^{*T-3} & (T-1)\rho^{*T-2} \\ 1 & 0 & 1 & \cdots & (T-3)\rho^{*T-4} & (T-2)\rho^{*T-3} \\ 2\rho^* & 1 & 0 & \cdots & (T-4)\rho^{*T-5} & (T-3)\rho^{*T-4} \\ \vdots & \vdots & \vdots & \ddots & \cdots & \cdots \\ (T-2)\rho^{*T-3} & (T-3)\rho^{*T-4} & (T-4)\rho^{*T-5} & \cdots & 0 & 1 \\ (T-1)\rho^{*T-2} & (T-2)\rho^{*T-3} & (T-3)\rho^{*T-4} & \cdots & 1 & 0 \end{pmatrix}$$

$$\boldsymbol{W}_{\rho^*}^{-1}\frac{\partial \boldsymbol{W}_{\rho^*}}{\partial \rho^*} = \frac{1}{1-\rho^{*2}}\begin{pmatrix} -\rho^* & 1 & \rho^* & \rho^{*2} & \rho^{*3} & \cdots & \rho^{*T-2} \\ 1-\rho^{*2} & -2\rho^* & 1-\rho^{*2} & \rho^*(1-\rho^{*2}) & \rho^{*2}(1-\rho^{*2}) & \cdots & \rho^{*T-3}(1-\rho^{*2}) \\ \rho^*(1-\rho^{*2}) & 1-\rho^{*2} & -2\rho^* & 1-\rho^{*2} & \rho^*(1-\rho^{*2}) & \cdots & \rho^{*T-4}(1-\rho^{*2}) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{*T-3}(1-\rho^{*2}) & \cdots & \rho^{*2}(1-\rho^{*2}) & \rho^*(1-\rho^{*2}) & 1-\rho^{*2} & -2\rho^* & 1-\rho^{*2} \\ \rho^{*T-2} & \cdots & \rho^{*3} & \rho^{*2} & \rho^* & 1 & -\rho^* \end{pmatrix}$$

$$\boldsymbol{W}_{\rho^*}^{-1}\boldsymbol{W}_{\rho} = \frac{1}{1-\rho^{*2}}\begin{pmatrix} 1-\rho^*\rho & \rho-\rho^* & \rho(\rho-\rho^*) & \rho^2(\rho-\rho^*) & \cdots & \rho_a^{(T-2)}(\rho_a-\rho^*) \\ \rho(1+\rho^{*2})-\rho^*(1+\rho^2) & (1+\rho^{*2})-2\rho^*\rho & \rho(1+\rho^{*2})-\rho^*(1+\rho^2) & \rho[\rho(1+\rho^{*2})-\rho^*(1+\rho^2)] & \cdots & \rho^{(T-3)}[\rho(1+\rho^{*2})-\rho^*(1+\rho^2)] \\ \rho^2[\rho(1+\rho^{*2})-\rho^*(1+\rho^2)] & \rho(1+\rho^{*2})-\rho^*(1+\rho^2) & (1+\rho^{*2})-2\rho^*\rho & \rho(1+\rho^{*2})-\rho^*(1+\rho^2) & \cdots & \rho^{(T-4)}[\rho(1+\rho^{*2})-\rho^*(1+\rho^2)] \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{(T-3)}[\rho(1+\rho^{*2})-\rho^*(1+\rho^2)] & \cdots & \rho[\rho(1+\rho^{*2})-\rho^*(1+\rho^2)] & \rho(1+\rho^{*2})-\rho^*(1+\rho^2) & (1+\rho^{*2})-2\rho^*\rho & \rho(1+\rho^{*2})-\rho^*(1+\rho^2) \\ \rho^{(T-2)}(\rho-\rho^*) & \cdots & \rho^2(\rho-\rho^*) & \rho(\rho-\rho^*) & \rho-\rho^* & 1-\rho^*\rho \end{pmatrix}$$

$$
\begin{aligned}
&\operatorname{tr}\left(\boldsymbol{W}_{\rho^*}^{-1}\frac{\partial \boldsymbol{W}_{\rho^*}}{\partial \rho^*}\boldsymbol{W}_{\rho^*}^{-1}\boldsymbol{\Sigma}_a\right)\\
=&\sum_{i=1}^{T}\sum_{j=1}^{T}\left(\left(\boldsymbol{W}_{\rho^*}^{-1}\frac{\partial \boldsymbol{W}_{\rho^*}}{\partial \rho^*}\right)_{ij}(\boldsymbol{W}_{\rho^*}^{-1}\boldsymbol{\Sigma}_a)_{ji}\right)\\
=&\frac{\sigma_a^2}{(1-\rho^{*2})^2}\Big\{-2\rho^*(1-\rho^*\rho_a)+2\rho^{*T-2}\rho_a^{T-2}(\rho_a-\rho^*)+\\
&\qquad 2\sum_{t=2}^{T-1}\rho^{*t-2}\rho_a^{t-2}\left(\rho_a(1+\rho^{*2})-\rho^*(1+\rho_a^2)\right)+\\
&\qquad 2(1-\rho^{*2})(\rho_a-\rho^*)\sum_{t=2}^{T-1}(\rho^*\rho_a)^{t-2}-\\
&\qquad 2(T-2)\rho^*(1+\rho^{*2}-2\rho^*\rho_a)+\\
&\qquad 2(1-\rho^{*2})\left[(1+\rho^{*2})\rho_a-\rho^*(1+\rho_a^2)\right]\sum_{k=1}^{T-3}\sum_{t=1}^{k}(\rho_a\rho^*)^{t-1}\Big\}\\
=&\frac{2\sigma_a^2}{(1-\rho^{*2})^2}\Big\{-\rho^*(1-\rho^*\rho_a)+\rho^{*T-2}\rho_a^{T-2}(\rho_a-\rho^*)+\\
&\qquad \left(\rho_a(1+\rho^{*2})-\rho^*(1+\rho_a^2)\right)\frac{1-\rho^{*T-2}\rho_a^{T-2}}{1-\rho^*\rho_a}+\\
&\qquad (1-\rho^{*2})(\rho_a-\rho^*)\frac{1-\rho^{*T-2}\rho_a^{T-2}}{1-\rho^*\rho_a}-\\
&\qquad (T-2)\rho^*(1+\rho^{*2}-2\rho^*\rho_a)+\\
&\qquad (1-\rho^{*2})\left[(1+\rho^{*2})\rho_a-\rho^*(1+\rho_a^2)\right]\frac{1}{1-\rho^*\rho_a}\left[(T-3)-\frac{\rho^*\rho_a-(\rho^*\rho_a)^{T-2}}{1-\rho^*\rho_a}\right]\Big\}\\
=&\frac{2\sigma_a^2}{(1-\rho^{*2})^2}\Big\{-\rho^*(1-\rho^*\rho_a)-\left(1-\rho^{*T-2}\rho_a^{T-2}\right)(\rho_a-\rho^*)+\rho_a-\rho^*-(T-2)\rho^*(1+\rho^{*2}-2\rho^*\rho_a)+\\
&\qquad \frac{1-\rho^{*T-2}\rho_a^{T-2}}{1-\rho^*\rho_a}\left(\rho_a(1+\rho^{*2})-\rho^*(1+\rho_a^2)+(1-\rho^{*2})(\rho_a-\rho^*)\right)+\\
&\qquad (1-\rho^{*2})(\rho_a-\rho^*)\left[(T-2)-\frac{1-(\rho^*\rho_a)^{T-2}}{1-\rho^*\rho_a}\right]\Big\}\\
=&\frac{2\sigma_a^2}{(1-\rho^{*2})^2}\Big\{-\rho^*(1-\rho^*\rho_a)-(T-2)\rho^*(1+\rho^{*2}-2\rho^*\rho_a)+(T-2)(1-\rho^{*2})(\rho_a-\rho^*)+\rho_a-\rho^*+\\
&\qquad \frac{1-\rho^{*T-2}\rho_a^{T-2}}{1-\rho^*\rho_a}\left(\rho_a(1+\rho^{*2})-\rho^*(1+\rho_a^2)-(\rho_a-\rho^*)(1-\rho^*\rho_a)\right)\Big\}\\
=&\frac{2(T-1)\sigma_a^2}{(1-\rho^{*2})^2}\left[-2\rho^*+\rho^{*2}\rho_a+\rho_a\right]
\end{aligned}
$$

B.4. Solving eq. 3.14 and eq. 3.15 for ρ^* and $\sigma^2_{\varepsilon^*}$

Eq. 3.15 can be further simplified using the auxiliary calculations in Appendix B.3:

$$
\begin{aligned}
& \operatorname{tr}\left(\boldsymbol{W}_{\rho^*}^{-1}\frac{\partial \boldsymbol{W}_{\rho^*}}{\partial \rho^*}\boldsymbol{W}_{\rho^*}^{-1}\boldsymbol{V}_T\right) = \frac{1}{T}\operatorname{tr}\left(\boldsymbol{W}_{\rho^*}^{-1}\frac{\partial \boldsymbol{W}_{\rho^*}}{\partial \rho^*}\right)\operatorname{tr}\left(\boldsymbol{W}_{\rho^*}^{-1}\boldsymbol{V}_T\right) \\
\Leftrightarrow\quad & \frac{2(T-1)\sigma_a^2}{(1-\rho^{*2})^2}\left[-2\rho^* + \rho^{*2}\rho_a + \rho_a\right] + \\
& \frac{2(T-1)\sigma_b^2}{(1-\rho^{*2})^2}\left[-2\rho^* + \rho^{*2}\rho_b + \rho_b\right] = \\
& -\frac{2(T-1)\rho^*}{T(1-\rho^{*2})^2}\left[\sigma_a^2\left(2(1-\rho^*\rho_a) + (T-2)(1+\rho^{*2}-2\rho^*\rho_a)\right) + \right. \\
& \qquad \left. \sigma_b^2\left(2(1-\rho^*\rho_b) + (T-2)(1+\rho^{*2}-2\rho^*\rho_b)\right)\right] \\
\Leftrightarrow\quad & \sigma_a^2\left[-(T-2)\rho^{*2}\rho_a + T\rho_a - T\rho^* + (T-2)\rho^{*3}\right] + \\
& \sigma_b^2\left[-(T-2)\rho^{*2}\rho_b + T\rho_b - T\rho^* + (T-2)\rho^{*3}\right] = 0 \\
\Leftrightarrow\quad & \sigma_a^2\left[\rho_a(T-(T-2)\rho^{*2}) - \rho^*(T-(T-2)\rho^{*2})\right] + \\
& \sigma_b^2\left[\rho_b(T-(T-2)\rho^{*2}) - \rho^*(T-(T-2)\rho^{*2})\right] = 0 \\
\Rightarrow\quad & \rho^* = \frac{\rho_a\sigma_a^2 + \rho_b\sigma_b^2}{\sigma_a^2 + \sigma_b^2} \\
& \sigma^2_{\varepsilon^*} = \sigma_a^2\frac{g^*_{1,T}(\rho_a)}{T(1-\rho^{*2})} + \sigma_b^2\frac{g^*_{1,T}(\rho_b)}{T(1-\rho^{*2})} \\
& \qquad = \frac{1}{T(1-\rho^{*2})}\left[\sigma_a^2 g^*_{1,T}(\rho_a) + \sigma_b^2 g^*_{1,T}(\rho_b)\right] \\
& \qquad = \frac{(\sigma_a^2+\sigma_b^2)^2}{T\left[(\sigma_a^2+\sigma_b^2)^2 - (\rho_a\sigma_a^2+\rho_b\sigma_b^2)^2\right]} \\
& \qquad\quad \left\{\sigma_a^2\frac{T(\sigma_a^2+\sigma_b^2)^2 - 2(T-1)\rho_a(\sigma_a^2+\sigma_b^2)(\rho_a\sigma_a^2+\rho_b\sigma_b^2) + (T-2)(\rho_a\sigma_a^2+\rho_b\sigma_b^2)^2}{(\sigma_a^2+\sigma_b^2)^2} + \right. \\
& \qquad\quad \left. \sigma_b^2\frac{T(\sigma_a^2+\sigma_b^2)^2 - 2(T-1)\rho_b(\sigma_a^2+\sigma_b^2)(\rho_a\sigma_a^2+\rho_b\sigma_b^2) + (T-2)(\rho_a\sigma_a^2+\rho_b\sigma_b^2)^2}{(\sigma_a^2+\sigma_b^2)^2}\right\} \\
& \qquad = \frac{(\sigma_a^2+\sigma_b^2)\left[T(\sigma_a^2+\sigma_b^2)^2 - 2(T-1)(\rho_a\sigma_a^2+\rho_b\sigma_b^2)^2 + (T-2)(\rho_a\sigma_a^2+\rho_b\sigma_b^2)^2\right]}{T\left[(\sigma_a^2+\sigma_b^2)^2 - (\rho_a\sigma_a^2+\rho_b\sigma_b^2)^2\right]} \\
& \qquad = \sigma_a^2 + \sigma_b^2
\end{aligned}
$$

with $g^*_{1,T}(\rho) := T - 2(T-1)\rho\rho^* + (T-2)\rho^{*2}$.

B.5. Simplification of equation 3.27 for $\rho^{X^{*\mathrm{B}}} = \rho^{\mathrm{B}} = \rho^{\mathrm{C}} = \rho = 0$

$\rho^{X^{*\mathrm{B}}} = \rho^{\mathrm{B}} = \rho^{\mathrm{C}} = \rho = 0$ implicates that also $\rho^{*\mathrm{C}} = 0$. Therefore,

$$g_T^{*\mathrm{C}}(0) = 1 - \frac{T}{T\left(T + \frac{\sigma^2_{\varepsilon^{*\mathrm{C}}}}{\sigma^2_{\tau^{*\mathrm{C}}}}\right)} = \frac{\sigma^2_{\varepsilon^{*\mathrm{C}}}\left[1 + (T-1)\frac{\sigma^2_{\tau^{*\mathrm{C}}}}{\sigma^2_{\varepsilon^{*\mathrm{C}}}}\right]}{T\sigma^2_{\tau^{*\mathrm{C}}} + \sigma^2_{\varepsilon^{*\mathrm{C}}}}$$
$$g_T^{*\mathrm{C}}(1) = 1 - \frac{T^2}{T\left(T + \frac{\sigma^2_{\varepsilon^{*\mathrm{C}}}}{\sigma^2_{\tau^{*\mathrm{C}}}}\right)} = \frac{\sigma^2_{\varepsilon^{*\mathrm{C}}}}{T\sigma^2_{\tau^{*\mathrm{C}}} + \sigma^2_{\varepsilon^{*\mathrm{C}}}}.$$

The attenuation factor results as

$$\begin{aligned}\lambda^{\mathrm{C}} &= \frac{\sigma^2_{X^{*\mathrm{B}}} g_T^{*\mathrm{C}}(0) + \sigma^2_{\nu^{\mathrm{B}}} g_T^{*\mathrm{C}}(1) + \sigma^2_{U^{\mathrm{B}}} g_T^{*\mathrm{C}}(0)}{\sigma^2_{X^{*\mathrm{B}}} g_T^{*\mathrm{C}}(0) + \sigma^2_{\nu^{\mathrm{B}}} g_T^{*\mathrm{C}}(1) + \sigma^2_{\nu^{\mathrm{C}}} g_T^{*\mathrm{C}}(1) + \sigma^2_{U^{\mathrm{B}}} g_T^{*\mathrm{C}}(0) + \sigma^2_{U^{\mathrm{C}}} g_T^{*\mathrm{C}}(0)} \\ &= \frac{\left[1 + (T-1)\frac{\sigma^2_{\tau^{*\mathrm{C}}}}{\sigma^2_{\varepsilon^{*\mathrm{C}}}}\right](\sigma^2_{X^{*\mathrm{B}}} + \sigma^2_{U^{\mathrm{B}}}) + \sigma^2_{\nu^{\mathrm{B}}}}{\left[1 + (T-1)\frac{\sigma^2_{\tau^{*\mathrm{C}}}}{\sigma^2_{\varepsilon^{*\mathrm{C}}}}\right]\left(\sigma^2_{X^{*\mathrm{B}}} + \sigma^2_{U^{\mathrm{B}}} + \sigma^2_{U^{\mathrm{C}}}\right) + \sigma^2_{\nu^{\mathrm{B}}} + \sigma^2_{\nu^{\mathrm{C}}}}.\end{aligned}$$

B.6. Derivation of $\lim_{T\to\infty} g_T^*(\rho)$

$g_T^*(\rho)$ is defined as

$$g_T^*(\rho) := \left[g_{1,T}^*(\rho) - \frac{\sigma^2_{\tau^*}}{\sigma^2_{\varepsilon^*}(1-\rho^{*2})d} g_{2,T}^*(\rho)\right]/T$$

with

$$\begin{aligned}g_{1,T}^*(\rho) :=& T - 2(T-1)\rho\rho^* + (T-2)\rho^{*2} \\ g_{2,T}^*(\rho) :=& 2(1-\rho^*)^2(1+\rho^{T-1}) + 4(1-\rho^*)^3 \sum_{t=1}^{T-2}\rho^t + (1-\rho^*)^4\left(T - 2 + 2\sum_{k=1}^{T-3}\sum_{t=1}^{k}\rho^t\right)\end{aligned}$$

and

$$d := 1 + \frac{\sigma^2_{\tau^*}}{\sigma^2_{\varepsilon^*}(1-\rho^{*2})}\left[(T-2)(1-\rho^*)^2 + 2(1-\rho^*)\right].$$

For $|\rho| < 1$, the geometric series can be transformed to

$$\sum_{t=1}^{T-2}\rho^t = \frac{\rho - \rho^{T-1}}{1-\rho}$$

and

$$\sum_{k=1}^{T-3}\sum_{t=1}^{k}\rho^t = \sum_{k=1}^{T-3}\frac{\rho - \rho^{k+1}}{1-\rho} = (T-3)\frac{\rho}{1-\rho} - \frac{1}{1-\rho}\sum_{k=1}^{T-3}\rho^{k+1} = (T-2)\frac{\rho}{1-\rho} - \frac{\rho - \rho^{T-1}}{(1-\rho)^2}.$$

Thus, the following approximations hold:

$$\begin{aligned}
g^*_{1,T}(\rho) =& \left[T - 2(T-1)\rho\rho^* + (T-2)\rho^{*2}\right] \\
\approx& T(1 - 2\rho\rho^* + \rho^{*2}) \\
g^*_{2,T}(\rho) =& (T-2)(1-\rho^*)^4 + 2(1+\rho^{T-1})(1-\rho^*)^2 + 4(1-\rho^*)^3\frac{\rho-\rho^{T-1}}{1-\rho} + \\
& 2(1-\rho^*)^4\left[(T-2)\frac{\rho}{1-\rho} - \frac{\rho-\rho^{T-1}}{(1-\rho)^2}\right] \\
\approx& (1-\rho^*)^2\left\{T(1-\rho^*)^2\left(1+2\frac{\rho}{1-\rho}\right) + 2 + 4(1-\rho^*)\frac{\rho}{1-\rho} - 2\rho\frac{(1-\rho^*)^2}{(1-\rho)^2}\right\} \\
\approx& (1-\rho^*)^4\frac{1+2\rho}{1-\rho}T \\
d =& 1 + \frac{\sigma^2_{\tau^*}}{\sigma^2_{\varepsilon^*}(1-\rho^{*2})}\left[(T-2)(1-\rho^*)^2 + 2(1-\rho^*)\right] \\
\approx& 1 + \frac{\sigma^2_{\tau^*}}{\sigma^2_{\varepsilon^*}(1-\rho^{*2})}T(1-\rho^*)^2 \\
=& 1 + \frac{\sigma^2_{\tau^*}}{\sigma^2_{\varepsilon^*}}\frac{1-\rho^*}{1+\rho^*}T \\
\lim_{T\to\infty} g^*_T(\rho) =& \frac{1}{T}g^*_{1,T}(\rho) - \frac{\sigma^2_{\tau^*}}{\sigma^2_{\varepsilon^*}(1-\rho^{*2})Td}g^*_{2,T}(\rho) \\
=& 1 - 2\rho\rho^* + \rho^{*2}.
\end{aligned}$$

For $\rho = 1$:

$$\begin{aligned}
g^*_{1,T}(1) \approx& T(1-\rho^*)^2 \\
g^*_{2,T}(1) \approx& T^2(1-\rho^*)^4 \\
d \approx& 1 + \frac{\sigma^2_{\tau^*}}{\sigma^2_{\varepsilon^*}(1-\rho^{*2})}T(1-\rho^*)^2 = 1 + T\frac{\sigma^2_{\tau^*}}{\sigma^2_{\varepsilon^*}}\frac{1-\rho^*}{1+\rho^*} \\
g^*_T(1) =& \frac{1}{T}g^*_{1,T}(1) - \frac{\sigma^2_{\tau^*}}{\sigma^2_{\varepsilon^*}\left(1-\rho^{*2}\right)Td}g^*_{2,T}(1) \\
\approx& (1-\rho^*)^2\left[1 - \frac{\sigma^2_{\tau^*}T(1-\rho^*)}{\sigma^2_{\varepsilon^*}(1+\rho^*) + \sigma^2_{\tau^*}T(1-\rho^*)}\right] \\
=& (1-\rho^*)^2\frac{\sigma^2_{\varepsilon^*}(1+\rho^*)}{\sigma^2_{\varepsilon^*}(1+\rho^*) + \sigma^2_{\tau^*}T(1-\rho^*)} \\
\lim_{T\to\infty} g^*_T(1) =& 0 \text{ for } \rho^* \neq 1.
\end{aligned}$$

C. Additional simulation results

C.1. Bias, variance and RMSE for the corrected estimations in the Scenarios (I)–(III) for $\rho^{X^{*\mathbf{B}}} = 0.5$

	SF of	Truth		Berkson		Classical		Mixture	
Scen.	$\sigma^2_{U\mathrm{B}}, \sigma^2_{U\mathrm{C}}$	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
(I)	0	-0.006	0.186	**-0.006**	**0.186**	-0.018	0.191	**-0.006**	**0.186**
	0.5	-0.002	0.168	-0.020	0.193	**-0.008**	0.197	-0.009	**0.178**
	1	0.003	0.153	-0.020	0.193	0.008	0.193	**0.001**	**0.174**
	2	0.006	0.132	-0.020	0.194	0.020	0.158	**0.007**	**0.147**
	5	0.008	0.099	-0.020	0.195	0.004	0.124	**0.001**	**0.118**
(II)	0	0.019	0.161	**0.021**	**0.162**	0.021	0.176	0.023	0.166
	0.5	-0.006	0.174	0.004	0.194	**0.000**	0.202	0.005	**0.193**
	1	-0.007	0.154	**0.005**	0.194	-0.028	0.199	-0.016	**0.183**
	2	-0.008	0.128	**0.005**	0.194	-0.010	0.172	-0.015	**0.164**
	5	-0.007	0.091	**0.006**	0.195	-0.008	0.120	-0.008	**0.120**
(III)	0	0.001	0.131	0.001	**0.131**	**-0.001**	0.153	0.002	0.138
	0.5	0.035	0.121	0.042	**0.136**	**0.021**	0.156	0.035	0.139
	1	0.030	0.110	0.042	**0.136**	0.034	0.141	**0.030**	0.142
	2	0.023	0.096	0.041	**0.135**	**0.022**	0.142	0.025	0.136
	5	0.014	0.073	0.041	0.135	**0.010**	**0.117**	0.025	0.118

TABLE C.1.: *Bias and RMSE of MOM–corrected effect estimations for Scenarios (I)-(III) with* $\rho^{X^{*B}} = 0.5$*; bold: lowest absolute bias or lowest RMSE of each row (disregrading the results based on the true data).*

C.2. Visualization of simulation results of Subsection 5.4.1

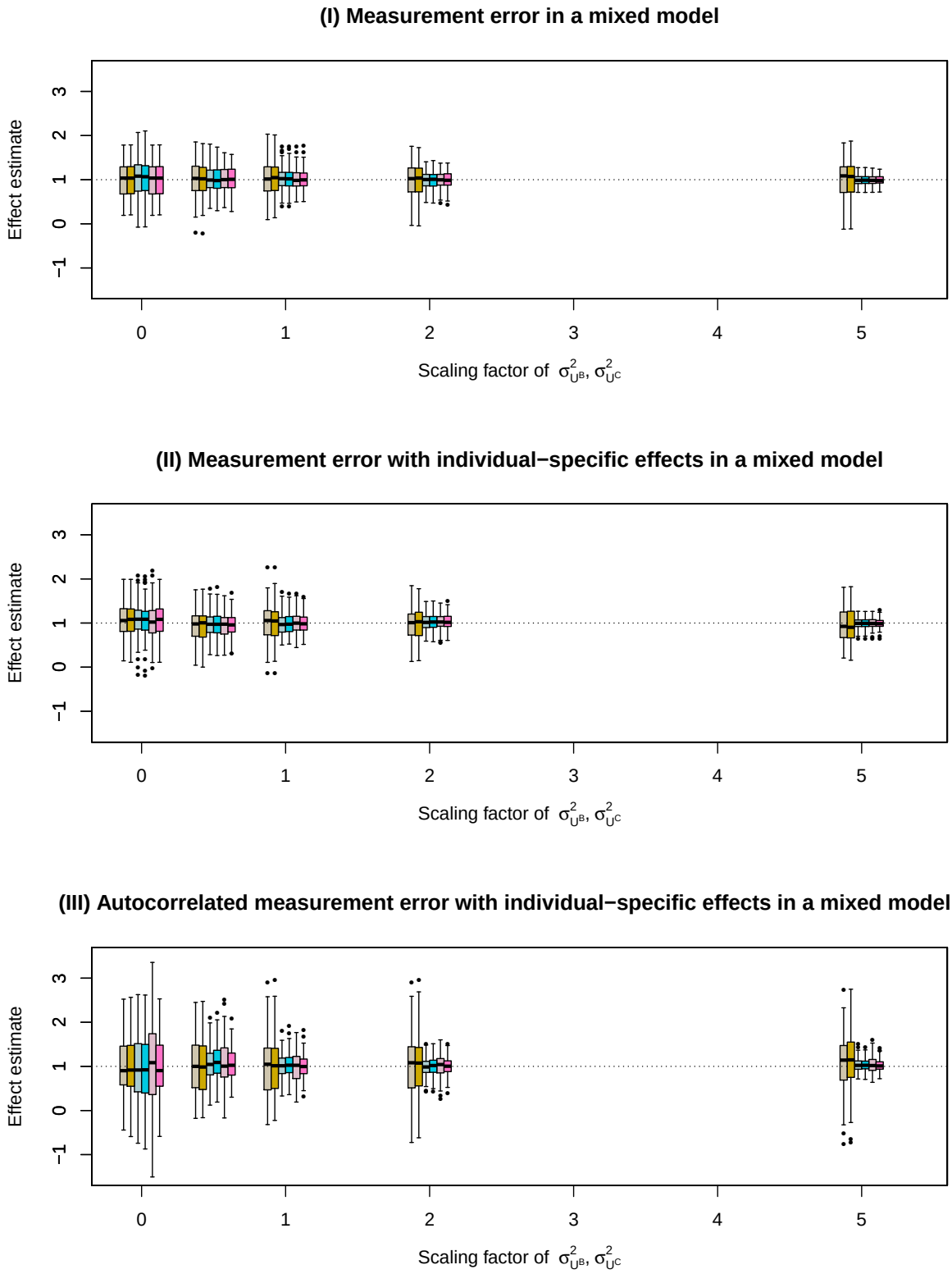


FIGURE C.1.: *Simulation results for measurement error corrected effect estimates for increasing size of measurement errors: Bayesian estimates based on measurements with Berkson error (beige), with classical measurement error (blue) and with mixture measurement error (pink); frequentist estimates are depicted with the corresponding lighter colors.*

C.3. Visualization of simulation results of Subsection 5.4.2

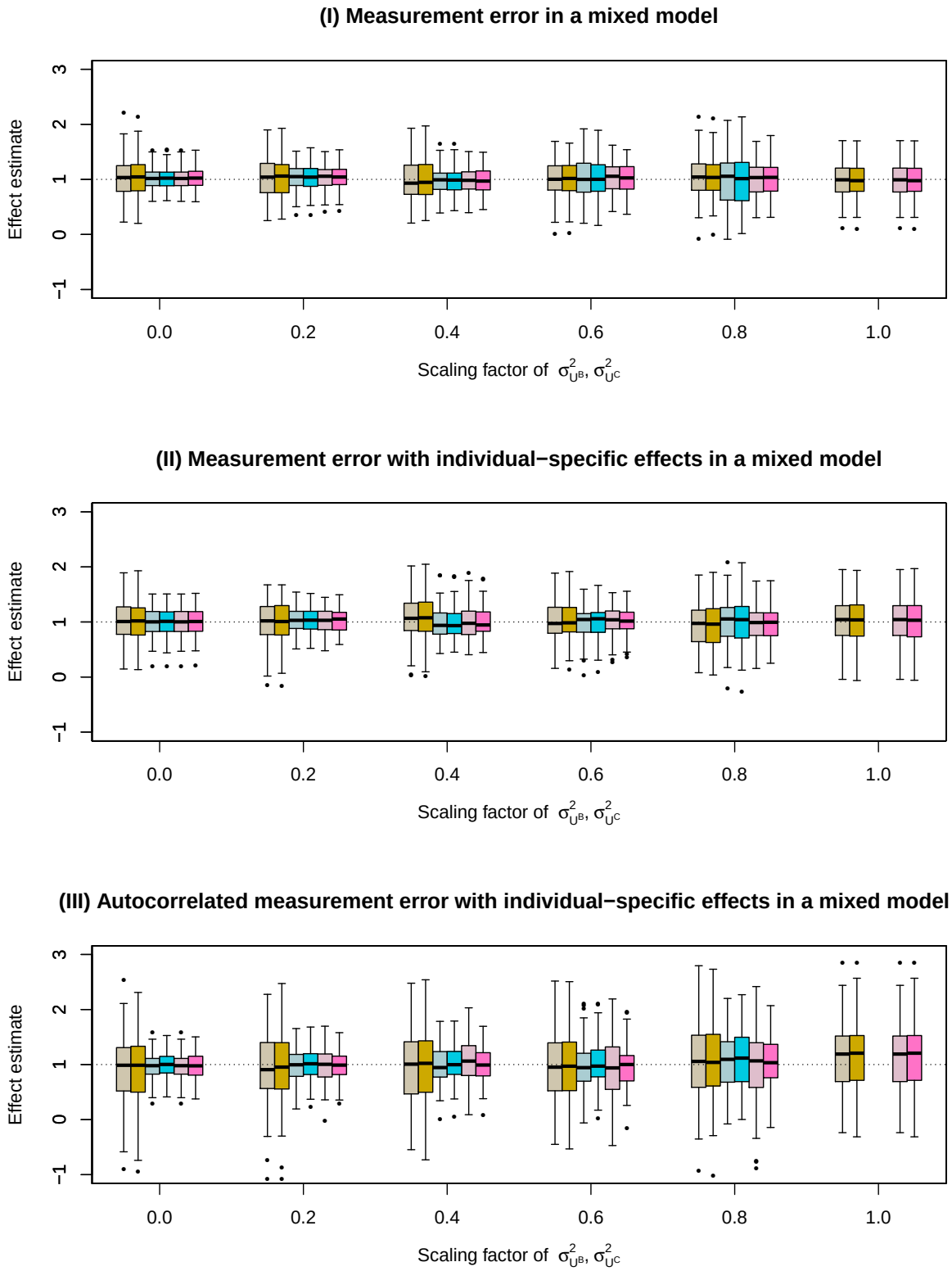


FIGURE C.2.: *Simulation results for measurement error corrected effect estimates for increasing percentage of missing personal measurements: Bayesian estimates based on measurements with Berkson error (beige), with classical measurement error (blue) and with mixture measurement error (pink); frequentist estimates are depicted with the corresponding lighter colors.*

D. Bayesian models for measurement error using conditionally conjugate prior distributions

In this section, full conditional distributions for model 3.1 and Berkson and classical measurement error (eq. 2.4 and eq. 2.5) are given.

D.1. Choice of priors

The following independent, conditionally conjugate – if possible – prior distributions are used for the unknown parameters:

- Effect coefficients: $\beta_0 \sim \mathrm{N}(\beta_0^{(0)}, \sigma^2_{\beta_0}), \beta_X \sim \mathrm{N}(\beta_X^{(0)}, \sigma^2_{\beta_X})$
- Variances: $\sigma^2_\tau \sim \mathrm{IG}(\delta_{\tau,1}, \delta_{\tau,2})$, $\sigma^2_\varepsilon \sim \mathrm{IG}(\delta_{\varepsilon,1}, \delta_{\varepsilon,2})$,
  $\sigma^2_{\nu^\mathrm{B}} \sim \mathrm{IG}(\delta_{\nu^\mathrm{B},1}, \delta_{\nu^\mathrm{B},2})$, $\sigma^2_{U^\mathrm{B}} \sim \mathrm{IG}(\delta_{U^\mathrm{B},1}, \delta_{U^\mathrm{B},2})$, $\sigma^2_{\nu^\mathrm{C}} \sim \mathrm{IG}(\delta_{\nu^\mathrm{C},1}, \delta_{\nu^\mathrm{C},2})$, $\sigma^2_{U^\mathrm{C}} \sim \mathrm{IG}(\delta_{U^\mathrm{C},1}, \delta_{U^\mathrm{C},2})$
- Correlation coefficients: $\rho \sim \mathrm{Unif}(\gamma_{\rho,1}, \gamma_{\rho,2}), \rho^\mathrm{B} \sim \mathrm{Unif}(\gamma_{\rho^\mathrm{B},1}, \gamma_{\rho^\mathrm{B},2})$, $\rho^\mathrm{C} \sim \mathrm{Unif}(\gamma_{\rho^\mathrm{C},1}, \gamma_{\rho^\mathrm{C},2})$

D.2. Likelihood

The likelihood for the data is:

$$
\begin{aligned}
f(\boldsymbol{Y}, \boldsymbol{X}^{*\mathrm{C}}, \boldsymbol{X}^{*\mathrm{B}} | \boldsymbol{X}, \boldsymbol{\theta}) \propto & f(\boldsymbol{Y}|\boldsymbol{X}, \boldsymbol{\theta}^H) f(\boldsymbol{X}^{*\mathrm{C}}|\boldsymbol{X}, \boldsymbol{\theta}^M) f(\boldsymbol{X}|\boldsymbol{X}^{*\mathrm{B}}, \boldsymbol{\theta}^{E_B}) \\
= & (2\pi)^{-\frac{3nT}{2}} |\boldsymbol{\Sigma}_\varepsilon|^{-\frac{n}{2}} |\boldsymbol{\Sigma}_{U^\mathrm{C}}|^{-\frac{n}{2}} |\boldsymbol{\Sigma}_{U^\mathrm{B}}|^{-\frac{n}{2}} \left(\sigma^2_\tau\right)^{-\frac{n}{2}} \left(\sigma^2_{\nu^\mathrm{B}}\right)^{-\frac{n}{2}} \left(\sigma^2_{\nu^\mathrm{C}}\right)^{-\frac{n}{2}} \\
& \times \prod_{i=1}^{n} \exp\Big\{ -\frac{1}{2} \Big[(\boldsymbol{Y}_i - \boldsymbol{\chi}_i \boldsymbol{\beta} - \tau_i \mathbf{1}_T)^\top \boldsymbol{\Sigma}_\varepsilon^{-1} (\boldsymbol{Y}_i - \boldsymbol{\chi}_i \boldsymbol{\beta} - \tau_i \mathbf{1}_T) + \\
& \qquad (\boldsymbol{X}_i^{*\mathrm{C}} - \boldsymbol{X}_i - \nu_i^\mathrm{C} \mathbf{1}_T)^\top \boldsymbol{\Sigma}_{U^\mathrm{C}}^{-1} (\boldsymbol{X}_i^{*\mathrm{C}} - \boldsymbol{X}_i - \nu_i^\mathrm{C} \mathbf{1}_T) + \\
& \qquad (\boldsymbol{X}_i - \boldsymbol{X}_i^{*\mathrm{B}} - \nu_i^\mathrm{B} \mathbf{1}_T)^\top \boldsymbol{\Sigma}_{U^\mathrm{B}}^{-1} (\boldsymbol{X}_i - \boldsymbol{X}_i^{*\mathrm{B}} - \nu_i^\mathrm{B} \mathbf{1}_T) + \\
& \qquad \frac{\tau_i^2}{\sigma^2_\tau} + \frac{\nu_i^{\mathrm{C}2}}{\sigma^2_{\nu^\mathrm{C}}} + \frac{\nu_i^{\mathrm{B}2}}{\sigma^2_{\nu^\mathrm{B}}} \Big] \Big\}
\end{aligned}
$$

with $\boldsymbol{\chi}_i = (\mathbf{1}_T, \boldsymbol{X}_i), \boldsymbol{\beta} = (\beta_0, \beta_X)^\top, \mathbf{1}_T = (1, \ldots, 1)$, a vector of length T.

D.3. Full conditional distributions

In the following, $\boldsymbol{\beta}^{(0)} = (\beta_0^{(0)}, \beta_X^{(0)})^\top$ and $\boldsymbol{\kappa}_{\boldsymbol{\beta}} = (\frac{1}{\sigma^2_{\beta_0}}, \frac{1}{\sigma^2_{\beta_X}})^\top$. The full conditional distributions are

$$f(\boldsymbol{\beta}|\cdot) \propto \exp\left\{-\frac{1}{2\sigma_\varepsilon^2}\left[\boldsymbol{\beta}^\top\left(\sum_{i=1}^n \boldsymbol{\chi}_i^\top \boldsymbol{W}_\rho^{-1}\boldsymbol{\chi}_i + \begin{pmatrix}\sigma_\varepsilon^2/\sigma_{\beta_0}^2 & 0\\ 0 & \sigma_\varepsilon^2/\sigma_{\beta_X}^2\end{pmatrix}\right)\boldsymbol{\beta} - \right.\right.$$
$$\left.\left. 2\sum_{i=1}^n (\boldsymbol{Y}_i - \tau_i\mathbf{1})^\top \boldsymbol{W}_\rho^{-1}\boldsymbol{\chi}_i\boldsymbol{\beta} - 2\boldsymbol{\beta}^{(0)\top}\begin{pmatrix}\sigma_\varepsilon^2/\sigma_{\beta_0}^2 & 0\\ 0 & \sigma_\varepsilon^2/\sigma_{\beta_X}^2\end{pmatrix}\boldsymbol{\beta}\right]\right\}$$
$$\Rightarrow \boldsymbol{\beta}|\cdot \sim \mathrm{N}\Bigg(\left[\sum_{i=1}^n \boldsymbol{\chi}_i^\top \boldsymbol{W}_\rho^{-1}\boldsymbol{\chi}_i + \begin{pmatrix}\sigma_\varepsilon^2/\sigma_{\beta_0}^2 & 0\\ 0 & \sigma_\varepsilon^2/\sigma_{\beta_X}^2\end{pmatrix}\right]^{-1}$$
$$\left[\sum_{i=1}^n (\boldsymbol{Y}_i - \tau_i\mathbf{1})^\top \boldsymbol{W}_\rho^{-1}\boldsymbol{\chi}_i + \boldsymbol{\beta}^{(0)\top}\begin{pmatrix}\sigma_\varepsilon^2/\sigma_{\beta_0}^2 & 0\\ 0 & \sigma_\varepsilon^2/\sigma_{\beta_X}^2\end{pmatrix}\right],$$
$$\sigma_\varepsilon^2\left[\sum_{i=1}^n \boldsymbol{\chi}_i^\top \boldsymbol{W}_\rho^{-1}\boldsymbol{\chi}_i + \begin{pmatrix}\sigma_\varepsilon^2/\sigma_{\beta_0}^2 & 0\\ 0 & \sigma_\varepsilon^2/\sigma_{\beta_X}^2\end{pmatrix}\right]^{-1}\Bigg)$$

$$f(\boldsymbol{X}_i|\cdot) \propto \exp\Big\{-\frac{1}{2}\Big[\beta_X^2\boldsymbol{X}_i^\top\boldsymbol{\Sigma}_\varepsilon^{-1}\boldsymbol{X}_i + \boldsymbol{X}_i^\top\boldsymbol{\Sigma}_{U^{\mathrm{C}}}^{-1}\boldsymbol{X}_i + \boldsymbol{X}_i^\top\boldsymbol{\Sigma}_{U^{\mathrm{B}}}^{-1}\boldsymbol{X}_i -$$
$$2\beta_X\boldsymbol{X}_i^\top\boldsymbol{\Sigma}_\varepsilon^{-1}(\boldsymbol{Y}_i - \beta_0\mathbf{1}_T - \tau_i\mathbf{1}_T) - 2\beta_X\boldsymbol{X}_i^\top\boldsymbol{\Sigma}_{U^{\mathrm{C}}}^{-1}(\boldsymbol{X}_i^{*\mathrm{C}} - \nu_i^{\mathrm{C}}\mathbf{1}_T) -$$
$$2\beta_X\boldsymbol{X}_i^\top\boldsymbol{\Sigma}_{U^{\mathrm{B}}}^{-1}(\boldsymbol{X}_i^{*\mathrm{B}} - \nu_i^{\mathrm{B}}\mathbf{1}_T)\Big]\Big\}$$
$$\Rightarrow \boldsymbol{X}_i|\cdot \sim \mathrm{N}\Bigg(\left[\beta_1^2\boldsymbol{\Sigma}_\varepsilon^{-1} + \boldsymbol{\Sigma}_{U^{\mathrm{C}}}^{-1} + \boldsymbol{\Sigma}_{U^{\mathrm{B}}}^{-1}\right]^{-1}$$
$$\left[\beta_X\boldsymbol{\Sigma}_\varepsilon^{-1}(\boldsymbol{Y}_i - \beta_0\mathbf{1}_T - \tau_i\mathbf{1}_T) + \boldsymbol{\Sigma}_{U^{\mathrm{C}}}^{-1}(\boldsymbol{X}_i^{*\mathrm{C}} - \nu_i^{\mathrm{C}}\mathbf{1}_T) + \boldsymbol{\Sigma}_{U^{\mathrm{B}}}^{-1}(\boldsymbol{X}_i^{*\mathrm{B}} + \nu_i^{\mathrm{B}}\mathbf{1}_T)\right],$$
$$\left[\beta_1^2\boldsymbol{\Sigma}_\varepsilon^{-1} + \boldsymbol{\Sigma}_{U^{\mathrm{C}}}^{-1} + \boldsymbol{\Sigma}_{U^{\mathrm{B}}}^{-1}\right]^{-1}\Bigg).$$

$$f(\tau_i|\cdot) \overset{eq.2.3}{\propto} \exp\left\{-\frac{1}{2\sigma_\varepsilon^2(1-\rho^2)}\left[\tau_i^2\left((T-2)(1-\rho)^2 - 2(1-\rho) + \frac{\sigma_\varepsilon^2(1-\rho^2)}{\sigma_\tau^2}\right) - \right.\right.$$
$$\left.\left. 2\tau_i\left((1-\rho)^2\sum_{t=2}^{T-1}(Y_{it} - \boldsymbol{\chi}_{it}\boldsymbol{\beta}) - (1-\rho)(\boldsymbol{Y}_{i1} + \boldsymbol{Y}_{iT} - \boldsymbol{\chi}_{i1}\boldsymbol{\beta} - \boldsymbol{\chi}_{iT}\boldsymbol{\beta})\right)\right]\right\}$$
$$\Rightarrow \tau_i|\cdot \sim \mathrm{N}\Bigg(\left[(T-2)(1-\rho)^2 - 2(1-\rho) + \frac{\sigma_\varepsilon^2(1-\rho^2)}{\sigma_\tau^2}\right]^{-1}$$
$$\left[(1-\rho)^2\sum_{t=2}^{T-1}(Y_{it} - \boldsymbol{\chi}_{it}\boldsymbol{\beta}) + (1-\rho)(\boldsymbol{Y}_{i1} + \boldsymbol{Y}_{iT} - \boldsymbol{\chi}_{i1}\boldsymbol{\beta} - \boldsymbol{\chi}_{iT}\boldsymbol{\beta})\right],$$
$$\sigma_\varepsilon^2(1-\rho^2)\left[(T-2)(1-\rho)^2 - 2(1-\rho) + \frac{\sigma_\varepsilon^2(1-\rho^2)}{\sigma_\tau^2}\right]^{-1}\Bigg)$$

Analogously follows

$$\nu_i^{\mathrm{C}}|\cdot \sim \mathrm{N}\Bigg(\left[(T-2)(1-\rho^{\mathrm{C}})^2 - 2(1-\rho^{\mathrm{C}}) + \frac{\sigma_{U^{\mathrm{C}}}^2(1-\rho^{\mathrm{C}^2})}{\sigma_{\nu^{\mathrm{C}}}^2}\right]^{-1}$$
$$\left[(1-\rho^{\mathrm{C}})^2\sum_{t=2}^{T-1}(X_{it}^{*\mathrm{C}} - X_{it}) + (1-\rho^{\mathrm{C}})(X_{i1}^{*\mathrm{C}} + X_{iT}^{*\mathrm{C}} - X_{i1} - X_{iT})\right]$$
$$\sigma_{U^{\mathrm{C}}}^2(1-\rho^{\mathrm{C}^2})\left[(T-2)(1-\rho^{\mathrm{C}})^2 - 2(1-\rho^{\mathrm{C}}) + \frac{\sigma_{U^{\mathrm{C}}}^2(1-\rho^{\mathrm{C}^2})}{\sigma_{\nu^{\mathrm{C}}}^2}\right]^{-1}\Bigg)$$
$$\nu_i^{\mathrm{B}}|\cdot \sim \mathrm{N}\Bigg(\left[(T-2)(1-\rho^{\mathrm{B}})^2 - 2(1-\rho^{\mathrm{B}}) + \frac{\sigma_{U^{\mathrm{B}}}^2(1-\rho^{\mathrm{B}})}{\sigma_{\nu^{\mathrm{B}}}^2}\right]^{-1}$$
$$\left[(1-\rho^{\mathrm{B}})^2\sum_{t=2}^{T-1}(X_{it} - X_{it}^{*\mathrm{B}}) + (1-\rho^{\mathrm{B}})(X_{i1} + X_{iT} - X_{i1}^{*\mathrm{B}} - X_{iT}^{*\mathrm{B}})\right]$$
$$\sigma_{U^{\mathrm{B}}}^2(1-\rho^{\mathrm{B}^2})\left[(T-2)(1-\rho^{\mathrm{B}})^2 - 2(1-\rho^{\mathrm{B}}) + \frac{\sigma_{U^{\mathrm{B}}}^2(1-\rho^{\mathrm{B}})}{\sigma_{\nu^{\mathrm{B}}}^2}\right]^{-1}\Bigg).$$

The full conditional distributions of the variance parameters are Inverse Gamma distributions:

$$f(\sigma_\varepsilon^2|\cdot) \propto (\sigma_\varepsilon^2)^{-(\frac{nT}{2}+\delta_{\varepsilon,1}+1)} \exp\left\{-\frac{1}{\sigma_\varepsilon^2}\left[\delta_{\varepsilon,2} + \frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{Y}_i - \boldsymbol{\chi}_i\boldsymbol{\beta} - \tau_i\mathbf{1}_T)^\top \boldsymbol{W}_\rho^{-1}(\boldsymbol{Y}_i - \boldsymbol{\chi}_i\boldsymbol{\beta} - \tau_i\mathbf{1}_T)\right]\right\}$$
$$\sigma_\varepsilon^2|\cdot \sim \mathrm{IG}\left(\delta_{\varepsilon,1} + \frac{nT}{2}, \delta_{\varepsilon,2} + \frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{Y}_i - \boldsymbol{\chi}_i\boldsymbol{\beta} - \tau_i\mathbf{1}_T)^\top \boldsymbol{W}_\rho^{-1}(\boldsymbol{Y}_i - \boldsymbol{\chi}_i\boldsymbol{\beta} - \tau_i\mathbf{1}_T)\right)$$
$$\sigma_{U^{\mathrm{C}}}^2|\cdot \sim \mathrm{IG}\left(\delta_{U^{\mathrm{C}},1} + \frac{nT}{2}, \delta_{U^{\mathrm{C}},2} + \frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{X}_i^{*\mathrm{C}} - \boldsymbol{X}_i - \nu_i^{\mathrm{C}}\mathbf{1}_T)^\top \boldsymbol{W}_{\rho^{\mathrm{C}}}^{-1}(\boldsymbol{X}_i^{*\mathrm{C}} - \boldsymbol{X}_i - \nu_i^{\mathrm{C}}\mathbf{1}_T)\right)$$
$$\sigma_{U^{\mathrm{B}}}^2|\cdot \sim \mathrm{IG}\left(\delta_{U^{\mathrm{B}},1} + \frac{nT}{2}, \delta_{U^{\mathrm{B}},2} + \frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{X}_i - \boldsymbol{X}_i^{*\mathrm{B}} - \nu_i^{\mathrm{B}}\mathbf{1}_T)^\top \boldsymbol{W}_{\rho^{\mathrm{B}}}^{-1}(\boldsymbol{X}_i - \boldsymbol{X}_i^{*\mathrm{B}} - \nu_i^{\mathrm{B}}\mathbf{1}_T)\right)$$

$$f(\sigma_\tau^2|\cdot) \propto \sigma_\tau^{2-(\frac{n}{2}+\delta_{\tau,1}+1)} \exp\left\{-\frac{1}{\sigma_\tau^2}\left[\delta_{\tau,2} + \frac{1}{2}\sum_{i=1}^{n}\tau_i^2\right]\right\}$$
$$\sigma_\tau^2|\cdot \sim \mathrm{IG}\left(\delta_{\tau,1} + \frac{n}{2}, \delta_{\tau,2} + \frac{1}{2}\sum_{i=1}^{n}\tau_i^2\right)$$
$$\sigma_{\nu^{\mathrm{B}}}^2|\cdot \sim \mathrm{IG}\left(\delta_{\nu^{\mathrm{B}},1} + \frac{n}{2}, \delta_{\nu^{\mathrm{B}},2} + \frac{1}{2}\sum_{i=1}^{n}\nu_i^{\mathrm{B}^2}\right)$$
$$\sigma_{\nu^{\mathrm{C}}}^2|\cdot \sim \mathrm{IG}\left(\delta_{\nu^{\mathrm{C}},1} + \frac{n}{2}, \delta_{\nu^{\mathrm{C}},2} + \frac{1}{2}\sum_{i=1}^{n}\nu_i^{\mathrm{C}^2}\right).$$

$$
\begin{aligned}
f(\rho|\cdot) \propto |\boldsymbol{W}_\rho|^{-\frac{n}{2}} \exp\Bigg\{ & -\frac{1}{2\sigma_\varepsilon^2} \sum_{i=1}^{n} \Big[(\boldsymbol{Y}_i - \boldsymbol{\chi}_i \boldsymbol{\beta} - \tau_i \mathbf{1}_T)^\top \boldsymbol{W}_\rho^{-1} \\
& (\boldsymbol{Y}_i - \boldsymbol{\chi}_i \boldsymbol{\beta} - \tau_i \mathbf{1}_T) \Big] \Bigg\} \boldsymbol{I}_{[\gamma_{\rho,1}, \gamma_{\rho,2}]}(\rho) \\
f(\rho^{\mathrm{C}}|\cdot) \propto |\boldsymbol{W}_\rho^{\mathrm{C}}|^{-\frac{n}{2}} \exp\Bigg\{ & -\frac{1}{2\sigma_{U^{\mathrm{C}}}^2} \sum_{i=1}^{n} \Big[(\boldsymbol{X}_i^{*\mathrm{C}} - \boldsymbol{X}_i - \nu_i^{\mathrm{C}} \mathbf{1}_T)^\top \boldsymbol{W}_{\rho^{\mathrm{C}}}^{-1} \\
& (\boldsymbol{X}_i^{*\mathrm{C}} - \boldsymbol{X}_i - \nu_i^{\mathrm{C}} \mathbf{1}_T) \Big] \Bigg\} \boldsymbol{I}_{[\gamma_{\rho^{\mathrm{C}},1}, \gamma_{\rho^{\mathrm{C}},2}]}(\rho^{\mathrm{C}}) \\
f(\rho^{\mathrm{B}}|\cdot) \propto |\boldsymbol{W}_\rho^{\mathrm{B}}|^{-\frac{n}{2}} \exp\Bigg\{ & -\frac{1}{2\sigma_{U^{\mathrm{B}}}^2} \sum_{i=1}^{n} \Big[(\boldsymbol{X}_i - \boldsymbol{X}_i^{*\mathrm{B}} - \nu_i^{\mathrm{B}} \mathbf{1}_T)^\top \boldsymbol{W}_{\rho^{\mathrm{B}}}^{-1} \\
& (\boldsymbol{X}_i - \boldsymbol{X}_i^{*\mathrm{B}} - \nu_i^{\mathrm{B}} \mathbf{1}_T) \Big] \Bigg\} \boldsymbol{I}_{[\gamma_{\rho^{\mathrm{B}},1}, \gamma_{\rho^{\mathrm{B}},2}]}(\rho^{\mathrm{B}})
\end{aligned}
$$

Bibliography

Alexeeff, S. E., R. J. Carroll, and B. Coull (2016). Spatial measurement error and correction by spatial simex in linear regression models when using predicted air pollution exposures. *Biostatistics 17*(2), 377–389.

Barnard, J., R. McCulloch, and X.-L. Meng (2000). Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Statistica Sinica 10*(4), 1281–1312.

Bartlett, M. S. (1946). On the theoretical specification and sampling properties of autocorrelated time-series. *Supplement to the Journal of the Royal Statistical Society*, 27–41.

Bartlett, M. S. (1951). An inverse matrix adjustment arising in discriminant analysis. *The Annals of Mathematical Statistics*, 107–111.

Bateson, T. F. and J. M. Wright (2010). Regression calibration for classical exposure measurement error in environmental epidemiology studies using multiple local surrogate exposures. *American journal of epidemiology*, kwq123.

Beckx, C., L. Int Panis, T. Arentze, D. Janssens, R. Torfs, S. Broekx, and G. Wets (2009). A dynamic activity-based population modelling approach to evaluate exposure to air pollution: Methods and application to a Dutch urban area. *Environmental Impact Assessment Review 29*(3), 179–185.

Bekö, G., C. J. Weschler, A. Wierzbicka, D. G. Karottki, J. Toftum, S. Loft, and G. Clausen (2013). Ultrafine particles: exposure and source apportionment in 56 Danish homes. *Environmental science & technology 47*(18), 10240–10248.

Berghmans, P., N. Bleux, L. I. Panis, V. Mishra, R. Torfs, and M. Van Poppel (2009). Exposure assessment of a cyclist to PM_{10} and ultrafine particles. *Science of the Total Environment 407*(4), 1286–1298.

Berhane, K., W. J. Gauderman, D. O. Stram, and D. C. Thomas (2004). Statistical Issues in Studies of the Long-Term Effects of Air Pollution: The Southern California Children's Health Study. *Statistical Science 19*(3), 414–449.

Bernardo, J. M. (1979). Reference posterior distributions for Bayesian inference. *Journal of the Royal Statistical Society. Series B (Methodological)*, 113–147.

Best, N. G., K. Ickstadt, and R. L. Wolpert (2000). Spatial Poisson Regression for Health and Exposure Data Measured at Disparate Resolutions. *Journal of the American Statistical Association 95*(452), 1076–1088.

Bi, J. (2012). A review of statistical methods for determination of relative importance of correlated predictors and identification of drivers of consumer liking. *Journal of Sensory Studies 27*(2), 87–101.

Blackwell, M., J. Honaker, and G. King (2015). A unified approach to measurement error and missing data overview and applications. *Sociological Methods & Research*, 0049124115585360.

Bliznyuk, N., C. J. Paciorek, J. Schwartz, and B. Coull (2014). Nonlinear predictive latent process models for integrating spatio-temporal exposure data from multiple sources. *The Annals of Applied Statistics 8*(3), 1538–1560.

Box, G. E. and G. C. Tiao (1973). *Bayesian Inference in Statistical Analysis*. Addison Wesley Publishing Co.

Brasche, S. and W. Bischof (2005). Daily time spent indoors in German homes–baseline data for the assessment of indoor exposure of German occupants. *International journal of hygiene and environmental health 208*(4), 247–253.

Brauer, M., R. Hirtle, A. Hall, and T. Yip (1999). Monitoring personal fine particle exposure with a particle counter. *Journal of Exposure Analysis and Environmental Epidemiology 9*(3), 228–236.

Breusch, T. S. and A. R. Pagan (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica: Journal of the Econometric Society*, 1287–1294.

Bühlmann, P. and T. Hothorn (2007). Boosting algorithms: Regularization, prediction and model fitting. *Statistical Science 22*(4), 477–505.

Buonanno, G., S. Marini, L. Morawska, and F. Fuoco (2012). Individual dose and exposure of italian children to ultrafine particles. *Science of the Total Environment 438*, 271–277.

Burns, C. J., J. M. Wright, J. B. Pierson, T. F. Bateson, I. Burstyn, D. A. Goldstein, J. E. Klaunig, T. J. Luben, G. Mihlan, L. Ritter, A. R. Schnatter, J. M. Symons, and K. D. Yi (2014). Evaluating Uncertainty to Strengthen Epidemiologic Data for Use in Human Health Risk Assessments. *Environmental Health Perspectives*.

Buzas, J. S., L. A. Stefanski, and T. D. Tosteson (2005). Measurement error. In W. Ahrens and I. Pigeot (Eds.), *Handbook of Epidemiology*, pp. 729–765. Springer Berlin Heidelberg.

Calder, C. A., C. H. Holloman, S. M. Bortnick, W. Strauss, and M. Morara (2008). Relating ambient particulate matter concentration levels to mortality using an exposure simulator. *Journal of the American Statistical Association 103*(481), 137–148.

Carroll, R., D. Ruppert, L. Stefanski, and C. Crainiceanu (2006). *Measurement Error in Nonlinear Models - A Modern Perspective* (2. ed.). Chapman & Hall/CRC, Boca Raton.

Carroll, R. J., J. D. Maca, and D. Ruppert (1999). Nonparametric regression in the presence of measurement error. *Biometrika 86*(3), 541–554.

Carroll, R. J., K. Roeder, and L. Wasserman (1999). Flexible parametric measurement error models. *Biometrics 55*(1), 44–54.

Carroll, R. J. and L. A. Stefanski (1990). Approximate quasi-likelihood estimation in models with surrogate predictors. *Journal of the American Statistical Association 85*(411), 652–663.

Cattaneo, A., G. Garramone, M. Taronna, C. Peruzzo, and D. Cavallo (2009). Personal exposure to airborne ultrafine particles in the urban area of Milan. In *Journal of Physics: Conference Series*, Volume 151, pp. 012039. IOP Publishing.

Chang, H. H., R. D. Peng, and F. Dominici (2011). Estimating the acute health effects of coarse particulate matter accounting for exposure measurement error. *Biostatistics 12*(4), 637–652.

Chesher, A. (1991). The effect of measurement error. *Biometrika 78*(3), 451–462.

Cohen Hubal, E., L. Sheldon, J. Burke, T. McCurdy, M. Berry, Z. V. Rigas, M.L., and N. Freeman (2000). Children's exposure assessment: a review of factors influencing children's exposure, and the data available to characterize and assess that exposure. *Environmental Health Perspectives 108*, 475–486.

Cook, J. R. and L. A. Stefanski (1994). Simulation-extrapolation estimation in parametric measurement error models. *Journal of the American Statistical association 89*(428), 1314–1328.

Crainiceanu, C., D. Ruppert, and M. P. Wand (2005). Bayesian analysis for penalized spline regression using WinBUGS. *Journal of Statistical Software 14*(14).

Cyrys, J., M. Pitz, H. Heinrich, J. Wichmann, and A. Peters (2008). Spatial and temporal variation of particle number concentration in Augsburg, Germany. *Science of the Total Environment 401*, 168–175.

Daniels, M. J. (1999). A prior for the variance in hierarchical models. *The Canadian Journal of Statistics/La Revue Canadienne de Statistique*, 567–578.

Daniels, M. J. and R. E. Kass (1999). Nonconjugate Bayesian Estimation of Covariance Matrices and Its Use in Hierarchical Models. *Journal of the American Statistical Association 94*(448), 1254.

de Vocht, F., H. Kromhout, G. Ferro, P. Boffetta, and I. Burstyn (2009). Bayesian modelling of lung cancer risk and bitumen fume exposure adjusted for unmeasured confounding by smoking. *Occupational and Environmental Medicine 66*(8), 502–508.

Deffner, V., H. Küchenhoff, V. Maier, M. Pitz, J. Cyrys, S. Breitner, A. Schneider, J. Gu, U. Geruschkat, and A. Peters (2016). Personal exposure to ultrafine particles: Two-level statistical modeling of background exposure and time-activity patterns during three seasons. *Journal of Exposure Science and Environmental Epidemiology 26*(1), 17–25.

Dellaportas, P. and D. A. Stephens (1995). Bayesian Analysis of Errors-in-Variables Regression Models. *Biometrics 51*(3), 1085.

Dominici, F., S. L. Zeger, and J. M. Samet (2000). A measurement error model for time-series studies of air pollution and mortality. *Biostatistics 1*(2), 157–175.

Ebelt, S. T., W. E. Wilson, and M. Brauer (2005). Exposure to ambient and nonambient components of particulate matter: a comparison of health effects. *Epidemiology 16*(3), 396–405.

Espino-Hernandez, G., P. Gustafson, and I. Burstyn (2011). Bayesian adjustment for measurement error in continuous exposures in an individually matched case-control study. *BMC Medical Research Methodology 11*(1), 67.

Fahrmeir, L., T. Kneib, and S. Lang (2007). *Regression*. Springer DE.

Fosdick, B. K. and A. E. Raftery (2012). Estimating the Correlation in Bivariate Normal Data With Known Variances and Small Sample Sizes. *The American Statistician 66*(1), 34–41.

Franck, U., O. Herbarth, B. Wehner, A. Wiedensohler, and M. Manjarrez (2003). How do the indoor size distributions of airborne submicron and ultrafine particles in the absence of significant indoor sources depend on outdoor distributions? *Indoor Air 13*(2), 174–181.

Freedman, L. S., D. Midthune, R. J. Carroll, and V. Kipnis (2008). A comparison of regression calibration, moment reconstruction and imputation for adjusting for covariate measurement error in regression. *Statistics in medicine 27*(25), 5195–5216.

Fromme, H., D. Twardella, S. Dietrich, D. Heitmann, R. Schierl, B. Liebl, and H. Rüden (2007). Particulate matter in the indoor air of classrooms—exploratory results from munich and surrounding area. *Atmospheric Environment 41*(4), 854–866.

Fuller, W. A. (1987). *Measurement Error Models*. New York: John Wiley and Sons.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian analysis 1*(3), 515–534.

Geman, S. and D. Geman (1984). Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-6*(6), 721–741.

Gerharz, L. E., O. Klemm, A. V. Broich, and E. Pebesma (2013). Spatio-temporal modelling of individual exposure to air pollution and its uncertainty. *Atmospheric Environment 64*, 56–65.

Gerharz, L. E., A. Krüger, and O. Klemm (2009). Applying indoor and outdoor modeling techniques to estimate individual exposure to PM2.5 from personal GPS profiles and diaries: A pilot study. *Science of The Total Environment 407*(18), 5184–5193.

Gilliland, F., E. Avol, P. Kinney, M. Jerrett, T. Dvonch, F. Lurmann, T. Buckley, P. Breysse, G. Keeler, T. de Villiers, et al. (2005). Air pollution exposure assessment for epidemiologic studies of pregnant women and children: lessons learned from the centers for children's environmental health and disease prevention research. *Environmental health perspectives*, 1447–1454.

Gokhale, D. V. and S. J. Press (1982). Assessment of a Prior Distribution for the Correlation Coefficient in a Bivariate Normal Distribution. *Journal of the Royal Statistical Society. Series A (General) 145*(2), 237.

Goldman, G. T., J. A. Mulholland, A. G. Russell, M. J. Strickland, M. Klein, L. A. Waller, and P. E. Tolbert (2011). Impact of exposure measurement error in air pollution epidemiology: effect of error type in time-series studies. *Environ Health 10*(61), 10–1186.

Gómez-Moreno, F., M. Pujadas, J. Plaza, J. Rodríguez-Maroto, P. Martínez-Lozano, and B. Artíñano (2011). Influence of seasonal factors on the atmospheric particle number concentration and size distribution in Madrid. *Atmospheric Environment 45*, 3169–3180.

Greenland, S. (1980). The effect of misclassification in the presence of covariates. *American journal of epidemiology 112*(4), 564–569.

Greenland, S. (1998). Probability logic and probabilistic induction. *Epidemiology (Cambridge, Mass.) 9*(3), 322–332.

Grömping, U. (2007). Estimators of relative importance in linear regression based on variance decomposition. *The American Statistician 61*(2).

Gryparis, A., B. A. Coull, and J. Schwartz (2007). Controlling for confounding in the presence of measurement error in hierarchical models: a Bayesian approach. *Journal of Exposure Science and Environmental Epidemiology 17*, S20–S28.

Gryparis, A., B. A. Coull, J. Schwartz, and H. H. Suh (2007). Semiparametric latent variable regression models for spatiotemporal modelling of mobile source particles in the greater Boston area. *Journal of the Royal Statistical Society: Series C (Applied Statistics) 56*(2), 183–209.

Gryparis, A., C. J. Paciorek, A. Zeka, J. Schwartz, and B. A. Coull (2009). Measurement error caused by spatial misalignment in environmental epidemiology. *Biostatistics 10*(2), 258–274.

Gu, J., U. Kraus, A. Schneider, R. Hampel, M. Pitz, S. Breitner, K. Wolf, O. Hänninen, A. Peters, and J. Cyrys (2015). Personal day-time exposure to ultrafine particles in different microenvironments. *International journal of hygiene and environmental health 218*(2), 188–195.

Gulliver, J. and D. Briggs (2011). STEMS-Air: A simple GIS-based air pollution dispersion model for city-wide exposure assessment. *Science of The Total Environment 409*(12), 2419–2429.

Guolo, A. (2008). Robust techniques for measurement error correction: a review. *Statistical Methods in Medical Research 17*(6), 555–580.

Gustafson, P. (2005). On Model Expansion, Model Contraction, Identifiability and Prior Information: Two Illustrative Scenarios Involving Mismeasured Variables. *Statistical Science 20*(2), 111–140.

Gustafson, P. (2015). *Bayesian Inference for Partially Identified Models: Exploring the Limits of Limited Data*, Volume 140. CRC Press.

Hampel, R., S. Breitner, A. Schneider, W. Zareba, U. Kraus, J. Cyrys, U. Geruschkat, P. Belcredi, M. Müller, H.-E. Wichmann, et al. (2012). Acute air pollution effects on heart rate variability are modified by SNPs involved in cardiac rhythm in individuals with diabetes or impaired glucose tolerance. *Environmental research 112*, 177–185.

Hampel, R., S. Breitner, W. Zareba, U. Kraus, M. Pitz, U. Geruschkat, P. Belcredi, A. Peters, A. Schneider, C. H. R. in the Region of Augsburg (KORA) Study Group, et al. (2012). Immediate ozone effects on heart rate and repolarisation parameters in potentially susceptible individuals. *Occupational and environmental medicine 69*(6), 428–436.

Hatch, M. and D. Thomas (1993). Measurement Issues in Environmental Epidemiology. *Environmental Health Perspectives 101*, 49.

He, C., L. Morawska, J. Hitchins, and D. Gilbert (2004). Contribution from indoor sources to particle number and mass concentrations in residential houses. *Atmospheric environment 38*(21), 3405–3415.

Horick, N., E. Weller, D. K. Milton, D. R. Gold, R. Li, and D. Spiegelman (2006). Home endotoxin exposure and wheeze in infants: correction for bias due to exposure measurement error. *Environmental health perspectives*, 135–140.

Hudda, N., K. Cheung, K. Moore, and C. Sioutas (2010). Inter-community variability in total particle number concentrations in the eastern Los Angeles air basin. *Atmospheric Chemistry and Physics 10*(23), 11385–11399.

Hussein, T., A. Puustinen, P. Aalto, J. Mäkelä, K. Hämeri, and M. Kulmala (2004). Urban aerosol number size distributions. *Atmospheric Chemistry and Physics 4*, 301–411.

Jaynes, E. T. (1982). On the rationale of maximum-entropy methods. *Proceedings of the IEEE 70*(9), 939–952.

Jeffreys, H. (1961). *Theory of probability*. New York: Oxford University Press.

Jurek, A. M., G. Maldonado, S. Greenland, and T. R. Church (2006). Exposure-measurement error is frequently ignored when interpreting epidemiologic study results. *European Journal of Epidemiology 21*(12), 871–876.

Kass, R. E. and L. Wasserman (1996). The selection of prior distributions by formal rules. *Journal of the American Statistical Association 91*(435), 1343–1370.

Kaur, S. and M. Nieuwenhuijsen (2009). Determinants of personal exposure to $PM_{2.5}$, ultrafine particle counts, and CO in a transport microenvironment. *Environmental Science & Technology 43*(13), 4737–4743.

Kearney, J., L. Wallace, M. MacNeill, X. Xu, K. VanRyswyk, H. You, R. Kulka, and A. Wheeler (2011). Residential indoor and outdoor ultrafine particles in Windsor, Ontario. *Atmospheric environment 45*(40), 7583–7593.

Keatinge, W. R., S. R. Coleshaw, J. C. Easton, F. Cotter, M. B. Mattock, and R. Chelliah (1986). Increased platelet and red cell counts, blood viscosity, and plasma cholesterol levels during heat stress, and mortality from coronary and cerebral thrombosis. *The American journal of medicine 81*(5), 795–800.

Kim, H.-J. (2015). A class of ratio distributions of dependent folded normals and its applications. *Statistics*, 1–21.

Kraus, U., S. Breitner, R. Hampel, K. Wolf, J. Cyrys, U. Geruschkat, J. Gu, K. Radon, A. Peters, and A. Schneider (2015). Individual daytime noise exposure in different microenvironments. *Environmental research 140*, 479–487.

Krudysz, M., K. Moore, M. Geller, C. Sioutas, and J. Froines (2009). Intra-community spatial variability of particulate matter size distributions in Southern California/Los Angeles. *Atmospheric Chemistry and Physics 9*(3), 1061–1075.

Küchenhoff, H., S. M. Mwalili, and E. Lesaffre (2006). A general method for dealing with misclassification in regression: The misclassification SIMEX. *Biometrics 62*(1), 85–96.

Lambert, P. C., A. J. Sutton, P. R. Burton, K. R. Abrams, and D. R. Jones (2005). How vague is vague? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS. *Statistics in Medicine 24*(15), 2401–2428.

Lanki, T., A. Ahokas, S. Alm, N. A. Janssen, G. Hoek, J. J. De Hartog, B. Brunekreef, and J. Pekkanen (2007). Determinants of personal and indoor $PM_{2.5}$ and absorbance among elderly subjects with coronary heart disease. *Journal of Exposure Science and Environmental Epidemiology 17*(2), 124–133.

Lanzinger, S., R. Hampel, S. Breitner, R. Rückerl, U. Kraus, J. Cyrys, U. Geruschkat, A. Peters, and A. Schneider (2014). Short-term effects of air temperature on blood pressure and pulse pressure in potentially susceptible individuals. *International journal of hygiene and environmental health 217*(7), 775–784.

Laplace, P. (1820). *Essai Philosophique sur les Probabilités.* English translation: *Philosophical Essays on Probabilities (1951).* New York:Dover.

Lee, D. and G. Shaddick (2010). Spatial Modeling of Air Pollution in Studies of Its Short-Term Health Effects. *Biometrics 66*(4), 1238–1246.

Li, R., E. Weller, D. W. Dockery, L. M. Neas, and D. Spiegelman (2006). Association of indoor nitrogen dioxide with respiratory symptoms in children: application of measurement error correction techniques to utilize data from multiple surrogates. *Journal of Exposure Science and Environmental Epidemiology 16*(4), 342–350.

Li, Y., A. Guolo, F. O. Hoffman, and R. J. Carroll (2007). Shared uncertainty in measurement error problems, with application to Nevada Test Site fallout data. *Biometrics 63*(4), 1226–1236.

Li, Y. and X. Lin (2003). Functional inference in frailty measurement error models for clustered survival data using the SIMEX approach. *Journal of the American Statistical Association 98*(461), 191–203.

Liechty, J. C. (2004). Bayesian correlation estimation. *Biometrika 91*(1), 1–14.

Liu, J., P. Gustafson, N. Cherry, and I. Burstyn (2009). Bayesian analysis of a matched case–control study with expert prior information on both the misclassification of exposure and the exposure–disease association. *Statistics in Medicine 28*(27), 3411–3423.

Lonati, G. and M. Giugliano (2006). Size distribution of atmospheric particulate matter at traffic exposed sites in the urban area of Milan (Italy). *Atmospheric Environment 40*, S264–S274.

Lütkepohl, H. (1984). Linear transformations of vector ARMA processes. *Journal of Econometrics 26*(3), 283–293.

Lyles, R. H. and L. L. Kupper (1999). A note on confidence interval estimation in measurement error adjustment. *The American Statistician 53*(3), 247–253.

Mage, D., W. Wilson, V. Hasselblad, and L. Grant (1999). Assessment of human exposure to ambient particulate matter. *Journal of the Air & Waste Management Association 49*(11), 1280–1291.

Mallick, B., F. Hoffman, and R. Carroll (2002). Semiparametric regression modeling with mixtures of Berkson and classical error, with application to fallout from the Nevada test site. *Biometrics 58*(1), 13–20.

McCracken, J. P., J. Schwartz, N. Bruce, M. Mittleman, L. M. Ryan, and K. R. Smith (2009). Combining individual-and group-level exposure information: child carbon monoxide in the Guatemala woodstove randomized control trial. *Epidemiology 20*(1), 127–136.

McShane, L. M., D. N. Midthune, J. F. Dorgan, L. S. Freedman, and R. J. Carroll (2001). Covariate measurement error adjustment for matched case–control studies. *Biometrics 57*(1), 62–73.

Meng, Q. Y., B. J. Turpin, L. Korn, C. P. Weisel, M. Morandi, S. Colome, J. Zhang, T. Stock, D. Spektor, A. Winer, et al. (2005). Influence of ambient (outdoor) sources on residential indoor and personal $PM_{2.5}$ concentrations: analyses of RIOPA data. *Journal of Exposure Science and Environmental Epidemiology 15*(1), 17–28.

Meng, X.-L. and A. M. Zaslavsky (2002). Single observation unbiased priors. *Annals of Statistics*, 1345–1375.

Messer, K. and L. Natarajan (2008). Maximum likelihood, multiple imputation and regression calibration for measurement error adjustment. *Statistics in medicine 27*(30), 6332.

Moberg, A. and G. Brattström (2011). Prediction intervals for climate reconstructions with autocorrelated noise—an analysis of ordinary least squares and measurement error methods. *Palaeogeography, Palaeoclimatology, Palaeoecology 308*(3), 313–329.

Molitor, J., M. Jerrett, C.-C. Chang, N.-T. Molitor, J. Gauderman, K. Berhane, R. McConnell, F. Lurmann, J. Wu, A. Winer, et al. (2007). Assessing uncertainty in spatial exposure models for air pollution health effects assessment. *Environmental Health Perspectives*, 1147–1153.

Molitor, J., N.-T. Molitor, M. Jerrett, R. McConnell, J. Gauderman, K. Berhane, and D. Thomas (2006). Bayesian modeling of air pollution health effects with missing exposure data. *American journal of epidemiology 164*(1), 69–76.

Monn, C. (2001). Exposure assessment of air pollutants: a review on spatial heterogeneity and indoor/outdoor/personal exposure to suspended particulate matter, nitrogen dioxide and ozone. *Atmospheric environment 35*(1), 1–32.

Morawska, L., C. He, J. Hitchins, K. Mengersen, and D. Gilbert (2003). Characteristics of particle number andmass concentrations in residential houses in Brisbane, Australia. *Atmosperic Environment 37*, 4195–4203.

Moschandreas, D. J. (1981). Exposure to pollutants and daily time budgets of people. *Bulletin of the New York Academy of Medicine 57*(10), 845.

Mudelsee, M. (2001). Note on the bias in the estimation of the serial correlation coefficient of AR(1) processes. *Statistical Papers 42*, 517–527.

Muff, S., A. Riebler, L. Held, H. Rue, and P. Saner (2015). Bayesian analysis of measurement error models using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series C (Applied Statistics) 64*(2), 231–252.

Needham, L. L., H. Özkaynak, R. M. Whyatt, D. B. Barr, R. Y. Wang, L. Naeher, G. Akland, T. Bahadori, A. Bradman, R. Fortmann, et al. (2005). Exposure assessment in the National Children's Study: introduction. *Environmental health perspectives*, 1076–1082.

Nguyen, J. L., M. S. Link, H. Luttmann-Gibson, F. Laden, J. Schwartz, B. S. Wessler, M. A. Mittleman, D. R. Gold, and D. W. Dockery (2015). Drier air, lower temperatures, and triggering of paroxysmal atrial fibrillation. *Epidemiology 26*(3), 374–380.

Oglesby, L., N. Künzli, M. Röösli, C. Braun-Fahrländer, P. Mathys, W. Stern, M. Jantunen, and A. Kousa (2000). Validity of ambient levels of fine particles as surrogate for personal exposure to outdoor air pollution—results of the european EXPOLIS-EAS Study (Swiss Center Basel). *Journal of the Air & Waste Management Association 50*(7), 1251–1261.

Pan, W., D. Zeng, and X. Lin (2009). Estimation in semiparametric transition measurement error models for longitudinal data. *Biometrics 65*(3), 728–736.

Papoulis, A. (1991). *Probability, Random Variables, and Stochastic Processes*. McGraw Hill.

Pekkanen, J., K. L. Timonen, J. Ruuskanen, A. Reponen, and A. Mirme (1997). Effects of ultrafine and fine particles in urban air on peak expiratory flow among children with asthmatic symptoms. *Environmental research 74*(1), 24–33.

Peng, R. D. and M. L. Bell (2010). Spatial misalignment in time series studies of air pollution and health data. *Biostatistics 11*(4), 720–740.

Penttinen, P., K. L. Timonen, P. Tiittanen, A. Mirme, J. Ruuskanen, and J. Pekkanen (2001). Number concentration and size of particles in urban air: effects on spirometric lung function in adult asthmatic subjects. *Environmental health perspectives 109*(4), 319.

Peters, A., R. Hampel, J. Cyrys, S. Breitner, U. Geruschkat, U. Kraus, W. Zareba, and A. Schneider (2015). Elevated particle number concentrations induce immediate changes in heart rate variability: a panel study in individuals with impaired glucose metabolism or diabetes. *Particle and Fibre Toxicology 12*(1).

Pham-Gia, T., N. Turkkan, and E. Marchand (2006). Density of the ratio of two normal random variables and applications. *Communications in Statistics - Theory and Methods 35*(9), 1569–1591.

Pinheiro, J. and D. Bates (2000). Mixed-effects models in S and S-PLUS. *Statistics and Computing. Springer-Verlag, Berlin, D.*

Pinheiro, J., D. Bates, S. DebRoy, D. Sarkar, and R Core Team (2013). *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-109.

Pitz, M., W. Birmili, O. Schmid, A. Peters, H. Wichmann, and J. Cyrys (2008). Quality control and quality assurance for particle size distribution measurements at an urban monitoring station in Augsburg, Germany. *Journal of Environmental Monitoring 10*, 1017–1024.

Pitz, M., J. Cyrys, E. Karg, A. Wiedensohler, H.-E. Wichmann, and J. Heinrich (2003). Variability of apparent particle density of an urban aerosol. *Environmental Science & Technology 37*(19), 4336–4342.

Pitz, M., O. Schmid, J. Heinrich, W. Birmili, J. Maguhn, R. Zimmermann, H. Wichmann, A. Peters, and J. Cyrys (2008). Seasonal and diurnal variation of $pm_{2.5}$ apparent particle density in urban air in Augsburg, Germany. *Environmental Science & Technology 42*, 5087–5093.

Plummer, M. (2003). Jags: A program for analysis of Bayesian graphical models using Gibbs sampling. In *Proceedings of the 3rd international workshop on distributed statistical computing*, Volume 124, pp. 125. Technische Universit at Wien.

Prescott, G. J. and P. H. Garthwaite (2005). Bayesian analysis of misclassified binary data from a matched case-control study with a validation sub-study. *Statistics in Medicine 24*(3), 379–401.

Rice, K. (2003). Full-likelihood approaches to misclassification of a binary exposure in matched case-control studies. *Statistics in Medicine 22*(20), 3177–3194.

Richardson, S. (1996). *Markov chain monte carlo*, Chapter Measurement error, pp. 401–418. London, UK: Chapman & Hall.

Richardson, S. and W. R. Gilks (1993). Conditional independence models for epidemiological studies with covariate measurement error. *Statistics in Medicine 12*(18), 1703–1722.

Richardson, S., L. Leblond, I. Jaussent, and P. J. Green (2002). Mixture models in measurement error problems, with reference to epidemiological studies. *Journal of the Royal Statistical Society: Series A (Statistics in Society) 165*(3), 549–566.

Rosner, B., W. Willett, and D. Spiegelman (1989). Correction of logistic regression relative risk estimates and confidence intervals for systematic within-person measurement error. *Statistics in medicine 8*(9), 1051–1069.

Rückerl, R., R. Hampel, S. Breitner, J. Cyrys, U. Kraus, J. Carter, L. Dailey, R. B. Devlin, D. Diaz-Sanchez, W. Koenig, et al. (2014). Associations between ambient air pollution and blood markers of inflammation and coagulation/fibrinolysis in susceptible populations. *Environment international 70*, 32–49.

Ruuskanen, J., T. Tuch, H. Ten Brink, A. Peters, A. Khlystov, A. Mirme, G. Kos, B. Brunekreef, H. Wichmann, G. Buzorius, et al. (2001). Concentrations of ultrafine, fine and $PM_{2.5}$ particles in three European cities. *Atmospheric Environment 35*(21), 3729–3738.

Sarnat, J. A., W. E. Wilson, M. Strand, J. Brook, R. Wyzga, and T. Lumley (2007). Panel discussion review: session 1—exposure assessment and related errors in air pollution epidemiologic studies. *Journal of Exposure Science and Environmental Epidemiology 17*, S75–S82.

Schafer, D. W. (1987). Covariate measurement error in generalized linear models. *Biometrika 74*(2), 385–391.

Schisterman, E. F., K. B. Moysich, L. J. England, and M. Rao (2003). Estimation of the correlation coefficient using the Bayesian Approach and its applications for epidemiologic research. *BMC medical research methodology 3*, 5.

Schwartz, J. and B. A. Coull (2003). Control for confounding in the presence of measurement error in hierarchical models. *Biostatistics 4*(4), 539–553.

Sherman, J. and W. J. Morrison (1950). Adjustment of an inverse matrix corresponding to a change in one element of a given matrix. *The Annals of Mathematical Statistics*, 124–127.

Smith, K. R., J. P. McCracken, L. Thompson, R. Edwards, K. N. Shields, E. Canuz, and N. Bruce (2010). Personal child and mother carbon monoxide exposures and kitchen levels: methods and results from a randomized trial of woodfired chimney cookstoves in Guatemala (RESPIRE). *Journal of Exposure Science and Environmental Epidemiology 20*(5), 406–416.

Spiegelhalter, D. J. (2001). Bayesian methods for cluster randomized trials with continuous responses. *Statistics in Medicine 20*(3), 435–452.

Spiegelman, D. (2010). Approaches to uncertainty in exposure assessment in environmental epidemiology. *Annual review of public health 31*, 149.

Spiegelman, D., B. Rosner, and R. Logan (2000). Estimation and inference for logistic regression with covariate misclassification and measurement error in main study/validation study designs. *Journal of the American Statistical Association 95*(449), 51–61.

Steenland, K. and S. Greenland (2004). Monte Carlo sensitivity analysis and Bayesian analysis of smoking as an unmeasured confounder in a study of silica and lung cancer. *American Journal of Epidemiology 160*(4), 384–392.

Stefanski, L. A. and R. J. Carroll (1985). Covariate measurement error in logistic regression. *The Annals of Statistics*, 1335–1351.

Stefanski, L. A. and R. J. Carroll (1987). Conditional scores and optimal scores for generalized linear measurement-error models. *Biometrika 74*(4), 703–716.

Strand, M., S. Vedal, C. Rodes, S. J. Dutton, E. W. Gelfand, and N. Rabinovitch (2006). Estimating effects of ambient $PM_{2.5}$ exposure on health using $PM_{2.5}$ component measurements and regression calibration. *Journal of Exposure Science and Environmental Epidemiology 16*(1), 30–38.

Sturm, P. J., U. Baltensperger, M. Bacher, B. Lechner, S. Hausberger, B. Heiden, D. Imhof, E. Weingartner, A. S. Prevot, R. Kurtenbach, et al. (2003). Roadside measurements of particulate matter size distribution. *Atmospheric Environment 37*(37), 5273–5281.

Suggs, J. C. and T. C. Curran (1983). An empirical Bayes method for comparing air pollution data to air quality standards. *Atmospheric Environment (1967) 17*(4), 837–841.

Szpiro, A. A., P. D. Sampson, L. Sheppard, T. Lumley, S. D. Adar, and J. D. Kaufman (2010). Predicting intra-urban variation in air pollution concentrations with complex spatio-temporal dependencies. *Environmetrics*, 606–631.

Thomas, D., D. Stram, and J. Dwyer (1993). Exposure Measurement Error: Influence on Exposure-Disease Relationships and Methods of Correction. *Annual Review of Public Health 14*(1), 69–93.

Thomas, D. C. (2000). Some Contributions of Statistics to Environmental Epidemiology. *Journal of the American Statistical Association 95*(449), 315–319.

Tosteson, T. D., J. P. Buonaccorsi, and E. Demidenko (1998). Covariate measurement error and the estimation of random effect parameters in a mixed model for longitudinal data. *Statistics in Medicine 17*(17), 1959–1971.

TSI Incorporated (2016a). Hand–held condensation particle counter model 3007. `http://www.tsi.com/uploadedFiles/_Site_Root/Products/Literature/Spec_Sheets/3007_1930032.pdf`. Accessed: 2016-04-13.

TSI Incorporated (2016b). Hand–held condensation particle counter model 3007. `http://www.tsi.com/uploadedFiles/_Site_Root/Products/Literature/Spec_Sheets/3025A.pdf`. Accessed: 2016-04-13.

Van Roosbroeck, S., R. Li, G. Hoek, E. Lebret, B. Brunekreef, and D. Spiegelman (2008). Traffic-related outdoor air pollution and respiratory symptoms in children: the impact of adjustment for exposure measurement error. *Epidemiology 19*(3), 409–416.

Walck, C. (2007). Handbook on statistical distributions for experimentalists.

Wallace, L. and W. Ott (2011). Personal exposure to ultrafine particles. *Journal of Exposure Science and Environmental Epidemiology 21*(1), 20–30.

Wang, C., N. Wang, and S. Wang (2000). Regression analysis when covariates are regression parameters of a random effects model for observed longitudinal measurements. *Biometrics 56*(2), 487–495.

Wang, C.-Y. and M. Sullivan Pepe (2000). Expected estimating equations to accommodate covariate measurement error. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 62*(3), 509–524.

Wang, N., R. Carroll, and K.-Y. Liang (1996). Quasilikelihood estimation in measurement error models with correlated replicates. *Biometrics*, 401–411.

Wang, N., X. Lin, R. Gutierrez, and R. Carroll (1998). Bias analysis and simex approach in generalized linear mixed measurement error models. *Journal of the American Statistical Association 93*(11), 249–261.

Wang, Y., S. Wang, and R. J. Carroll (2015). The direct integral method for confidence intervals for the ratio of two location parameters. *Biometrics 71*(3), 704–713.

Wansbeek, T. and A. Kapteyn (1985). Estimation in a linear model with serially correlated errors when observations are missing. *International Economic Review*, 469–490.

Wehner, B. and A. Wiedensohler (2003). Long term measurements of submicrometer urban aerosols: statistical analysis for correlations with meteorological conditions and trace gases. *Atmospheric Chemistry and Physics 3*(3), 867–879.

Wheeler, A. J., L. A. Wallace, J. Kearney, K. Van Ryswyk, H. You, R. Kulka, J. R. Brook, and X. Xu (2011). Personal, indoor, and outdoor concentrations of fine and ultrafine particles using continuous monitors in multiple residences. *Aerosol Science and Technology 45*(9), 1078–1089.

White, E., B. K. Armstrong, and R. Saracci (2008). *Principles of exposure measurement in epidemiology: collecting, evaluating and improving measures of disease risk factors*. OUP Oxford.

Whittemore, A. S. and J. B. Keller (1988). Approximations for regression with covariate measurement error. *Journal of the American Statistical Association 83*(404), 1057–1066.

Willett, W. (1989). An overview of issues related to the correction of non-differential exposure measurement error in epidemiologic studies. *Statistics in Medicine 8*(9), 1031–1040.

Wilson, W. E. and M. Brauer (2006). Estimation of ambient and non-ambient components of particulate matter exposure from a personal monitoring panel study. *Journal of Exposure Science and Environmental Epidemiology 16*(3), 264–274.

Wilson, W. E. and H. H. Suh (1997). Fine particles and coarse particles: concentration relationships relevant to epidemiologic studies. *Journal of the Air & Waste Management Association 47*(12), 1238–1249.

Yi, G. (2008). A simulation-based marginal method for longitudinal data with dropout and mismeasured covariates. *Biostatistics 9*(3), 501–512.

Zeger, S. L., D. Thomas, F. Dominici, J. M. Samet, J. Schwartz, D. Dockery, and A. Cohen (2000). Exposure measurement error in time-series studies of air pollution: concepts and consequences. *Environmental health perspectives 108*(5), 419.

Zhu, Y., W. C. Hinds, S. Kim, and C. Sioutas (2002). Concentration and size distribution of ultrafine particles near a major highway. *Journal of the air & waste management association 52*(9), 1032–1042.

Zięba, A. (2010). Effective number of observations and unbiased estimators of variance for autocorrelated data – an overview. *Metrology and Measurement Systems 17*, 3–16.

List of Figures

List of Tables

Eidesstattliche Versicherung

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

München, den 06.06.2016

Veronika Deffner