A Bayesian semiparametric latent variable model for binary, ordinal and continuous response

Alexander Wolf Raach

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

München, 9. November 2005

A Bayesian semiparametric latent variable model for binary, ordinal and continuous response

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

vorgelegt von
Alexander Wolf Raach
aus Stuttgart – Bad Cannstatt

München, 9. November 2005

Erstgutachter: Prof. Dr. Ludwig Fahrmeir
Zweitgutachter: Prof. Dr. Hans Schneeweiß
Drittgutachter: Prof. Dr. Gerhard Arminger

Rigorosum: 26. Januar 2006

Danksagung

Ich bin meinem Betreuer Prof. Dr. Fahrmeir zu tiefstem Dank verpflichtet. Er hat das Risiko auf sich genommen und mich als Doktoranden akzeptiert, obwohl ich als Externer und Fachfremder diese Arbeit durchführen wollte. Weiterhin hatte er immer ausführlich Zeit zur gemeinsamen Diskussion und thematischen Weiterentwicklung, trotz teilweise kurzfristiger Terminanfragen meinerseits.

Ebenfalls möchte ich Prof. Dr. Schneeweiß und Prof. Dr. Arminger für die Bereitschaft danken, als Gutachter für meine Arbeit tätig gewesen zu sein. Prof. Dr. Schneeweiß hat darüberhinaus durch sein ausführliches Korrekturlesen zur Qualitätssteigerung der Dissertation beigetragen.

Mein spezieller Dank gilt Alexander Jerak, der den Kontakt zum Lehrstuhl herstellte und mir in etlichen Gesprächen wertvolle "Statistikertips" übermittelte, die mir die Eingewöhnung in die statistische Materie stark erleichterten. Andreas Brezgers Hilfestellungen ermöglichten mir eine zügige Implementation der P-Splines-Funktionalität. Rainer Kiefer möchte ich für die Bereitstellung des Datensatzes danken, welcher die Basis für die vorliegende Arbeit bildete.

Meinen Eltern sei gedankt dafür, dass sie mich doch noch auf die richtige Schule geschickt und mir alle Möglichkeiten zur persönlichen Entwicklung geschaffen haben (hier betone ich bereits Gesagtes).

Abschließend danke ich insbesondere meiner Freundin, die mich in allen schlechten Phasen hat aushalten müssen, die mir durch Ihre Unterstützung (v. a. Korrekturlesen) sehr geholfen und mir durch viele andere Dinge die Doktorarbeitszeit immens versüßt hat.

Zusammenfassung

Diese Arbeit diskutiert ein Latentes-Variablen-Modell (LVM), welches in einem Bayesianischem Ansatz formuliert und mit Markov chain Monte Carlo Methoden (MCMC) geschätzt wird. Dieses Modell erweitert die klassische Faktoranalyse, indem es nicht nur normalverteilte metrische manifeste Variablen zulässt, sondern auch binäre und ordinale Indikatoren integriert, welche in vielen Anwendungsfeldern (z. B. Psychologie, Soziologie) verbreitet sind. Weiterhin wird ein semiparametrischer Prädiktor eingeführt, welcher den Einfluss von Kovariaten auf die latenten Variablen beschreibt. Der Prädiktor kann parametrische Effekte, glatte Funktionen von metrischen Kovariaten (modelliert durch Random Walks und P-Splines), räumliche Effekte (modelliert durch Markov Random Fields) und Interaktionen von metrischen und kategorialen Kovariaten beinhalten. Eine Integration von zeitlichen Effekten wäre leicht möglich. Somit kann der Einfluss von Kovariaten auf die latenten Variablen wesentlich detaillierter als mit bisherigen Methoden untersucht werden.

Ein Schwerpunkt der Arbeit ist die Entwicklung eines effizienten MCMC Algorithmus mit guten Schätzeigenschaften (insbesondere für die Schwellenwerte der ordinalen Indikatoren) und dessen Implementierung im Standardsoftwarepaket R. Ebenso steht die Demonstration der Anwendbarkeit des Modells an einer Internetumfrage im Mittelpunkt. Hierzu werden zahlreiche Modelle mit unterschiedlich strukturierten Prädiktoren analysiert und erste Ansätze zur Modellwahl vorgestellt.

Abstract

This thesis discusses a latent variable model (LVM) which is based on a Bayesian approach and is estimated by Markov chain Monte Carlo methods (MCMC). The model extends classic factor analysis by allowing not only for gaussian metric manifest variables, but also for binary and ordinal indicators which are very common in many areas of application (e. g. psychology, sociology). Furthermore, a semiparametric predictor is introduced which describes the influence of covariates on the latent variables. The predictor may contain parametric effects, smooth functions of metric covariates (modeled by random walks and P-splines), spatial effects (modeled by Markov random fields) and interactions of metric and categorical covariates. The integration of temporal effects is easily possible. Consequently, the influence of covariates on the latent variables can be analyzed in much more detail than with currently available methods.

One emphasis of this work is the development of an efficient MCMC algorithm with good estimation properties (in particular concerning the cutpoints of ordinal indicators) and its implementation in the standard software package R. Another focus lies on the demonstration of the model's applicability using data from an internet survey. Several models with differently structured predictors are analyzed and first ideas for model selection are presented.

Contents

Chapter 1

Introduction

In many scientific fields, the variable of interest cannot be observed. This is especially true for the fields of social sciences and psychology where many theoretical concepts (e. g. intelligence) cannot be measured directly; therefore those types of variables or constructs are termed latent. In order to gain a better understanding of the latent variable, the researcher collects a multidimensional set of response variables – called indicators – which are used to measure the latent variable indirectly. It is assumed that the latent variable itself influences the response given for the indicators. Latent variable models (LVM) represent a tool to analyze the multidimensional response variables in such a way to reveal the indicators' underlying relationships that are caused by the latent variables. The most famous example of a LVM is the factor analysis model which requires the indicators to be measured on a continuous scale.

The classic factor analysis model can be extended in two different ways. Firstly, the augmented model also allows for binary and ordinal indicators. Very often the data is not of a continuous type but of binary or ordinal structure, hence the model should be able to properly account for these types of indicators. Typically researchers have used ordinal data in classic factor analysis models, and thus assumed the ordinal data to be normally distributed which is a rigid assumption – in particular for indicators with a low number of categories – and leads to distorted parameter estimates. Secondly, the extended LVM includes covariates that influence the indicators or the latent variables. The models presented in the statistical literature so far assume that the influence of the covariates on both the indicators and the latent variables is strictly linear. In most research settings this is a very restrictive assumption, and hence the analysis of these models does not reveal the true functional relationship between the covariates and the latent variables.

The LVM presented in this thesis covers both mentioned extensions. It allows for the inclusion of continuous, binary and ordinal indicators in arbitrary numbers and combinations; for example, the estimation of a model containing all types of indicators – continuous, binary, and ordinal – is possible. Furthermore, the LVM includes two types of covariates: the first type of covariates influence the indicators, and the second set of covariates influence the latent variables. Both sets of covariates can be modeled to exhibit a standard linear influence. Additionally our LVM allows the second set of covariates modifying the latent variables to be modeled in a more flexible way: metric covariates can exhibit a non-linear influence on the latent variables that is modeled as a smooth function; spatial covariates can be included in the model which allows for the analysis of regional effects altering the value of the latent variables; finally an interaction of metric and categorical covariates can be incorporated in the model which allows for the estimation of smooth functions of the metric covariate for separate categories of the categorical covariate. Hence this model enables the applied researcher to conduct much more detailed analyses of the covariate effects on the latent variables compared to currently available models.

Traditionally all LVM have been estimated by maximum-likelihood methods, and nowadays this predominance of maximum-likelihood methods still prevails in the area of latent variable modeling. However, the use of Bayesian models and estimation methods became more popular in recent years. For our LVM, a fully Bayesian approach is employed where all unknown population parameters are considered to be random variables. This includes the specification of prior distributions for all parameters that have to be estimated. The posterior distribution of those parameters is obtained by using Markov chain Monte Carlo (MCMC) methods.

This thesis is structured in the following way. Chapter 2.1 starts with the explanation of what a latent variable is supposed to represent. We think that such a basic introduction might be useful for many statisticians who are not so familiar with the concept of latent variables and with latent variable modeling. After that we give an overview of the different types of LVM classified into three groups – the LVM without covariate effects, the LVM including covariate effects, and structural equation modeling. The model of this thesis belongs to the second group of LVM including covariate effects. This classification into three groups helps in understanding how the various and numerous latent variable models relate to each other. We conclude this chapter with a summary of the Bayesian developments that occurred in the fragmented landscape of latent variable modeling.

In Chapter 3 the statistical model of our LVM is presented in full detail. The model consists of two parts: the measurement model relates the latent variables and direct covariates to the indicators; the structural equation model links the indirect covariates to the latent variables. In our model only the covariates influencing the latent variables, i. e. the covariates of the structural equation model, can be modeled by smooth non-linear functions. The discussion of the model also involves the establishment of identification restrictions which fix some parameters to certain values and hence ensure the identifiability of all parameters of the model. The end of this chapter gives a short summary of the whole model including a table showing all employed variables and notational indices.

Chapter 4 gives an overview of Bayesian methodology. After having explained how a Bayesian model is setup, we continue the discussion with an introduction to Markov chain

Monte Carlo (MCMC) methods. This includes a short review on Monte Carlo methods, Markov chains, the Metropolis-Hastings algorithm, and the Gibbs sampler. This very basic overview is provided for practitioners of latent variable modeling who are not familiar with the concepts of Bayesian modeling and MCMC. The chapter finishes with a short account of sampler convergence, and the theoretical background of the Generalized Gibbs sampler which is able to improve the MCMC sampler convergence for our model.

Chapter 5 combines the statistical model of Chapter 3 and the Bayesian theory of Chapter 4, and hence describes how our LVM is formulated as a Bayesian model. This includes the full prior specification of all involved model parameters with a focus on the priors of the smooth functions related to the covariates modifying the latent variables. After that, likelihood and posterior distribution of the LVM are set out. Based on the posterior distribution, three different MCMC sampling algorithms are derived: the first one solely contains Gibbs steps, the second one replaces the Gibbs step for the cutpoints of the ordinal indicators of the first sampler with a Metropolis-Hastings step, and the third one equals the first sampler plus an additional Generalized Gibbs move at the end of each iteration, whose theoretical foundation is given in the end of Chapter 4. We employ three different MCMC samplers due to their different convergence properties, regarding those of the cutpoint parameters in particular.

Chapter 6 describes various simulation studies which are structured in three parts. The first part looks at the convergence properties of the three different samplers. The second part examines the quality of parameter estimates for models without the use of covariates influencing the latent variables, and the third part shows how smooth functions of covariates (metric, spatial) modifying the latent variables are estimated. Chapter 6 clearly demonstrates that it is possible to estimate smooth functions of covariates in a LVM.

Chapter 7 deals with the analysis of a real data set. Here we examine parts of the data set collected in the internet survey "Perspektive Deutschland 1"[1] in the context of social sciences. This data offers a variety of different latent variables to be analyzed; e. g. we take a closer look at the attitude of German citizens if provisions (e. g. health insurance) should be managed by each citizen itself or by the state. After having described the data in an explorative way, we conduct several analyses that show that the estimation of smooth functions of covariates (metric and spatial) is possible and useful for a real data set, and that mistakes would be made if only a linear effect of the covariates on the latent variables was considered. In order to identify the best fitting model, two versions of the deviance information criterion (DIC) are calculated. We conclude the discussion by looking at the actual values of the latent variables in the population, and the probability of a certain observation having a higher latent score than another observation.

Chapter 8 finally gives a short summary of the findings and an outlook including potential developments that could be pursued in order to extend the presented LVM in future.

[1]English translation: "Prospect Germany 1"

Chapter 2

Latent Variable Models (LVM)

2.1 Description of latent variables

In many applied research settings it is not possible to measure the variable (or variables) of interest directly. In order to conduct analyses of this variable, several measurable quantities are collected and used to extract information about the primary variable of interest. Therefore we distinguish two types of variables: the unobservable variable of interest which is also called the latent variable, the latent construct or the latent factor; and the observable variables which are termed manifest variables or indicators. Hence a latent variable is a hypothetical construct which is measured by several indicators.

LVM have been introduced by psychologists, and are used most extensively in the fields of psychology (psychometrics) and social sciences. The reason for this lies in the fact that most variables of interest in these areas cannot be measured directly, and thus are represented by latent variables. In natural sciences LVM have not been used as frequently because most variables could be measured directly. Nevertheless, LVM have been applied in a broad range of disciplines (including natural sciences) as diverse as chemistry, economics, geology, medicine and politics. Historically LVM have been mistrusted to some extent because researchers have sometimes attempted to interpret the latent variable as a true quantity based on the mathematical correlations alone (Gould, 1981). Today LVM provide an important element in the applied statistician's tool kit for the analysis of multivariate data. In the following, we give examples of latent variables in various fields to gain a more intuitive understanding of hypothetical constructs.

Probably the most famous latent construct is the notion of intelligence which was first introduced by Spearman (1904) in the field of psychology. Obviously the magnitude of a person's intelligence cannot be obtained directly and therefore represents a latent variable. Accordingly, psychologists tried to measure intelligence with a number of mental tests whose results are used as indicators, e. g. tests of spatial sense.

In medicine a syndrome can be regarded as a latent variable. Syndromes are diseases where a group of symptoms manifests itself simultaneously, hence the patient suffers from a variety of effects (e. g. fetal alcohol syndrome, Downs syndrome). The appearance of a single effect is not a reliable indicator if a patient is affected by the syndrome or not. Holmes et al. (1987) suggest that many teratology studies can gain power by looking at several different effects, i. e. indicators, on each subject.

An example of a latent variable in politics is provided by Quinn (2004) where political-economic risk is evaluated by five indicators which measure the independence of the national judiciary, the black-market premium, the expropriation risk, corruption and productivity for 62 countries. Other latent variables in politics include democracy, and economic freedom.

The broad range of hypothetical, unobservable variables of interest in various disciplines demonstrates the need for LVM, and their further development in applied statistical research.

The next section provides an introduction to the three different types of LVM which are used by applied researchers. The final section of this chapter summarizes the Bayesian developments in LVM and highlights advantages of the Bayesian estimation scheme for our model.

2.2 LVM model types

In this section we give a brief overview of the current landscape of LVM – for a full account we refer to Skrondal and Rabe-Hesketh (2004); a book about LVM in the context of social sciences is provided by Arminger et al. (1995). Firstly, we look at LVM excluding covariate effects. Secondly, we introduce LVM including covariate effects which is a natural expansion of LVM without covariate effects. This means that observed covariates can modify the indicators or the latent variables like in a standard parametric regression model. The main goal of this thesis is to extend the usual parametric predictor to a semiparametric predictor for covariates which influence the latent variables. Thirdly, to provide a complete picture, a short introduction to structural equation modeling (SEM) is given. SEM extends standard LVM by relating the latent variables to each other and to covariates in a system of linear regressions. Schematic path diagrams of the three models are depicted in Table 2.1. Each subsection gives an overview of existing models, brief historical background, references to relevant literature, and explains how the respective model class relates to the model in this thesis.

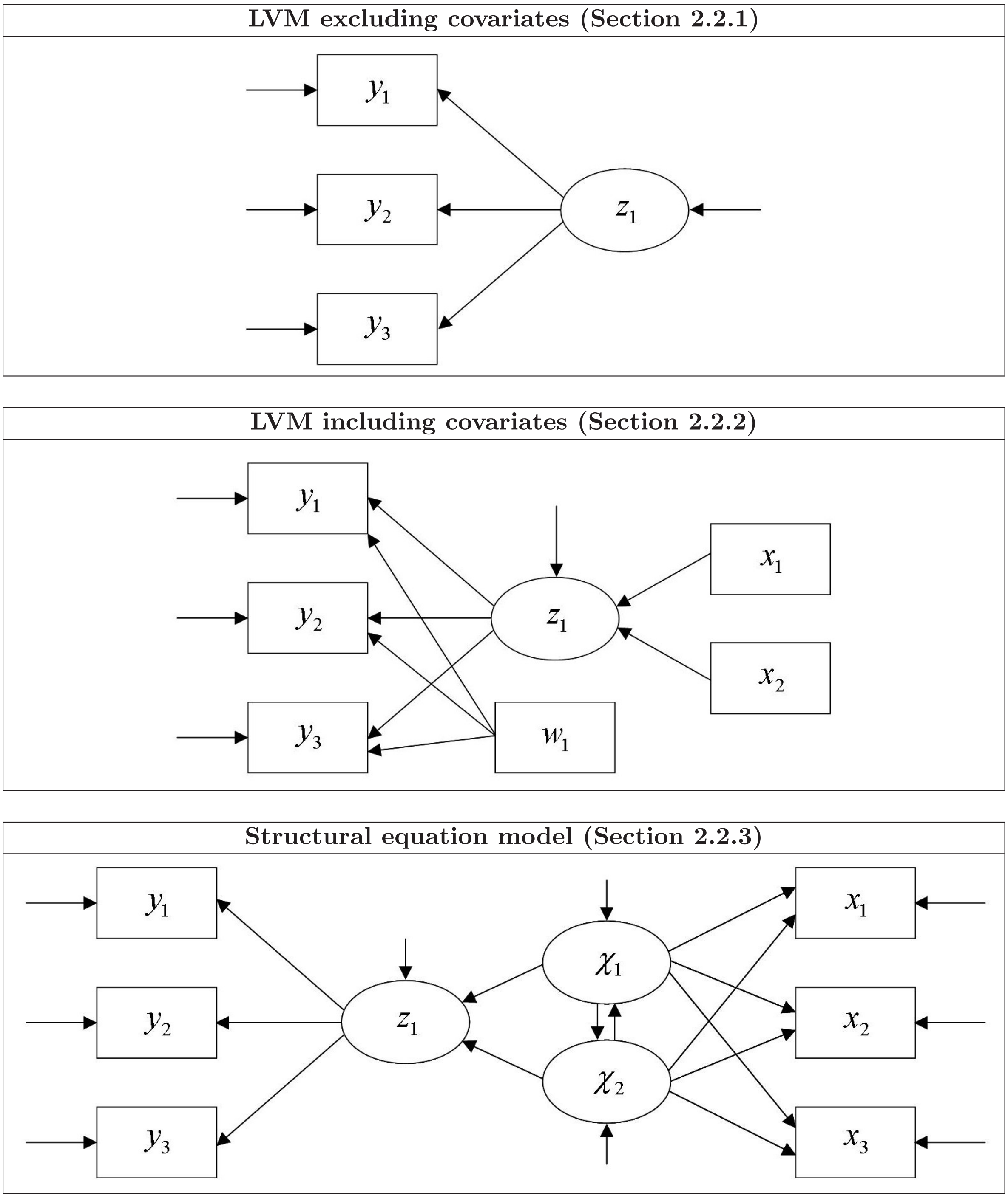


Table 2.1: *Schematic path diagrams of the LVM excluding covariates (top), the LVM including covariates (middle), and the structural equation model (bottom). Arrows denote the direction of influence, arrows without an origin symbolize error terms. Latent variables are indicated by circles, and manifest variables are represented by boxes.*

2.2.1 LVM excluding covariate effects

The statistical literature classifies LVM excluding covariates according to the metric (continuous) or categorical (discrete) type of both the latent variables and the manifest variables, as outlined in Table 2.2. Historically the different LVM of Table 2.2 originated from various sources and were developed rather independently from each other. Only recently the different models have been brought together in a unified approach which is demonstrated in more detail by Bartholomew (1987) or Skrondal and Rabe-Hesketh (2004). With regard to LVM, those textbooks also give a comprehensive view, theoretical account and literature references. The shaded areas of Table 2.2 indicate those methods on which the model of this thesis is based. Therefore we restrict our discussion to those models with metric, i. e. continuous latent variables.

Chronologically, factor analysis was the first LVM and was introduced by the psychologist Spearman (1904). He articulated the idea that several measurable intellectual capabilities or manifest variables $\boldsymbol{y}$ have a common fundamental function or latent factor z, in this case intelligence. The basic idea of factor analysis states that the multidimensional vector of p manifest variables $\boldsymbol{y}$ is represented by one or more latent factors $\boldsymbol{z}$ with a much lower dimension q. The basic factor analysis model consists of the so-called measurement model

$$\boldsymbol{y}_i = \boldsymbol{\Lambda} \boldsymbol{z}_i + \boldsymbol{\varepsilon}_i \,,$$

for each observation i. The term $\boldsymbol{\varepsilon}_i$ represents a p-dimensional error term, and $\boldsymbol{\Lambda}$ is a $(p \times q)$- dimensional matrix of regression coefficients named factor loadings that indicate the strengths of relationship between the manifest variables $\boldsymbol{y}_i$ and latent variables $\boldsymbol{z}_i$. The next chapter tells more about the distributional assumptions on the latent factors $\boldsymbol{z}_i$. Consequently, factor analysis reduces the dimensionality of the data in such a way that the interrelationships among the indicators are preserved as far as possible. The mutual correlation of the manifest variables should be solely explained by their common dependence on the latent variables. For that reason, factor analysis always involved the analysis of the correlation matrix of the manifest variables in its early times. The basic model is illustrated as a schematic path diagram in the upper part of Table 2.1. We see that

<table>
<tr><th></th><th></th><th colspan="2">Manifest variables</th></tr>
<tr><th></th><th></th><th>Metric</th><th>Categorical</th></tr>
<tr><td rowspan="2">Latent variables</td><td>Metric</td><td>Factor analysis</td><td>Factor analysis of categorical data
• Underlying variable approach
• Item response theory approach (latent trait analysis)</td></tr>
<tr><td>Categorical</td><td>Latent profile analysis</td><td>Latent class analysis</td></tr>
<tr><td></td><td></td><td colspan="2">Analysis of mixtures</td></tr>
</table>

Table 2.2: *Classification of LVM excluding covariates (adapted from Bartholomew, 1987). The model of this thesis covers the shaded areas.*

one single latent variable z_1 influences three manifest variables $\{y_1, y_2, y_3\}$, the direction of influence indicated by arrows. Additionally, there are individual error terms added to the observed manifest variables, illustrated by arrows without origins. Of course, there might be more than one latent variable accounting for the interdependencies among the indicators in general.

The first comprehensive treatment of factor analysis in a sound statistical way – including a proper probability model – has been written by Lawley and Maxwell (1963) in their well-known book *Factor analysis as a statistical method*. Here, the starting point was a linear model in which the manifest variables were expressed as linear functions of the latent factors plus random errors. Now the theory of statistical inference, especially estimation and hypothesis testing using the likelihood function, could be fully employed. This was accompanied by a shift to analyzing the covariance matrix instead of the correlation matrix, and laid the foundations for covariance structure analysis which is used extensively in structural equation modeling.

Traditional factor analysis only allows the use of continuous manifest variables. This is a severe restriction because most analyses in psychometrics and social sciences rely on binary or ordinal data. Treating ordinal data as normally distributed data leads to wrong parameter estimates, especially if the number of categories is low. To solve this issue, two different approaches have emerged in the literature to deal with factor analysis of categorical data:

- **Underlying variable approach (UVA)**:
  This paradigm assumes that the categorical observed variables $\boldsymbol{y}_i$ are created by underlying unobserved continuous variables which are normally distributed. Muthén (1984) proposed a method for dealing with ordinal variables, as did Jöreskog (1994). Other notable contributions regarding this approach have been made by Arminger and Küsters (1988; 1989), and Lee, Poon and Bentler (1992; 1995). Frequentist estimation methods typically are maximum likelihood (ML), generalized least squares (GLS), or weighted least squares (WLS) with two- or three-stage estimation algorithms that are described in the literature in more detail.

- **Item response theory approach (IRT)**:
  The IRT approach specifies the conditional distribution of the full dimensional response pattern of the manifest variables as a function of the latent variable (typically only one in IRT). This means that there is no distributional assumption made about the latent variables. IRT emanated from educational testing where responses typically are binary or ordinal. Most prominent IRT models include the one-parameter model from Rasch in 1960, the two-parameter model from Lord and Birnbaum in 1952, the partial credit model by Masters in 1982, and the graded response model from Samejima in 1969. These and many additional models are thoroughly described in the standard work *Handbook of Modern Item Response Theory* by van der Linden and Hambleton (1997). Another recent paper about a LVM for ordinal variables in

the IRT setting is written by Moustaki (2000). Inference is typically done by joint ML or marginal ML estimation with an EM algorithm.

Although UVA and IRT approaches seem to be incompatible, Takane and de Leeuw (1987) demonstrated their parameter equivalence. Specifically, factor analysis for ordinal responses with the UVA approach is equivalent to the two-parameter item response model with probit link. Still, parameter estimation processes are rather different for UVA and IRT. Jöreskog and Moustaki (2001) describe four different approaches to factor analysis of ordinal variables in both the UVA and IRT setting. A historical account of the relationship between UVA and IRT is given in Reckase (1997).

Latent structure analysis which comprises models with categorical latent variables (latent profile and latent class analysis) originated with Lazarsfeld (1968) as a tool for sociological analysis. A more up-to-date account is delivered in Everitt (1984). Historically considered as two completely separate model classes, factor analysis and latent structure analysis can be brought together in a unified framework as demonstrated in detail by Bartholomew (1987), or Skrondal and Rabe-Hesketh (2004).

The model in this thesis uses the UVA for the factor analysis of ordinal indicators. One reason for this is that the underlying normally distributed variables behind ordinal responses can be quite naturally incorporated into a Bayesian estimation approach, as shown in Chapter 4. Our model extends the factor analysis model by incorporating covariates. Therefore, we proceed with the description of LVM with covariate effects.

2.2.2 LVM including covariate effects

Hitherto, the latent variables solely account for the dependencies and relationships among the manifest variables, although the dimension of the latent variables is much lower than the number of manifest variables. There are two reasons why this model should be extended. On the one hand, there are applications where it is useful to allow explanatory variables to directly effect the observed manifest variables in addition to the latent factors. These explanatory variables are named direct effects $\boldsymbol{w}$. On the other hand, one might be interested in how explanatory variables modify the latent factors, and hence affect the observed variables indirectly. Therefore, those explanatory variables are called indirect effects $\boldsymbol{x}$. In this study we put a strong focus on the indirect effects as we are interested in how explanatory variables, e. g. demographic variables, affect the latent construct. Now the model consists of two parts. Firstly, the measurement model

$$\boldsymbol{y}_i = \boldsymbol{\Lambda} \boldsymbol{z}_i + \boldsymbol{A} \boldsymbol{w_i} + \boldsymbol{\varepsilon}_i$$

extends the classic measurement model with direct effects $\boldsymbol{w}_i$ plus the matrix of regression coefficients $\boldsymbol{A}$. Secondly, the structural equation model links the latent variables and the indirect covariates according to

$$\boldsymbol{z}_i = \boldsymbol{\gamma} \boldsymbol{x}_i + \boldsymbol{\xi}_i.$$

The matrix $\boldsymbol{\gamma}$ contains the regression coefficients of the indirect covariates $\boldsymbol{x}$, and $\boldsymbol{\xi}$ represents a q-dimensional error term. A schematic path diagram of this extended model can be found in the middle part of Table 2.1. In addition to the basic model, the direct covariate w_1 explains correlations between the manifest variables y_i, whereas indirect covariates x_1 and x_2 modify the latent variable z_1. Again, two or more latent variables might be included in the model.

Equivalent to the case of LVM excluding covariate effects, the landscape is divided into the two paradigms UVA (including standard continuous factor analysis) and IRT. The UVA approach is used within the general framework of structural equation modeling (SEM), and therefore all currently available SEM packages are able to estimate LVM with covariate effects as a special case of a SEM. We refer to the next section for more details on available algorithms and packages. Hence the LVM including covariate effects can be viewed as both an extension of the LVM without covariate effects, or a special case of the more general class of SEM.

Furthermore, the special case of a LVM with covariate effects also appeared independently in the literature. Jöreskog and Goldberger (1975) named and discussed a multiple indicators and multiple causes model (MIMIC) which included continuous normally distributed manifest variables with a single latent factor, direct and indirect parametric effects. This type of model has already been treated before by Zellner (1970), and Hauser and Goldberger (1971). Muthén's (1989) work extended the MIMIC model to include binary and ordinal manifest variables. A good summary of modeling and estimation issues of this model can be found in Browne and Arminger (1995). Recently Zhu, Eickhoff and Yan (2005) discussed a LVM for Gaussian and non-Gaussian manifest variables which may vary both across space and over time.

Within the IRT context, traditionally there was no focus on the inclusion of covariates. However, several authors made contributions in this area recently. Verhelst, Glas and Verstralen (1994), Zwinderman (1997), and Glas (2001) discussed the one-parameter logistic model with covariate effects. Sammel and Ryan (1996) and Sammel, Ryan and Legler (1997) discussed a LVM with covariates for mixed outcomes, whereas Moustaki's (2003) work solely allows for ordinal variables. A comparison of UVA and IRT methods for ordinal variables including covariate effects is given in Moustaki, Jöreskog and Mavridis (2004) in the frequentist setting.

All the papers mentioned in this section (apart from the work of Zhu et al., 2005) have one thing in common: only parametric effects are considered to modify the manifest and latent variables. The model presented in this thesis resolves the restrictions imposed by parametric effects, and introduces nonparametric effects on the latent variables. We might also include a nonparametric predictor which influences the manifest variables directly but we consider a more detailed and thorough analysis of covariates on the latent variables as much more illuminating and interesting from an applied researcher's view. The predictor affecting the latent variables comprises additive models (AMs), geoadditive models, and varying coefficients models (VCMs), as detailed out in Section 3.2. Hence we develop a LVM with a more versatile predictor including non-parametric components which are already available and widely used in semiparametric extensions of standard and generalized regressions.

To complete the picture of LVM, we conclude with a brief summary of structural equation modeling.

2.2.3 Structural equation models

Structural equation models (SEM) represent a further extension of factor analysis by relating latent variables to each other and to covariates in a system of linear regressions (see Bollen, 1989). A SEM consists of two measurement models, i. e.

$$\begin{aligned} \boldsymbol{y}_i &= \boldsymbol{\Lambda}_y \boldsymbol{z}_i + \boldsymbol{\varepsilon}_i \,, \\ \boldsymbol{x}_i &= \boldsymbol{\Lambda}_x \boldsymbol{\chi}_i + \boldsymbol{\delta}_i \,, \end{aligned}$$

with latent factors $\boldsymbol{z}_i$ and $\boldsymbol{\chi}_i$, matrices $\boldsymbol{\Lambda}_y$ and $\boldsymbol{\Lambda}_x$ of regression coefficients, and error terms $\boldsymbol{\varepsilon}_i$ and $\boldsymbol{\delta}_i$. Furthermore, the structural equation links the latent variables of both measurement models according to

$$\boldsymbol{z}_i = \boldsymbol{B}\boldsymbol{z}_i + \boldsymbol{\Gamma}\boldsymbol{\chi}_i + \boldsymbol{\xi}_i \,,$$

where $\boldsymbol{B}$ and $\boldsymbol{\Gamma}$ are parameter matrices containing regression coefficients, and $\boldsymbol{\xi}_i$ represents another error term. A schematic illustration of a structural equation model is depicted in the path diagram at the bottom of Table 2.1. The manifest variables $\boldsymbol{x}$ are influenced by the latent variables $\boldsymbol{\chi}$, representing one factor analysis model. The latent variables $\boldsymbol{\chi}$ can modify themselves and build a regression relationship with the target latent variable z, forming the structural equations part of the model. Finally, the latent variable z is measured by further indicators $\boldsymbol{y}$ in the second measurement model.

Although SEM are extensively employed in applied research (e. g. organizational research), it is still not fully accepted in mainstream statistics. One reason for this is given by the fact that SEM heavily relies on latent variables which could not be directly measured; this property was and still is somewhat suspicious for statisticians and mathematicians alike. However, it is rather intriguing that there exist many connections between LVM (including SEM) and established statistical concepts such as random effects, missing data, finite mixtures and others (see Muthén, 2002; Skrondal and Rabe-Hesketh, 2004).

Due to its very applied nature, all SEM algorithms have been implemented in commercial software packages and are known by those packages' names such as LISREL (Jöreskog and Sörbom, 1996), Mplus (Muthén and Muthén, 1998-2001), EQS (Bentler, 1995) and MECOSA (Arminger, Wittenberg and Schepers, 1996). These packages can deal with both full-scale SEMs and the special case of a LVM including covariate effects in the UVA setting as described in the last section.

Finally, we want to give an overview of Bayesian developments which have occurred in the relevant field of factor analysis with continuous, binary and ordinal response based on the UVA.

2.3 Bayesian accounts on LVM

The literature referred so far is almost exclusively based on frequentist approaches. The reason for this is that the other big statistic paradigm, the Bayesian framework (see Chapter 4 for details), has been rarely employed in applications until the early 1990s due to a lack of computing power and suitable numerical methods (e. g. MCMC). Therefore, most of the evolvement and historical development of LVM occurred without the consideration of Bayesian models. During the last 15 years Bayesian contributions have been made to LVM (especially in the field of IRT) – an overview of the recent developments including literature references is provided by Rupp, Dey and Zumbo (2004). Here we want to focus on the UVA that is relevant to our model. Nowadays Bayesian methods are a viable alternative to traditional frequentist approaches, moreover they have several advantages which are outlined below. Furthermore, Bayesian models are sometimes the only way to analyze more complex data structures, and to estimate more sophisticated statistical models.

Continuous factor analysis has been discussed by several authors. Lee (1981) estimates factor loadings and error variances by using a Newton-Raphson algorithm in the context of a confirmatory factor analysis. Bartholomew (1981) performed the estimation of factor scores, as did Press and Shigemasu (1989) who also gave an overview of Bayesian approaches in factor analysis. Shi and Lee's (1997) work calculated factor scores for polytomous and missing data. Mixed continuous and ordinal responses are incorporated in the model of Quinn (2004).

Models including covariates have been rarely treated so far. Sammel and Ryan (1996) and Sammel, Ryan and Legler (1997) presented models allowing for the inclusion of covariates by using empirical Bayes estimation procedures.

In the area of SEM, an overview of Bayesian estimation and testing is provided by Scheines, Hoijtink and Boomsma (1999). Arminger and Muthén's (1998) paper extends traditional SEM by allowing nonlinear relationships between latent variables.

Bayesian models are especially useful when used for problems which cannot be dealt with easily by frequentist approaches, such as the just mentioned work of Arminger and Muthén (1998), the estimation of factor analysis for multilevel binary responses (Ansari and Jedidi, 2000), dynamic factor components with time series (Aguilar and West, 2000), or determining the right number of latent variables (Lopes and West, 2004).

We employ the Bayesian approach because it offers the following advantages compared to a frequentist approach:

- The factor scores (values of the latent variables) are automatically estimated. In a frequentist approach, however, factor scores have to be calculated separately after model estimation. There has been a long and controversial debate in the statistical community on how to estimate factor scores, and as a result several estimation methods are available, e. g. Bartlett's scores or regression scores (Bartholomew, 1987). This issue is neatly resolved in the Bayesian context, where the factor scores are treated as random variables and are automatically obtained as a byproduct of the estimation process.

- The marginal distribution of all parameters and factor scores are obtained. Hence uncertainty and range of parameter values can be immediately analyzed after the estimation process by looking at the marginal distributions. This also makes classic hypothesis testing obsolete. Additionally, probabilistic statements about factor scores are possible, for example we can easily calculate the probability that one observation has a higher latent score than another observation.

- The full posterior distribution of the model is analyzed, and hence the complete information of the interrelationships among the indicators is incorporated in the estimation process. This is in contrast to most of the UVA algorithms used by frequentist approaches, where only univariate and bivariate marginal likelihoods of response variables are evaluated (Jöreskog and Moustaki, 2001).

- A Bayesian approach is very powerful for the estimation of nonparametric effects including the estimation of the degree of smoothing (see Section 5.4.1).

A LVM with nonparametric effects is a very complex model with a high number of parameters. Therefore and due to the just mentioned reasons, the Bayesian approach is especially suited for estimation purposes.

Chapter 3

Statistical Model

The LVM with covariate effects consists of two components. Firstly, the measurement model is discussed for continuous, binary and ordinal response with direct effects, also called measurement model. In this part of the model, the influences of latent variables and direct effects on the manifest variables are formulated. Secondly, structural equations explain the modification of the latent variables by indirect effects. This thesis breaks new ground in the structural equation part of the model by introducing nonparametric effects influencing the latent variables. Thus the next two sections describe the measurement model and the structural equation model, respectively.

3.1 The measurement model

3.1.1 Measurement model excluding direct effects

In all LVM settings, we observe p different indicators (i. e. manifest variables) for n observations. Each manifest variable j $(1 \leq j \leq p)$ can be continuous, binary or ordinal. In future, we always include binary variables when referring to ordinal variables. The specific response value of indicator j of individuum i $(1 \leq i \leq n)$ is denoted by y_{ij}. All indicators of a single individuum are contained in the $(p \times 1)$-dimensional vector $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{ip})'$. For notational convenience, we sort the manifest variables in such a way that the first p_1 indicators are binary or ordinal, and the remaining $p_2 = p - p_1$ indicators are continuous. The full response for a single manifest variable j is arranged in the $(n \times 1)$-dimensional vector $\boldsymbol{y_j} = (y_{1j}, \ldots, y_{nj})'$. In general, a quantity Q, whose elements are contained in a $(n \times p)$-dimensional matrix $\boldsymbol{Q} = \{q_{ij}\}$, can be arranged in two different ways: the vector $\boldsymbol{q}_i$ contains all elements of the i-th row, $\boldsymbol{q}_i = (q_{i1}, q_{i2}, \ldots, q_{ip})'$, and the vector $\boldsymbol{q}_j$ contains all elements of the j-th column, $\boldsymbol{q}_j = (q_{1j}, q_{2j}, \ldots, q_{nj})'$.

In order to incorporate ordinal indicators into our model, we postulate an underlying unobserved variable y^*_{ij} associated with each response[1]. In the literature, the underlying variable is often termed "latent variable". We do not use this term in this context since it is already reserved for the latent factors. The underlying variable (UV) is typically assumed to be drawn from a continuous distribution centered at a mean value μ which in our case is obtained by the latent factors and direct effects as given below. The discrete values of an ordinal indicator are obtained by partitioning the domain of the UV at specific cutpoints. Let us assume that ordinal indicator j has K_j categories and its cutpoints are denoted by τ_{jc} $(0 \leq c \leq K_j)$. The discrete value of an ordinal indicator y_{ij} is generated by the UV y^*_{ij} according to

$$y_{ij} = k \quad \Longleftrightarrow \quad \tau_{j,k-1} < y^*_{ij} \leq \tau_{jk}\,, \tag{3.1}$$

with $1 \leq i \leq n$, $1 \leq j \leq p_1$, and $1 \leq k \leq K_j$. Since ordinal categories are ordered, we have to impose an order restriction on the cutpoints as stated by

$$-\infty =: \tau_{j0} < \tau_{j1} < \tau_{j2} < \ldots < \tau_{jK_j} := \infty\,.$$

The distribution of the UV is governed by the equation

$$y^*_{ij} = \mu_{ij} + \varepsilon_{ij}\,,$$

where ε_{ij} is a random error variable drawn from a continuous distribution f. Let F denote the respective cumulative distribution function. Using (3.1) the probability p_{ijk} of observing category k for individual i and indicator j leads to

$$\begin{aligned} p_{ijk} = P(y_{ij} = k|\mu_{ij}) &= P(\tau_{j,k-1} < y^*_{ij} \leq \tau_{jk}|\mu_{ij}) \\ &= \int_{\tau_{k-1}}^{\tau_k} f(y^*_{ij} - \mu_{ij})\, dy^*_{ij} \\ &= F(\tau_{jk} - \mu_{ij}) - F(\tau_{j,16gk-1} - \mu_{ij})\,. \end{aligned} \tag{3.2}$$

In standard ordinal regression models, the mean μ_{ij} is given by the predictor comprising covariates and regression coefficients. In our model, however, the mean μ_{ij} is determined by the factor loadings and latent variables as given below in the measurement model. Finally, we can write (3.2) in compact form by stating the cumulative probability for category k summing up all individual probabilities for categories 1 to k

$$P(y_{ij} \leq k|\mu_{ij}) = p_{ij1} + p_{ij2} + \ldots + p_{ijk} = F(\tau_{jk} - \mu_{ij}). \tag{3.3}$$

Alternatively this can be written by $F^{-1}\{P(y_{ij} \leq k|\mu_{ij})\} = \tau_{jk} - \mu_{ij}$ where F^{-1} denotes the link function. The two most common functions in practice are the probit model with probit link for normally distributed errors and $F = \Phi$ as the cumulative distribution function; and the logit model with logit link and F as the logistic distribution function. The logit model is very popular in the IRT approach. Although parameter estimates for both models seem

[1]This postulate names the underlying variable approach (UVA).

to differ, both models lead to very similar results in prediction (Moustaki, 2003). So most of the times, the choice between the two models is a matter of convenience. Other link functions appearing in the literature are the log-log function, the complementary log-log function, and the inverse Cauchy function. In this thesis, the probit link is employed which allows the use of a standard MCMC scheme, the Gibbs sampler.

In order to illustrate the UV concept, we have drawn three figures for an indicator j with four categories and cutpoints $\tau_{j1} = 0$, $\tau_{j2} = 2$, $\tau_{j3} = 3$ in Figure 3.1. The top figure shows the relationship between the normally distributed UV, and the partitioning into the four response categories for $\mu_{ij} = 1.25$. The middle figure shows the response probabilities for all 4 categories over the value range of the underlying variable, and the bottom figure displays the corresponding cumulative probabilities. Accounts about modeling of ordinal variables are given in Johnson and Albert (1999), and Fahrmeir and Tutz (2001).

Unlike ordinal indicators, continuous variables are observed directly, hence the underlying variable is obsolete, i. e.

$$y_{ij}^* = y_{ij} \,,$$

for $p_1 < j \leq p$. The UV for ordinal indicators and the observed variables for continuous indicators for each observation $\boldsymbol{y_i}$ are considered to be independently and identically distributed (i. i. d.) with a p-dimensional multivariate normal distribution

$$\boldsymbol{y}_i^* \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Phi}) \,, \tag{3.4}$$

and a $p \times p$ non-singular covariance matrix $\boldsymbol{\Phi}$. Equation (3.4) represents a marginal model for $\boldsymbol{y}_i^*$ in which the latent variables $\boldsymbol{z}_i$ do not appear. Both expectation vector and covariance matrix depend on a parameter vector $\boldsymbol{\theta}$ which yields $\boldsymbol{\mu} = \boldsymbol{\mu}(\boldsymbol{\theta})$ and $\boldsymbol{\Phi} = \boldsymbol{\Phi}(\boldsymbol{\theta})$, respectively. The elements of the parameter vector $\boldsymbol{\theta}$ follow from the parameterization of the statistical model.

The relationship between the $\boldsymbol{y}_i^*$ variables and the latent variables $\boldsymbol{z}_i$ is given by the measurement model representing the conditional model $\boldsymbol{y}_i^*|\boldsymbol{z}_i$, i. e.

$$\boldsymbol{y}_i^* = \boldsymbol{\lambda}_0 + \boldsymbol{\Lambda}\boldsymbol{z}_i + \boldsymbol{\varepsilon}_i \,, \tag{3.5}$$

with $\boldsymbol{\varepsilon}_i \sim N_p(\boldsymbol{0}, \boldsymbol{\Sigma})$, $\boldsymbol{\Sigma} = \text{diag}(\sigma_1^2, \ldots, \sigma_p^2)$, and $\boldsymbol{z}$ and $\boldsymbol{\varepsilon}$ being stochastically independent. The q-dimensional vector $\boldsymbol{z}_i$ contains q latent factors which explain the relationships between the indicators $\boldsymbol{y}_i^*$. The p indicators $\boldsymbol{y}_i^*$ result from a linear combination of q latent factors plus individual error terms for each indicator. The $p \times q$ matrix $\boldsymbol{\Lambda}$ is composed of the factor loadings indicating the strength of relationship between latent variables and indicators. $\boldsymbol{\lambda}_0$ is a p-dimensional intercept vector. Let us suppose that the latent variables $\boldsymbol{z}_i$ are i.i.d. with $\boldsymbol{z}_i \sim N_q(\boldsymbol{0}, \boldsymbol{\Psi})$.

Since we assumed $\boldsymbol{\Sigma}$ to be diagonal, the model implies that the latent variables $\boldsymbol{z}$ are solely responsible for the correlations between the manifest variables $\boldsymbol{y}_i^*$ (see Bartholomew, 1987), hence the latent variables explain all the dependence structure among the q indicators.

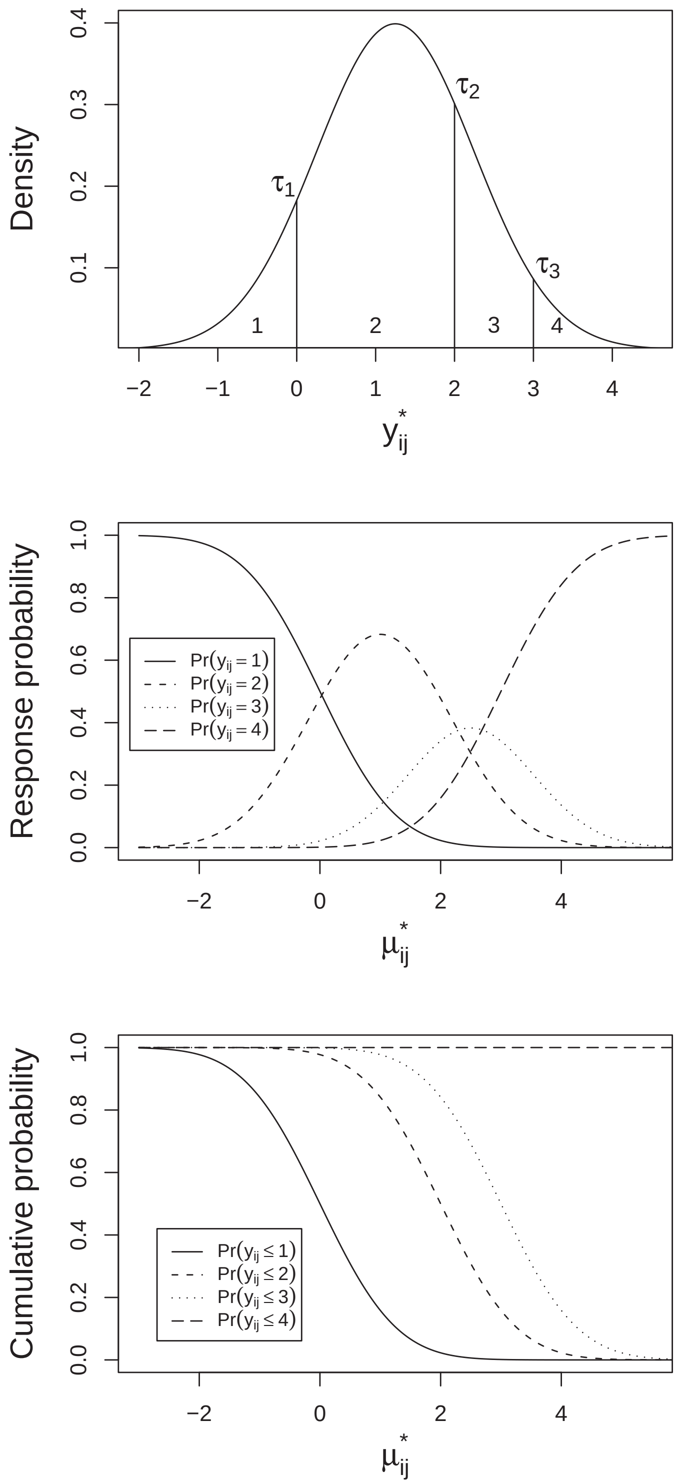


Figure 3.1: ***Top***: *Partitioning of the latent scale along the underlying variable y^*_{ij} into 4 categories for $\mu_{ij} = 1.25$, $\tau_{ij1} = 0$, $\tau_{ij2} = 2$, and $\tau_{ij3} = 3$.* ***Middle***: *Individual response probabilities.* ***Bottom***: *Cumulative response probabilities.*

When there is still a dependence structure among the $\boldsymbol{y}^*$ variables given the latent factors – for example when the components of the residual vector $\boldsymbol{\varepsilon}_i = \boldsymbol{y}_i^* - \boldsymbol{\lambda}_0 - \boldsymbol{\Lambda}\boldsymbol{z}_i$ are correlated – it might be necessary to include an additional latent variable which could absorb this remaining correlation.

For the remainder of this passage we assume the covariance matrix of the latent variables to be the identity matrix, i. e. $\boldsymbol{\Psi} = I$; the reason for this assumption is explained in the next subsection. Then we can conveniently describe the properties of the factor analysis model by its characterizing moments of two arbitrary indicators y_{ij}^* and y_{il}^*

- $\mathrm{Var}(y_{ij}^*|\boldsymbol{z}) = \sigma_j^2$,
- $\mathrm{Var}(y_{ij}^*) = \sum_{r=1}^{q} \lambda_{jr}^2 + \sigma_j^2$,
- $\mathrm{Cov}(y_{ij}^*, y_{il}^*|\boldsymbol{z}) = 0$,
- $\mathrm{Cov}(y_{ij}^*, y_{il}^*) = \sum_{r=1}^{q} \lambda_{jr}\lambda_{lr}$,
- $\boldsymbol{\Phi} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}' + \boldsymbol{\Sigma}$.

The communality $\widetilde{y}_j$ of an indicator which is defined as the proportion of variance accounted for by all latent factors results in

$$\widetilde{y}_j = \frac{\sum_{r=1}^{q} \lambda_{jr}^2}{\sum_{r=1}^{q}\lambda_{jr}^2 + \sigma_j^2} .$$

Accordingly the amount of an indicator's variance accounted for by the error term $\boldsymbol{\varepsilon}_j$ yields $1 - \widetilde{y}_j$, the complement of communality.

The factor analysis model presented in this section contains two important models as special cases. When all indicators are continuous, the traditional factor analysis model is obtained (Lawley and Maxwell, 1971). In that context the indicators typically are standardized; therefore, the p-dimensional intercept λ_0 can be omitted. When all indicators are dichotomous or ordinal, we obtain the two-parameter item response model for binary or ordinal data, respectively (Johnson and Albert, 1999).

3.1.2 Identification restrictions of the factor analysis model

The factor analysis model faces two sources of identification problems; the first problem is associated with the modeling of ordinal variables, and the second problem is related to the uniqueness of the factor loadings matrix $\boldsymbol{\Lambda}$ and factor scores. We demonstrate the identification problems on the model without covariates, because the problems become clearer and remain equal if covariates are included.

Ordinal modeling

Using the normality assumption for y^*_{ij} and Equations (3.1), (3.3) and (3.5) we get

$$P(y_{ij} \leq k) = P(y^*_{ij} \leq \tau_{jk}) = \mathbf{\Phi}\left(\frac{\tau_{jk} - \lambda_{j0}}{\sqrt{\Phi_{jj}}}\right) \tag{3.6}$$

if latent scores and errors are assumed to be zero. Two problems arise with this probability statement:

- The probability on the right hand side of Equation (3.6) does not change if we add a constant c_1 to both the cutpoints τ_{jk} and the intercept λ_{j0} because the value of the nominator remains constant. Only the difference between the cutpoints τ_{jk} and the intercept λ_{j0} is identified. We solve this issue by fixing the first cutpoint of all indicators to zero, $\tau_{j1} = 0$, $j = 1, \ldots, p_1$. This is a standard procedure in the literature (e. g. Johnson and Albert, 1999). Ergo, only $K_j - 2$ cutpoints for each indicator j have to be estimated.

- The probability is also not altered when we multiply the nominator and the denominator by a constant c_2, implying that τ_{jk}, λ_{j0} and ϕ_{jj} are only identified up to a multiplicative constant. In our model, we dispose of this identification problem by setting the error variances $V(\varepsilon_{ij}) = \sigma^2_j = 1$. In the frequentist literature instead, it is common to fix the total variance to one, i. e. $V(y^*_{ij}) = \phi_{jj} = 1$, which explains the use of the standard normal distribution function when demonstrating the partitioning of the latent scale for ordinal categories (top picture of Figure 3.1). For that reason the parameter estimates of the Bayesian and the frequentist models cannot be compared directly. More about the conversion of frequentist and Bayesian parameter estimates, and the relation between those two approaches is provided in Section 3.1.4. The reason why we proceed in a different way in the Bayesian context lies in the fact that full conditional distributions are of an easier form for the constraint $V(\varepsilon_{ij}) = \sigma^2_j = 1$, as can be seen in Section 5.4.1.

Uniqueness of factor loadings and scores

The measurement model possesses two further identification problems with regard to the parameters of the factor loadings and scores.

- Consider the transformation of Equation (3.5) with a $q \times q$ non-singular matrix $\boldsymbol{T}$ (e. g. Bartholomew, 1987), i. e.

$$\boldsymbol{y}^*_i = \boldsymbol{\lambda}_0 + \boldsymbol{\Lambda}\boldsymbol{T}^{-1}\boldsymbol{T}\boldsymbol{z}_i + \boldsymbol{\varepsilon}_i\,, \tag{3.7}$$

with the new factor loadings matrix $\boldsymbol{\Lambda T^{-1}}$, new latent scores $\boldsymbol{T z}_i$ and thus $V(\boldsymbol{z_i}) = \boldsymbol{T\Psi T'}$. If there are no restrictions for $\boldsymbol{\Lambda}$ or $\boldsymbol{\Psi}$, an indefinite number of different models could be created. Since the matrix $\boldsymbol{T}$ consists of q^2 elements, we have to impose q^2 restrictions in the model. Without loss of generality – in the tradition of exploratory factor analysis – we choose the variance matrix of the latent scores to be the q-dimensional identity matrix I_q, leading to $\boldsymbol{z}_i \sim N_q(\boldsymbol{0}, \boldsymbol{I_q})$. Hence the latent scores have a standard normal distribution, and no correlations among the latent variables exist.

- After fixing the variance matrix of the latent factors, there is still an indeterminacy with respect to the factor loadings matrix and factor scores. The model is invariant under transformations with any orthogonal $q \times q$ matrix $\boldsymbol{V}$ of the form $\widetilde{\boldsymbol{\Lambda}} = \boldsymbol{\Lambda V'}$ and $\widetilde{\boldsymbol{z}}_i = \boldsymbol{V z}_i$, because this transformation keeps the variance of the latent scores unchanged ($V(\boldsymbol{z}_i) = \boldsymbol{V I}_k \boldsymbol{V'} = \boldsymbol{\Psi}$). An indefinite number of models exists again since all orthogonal rotations of the latent space could occur. The solution lies in the restriction of parameters of the factor loadings matrix $\boldsymbol{\Lambda}$ in a suitable way (e. g. Seber, 1984). We choose the factor loadings matrix $\boldsymbol{\Lambda}$ to be a lower block triangular matrix of full rank and positive diagonal elements (see Geweke and Zhou, 1996; Aguilar and West, 2000) with $f = pq - q(q-1)/2$ free parameters. This guarantees identification and provides, most of the times, useful interpretation of the factor model. This restriction is typically not necessary for our analyses which only use one single latent factor. More about the reason for this choice can be found in Section 3.1.5. If one latent variable is employed there is only one possible orthogonal transformation, that is the change of the sign for factor loadings and factor scores. If the sign is not appropriate, the problem can be easily resolved by multiplying factor loadings and factor scores by -1 after the analysis.

3.1.3 Measurement model including direct effects

We extend the model of Equation (3.5) to allow for covariate effects influencing the indicators, leading to the measurement model

$$\boldsymbol{y}_i^* = \boldsymbol{\lambda}_0 + \boldsymbol{\Lambda z}_i + \boldsymbol{A w}_i + \boldsymbol{\varepsilon}_i \,,$$

with a d-dimensional vector of covariates $\boldsymbol{w}_i = (w_{i1}, \ldots, w_{id})'$ for each individuum i and a $(p \times d)$-dimensional matrix $\boldsymbol{A}$ of regression coefficients. We call the covariates $\boldsymbol{w}_i$ direct effects because they directly modify the indicators $\boldsymbol{y}_i^*$. Although direct effects are typically not in the focus of an analysis[2], they still provide additional information about the data structure and increase the strength of dimensionality by including associations between

[2]Most of the times the researcher is more interested in the factor loadings, the latent scores, and the influence of covariates on the latent variable. If direct effects were mainly responsible for the correlations among the manifest variables, a factor analysis would not be necessary in the first place.

		Total variance	Error variance	Communality
Ordinal indicators	*Bayesian solution*	$\phi_{jj}^2 = \sum_{r=1}^{q} \lambda_{jr}^2 + 1$	$\sigma_j^2 = 1$	$\frac{\sum_{r=1}^{q} \lambda_{jr}^2}{\sum_{r=1}^{q} \lambda_{jr}^2 + 1}$
	Standardized solution	$\bar{\phi}_{jj}^2 = 1$	$\bar{\sigma}_j^2 = 1 - \sum_{r=1}^{q} \bar{\lambda}_{jr}^2$	$\sum_{r=1}^{q} \bar{\lambda}_{jr}^2$
Metric indicators	*Bayesian solution*	$\phi_{jj}^2 = \sum_{r=1}^{q} \lambda_{jr}^2 + \sigma_j^2$	$\sigma_j^2 = \sigma_j^2$ (*estimated*)	$\frac{\sum_{r=1}^{q} \lambda_{jr}^2}{\sum_{r=1}^{q} \lambda_{jr}^2 + \sigma_j^2}$
	Standardized solution	$\bar{\phi}_{jj}^2 = 1$	$\bar{\sigma}_j^2 = 1 - \sum_{r=1}^{q} \bar{\lambda}_{jr}^2$	$\sum_{r=1}^{q} \bar{\lambda}_{jr}^2$

Table 3.1: *Total variances, error variances, and communalities for the Bayesian and the standardized solution, and for ordinal and metric indicators, respectively. Standardized parameters are denoted by a bar on top.*

indicators $\boldsymbol{y}_i$ and covariates $\boldsymbol{w}_i$, and not only among the $\boldsymbol{y}_i$ themselves (see Muthén, 1989). Now both latent variables and covariates explain the correlations between the $\boldsymbol{y}_i$ variables, instead of the latent variables alone.

An illustrative example for the inclusion of direct effects is provided by Sammel and Ryan (1996). The authors examined a data set where the health of infants after birth is measured by several indicators. The main interest of this study is answering the question if the dispensation of a certain drug influences the infants' state of health. Since it is not expected that the effect of the drug is gender specific, it is still possible that some indicators, e. g. body length of the infant, depend on gender. Therefore, gender is included as a direct effect in the model.

3.1.4 Standardization of parameters

Using the terminology of SEM, we call parameter values standardized if the total variance of an indicator given covariates $\boldsymbol{w}_i$ is one, i. e. $V(y_{ij}^*|\boldsymbol{w}_i) = \phi_{jj} = 1$. In our model, both ordinal and continuous indicators are not standardized. For ordinal indicators we set $\sigma_j = 1$ and have to add the variance of the latent variables leading to a total indicator variance greater than one. For continuous variables, the variance parameter σ_j is estimated, and it would be a mere coincidence if the sum of variance parameter and squared factor loadings resulted exactly in one. In order to make parameter values comparable between ordinal and continuous variables, or between the Bayesian and the frequentist approaches, a standardization of parameters is advisable. We summarize the properties and the conversion formulas in Tables 3.1 and 3.2 for the non-standardized and standardized solutions, respectively. Another advantage of the standardized solution is that communalities can be obtained easily, e. g. in a model with one latent variable the communality of an indicator j is λ_{j1}^2.

	Parameters	Bayesian to Standardized	Standardized to Bayesian
Ordinal indicators	*Factor loadings*	$\bar{\lambda}_{jr} = \frac{\lambda_{jr}}{\sqrt{\sum_{r=1}^{q} \lambda_{jr}^2 + 1}}$	$\lambda_{jr} = \frac{\bar{\lambda}_{jr}}{\sqrt{1 - \sum_{r=1}^{q} \bar{\lambda}_{jr}^2}}$
	Intercepts	$\bar{\lambda}_{j0} = \frac{\lambda_{j0}}{\sqrt{\sum_{r=1}^{q} \lambda_{jr}^2 + 1}}$	$\lambda_{j0} = \frac{\bar{\lambda}_{j0}}{\sqrt{1 - \sum_{r=1}^{q} \bar{\lambda}_{jr}^2}}$
	Cutpoints ($k \geq 2$)	$\bar{\tau}_{jk} = \frac{\tau_{jk}}{\sqrt{\sum_{r=1}^{q} \lambda_{jr}^2 + 1}}$	$\tau_{jk} = \frac{\bar{\tau}_{j,k}}{\sqrt{1 - \sum_{r=1}^{q} \bar{\lambda}_{jr}^2}}$
	Direct effects	$\bar{a}_{jc} = \frac{a_{jc}}{\sqrt{\sum_{r=1}^{q} \lambda_{jr}^2 + 1}}$	$a_{jc} = \frac{\bar{a}_{jc}}{\sqrt{1 - \sum_{r=1}^{q} \bar{\lambda}_{jr}^2}}$
	Error variance	$\bar{\sigma}_j^2 = \frac{1}{\sum_{r=1}^{q} \lambda_{jr}^2 + 1}$	$\sigma_j^2 = 1$
Metric indicators	*Factor loadings*	$\bar{\lambda}_{jr} = \frac{\lambda_{jr}}{\sqrt{\sum_{r=1}^{q} \lambda_{jr}^2 + \sigma_j^2}}$	not possible due to lack of information about variance
	Intercepts	$\bar{\lambda}_{j0} = \frac{\lambda_{j0}}{\sqrt{\sum_{r=1}^{q} \lambda_{jr}^2 + \sigma_j^2}}$	
	Cutpoints ($k \geq 2$)	$\bar{\tau}_{j,k} = \frac{\tau_{jk}}{\sqrt{\sum_{r=1}^{q} \lambda_{jr}^2 + \sigma_j^2}}$	
	Direct effects	$\bar{a}_{jc} = \frac{a_{jc}}{\sqrt{\sum_{r=1}^{q} \lambda_{jr}^2 + \sigma_j^2}}$	
	Error variance	$\bar{\sigma}_j^2 = \frac{\sigma_j^2}{\sum_{r=1}^{q} \lambda_{jr}^2 + \sigma_j^2}$	

Table 3.2: *Formulas for the conversion of model parameters from the Bayesian solution to the standardized solution, and vice versa.*

3.1.5 Exploratory versus confirmatory factor analysis

LVM without covariates have been traditionally used in two different settings, called the exploratory and the confirmatory factor analysis.

In the exploratory setting, researchers do not have much information about the latent variables and the data structure at the beginning of his analysis. They are confronted with a high number of indicators and want to identify the latent variables so that the correlation structure between the indicators is best explained. In this context, identifying the appropriate number of latent variables (e. g. Lopes and West, 2004) is a key issue and the number of latent factors is typically higher than one. After the analysis, the strengths of relationships between latent variables and indicators are known. Very often this solution can be rotated in a suitable way to allow for convenient interpretation of the latent variables (see Bartholomew, 1987). In this sense, the results of an exploratory factor analysis enable the researcher to generate hypotheses which can be tested by other multivariate methods.

In the confirmatory setting however, researchers have knowledge about the factorial nature of the variables so they are able to tell which indicators depend on which latent variables, for example the latent construct might be defined by indicators well-known in the literature.

Typically, each manifest variable solely depends on one latent factor so that there is only one non-zero entry in each row of the factor loadings matrix $\boldsymbol{\Lambda}$. It can even be argued that it is better to create submodels with one single latent variable including the relevant manifest variables if it is a priori known which sets of manifest variables are associated to which latent variables.

We clearly see our model in the confirmatory setting. After having generated ideas and notions about the interdependence structure between latent variables and manifest variables – e. g. with the help of an exploratory factor analysis, or with a carefully selected set of manifest variables targeted to describe a specific latent factor – our model can be used to analyze the effect of covariates on the latent variable like in a standard regression model. This is the reason why most analyses conducted in this thesis incorporate one single latent factor. Although our model is also suited for performing exploratory analysis, we see its power and usefulness in the detailed analysis of covariate effects on the latent variable as explained in the next paragraph.

3.2 The structural equation

As described in the last section, the latent variables are distributed $\boldsymbol{z}_i \sim N_q(\boldsymbol{0}, \boldsymbol{I}_q)$ in models excluding covariates. Now we allow covariates or indirect effects to modify the latent variables by introducing the structural equation part of the model, i. e.

$$\boldsymbol{z}_i = \boldsymbol{\eta}_i + \boldsymbol{\xi}_i \,,$$

with $\boldsymbol{\xi}_i = \boldsymbol{z}_i - \boldsymbol{\eta}_i | \boldsymbol{\eta}_i \sim N_q(\boldsymbol{0}, \boldsymbol{I}_q)$, a linear predictor vector $\boldsymbol{\eta}_i$, and $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$ being stochastically independent. In the literature (e. g. Moustaki, Jöreskog and Mavridis, 2004; Sammel and Ryan, 1996), the linear predictor $\boldsymbol{\eta}_i$ is always of the form

$$\boldsymbol{\eta}_i = \boldsymbol{\gamma} \boldsymbol{u}_i \,,$$

where $\boldsymbol{\gamma}$ is a $q \times m$ matrix of regression coefficients, and $\boldsymbol{u}_i$ is a $m \times 1$ vector of fixed covariates of observation i which are summarized in the $n \times m$ matrix $\boldsymbol{U} = \{u_{il}\}$. These parametric effects imply that the means of the latent variables are linearly dependent on the covariates $\boldsymbol{u}_i$ which is a severe restriction for most real-life research settings:

- For continuous covariates, the assumption of a strictly linear effect on the predictor may not be appropriate. Additionally, effects of continuous covariates may vary for different subgroups of the population.
- Latent variables might be spatially correlated.

To incorporate those issues, we employ a more versatile predictor

$$\eta_{ir} = f_{r1}(x_{i1}) + \ldots + f_{rg}(x_{ig}) + \boldsymbol{\gamma}_r' \boldsymbol{u}_i \,, \tag{3.8}$$

where g denotes the number of different nonparametric functions f, the x_{ih} $(1 \leq h \leq g)$ denote different covariates, f_{rh} are functions of the covariates belonging to latent variable r, and $\boldsymbol{\gamma}_r$ is the vector of values in the r-th row of the $(q \times m)$-dimensional matrix $\boldsymbol{\gamma}$ of standard regression coefficients. We recognize that a separate function per covariate has to be estimated for each of the latent variables. We assume the structure of the linear predictor to be equal for all latent variables[3] r, leading to

$$\boldsymbol{\eta}_i = \boldsymbol{f}_1(x_{i1}) + \ldots + \boldsymbol{f}_g(x_{ig}) + \boldsymbol{\gamma}\boldsymbol{u}_i\,, \tag{3.9}$$

where $\boldsymbol{f}_h$ are now q-dimensional vector valued functions. In component notation, the predictor of Equation (3.9) yields

$$\boldsymbol{\eta}_i = \begin{pmatrix} \eta_{i1} \\ \eta_{i2} \\ \vdots \\ \eta_{iq} \end{pmatrix} = \begin{pmatrix} f_{11}(x_{i1}) \\ f_{21}(x_{i1}) \\ \vdots \\ f_{q1}(x_{i1}) \end{pmatrix} + \ldots + \begin{pmatrix} f_{1g}(x_{ig}) \\ f_{2g}(x_{ig}) \\ \vdots \\ f_{qg}(x_{ig}) \end{pmatrix} + \begin{pmatrix} \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1m} \\ \gamma_{21} & \gamma_{22} & \cdots & \gamma_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{q1} & \gamma_{q2} & \cdots & \gamma_{qm} \end{pmatrix} \cdot \begin{pmatrix} u_{i1} \\ u_{i2} \\ \vdots \\ u_{im} \end{pmatrix}.$$

For each function f_{rh}, a comparably large number of parameters d_h have to be estimated. Let $\boldsymbol{\beta}_{rh} = (\beta_{rh,1}, \beta_{rh,2}, \ldots, \beta_{rh,d_h})'$ denote the vector of function parameters of function f_{rh}. The vectors of function evaluations $\boldsymbol{\beta}_{rh}$ allow a feasible notation of the vector $\boldsymbol{\eta}_r$ which contains the predictor values of all observations i in the following way:

$$\boldsymbol{\eta}_r = (\eta_{1r}, \eta_{2r}, \ldots, \eta_{nr})' = \boldsymbol{X}_1\boldsymbol{\beta}_{r1} + \ldots + \boldsymbol{X}_g\boldsymbol{\beta}_{rg} + \boldsymbol{U}\boldsymbol{\gamma}_r\,, \tag{3.10}$$

with suitably defined $(n \times d_h)$-dimensional design matrices $\boldsymbol{X}_h$ whose entries depend on the type of function and modeling approach. The most compact form of the predictor $\boldsymbol{\eta}$ is then obtained by

$$\boldsymbol{\eta} = \begin{pmatrix} \eta_{11} & \cdots & \eta_{1q} \\ \vdots & \cdots & \vdots \\ \eta_{n1} & \cdots & \eta_{nq} \end{pmatrix} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \ldots + \boldsymbol{X}_g\boldsymbol{\beta}_g + \boldsymbol{U}\boldsymbol{\gamma},$$

where $(d_h \times q)$-dimensional $\boldsymbol{\beta}_h$ contains all vectors of function evaluations $\boldsymbol{\beta}_{rh}$ according to $\boldsymbol{\beta}_h = (\boldsymbol{\beta}_{1h}, \boldsymbol{\beta}_{2h}, \ldots, \boldsymbol{\beta}_{qh})$. Besides one smoothing parameter κ_{rh} has to be estimated for each function f_{rh} which is further explained in Section 5.2.2. In total, the model estimates $q \cdot g$ functions plus $q \cdot m$ regression coefficients plus $q \cdot g$ smoothing parameters, adding up to $q \cdot (\sum_{h=1}^{g} d_h + m + g)$ parameters. The functions considered in our approach are classified according to their respective model name in the literature:

- *Additive models (AM)*
  An AM is obtained if the covariates x_g are univariate and continuous and the $\boldsymbol{f}_g$ are smooth functions. The AM can be considered as the special case of a generalized

[3] In general, a different structure of the predictor might be used for each latent variable.

additive model (GAM) with normally distributed errors. More about the class of GAM is written in Hastie and Tibshirani (1990). In our approach, the smooth functions $\boldsymbol{f}_g$ can be modeled as both random walks and P-splines as explained in Section 5.2.2. A typical example of an additive effect is the nonlinear influence of the covariate age on the latent variables.

- *Geoadditive models*
  A geoadditive model (see Kammann and Wand, 2003) is obtained when the covariate x_{ig} denotes the link of observation i to a certain geographical region. In our approach, the smooth spatial effect $\boldsymbol{f}_g$ is modeled by Markov random fields (Besag, York and Mollie, 1991). A common example would be the analysis of a geographical effect on the latent variable when the federal state for each observed person i is known.

- *Varying coefficient models (VCM)*
  A VCM as proposed by Hastie and Tibshirani (1993) is obtained when all functions are of the form

  $$f_{rh}(x_{ih}) = f_{rh}(\tilde{x}_{ih}, v_{ih}) = g_{rh}(\tilde{x}_{ih})v_{ih}\,,$$

  where the effect modifiers $\tilde{x}_{ih}$ are continuous covariates, and the interacting variables v_{ih} are continuous or categorical. For our purposes, we restrict the interacting variables to be categorical. The continuous functions g_{rh} are also modeled by random walks and P-splines. For example, we could analyze the effect of the continuous covariate age $\tilde{x}_{i1}$ on the latent variables for the different genders v_{ih} as effect modifier.

All these three types of functions plus parametric effects can be used in our LVM in any combination and any number, although it is generally useful to employ only one single geographical effect. Hence we obtain a predictor which embodies a wide range of models, and can be viewed as an extract of the structured additive predictor (STAR) proclaimed by Fahrmeir, Kneib and Lang (2004).

It has to be noted that there is no constant intercept allowed in the predictor, and thus in its parametric part $\boldsymbol{\gamma u}_i$ due to identification restrictions. This can be readily seen by assuming a simple model with one latent factor z_i, no direct effects, an intercept γ_0 in the predictor of the structural equations and no further indirect effects. This leads to

$$\mathrm{E}(y_{ij}^*) = \lambda_{j0} + \lambda_{j1}\gamma_0\,.$$

Accordingly, the intercept γ_0 in the structural equation cannot be estimated independently from the intercepts λ_{j0} in the measurement model because adding a constant c_3 to γ_0 can be offset by reducing λ_{j0} with $c_3\lambda_{j1}$. This is also the reason why in the classic factor analysis model without covariates the means of the latent variables are not estimated but fixed to zero. Another consequence of this restriction is the necessity to centre all functions f_{rh} around zero.

The issue of standardization (see Section 3.1.4) has no influence on the parameter estimates of the structural part of the model because model formulation, parameter values and assumptions are equal in the standardized and Bayesian approaches.

3.3 Summary of model formulation

Since a LVM including covariates is very complex with a high number of parameters, a concise view of the model including variable and parameter names is presented in this section. All manifest variables, covariates and parameters are outlined in Table 3.3.

- **Modeling of ordinal indicators**
  Postulation of underlying variables y^*_{ij} linked to the actual response y_{ij} according to

  $$y_{ij} = k \quad \Longleftrightarrow \quad \tau_{j,k-1} < y^*_{ij} \leq \tau_{jk}\,, \tag{3.11}$$

  and cumulative probability statement with probit link, i. e.

  $$P(y_{ij} \leq k|\mu_{ij}) = \mathbf{\Phi}(\tau_{jk} - \mu_{ij})\,.$$

- **Measurement model including direct effects**

  $$\boldsymbol{y}^*_i = \boldsymbol{\lambda}_0 + \boldsymbol{\Lambda}\boldsymbol{z}_i + \boldsymbol{A}\boldsymbol{w}_i + \boldsymbol{\varepsilon}_i\,,$$

  with $\boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \mathbf{\Sigma})$ and $\mathbf{\Sigma} = \mathrm{diag}(1, \ldots, 1, \sigma^2_{p_1+1}, \ldots, \sigma^2_p)$ where σ^2_j is restricted to 1 for all p_1 ordinal indicators. In the form of individual components the measurement model results in

  $$y^*_{ij} = \lambda_{j0} + \lambda_{j1}z_{i1} + \ldots + \lambda_{jq}z_{iq} + a_{j1}w_{i1} + \ldots + a_{jd}w_{id} + \varepsilon_{ij}\,. \tag{3.12}$$

- **Structural equation model including indirect effects**

  $$\boldsymbol{z}_i = \boldsymbol{\eta}_i + \boldsymbol{\xi}_i = \boldsymbol{f}_1(x_{i1}) + \ldots + \boldsymbol{f}_g(x_{ig}) + \boldsymbol{u}_i\boldsymbol{\gamma}, +\boldsymbol{\xi}_i\,, \tag{3.13}$$

  with $\boldsymbol{\xi}_i \sim N_q(\mathbf{0}, \boldsymbol{I}_q)$. Each single component is given by

  $$z_{ir} = \eta_{ir} + \xi_{ir} = f_{r1}(x_{i1}) + \ldots + f_{rg}(x_{ig}) + \gamma_{r1}u_{i1} + \ldots + \gamma_{rm}u_{im} + \xi_{ir}\,.$$

- **Vector of parameters to be estimated**
  All parameters are arranged in the parameter vector $\boldsymbol{\theta}$ as follows:

  $$\boldsymbol{\theta} = \mathrm{vec}\{\boldsymbol{\lambda}_0, \boldsymbol{\Lambda}, \boldsymbol{A}, \mathbf{\Sigma}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\tau}\}\,. \tag{3.14}$$

  Parameters which are fixed to a certain value, e. g. the first cutpoints of ordinal indicators, are excluded from this vector.

Variable name	**Source**	**Part of model**	**Description**	**Running index**
Response				
n	D	M, S	Number of observations	$i = 1, \ldots, n$
p	D, R	M	Total number of indicators	$j = 1, \ldots, p$
p_1	D	M	Number of ordinal indicators	$j = 1, \ldots, p_1$
p_2	D	M	Number of metric indicators	$j = p_1 + 1, \ldots, p$
$\boldsymbol{y} = \{y_{ij}\}$	D	M	Observed response	
Ordinal indicators				
K_j	D	M	Number of categories	$j = 1, \ldots, p_1$
$\boldsymbol{\tau} = \{\tau_{jk}\}$	E	M	Cutpoints (values fixed for $k = \{0, 1, K_j\}$)	$j = 1, \ldots, p_1$ $k = 0, \ldots, K_j$
$\boldsymbol{y^*} = \{y^*_{ij}\}$	E	M	Underlying variables	
Direct parametric effects				
d	D, R	M	Number of par. direct effects	$c = 1, \ldots, d$
$\boldsymbol{A} = \{a_{jc}\}$	E	M	Direct regression coefficients	
$\boldsymbol{W} = \{w_{ic}\}$	D	M	Direct covariates	
Intercept and error variances				
$\boldsymbol{\lambda}_0 = \{\lambda_{j0}\}$	E	M	Intercepts	
$V(\boldsymbol{\varepsilon}_i) = \boldsymbol{\Sigma} = \{\sigma^2_j\}$	fixed, E	M	Error variances	
$\boldsymbol{\Phi} = \{\phi_{jj}\}$	E	M	Total variance	
Latent variables				
q	D, R	M, S	Number of latent variables	$r = 1, \ldots, q$
$\boldsymbol{z} = \{z_{ir}\}$	E	M, S	Latent variables	
$\boldsymbol{\Lambda} = \{\lambda_{jr}\}$	E	M	Factor loadings matrix	
$V(\boldsymbol{\xi}_i) = \boldsymbol{\Psi} = \boldsymbol{I}_q$	fixed	S	Error variances of latent var.	
Indirect nonparametric effects				
g	R	S	Number of nonpar. effects	$h = 1, \ldots, g$
$\boldsymbol{x} = \{x_{ih}\}$	D	S	Covariates for nonpar. effects	
$\boldsymbol{\beta}_{rh}$	E	S	Evaluation of nonpar. function	
d_h	D, R	S	No. of function parameters	$c_h = 1, \ldots, d_h$
$\boldsymbol{X}_h$	D, R	S	Design matrix for function h	
κ_{rh}	E	S	Smoothing parameter	
Indirect parametric effects				
m	D, R	S	Number of par. indirect effects	$l = 1, \ldots, m$
$\boldsymbol{\gamma} = \{\gamma_{rl}\}$	E	S	Matrix of regression coefficients	
$\boldsymbol{U} = \{u_{il}\}$	D	S	Indirect par. covariates	

Table 3.3: *Overview of variables and parameters of the LVM. The column "Source" indicates whether the quantity is obtained by the estimation process (E), originates from the data (D), or is chosen deliberately by the researcher (R). The column "Part of model" denotes if the quantity is used in the measurement model (M), or in the structural equation (S).*

One might wonder why the measurement model does not include a semiparametric predictor but the structural equation does. The restriction of a parametric measurement model is deliberate due to several reasons. First of all the main targets of interest of the applied researcher are the latent factors and the latent scores. Therefore we introduced a semiparametric predictor in the structural equation in order to conduct more detailed analyses on the dependence structure of the latent factors and indirect covariates. Direct effects, however, are generally of secondary interest to the researcher, and are probably not included in many analyses in the first place. The reason for that lies in the fact that the dependence structure among the indicators is supposed to be explained primarily by the latent factors, and only secondary by the direct effects. If this was not the case one might just conduct p univariate semiparametric regressions with the indicators as response variables, and a latent variable model would be unnecessary. Furthermore, if smooth functions of covariates were included in the measurement model, the number of parameters would rise remarkably because one function would have to be estimated for each indicator and direct covariate, and the number of indicators is always much higher than the number of latent factors. This high number of parameters would require a high number of observations in order to be able to estimate all smooth functions in the measurement model with narrow confidence intervals. However, the inclusion of a semiparametric predictor in the measurement model is conceptionally possible and straightforward, and might be useful for specific analyses.

Chapter 4

Bayesian Inference

In this chapter we want to give an introduction to the Bayesian methodology and the simulation methods employed for the estimation of Bayesian models, i. e. Markov chain Monte Carlo (MCMC) algorithms. Our account is kept brief, and we frequently refer to the abundant literature in these areas.

4.1 Basics of Bayesian methodology

A comprehensive treatment of Bayesian methodology is provided by Gelman et al. (2004). More introductory textbooks about this topic include Lee (2004) and Berry (1996). A thorough theoretical account is given in Bernardo and Smith (1994). A primer on Bayesian methodology within the context of psychometrics is written by Rupp, Dey and Zumbo (2004).

In the classic frequentist paradigm, unknown parameters in a statistical model (e. g. regression coefficients, underlying variables of ordinal responses) are treated as unknown but fixed quantities. From a Bayesian perspective, all unknown population parameters are considered as random variables which follow a certain probability distribution given the data. Let us assume our model requires the estimation of t unknown parameters or random variables which are summarized in the vector

$$\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_t)\,,$$

and the observed data $\boldsymbol{y}$ consists of n observations

$$\boldsymbol{y} = (y_1, y_2, \ldots, y_n)\,.$$

The target of interest is the probability distribution of the unknown parameters given the data $p(\boldsymbol{\theta}|\boldsymbol{y})$, also called the posterior distribution. Applying the Bayes theorem for vector-valued random variables, we obtain

$$p(\boldsymbol{\theta}|\boldsymbol{y}) = \frac{p(\boldsymbol{\theta}, \boldsymbol{y})}{p(\boldsymbol{y})} = \frac{p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\boldsymbol{y})} = \frac{p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{\int p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}}. \tag{4.1}$$

The value of the denominator in Equation (4.1) is constant because the denominator does not depend on $\boldsymbol{\theta}$. Therefore, the Bayes theorem is often stated in the following abbreviated form, i. e.

$$p(\boldsymbol{\theta}|\boldsymbol{y}) \propto p(\boldsymbol{\theta})p(\boldsymbol{y}|\boldsymbol{\theta}). \tag{4.2}$$

The first term in the product $p(\boldsymbol{\theta})$ is called the prior distribution of the unknown parameters $\boldsymbol{\theta}$, and contains our prior beliefs about the distribution of $\boldsymbol{\theta}$. The right term in the product of Equation (4.2) states the probability distribution of the data conditional on fixed parameters $\boldsymbol{\theta}$. We can also view $p(\boldsymbol{y}|\boldsymbol{\theta})$ the other way round, hence consider it as a function of $\boldsymbol{\theta}$ where the data $\boldsymbol{y}$ is obtained for given $\boldsymbol{\theta}$. In this context $p(\boldsymbol{y}|\boldsymbol{\theta})$ is called the likelihood function as in the frequentist setting and can be written as

$$l(\boldsymbol{\theta}|\boldsymbol{y}) = p(\boldsymbol{y}|\boldsymbol{\theta}).$$

This leads to the probably most memorable expression of Bayes theorem:

$$\text{posterior} \propto \text{prior} \times \text{likelihood}.$$

Distributions are called proper if they integrate to a finite value, otherwise they are improper. Intriguingly it turns out that proper posterior densities are often obtained when improper priors are combined with ordinary likelihoods based on the data. Nevertheless, this property is not guaranteed, and it is always wise to check the propriety of the posterior when improper priors are employed. A common reproach of Bayesian critics states that Bayesian estimation is not unbiased because two different researchers might include different prior information into the same model, and thus obtain different parameter estimates. This reasoning is not valid because one could use prior distributions which contain no or very little information about the parameters $\boldsymbol{\theta}$, and therefore do not influence the posterior distribution. Such reference prior distributions are termed noninformative, diffuse or vague (e. g. Jeffreys' prior; see Jeffreys, 1961). In these cases the likelihood solely determines the posterior distribution, and the Bayesian parameter estimates equal the maximum likelihood estimates.

A prior distribution is called conjugate with respect to the likelihood if the posterior distribution follows the same parametric form as the prior distribution. For a likelihood belonging to the exponential family (Fahrmeir and Tutz, 2001), a conjugate prior distribution always exists. Very often the parameterization of the prior distribution is given – for example taken from a conjugate family – but the exact form determined by the parameter values of the corresponding parameterization is not known. Such parameters are called hyperparameters which also have to be estimated in a fully Bayesian approach. Models including hyperparameters are called hierarchical models that are explained in full detail by Gelman et al. (2004).

In a Bayesian analysis, the full posterior distribution of $p(\boldsymbol{\theta}|\boldsymbol{y})$ can be conveniently described by point summaries such as the mean, median, mode and variance of the parameters $\theta_1, \ldots, \theta_t$. Additionally, we define the central posterior interval or credible region $I_{cred} := [I_{left}, I_{right}]$ of a parameter θ_p to be that region of the posterior which contains $100(1-\alpha)\%$ percent of the probability mass, and $100 \cdot \alpha/2\%$ of the probability mass is located to the left and right of that region, respectively. This definition yields the following mathematical equations

$$\int_{I_{left}}^{I_{right}} p(\theta_p|\boldsymbol{y})\, d\theta_p = 1-\alpha\,, \quad \text{and} \quad \int_{-\infty}^{I_{left}} p(\theta_p|\boldsymbol{y})\, d\theta_p = \int_{I_{right}}^{\infty} p(\theta_p|\boldsymbol{y})\, d\theta_p = \frac{\alpha}{2}\,,$$

ergo the boundaries of the central posterior interval are equal to the posterior $\alpha/2$ and $1-\alpha/2$ quantiles. Accordingly, the Bayesian approach allows statements about the probability of a specific parameter lying in a certain region. Such statements are by no means possible in the frequentist paradigm where hypotheses have to be formulated instead to identify confidence intervals, in which a specific parameter is situated in e. g. 95 cases out of 100 repeated experiments. In most research settings, it is not possible to repeat experiments, and furthermore the confidence interval of the frequentist approach does not allow the intuitive understanding that a parameter is located in a certain region with a certain probability. The Bayesian approach eliminates these serious drawbacks of the frequentist paradigm, and makes the often unhelpful and unintuitive classic hypothesis testing obsolete.

In order to compare different models and to identify the best-fitting model, Spiegelhalter, Best, Carlin and van der Linde (2002) introduced the deviance information criterion (DIC). The DIC is defined as

$$DIC = D(\overline{\boldsymbol{\theta}}) + 2p_D = \overline{D(\boldsymbol{\theta})} + p_D\,, \tag{4.3}$$

where $D(\overline{\boldsymbol{\theta}})$ is the deviance of the model evaluated at the posterior mean estimate $\overline{\boldsymbol{\theta}}$, $\overline{D(\boldsymbol{\theta})}$ is the posterior mean of the deviance, and $p_D = \overline{D(\boldsymbol{\theta})} - D(\overline{\boldsymbol{\theta}})$ is the effective number of parameters. We use the unstandardized deviance $D(\boldsymbol{\theta}) = -2 \cdot$log-likelihood. All quantities can be calculated easily in a Markov chain Monte Carlo analysis as described in the next section. For LVM it is still argued in the statistical community whether the DIC is an adequate and a useful measure for model comparison – nevertheless we incorporate the DIC in order to check the results for its applicability.

Next the discussion is continued with standard algorithms employed to obtain the posterior density of Equations (4.1) and (4.2).

4.2 Markov Chain Monte Carlo (MCMC)

The Bayesian approach has been rarely used by applied statisticians until the early 1990s, and the main reason for that was the practical insolubility of the posterior density in

Equation (4.1), and its multidimensional integral in the denominator in particular. Before the early 1990s, Bayesian estimation methods relied on numerical optimization routines such as the EM algorithm, Gauss-Hermite quadrature or the Newton-Raphson method which were even more cumbersome than the numerical estimation procedures for frequentist analyses. Bayesian models became much more popular when the statistical community became aware of the tool of MCMC by the publication of Gelfand and Smith (1990), and increasing computer power necessary for MCMC facilitated the dissemination of Bayesian modeling approaches. The fundamental difference between classic numerical techniques and the new simulation based approaches is that the latter rely heavily on the generation of random numbers. An exhaustive account on Monte Carlo methods including MCMC and random number generation is provided by Robert and Casella (2004), while Gilks, Richardson and Spiegelhalter (1996) focus on various applications of MCMC.

The next sections discuss the very basics of Monte Carlo integration and Markov chains, and present the most prominent MCMC algorithms: the Metropolis-Hastings algorithm and the Gibbs sampler. We conclude with the description of a method how to improve sampling convergence which proves to be useful for our algorithm set out in the next chapter.

4.2.1 Classic Monte Carlo integration

Simulation-based Monte Carlo integration solves the fundamental problem to evaluate an integral of the form

$$\mathrm{E}_f[h(X)] = \int_{\mathcal{X}} h(x) f(x) dx \,, \tag{4.4}$$

where X is a random variable with density f, $\mathcal{X}$ denotes the probability space, and $h(X)$ is an arbitrary function of X. This integral can be approximated by generating a random sample $x^{(1)}, \ldots, x^{(M)}$ from the density f and calculating the empirical average

$$\bar{h}_M = \frac{1}{M} \sum_{m=1}^{M} h(x^{(m)}) \,,$$

since $\bar{h}_M$ converges almost surely to $\mathrm{E}_f[h(X)]$ by the strong law of large numbers (see Breiman, 1992). Furthermore, when h^2 has a finite expectation under f and M is large,

$$\frac{\bar{h}_M - \mathrm{E}_f[h(X)]}{\sqrt{v_M}} \tag{4.5}$$

converges to the standard normal distribution N(0,1), with $v_M = M^{-2} \sum_{m=1}^{M} [h(x^{(m)}) - \bar{h}_M]^2$ as the sample variance. Hence Equation (4.5) can be utilized for the construction of confidence bounds on the approximation of $\mathrm{E}_f[h(X)]$.

In Bayesian inference, many posterior quantities of interest follow the form of Equation (4.4), with $f(x)$ equaling the posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{y})$. For example, the posterior mean of the parameter θ_p is then obtained by using $h(\boldsymbol{\theta}) = \theta_p$, i. e.

$$\mathrm{E}(\theta_p) = \int \theta_p \, p(\boldsymbol{\theta}|\boldsymbol{y}) \, d\boldsymbol{\theta} \,,$$

which will be approximated by

$$\bar{\theta_p}^M = \frac{1}{M} \sum_{m=1}^{M} \theta_p^{(m)} \,, \tag{4.6}$$

where $\theta_t^{(1)}, \ldots, \theta_t^{(M)}$ is a random sample from the posterior distribution of $p(\boldsymbol{\theta}|\boldsymbol{y})$. The main difficulty consists in generating a random sample of the posterior distribution. For this purpose we make use of the properties of Markov chains as described in the next section.

4.2.2 Markov chains

We want to give a very short description of the essential properties of Markov chains and how they are used in the context of MCMC. Markov chains are the simplest mathematical models for random processes evolving in time. Their simple structure makes it possible to reveal a lot about their properties. Nevertheless the class of Markov chains is rich enough to serve in many applications, for example in the field of MCMC. For a thorough theoretical account we refer to Tierney (1994) or Robert and Casella (2004).

Let us assume there is a state space Ω, and the Markov chain jumps from one state to another with a certain probability. The random variable or random vector $\boldsymbol{X}_i$ holds the state of the Markov chain at time i. The most important property of a Markov chain is that is has no memory of where it has been in the past. Hence a Markov chain is fully characterized by the transition kernel which gives the probability of jumping from the state $\boldsymbol{x} \in \Omega$ to a state region $A \subset \Omega$, i. e.

$$P(\boldsymbol{x}, A) = P(\boldsymbol{X}_{i+1} \in A | \boldsymbol{X}_i = \boldsymbol{x}) \,.$$

It follows that the probability of jumping to a state region A only depends on the current state $\boldsymbol{x}$ of the Markov chain. In the Bayesian setting, the state space corresponds to a probability space with the dimension of the posterior density.

If certain conditions are fulfilled by a Markov chain[1] – and they usually are in the MCMC setting by construction – the Markov chain possesses an invariant or stationary distribution π which satisfies

$$\pi(d\boldsymbol{y}) = \int_\Omega P(\boldsymbol{x}, d\boldsymbol{y}) \pi(\boldsymbol{x}) \, d\boldsymbol{x} \,.$$

This states that if $\boldsymbol{X}_i$ is distributed according to the invariant distribution π, $\boldsymbol{X}_{i+1}$ is also distributed like π. If a Markov chain has an invariant distribution, the random variables

[1]Essentially, the Markov chain has to satisfy the irreducibility condition. For more details see Robert and Casella (2004).

$\boldsymbol{X}_i$ of a Markov chain are distributed according to the stationary distribution π as i goes to infinity. This also implies that the first realizations of a Markov chain are not distributed according to π.

In classic Markov chain theory, you typically start with a transition kernel, and the invariant distribution has to be derived from that kernel. In MCMC however, the procedure is the other way round: you want to obtain a Markov chain with the posterior distribution of the Bayesian model as the invariant distribution – the transition kernel just has to be suitably constructed in such a way. This can be ensured by setting up a transition kernel which fulfills the detailed balance condition. This condition states that the probability of being in a state $\boldsymbol{x}$ and jumping to a state $\boldsymbol{y}$ is equal to the probability of being in the state $\boldsymbol{y}$ and jumping to $\boldsymbol{x}$ for all $\boldsymbol{x}, \boldsymbol{y} \in \Omega$. It can be shown that the probability distribution $f(\boldsymbol{x})$ of being in the state $\boldsymbol{x}$ equals the invariant distribution $\pi(\boldsymbol{x})$ if the detailed balance condition holds. Hence the transition kernel is constructed in such a way that the detailed balance condition holds for the posterior density as the invariant distribution. The well-known MCMC algorithms – the Metropolis-Hastings algorithm and the Gibbs sampler – automatically fulfill the detailed balance condition by definition, so that this property has not to be checked before their application.

After having constructed a transition kernel of a Markov chain $\{\boldsymbol{X}_m\}$ with the invariant distribution π, it follows that

$$\frac{1}{M}\sum_{m=1}^{M} h(\boldsymbol{X_m}) \longrightarrow \int h(\boldsymbol{x})\pi(\boldsymbol{x})\,d\boldsymbol{x} \text{ , as } M \to \infty \, , \tag{4.7}$$

which links Markov chains to Monte Carlo methods. This and more convergence results on Markov chains are proven in Robert and Casella (2004). Now we can easily calculate posterior quantities using Equation (4.7) because the invariant distribution π is equal to the posterior density $p(\boldsymbol{\theta}|\boldsymbol{y})$. The next two subsections describe the most common MCMC algorithms which both ensure that the invariant distribution of the underlying Markov chain is identical to the posterior density.

4.2.3 Metropolis-Hastings algorithm

The fundamental idea of the Metropolis-Hastings (MH) algorithm can be traced back to the physicists Metropolis et al. (1953), and was generalized by Hastings (1970). The MH algorithm enables the generation of random samples of a multivariate distribution which is typically not tractable by numerical or analytical integration. In the Bayesian setting, the multivariate distribution corresponds to the posterior density. The MH algorithm draws random values from a suitable proposal density; afterwards those random values are accepted with a certain probability – the MH acceptance probability – in such a way that the detailed balance condition holds. Therefore, the invariant distribution of the resulting Markov chain equals the target density. Common choices for proposal densities include random walk proposals, and fixed proposals that do not depend on the current state of the Markov chain.

Let $p(\boldsymbol{\theta})$ be the parameter vector's target posterior density[2]. Beginning with the starting value $\boldsymbol{\theta}^{(0)}$, we construct a Markov chain $\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \boldsymbol{\theta}^{(3)}, \ldots, \boldsymbol{\theta}^{(M)}$ of length M. After having converged to the stationary distribution, the random draws of the Markov chain can be considered to be distributed according to the posterior density $p(\boldsymbol{\theta})$. The MH algorithm reads as follows:

The Metropolis-Hastings algorithm

1. Choose a starting value $\boldsymbol{\theta}^{(0)}$.
2. Repeat for $m = 0, 1, \ldots, M-1$
   - Draw a proposal value $\widetilde{\boldsymbol{\theta}}$ from the proposal density $q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(m)})$.
   - Set the new value of the Markov chain

$$\boldsymbol{\theta}^{(m+1)} = \begin{cases} \widetilde{\boldsymbol{\theta}} & \text{with probability} \quad \rho(\boldsymbol{\theta}^{(m)}, \widetilde{\boldsymbol{\theta}}) \\ \boldsymbol{\theta}^{(m)} & \text{with probability} \quad 1 - \rho(\boldsymbol{\theta}^{(m)}, \widetilde{\boldsymbol{\theta}}) \end{cases}$$

   with the MH acceptance probability

$$\rho(\boldsymbol{\theta}^{(m)}, \widetilde{\boldsymbol{\theta}}) = \min\left\{ \frac{p(\widetilde{\boldsymbol{\theta}})}{p(\boldsymbol{\theta}^{(m)})} \frac{q(\boldsymbol{\theta}^{(m)}|\widetilde{\boldsymbol{\theta}})}{q(\widetilde{\boldsymbol{\theta}}|\boldsymbol{\theta}^{(m)})}, 1 \right\} . \tag{4.8}$$

   - Return the values $\{\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \ldots, \boldsymbol{\theta}^{(M)}\}$.

We recognize that the MH acceptance probability (4.8) contains the ratio of the posterior densities $p(\boldsymbol{\theta})$, hence we can drop the constant denominator in (4.1), and Equation (4.2) suffices for generating random samples of the posterior distribution. For that reason the difficult calculation of the normalizing constant in (4.1) is not necessary. If the proposal density is symmetric, i. e. $q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(m)}) = q(\boldsymbol{\theta}^{(m)}|\boldsymbol{\theta})$, the MH acceptance probability simplifies to

$$\rho(\boldsymbol{\theta}^{(m)}, \widetilde{\boldsymbol{\theta}}) = \min\left\{ \frac{p(\widetilde{\boldsymbol{\theta}})}{p(\boldsymbol{\theta}^{(m)})}, 1 \right\} . \tag{4.9}$$

This special case in (4.9) is called Metropolis algorithm, named after its inventor Metropolis (1953), and was generalized by Hastings (1970) for non-symmetric proposal densities according to Equation (4.8).

Although it is theoretically possible to generate random samples of arbitrary multivariate densities using the MH algorithm, it is practically very difficult or even impossible to find suitable and efficient proposal densities for high-dimensional multivariate target distributions. For that reason it is recommended to partition the parameter vector $\boldsymbol{\theta}$ into B

[2] To obtain a shorter account, the abbreviation $p(\boldsymbol{\theta})$ instead of $p(\boldsymbol{\theta}|\boldsymbol{y})$ is used.

disjoint parameter blocks $\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_B$. Let $\boldsymbol{\theta}_{-b}$ denote the parameter vector without the parameter block $\boldsymbol{\theta}_b$, i. e. $\boldsymbol{\theta}_{-b} := (\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_{b-1}, \boldsymbol{\theta}_{b+1}, \dots, \boldsymbol{\theta}_B)$. Furthermore, $p_{b|-b}(\boldsymbol{\theta}_b|\boldsymbol{\theta}_{-b})$ denotes the conditional density of the block b given all other blocks. Then we can apply the MH algorithm on each of the B blocks sequentially which also leads to the posterior distribution as the invariant distribution (see Robert and Casella, 2004).

Multiple-block Metropolis-Hastings algorithm

1. Choose a starting value $\boldsymbol{\theta}^{(0)} = (\boldsymbol{\theta}_1^{(0)}, \dots, \boldsymbol{\theta}_B^{(0)})$.

2. Repeat for $m = 0, 1, \dots, M-1$
   Repeat for $b = 1, 2, \dots, B$

   - Draw a proposal value $\widetilde{\boldsymbol{\theta}}_b$ from the proposal density $q_b(\boldsymbol{\theta}|\boldsymbol{\theta}_b^{(m)}, \boldsymbol{\theta}_{-b}^{(m+1)})$ with
   $$\boldsymbol{\theta}_{-b}^{(m+1)} := (\boldsymbol{\theta}_1^{(m+1)}, \dots, \boldsymbol{\theta}_{b-1}^{(m+1)}, \boldsymbol{\theta}_{b+1}^{(m)}, \dots, \boldsymbol{\theta}_B^{(m)}).$$
   - Set the new value of the Markov chain
   $$\boldsymbol{\theta}_b^{(m+1)} = \begin{cases} \widetilde{\boldsymbol{\theta}}_b & \text{with probability} \quad \rho(\boldsymbol{\theta}_b^{(m)}, \widetilde{\boldsymbol{\theta}}_b) \\ \boldsymbol{\theta}_b^{(m)} & \text{with probability} \quad 1 - \rho(\boldsymbol{\theta}_b^{(m)}, \widetilde{\boldsymbol{\theta}}_b) \end{cases}$$
   with the MH acceptance probability
   $$\rho(\boldsymbol{\theta}_b^{(m)}, \widetilde{\boldsymbol{\theta}}_b) = \min\left\{ \frac{p_{b|-b}(\widetilde{\boldsymbol{\theta}}_b|\boldsymbol{\theta}_{-b}^{(m+1)})}{p_{b|-b}(\boldsymbol{\theta}_b^{(m)}|\boldsymbol{\theta}_{-b}^{(m+1)})} \frac{q(\boldsymbol{\theta}_b^{(m)}|\widetilde{\boldsymbol{\theta}}_b, \boldsymbol{\theta}_{-b}^{(m+1)})}{q(\widetilde{\boldsymbol{\theta}}_b|\boldsymbol{\theta}_b^{(m)}, \boldsymbol{\theta}_{-b}^{(m+1)})}, 1 \right\}. \tag{4.10}$$

3. Return the values $\{\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \dots, \boldsymbol{\theta}^{(M)}\}$.

4.2.4 Gibbs sampler

The Gibbs sampler is obtained as a special case of the multiple-block MH algorithm when the proposal densities q_b are chosen to be the full conditional distributions $p_{b|-b}(\boldsymbol{\theta}_b|\boldsymbol{\theta}_{-b})$. In this case the MH acceptance probability in Equation (4.10) equals 1 at all times because the terms in the ratio cancel out. Hence the Gibbs sampler is structured the following way:

Gibbs sampler

1. Choose a starting value $\boldsymbol{\theta}^{(0)} = (\boldsymbol{\theta}_1^{(0)}, \dots, \boldsymbol{\theta}_B^{(0)})$.

2. Repeat for $m = 0, 1, \dots, M-1$

   - Draw $\boldsymbol{\theta}_1^{(m+1)} = p_{1|-1}(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2^{(m)}, \dots, \boldsymbol{\theta}_B^{(m)})$,
   - Draw $\boldsymbol{\theta}_2^{(m+1)} = p_{2|-2}(\boldsymbol{\theta}_2|\boldsymbol{\theta}_1^{(m+1)}, \boldsymbol{\theta}_3^{(m)}, \dots, \boldsymbol{\theta}_B^{(m)})$,

   $\vdots$

- Draw $\boldsymbol{\theta}_b^{(m+1)} = p_{b|-b}(\boldsymbol{\theta}_b|\boldsymbol{\theta}_1^{(m+1)}, \ldots, \boldsymbol{\theta}_{b-1}^{(m+1)}, \boldsymbol{\theta}_{b+1}^{(m)}, \ldots, \boldsymbol{\theta}_B^{(m)})$,

  $\vdots$
- Draw $\boldsymbol{\theta}_B^{(m+1)} = p_{B|-B}(\boldsymbol{\theta}_B|\boldsymbol{\theta}_1^{(m+1)}, \ldots, \boldsymbol{\theta}_{B-1}^{(m+1)})$.

3. Return the values $\{\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \ldots, \boldsymbol{\theta}^{(M)}\}$.

In practice, the Gibbs sampler is often preferred to the MH algorithm due to its easier structure without the need of acceptance probabilities which have to be adjusted carefully before each simulation run in order to obtain an efficient sampling of the parameters.

4.2.5 Sampler convergence and its improvement

In practice the MCMC sampler is run for a certain number of iterations. The first samples of a MCMC run – the so-called burnin phase – are discarded, and after the burnin phase the Markov chain is supposed to have converged and the random samples are considered to be drawn from the posterior distribution. The main question arising is when convergence has occurred. This question is not answered easily, and unfortunately there is no single tool which can confirm or reject the assumption of convergence for all MCMC samplers. One good indicator of convergence is the amount of autocorrelations of subsequent random samples of a single parameter. Typically low to medium autocorrelations are not a problem since the samples still belong to the invariant distribution. High autocorrelations, however, can cause inefficient parameter estimates and might imply that convergence has not occurred. One possibility of dealing with moderate autocorrelations is thinning out the random samples so that only a fraction of all samples is retained for the calculation of posterior quantities. A thinning parameter of 5, for example, means that only the parameter samples of every fifth iteration are stored and used for the posterior analysis. Furthermore, there exists a number of convergence diagnostic algorithms, such as the Raftery and Lewis' convergence rate estimator or spectral methods. An overview of convergence diagnostic methods is provided by Cowles and Carlin (1996), and Brooks and Roberts (1998).

Standard MCMC algorithms like the Gibbs sampler tend to exhibit slow convergence properties, especially for high-dimensional parameter spaces. Several suggestions have been made by the statistical community to improve MCMC convergence in such cases.

Instead of sampling from one- or low-dimensional full conditionals, block updates could be used where several low-dimensional components of the Gibbs sampler are merged into a high-dimensional parameter block. Especially the convergence of highly correlated parameters profits from this procedure, as shown by Gilks, Richardson and Spiegelhalter (1996), and Roberts and Sahu (1997). Often blocking is not a viable option because sampling of high-dimensional parameter spaces would require a MH step with a suitable proposal density which is difficult to find.

Another choice is the reparametrization of highly-correlated parameter blocks in order to reduce the correlations between those blocks. This strategy is pursued by the concept of hierarchical centering introduced by Gelfand, Sahu and Carlin (1995; 1996) which is especially useful for mixed models. Another method is the approximate orthogonalisation of parameters which unfortunately can rarely be done for high-dimensional posterior densities.

Other convergence improvement measures include the collapsed Gibbs sampler (Liu, 1994) which employs sampling from partially marginalised distributions, Langevin diffusion MCMC (Roberts and Tweedie, 1996), Slice sampling (Neal, 2003), improved hybrid MC with leapfrog algorithm (see Neal, 1994; Duane et al., 1987) and simulated tempering.

The reason why convergence improvement measures are discussed here is because the standard sampling method for ordinal regression introduced by Albert and Chib (1993) exhibits mediocre convergence properties for the cutpoints. Two improved algorithms which solve this problem have been contributed by Cowles (1996) who makes use of the idea of collapsed sampling, and Nandram and Chen (1996) who additionally include a reparameterization of the cutpoints.

In this thesis, we employ and examine three different samplers: the standard algorithm of Albert and Chib (1993), the well established Cowles algorithm (1996), and a new method proposed by Liu and Sabatti (2000) which is based on the following theorem:

Theorem 1 *(Liu and Sabatti, 2000)*

Let Γ be a locally compact group of transformations on the sample space $\boldsymbol{S}$, L be its left-Haar measure, and $\boldsymbol{\rho} \in \boldsymbol{S}$ follows a distribution with density $\pi(\boldsymbol{\rho})$. If $\gamma \in \Gamma$ is drawn from

$$p_x(\gamma)\, d\gamma \sim \pi(\gamma(\boldsymbol{\rho}))\, |J_\gamma(\boldsymbol{\rho})|\, L(d\gamma)\,,$$

where $J_\gamma(\boldsymbol{\rho}) = det(\partial\gamma(\boldsymbol{\rho})/\partial\boldsymbol{\rho})$ is the Jacobian of the transformation, then $\boldsymbol{x'} = \gamma(\boldsymbol{\rho})$ also has the density π.

Here a set $\Gamma = \{\gamma\}$ of transformations on $\boldsymbol{S}$ is called a locally compact group if

- Γ is a locally compact space;
- The elements in Γ form a group with respect to the composition operation;
- The group operations $(\gamma_1, \gamma_2) \to \gamma_1\gamma_2$ and $\gamma \to \gamma^{-1}$ are continuous.

A measure L is called a left-Haar invariant measure if $L(B) = L(\gamma B)$ for every γ and for all measurable sets $B \in \Gamma$. More details about transformation groups, compact spaces and left-Haar measures can be found in Rao (1987). The transformation group later employed is the scale group which is defined as

$$\Gamma = \left\{ \gamma \in \mathbb{R}^1 : \gamma(\boldsymbol{\rho}) = \gamma\boldsymbol{\rho} = (\gamma\rho_1, \ldots, \gamma\rho_d) \right\} ,$$

with the left-Haar measure $L(d\gamma) = \gamma^{-1}d\gamma$. This leads to the sampling distribution of a transformation element $p_x(\gamma) \propto |\gamma|^{d-1}\pi(\gamma\boldsymbol{\rho})$. Other possible choices of transformation groups are the translation group, the affine transformation group, and the orthonormal transformation group.

In the MCMC context, the theorem of Liu and Sabatti (2000) can be used by applying the transformation γ to a group $\boldsymbol{\rho}$ of parameters. Since the transformation γ maintains the distribution of $\boldsymbol{\rho}$, i. e. $\pi(\gamma(\boldsymbol{\rho})) = \pi(\boldsymbol{\rho})$, the stationary posterior distribution is not altered by this procedure. A Gibbs sampler with such a transformation is called a Generalized Gibbs sampler (GGS) with a Generalized Gibbs move $\boldsymbol{\rho} \to \gamma(\boldsymbol{\rho})$. Whether a GGS shows a better convergence than its parent MCMC algorithm depends crucially on the choice of transformation group Γ, and the form of the posterior distribution π. Clearly the slowly converging parameters should be transformed by the GGM in order to improve convergence, and sampling from the distribution $\pi(\gamma(\boldsymbol{\rho}))\,|J_\gamma(\boldsymbol{\rho})|\,L(d\gamma)$ should be straightforward and fast.

Note that in practice it is often difficult to find a suitable transformation group which both improves convergence and allows efficient sampling of the transformation members γ. However, this method has shown to enhance the convergence of cutpoint parameters in ordinal probit models with an underlying variable (e. g. see Liu and Sabatti, 2000).

Chapter 5

Bayesian formulation of the LVM

This chapter gives a full account on how random samples of the posterior distribution are obtained. Firstly, we introduce the Bayesian model setup, continue with the specification of prior distributions for all parameters, and conclude with the full formula of the posterior distribution including the likelihood function. After that the corresponding MCMC algorithms are formulated. We present three different MCMC algorithms which essentially differ in the way of estimating the cutpoints of ordinal indicators.

5.1 Bayesian model setup

The posterior distribution is obtained via Bayes formula in Equation (4.2). All variables are naturally divided into given data and parameters that are to be estimated. In our model the given data includes indicators $\boldsymbol{y}_i$, direct effects $\boldsymbol{w}_i$, indirect nonparametric effects $\boldsymbol{x}_i$, and indirect parametric effects $\boldsymbol{u}_i$ for $i = 1, \ldots, n$. The parameter vector that has to be estimated stems from Equation (3.14) and yields $\boldsymbol{\theta} = \text{vec}\{\boldsymbol{\lambda}_0, \boldsymbol{\Lambda}, \boldsymbol{A}, \boldsymbol{\Sigma}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\tau}\}$. Hence the posterior distribution leads to

$$p(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{w}, \boldsymbol{x}, \boldsymbol{u}) \propto p(\boldsymbol{\theta})\, p(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{w}, \boldsymbol{x}, \boldsymbol{u})\,.$$

It turns out that the resulting posterior distribution cannot be estimated by a convenient MCMC algorithm because the likelihood part of the posterior contains high-dimensional integrals which cannot be sampled in a neat and efficient way. This problem can be resolved by extending the parameter space with nonobservable data; in our model this relates to the underlying variables $\boldsymbol{y}^*$ for ordinal indicators, and the latent variables $\boldsymbol{z}$. This is a common approach in Bayesian methodology, called data augmentation, which was introduced by Tanner and Wong (1987). Albert and Chib (1993) implemented this approach for the estimation of regression parameters and cutpoints for ordinal response. Consequently, the complete parameter vector is obtained by adding the underlying variables and latent variables to the parameter vector $\boldsymbol{\theta}$, i. e. $\{\boldsymbol{\theta}, \boldsymbol{y}^*, \boldsymbol{z}\}$, leading to the posterior distribution

$$p(\boldsymbol{\theta}, \boldsymbol{y}^*, \boldsymbol{z}|\boldsymbol{y}, \boldsymbol{w}, \boldsymbol{x}, \boldsymbol{u}) \propto p(\boldsymbol{\theta})\, p(\boldsymbol{y}, \boldsymbol{y}^*, \boldsymbol{z}|\boldsymbol{\theta}, \boldsymbol{w}, \boldsymbol{x}, \boldsymbol{u})\,. \tag{5.1}$$

The augmented posterior distribution is not just a mere technicality to enable efficient and easy sampling, it also empowers the interpretation of latent variables because the values of all latent variables z_{ir} are automatically estimated. For example, this enables to rank observations according to their respective latent variable value, and statements can be made about the probability of observation i_1 having a higher latent variable value than observation i_2. This property is an important advantage of the Bayesian approach compared to the frequentist approach. The underlying variables y^*_{ij} are also automatically estimated but usually are of minor importance to the researcher.

After having set up the Bayesian model in Equation (5.1), the next section gives a full description of all prior distributions $p(\boldsymbol{\theta})$.

5.2 Prior distributions

In this section a complete specification of the prior distributions for all model parameters is provided. Since the prior distributions of the underlying variables $\boldsymbol{y}^*$ and the latent variables $\boldsymbol{z}$ are implicitly determined by the prior distributions of all other parameters and the distributional assumptions about $\boldsymbol{\varepsilon}_i$ and $\boldsymbol{\xi}_i$, we only have to specify prior distributions for the parameter vector $\boldsymbol{\theta}$. From now on, let us assume that the individual parts of the model are stochastically independent. This is not a severe restriction because the researcher generally does not have prior information about the relationships between parameter sets of different parts of the model, and in any case highly diffuse priors are chosen whenever possible. Thus the prior distribution yields

$$p(\boldsymbol{\theta}) = p(\boldsymbol{\lambda}_0, \boldsymbol{\Lambda}, \boldsymbol{A}) \cdot p(\boldsymbol{\Sigma}) \cdot p(\boldsymbol{\beta}, \gamma) \cdot p(\boldsymbol{\tau})\,. \tag{5.2}$$

The following two sections treat the prior distributions $p(\boldsymbol{\lambda}_0, \boldsymbol{\Lambda}, \boldsymbol{A})$, $p(\boldsymbol{\Sigma})$ and $p(\boldsymbol{\tau})$ of the measurement model, and $p(\boldsymbol{\beta}, \boldsymbol{\gamma})$ of the structural equation model, respectively. The focus of the account lies on the prior distributions of the nonparametric indirect effects because they define the form of the nonparametric functions $\boldsymbol{f}$.

5.2.1 Prior distributions of the measurement model

Prior distribution of intercepts, factor loadings and direct effects $\boldsymbol{p}(\boldsymbol{\lambda}_0, \boldsymbol{\Lambda}, \boldsymbol{A})$

We define a $(p \cdot (1 + q + d) \times 1)$-dimensional vector $\bar{\boldsymbol{\lambda}}$ which contains all parameters of $\boldsymbol{\lambda}_0$, $\boldsymbol{\Lambda}$ and $\boldsymbol{A}$ arranged as follows:

$$\begin{aligned} \bar{\boldsymbol{\lambda}} := (&\lambda_{10}, \lambda_{11}, \ldots, \lambda_{1q}, a_{11}, \ldots, a_{1d}, \\ &\lambda_{20}, \lambda_{21}, \ldots, \lambda_{2q}, a_{21}, \ldots, a_{2d}, \\ &\quad\vdots \qquad\qquad \vdots \qquad\qquad \vdots \\ &\lambda_{p0}, \lambda_{p1}, \ldots, \lambda_{pq}, a_{p1}, \ldots, a_{pd}) \,. \end{aligned} \tag{5.3}$$

The prior distribution selected for $\bar{\boldsymbol{\lambda}}$ is a $p \cdot (1 + q + d)$-dimensional multivariate normal density with the mean $\bar{\boldsymbol{\lambda}}^*$ and the precision matrix $\bar{\boldsymbol{\Lambda}}^*$ which are chosen by the researcher according to his prior information about the parameters, i. e.

$$\bar{\boldsymbol{\lambda}} \sim N(\bar{\boldsymbol{\lambda}}^*, \bar{\boldsymbol{\Lambda}}^{*-1}) \,.$$

If no prior information is available, a noninformative prior is used which corresponds to a precision matrix with all values set to zero. Hence $p(\bar{\boldsymbol{\lambda}}) \propto \textit{constant}$, and the prior distribution is improper. In order to compare Bayesian parameter estimates with frequentist estimates, it is recommended to use noninformative priors because then the posterior solely depends on the likelihood part, and both parameter estimates coincide.

For our purposes we choose noninformative priors for the intercepts $\boldsymbol{\lambda}_0$ and the regression coefficients $\boldsymbol{A}$ of the direct effects. However, we are forced to include prior information for the factor loadings for ordinal indicators in order to prevent the occurrence of Heywood cases in the Bayesian setting. A Heywood case appears when one factor loads up completely on one (sometimes even more) indicator(s), hence the latent variable accounts for the full variability of the respective indicator, and the corresponding communality equals 1. Since this result is highly implausible, we choose informative priors with a normal density centered at zero with a certain variance. A standard choice in applications (Lopes and West, 2004; Quinn, 2004) is a prior variance of one because this prevents the occurrence of Heywood cases, is highly diffuse and therefore allows to obtain high factor loadings. Since the number of occurrences of Heywood cases depends on the MCMC algorithm used, we use three different prior settings, weak, standard and strong (see Table 5.1) in our simulation studies to analyze this effect. Hence our prior precision matrix $\bar{\boldsymbol{\Lambda}}^*$ equals zero for the off-diagonal elements, and the diagonal elements also equal zero except for the factor loadings λ_{jr} which are set to one of the values 0.5 (weak prior), 1.0 (standard prior) or 4.0 (strong prior). More information about Heywood cases can be found in appendix A.

Prior settings for factor loadings		
Prior strength	**Value on diagonal in precision matrix $\bar{\mathbf{\Lambda}}^*$**	**Standard deviation of factor loadings**
Weak	0.50	$\sqrt{2} \approx 1.41$
Standard	1.00	1.00
Strong	4.00	0.50

Table 5.1: *Three different prior strengths are used for the factor loading parameters in the simulation studies. All models based on a real data set in Chapter 7 employ the standard prior setting.*

Prior distributions of error variances $p(\mathbf{\Sigma})$

For ordinal indicators, error variances are not estimated but fixed to one, as described in Section 3.1.2 due to identification restrictions. For continuous indicators however, error variances have to be estimated. Since the error variance matrix $\mathbf{\Sigma}$ is diagonal, its prior distribution can be given by specifying the individual prior distributions of σ_j^2. The standard conjugate prior choice for error variances in a linear model with normally distributed errors is the inverse Gamma distribution, hence

$$\sigma_j^2 \sim IG^1\left(\frac{\nu}{2}, \frac{\nu s^2}{2}\right) \text{ , for } j = p_1 + 1, \ldots, p\,,$$

with two hyperparameters; the degrees of freedom $\nu > 0$, and the scale parameter $s > 0$ (see Gelman et al., 2004). The prior density yields

$$p(\sigma_j^2) \propto (\sigma_j^2)^{-(\nu/2+1)} \exp(-\nu s^2/(2\sigma_j^2)) \text{ , } \sigma_j^2 > 0\,.$$

A noninformative prior distribution is obtained for $\nu \longrightarrow 0$ and $\nu s^2 \longrightarrow 0$, and results in the improper prior distribution

$$p(\sigma_j^2) \propto \frac{1}{\sigma_j^2} \quad \text{with} \quad \sigma_j^2 \in [0; \infty[.$$

We avoid the use of improper priors for the error variances σ_j^2 to prevent Heywood cases. Hyperparameters ν and s should be chosen in such a way to correctly include prior information if available. If noninformative priors are to be used, the prior distributions should have a vanishing probability for $\sigma_j^2 \longrightarrow 0$ to prevent Heywood cases.

Prior distributions of cutpoints for ordinal indicators $p(\boldsymbol{\tau})$

We choose noninformative, diffuse prior distributions for the cutpoints $\boldsymbol{\tau}$ which have to satisfy the order condition

$$0 < \tau_{j2} < \tau_{j3} < \ldots < \tau_{j,K_j-1} < \infty \qquad \text{, for } j = 1, \ldots, p\,.$$

[1]This corresponds to a scaled inverse-chi-square distribution with $\sigma_j^2 \sim \text{Inv-}\chi^2(\nu, s^2)$.

5.2.2 Prior distributions of the structural equation

In this section, priors for the nonparametric function parameters $\boldsymbol{\beta}$ and for the parametric effects $\boldsymbol{\gamma}$ are specified. We assume the independence of prior specifications between separate functions and parametric effects, and between functions and parametric effects of different latent variables; thus we obtain

$$p(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \prod_{r=1}^{q} \prod_{h=1}^{g} p(\boldsymbol{\beta}_{rh}) \cdot \prod_{r=1}^{q} p(\boldsymbol{\gamma}_r) \,.$$

Priors for metric covariates (additive models, varying coefficient models) are based on Gaussian smoothness priors (see Fahrmeir and Tutz, 2001), and priors for spatial covariates (geoadditive models) are based on Markov random fields[2] (see Besag, 1974; Besag and Kooperberg, 1995). Conveniently, in the Bayesian approach both types of covariates can be treated in a unifying framework involving the use of a penalty matrix $\boldsymbol{K}$.

Nonparametric effects of metric covariates

In this thesis, nonparametric effects for metric covariates can be modeled in three different ways: first-order random walks, second-order random walks, and P-splines. In order to simplify the notation in this section, we drop the indices of most of the variables, hence f denotes the nonparametric function, $\boldsymbol{\beta}$ is the vector of function parameters, x represents the metric covariate, and d denotes the dimension of the vector of function parameters.

1. *First-order random walk*
   First we consider the case of a metric covariate x with equally spaced observations $x_{(t)}$, $t = 1, \ldots, d$, $d \leq n$. The unique observations $x_{(t)}$ are sorted according to $x_{(1)}, \ldots, x_{(t)}, \ldots, x_{(d)}$, and thus define an equidistant grid on the x-axis. A classic example is the covariate age, ranging from the age of 20 to 80 in a social survey, hence $d = 61$. Let us set $\beta_t := f(x_{(t)})$ and let

   $$\boldsymbol{\beta} = (\beta_1, \ldots, \beta_t, \ldots, \beta_d)'$$

   denote the vector of function evaluations according to Section 3.2. The first-order random walk is defined as

   $$\beta_t = \beta_{t-1} + u_t \quad \text{with} \quad u_t \sim N(0, \kappa^2) \,,$$

   $t = 2, \ldots, d$ and a diffuse prior $\beta_1 \propto$ *constant*. The parameter β_t is determined by the previous value β_{t-1} plus a normally distributed random error u_t with mean 0 and variance κ^2, i. e. $\beta_t | \beta_{t-1}, \kappa^2 \sim N(\beta_{t-1}, \kappa^2)$. The expected value of β_t coincides

[2] Other spatial modeling methods might be included in the future, e. g. two-dimensional P-splines.

with the expected value of β_{t-1}, and consequently this prior specification penalizes value differences between two successive observations. The entire prior distribution of a function f with a vector of function parameters $\boldsymbol{\beta}$ for a first-order random walk follows as

$$p(\boldsymbol{\beta}) = \prod_{t=2}^{d} p(\beta_t|\beta_{t-1},\kappa^2) \propto \exp(-\frac{1}{2\kappa^2}\sum_{t=2}^{d}(\beta_t-\beta_{t-1})^2) \\ = \exp(-\frac{1}{2\kappa^2}\boldsymbol{\beta}'\boldsymbol{K}\boldsymbol{\beta}),$$

with the penalty matrix

$$\boldsymbol{K} = \begin{pmatrix} 1 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 1 \end{pmatrix}.$$

The generalization for non-equidistant observations is straightforward and is detailed in appendix B.

After parameters have been estimated, the function evaluations of all observations i are given by $\boldsymbol{X\beta}$ with the $(n \times d)$-dimensional design matrix $\boldsymbol{X}$. Each row of $\boldsymbol{X}$ contains the value 1 in that column number corresponding to the respective observation, all other columns in that row are zero. To continue the example from above, we observe ages from 20 to 80 and thus the design matrix $\boldsymbol{X}$ has 61 columns; if observation i has age 40, the i-th row of $\boldsymbol{X}$ contains a one in the 21st column, and zeros in all other columns.

2. *Second-order random walk*
   Definitions and notation is identical to the case of the first-order random walk. The second-order random walk is defined as

$$\beta_t = 2\beta_{t-1} - \beta_{t-2} + u_t \quad \text{with} \quad u_t \sim N(0,\kappa^2), \tag{5.4}$$

   $t = 3,\ldots,d$, and diffuse priors $\beta_1 \propto$ *constant* and $\beta_2 \propto$ *constant*. The parameter β_t is determined by the doubled previous value $2\beta_{t-1}$ minus the value β_{t-2} plus a normally distributed random error u_t with mean 0 and variance κ^2, i. e. $\beta_t|\beta_{t-1},\beta_{t-2},\kappa^2 \sim N(2\beta_{t-1}-\beta_{t-2},\kappa^2)$. Since the expected value of β_t can be interpreted as the extrapolated value of the straight line through β_{t-1} and β_{t-2}, the second-order random walk prior penalizes deviations from the linear trend. Typically, the second-order random walk generates visually smoother functions f than the first-order random walk. The entire prior distribution of a function f with second-order random walk follows as

$$p(\boldsymbol{\beta}) = \prod_{t=3}^{d} p(\beta_t|\beta_{t-1}, \beta_{t-2}, \kappa^2) \propto \exp(-\frac{1}{2\kappa^2}\sum_{t=3}^{d}(\beta_t - 2\beta_{t-1} + \beta_{t-2})^2)$$
$$= \exp(-\frac{1}{2\kappa^2}\boldsymbol{\beta}'\boldsymbol{K}\boldsymbol{\beta}),$$

with the penalty matrix

$$\boldsymbol{K} = \begin{pmatrix} 1 & -2 & 1 & & & & \\ -2 & 5 & -4 & 1 & & & \\ 1 & -4 & 6 & -4 & 1 & & \\ & \ddots & \ddots & \ddots & \ddots & \ddots & \\ & & 1 & -4 & 6 & -4 & 1 \\ & & & 1 & -4 & 5 & -2 \\ & & & & 1 & -2 & 1 \end{pmatrix}.$$

The generalization for non-equidistant observations is also detailed in appendix B. The design matrix $\boldsymbol{X}$ for second-order random walks is constructed in the same way as the design matrix of first-order random walks.

3. *P-splines*

Both random walks require the estimation of one parameter for each unique observation $x_{(t)}$, leading to a high number of parameters and occasionally to an overfit of the data. For that reasons, other ways of estimating smooth functions exist in the literature, see Fahrmeir and Tutz (2001). One method which became quite popular in recent years for smoothing in semiparametric models is the approach of penalized splines (P-splines). In the statistical community two different types of P-splines are basically discussed. The first approach stems from the work of Eilers and Marx (1996) who promote the use of a B-spline basis, equally-spaced knots and difference penalties, whereas Ruppert, Wand and Carroll (2003) employ truncated power functions, knots based on quantiles of the covariate of interest x and a ridge penalty. Eilers and Marx (2004) compared the two approaches along several dimensions (e. g. numerical stability, quality of fit) and concluded that the first approach is to be preferred. For that reason, we restrict our discussion to B-splines with difference penalties and equally spaced knots. The literature mentioned above covers the frequentist treatment of P-splines whereas our Bayesian approach is based on the work of Lang and Brezger (2004) and Brezger and Lang (2005), who give a detailed account on Bayesian P-splines in various settings.

The unknown function f of a metric covariate x is approximated by a polynomial spline of degree D, defined on a set of equally spaced knots $x^{min} = \varrho_0 < \varrho_1 < \ldots < \varrho_{I-1} < \varrho_I = x^{max}$ with I intervals. This polynomial spline is constructed by a linear combination of $d = D + I$ B-spline basis functions B_c in the following way:

$$f(x) = \sum_{c=1}^{d} \beta_c B_c(x).$$

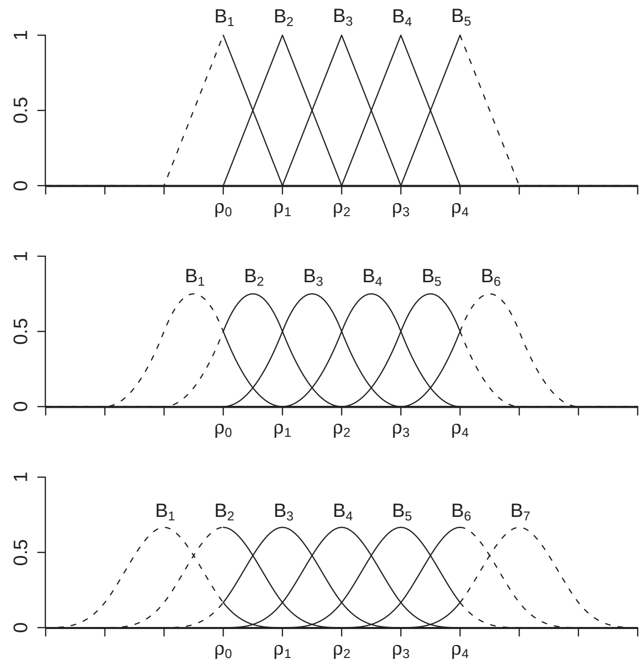


Figure 5.1: *Illustration of B-splines basis functions with $D = 1$ (top), $D = 2$ (middle) and $D = 3$ (bottom) for $I = 4$ intervals and $I + 1 = 5$ knots $\kappa_0, \ldots, \kappa_4$.*

The vector of function parameters now contains the regression coefficients or weights of the individual B-spline basis functions, i. e. $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_{D+I})'$. Note that $\boldsymbol{\beta}$ does not contain function evaluations as is the case for random walk models. A B-spline basis function has the following characteristic properties:

- A B-spline of degree D consists of $D+1$ polynomial pieces connected at D inner knots;
- Continuous derivatives to the order $D-1$ exist at the knots;
- A B-spline covers $D+2$ knots or $D+1$ regions between knots, and overlaps with $2D$ adjacent B-splines;
- At each point on the covariate axis (apart from the knots), $D+1$ B-splines have a non-zero value.

More information about B-splines can be found in the mentioned literature. Figure 5.1 shows two examples of B-spline basis functions for the degrees $D = \{1, 2, 3\}$, and $I = 4$ intervals with $I + 1 = 5$ knots.

The design matrix $\boldsymbol{X}$ for P-splines is more intricate than in the case of random walk priors. Each row i of $\boldsymbol{X}$ contains the values of the B-spline basis functions evaluated at x_i, hence $X_{ic} = B_c(x_i)$. In accordance with the fourth property of B-splines, each row in X has $D+1$ non-zero values. Thus the vector of function evaluations for all observations i is given by $\boldsymbol{X\beta}$.

A crucial question is the determination of the number of knots. The number of knots should be high enough to adapt to the underlying function f; however, it should not be too high to overfit the data. Eilers and Marx recommend the number of knots to range between 20 and 40, and introduce a penalization of the differences between regression coefficients of adjacent B-spline basis functions in order to generate a smoothing effect. Ergo the smoothness of the function f is achieved through penalizing too high differences of coefficients of adjacent B-splines, but not by altering the number of knots. In a Bayesian approach, this penalization is incorporated conveniently by applying a random walk prior to the B-splines regression coefficients f. In our analyses, we typically choose B-splines of degree $D = 3$ with $I = 10$ intervals, and a second-order random walk prior on the B-splines regression coefficients.

Nonparametric effects of interactions (VCM)

The models discussed so far are not suitable for modeling interactions. As introduced in Section 3.2, in a VCM the function is of the form

$$f(x_i) = f(\tilde{x}_i, v_i) = g(\tilde{x}_i)v_i \,,$$

where the effect modifiers $\tilde{x}_i$ are continuous covariates, and the interacting variables v_i are metric or categorical. We restrict our model to cope with categorical interacting variables. Since the differences between two categories of an ordinal or categorical variable are not interpretable, we apply dummy coding for v (see Fahrmeir and Tutz, 2001). Let us assume that v has K categories, then we define

$$v_i^{(k)} = \begin{cases} 1, & \text{if sample } i \text{ observes category } k \\ 0, & \text{else} \end{cases} \qquad , \; k = 1, \ldots, K \,.$$

The dummy coding implies the estimation of K different functions $f^{(k)}$ with function parameter values $\boldsymbol{\beta}^{(k)}$, so that the total part of the predictor for the function f results in

$$f = f^{(1)} + \ldots + f^{(K)} = X^*\boldsymbol{\beta}^{(1)} + diag(v_1^{(2)}, \ldots, v_n^{(2)})X^*\boldsymbol{\beta}^{(2)} + \ldots + diag(v_1^{(K)}, \ldots, v_n^{(K)})X^*\boldsymbol{\beta}^{(K)}.$$

Here the reference category was set to category 1, but arbitrary reference categories are possible. The design matrix X^* is the usual design matrix associated with the continuous function $g(\tilde{x})$ which can be modeled by either a random walk prior or a P-splines prior. For example, we could identify two separate nonparametric functions of the continuous covariate age for the effect modifier gender with two categories.

Nonparametric effects of spatial covariates

In this section we discuss the prior distribution of spatial covariates. Let us assume covariate x_i denotes the region c of observation i, and the vector of function evaluations

$\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_d)$ contains the estimates of the d different regions. The spatial function evaluations of all observations i can be written as $\boldsymbol{X\beta}$ with the $(n \times d)$-dimensional design matrix $\boldsymbol{X}$, where $X_{ic} = 1$ if observation i is associated to region c; all other values of row i equal zero. It can be quite useful to include a spatial covariate in an analysis in order to examine the geographical variation of the latent variables. Of course, the region itself has typically not a direct effect on the values of the latent variables, but there are certain underlying characteristics of each region which could readily influence their values. For example, the latent variable "satisfaction with living conditions" surely depends on the existence of heavy industry polluting the air or on the local unemployment rate, both of which are varying across regions. The basic assumption is that adjacent regions should have a similar impact on latent values, while two regions far apart from each other do not exhibit such a similarity. In order to make a prior specification, the full neighborhood structure for each region has to be known. In our context, two regions are considered neighbors when they share a common boundary. Other definitions of neighborhood are described by Besag et al. (1991). We apply the following spatial smoothness prior to the function evaluations β_c $(c = 1, \ldots, d)$ for all d regions:

$$\beta_c | \beta_e, e \neq c, \kappa^2 \sim N\left(\sum_{e \in \partial_c} \frac{\beta_e}{N_c}, \frac{\kappa^2}{N_c}\right), \tag{5.5}$$

where N_c indicates the number of adjacent sites of region c, and $e \in \partial_c$ denotes all regions e being neighbors of region c. Hence the conditional mean of β_c is an unweighted average of the function values of all adjacent regions. Since spatial data, e. g. regions, does not inhibit a natural ordering, a symmetric conditioning is applied. A more general prior including Equation (5.5) as a special case is given by

$$\beta_c | \beta_e, e \neq c, \kappa^2 \sim N\left(\sum_{e \in \partial_c} \frac{w_{ce}}{w_{c+}} \beta_e, \frac{\kappa^2}{w_{c+}}\right),$$

where the weights w_{ce} are not necessarily equal, and $w_{c+} = \sum_{e \in \partial_c} w_{ce}$. For example, the weights could depend on the length of the border or distance between two adjacent regions. In our analyses, the specialised prior of Equation (5.5) is always used. The entire prior distribution follows as

$$p(\boldsymbol{\beta}) \propto \exp\left(-\sum_{c=1}^{d} \frac{w_{c+}}{2\kappa^2}\left(\beta_c - \sum_{e \in \partial_c} \frac{w_{ce}}{w_{c+}}\right)^2\right) = \exp(-\frac{1}{2\kappa^2}\boldsymbol{\beta}' K \boldsymbol{\beta}),$$

with the d-dimensional penalty matrix $\boldsymbol{K}$ whose entries are

$$k_{cc} = w_{c+} \qquad \text{and} \qquad k_{ce} = \begin{cases} -w_{ce} & , e \in \partial_c, \\ 0 & , \text{otherwise}. \end{cases}$$

To conclude, we want to emphasize that priors of all nonparametric effects (metric, spatial, and interaction) can me modeled in a unifying framework with $p(\boldsymbol{\beta}) \propto \exp(-\frac{1}{2\kappa^2}\boldsymbol{\beta}' K \boldsymbol{\beta})$ and a suitably defined penalty matrix $\boldsymbol{K}$.

Hyperpriors of variances $\boldsymbol{\kappa}^2$ of nonparametric effects

We have defined all priors for nonparametric functions conditional on the variance κ^2, i. e. $p(\boldsymbol{\beta}) = p(\boldsymbol{\beta}|\kappa^2)p(\kappa^2)$. The variance κ^2 determines the smoothness of the resulting function f, and is therefore called smoothing parameter. It is automatically estimated in our Bayesian approach. To complete the prior specification for nonparametric effects, we define the prior of the hyperparameter κ^2 to be

$$p(\kappa^2) \propto \frac{1}{(\kappa^2)^{a+1}} \exp(-b/\kappa^2)\,,$$

where $a \in \mathbb{R}$ and $b > 0$. If $a > 0$, this expression corresponds to an inverse Gamma distribution IG(a,b). The parameters a and b have to be chosen appropriately. Common choices include $a = b = 0.001$ leading to an almost noninformative prior for κ^2; or $a = 1$ and b equal a very small value, e. g. $b = 0.005$ as proposed by Besag et al. (1995). The choice of such highly vague but proper priors prevent problems associated with noninformative priors, such as the nonconvergence of the Gibbs sampler (see Hobert and Casella, 1996). If a noninformative and improper prior is used for the variances κ^2, the resulting posterior can be improper which is not necessarily indicated by the sampling chains of the Gibbs sampler. In such a case the Gibbs sampler would yield random draws of a nonexistent posterior distribution. All simulation studies and data analyses in this thesis use a vague, but informative prior setting with $a = b = 0.001$.

However, Sun, Tsutakawa and He (2001) pointed out that the posterior can be proper when improper diffuse priors for the variance component κ^2 are chosen if certain conditions hold. There are two types of improper priors that can be employed. The first improper prior is obtained for $a = -1$, $b = 0$ which yields $p(\kappa^2) \propto 1$; the parameters $a = -0.5$, $b = 0$ lead to the second improper prior $p(\kappa^2) \propto (\kappa^2)^{-1/2}$. Since some statisticians argue that the use of highly diffuse but proper priors influence the parameter estimates in a significant way, we check the effect of these two improper priors in Sections 6.3.2 and 6.3.3 where smooth functions of metric and spatial covariates are estimated in simulation studies.

Parametric effects

The conjugate prior distribution of the vector of regression coefficients $\boldsymbol{\gamma}_r$ is a m-dimensional multivariate normal density with the mean $\boldsymbol{\gamma}_r^*$ and the precision matrix $\boldsymbol{\Gamma}_r^*$ which are chosen by the researcher according to his prior information about the parameters, i. e.

$$\boldsymbol{\gamma}_r \sim N(\boldsymbol{\gamma}_r^*, \boldsymbol{\Gamma}_r^{*-1})\,.$$

In our analyses, we always choose noninformative priors for all regression parameters $\boldsymbol{\gamma}_r$, hence all values of $\boldsymbol{\Gamma}_r^*$ are set to zero.

Full prior distribution of the structural equation model

Finally, the full specification of prior distributions for the structural equation is given by

$$p(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \prod_{r=1}^{q}\prod_{h=1}^{g} p(\boldsymbol{\beta}_{rh}) \cdot \prod_{r=1}^{q} p(\boldsymbol{\gamma}_r) \propto \prod_{r=1}^{q}\prod_{h=1}^{g} \exp(-\frac{1}{2\kappa_{rh}^2}\boldsymbol{\beta}_{rh}'\boldsymbol{K}_{rh}\boldsymbol{\beta}_{rh})p(\kappa_{rh}^2) \cdot \prod_{r=1}^{q} p(\boldsymbol{\gamma}_r)\,,$$

where the penalty matrices $\boldsymbol{K}_{rh}$, smoothing parameters κ_{rh}^2, and parametric effects $\boldsymbol{\gamma}_r$ are defined in the last paragraphs.

5.3 Likelihood, posterior distribution and DIC

The posterior distribution is obtained by multiplying the prior distributions with the likelihood, leading to

$$\begin{aligned} p(\boldsymbol{\theta}, \boldsymbol{y}^*, \boldsymbol{z}|\boldsymbol{y}, \boldsymbol{w}, \boldsymbol{x}, \boldsymbol{u}) &\propto p(\boldsymbol{\theta}) \cdot p(\boldsymbol{y}, \boldsymbol{y}^*, \boldsymbol{z}|\boldsymbol{\theta}, \boldsymbol{w}, \boldsymbol{x}, \boldsymbol{u}) \\ &= p(\boldsymbol{\theta}) \cdot p(\boldsymbol{y}^*, \boldsymbol{z}|\boldsymbol{\theta}, \boldsymbol{w}, \boldsymbol{x}, \boldsymbol{u}) \cdot p(\boldsymbol{y}|\boldsymbol{y}^*, \boldsymbol{z}, \boldsymbol{\theta}, \boldsymbol{w}, \boldsymbol{x}, \boldsymbol{u}) \\ &= p(\boldsymbol{\theta}) \cdot p(\boldsymbol{y}^*, \boldsymbol{z}|\boldsymbol{\theta}, \boldsymbol{w}, \boldsymbol{x}, \boldsymbol{u}) \cdot p(\boldsymbol{y}|\boldsymbol{y}^*, \boldsymbol{\tau}). \end{aligned}$$

The simplification from the second to the third row is induced by Equation (3.1) which states that the ordinal response $\boldsymbol{y}_{ij}$ is solely determined by the corresponding underlying variable $\boldsymbol{y}_{ij}^*$ and the cutpoints $\boldsymbol{\tau}_j$. Accordingly, the likelihood splits into two separate parts, the joint distribution of the underlying and latent variables given the parameters and covariates, and the distribution of the actual response given the underlying variables and the cutpoints.

The joint distribution of $\boldsymbol{y}_i^*$ and $\boldsymbol{z}_i$ results from a combination of the measurement model and the structural equation, and is easily obtained by standard statistical calculus for multivariate densities, hence $\boldsymbol{y}_i^*, \boldsymbol{z}_i|\boldsymbol{\theta}, \boldsymbol{w}_i, \boldsymbol{x}_i, \boldsymbol{u}_i$ results in

$$\begin{bmatrix} \boldsymbol{y}_i^* \\ \boldsymbol{z}_i \end{bmatrix} \sim N\left(\begin{bmatrix} \boldsymbol{\lambda_0} + \boldsymbol{A}\boldsymbol{w}_i + \boldsymbol{\Lambda}\boldsymbol{\eta}_i \\ \boldsymbol{\eta}_i \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Lambda}\boldsymbol{\Lambda'} + \boldsymbol{\Sigma} & \boldsymbol{\Lambda} \\ \boldsymbol{\Lambda'} & \boldsymbol{I_{m\times m}} \end{bmatrix} \right)\,, \tag{5.6}$$

where the predictor $\boldsymbol{\eta}_i$ is a function of $\boldsymbol{x}_i$ and $\boldsymbol{u}_i$ according to Equation (3.9)

$$\boldsymbol{\eta}_i = \boldsymbol{f}_1(x_{i1}) + \ldots + \boldsymbol{f}_g(x_{ig}) + \boldsymbol{\gamma}\boldsymbol{u}_i\,.$$

More details about the form of the predictor are outlined in Section 3.2.

The second part of the likelihood is trivial for continuous indicators. The probability density is constant because no underlying variable is necessary, and $\boldsymbol{y}_i^*$ equals the actual response $\boldsymbol{y}_i$. For ordinal indicators however, the response y_{ij} is unambiguously determined by the underlying variable y_{ij}^* and the cutpoints $\boldsymbol{\tau}_j$ through Equation (3.1). For that reason

$p(y_{ij}|y^*_{ij}, \boldsymbol{\tau}_j)$ results in 0 or 1 if viewed as a function of y_{ij}. In mathematical terms this yields

$$p(y_{ij}|y^*_{ij}, \boldsymbol{\tau}_j) = \sum_{k=1}^{K_j} \mathbb{1}_{\tau_{j,k-1}<y^*_{ij}\le\tau_{jk}} \mathbb{1}_{y_{ij}=k} \text{ , for } j = 1, \ldots, p_1 \,,$$

where $\mathbb{1}_x$ denotes the indicator function of x. Since the data y_{ij} is known, this density is viewed as a function of the underlying variable y^*_{ij} which is therefore constrained by an interval bounded by the two corresponding cutpoints. The density for all ordinal indicators j of a single observation i leads to

$$p(\boldsymbol{y}_i|\boldsymbol{y}^*_{ij}, \boldsymbol{\tau}) = \prod_{j=1}^{p_1} \sum_{k=1}^{K_j} \mathbb{1}_{\tau_{j,k-1}<y^*_{ij}\le\tau_{jk}} \mathbb{1}_{y_{ij}=k} \,.$$

After having specified all prior distributions and the two components of the likelihood, we are ready to specify the full posterior distribution. Assuming independently and identically distributed observations, we obtain

$$p(\boldsymbol{\theta}, \boldsymbol{y}^*, \boldsymbol{z}|\boldsymbol{y}, \boldsymbol{w}, \boldsymbol{x}, \boldsymbol{u}) \propto$$
$$p(\boldsymbol{\theta}) \prod_{i=1}^{n} \left[p(\boldsymbol{y}^*_i, \boldsymbol{z}_i|\boldsymbol{\theta}, \boldsymbol{w}_i, \boldsymbol{x}_i, \boldsymbol{u}_i) \times \left(\prod_{j=1}^{p_1} \sum_{k=1}^{K_j} \mathbb{1}_{\tau_{j,k-1}<y^*_{ij}\le\tau_{jk}} \mathbb{1}_{y_{ij}=k} \right) \right] \,,$$

with a prior distribution $p(\boldsymbol{\theta})$ according to Equation (5.2) and $p(\boldsymbol{y}^*_i, \boldsymbol{z}_i|\boldsymbol{\theta}, \boldsymbol{w}_i, \boldsymbol{x}_i, \boldsymbol{u}_i)$ as given in Equation (5.6). Obviously the analytical calculation of this high-dimensional density is impossible, and direct sampling is a difficult task to do and highly inefficient. For that reason we introduce three MCMC algorithms which sample the parameters in a sequential fashion as outlined in Section 4.2. In particular, we will make extensive use of the Gibbs sampler that is well suited to sample from normal densities.

Having defined the likelihood, the deviance information criterion (DIC) defined in Equation (4.3) can be calculated. As mentioned in Section 4.1, the applicability of the DIC for models involving latent variables is still debated in the statistical community. One reason lies in the fact that there exist several possibilities to define the DIC in LVM, as DeIorio and Robert mentioned in the discussion of the paper Spiegelhalter, Best, Carlin and van der Linde (2002) – one version of the DIC might include the latent variables whereas another DIC version excludes them, and different results can be obtained for different DIC versions. The model presented in this thesis further complicates the issue by containing two types of latent or nonobservable variables, i. e. the underlying variables $\boldsymbol{y}^*$ for ordinal indicators and the latent scores $\boldsymbol{z}$. We want to test two different versions of the DIC: one version includes the estimated latent scores $\boldsymbol{z}$, whereas the other version excludes them – both versions do not use the underlying variables $\boldsymbol{y}^*$. Then the first version DIC_1 including the latent scores $\boldsymbol{z}$ yields

$$DIC_1 = D_1(\overline{\boldsymbol{\theta}}, \overline{\boldsymbol{z}}) + 2p_D = \overline{D_1(\boldsymbol{\theta}, \boldsymbol{z})} + p_D \,. \tag{5.7}$$

The deviance of the model evaluated at the posterior mean estimate $\overline{\boldsymbol{\theta}}$ yields

$$\begin{aligned}
D_1(\overline{\boldsymbol{\theta}}, \overline{\boldsymbol{z}}) &= -2 \cdot \log l(\overline{\boldsymbol{\theta}}, \overline{\boldsymbol{z}} | \boldsymbol{y}) = -2 \cdot \log p(\boldsymbol{y} | \overline{\boldsymbol{\theta}}, \overline{\boldsymbol{z}}) \\
&= -2 \cdot \log \prod_{i=1}^{n} \left[\prod_{j=1}^{p_1} \left(\Phi(\overline{\tau}_{j,k_{ij}} - \overline{\mu}_{ij}^1) - \Phi(\overline{\tau}_{j,k_{ij}-1} - \overline{\mu}_{ij}^1) \right) \times \right. \\
&\qquad \left. \prod_{j=p_1+1}^{p} \frac{1}{\sqrt{2\pi}\overline{\sigma}_j} \exp\left(-\frac{(y_{ij} - \overline{\mu}_{ij}^1)^2}{2\overline{\sigma}_j^2} \right) \right],
\end{aligned} \tag{5.8}$$

with k_{ij} equaling the number of the observed category of observation i for indicator j, and the expected value $\overline{\mu}_{ij}^1$ of the measurement model is based on the estimated latent scores $\boldsymbol{z}$, i. e.

$$\overline{\mu}_{ij}^1 = \overline{\lambda}_{j0} + \overline{\lambda}_{j1}\overline{z}_{i1} + \ldots + \overline{\lambda}_{jq}\overline{z}_{iq} + \overline{a}_{j1}w_{i1} + \ldots + \overline{a}_{jd}w_{id}\,. \tag{5.9}$$

The posterior mean of the deviance $\overline{D_1(\boldsymbol{\theta}, \boldsymbol{z})}$ is based on the deviance in each iteration of the MCMC sampler, i. e.

$$\overline{D_1(\boldsymbol{\theta}, \boldsymbol{z})} = \frac{1}{M} \sum_{m=1}^{M} D_1(\boldsymbol{\theta}^{(m)}, \boldsymbol{z}^{(m)})\,,$$

where the deviance $D_1(\boldsymbol{\theta}^{(m)}, \boldsymbol{z}^{(m)})$ in each iteration is calculated in the same way as the deviance of the posterior mean in Equations (5.8) and (5.9) but the posterior mean parameter values are replaced by the respective parameter values in iteration m of the MCMC sampler. Finally the number of effective parameters p_{D_1} can be calculated by $p_{D_1} = \overline{D_1(\boldsymbol{\theta}, \boldsymbol{z})} - D_1(\overline{\boldsymbol{\theta}}, \overline{\boldsymbol{z}})$.

The second version of the DIC replaces the latent scores $\boldsymbol{z}$ with the respective value of the predictor of the structural equation $\boldsymbol{\eta}$, i. e.

$$DIC_2 = D_2(\overline{\boldsymbol{\theta}}) + 2p_D = \overline{D_2(\boldsymbol{\theta})} + p_D\,. \tag{5.10}$$

The deviance is calculated almost identically to the deviance of the model with latent variables according to

$$\begin{aligned}
D_2(\overline{\boldsymbol{\theta}}) &= -2 \cdot \log l(\overline{\boldsymbol{\theta}} | \boldsymbol{y}) = -2 \cdot \log p(\boldsymbol{y} | \overline{\boldsymbol{\theta}}) \\
&= -2 \cdot \log \prod_{i=1}^{n} \left[\prod_{j=1}^{p_1} \left(\Phi(\overline{\tau}_{j,k_{ij}} - \overline{\mu}_{ij}^2) - \Phi(\overline{\tau}_{j,k_{ij}-1} - \overline{\mu}_{ij}^2) \right) \times \right. \\
&\qquad \left. \prod_{j=p_1+1}^{p} \frac{1}{\sqrt{2\pi}\overline{\sigma}_j} \exp\left(-\frac{(y_{ij} - \overline{\mu}_{ij}^2)^2}{2\overline{\sigma}_j^2} \right) \right],
\end{aligned} \tag{5.11}$$

but now the expected value of the measurement model is not based on the estimated latent scores $\boldsymbol{z}$ but on the expected value η_{ij} (see Equation 3.8) of the predictor of the structural equation, i. e.

$$\overline{\mu}_{ij}^2 = \overline{\lambda}_{j0} + \overline{\lambda}_{j1}\overline{\eta}_{i1} + \ldots + \overline{\lambda}_{jq}\overline{\eta}_{iq} + \overline{a}_{j1}w_{i1} + \ldots + \overline{a}_{jd}w_{id}\,. \tag{5.12}$$

The posterior mean of the deviance $\overline{D_2(\boldsymbol{\theta}, \boldsymbol{z})}$ is obtained in the same way as for D_1 above.

The results of both DIC versions for the analyses with one latent variable of the PD1 dataset are presented in Section 7.3.6.

We want to emphasize that it is not the goal of the model selection and the DIC to identify the correct number of latent factors. For classic factor analysis this issue has been intensively discussed in the statistic literature (e. g. Lopes and West, 2004). In our context the number of indicators and latent variables remains fixed, and we want to identify the predictor of covariates that delivers the best fitting model; hence we want to answer the question which indirect covariates should be integrated into the predictor of the structural equation and which covariates ought to be left out.

5.4 MCMC implementation

In this section three different MCMC algorithms for the parameter estimation are presented. The sampling steps performed in each iteration are very similar for all three samplers, but some changes occur at certain steps. The first sampler, called the standard sampler, solely consists of Gibbs sampling steps, and the sampling of ordinal indicators follows the method introduced by Albert and Chib (1993). Unfortunately this sampler shows very bad convergence properties for the cutpoints, and thus for other parameters of the model, too. This issue was discussed by Cowles (1996) who proposed an alternative Metropolis-Hastings step for the sampling of the cutpoints. We implemented this algorithm in our second approach, and therefore call it the MH sampler (MHS). The convergence of the MHS is good but could still be improved; furthermore computational costs are high – therefore we introduce a third algorithm based on the work of Liu and Sabatti (2000) who propose the use of a Generalized Gibbs move in each iteration for all ordinal indicators after all sampling steps of the standard sampler are carried out. An overview of the samplings steps of all three MCMC algorithms is given in Table 5.2. The convergence properties of these samplers are examined in simulation studies described in Chapter 6.1.

Step	Parameters	Draw	No. of drawn parameters	Sampler types: Stand.	MHS	GGS
A	Underlying variables	y^*_{ij}	np_1	•	•	•
B	Latent variables	$\boldsymbol{z}_i$	npq	•	•	•
C	Nonparametric indirect effects	$\boldsymbol{\beta}_{rh}$ κ_{rh}	qgd_h qg	• •	• •	• •
D	Parametric indirect effects	$\boldsymbol{\gamma}_r$	qm	•	•	•
E	Intercepts, direct effects, and factor loadings	$\{\boldsymbol{\lambda}_0, \boldsymbol{A}, \boldsymbol{\Lambda}\}$	$p + pd + pq$ - q(q-1)/2	•	•	•
F	Error variances	σ^2_j	p_2	•	•	•
G1	Cutpoints (Gibbs)	τ_{jk}	$\sum_{j=1}^{p_1} K_j - 2p_1$	•		•
G2	Cutpoints (MH)	τ_{jk}	$\sum_{j=1}^{p_1} K_j - 2p_1$		•	
H	Transformation par.	γ_j	p_1			•

Table 5.2: *Overview of the sampling steps used in each iteration for the three different MCMC algorithms (indicated by black dots).*

5.4.1 Algorithm 1: the standard sampler

The conditional Gibbs steps derived from the posterior density are:

A. Draw the underlying variables from $p(\boldsymbol{y}^*|\boldsymbol{\theta}, \boldsymbol{z}, \boldsymbol{y}, \boldsymbol{w}, \boldsymbol{x}, \boldsymbol{u})$,

B. Draw the latent variables from $p(\boldsymbol{z}|\boldsymbol{\theta}, \boldsymbol{y}^*, \boldsymbol{y}, \boldsymbol{w}, \boldsymbol{x}, \boldsymbol{u})$,

C. Draw the nonparametric indirect effects from $p(\boldsymbol{\beta}|\boldsymbol{\theta} \setminus \{\boldsymbol{\beta}\}, \boldsymbol{y}^*, \boldsymbol{z}, \boldsymbol{y}, \boldsymbol{w}, \boldsymbol{x}, \boldsymbol{u})$,

D. Draw the parametric indirect effects from $p(\boldsymbol{\gamma}|\boldsymbol{\theta} \setminus \{\boldsymbol{\gamma}\}, \boldsymbol{y}^*, \boldsymbol{z}, \boldsymbol{y}, \boldsymbol{w}, \boldsymbol{x}, \boldsymbol{u})$,

E. Draw the intercepts, direct effects and factor loadings from $p(\boldsymbol{\lambda}_0, \boldsymbol{\Lambda}, \boldsymbol{A}|\boldsymbol{\theta} \setminus \{\boldsymbol{\lambda}_0, \boldsymbol{\Lambda}, \boldsymbol{A}\}, \boldsymbol{y}^*, \boldsymbol{z}, \boldsymbol{y}, \boldsymbol{w}, \boldsymbol{x}, \boldsymbol{u})$,

F. Draw the error variances from $p(\boldsymbol{\sigma}^2|\boldsymbol{\theta} \setminus \{\boldsymbol{\sigma}^2\}, \boldsymbol{y}^*, \boldsymbol{z}, \boldsymbol{y}, \boldsymbol{w}, \boldsymbol{x}, \boldsymbol{u})$,

G. Draw the cutpoints from $p(\boldsymbol{\tau}|\boldsymbol{\theta} \setminus \{\boldsymbol{\tau}\}, \boldsymbol{y}^*, \boldsymbol{z}, \boldsymbol{y}, \boldsymbol{w}, \boldsymbol{x}, \boldsymbol{u})$.

The full conditionals are standard distributions such as the multivariate normal density or the uniform density, and can be sampled easily and efficiently as opposed to the full posterior distribution. In the following sections, the full conditional distributions are presented in their unabridged detail.

A. Full conditionals of the underlying variables

For continuous indicators this step can be ommitted because there is no underlying variable y^*_{ij}, hence $y^*_{ij} = y_{ij}$. For ordinal indicators, the problem of sampling $\boldsymbol{y}^*$ simplifies to the sampling of the individual $\boldsymbol{y}^*_i$ because the observations are independently and identically distributed. Furthermore, the multivariate vector $\boldsymbol{y}^*_i$ can be obtained by sampling the individual components y^*_{ij} because the conditional covariance matrix $V(\boldsymbol{y}^*_i|\boldsymbol{\theta}, \boldsymbol{z}_i, \boldsymbol{y}_i, \boldsymbol{w}_i, \boldsymbol{x}_i, \boldsymbol{u}_i) = \boldsymbol{\Sigma}$ is diagonal. The full conditional follows from Equations (3.1) and (3.12) which contain all information about z_{ij}, i. e.

$$\begin{aligned} y^*_{ij}|\boldsymbol{z}_i, \boldsymbol{\theta} \setminus \{\boldsymbol{\beta}, \boldsymbol{\gamma}\}, y_{ij}, \boldsymbol{w}_i \quad &\sim \quad N\left(\lambda_{j0} + \sum_{c=1}^{d} a_{jc} w_{ic} + \sum_{r=1}^{q} \lambda_{jr} z_{ir}, 1\right) \times \\ &\quad\;\; \sum_{k=1}^{K_j} \mathbb{1}_{\tau_{j,k-1} < y^*_{ij} \leq \tau_{jk}} \mathbb{1}_{y_{ij}=k}\,. \end{aligned} \tag{5.13}$$

We recognize that this distribution does not depend on the parameters of indirect effects $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$. Generating random values of the truncated normal distribution (5.13) can be problematic when using the inverse cumulative distribution, especially for values in the margins of the distribution. Geweke (1991) proposed an algorithm which solves those issues and increases computational speed.

B. Full conditionals of the latent variables

The full conditional $p(\boldsymbol{z}|\boldsymbol{\theta}, \boldsymbol{y}^*, \boldsymbol{y}, \boldsymbol{w}, \boldsymbol{x}, \boldsymbol{u}) = p(\boldsymbol{z}|\boldsymbol{\theta}, \boldsymbol{y}^*, \boldsymbol{w}, \boldsymbol{x}, \boldsymbol{u})$ springs from the joint distribution of $\boldsymbol{y}^*_i$ and $\boldsymbol{z}_i$ given $\boldsymbol{\theta}$, $\boldsymbol{x}_i$, $\boldsymbol{u}_i$ and $\boldsymbol{w}_i$ in (5.6), and is a multivariate normal distribution (e. g. see Tong, 1990) with the expectation vector

$$E(\boldsymbol{z}_i|\boldsymbol{\theta}, \boldsymbol{y}^*_i, \boldsymbol{w}_i, \boldsymbol{x}_i, \boldsymbol{u}_i) = \boldsymbol{\eta}_i + \boldsymbol{\Lambda}'(\boldsymbol{\Lambda}\boldsymbol{\Lambda}' + \boldsymbol{\Sigma})^{-1}(\boldsymbol{y}^*_i - \boldsymbol{\lambda}_0 - \boldsymbol{A}\boldsymbol{w}_i - \boldsymbol{\Lambda}\boldsymbol{\eta}_i)\,, \tag{5.14}$$

and covariance matrix

$$V(\boldsymbol{z}_i|\boldsymbol{\theta}, \boldsymbol{y}^*_i, \boldsymbol{w}_i, \boldsymbol{x}_i, \boldsymbol{u}_i) = I_{q\times q} - \boldsymbol{\Lambda}'(\boldsymbol{\Lambda}\boldsymbol{\Lambda}' + \boldsymbol{\Sigma})^{-1}\boldsymbol{\Lambda}\,, \tag{5.15}$$

where $\boldsymbol{\eta}_i$ denotes the predictor of the structural part of the model as defined in (3.9). A conditioning on $\boldsymbol{y}_i$ is not necessary since $\boldsymbol{y}_i$ is implicitly known through $\boldsymbol{y}^*_i$.

Random samples from this multivariate density are generated by sampling from a multivariate standard normal density with dimension q, multiplying the result with the Cholesky matrix of the covariance matrix (5.15), and adding the expectation vector (5.14). Since the number of indicators p is typically much higher than the number of latent variables q, it is computationally more efficient to calculate the inversion of $\boldsymbol{\Lambda}\boldsymbol{\Lambda}' + \boldsymbol{\Sigma}$ by applying Woodbury's identity (see Seber, 1984), i. e.

$$(\boldsymbol{\Lambda}\boldsymbol{\Lambda}' + \boldsymbol{\Sigma})^{-1} = \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\boldsymbol{\Lambda}(\boldsymbol{I}_{q\times q} + \boldsymbol{\Lambda}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Lambda})^{-1}\boldsymbol{\Lambda}'\boldsymbol{\Sigma}^{-1}\,.$$

Now, only the $q \times q$ matrix $\boldsymbol{I}_{q\times q} + \boldsymbol{\Lambda}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Lambda}$ has to be inverted instead of the $p \times p$ matrix $\boldsymbol{\Lambda}\boldsymbol{\Lambda}' + \boldsymbol{\Sigma}$, and the diagonal matrix $\boldsymbol{\Sigma}$ is inverted easily.

C. Full conditionals of the nonparametric indirect effects

The structural part of the model in Equation (3.10) or (3.13) forms the basis of the full conditional for all indirect effects, as all other specifications of the model do not provide any additional information necessary for the estimation of indirect effects. Since the error variance matrix of the latent variables $\boldsymbol{z}_i$ is diagonal and priori information about the nonparametric indirect effects is defined per latent variable and function, we can draw the parameter vectors $\boldsymbol{\beta}_{rh}$ of functions f_{rh} sequentially. The conditional expectation vector yields

$$E(\boldsymbol{\beta}_{rh}|\boldsymbol{\beta}_r \setminus \{\boldsymbol{\beta}_{rh}\}, \boldsymbol{z}_r, \boldsymbol{x}_r, \boldsymbol{u}) = \left(\boldsymbol{X}_h'\boldsymbol{X}_h + \frac{1}{\kappa_h}\boldsymbol{K}_h\right)^{-1} \boldsymbol{X}_h'(\boldsymbol{z}_r - \tilde{\boldsymbol{\eta}}_r)\,, \tag{5.16}$$

and the covariance matrix is given by

$$V(\boldsymbol{\beta}_{rh}|\boldsymbol{\beta}_r \setminus \{\boldsymbol{\beta}_{rh}\}, \boldsymbol{z}_r, \boldsymbol{x}_r, \boldsymbol{u}) = M_h^{-1} = \left(\boldsymbol{X}_h'\boldsymbol{X}_h + \frac{1}{\kappa_h}\boldsymbol{K}_h\right)^{-1}. \tag{5.17}$$

The term $\tilde{\boldsymbol{\eta}}_r$ contains the sum of all remaining parts of the predictor, i. e.

$$\tilde{\boldsymbol{\eta}}_r = (\tilde{\eta}_{1r}, \tilde{\eta}_{2r}, \ldots, \tilde{\eta}_{nr})' = X_1\boldsymbol{\beta}_{r1} + \ldots + X_{h-1}\boldsymbol{\beta}_{r,h-1} + X_{h+1}\boldsymbol{\beta}_{r,h+1} + \ldots + X_g\boldsymbol{\beta}_{rg} + \boldsymbol{U}\boldsymbol{\gamma}_r\,.$$

The terms $\boldsymbol{X}_h$, $\boldsymbol{K}_h$, and κ_h denote the design matrix, penalty matrix, and variance of the nonparametric function h as explained in Section 5.2.2. Although the parameter vector $\boldsymbol{\beta}_{rh}$ is drawn from a standard multivariate normal distribution, efficient sampling is not straightforward because linear equation systems with high-dimensional precision matrices M_h have to be solved in every iteration of the MCMC algorithm. We implemented an approach presented by Rue (2001) where first the Cholesky composition $M_h = C_hC_h'$ is calculated. Then the linear equation system $C_h'\boldsymbol{\beta}_{rh} = \boldsymbol{g}$ is solved where $\boldsymbol{g}$ is a vector of independent standard Gaussians, so $\boldsymbol{\beta}_{rh} \sim N(\boldsymbol{0}, C_h^{-1})$. We proceed by computing the mean of $\boldsymbol{\beta}_{rh}$ by solving $C_hE(\boldsymbol{\beta}_{rh}|\cdot) = X_h'(\boldsymbol{z}_r - \tilde{\boldsymbol{\eta}}_r)$ which is done by standard sequential forward and backward substitution. We obtain the final value of $\boldsymbol{\beta}_{rh}$ by adding the mean to the firstly calculated value of $\boldsymbol{\beta}_{rh}$.

Since the parameter vector is high-dimensional and often contains several hundred parameters, the solving of linear equations has to be done in a very efficient way to limit the computational task. Therefore all penalty matrices $\boldsymbol{K}_h$ should be transformed into a band matrix like structure, and all computational operations concerned with solving linear equations should be optimized for band matrices in an appropriate way. For metric covariates with random walk and P-spline priors, the penalty matrix $\boldsymbol{K}_h$ is already in band form; for spatial covariates however, the penalty matrix typically contains non-zero entries spread along the two dimensions of the matrix. For that reason, it is strongly recommended to transform the spatial penalty matrix $\boldsymbol{K}_h$ into band form. We choose the popular reverse Cuthill-McKee algorithm (e. g. see George and Liu, 1981) to perform this transformation.

After having sampled the parameter vector $\boldsymbol{\beta}_{rh}$, it is necessary to center the sampled parameter vector appropriately around zero because there is no intercept allowed in the structural part of the model due to identification restrictions. For random walk and spatial priors, centering is performed by calculating the mean of all function values $\boldsymbol{\beta}_{rh}$ and subtracting it from all function values $\boldsymbol{\beta}_{rh}$. For Bayesian P-splines, a weighted mean of the corresponding regression coefficients in the observed interval is computed and deducted. If there is no proper centering, different offsets will appear in all nonparametric functions in the predictor, leading to highly fluctuating or even non-converging parameter estimates.

We proceed by sampling the smoothing parameters κ_{rh}. As described above, the smoothing parameter is a priori inverse Gamma distributed according to $\kappa_{rh} \sim IG(a_{rh}, b_{rh})$. Then the full conditional distributions are also inverse Gamma distributed, i. e.

$$\kappa_{rh} \sim IG(a'_{rh}, b'_{rh}) \,,$$

with

$$a'_{rh} = a_{rh} + \frac{\operatorname{rank}(\boldsymbol{K}_h)}{2} \,, \qquad \text{and} \qquad b'_{rh} = b_{rh} + \frac{1}{2}\boldsymbol{\beta}'_{rh}\boldsymbol{K}_{rh}\boldsymbol{\beta}_{rh} \,.$$

Hyperpriors a_{rh} and b_{rh} of the smoothing parameter κ_{rh} are typically chosen to be highly diffuse but informative in order to ensure a proper posterior. Furthermore, we conduct several simulation studies for metric functions (Section 6.3.2) and spatial functions (Section 6.3.3) to examine the effect of two different noninformative prior distributions parameterized by $a_{rh} = -1$, $b_{rh} = 0$, and $a_{rh} = -0.5$, $b_{rh} = 0$, respectively.

D. Full conditionals of the parametric indirect effects

Equivalently to the nonparametric indirect effects, Equation (3.13) is the basis for the estimation of parametric indirect effects. Since the priori information of the parametric indirect effects is also defined for each latent variable r, $\boldsymbol{\gamma}_r \sim N(\boldsymbol{\gamma}_r^*, \boldsymbol{\Gamma}_r^{*-1})$, we can again estimate the regression parameters $\boldsymbol{\gamma}_r$ sequentially due to the diagonal covariance matrix of $\boldsymbol{\xi}_i$. Ergo for each latent variable r, we obtain a linear model with Gaussian response, and the full conditional distribution for the parametric indirect effects yields the expectation vector

$$E(\boldsymbol{\gamma}_r|\boldsymbol{\beta}_r, \boldsymbol{z}_r, \boldsymbol{x}_r, \boldsymbol{u}) = (\boldsymbol{\Gamma}_r^* + \boldsymbol{U}'\boldsymbol{U})^{-1} \left(\boldsymbol{\Gamma}_r^*\boldsymbol{\gamma}_r^* + \boldsymbol{U}'(\boldsymbol{z}_r - \tilde{\boldsymbol{\eta}}_r)\right) \,,$$

and the covariance matrix

$$V(\boldsymbol{\gamma}_r|\boldsymbol{\beta}_r, \boldsymbol{z}_r, \boldsymbol{x}_r, \boldsymbol{u}) = (\boldsymbol{\Gamma}_r^* + \boldsymbol{U}'\boldsymbol{U})^{-1} \,,$$

with the $(n \times m)$-dimensional design matrix $\boldsymbol{U}$ defined in the usual way, containing the covariates u_{il}. The term $\tilde{\boldsymbol{\eta}}_r$ contains the sum of all values of the remaining parts of the predictor, i. e.

$$\tilde{\boldsymbol{\eta}}_r = (\tilde{\eta}_{1r}, \tilde{\eta}_{2r}, \ldots, \tilde{\eta}_{nr})' = X_1\boldsymbol{\beta}_{r1} + \ldots + X_g\boldsymbol{\beta}_{rg} \,.$$

We always use diffuse priori distributions for parametric indirect effects leading to the following simplified expressions

$$E(\boldsymbol{\gamma}_r|\boldsymbol{\beta}_r, \boldsymbol{z}_r, \boldsymbol{x}_r, \boldsymbol{u}) = (\boldsymbol{U}'\boldsymbol{U})^{-1}\boldsymbol{U}'(\boldsymbol{z}_r - \tilde{\boldsymbol{\eta}}_r)\,, \qquad \text{and}$$
$$V(\boldsymbol{\gamma}_r|\boldsymbol{\beta}_r, \boldsymbol{z}_r, \boldsymbol{x}_r, \boldsymbol{u}) = (\boldsymbol{U}'\boldsymbol{U})^{-1}\,.$$

E. Full conditionals of the intercepts, direct effects and factor loadings

As defined in Section 5.2.1, the vector $\bar{\boldsymbol{\lambda}}$ contains all parameters $\boldsymbol{\lambda}_0$, $\boldsymbol{\Lambda}$ and $\boldsymbol{A}$ as in Equation (5.3). It is sufficient to examine the measurement model (3.11) because the parameters of the structural part of the model do not convey additional information when $\boldsymbol{z}_i$ is known. For an informative prior distribution $N(\bar{\boldsymbol{\lambda}}^*, \bar{\boldsymbol{\Lambda}}^{*-1})$, the full conditional $p(\bar{\boldsymbol{\lambda}}|\boldsymbol{\theta} \setminus \{\bar{\boldsymbol{\lambda}}\}, \boldsymbol{y}^*, \boldsymbol{z}, \boldsymbol{y}, \boldsymbol{w}, \boldsymbol{x}, \boldsymbol{u}) = p(\bar{\boldsymbol{\lambda}}|\boldsymbol{\Sigma}, \boldsymbol{y}^*, \boldsymbol{z}, \boldsymbol{w})$ is a $p \cdot (1+d+q)$-dimensional multivariate normal distribution with expectation vector

$$E(\bar{\boldsymbol{\lambda}}|\boldsymbol{\Sigma}, \boldsymbol{y}^*, \boldsymbol{z}, \boldsymbol{w}) = \left(\bar{\boldsymbol{\Lambda}}^* + \sum_{i=1}^{n} \boldsymbol{L}_i'\boldsymbol{\Sigma}^{-1}\boldsymbol{L}_i\right)^{-1} \left(\bar{\boldsymbol{\Lambda}}^*\bar{\boldsymbol{\lambda}}^* + \sum_{i=1}^{n} \boldsymbol{L}_i'\boldsymbol{\Sigma}^{-1}\boldsymbol{y}_i^*\right),$$

and covariance matrix

$$V(\bar{\boldsymbol{\lambda}}|\boldsymbol{\Sigma}, \boldsymbol{y}^*, \boldsymbol{z}, \boldsymbol{w}) = \left(\bar{\boldsymbol{\Lambda}}^* + \sum_{i=1}^{n} \boldsymbol{L}_i'\boldsymbol{\Sigma}^{-1}\boldsymbol{L}_i\right)^{-1}.$$

The $(p \times p(1+d+q))$-dimensional matrix $\boldsymbol{L}_i$ is defined as $\boldsymbol{L}_i = \boldsymbol{I}_{p\times p} \otimes \boldsymbol{l}_i$ with $\boldsymbol{l}_i = (1, w_{i1}, \ldots, w_{id}, z_{i1}, \ldots, z_{iq})$.

For diffuse priors, priors with a diagonal prior precision matrix $\bar{\boldsymbol{\Lambda}}^*$, and priors with a precision matrix $\bar{\boldsymbol{\Lambda}}^*$ with off-diagonal zero-entries for parameters for different indicators, the full conditional can be calculated sequentially for each indicator because $V(\boldsymbol{\varepsilon}_i) = \boldsymbol{\Sigma}$ is diagonal. Let the vector $\bar{\boldsymbol{\lambda}}^j$ contain the parameters of row j, i. e. $\bar{\boldsymbol{\lambda}}^j = (\lambda_{j0}, a_{j1}, \ldots, a_{jd}, \lambda_{j1}, \ldots, \lambda_{jm})'$. Ergo the full conditional distribution of $\bar{\boldsymbol{\lambda}}^j$ is a $(1+m+q)$-dimensional multivariate normal distribution with expectation vector

$$E(\bar{\boldsymbol{\lambda}}^j|\sigma_j^2, \boldsymbol{y}_j^*, \boldsymbol{z}, \boldsymbol{w}) = (\boldsymbol{L}'\boldsymbol{L})^{-1}\boldsymbol{L}'\boldsymbol{y}_j^*\,,$$

and covariance matrix

$$V(\bar{\boldsymbol{\lambda}}^j|\sigma_j^2, \boldsymbol{y}_j^*, \boldsymbol{z}, \boldsymbol{w}) = \sigma_j^2(\boldsymbol{L}'\boldsymbol{L})^{-1}\,.$$

The $(n \times (1+d+q))$-dimensional matrix $\boldsymbol{L}$ is defined by $\boldsymbol{L} = (\boldsymbol{l}_1, \ldots, \boldsymbol{l}_n)'$ with the rows $\boldsymbol{l}_i = (1, w_{i1}, \ldots, w_{id}, z_{i1}, \ldots, z_{iq})$. For our choice of priors, we can always use the second way of sampling the parameters which provides faster computation.

Here it becomes clear why the error variances for ordinal indicators are fixed to 1. In order to obtain a standardized solution, we had to sample the factor loadings under the restriction $\sum_{r=1}^{q} \lambda_{jr}^2 + \sigma_j^2 = 1$. Since this cannot be achieved by a standard full conditional, we fix the error variances of ordinal manifest variables to 1 instead. A standardization is easily possible after the simulation run by using the formulas provided by Table 3.2.

F. Full conditionals of the error variances

For ordinal variables, the sampling of error variances is ommitted because error variances are set to 1 due to identification restrictions of the model. For continuous indicators however, error variances $\boldsymbol{\sigma}^2$ have to be estimated. Due to a diagonal error variance matrix $\boldsymbol{\Sigma}$, parameters σ_j^2 can be sampled sequentially based on Equation (3.12). Since error variances are distributed normally, the full conditional $p(\sigma_j^2|\boldsymbol{\theta} \setminus \{\boldsymbol{\sigma}^2\}, \boldsymbol{y}_i^*, \boldsymbol{z}, \boldsymbol{w})$ results to an inverse gamma distribution with $n+\nu$ degrees of freedom and scale parameter s^2, i. e.

$$\sigma_j^2|\boldsymbol{\theta} \setminus \{\boldsymbol{\sigma}^2\}, \boldsymbol{y}_i^*, \boldsymbol{z}, \boldsymbol{w} \quad \sim \quad IG\left(\frac{n+\nu}{2}, \frac{(n+\nu)S^2}{2}\right), \tag{5.18}$$

with

$$S^2 = \frac{1}{n+\nu}\left(\sum_{i=1}^{n}\left(y_{ij}^* - \lambda_{j0} - \sum_{c=1}^{d} a_{jc}w_{ic} - \sum_{r=1}^{q}\lambda_{jr}z_{ir}\right)^2 + \nu s^2\right), \tag{5.19}$$

where σ_j^2 is a priori $IG(\nu/2, \nu s^2/2)$ distributed. In the case of a noninformative prior distribution for σ_j^2, the full conditional distribution is obtained by setting the values of ν and s in (5.18) and (5.19) to zero.

Random samples can be drawn efficiently by drawing a random value R from a χ^2 distribution with $n+\nu$ degrees of freedom, and setting $\sigma_j^2 = (n+\nu)S^2/R$.

G. Full conditionals of the cutpoints

For continuous indicators, there are no cutpoints which have to estimated. The full conditional of τ_{jk} $(2 \leq k \leq K_j - 1)$ given $\boldsymbol{y}_j$, $\boldsymbol{y}_j^*$ and $\boldsymbol{\theta} \setminus \{\tau_{jk}\}$ denotes

$$\tau_{jk}|\boldsymbol{y}_j, \boldsymbol{y}_j^*, \boldsymbol{\theta} \setminus \{\tau_{jk}\} \propto \prod_{i=1}^{n}\left[\mathbb{1}_{\tau_{j,k-1}<y_{ij}^*\leq\tau_{jk}}\mathbb{1}_{y_{ij}=k} + \mathbb{1}_{\tau_{jk}<y_{ij}^*\leq\tau_{j,k+1}}\mathbb{1}_{y_{ij}=k+1}\right]. \tag{5.20}$$

Hence, all of the following conditions have to be fulfilled in order to gain a non-zero posteriori distribution:

$$\begin{aligned}
\tau_{jk} &\geq y_{ij}^*, \qquad i=1,\ldots,n, \text{ where } y_{ij}=k, \\
\tau_{jk} &\geq \tau_{j,k-1}, \\
\tau_{jk} &\leq y_{ij}^*, \qquad i=1,\ldots,n, \text{ where } y_{ij}=k+1, \\
\tau_{jk} &\leq \tau_{j,k+1}.
\end{aligned}$$

This corresponds to a uniform distribution for τ_{jk} in the interval $[l_{\tau k}, r_{\tau k}]$, with the interval borders

$$
\begin{aligned}
l_{\tau,k} :=& \max\left\{\tau_{j,k-1}, \max_{i=1,\ldots,n}\{y^*_{ij}|y_{ij}=k\}\right\},\\
r_{\tau,k} :=& \min\left\{\tau_{j,k+1}, \min_{i=1,\ldots,n}\{y^*_{ij}|y_{ij}=k+1\}\right\},
\end{aligned}
$$

where $\max(\emptyset) = -\infty$ and $\min(\emptyset) = \infty$. Drawing random numbers from uniform distributions is trivial. Now it becomes very clear why this sampling algorithm exhibits bad convergence properties. The cutpoints τ_{jk} have almost no room to move in the small interval $[l_{\tau k}, r_{\tau k}]$, especially for a moderate or high number of observations. This also leads to a poor convergence of some other parameters of the model, especially the intercepts $\boldsymbol{\lambda}_0$. The convergence properties can be improved by introducing a MH step for the sampling of the cutpoints. This is implemented in our second algorithm presented in the following section.

5.4.2 Algorithm 2: the MH sampler (MHS)

As mentioned before, the standard sampler exhibits bad convergence properties, especially for the cutpoint and intercept parameters. Cowles (1996) proposed an algorithm which improves convergence and still has a convenient structure. The overall algorithm of the standard sampler remains the same, only the sampling step G for the cutpoint parameters is a MH step instead of a Gibbs step; therefore this sampling algorithm is named the MH sampler (MHS). The algorithm can be found in the paper of Cowles (1996), and Johnson and Albert (1999). Another notable algorithm how to update the cutpoints has been contributed by Nandram and Chen (1996) who included a reparameterization of the cutpoints.

The MH step replacing the Gibbs step G involves three parts: first a proposal of cutpoint values for one indicator is drawn, second the MH acceptance probability is calculated and finally the proposed cutpoint values are accepted with that acceptance probability. We proceed with a description of these three steps which have to be performed sequentially for all ordinal indicators at iteration t:

Repeat for $j = 1, \ldots, p_1$

- *Draw a set of proposal cutpoints* $\tilde{\boldsymbol{\tau}}$
  For $k = 2, \ldots, K_j - 1$, sample $\tilde{\tau}_k \sim N(\tau_{jk}^{(t-1)}, \sigma_{MH})$ truncated to the interval $[\tilde{\tau}_{k-1}, \tau_{j,k+1}^{(t-1)}]$.

- *Compute the acceptance ratio R*

$$R = \prod_{i=1}^{n} \frac{\Phi(\tilde{\tau}_{y_i} - \mu_{ij}^{(t-1)}) - \Phi(\tilde{\tau}_{y_i-1} - \mu_{ij}^{(t-1)})}{\Phi(\tau_{y_i}^{(t-1)} - \mu_{ij}^{(t-1)}) - \Phi(\tau_{y_i-1}^{(t-1)} - \mu_{ij}^{(t-1)})}$$
$$\times \prod_{k=2}^{K_j-1} \frac{\Phi((\tau_{k+1}^{(t-1)} - \tau_k^{(t-1)})/\sigma_{MH}) - \Phi((\tilde{\tau}_{k-1} - \tau_k^{(t-1)})/\sigma_{MH})}{\Phi((\tilde{\tau}_{k+1} - \tilde{\tau}_k)/\sigma_{MH}) - \Phi((\tau_{k-1}^{(t-1)} - \tilde{\tau}_k)/\sigma_{MH})} .$$

  $\tilde{\tau}_{y_i}$ denotes the cutpoint proposal corresponding to the ordinal value of observation y_i, similarly $\tau_{y_i}^{(t-1)}$ is the actual value of the cutpoint corresponding to the observed ordinal category at iteration $t-1$. The term $\mu_i^{(t-1)}$ denotes the value of the linear predictor in the measurement model at iteration $t-1$ for observation i and indicator j resulting in $\mu_i^{(t-1)} = \lambda_{j0}^{(t-1)} + \sum_{c=1}^{d} a_{jc}^{(t-1)} w_{ic} + \sum_{r=1}^{q} \lambda_{jr}^{(t-1)} z_{ir}^{(t-1)}$.

- *Accept or reject proposal value* $\tilde{\boldsymbol{\tau}}$
  Set $\boldsymbol{\tau}_j^{(t)} = \tilde{\boldsymbol{\tau}}$ with probability R, otherwise set $\boldsymbol{\tau}_j^{(t)} = \boldsymbol{\tau}_j^{(t-1)}$.

We emphasize that the current value of the underlying variable y_{ij}^* does not play a role neither in the calculation of the proposal values nor in the computation of the acceptance ratio. The value σ_{MH} can be considered a tuning parameter and has to be set by the researcher before starting the simulation. A rule of thumb recommends setting $\sigma_{MH} = 0.05/K_j$ which should lead to acceptance ratios of 25-50%. If necessary, a different σ_{MH}^j for each indicator could be employed to achieve proper acceptance ratios for all indicators j, for example when the number of ordinal categories varies across indicators.

This popular standard algorithm of sampling the cutpoints still inhibits some drawbacks. Convergence is still not optimal, a MH step has to be employed with the serious drawback of setting and adjusting the tuning parameter σ_{MH}, and the calculation of the acceptance ratio is computationally demanding, especially in an analysis involving many ordinal indicators.

5.4.3 Algorithm 3: the Generalized Gibbs sampler (GGS)

In order to receive even better convergence properties than the MHS and compensate for its drawbacks, we introduce a third algorithm based on the work of Liu and Sabatti (2000) presented in Section 4.2.5. The general idea is to find a suitable transformation – the so-called Generalized Gibbs move – which transforms some or all of the parameters in such a way that convergence increases for ordinal indicators. In the end, the Generalized Gibbs sampler is identical to the standard Gibbs sampler, but furthermore applies p_1 different transformations to p_1 different sets of parameters at the end of each iteration. In this section we present the p_1 subsets of parameters and derive the transformation applied to them, including the corresponding probability distribution for the transformation variable. We start with a reminder of the total parameter vector of the LVM, i. e. $\text{vec}\,\{\{\boldsymbol{\lambda_0}, \boldsymbol{\Lambda}, \boldsymbol{A}, \boldsymbol{\Sigma}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\tau}\} \cup \{\boldsymbol{y}^*, \boldsymbol{z}\}\}$.

The difficulty is to find a suitable transformation group Γ so that the resulting distribution allows a fast sampling of the transformation members[3] γ. Since a Generalized Gibbs move for the whole posterior cannot be deducted, we develop an individual Generalized Gibbs move for each of the p_1 linear submodels in the measurement model for all ordinal indicators. This is possible due to the diagonal form of the error variance matrix $\boldsymbol{\Sigma}$. We identified the partial scale group on $\boldsymbol{S}$ to be a suitable transformation group:

$$\Gamma_v := \{\gamma > 0 : \gamma(\boldsymbol{\theta}) = (\gamma\theta_1, \ldots, \gamma\theta_v, \theta_{v+1}, \ldots, \theta_{dim})\} .$$

Here only v components are transformed, the others remain fixed. The left-Haar measure for this group is $\gamma^{-1}d\gamma$ as for the total scale group. The determinant of the Jacobian is $\det(\partial\gamma(\boldsymbol{\theta})/\partial\boldsymbol{\theta}) = \gamma^v$. This yields

$$\pi(\gamma(\boldsymbol{\theta}))\,|J_\gamma(\boldsymbol{\theta})|\,L(d\gamma) = \gamma^{v-1}\pi(\gamma(\boldsymbol{\theta}))\,d\gamma .$$

Now we specify suitable subsets $\boldsymbol{\theta}_j$ of the total parameter vector $\boldsymbol{\theta}$, so that we can transform these subsets' parameters for each indicator j. Accordingly we define the p_1 parameter vectors $\boldsymbol{\theta}_j$ per ordinal indicator j as

$$\boldsymbol{\theta}_j = (y^*_{j1}, \ldots, y^*_{jn}, \lambda_{j0}, a_{j1}, \ldots, a_{jd}, \lambda_{j1}, \ldots, \lambda_{jm}, \tau_{j2}, \ldots, \tau_{jK_{j-1}}) ,$$

each of which contains $v = n + d + m + K_j - 1$ parameters. Thus we get p_1 different Generalized Gibbs moves

$$\gamma_j(\boldsymbol{\theta}_j) = (\gamma_j y^*_{j1}, \ldots, \gamma_j y^*_{jn}, \gamma_j\lambda_{j0}, \gamma_j a_{j1}, \ldots, \gamma_j a_{jd}, \gamma_j\lambda_{j1}, \ldots, \gamma_j\lambda_{jm}, \gamma_j\tau_{j2}, \ldots, \gamma_j\tau_{jK_{j-1}}) ,$$

that transform the corresponding parameter set $\boldsymbol{\theta}_j$. All other parameters of the full parameter vector remain constant and are not transformed. The individual parameter sets $\boldsymbol{\theta}_j$

[3]Note that the transformation members γ and γ_j in this section are not associated with the regression coefficients $\boldsymbol{\gamma}$ of the structural equation in Section 3.2.

can be derived by arranging the posterior distribution in the following way:

$$
\begin{aligned}
&p(\boldsymbol{\theta})\prod_{i=1}^{n}\left[p(\boldsymbol{y}_i^*,\boldsymbol{z}_i|\boldsymbol{\theta},\boldsymbol{w}_i,\boldsymbol{x}_i,\boldsymbol{u}_i)\times\left(\prod_{j=1}^{p_1}\sum_{k=1}^{K_j}\mathbb{1}_{\tau_{j,k-1}<y_{ij}^*\le\tau_{jk}}\mathbb{1}_{y_{ij}=k}\right)\right]\\
=&p(\boldsymbol{\theta})\prod_{i=1}^{n}\left[p(\boldsymbol{z}_i|\boldsymbol{\theta},\boldsymbol{w}_i,\boldsymbol{x}_i,\boldsymbol{u}_i)\cdot p(\boldsymbol{y}_i^*|\boldsymbol{\theta},\boldsymbol{z}_i,\boldsymbol{w}_i,\boldsymbol{x}_i,\boldsymbol{u}_i)\times\left(\prod_{j=1}^{p_1}\sum_{k=1}^{K_j}\mathbb{1}_{\tau_{j,k-1}<y_{ij}^*\le\tau_{jk}}\mathbb{1}_{y_{ij}=k}\right)\right]\\
=&p(\boldsymbol{\theta})\prod_{i=1}^{n}\left[p(\boldsymbol{z}_i|\boldsymbol{\theta},\boldsymbol{w}_i,\boldsymbol{x}_i,\boldsymbol{u}_i)\times\prod_{j=p_1+1}^{p}p(y_{ij}^*|\boldsymbol{z}_i,\boldsymbol{\theta}_j\setminus\{\boldsymbol{\beta},\boldsymbol{\gamma}\},\boldsymbol{w}_i)\right.\\
&\qquad\left.\times\prod_{j=1}^{p_1}\left(p(y_{ij}^*|\boldsymbol{z}_i,\boldsymbol{\theta}_j\setminus\{\boldsymbol{\beta},\boldsymbol{\gamma}\},\boldsymbol{w}_i)\sum_{k=1}^{K_j}\mathbb{1}_{\tau_{j,k-1}<y_{ij}^*\le\tau_{jk}}\mathbb{1}_{y_{ij}=k}\right)\right]\\
=&p(\boldsymbol{\theta})\prod_{i=1}^{n}p(\boldsymbol{z}_i|\boldsymbol{\theta},\boldsymbol{w}_i,\boldsymbol{x}_i,\boldsymbol{u}_i)\times\prod_{i=1}^{n}\prod_{j=p_1+1}^{p}p(y_{ij}^*|\boldsymbol{z}_i,\boldsymbol{\theta}_j\setminus\{\Gamma\},\boldsymbol{w}_i)\\
&\qquad\times\prod_{j=1}^{p_1}\left[\underbrace{\prod_{i=1}^{n}\left(p(y_{ij}^*|\boldsymbol{z}_i,\boldsymbol{\theta}_j\setminus\{\Gamma\},\boldsymbol{w}_i)\right)\sum_{k=1}^{K_j}\mathbb{1}_{\tau_{j,k-1}<y_{ij}^*\le\tau_{jk}}\mathbb{1}_{y_{ij}=k}}\right].
\end{aligned}
$$

Thus we deploy p_1 distinct Generalized Gibbs moves for the p_1 components of the posterior distribution which are underbraced in the above formula. Based on the underbraced part of the posterior, we can formulate the densities $\gamma_j^{v-1}\pi(\gamma_j\boldsymbol{\theta}_j|\cdot)$ to be proportional to

$$
\begin{aligned}
&\gamma_j^{v-1}\exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\left(\gamma_j y_{ij}^*-\gamma_j\lambda_{j0}-\sum_{c=1}^{d}\gamma_j a_{jc}w_{ic}-\sum_{r=1}^{q}\gamma_j\lambda_{jr}z_{ir}\right)^2\right\}\\
=&\left(\gamma_j^2\right)^{\frac{v-1}{2}}\exp\left\{-\frac{1}{2}\gamma_j^2\sum_{i=1}^{n}\left(y_{ij}^*-\lambda_{j0}-\sum_{c=1}^{d}a_{jc}w_{ic}-\sum_{r=1}^{q}\lambda_{jr}z_{ir}\right)^2\right\}.
\end{aligned}
\tag{5.21}
$$

We dropped the right hand side of the underbraced formula because the right term remains constant under the transformation according to

$$
\sum_{k=1}^{K_j}\mathbb{1}_{\gamma_j\tau_{j,k-1}<\gamma_j\bar{y}_{ij}\le\gamma_j\tau_{jk}}\mathbb{1}_{y_{ij}=k}\quad\Longleftrightarrow\quad\sum_{k=1}^{K_j}\mathbb{1}_{\tau_{j,k-1}<\bar{y}_{ij}\le\tau_{jk}}\mathbb{1}_{y_{ij}=k}\,.
$$

It can be seen from Equation (5.21) that γ_j^2 follows a Gamma distribution $\Gamma(a_j,b_j)$ with parameters

$$
\begin{aligned}
a_j&=\frac{v+1}{2}=\frac{n+d+m+K_j}{2},\\
b_j&=\frac{\sum_{i=1}^{N}\left(y_{ij}^*-\lambda_{j0}-\sum_{c=1}^{d}a_{jc}w_{ic}-\sum_{r=1}^{q}\lambda_{jr}z_{ir}\right)^2}{2},
\end{aligned}
\tag{5.22}
$$

and the density of $\Gamma(a,b)$ is given by $f(x|a,b)=b^a x^{a-1}e^{-bx}/\Gamma(a)$ for $x\ge 0$. Finally, the modified sampling algorithm for the Generalized Gibbs sampler comprises of the following steps:

- Steps A–G are equivalent to the standard sampler.

- Introduction of a Generalized Gibbs step H which comprises of the sequential execution of p_1 Generalized Gibbs moves $j = 1, \ldots, p_1$:

  Draw γ_j^2 from $\Gamma(a_j, b_j)$ with a_j and b_j as defined in Equations (5.22), and update all parameters of the subset $\boldsymbol{\theta}_j$ in the following way:

$$\begin{aligned} y^{*\,new}_{\cdot j} &\longleftarrow \gamma_j \, y^{*\,current}_{\cdot j} \, , \\ \lambda^{new}_{j\cdot} &\longleftarrow \gamma_j \, \lambda^{current}_{j\cdot} \, , \\ a^{new}_{j\cdot} &\longleftarrow \gamma_j \, a^{current}_{j\cdot} \, , \\ \tau^{new}_{j\cdot} &\longleftarrow \gamma_j \, \tau^{current}_{j\cdot} \, . \end{aligned}$$

5.4.4 Starting values

For all three MCMC algorithms, we have to specify starting values for all parameters $\boldsymbol{\theta} = \{\boldsymbol{\lambda_0}, \boldsymbol{\Lambda}, \boldsymbol{A}, \boldsymbol{\Sigma}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\tau}\}$. After fixing the starting values, the latent variables $\boldsymbol{z}$ can be calculated by using the parameter values of the structural part of the model. A common strategy of specifying starting values is the use of the respective maximum likelihood estimates of the model. Unfortunately, this approach is not possible here because no frequentist equivalent to our model exists. In general it is desirable that the MCMC algorithm converges to the posterior distribution regardless of the specific choice of starting values. Therefore we make straightforward choices for all parameters, and assume that they do not influence the parameter estimation process. Intercepts $\boldsymbol{\lambda}_0$ and regression parameters of direct effects $\boldsymbol{A}$ are set to zero. The first free factor loading parameter of each latent variable is set to 0.7 to promote a positive solution for those parameters, all other factor loadings are also set to zero. Variances σ_j^2 for continuous indicators[4] start at the value 1. In the structural part of the model, nonparametric function parameters $\boldsymbol{\beta}$ and standard regression coefficients $\boldsymbol{\gamma}$ start at zero. Cutpoints τ_{jk} are set to be $1, \ldots, K_j - 2$ for $k = 2, \ldots, K_j - 1$. This choice of starting values implies values of zero for the latent variables $\boldsymbol{z}$, while specifying starting values for the underlying variables $\boldsymbol{y}^*$ is not necessary because they are sampled in the first iteration of all three sampling schemes. We performed a sensitivity analysis and tested variations of different starting values for all involved parameters. The results show that starting values do not affect the resulting parameter estimates, as long as we refrain from using implausible and far-fetched starting values (e. g. factor loadings higher than 5).

[4]For ordinal indicators, variances are fixed to 1 for all iterations.

5.4.5 Consideration of parameter restrictions

MCMC algorithms allow the inclusion of various types of parameter restrictions, three of which are presented here. Firstly, some parameters might have to be fixed to certain values; secondly, some parameters underly unequality constraints; and thirdly, certain parameters might have to be equal to other parameters of the same type. For example, certain factor loading parameters have to be fixed to zero, others are restrained to be greater than zero (see Section 3.1.2), and sometimes a specific factor loading parameter should be equal to another factor loading parameter. For our analyses, we only make use of the first two types of restrictions.

- *Fixing parameters to predetermined values*
  If the parameter is sampled individually, the starting value of the parameter is set to the desired value, and all subsequent Gibbs steps are skipped. If the parameter is sampled together with other parameters, e. g. by a multivariate normal distribution, sample the parameter block as usual, and set the target parameter to the desired value after each iteration of the respective full conditional.

- *Secure unequality constraints for parameters*
  A common way of dealing with unequality constraints requires the full run of the MCMC algorithm. After the simulation, all samples which do not satisfy the unequality constraint are dropped. Of course, this method is only applicable and efficient if not too many samples have to be dropped. Another possibility to influence the target region of a parameter is the employment of suitable starting value. For example, to ensure that a certain factor loading parameter is greater than zero, we choose a starting value of 0.7 so that the probability of obtaining a negative factor loading after the burnin phase is lowered. Often, unequality constraints are not compatible with the underlying analysis, e. g. it is not useful to expect and restrict two factor loadings for two different indicators to be greater than zero if the latent variable influences those two indicators in opposite ways.

- *Guarantee equality of parameters*
  Equality of two or more parameters of the same type can be ensured by adding the corresponding columns in the design matrix. For example, in order to guarantee equality of two regression parameters of direct effects, add the two corresponding columns of the design matrix $\boldsymbol{L}$.

Chapter 6

Simulation Studies

In this chapter several simulation studies are performed in order to examine the convergence and the estimation properties of the three MCMC samplers for various models and parameter sets of the LVM. The chapter is divided into three sections. In the first section, the convergence properties of the three MCMC samplers regarding the measurement model are investigated. At this stage we restrict the discussion to the case of ordinal indicators because they appear more often in applications than metric indicators, and MCMC convergence of the traditional factor analysis model with metric indicators is generally fine as reported in the literature and confirmed by our own studies. After that, we analyze the estimation properties of the MHS and the GGS without indirect effects by simulating five data sets with different types of indicators. Again the focus lies on data sets including ordinal response due to the same reasons as just mentioned. In the third section, the estimation properties of indirect effects such as parametric covariates, metric covariates, spatial effects, and VCM are studied. The additional estimation of indirect effects exhibits almost no influence on the estimation of parameters belonging to the measurement model. For that reason, we split the discussion into the parameter estimation for the measurement model and structural equation, respectively.

6.1 Convergence comparison of the three MCMC algorithms

In this section, the convergence properties of the three different samplers regarding factor loadings, intercepts and cutpoints are examined. Firstly, we look at the number of iterations it takes relevant parameters to reach their true values. Secondly, autocorrelations are calculated and plotted to indicate poor or good convergence.

We simulated a data set with $N = 5000$ observations, five ordinal indicators (with five categories each), one latent variable, and without direct and indirect effects[1]. The data is generated according to the following model

$$z_{i1} = \xi_{i1} \ , \text{with} \quad \xi_{i1} \sim N(0,1) \, ,$$

$$\begin{pmatrix} y^*_{i1} \\ y^*_{i2} \\ y^*_{i3} \\ y^*_{i4} \\ y^*_{i5} \end{pmatrix} = \begin{pmatrix} 0.943 \\ 1.039 \\ 1.125 \\ 1.260 \\ 2.065 \end{pmatrix} + \begin{pmatrix} 0.314 \\ 0.577 \\ 0.750 \\ 0.980 \\ 2.065 \end{pmatrix} z_{i1} + \begin{pmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \varepsilon_{i3} \\ \varepsilon_{i4} \\ \varepsilon_{i5} \end{pmatrix} \quad , \text{with} \quad \varepsilon_{ij} \sim N(0,1) \, ,$$

$$y_{i1} = \begin{cases} 1, & \text{for } y^*_{i1} \le 0 \\ 2, & \text{for } 0 < y^*_{i1} \le 0.629 \\ 3, & \text{for } 0.629 < y^*_{i1} \le 1.258 \\ 4, & \text{for } 1.258 < y^*_{i1} \le 1.887 \\ 5, & \text{for } y^*_{i1} > 1.887 \end{cases} \qquad y_{i2} = \begin{cases} 1, & \text{for } y^*_{i2} \le 0 \\ 2, & \text{for } 0 < y^*_{i2} \le 0.693 \\ 3, & \text{for } 0.693 < y^*_{i2} \le 1.386 \\ 4, & \text{for } 1.386 < y^*_{i2} \le 2.078 \\ 5, & \text{for } y^*_{i2} > 2.078 \end{cases} \qquad (6.1)$$

$$y_{i3} = \begin{cases} 1, & \text{for } y^*_{i3} \le 0 \\ 2, & \text{for } 0 < y^*_{i3} \le 0.750 \\ 3, & \text{for } 0.750 < y^*_{i3} \le 1.500 \\ 4, & \text{for } 1.500 < y^*_{i3} \le 2.250 \\ 5, & \text{for } y^*_{i3} > 2.250 \end{cases} \qquad y_{i4} = \begin{cases} 1, & \text{for } y^*_{i4} \le 0 \\ 2, & \text{for } 0 < y^*_{i4} \le 0.840 \\ 3, & \text{for } 0.840 < y^*_{i4} \le 1.680 \\ 4, & \text{for } 1.680 < y^*_{i4} \le 2.521 \\ 5, & \text{for } y^*_{i4} > 2.521 \end{cases}$$

$$y_{i5} = \begin{cases} 1, & \text{for } y^*_{i5} \le 0 \\ 2, & \text{for } 0 < y^*_{i5} \le 1.376 \\ 3, & \text{for } 1.376 < y^*_{i5} \le 2.753 \\ 4, & \text{for } 2.753 < y^*_{i5} \le 4.129 \\ 5, & \text{for } y^*_{i5} > 4.129 \, . \end{cases}$$

The dataset used here corresponds to the parameter set C presented in the next section in Tables 6.1 and 6.2. The starting values are chosen according to Section 5.4.4 apart from the factor loadings which all start at 0 in order to reveal how fast the sampled factor loadings approach their true values. Priori information is diffuse except for the factor loadings which receive a standard normal prior as outlined in Section 5.2.1. Simulations with other parameter sets were also conducted to find out if the properties of the three samplers are similar for various data sets, for example with mixed ordinal and metric indicators; however, the general behaviour of the three MCMC algorithms is equivalent to the results of the paremeter set employed here. Therefore we restrict the discussion to one simulated dataset.

[1]The inclusion of direct or indirect effects has no influence on the convergence of the parameters of the measurement model (factor loadings, intercepts, and cutpoints), as simulation studies have shown.

First we demonstrate how fast the parameters approach the stationary distribution. For that purpose one simulation with 500 iterations and no burnin phase is run. Exemplary, the sampling paths of the critical parameters – cutpoints, intercepts and factor loadings – are plotted for two indicators, one with a low and one with a high factor loading each. Figures 6.1 and 6.2 show the cutpoints for indicators 2 and 4, respectively. The corresponding figures for the intercepts and factor loadings of indicators 2 and 4 are depicted in Figures 6.3 and 6.4. In all four illustrations, sampling paths of the respective parameters are drawn for the three different MCMC algorithms, and true parameters are drawn as horizontal thin lines.

The results are very clear. The standard sampler exhibits extremely bad convergence properties. The sampling paths approach the true parameter values very slowly, or almost not at all. Even when simulating more than 50,000 iterations, the sampled parameters are not in the region of the true parameters. Due to its bad convergence properties, the standard sampler is not discussed in the subsequent sections. The MHS, with an acceptance ratio of about 40%, shows a much better convergence. For example, all parameters of indicator 2 meet their true values at about iteration 150. However, we recognize that convergence for the parameters of indicator 4 with a higher factor loading is not as good as for indicator 2. It is a general result that the convergence of parameters for indicators with higher factor loadings is worse than for indicators with lower factor loadings. Finally, the GGS demonstrates a very fast approach of the sampling paths to the respective true values for both indicators 2 and 4. After about 30 iterations from the start, samples around the true parameter values are drawn. This implicates a far superior convergence property of the GGS compared to the MHS; however, there is a drawback of the GGS as explained in the following.

We proceed by analyzing the autocorrelations of the sampling paths which are plotted in Figure 6.5 for the cutpoints, intercepts and factor loadings of indicators 2 and 4. As expected, the standard sampler exhibits very bad convergence which is indicated by a very high autocorrelation. For the MHS and the GGS, autocorrelation curves are quite similar, except for the autocorrelations of the second cutpoints of the GGS which are rather high. In order to check if this behaviour of the GGS is universal, autocorrelations of the cutpoints, intercepts and factor loadings of indicator 5, and for the second cutpoint of indicator 3 are drawn in Figure 6.6. Again, autocorrelations of the second cutpoints for the GGS are high, although higher for the third indicator than for the fifth one with a higher factor loading. We also recognize that autocorrelations of the MHS for all parameters are the poorest for the indicator with the highest factor loading. This also is a general property of the MHS. In order to examine the peculiar behaviour of the autocorrelations for the second cutpoints of the GGS in more detail, sampling paths and density functions of the second cutpoints for indicators 3 and 4 for the MHS and the GGS are depicted in Figure 6.7. Sampling paths of the MHS look fine, density functions, however, are not completely smooth. For the GGS, we observe sampling paths oscillating in a wavy fashion, and rather smooth density functions. This behaviour of the sampling paths is characteristic for the second

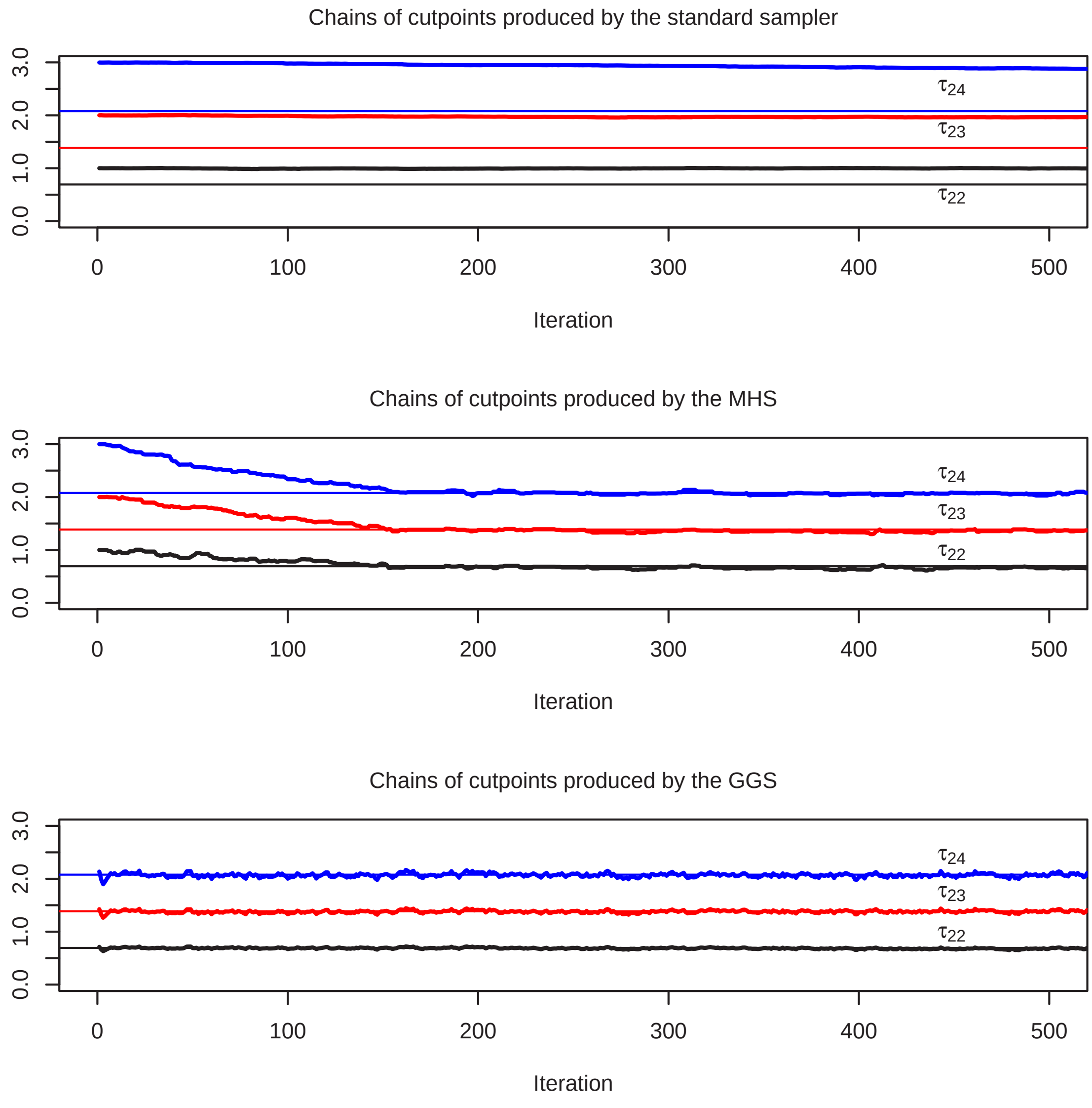


Figure 6.1: *First 500 sampling iterations of the cutpoints of indicator 2 (factor loading of 0.577) produced by the standard Gibbs sampler (top), the MHS (middle) and the GGS (bottom). The horizontal thin lines indicate the true values.*

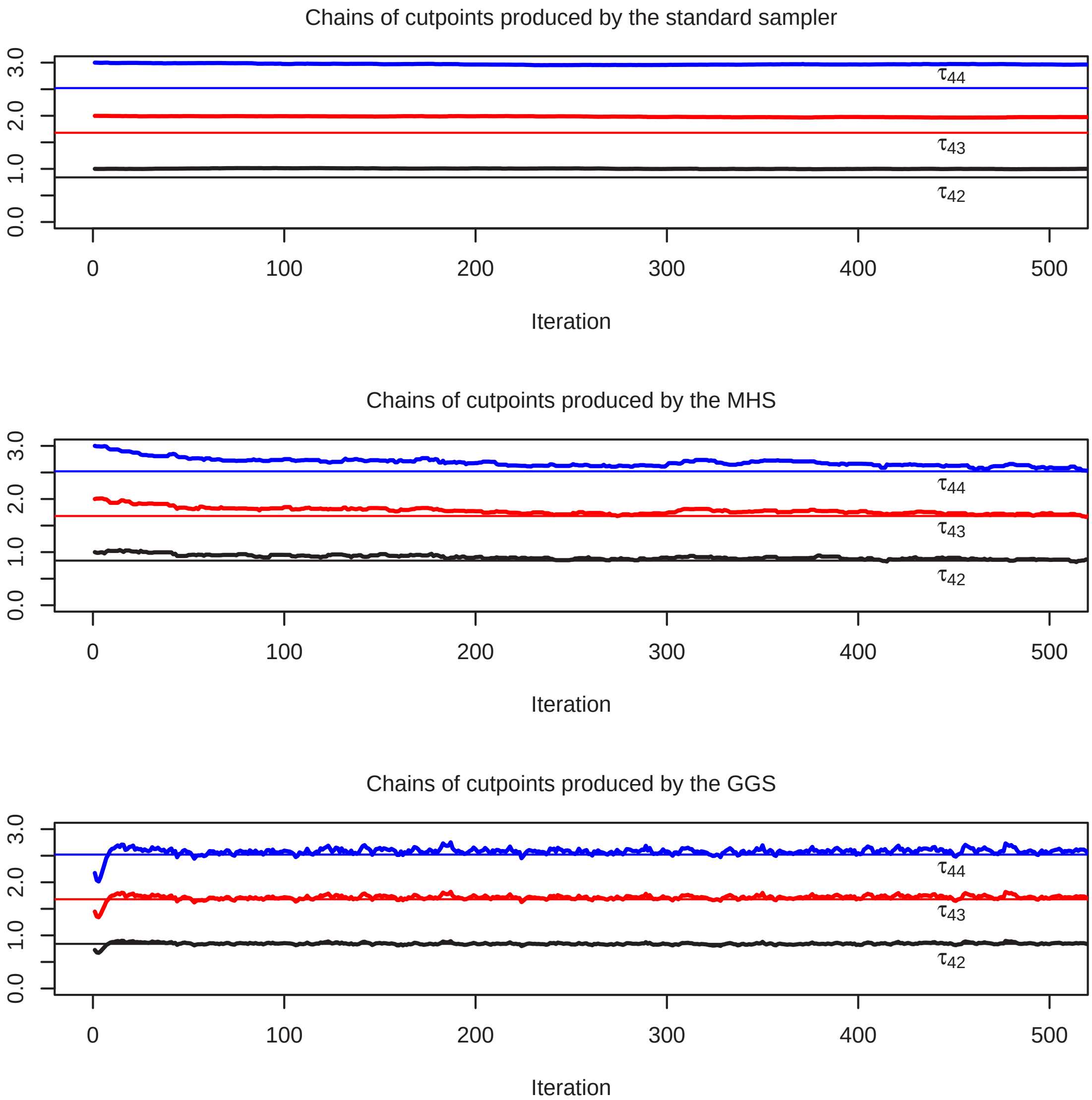


Figure 6.2: *First 500 sampling iterations of the cutpoints of indicator 4 (factor loading of 0.980) produced by the standard Gibbs sampler (top), the MHS (middle), and the GGS (bottom). The horizontal thin lines indicate the true values.*

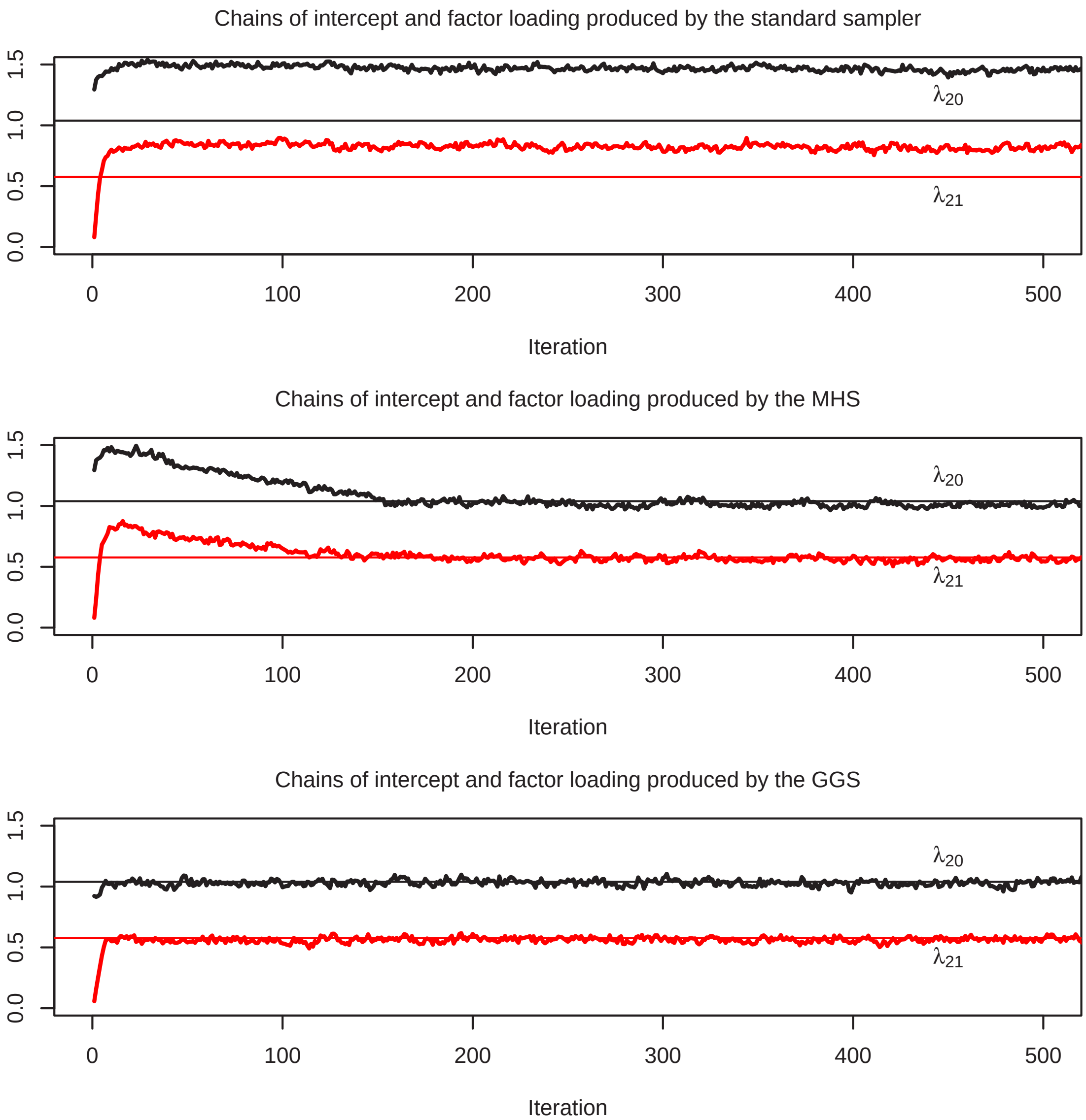


Figure 6.3: *First 500 sampling iterations of the intercept and factor loading of indicator 2 (factor loading of 0.577) produced by the standard Gibbs sampler (top), the MHS (middle), and the GGS (bottom). The horizontal thin lines indicate the true values.*

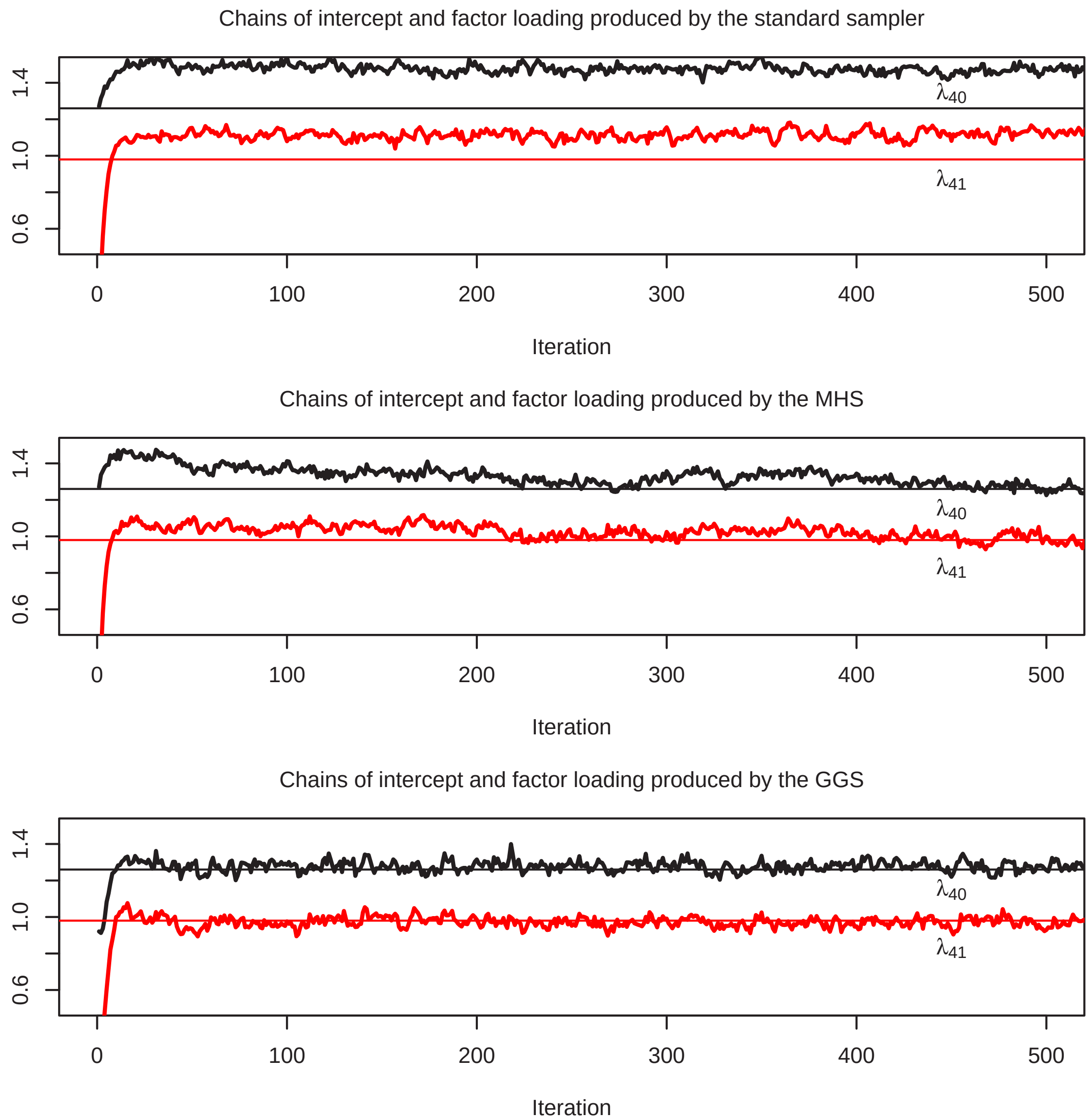


Figure 6.4: *First 500 sampling iterations of the intercept and factor loading of indicator 4 (factor loading of 0.980) produced by the standard Gibbs sampler (top), the MHS (middle), and the GGS (bottom). The horizontal thin lines indicate the true values.*

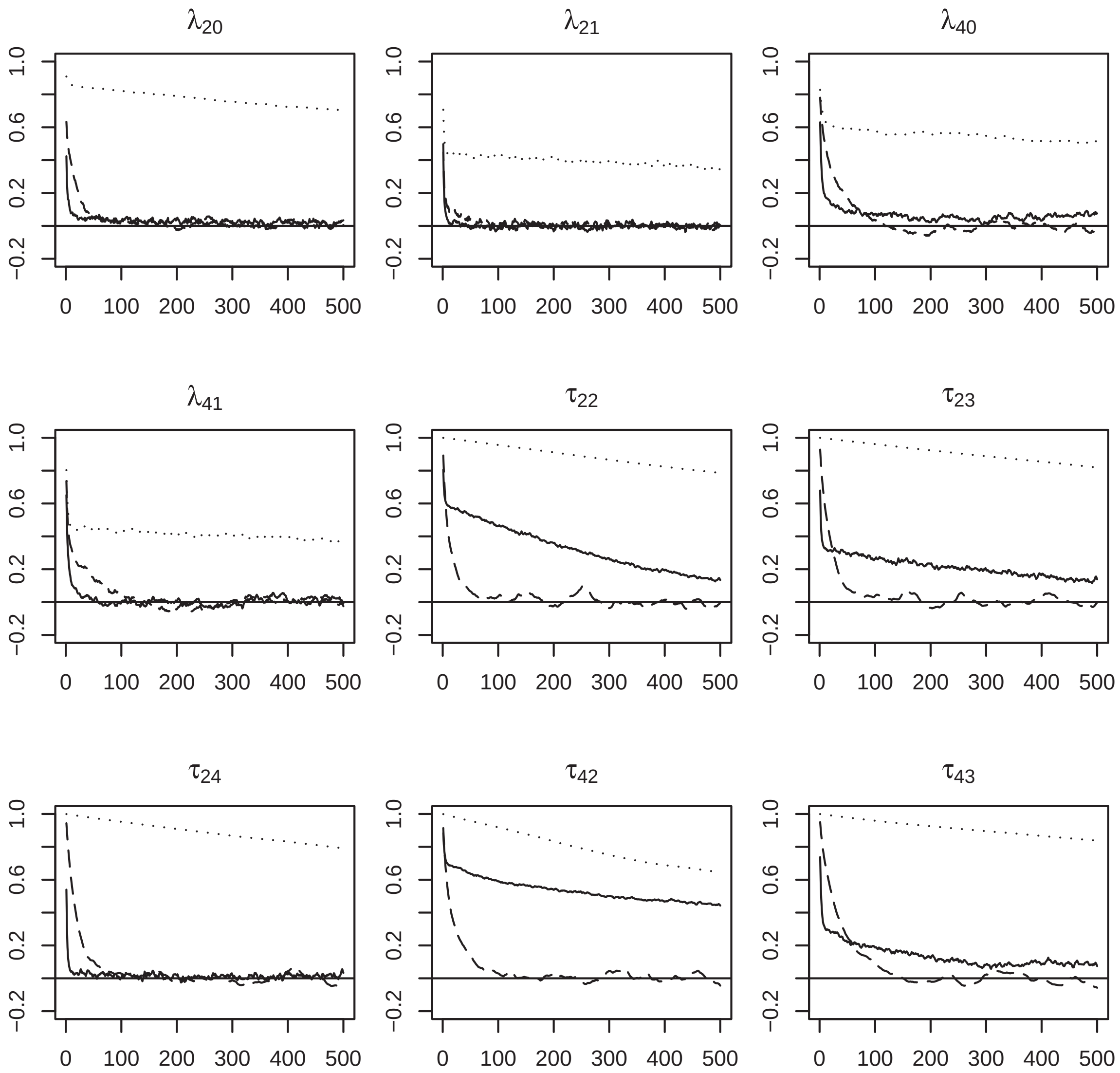


Figure 6.5: *Autocorrelations produced by the standard Gibbs sampler (dotted), the MHS (dashed), and the GGS (solid) for the intercepts, factor loadings, and cutpoints of indicators 2 and 4. The x-axis denotes the lag.*

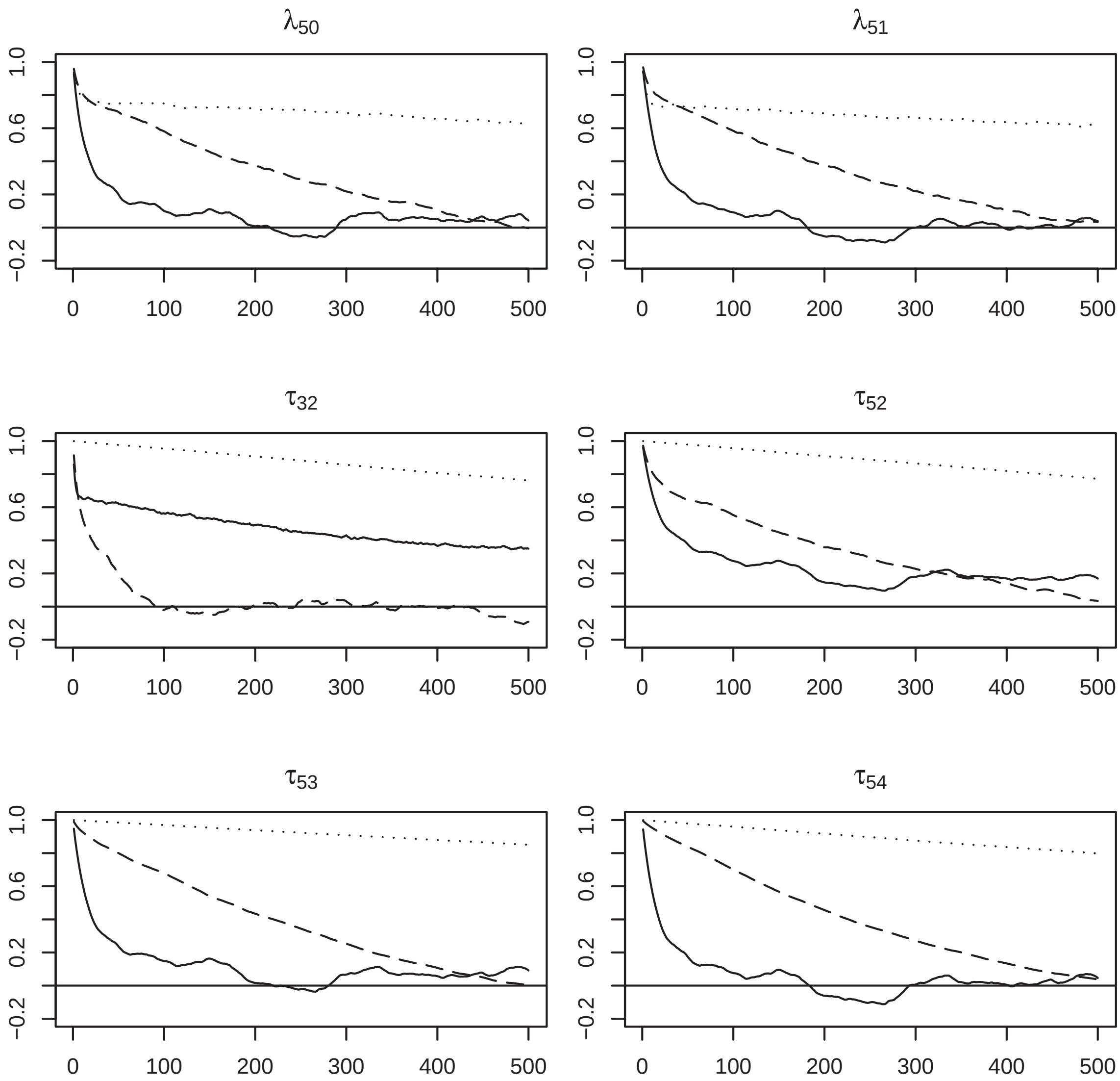


Figure 6.6: *Autocorrelations produced by the standard Gibbs sampler (dotted), the MHS (dashed), and the GGS (solid) for the intercepts, factor loadings, and cutpoints of indicators 2 and 4. The x-axis denotes the lag.*

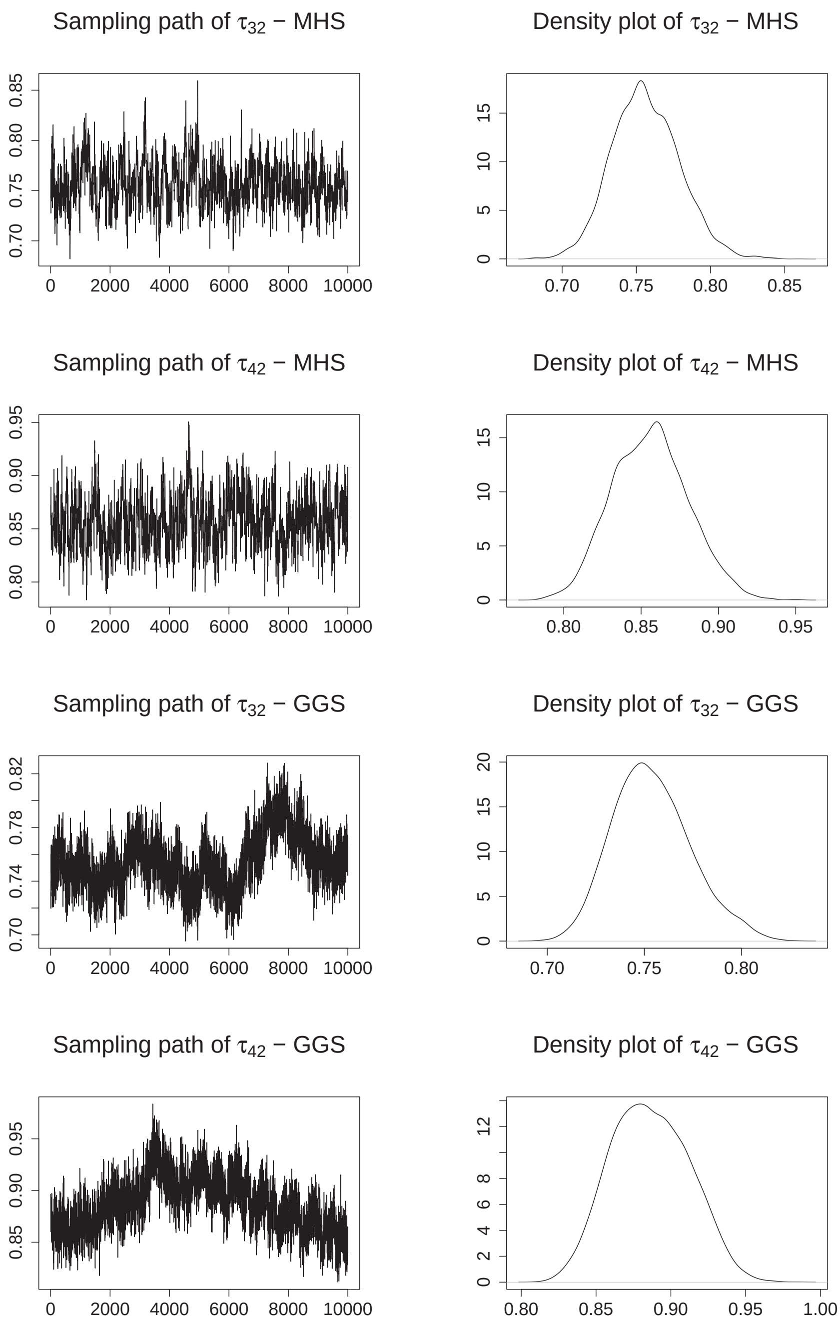


Figure 6.7: *Comparison of sampling paths (left column) and density plots (right column) of two selected cutpoint parameters for the MHS (upper two rows), and the GGS (bottom two rows).*

cutpoints, and is not observed for any of the other parameters which have good mixing sampling paths. The reason for the wavy sampling paths is not clear and requires further examination. Since the GGS still estimates the true parameter of the second cutpoints correctly, the corresponding density functions look fine, and the cutpoint parameters are typically of minor interest, this discrepancy of the GGS can be tolerated.

After having looked at the convergence properties of the three samplers, the estimation qualities of the MHS and the GGS with respect to the parameters of the measurement model are compared in the next section.

6.2 LVM excluding indirect effects

In order to demonstrate the estimation quality of the MHS and the GGS, simulations for five simulated parameter sets are carried out which cover the whole range of degrees of freedom in the measurement model. The simulated data sets vary along the following dimensions: number of observations, number of indicators, types of indicators (metric, binary, ordinal), frequency distribution of observations for binary and ordinal indicators, number of factors, prior information of factor loadings, values of factor loadings, and incorporation of fixed effects. Since the estimation of parameters of the measurement model and the structural part of the model are rather independent, indirect effects are not included at this stage but are examined in the next section. As parameter estimates for binary and ordinal indicators are very frequently used in applications and their estimation process is more intricate, our simulations focus on those types of indicators.

The general characteristics of the five parameter sets are layed out in Table 6.1. The column "Frequencies in categories" requires further explanation. In the case of equal frequencies, the probability of observation i choosing a certain category of indicator j is identical for all categories. For example, both categories of a binary indicator are selected by about 50% of the population, and the five categories of an ordinal indicator are occupied with about 20% of the population, each. For parameter sets B and D, different frequencies for

Dataset	**Indicators**	**Frequencies in categories**	**Factors**	**Factor loadings**	**Direct effects**
A	5 binary	similar	1	very different	no
B	5 binary	very different	1	identical	no
C	5 ordinal	similar	1	very different	no
D	5 ordinal	very different	1	identical	no
E	2 binary 6 ordinal 2 metric	similar similar –	2	very different	yes

Table 6.1: *Characteristics of the 5 parameter sets A, B, C, D and E.*

	j	λ_{j1}	σ_j^2	λ_{j0}	τ_{j2}	τ_{j3}	τ_{j4}
	1	0.314 (0.3)	1 (0.91)	0.0	-	-	-
	2	0.577 (0.5)	1 (0.75)	0.0	-	-	-
A	3	0.750 (0.6)	1 (0.64)	0.0	-	-	-
	4	0.980 (0.7)	1 (0.51)	0.0	-	-	-
	5	2.065 (0.9)	1 (0.19)	0.0	-	-	-
	1	0.980 (0.7)	1 (0.51)	-1.400 (-1.0)	-	-	-
	2	0.980 (0.7)	1 (0.51)	-0.980 (-0.7)	-	-	-
B	3	0.980 (0.7)	1 (0.51)	-0.420 (-0.3)	-	-	-
	4	0.980 (0.7)	1 (0.51)	0.700 (0.5)	-	-	-
	5	0.980 (0.7)	1 (0.51)	1.400 (1.0)	-	-	-
	1	0.314 (0.3)	1 (0.91)	0.943 (0.9)	0.629 (0.6)	1.258 (1.2)	1.887 (1.8)
	2	0.577 (0.5)	1 (0.75)	1.039 (0.9)	0.693 (0.6)	1.386 (1.2)	2.078 (1.8)
C	3	0.750 (0.6)	1 (0.64)	1.125 (0.9)	0.750 (0.6)	1.500 (1.2)	2.250 (1.8)
	4	0.980 (0.7)	1 (0.51)	1.260 (0.9)	0.840 (0.6)	1.680 (1.2)	2.521 (1.8)
	5	2.065 (0.9)	1 (0.19)	2.065 (0.9)	1.376 (0.6)	2.753 (1.2)	4.129 (1.8)
	1	0.980 (0.7)	1 (0.51)	0.840 (0.6)	1.540 (1.1)	2.380 (1.7)	3.220 (2.3)
	2	0.980 (0.7)	1 (0.51)	0.840 (0.6)	1.540 (1.1)	2.380 (1.7)	3.220 (2.3)
D	3	0.980 (0.7)	1 (0.51)	0.840 (0.6)	1.540 (1.1)	2.380 (1.7)	3.220 (2.3)
	4	0.980 (0.7)	1 (0.51)	2.380 (1.7)	0.840 (0.6)	1.680 (1.2)	3.220 (2.3)
	5	0.980 (0.7)	1 (0.51)	2.380 (1.7)	0.840 (0.6)	1.680 (1.2)	3.220 (2.3)

Table 6.2: *Parameter values of the parameter sets A, B, C and D (standardized values depicted in parentheses).*

the categories are chosen; for parameter set B, some binary indicators only have a 15% response rate for one of the two categories; for parameter set D, one category has about 40% of the response, while the lowest occupation rate is a response rate of about 5% for all of the ordinal indicators. The simulated data is generated in the same way as shown in Equations (6.1) but with different parameter settings. Detailed parameter specifications for parameter sets A–D and E can be found in Tables 6.2 and 6.3, respectively.

For each parameter set 9 simulation runs are conducted. Firstly, we use three different numbers of observations ($N_1 = 300$, $N_2 = 1000$, $N_3 = 5000$). Secondly, as explained in Section 5.2.1, for each number of observations three different priori strengths on the factor loadings are employed, in order to examine the appearance of Heywood cases in more detail; thus independent normal priors with $\mu_{weak} = \mu_{std} = \mu_{strong} = 0$, and $\sigma_{weak} = \sqrt{2}$, $\sigma_{std} = 1$, $\sigma_{strong} = 0.5$ are employed for the factor loadings. The burnin phase consists of 5000 iterations, and the sampling phase contains 10,000 iterations for simulations of parameter sets A, B, C, and D. We have also performed simulations with less iterations for both the burnin and the sampling phase, and results show that parameter estimates are equal if no Heywood cases occur. For that reason, the burnin phase is set to 2,000 iterations, and the sampling phase to 5,000 iterations for simulations of parameter set E.

	j	λ_{j1}	λ_{j2}	σ_j^2	λ_{j0}	τ_{j2}	τ_{j3}	τ_{j4}	τ_{j5}	a_{j1}	a_{j2}
E	1	1.33 (0.80)	0.0 (0.0)	1.00 (0.36)	-0.50 (-0.30)	–	–	–	–	0.83 (0.50)	-0.83 (-0.50)
	2	0.0 (0.0)	1.33 (0.8)	1.00 (0.36)	0.50 (0.3)	–	–	–	–	-0.83 (-0.5)	0.83 (0.5)
	3	1.33 (0.8)	0.0 (0.0)	1.00 (0.36)	1.17 (0.7)	2.33 (1.4)	–	–	–	0.83 (0.5)	-0.83 (-0.5)
	4	0.71 (0.5)	0.71 (0.5)	1.00 (0.50)	0.99 (0.7)	1.98 (1.4)	–	–	–	-0.71 (-0.5)	0.71 (0.5)
	5	0.0 (0.0)	1.33 (0.8)	1.00 (0.36)	1.17 (0.7)	2.33 (1.4)	–	–	–	0.83 (0.5)	-0.83 (-0.5)
	6	1.33 (0.8)	0.0 (0.0)	1.00 (0.36)	0.83 (0.5)	0.83 (0.5)	1.83 (1.1)	2.83 (1.7)	3.83 (2.3)	-0.83 (-0.5)	0.83 (0.5)
	7	0.71 (0.5)	0.71 (0.5)	1.00 (0.50)	0.71 (0.5)	0.71 (0.5)	1.56 (1.1)	2.40 (1.7)	3.25 (2.3)	0.71 (0.5)	-0.71 (-0.5)
	8	0.0 (0.0)	1.33 (0.8)	1.00 (0.36)	0.83 (0.5)	0.83 (0.5)	1.83 (1.1)	2.83 (1.7)	3.83 (2.3)	-0.83 (-0.5)	0.83 (0.5)
	9	3.00 (0.60)	0.0 (0.0)	4.00 (0.80)	2.0 (0.4)	–	–	–	–	0.5 (0.1)	-0.5 (-0.1)
	10	0.0 (0.0)	4.0 (0.80)	3.00 (0.60)	-5.0 (-1.0)	–	–	–	–	-0.5 (-0.1)	0.5 (0.1)

Table 6.3: *Parameter values of the parameter set E (standardized values depicted in parentheses).*

In order to assess the quality of parameter estimation and the occurrence of Heywood cases, $S = 100$ different data sets are simulated for each of the three different numbers of observations per parameter set. After the simulation of 100 different data sets, characteristic numbers can be calculated to indicate the quality of the estimation process. Let θ denote one specific but arbitrary parameter, and let θ^{true} indicate its true value. $\hat{\theta}_s^{mean}$ denotes the mean, and $\hat{\theta}_s^{std}$ the standard error of the respective parameter estimate for data set s (e. g. see Equation 4.6). Now we can calculate the estimated mean and standard deviation of parameter θ of all $S = 100$ simulation runs according to

$$\text{MEAN}_\theta = \frac{1}{S}\sum_{s=1}^{S}\hat{\theta}_s^{mean} \text{ , and } \quad \text{STD}_\theta = \frac{1}{S}\sum_{s=1}^{S}\hat{\theta}_s^{std} \, . \tag{6.2}$$

In a similar way, we define the bias and the mean squared error (MSE) indicating the deviation of the estimated value from its true value, i. e.

$$\text{BIAS}_\theta = \frac{1}{S}\sum_{s=1}^{S}(\hat{\theta}_s^{mean} - \theta^{true}) \text{ , and } \quad \text{MSE}_\theta = \frac{1}{S}\sum_{s=1}^{S}(\hat{\theta}_s^{mean} - \theta^{true})^2 \, . \tag{6.3}$$

Furthermore, the 95% coverage is obtained by counting the simulation runs for which the true parameter θ^{true} is located in the respective 95% credible region. Hence the coverage should be in the region of around 95 for all parameters when 100 data sets are estimated in one simulation run. So five characteristic numbers (MEAN, STD, BIAS, MSE, coverage) are calculated for all 9 simulation runs, each of which consists of the simulation of 100 different data sets.

All simulations are run for the MHS and the GGS to compare the estimation properties of the two MCMC schemes. The MEAN, STD, BIAS, the natural logarithm of the MSE, and the coverage for parameter set A can be found in Tables 6.4 and 6.5; results of parameter set D are given in Tables 6.6, 6.7, 6.8, 6.9, and 6.10. Simulation results of the other three parameter sets B, C, and E are outlined in appendix C.

We start the discussion by examining the parameter estimates of the MHS which is identical to the standard sampler for pure binary and metric indicators. For binary data, parameter estimates are very satisfactory. If the factor loadings are very different (parameter sets A and C), the MSE is lower for parameters associated with low factor loadings as expected. In general, the MSE decreases and thus the estimation quality improves for an increasing number of observations. If the parameter set contains a very high factor loading (parameter sets A and C), the strong priori information can influence the estimates for parameters belonging to the high factor loading in a negative way, resulting in parameter estimates which are too low compared to their true values. If none of the factor loadings is very high (parameter sets B, D and E), priori information is of minor importance which is also true for data sets with a high number of observations. For that reasons, the standard prior for factor loadings is generally a good choice for the MHS. Regarding the cutpoints, parameter estimates are inferior for a low number of observations ($N_1 = 300$), but improve fast for a higher number of observations. This behaviour is very similar to the estimation of cutpoint parameters in a standard ordinal regression setting. In general, occupation rates do not exert a significant influence on the quality of parameter estimates, but if some occupation rates are below a certain percentage, parameter estimates (especially concerning the cutpoints) can be problematic for data sets with a low number of observations. This behaviour manifests itself in high fluctuations of the respective sampling paths. One possible solution is to merge two adjacent categories so that sufficient observations fall into the resulting category.

Discussing the properties of the GGS, we focus on the differences compared to the properties of the MHS. Examining the coverage and the logarithm of the MSE, it becomes apparent that the parameter estimates of the GGS are very poor for data sets with a low number of observations, and for simulations with weak prior information. The estimates become better for a medium number of observations ($N_2 = 1000$), and are virtually equal to the MHS for a high number of observations ($N_3 = 5000$). The explanation for this phenomenon is the Heywood case which appears much more often for the GGS than for the MHS. In a Heywood case, one factor loading parameter (usually the one with the highest or second-highest value) diverges and tends to go to infinity for ordinal indicators (see appendix A for more information on Heywood cases). As mentioned in Section 5.2.1, a proper priori has to be applied in order to prevent Heywood cases. As we see, the prob-

Simulations of parameter set A – standard sampler/MHS									
N	$N_1 = 300$			$N_2 = 1000$			$N_3 = 5000$		
Priori	Weak	Std.	Strong	Weak	Std.	Strong	Weak	Std.	Strong
					MEAN				
λ_{10}	-0.012	-0.012	-0.012	-0.006	-0.006	-0.006	0.000	0.001	0.000
λ_{20}	-0.003	-0.004	-0.004	-0.008	-0.008	-0.008	-0.001	-0.001	-0.001
λ_{30}	-0.001	-0.001	-0.001	0.008	0.008	0.008	0.002	0.002	0.002
λ_{40}	0.002	0.001	0.000	0.003	0.003	0.003	-0.002	-0.002	-0.002
λ_{50}	0.007	0.006	0.005	-0.003	-0.002	-0.001	-0.005	-0.005	-0.004
λ_{11}	0.327	0.309	0.308	0.311	0.311	0.316	0.316	0.317	0.317
λ_{21}	0.611	0.571	0.566	0.580	0.582	0.592	0.575	0.576	0.577
λ_{31}	0.767	0.719	0.699	0.759	0.762	0.773	0.749	0.750	0.753
λ_{41}	1.057	0.991	0.924	0.966	0.973	0.990	0.979	0.982	0.990
λ_{51}	2.070	1.653	1.290	2.117	1.933	1.627	2.123	2.072	1.907
					STD				
λ_{10}	0.076	0.077	0.076	0.043	0.043	0.043	0.019	0.019	0.019
λ_{20}	0.082	0.083	0.082	0.045	0.045	0.045	0.022	0.022	0.022
λ_{30}	0.089	0.089	0.088	0.046	0.046	0.046	0.022	0.022	0.022
λ_{40}	0.098	0.098	0.094	0.055	0.055	0.056	0.024	0.024	0.024
λ_{50}	0.177	0.154	0.125	0.106	0.099	0.084	0.042	0.042	0.039
λ_{11}	0.110	0.151	0.128	0.083	0.083	0.055	0.027	0.027	0.027
λ_{21}	0.163	0.269	0.225	0.138	0.138	0.063	0.027	0.026	0.026
λ_{31}	0.194	0.331	0.292	0.164	0.165	0.080	0.036	0.036	0.036
λ_{41}	0.325	0.464	0.386	0.220	0.222	0.088	0.046	0.046	0.045
λ_{51}	0.661	0.846	0.504	0.594	0.483	0.136	0.207	0.184	0.123
					BIAS				
λ_{10}	-0.012	-0.012	-0.012	-0.006	-0.006	-0.006	0.000	0.001	0.000
λ_{20}	-0.003	-0.004	-0.004	-0.008	-0.008	-0.008	-0.001	-0.001	-0.001
λ_{30}	-0.001	-0.001	-0.001	0.008	0.008	0.008	0.002	0.002	0.002
λ_{40}	0.002	0.001	0.000	0.003	0.003	0.003	-0.002	-0.002	-0.002
λ_{50}	0.007	0.006	0.005	-0.003	-0.002	-0.001	-0.005	-0.005	-0.004
λ_{11}	0.012	-0.005	-0.006	-0.004	-0.004	0.001	0.002	0.002	0.003
λ_{21}	0.033	-0.006	-0.011	0.003	0.004	0.015	-0.002	-0.002	0.000
λ_{31}	0.017	-0.031	-0.051	0.009	0.012	0.023	-0.001	0.000	0.003
λ_{41}	0.077	0.011	-0.056	-0.014	-0.007	0.010	-0.001	0.002	0.009
λ_{51}	0.005	-0.412	-0.775	0.053	-0.132	-0.437	0.058	0.008	-0.157
					ln(MSE)				
λ_{10}	-5.129	-5.121	-5.132	-6.279	-6.289	-6.281	-7.934	-7.931	-7.938
λ_{20}	-5.005	-4.995	-5.006	-6.192	-6.204	-6.181	-7.682	-7.681	-7.681
λ_{30}	-4.853	-4.852	-4.864	-6.158	-6.147	-6.158	-7.629	-7.620	-7.616
λ_{40}	-4.655	-4.651	-4.729	-5.801	-5.794	-5.780	-7.440	-7.430	-7.431
λ_{50}	-3.470	-3.744	-4.171	-4.495	-4.641	-4.965	-6.326	-6.347	-6.492
λ_{11}	-4.415	-3.793	-4.123	-4.983	-4.984	-5.805	-7.263	-7.259	-7.253
λ_{21}	-3.601	-2.639	-2.995	-3.964	-3.964	-5.496	-7.260	-7.275	-7.284
λ_{31}	-3.279	-2.212	-2.443	-3.626	-3.604	-4.976	-6.652	-6.658	-6.676
λ_{41}	-2.201	-1.543	-1.895	-3.030	-3.015	-4.855	-6.166	-6.171	-6.179
λ_{51}	-0.839	-0.130	-0.161	-1.045	-1.395	-1.562	-3.087	-3.399	-3.223
					Coverage				
λ_{10}	93	93	94	93	93	93	96	95	95
λ_{20}	96	96	96	97	96	96	94	95	94
λ_{30}	98	98	98	95	96	95	96	96	96
λ_{40}	96	96	96	93	92	93	97	97	97
λ_{50}	95	95	95	96	95	96	93	92	93
λ_{11}	95	92	94	96	96	96	94	94	94
λ_{21}	95	92	95	95	96	96	94	94	94
λ_{31}	98	95	96	97	97	98	94	94	94
λ_{41}	94	92	95	97	98	99	94	95	93
λ_{51}	97	90	16	98	94	47	96	95	81

Table 6.4: *Estimates of parameter set A simulated by the standard sampler and MHS – MEAN, STD, BIAS, ln(MSE), and coverage of estimated parameters obtained by simulations of 100 different data sets.*

Simulations of parameter set A – GGS									
N	$N_1 = 300$			$N_2 = 1000$			$N_3 = 5000$		
Priori	Weak	Std.	Strong	Weak	Std.	Strong	Weak	Std.	Strong
	MEAN								
λ_{10}	0.014	-0.009	-0.012	-0.006	-0.005	-0.006	0.000	0.000	0.000
λ_{20}	0.015	-0.002	-0.004	-0.008	-0.006	-0.008	-0.001	-0.001	-0.001
λ_{30}	0.038	0.007	-0.001	0.009	0.010	0.008	0.002	0.002	0.002
λ_{40}	11.504	0.015	0.001	0.004	0.005	0.003	-0.002	-0.002	-0.002
λ_{50}	27.874	1.026	0.009	0.867	0.421	-0.004	-0.005	-0.005	-0.005
λ_{11}	0.161	0.278	0.299	0.299	0.307	0.315	0.316	0.316	0.316
λ_{21}	0.280	0.492	0.549	0.549	0.570	0.588	0.575	0.574	0.575
λ_{31}	0.334	0.609	0.691	0.708	0.734	0.766	0.748	0.748	0.749
λ_{41}	3.040	1.214	0.933	0.885	0.923	0.981	0.976	0.976	0.979
λ_{51}	110.638	23.076	1.753	66.828	19.624	2.079	2.205	2.198	2.091
	STD								
λ_{10}	0.235	0.125	0.077	0.065	0.054	0.043	0.019	0.019	0.019
λ_{20}	0.369	0.194	0.083	0.102	0.074	0.045	0.022	0.022	0.022
λ_{30}	0.432	0.222	0.091	0.119	0.085	0.046	0.022	0.022	0.022
λ_{40}	103.154	0.564	0.100	0.148	0.106	0.056	0.024	0.024	0.024
λ_{50}	553.997	88.854	0.189	49.938	10.439	0.108	0.044	0.044	0.042
λ_{11}	0.099	0.138	0.155	0.058	0.056	0.055	0.027	0.027	0.027
λ_{21}	0.154	0.233	0.278	0.075	0.069	0.063	0.026	0.027	0.026
λ_{31}	0.180	0.289	0.343	0.093	0.088	0.079	0.036	0.036	0.036
λ_{41}	20.006	2.329	0.495	0.127	0.108	0.091	0.048	0.048	0.045
λ_{51}	63.030	17.529	1.016	49.171	18.549	0.412	0.282	0.279	0.196
	BIAS								
λ_{10}	0.014	-0.009	-0.012	-0.006	-0.005	-0.006	0.000	0.000	0.000
λ_{20}	0.015	-0.002	-0.004	-0.008	-0.006	-0.008	-0.001	-0.001	-0.001
λ_{30}	0.038	0.007	-0.001	0.009	0.010	0.008	0.002	0.002	0.002
λ_{40}	11.504	0.015	0.001	0.004	0.005	0.003	-0.002	-0.002	-0.002
λ_{50}	27.874	1.026	0.009	0.867	0.421	-0.004	-0.005	-0.005	-0.005
λ_{11}	-0.153	-0.036	-0.016	-0.015	-0.007	0.000	0.002	0.002	0.002
λ_{21}	-0.298	-0.086	-0.029	-0.028	-0.007	0.011	-0.003	-0.003	-0.003
λ_{31}	-0.416	-0.141	-0.059	-0.042	-0.016	0.016	-0.002	-0.002	-0.001
λ_{41}	2.059	0.233	-0.047	-0.095	-0.058	0.000	-0.004	-0.004	-0.001
λ_{51}	108.573	21.012	-0.312	64.764	17.559	0.015	0.140	0.133	0.027
	ln(MSE)								
λ_{10}	-2.902	-4.163	-5.110	-5.454	-5.839	-6.278	-7.936	-7.925	-7.938
λ_{20}	-2.004	-3.289	-4.982	-4.578	-5.222	-6.179	-7.684	-7.684	-7.679
λ_{30}	-1.682	-3.016	-4.800	-4.257	-4.934	-6.145	-7.628	-7.622	-7.628
λ_{40}	9.275	-1.156	-4.611	-3.830	-4.503	-5.784	-7.430	-7.439	-7.437
λ_{50}	12.627	8.964	-3.340	7.812	4.683	-4.457	-6.254	-6.258	-6.328
λ_{11}	-3.407	-3.906	-3.734	-5.635	-5.771	-5.813	-7.261	-7.265	-7.260
λ_{21}	-2.190	-2.796	-2.563	-5.053	-5.348	-5.496	-7.262	-7.253	-7.275
λ_{31}	-1.585	-2.281	-2.123	-4.585	-4.833	-5.045	-6.637	-6.648	-6.674
λ_{41}	5.993	1.691	-1.409	-3.683	-4.205	-4.795	-6.093	-6.093	-6.189
λ_{51}	9.663	6.614	0.113	8.793	6.475	-1.781	-2.317	-2.359	-3.253
	Coverage								
λ_{10}	44	86	94	86	88	93	95	95	95
λ_{20}	24	84	96	76	91	96	95	95	94
λ_{30}	19	83	98	76	91	95	96	96	96
λ_{40}	17	78	96	70	83	92	97	97	97
λ_{50}	12	67	94	47	73	95	93	93	93
λ_{11}	45	88	92	89	94	97	94	94	94
λ_{21}	28	89	92	82	96	97	94	94	94
λ_{31}	28	83	93	83	91	97	94	95	94
λ_{41}	25	79	91	64	81	99	93	94	93
λ_{51}	2	30	88	42	58	96	93	94	98

Table 6.5: *Estimates of parameter set A simulated by the GGS – MEAN, STD, BIAS, ln(MSE), and coverage of estimated parameters obtained by simulations of 100 different data sets.*

Simulations of parameter set D – MHS									
N	$N_1 = 300$			$N_2 = 1000$			$N_3 = 5000$		
Priori	**Weak**	**Std.**	**Strong**	**Weak**	**Std.**	**Strong**	**Weak**	**Std.**	**Strong**
	MEAN								
λ_{10}	0.862	0.863	0.850	0.839	0.840	0.836	0.843	0.843	0.843
λ_{20}	0.854	0.855	0.846	0.854	0.852	0.850	0.846	0.845	0.845
λ_{30}	0.850	0.848	0.837	0.846	0.844	0.842	0.842	0.842	0.842
λ_{40}	2.431	2.421	2.412	2.409	2.400	2.394	2.381	2.381	2.378
λ_{50}	2.430	2.441	2.410	2.395	2.395	2.386	2.396	2.395	2.395
λ_{11}	1.033	1.025	0.969	0.984	0.983	0.966	0.984	0.983	0.980
λ_{21}	1.025	1.019	0.968	1.001	0.996	0.982	0.984	0.984	0.981
λ_{31}	1.005	0.995	0.944	0.998	0.994	0.979	0.979	0.980	0.976
λ_{41}	1.001	0.989	0.948	1.004	0.999	0.984	0.979	0.979	0.976
λ_{51}	1.002	0.997	0.948	0.984	0.981	0.967	0.984	0.983	0.980
τ_{12}	1.584	1.587	1.563	1.542	1.543	1.536	1.540	1.540	1.540
τ_{13}	2.440	2.441	2.406	2.382	2.383	2.374	2.386	2.385	2.385
τ_{14}	3.334	3.333	3.283	3.235	3.235	3.222	3.229	3.227	3.227
τ_{22}	1.568	1.569	1.550	1.569	1.565	1.561	1.545	1.544	1.544
τ_{23}	2.421	2.423	2.395	2.413	2.408	2.402	2.387	2.386	2.385
τ_{24}	3.316	3.321	3.285	3.281	3.274	3.265	3.233	3.231	3.230
τ_{32}	1.580	1.576	1.555	1.548	1.545	1.541	1.538	1.538	1.537
τ_{33}	2.426	2.419	2.388	2.412	2.407	2.401	2.374	2.375	2.373
τ_{34}	3.302	3.297	3.257	3.272	3.265	3.255	3.218	3.219	3.217
τ_{42}	0.878	0.875	0.875	0.855	0.850	0.847	0.842	0.842	0.841
τ_{43}	1.726	1.719	1.717	1.712	1.705	1.700	1.679	1.678	1.676
τ_{44}	3.284	3.272	3.259	3.268	3.257	3.249	3.217	3.217	3.213
τ_{52}	0.882	0.889	0.877	0.860	0.861	0.856	0.847	0.846	0.847
τ_{53}	1.732	1.742	1.718	1.700	1.701	1.694	1.692	1.691	1.692
τ_{54}	3.276	3.289	3.249	3.235	3.234	3.223	3.234	3.233	3.233
	STD								
λ_{10}	0.149	0.150	0.141	0.066	0.065	0.065	0.028	0.027	0.028
λ_{20}	0.112	0.110	0.116	0.055	0.054	0.052	0.025	0.025	0.025
λ_{30}	0.104	0.107	0.101	0.066	0.066	0.067	0.029	0.030	0.029
λ_{40}	0.220	0.207	0.195	0.123	0.120	0.122	0.053	0.054	0.054
λ_{50}	0.231	0.249	0.231	0.103	0.106	0.100	0.047	0.049	0.047
λ_{11}	0.164	0.149	0.130	0.065	0.065	0.062	0.029	0.029	0.029
λ_{21}	0.114	0.118	0.104	0.068	0.067	0.064	0.032	0.033	0.033
λ_{31}	0.141	0.139	0.123	0.062	0.062	0.058	0.030	0.030	0.030
λ_{41}	0.132	0.132	0.118	0.075	0.074	0.071	0.029	0.029	0.029
λ_{51}	0.125	0.125	0.111	0.058	0.057	0.055	0.030	0.030	0.030
τ_{12}	0.178	0.178	0.167	0.069	0.069	0.069	0.036	0.036	0.037
τ_{13}	0.216	0.205	0.192	0.095	0.097	0.095	0.050	0.050	0.051
τ_{14}	0.318	0.299	0.278	0.137	0.133	0.132	0.059	0.059	0.060
τ_{22}	0.137	0.134	0.130	0.071	0.068	0.068	0.032	0.033	0.033
τ_{23}	0.180	0.183	0.179	0.093	0.090	0.087	0.042	0.044	0.043
τ_{24}	0.261	0.260	0.260	0.135	0.128	0.126	0.060	0.062	0.061
τ_{32}	0.140	0.136	0.125	0.068	0.068	0.067	0.033	0.034	0.033
τ_{33}	0.187	0.188	0.158	0.104	0.104	0.101	0.046	0.048	0.046
τ_{34}	0.245	0.249	0.231	0.136	0.139	0.134	0.061	0.062	0.060
τ_{42}	0.156	0.154	0.143	0.097	0.092	0.096	0.041	0.041	0.042
τ_{43}	0.193	0.178	0.171	0.108	0.103	0.105	0.048	0.049	0.050
τ_{44}	0.262	0.251	0.231	0.131	0.123	0.125	0.059	0.059	0.059
τ_{52}	0.194	0.195	0.183	0.079	0.083	0.077	0.040	0.041	0.041
τ_{53}	0.226	0.235	0.221	0.094	0.099	0.091	0.047	0.049	0.048
τ_{54}	0.283	0.302	0.275	0.122	0.125	0.116	0.055	0.057	0.055

Table 6.6: *Estimates of parameter set D simulated by the MHS – MEAN and STD of estimated parameters obtained by simulations of 100 different data sets.*

Simulations of parameter set D – MHS									
N	$N_1 = 300$			$N_2 = 1000$			$N_3 = 5000$		
Priori	Weak	Std.	Strong	Weak	Std.	Strong	Weak	Std.	Strong
	BIAS								
λ_{10}	0.021	0.023	0.010	-0.001	-0.001	-0.004	0.003	0.003	0.003
λ_{20}	0.014	0.015	0.006	0.014	0.012	0.010	0.005	0.005	0.005
λ_{30}	0.010	0.008	-0.003	0.006	0.004	0.002	0.002	0.002	0.001
λ_{40}	0.051	0.041	0.032	0.028	0.020	0.014	0.001	0.000	-0.002
λ_{50}	0.050	0.060	0.029	0.015	0.014	0.005	0.015	0.014	0.014
λ_{11}	0.053	0.045	-0.011	0.004	0.002	-0.014	0.003	0.003	-0.000
λ_{21}	0.045	0.039	-0.012	0.021	0.016	0.002	0.004	0.003	0.001
λ_{31}	0.025	0.014	-0.036	0.018	0.014	-0.001	-0.001	-0.001	-0.004
λ_{41}	0.021	0.009	-0.033	0.024	0.019	0.004	-0.001	-0.001	-0.005
λ_{51}	0.022	0.017	-0.032	0.004	0.001	-0.013	0.004	0.003	0.000
τ_{12}	0.044	0.046	0.023	0.002	0.002	-0.004	0.000	-0.001	-0.001
τ_{13}	0.059	0.061	0.026	0.002	0.003	-0.007	0.006	0.005	0.004
τ_{14}	0.113	0.113	0.062	0.014	0.014	0.002	0.008	0.007	0.006
τ_{22}	0.028	0.028	0.010	0.028	0.025	0.021	0.005	0.004	0.003
τ_{23}	0.041	0.043	0.014	0.033	0.028	0.022	0.007	0.006	0.005
τ_{24}	0.096	0.101	0.064	0.060	0.053	0.045	0.012	0.010	0.010
τ_{32}	0.040	0.035	0.014	0.008	0.005	0.001	-0.002	-0.002	-0.003
τ_{33}	0.045	0.038	0.008	0.031	0.027	0.020	-0.006	-0.006	-0.007
τ_{34}	0.081	0.076	0.036	0.051	0.044	0.034	-0.002	-0.001	-0.004
τ_{42}	0.038	0.035	0.035	0.015	0.010	0.007	0.002	0.002	0.001
τ_{43}	0.046	0.039	0.036	0.032	0.025	0.020	-0.002	-0.002	-0.004
τ_{44}	0.064	0.052	0.039	0.047	0.037	0.029	-0.004	-0.004	-0.007
τ_{52}	0.042	0.049	0.037	0.020	0.021	0.015	0.007	0.006	0.007
τ_{53}	0.051	0.062	0.038	0.020	0.021	0.013	0.012	0.011	0.011
τ_{54}	0.056	0.068	0.029	0.014	0.013	0.002	0.014	0.012	0.012
	ln(MSE)								
λ_{10}	-3.801	-3.782	-3.929	-5.438	-5.470	-5.459	-7.183	-7.191	-7.166
λ_{20}	-4.371	-4.398	-4.319	-5.745	-5.781	-5.900	-7.352	-7.324	-7.376
λ_{30}	-4.531	-4.468	-4.595	-5.450	-5.432	-5.424	-7.060	-7.038	-7.067
λ_{40}	-2.990	-3.124	-3.251	-4.147	-4.226	-4.209	-5.870	-5.864	-5.855
λ_{50}	-2.892	-2.733	-2.922	-4.531	-4.486	-4.608	-6.015	-5.978	-6.031
λ_{11}	-3.529	-3.729	-4.084	-5.485	-5.473	-5.513	-7.047	-7.070	-7.061
λ_{21}	-4.207	-4.183	-4.522	-5.292	-5.366	-5.505	-6.850	-6.817	-6.860
λ_{31}	-3.892	-3.944	-4.119	-5.486	-5.508	-5.695	-7.013	-7.004	-7.026
λ_{41}	-4.031	-4.048	-4.215	-5.090	-5.147	-5.290	-7.079	-7.103	-7.063
λ_{51}	-4.141	-4.151	-4.316	-5.706	-5.754	-5.749	-7.011	-6.985	-7.028
τ_{12}	-3.406	-3.396	-3.566	-5.363	-5.351	-5.360	-6.660	-6.673	-6.630
τ_{13}	-3.006	-3.099	-3.293	-4.714	-4.672	-4.722	-5.973	-6.001	-5.953
τ_{14}	-2.181	-2.293	-2.519	-3.977	-4.027	-4.066	-5.653	-5.662	-5.632
τ_{22}	-3.944	-3.983	-4.083	-5.144	-5.273	-5.307	-6.853	-6.805	-6.840
τ_{23}	-3.388	-3.354	-3.441	-4.650	-4.732	-4.824	-6.302	-6.233	-6.284
τ_{24}	-2.571	-2.564	-2.648	-3.833	-3.961	-4.031	-5.591	-5.539	-5.588
τ_{32}	-3.866	-3.929	-4.155	-5.378	-5.383	-5.407	-6.806	-6.759	-6.795
τ_{33}	-3.304	-3.310	-3.694	-4.445	-4.480	-4.560	-6.158	-6.087	-6.153
τ_{34}	-2.720	-2.701	-2.919	-3.868	-3.866	-3.961	-5.600	-5.559	-5.621
τ_{42}	-3.664	-3.700	-3.848	-4.652	-4.762	-4.693	-6.405	-6.375	-6.343
τ_{43}	-3.248	-3.414	-3.496	-4.373	-4.508	-4.481	-6.071	-6.041	-6.014
τ_{44}	-2.628	-2.733	-2.910	-3.949	-4.122	-4.117	-5.651	-5.661	-5.639
τ_{52}	-3.246	-3.215	-3.362	-5.035	-4.933	-5.100	-6.413	-6.364	-6.388
τ_{53}	-2.932	-2.841	-2.996	-4.694	-4.599	-4.786	-6.046	-5.993	-6.039
τ_{54}	-2.496	-2.353	-2.582	-4.209	-4.166	-4.320	-5.767	-5.707	-5.768

Table 6.7: *Estimates of parameter set D simulated by the MHS – BIAS and ln(MSE) of estimated parameters obtained by simulations of 100 different data sets.*

Simulations of parameter set D – MHS									
N	**$N_1 = 300$**			**$N_2 = 1000$**			**$N_3 = 5000$**		
Priori	**Weak**	**Std.**	**Strong**	**Weak**	**Std.**	**Strong**	**Weak**	**Std.**	**Strong**
	Coverage								
λ_{10}	92	92	93	91	91	91	92	93	92
λ_{20}	96	93	93	99	99	100	97	95	96
λ_{30}	98	96	97	95	93	93	95	94	95
λ_{40}	90	94	92	96	94	93	94	95	93
λ_{50}	89	88	88	96	96	95	94	95	96
λ_{11}	90	91	91	95	94	95	94	94	94
λ_{21}	96	97	98	92	93	97	93	94	93
λ_{31}	95	90	89	97	96	97	96	94	96
λ_{41}	89	93	92	91	90	89	98	97	97
λ_{51}	93	92	92	97	98	96	95	96	95
τ_{12}	85	88	85	96	96	95	92	92	92
τ_{13}	94	92	92	94	92	94	93	91	90
τ_{14}	89	88	91	89	94	93	94	95	94
τ_{22}	94	94	94	96	97	98	95	96	97
τ_{23}	94	91	90	94	97	98	95	96	96
τ_{24}	94	94	91	92	96	95	92	92	93
τ_{32}	92	93	93	94	95	95	92	92	92
τ_{33}	90	90	94	92	93	93	93	93	94
τ_{34}	93	93	95	91	94	92	92	93	94
τ_{42}	92	94	96	90	92	93	94	95	92
τ_{43}	90	91	93	94	95	95	94	95	95
τ_{44}	91	88	93	93	93	97	97	96	95
τ_{52}	86	85	91	96	95	97	94	93	93
τ_{53}	81	83	83	97	96	95	94	94	93
τ_{54}	88	84	87	94	97	94	95	97	95

Simulations of parameter set D – GGS									
N	**$N_1 = 300$**			**$N_2 = 1000$**			**$N_3 = 5000$**		
Priori	**Weak**	**Std.**	**Strong**	**Weak**	**Std.**	**Strong**	**Weak**	**Std.**	**Strong**
	Coverage								
λ_{10}	94	93	93	90	91	91	92	94	92
λ_{20}	94	94	92	99	99	99	96	96	96
λ_{30}	95	95	95	94	94	93	94	94	95
λ_{40}	96	95	94	94	95	94	93	94	94
λ_{50}	96	94	93	94	96	96	88	90	91
λ_{11}	92	92	92	94	94	94	94	95	95
λ_{21}	97	96	97	94	95	96	94	95	93
λ_{31}	95	94	92	96	97	98	95	95	95
λ_{41}	91	90	94	90	90	88	96	96	97
λ_{51}	94	94	95	97	98	97	95	97	95
τ_{12}	90	90	90	97	97	97	87	90	89
τ_{13}	96	96	96	95	94	95	91	93	94
τ_{14}	90	92	92	92	93	94	94	94	95
τ_{22}	95	93	94	98	98	97	97	93	97
τ_{23}	93	92	94	96	96	97	95	95	94
τ_{24}	95	94	96	96	96	96	92	93	93
τ_{32}	95	94	95	95	94	96	93	93	92
τ_{33}	97	96	97	93	94	95	94	94	94
τ_{34}	97	97	99	94	93	93	94	93	93
τ_{42}	95	96	94	92	91	91	82	83	82
τ_{43}	97	98	98	92	93	95	88	85	87
τ_{44}	98	97	96	93	93	93	96	95	96
τ_{52}	89	88	90	94	96	96	81	81	80
τ_{53}	87	88	89	94	97	96	76	87	81
τ_{54}	89	88	91	97	98	98	93	96	95

Table 6.8: *Estimates of parameter set D simulated by the MHS (top) and the GGS (bottom) – coverage of estimated parameters obtained by simulations of 100 different data sets.*

Simulations of parameter set D – GGS									
N	$N_1 = 300$			$N_2 = 1000$			$N_3 = 5000$		
Priori	**Weak**	**Std.**	**Strong**	**Weak**	**Std.**	**Strong**	**Weak**	**Std.**	**Strong**
	MEAN								
λ_{10}	9.131	35.676	0.881	0.843	0.842	0.843	0.844	0.844	0.844
λ_{20}	0.870	0.871	0.874	0.856	0.855	0.857	0.846	0.846	0.846
λ_{30}	0.861	0.863	0.866	0.849	0.848	0.850	0.843	0.843	0.843
λ_{40}	2.464	2.466	2.477	2.414	2.412	2.415	2.392	2.388	2.391
λ_{50}	2.475	2.474	2.484	2.403	2.400	2.401	2.407	2.403	2.407
λ_{11}	2.496	3.584	0.996	0.987	0.984	0.972	0.984	0.984	0.981
λ_{21}	1.035	1.023	0.989	1.002	0.998	0.986	0.985	0.984	0.982
λ_{31}	1.009	0.999	0.966	1.001	0.998	0.985	0.980	0.980	0.977
λ_{41}	1.006	0.995	0.963	1.005	1.002	0.990	0.980	0.979	0.977
λ_{51}	1.011	1.000	0.966	0.986	0.983	0.971	0.984	0.983	0.981
τ_{12}	10.506	37.597	1.615	1.548	1.547	1.548	1.542	1.542	1.541
τ_{13}	12.202	40.249	2.475	2.390	2.388	2.389	2.388	2.387	2.387
τ_{14}	14.246	43.072	3.362	3.243	3.239	3.238	3.230	3.229	3.230
τ_{22}	1.590	1.591	1.598	1.573	1.570	1.572	1.545	1.545	1.546
τ_{23}	2.453	2.451	2.456	2.417	2.414	2.415	2.388	2.388	2.388
τ_{24}	3.352	3.349	3.347	3.283	3.278	3.278	3.235	3.234	3.234
τ_{32}	1.597	1.597	1.603	1.554	1.552	1.555	1.538	1.540	1.539
τ_{33}	2.444	2.444	2.450	2.417	2.416	2.417	2.377	2.376	2.375
τ_{34}	3.325	3.323	3.323	3.278	3.274	3.272	3.220	3.219	3.220
τ_{42}	0.900	0.903	0.921	0.858	0.857	0.862	0.854	0.850	0.854
τ_{43}	1.756	1.759	1.779	1.717	1.716	1.722	1.693	1.689	1.692
τ_{44}	3.324	3.324	3.339	3.275	3.272	3.274	3.226	3.223	3.225
τ_{52}	0.912	0.913	0.928	0.865	0.863	0.866	0.860	0.856	0.861
τ_{53}	1.773	1.774	1.791	1.708	1.706	1.708	1.707	1.702	1.708
τ_{54}	3.330	3.328	3.340	3.245	3.241	3.243	3.244	3.240	3.244
	STD								
λ_{10}	82.621	348.063	0.153	0.066	0.066	0.066	0.028	0.028	0.028
λ_{20}	0.114	0.114	0.114	0.053	0.053	0.053	0.025	0.025	0.025
λ_{30}	0.104	0.105	0.105	0.066	0.066	0.066	0.030	0.029	0.030
λ_{40}	0.206	0.208	0.210	0.122	0.123	0.122	0.054	0.055	0.054
λ_{50}	0.240	0.236	0.237	0.105	0.099	0.102	0.053	0.049	0.050
λ_{11}	14.597	25.557	0.155	0.065	0.064	0.063	0.029	0.029	0.029
λ_{21}	0.131	0.150	0.111	0.067	0.067	0.066	0.033	0.033	0.033
λ_{31}	0.152	0.167	0.130	0.061	0.061	0.060	0.030	0.030	0.030
λ_{41}	0.142	0.159	0.123	0.075	0.074	0.073	0.029	0.029	0.029
λ_{51}	0.140	0.157	0.120	0.058	0.057	0.057	0.030	0.030	0.030
τ_{12}	89.064	359.959	0.184	0.070	0.069	0.070	0.038	0.036	0.037
τ_{13}	97.408	377.865	0.219	0.095	0.096	0.094	0.051	0.051	0.050
τ_{14}	108.958	397.222	0.319	0.133	0.133	0.132	0.059	0.059	0.059
τ_{22}	0.133	0.138	0.132	0.070	0.069	0.070	0.033	0.033	0.033
τ_{23}	0.184	0.187	0.180	0.088	0.088	0.089	0.043	0.044	0.044
τ_{24}	0.261	0.264	0.251	0.128	0.127	0.127	0.062	0.061	0.060
τ_{32}	0.137	0.140	0.134	0.068	0.067	0.067	0.034	0.033	0.034
τ_{33}	0.183	0.188	0.178	0.102	0.102	0.102	0.047	0.047	0.047
τ_{34}	0.249	0.256	0.238	0.135	0.134	0.134	0.062	0.063	0.062
τ_{42}	0.147	0.148	0.152	0.096	0.097	0.097	0.044	0.046	0.046
τ_{43}	0.176	0.177	0.181	0.108	0.104	0.107	0.051	0.053	0.053
τ_{44}	0.249	0.254	0.249	0.127	0.127	0.126	0.058	0.060	0.060
τ_{52}	0.195	0.192	0.193	0.082	0.075	0.081	0.047	0.044	0.047
τ_{53}	0.228	0.225	0.225	0.098	0.090	0.096	0.056	0.051	0.052
τ_{54}	0.292	0.293	0.287	0.123	0.118	0.120	0.058	0.055	0.057

Table 6.9: *Estimates of parameter set D simulated by the GGS – MEAN and STD of estimated parameters obtained by simulations of 100 different data sets.*

Simulations of parameter set D – GGS									
N	**$N_1 = 300$**			**$N_2 = 1000$**			**$N_3 = 5000$**		
Priori	**Weak**	**Std.**	**Strong**	**Weak**	**Std.**	**Strong**	**Weak**	**Std.**	**Strong**
	BIAS								
λ_{10}	8.291	34.836	0.041	0.003	0.002	0.003	0.004	0.004	0.004
λ_{20}	0.029	0.031	0.034	0.016	0.015	0.016	0.006	0.006	0.006
λ_{30}	0.021	0.022	0.025	0.009	0.008	0.010	0.002	0.003	0.002
λ_{40}	0.083	0.085	0.097	0.034	0.031	0.035	0.011	0.008	0.011
λ_{50}	0.094	0.094	0.103	0.023	0.019	0.021	0.026	0.023	0.027
λ_{11}	1.516	2.604	0.016	0.007	0.004	-0.008	0.004	0.004	0.001
λ_{21}	0.055	0.043	0.009	0.021	0.018	0.006	0.005	0.004	0.002
λ_{31}	0.029	0.018	-0.015	0.021	0.018	0.005	0.000	-0.000	-0.003
λ_{41}	0.026	0.015	-0.017	0.025	0.022	0.009	-0.000	-0.001	-0.003
λ_{51}	0.031	0.020	-0.014	0.006	0.003	-0.009	0.004	0.003	0.001
τ_{12}	8.965	36.057	0.075	0.008	0.006	0.008	0.002	0.002	0.001
τ_{13}	9.822	37.868	0.095	0.010	0.007	0.008	0.007	0.007	0.006
τ_{14}	11.026	39.851	0.141	0.022	0.019	0.017	0.009	0.009	0.009
τ_{22}	0.049	0.050	0.057	0.032	0.030	0.032	0.005	0.005	0.005
τ_{23}	0.073	0.071	0.076	0.036	0.033	0.035	0.008	0.008	0.007
τ_{24}	0.131	0.128	0.126	0.062	0.057	0.057	0.014	0.013	0.013
τ_{32}	0.056	0.057	0.063	0.013	0.012	0.014	-0.002	-0.001	-0.001
τ_{33}	0.063	0.063	0.069	0.037	0.036	0.036	-0.003	-0.005	-0.005
τ_{34}	0.104	0.102	0.102	0.058	0.053	0.051	-0.000	-0.001	-0.000
τ_{42}	0.060	0.063	0.081	0.018	0.017	0.022	0.013	0.010	0.014
τ_{43}	0.076	0.079	0.099	0.037	0.036	0.041	0.013	0.008	0.012
τ_{44}	0.103	0.104	0.119	0.054	0.051	0.054	0.005	0.002	0.005
τ_{52}	0.071	0.072	0.088	0.024	0.022	0.026	0.020	0.016	0.021
τ_{53}	0.093	0.093	0.110	0.028	0.025	0.027	0.026	0.021	0.028
τ_{54}	0.109	0.107	0.119	0.024	0.020	0.022	0.023	0.020	0.023
	ln(MSE)								
λ_{10}	8.829	11.705	-3.695	-5.446	-5.456	-5.451	-7.150	-7.171	-7.163
λ_{20}	-4.294	-4.273	-4.274	-5.785	-5.818	-5.781	-7.346	-7.336	-7.306
λ_{30}	-4.496	-4.474	-4.467	-5.427	-5.428	-5.422	-7.048	-7.056	-7.048
λ_{40}	-3.019	-2.991	-2.936	-4.137	-4.136	-4.137	-5.809	-5.797	-5.818
λ_{50}	-2.718	-2.746	-2.712	-4.464	-4.591	-4.530	-5.677	-5.844	-5.748
λ_{11}	5.362	6.482	-3.734	-5.475	-5.497	-5.528	-7.051	-7.051	-7.077
λ_{21}	-3.915	-3.728	-4.399	-5.310	-5.344	-5.442	-6.839	-6.846	-6.860
λ_{31}	-3.749	-3.576	-4.076	-5.494	-5.534	-5.646	-7.013	-7.012	-7.007
λ_{41}	-3.880	-3.678	-4.176	-5.087	-5.120	-5.223	-7.095	-7.107	-7.099
λ_{51}	-3.887	-3.701	-4.236	-5.697	-5.722	-5.731	-7.006	-7.019	-7.037
τ_{12}	8.979	11.772	-3.243	-5.304	-5.359	-5.322	-6.572	-6.637	-6.610
τ_{13}	9.158	11.869	-2.878	-4.703	-4.702	-4.728	-5.951	-5.961	-5.991
τ_{14}	9.382	11.969	-2.117	-4.011	-4.031	-4.047	-5.655	-5.651	-5.655
τ_{22}	-3.914	-3.851	-3.889	-5.134	-5.194	-5.129	-6.830	-6.792	-6.782
τ_{23}	-3.246	-3.225	-3.271	-4.705	-4.727	-4.713	-6.269	-6.233	-6.243
τ_{24}	-2.470	-2.458	-2.543	-3.905	-3.944	-3.956	-5.521	-5.570	-5.582
τ_{32}	-3.825	-3.785	-3.827	-5.340	-5.377	-5.372	-6.774	-6.855	-6.762
τ_{33}	-3.292	-3.245	-3.325	-4.451	-4.463	-4.454	-6.120	-6.127	-6.120
τ_{34}	-2.626	-2.588	-2.711	-3.848	-3.882	-3.897	-5.572	-5.553	-5.558
τ_{42}	-3.686	-3.668	-3.521	-4.667	-4.654	-4.634	-6.148	-6.116	-6.082
τ_{43}	-3.317	-3.288	-3.165	-4.353	-4.422	-4.341	-5.889	-5.867	-5.852
τ_{44}	-2.631	-2.595	-2.584	-3.969	-3.988	-3.978	-5.680	-5.630	-5.635
τ_{52}	-3.150	-3.180	-3.111	-4.918	-5.096	-4.934	-5.980	-6.135	-5.947
τ_{53}	-2.814	-2.834	-2.777	-4.580	-4.744	-4.623	-5.572	-5.803	-5.672
τ_{54}	-2.342	-2.339	-2.348	-4.157	-4.262	-4.218	-5.544	-5.673	-5.596

Table 6.10: *Estimates of parameter set D simulated by the GGS – BIAS and ln(MSE) of estimated parameters obtained by simulations of 100 different data sets.*

ability of Heywood cases is much higher for the GGS than for the MHS for a low number of observations. For a high number of observations, the Heywood case will not occur anymore for all three prior settings, although the use of a standard prior is still recommended, especially if very high factor loadings ($\lambda_j \approx 2.0$) are involved in the analysis. Additionally the probability of Heywood cases can be lowered by the use of strong priors which often improves parameter estimates for low and medium number of observations. Hence for a medium number of observations and strong priors, and for a high number of observations and standard or strong priors, the results of the GGS are comparable to the MHS. The high sensitivity to Heywood cases of the GGS sampler can be explained by the execution of the Generalized Gibbs move in each iteration. The transformation member randomly drawn in the Generalized Gibbs move does not depend on the priori information, but the priori information is responsible for the prevention of Heywood cases. For that reason, the Generalized Gibbs move causes factor loadings to reach high values promoting the factor loadings to increase further, and so the probability to obtain Heywood cases rises. Naturally, the problem of Heywood cases disappears for a high number of observations where they naturally do not occur and prior information does not exert a big influence on the estimation process. Furthermore, the occurrence of Heywood cases is much rarer if none of the factor loadings appears to be of a very high value.

To sum up, the use of the standard sampler is recommended for pure binary data. In this case, no cutpoints have to be estimated whose convergence should be improved. For a high number of observations however, the GGS obtains the same results as the standard sampler without the occurrence of Heywood cases. For ordinal response, the sampler choice critically depends on the number of observations. For a low number of observations, the employment of the MHS is recommended because the probability of a Heywood case is too high for the GGS in this setting. For a medium number of observations, the MHS can also be the natural choice, although the GGS sampler obtains good results most of the times for strong prior settings. Still, sampling paths of the factor loadings should be analyzed to check for the appearance of a Heywood case. It must be said that Heywood cases appear very likely if one or more factor loadings are very high ($\lambda_j \sim 2.0$), and in such cases the investigation of sampling paths should be always part of the analysis. Naturally, such high factor loadings rarely occur and in this sense, the parameter sets B and D represent extreme parameterizations. For a high number of observations, the GGS is the preferred choice, because the GGS algorithm is computationally much more efficient (about twice as fast) than the MHS, and does not entail the tweaking of the MH tuning parameter. The real data set "PD1" will be analyzed with the GGS in the next chapter since a high number of responses (larger than 5,000) will be included in the estimation process.

After having analyzed the properties of parameter estimates of the measurement model, the next section explores the more intriguing question whether a semiparametric predictor based on indirect covariates can be estimated in a LVM in analogy to standard semiparametric regression.

Section	Model	Indirect effects	Covariates
6.3.1	S1	Parametric	1 metric plus 1 categorical
6.3.2	S2a	Nonparametric	1 metric (function 1)
	S2b	Nonparametric	1 metric (function 2)
6.3.3	S3	Nonparametric	1 spatial
6.3.4	S4a	Nonparametric	• 1 metric effect modifier • 1 categorical interacting variable
	S4b	Nonparametric	• 1 metric effect modifier • 1 categorical interacting variable

Table 6.11: *Overview of the simulated LVM including indirect effects analyzed in Section 6.3.*

6.3 LVM including indirect effects

In this section several simulations are run to check the estimation quality of indirect effects. In each subsection, the following types of indirect effects are analyzed sequentially: parametric effects, metric covariates, spatial covariates, and interactions (VCM). All simulations carried out in this section – which are outlined in Table 6.11 – are based on parameter set D with the factor loadings of parameter set C (see Table 6.2) in the measurement model. We have also conducted simulations with the other parameter sets applied to the measurement model; since the results for other parameter sets are virtually identical, we restrict our discussion and analysis to the parameter set defined here.

For each simulation run, 100 different data sets are produced again for each of the different numbers of observations ($N_1 = 300$, $N_2 = 1000$, $N_3 = 5000$). Throughout all simulations of this section, the MHS is employed due to its more reliable estimation properties for a low number of observations so that we can focus our attention on the estimation quality of indirect effects. Prior strengths for factor loadings are always standard. After each simulation run including 100 data sets, characteristic numbers such as MEAN, STD, BIAS, and MSE are calculated, and nonparametric function estimates are compared with the true underlying functions. Sampling paths and convergence are excellent for all parameters belonging to the structural equation, therefore we refrain from a detailed illustration.

6.3.1 Parametric effects of metric and categorical covariates

We start with the most basic of indirect effects, that is parametric effects. We introduce model S1 with two categorical covariates; one metric variable u_1 which is standard normally distributed, and one ordinal variable u_2 with four categories, all of which have an occupation rate of 25%. For the covariate u_2 standard dummy coding is used, and category 1 represents the reference category. An intercept is not estimated in the structural equation due to identification restrictions. The true values of all regression parameters plus the results of the simulation runs are outlined in Table 6.12.

	Cat.	True values	$N_1 = 300$			$N_2 = 1000$			$N_3 = 5000$		
			MEAN	ln(MSE)	Cov.	MEAN	ln(MSE)	Cov.	MEAN	ln(MSE)	Cov.
$\mathbf{u_1}$	–	-0.5	-0.511	-5.370	97	-0.500	-6.599	96	-0.501	-8.027	93
$\mathbf{u_2}$	1	0.0	–	–	–	–	–	–	–	–	–
	2	0.2	0.215	-3.310	96	0.201	-4.833	98	0.202	-6.053	92
	3	0.4	0.415	-3.654	99	0.397	-4.561	93	0.405	-5.862	91
	4	0.8	0.824	-3.259	92	0.792	-4.407	92	0.807	-5.843	87

	Cat.	True values	$N_1 = 300$		$N_2 = 1000$		$N_3 = 5000$	
			BIAS	STD	BIAS	STD	BIAS	STD
$\mathbf{u_1}$	–	-0.5	-0.011	0.068	0.000	0.037	-0.001	0.018
$\mathbf{u_2}$	1	0.0	–	–	–	–	–	–
	2	0.2	0.015	0.191	0.001	0.090	0.002	0.049
	3	0.4	0.015	0.161	-0.003	0.103	0.005	0.053
	4	0.8	0.024	0.196	-0.008	0.111	0.007	0.054

Table 6.12: *Results of model S1. The true values, MEAN, STD, BIAS, ln(MSE), and coverage are given for the four estimated parameters, obtained by simulating 100 different data sets using the MHS.*

We see that all parameters are estimated correctly. As expected the logarithm of the MSE decreases for a higher number of observations, and it is lower for the metric covariate 1 than for the ordinal covariate 2. Unfortunately the coverage for $N_3 = 5000$ is not as good as for the lower number of observations. One reason for this could be that the MHS does not exhibit good convergence for that high number of observations as the GGS. To compare the result with the GGS, the simulation was also carried out with $N_3 = 5000$ for the GGS, and results show similar MEAN, ln(MSE), and BIAS as for the MHS. However, standard deviation and coverage is slightly better for the GGS; for example the coverage of the regression parameters are 94/93/92/92 for the GGS compared to 93/92/91/87 for the MHS.

We have also conducted simulations for other parameter settings in the measurement model. In general, parameter estimates of the parametric indirect effects are fine; but if several unfavourable conditions hold, correct parameter estimation can fail. Those conditions are met for ordinal indicators with very unevenly spread occupation rates, a low number of observations, and many categorical indirect covariates. In those rare cases, regression coefficients of indirect covariates can be tremendously high, while factor loadings approach zero. Obviously the sampling algorithm is not able to differentiate between the intercept in the measurement model and the categorical regression parameters in the structural part of the model. Of course, such results are highly implausible and can be sometimes corrected by the use of an informative priori on the regression coefficients in the structural part of the model. It is apparent that the complex LVM approaches its limitations, and cannot always be estimated correctly for a very low number of observations and unfavourable general conditions.

6.3.2 Nonparametric effects of metric covariates

We proceed by estimating smooth functions of metric covariates. The functions can be modeled by a first-order random walk, a second-order random walk, or a P-splines prior with a predefined number of knots and degree. Two different models S2a and S2b with two different simulated functions are employed to check the quality of parameter estimates for different functional forms. The first function $f_1(x_1)$ in model S2a increases slowly and smoothly over the range of covariate values, and the second function $f_2(x_2)$ in model S2b rises and drops with a high curvature. The functions and value ranges are

$$
\begin{aligned}
f_1(x_1) &= \frac{x_1}{1500}\sqrt{x_1^2} \qquad &&, \text{ for } x_1 = -30, -29, \ldots, 30\,, \\
f_2(x_2) &= \sin\left(\frac{2\pi x_2}{20}\right) \qquad &&, \text{ for } x_2 = 0, 1, \ldots, 20\,.
\end{aligned}
$$

Simulations S2a and S2b are run for three different numbers of observations, and different priors for the nonparametric effects; the used priors include a first-order random walk (RW1), a second-order random walk (RW2), and a P-splines prior with ten intervals and degrees of two (P2) and three (P3). We have also run simulations with P-splines priors with a higher and lower number of knots and with degree 1, and parameter estimation works equally well for those cases, so their illustration was dropped. Hyperpriors of the smoothing parameter κ are chosen to be $a = 0.001$ and $b = 0.001$ which leads to a highly diffuse but proper prior distribution for κ.

For each of the simulation runs including 100 data sets, the MEAN, the average of the 10% and 90% quantiles, the MSE and the STD of all estimated function parameters are calculated. The true underlying function, the resulting MEAN, and 10% and 90% quantiles for functions f_1 and f_2 are plotted in Figures 6.8 and 6.9, respectively. MSE and STD for all parameters of both functions are drawn as boxplots in Figure 6.10.

In general it is observed that the nonparametric estimates of functions f_1 and f_2 fit the true underlying functions very well (see Figures 6.8 and 6.9). As expected the quality of the estimated functions improves significantly for an increasing number of observations: the bias decreases, the 80% credible interval narrows down, and the MSE and STD decrease significantly. For example for $N_1 = 300$, function f_1 shows a clearly visible bias for the first-order random walk prior and both P-splines priors, especially at the borders of the observed values. The MSE and STD plots in Figure 6.10 indicate that the P-splines deliver a slightly better fitting function estimate than the random walk priors. This can be explained by the fact that the random walks estimate one parameter for each observations, and therefore tend to overfit the data which leads to slightly higher MSE and STD values. The simulations clearly demonstrate that the estimation of a nonlinear function of a metric covariate influencing the latent variable is possible. This is a remarkable result since the covariate influences the latent scores which are not observed directly but have to be estimated.

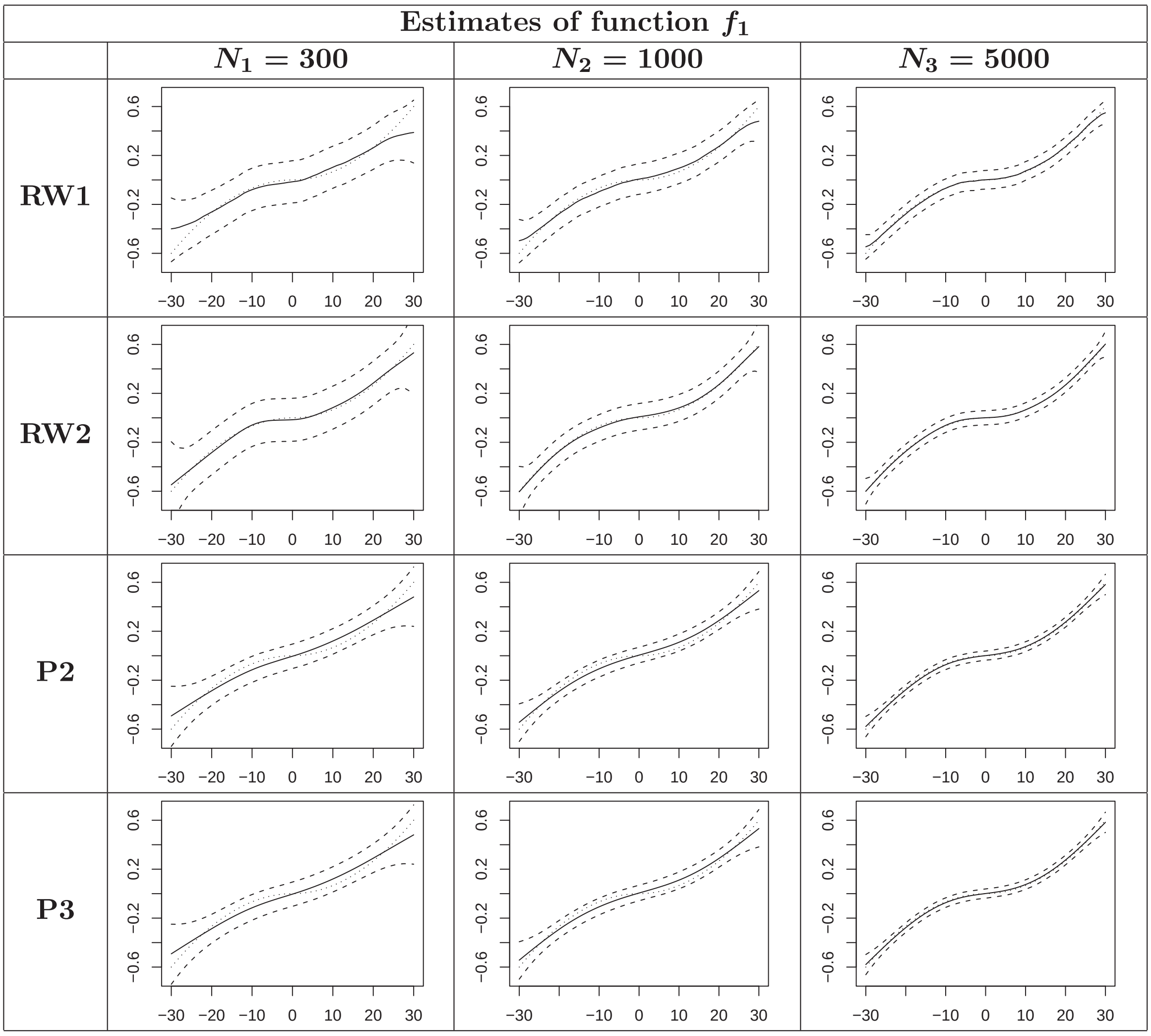


Figure 6.8: *Results of model S2a. The function estimates of f_1 show the MEAN (solid line), 10%- and 90%-quantiles (dashed lines) for different prior settings, and different numbers of observations. The true function f_1 is drawn by a dotted line.*

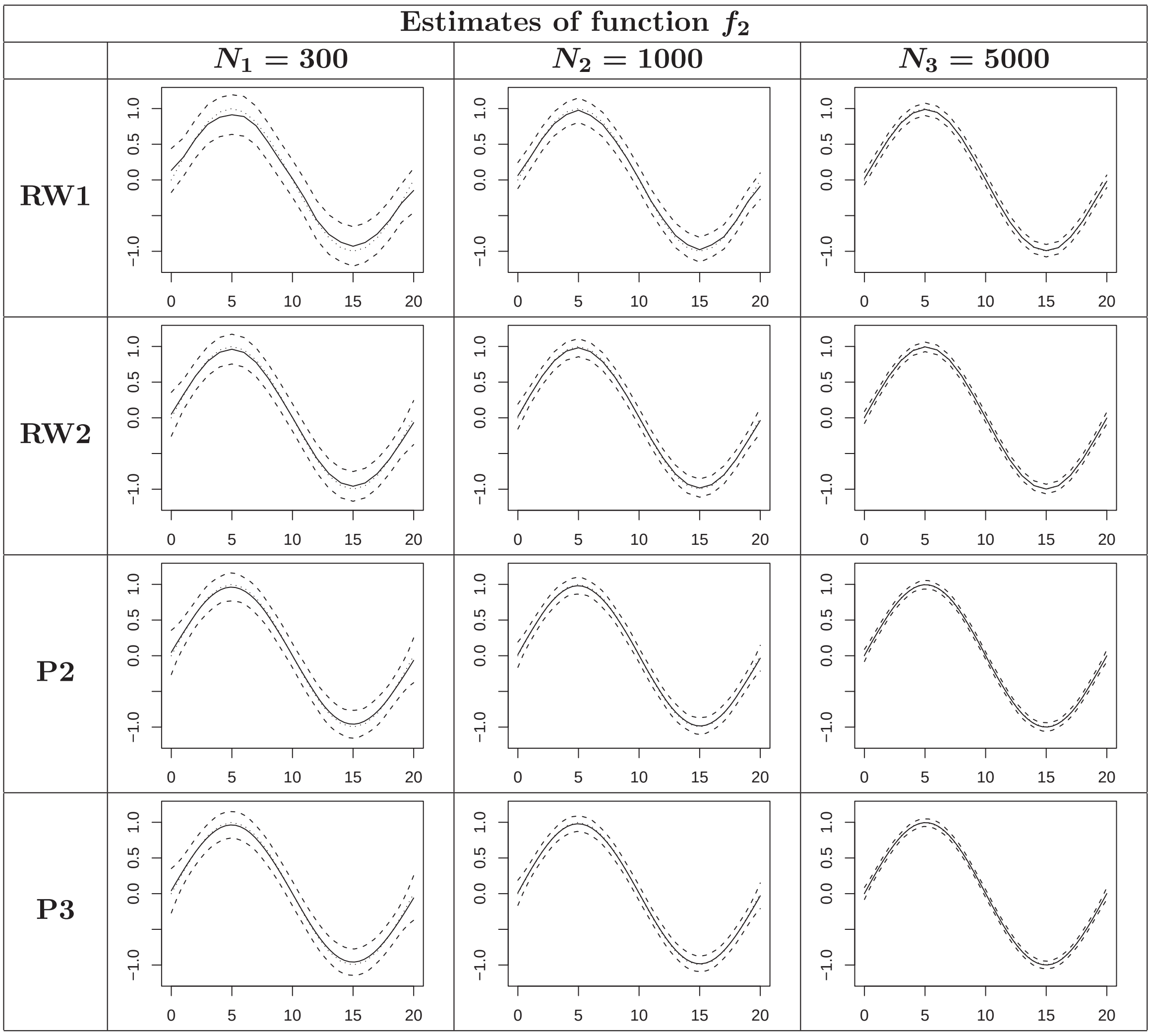


Figure 6.9: *Results of model S2b. The function estimates of f_2 show the MEAN (solid line), 10%- and 90%-quantiles (dashed lines) for different prior settings, and different numbers of observations. The true function f_2 is drawn by a dotted line.*

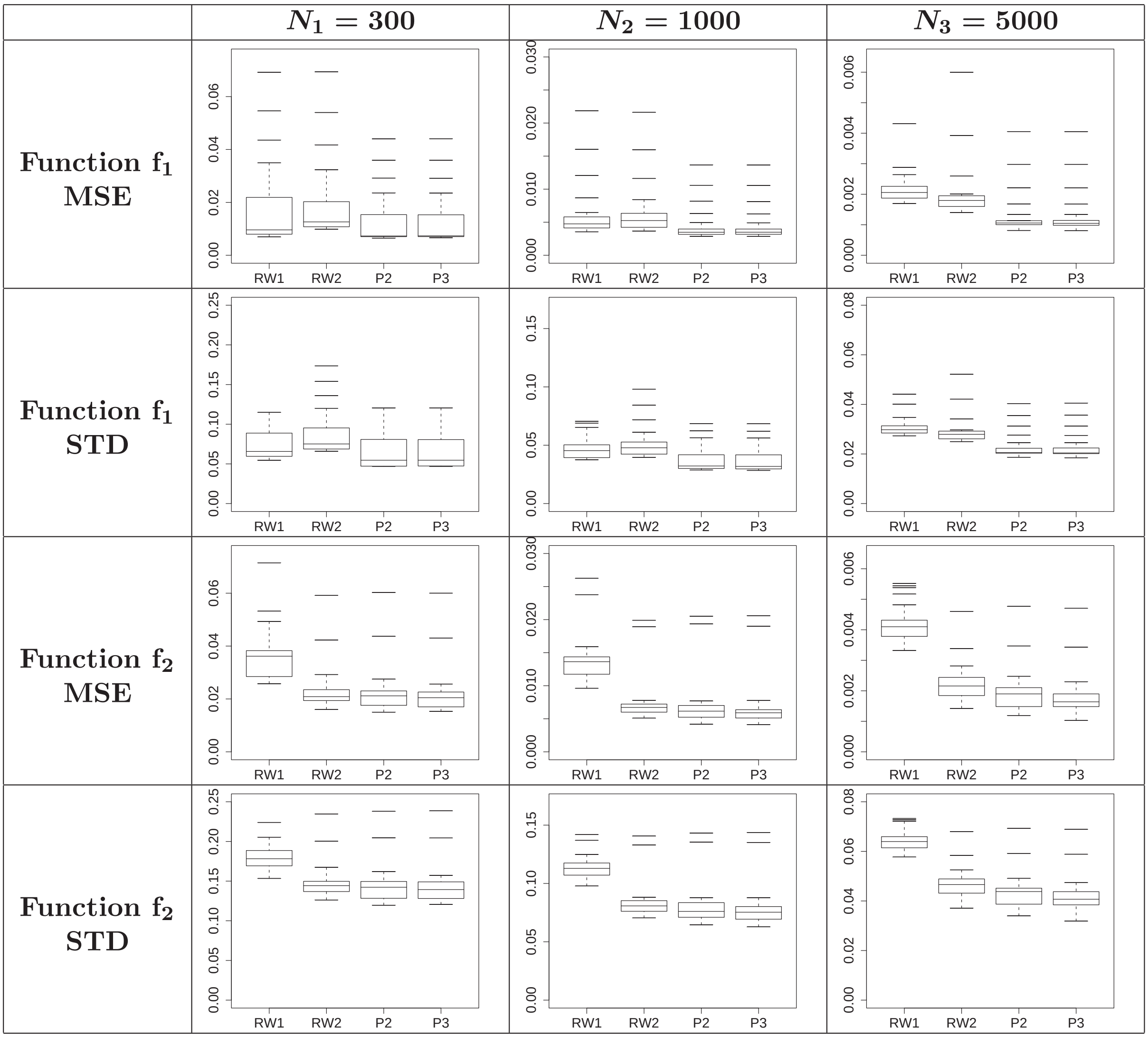


Figure 6.10: *Results of models S2a and S2b. MSE and STD plots for the estimated parameters of functions f_1 (top two rows) and f_2 (bottom two rows), and for different numbers of observations.*

Estimates of smoothing parameters κ for f_1							
	Prior of κ	$N_1 = 300$		$N_2 = 1000$		$N_3 = 5000$	
		MEAN	STD	MEAN	STD	MEAN	STD
RW1	$a = 0.001$, $b = 0.001$	$1.07 \cdot 10^{-2}$	$3.74 \cdot 10^{-3}$	$7.14 \cdot 10^{-3}$	$1.43 \cdot 10^{-3}$	$4.34 \cdot 10^{-3}$	$5.71 \cdot 10^{-4}$
	$a = -0.5$, $b = 0.0$	$1.63 \cdot 10^{-2}$	$4.71 \cdot 10^{-3}$	$8.23 \cdot 10^{-3}$	$1.74 \cdot 10^{-3}$	$4.60 \cdot 10^{-3}$	$6.40 \cdot 10^{-4}$
	$a = -1.0$, $b = 0.0$	$2.29 \cdot 10^{-2}$	$6.57 \cdot 10^{-3}$	$1.01 \cdot 10^{-2}$	$2.09 \cdot 10^{-3}$	$5.08 \cdot 10^{-3}$	$6.88 \cdot 10^{-4}$
RW2	$a = 0.001$, $b = 0.001$	$1.25 \cdot 10^{-3}$	$2.34 \cdot 10^{-4}$	$8.21 \cdot 10^{-4}$	$1.19 \cdot 10^{-4}$	$6.08 \cdot 10^{-4}$	$5.63 \cdot 10^{-5}$
	$a = -0.5$, $b = 0.0$	$3.40 \cdot 10^{-4}$	$2.12 \cdot 10^{-4}$	$1.21 \cdot 10^{-4}$	$6.16 \cdot 10^{-5}$	$7.82 \cdot 10^{-5}$	$2.73 \cdot 10^{-5}$
	$a = -1.0$, $b = 0.0$	$9.84 \cdot 10^{-4}$	$3.92 \cdot 10^{-4}$	$3.13 \cdot 10^{-4}$	$1.43 \cdot 10^{-4}$	$1.32 \cdot 10^{-4}$	$4.87 \cdot 10^{-5}$
P2	$a = 0.001$, $b = 0.001$	$2.71 \cdot 10^{-2}$	$1.65 \cdot 10^{-2}$	$1.61 \cdot 10^{-2}$	$6.64 \cdot 10^{-3}$	$1.32 \cdot 10^{-2}$	$3.80 \cdot 10^{-3}$
	$a = -0.5$, $b = 0.0$	$8.63 \cdot 10^{-2}$	$5.34 \cdot 10^{-2}$	$2.92 \cdot 10^{-2}$	$1.42 \cdot 10^{-2}$	$2.00 \cdot 10^{-2}$	$6.21 \cdot 10^{-3}$
	$a = -1.0$, $b = 0.0$	$2.65 \cdot 10^{-1}$	$1.00 \cdot 10^{-1}$	$8.49 \cdot 10^{-2}$	$3.28 \cdot 10^{-2}$	$3.48 \cdot 10^{-2}$	$1.08 \cdot 10^{-2}$
P3	$a = 0.001$, $b = 0.001$	$3.01 \cdot 10^{-2}$	$2.30 \cdot 10^{-2}$	$1.67 \cdot 10^{-2}$	$6.97 \cdot 10^{-3}$	$1.41 \cdot 10^{-2}$	$3.96 \cdot 10^{-3}$
	$a = -0.5$, $b = 0.0$	$9.56 \cdot 10^{-2}$	$6.06 \cdot 10^{-2}$	$3.26 \cdot 10^{-2}$	$1.66 \cdot 10^{-2}$	$2.18 \cdot 10^{-2}$	$6.71 \cdot 10^{-3}$
	$a = -1.0$, $b = 0.0$	$3.13 \cdot 10^{-1}$	$1.27 \cdot 10^{-1}$	$9.85 \cdot 10^{-2}$	$4.33 \cdot 10^{-2}$	$3.99 \cdot 10^{-2}$	$1.19 \cdot 10^{-2}$

Estimates of smoothing parameters κ for f_2							
	Prior of κ	$N_1 = 300$		$N_2 = 1000$		$N_3 = 5000$	
		MEAN	STD	MEAN	STD	MEAN	STD
RW1	$a = 0.001$, $b = 0.001$	0.146	0.045	0.099	0.016	0.070	$5.44 \cdot 10^{-3}$
	$a = -0.5$, $b = 0.0$	0.168	0.051	0.108	0.017	0.075	$5.85 \cdot 10^{-3}$
	$a = -1.0$, $b = 0.0$	0.194	0.058	0.117	0.019	0.080	$6.20 \cdot 10^{-3}$
RW2	$a = 0.001$, $b = 0.001$	0.044	0.017	0.028	$5.50 \cdot 10^{-3}$	0.018	$2.85 \cdot 10^{-3}$
	$a = -0.5$, $b = 0.0$	0.060	0.027	0.034	$7.33 \cdot 10^{-3}$	0.020	$3.48 \cdot 10^{-3}$
	$a = -1.0$, $b = 0.0$	0.085	0.041	0.043	$9.79 \cdot 10^{-3}$	0.023	$4.28 \cdot 10^{-3}$
P2	$a = 0.001$, $b = 0.001$	0.395	0.140	0.269	0.048	0.189	0.032
	$a = -0.5$, $b = 0.0$	0.580	0.223	0.347	0.068	0.226	0.041
	$a = -1.0$, $b = 0.0$	0.898	0.392	0.472	0.100	0.278	0.055
P3	$a = 0.001$, $b = 0.001$	0.433	0.156	0.297	0.056	0.214	0.034
	$a = -0.5$, $b = 0.0$	0.660	0.288	0.394	0.083	0.262	0.044
	$a = -1.0$, $b = 0.0$	1.083	0.580	0.558	0.144	0.334	0.067

Table 6.13: *Results of models S2a and S2b. MEAN and STD of the smoothing parameters κ are printed for different function types, different prior strengths of the smoothing parameter, and different numbers of observations.*

In order to compare the broadness of the 80% credible interval in our LVM setting with a standard linear model, we performed several regression simulations where the latent variable is used as the response and is influenced by the predictor of indirect effects. Of course, this is only possible because the actual values of the latent variables are known in a simulation study; in a real-life research setting however, the values of the latent variables have to be estimated themselves. The results of these analyses show that the broadness of the estimated functions in the LVM setting is slightly higher than in the case of a direct regression. This is obvious because the latent values have to be estimated, and therefore they fluctuate around the true latent values in each iteration of the MCMC sampler. Nevertheless the difference in broadness is very small which demonstrates that it is absolutely possible and valid to estimate smooth nonparametric functions in a LVM with narrow credible intervals.

So far we have used a highly diffuse but proper prior distribution for the smoothing parameter κ with $a = 0.001$ and $b = 0.001$. In order to check if proper function estimates are also obtained for noninformative prior distributions, all simulations were repeated for the two noninformative parameterizations introduced in Section 5.2.2, i. e. a=-0.5, b=0, and

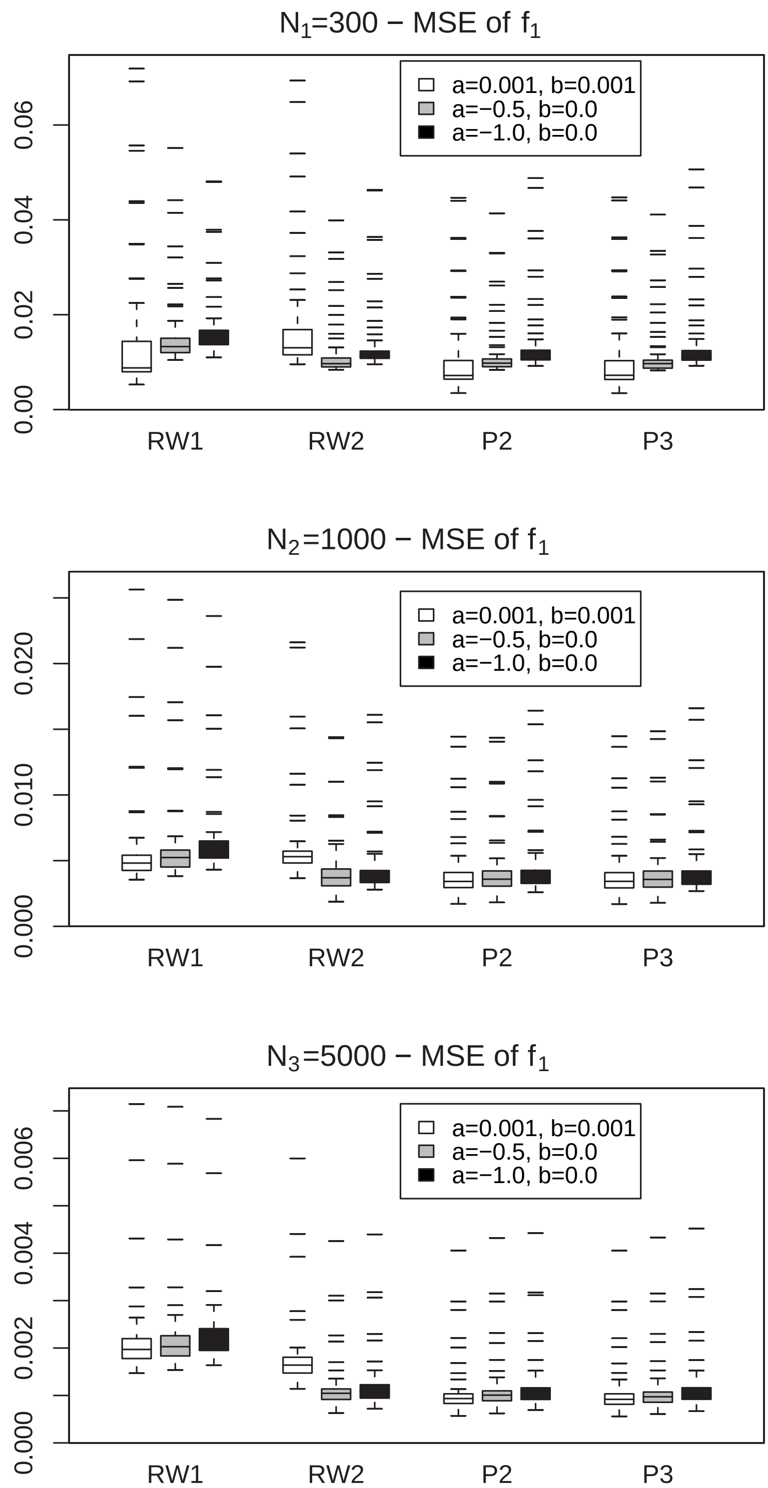


Figure 6.11: *Results of model S2a. MSE of the estimates of function f_1 are plotted for different function types, and different prior settings of the smoothing parameter κ.*

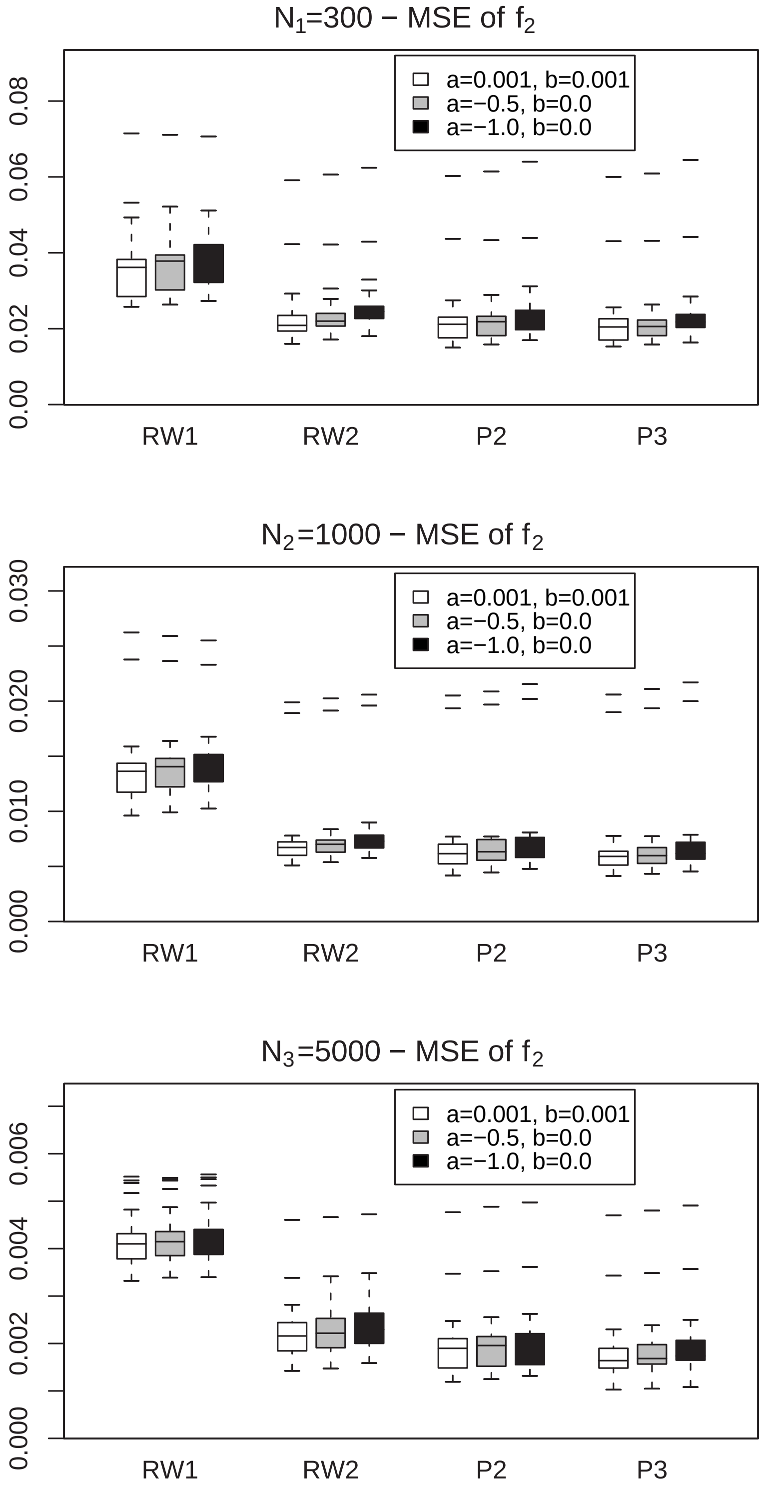


Figure 6.12: *Results of model S2b. MSE of the estimates of function f_2 are plotted for different function types, and different prior settings of the smoothing parameter κ.*

a=-1.0, b=0.

The estimated values for κ for f_1 and f_2 are summarized in Table 6.13. In general the value of the estimated smoothing parameters κ decreases as the number of observations n increases. Furthermore, the MEAN and the STD of κ increase as the prior distribution becomes more diffuse – this is a universal observation except for the second-order random walk of f_1. The increase in percentage is lower for data sets with a higher number of observations; it is also lower for all simulations with the more curvy function f_2 compared to the more flat function f_1.

Even more important than the mere values of the smoothing parameter is the effect of different prior settings of κ on the quality of the estimated function. Therefore the MSE of the estimated function values are plotted in Figures 6.11 and 6.12 for f_1 and f_2, respectively. Apparently the different prior settings of κ have a very small but observable effect on the MSE – the MSE of the function parameters rise slightly for more diffuse prior settings (again except for the second-order random walk of f_1). This clearly demonstrates that the use of improper priors on the smoothing parameter κ still results in proper function estimates of the metric covariate, hence improper priors on κ can be used in our LVM model.

6.3.3 Nonparametric effects of spatial covariates

To simulate a spatial covariate, we introduce a function based on a two-dimensional grid. The spatial form is a simple rectangular structure spreading along the x- and y-axis and consisting of 20×20 regions. The function is defined as

$$f_{spat} = \sin\left(\frac{2\pi x}{20}\right) \cdot \cos\left(\frac{2\pi y}{20}\right) \qquad , \text{for } \{x, y\} = 0, 1, \ldots, 20\,.$$

The true functional values with the spatial structure are drawn in Figure 6.13. Each region in the interior has four direct neighbors, located to the north, east, south, and west of each region; accordingly regions on the border or at the corners have less neighbors. All regions occur with the same probability. Equivalent to the case for metric covariates, the hyperpriors of the smoothing parameter κ are chosen to be highly diffuse but proper with $a = 0.001$ and $b = 0.001$.

Again 100 different data sets are generated for each of the number of observations $N_1 = 300$, $N_2 = 1000$, and $N_3 = 5000$. For each of the number observations, we have plotted the estimated values of one specific data set, the MEAN and the BIAS of all estimated regional function parameters in Figure 6.14. For all parameters, the boxplots of the MSE and the boxplots of the STD of all regions are depicted in Table 6.15.

The two-dimensional plots show that the estimation of a spatial effect is still poor for a low number of observations compared to the number of regions. Although the graph

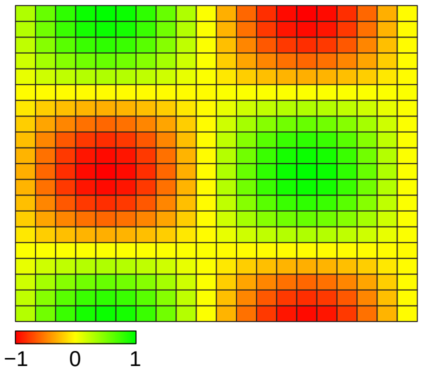


Figure 6.13: *S3: True functional values and structure of the spatial function f_{spat}.*

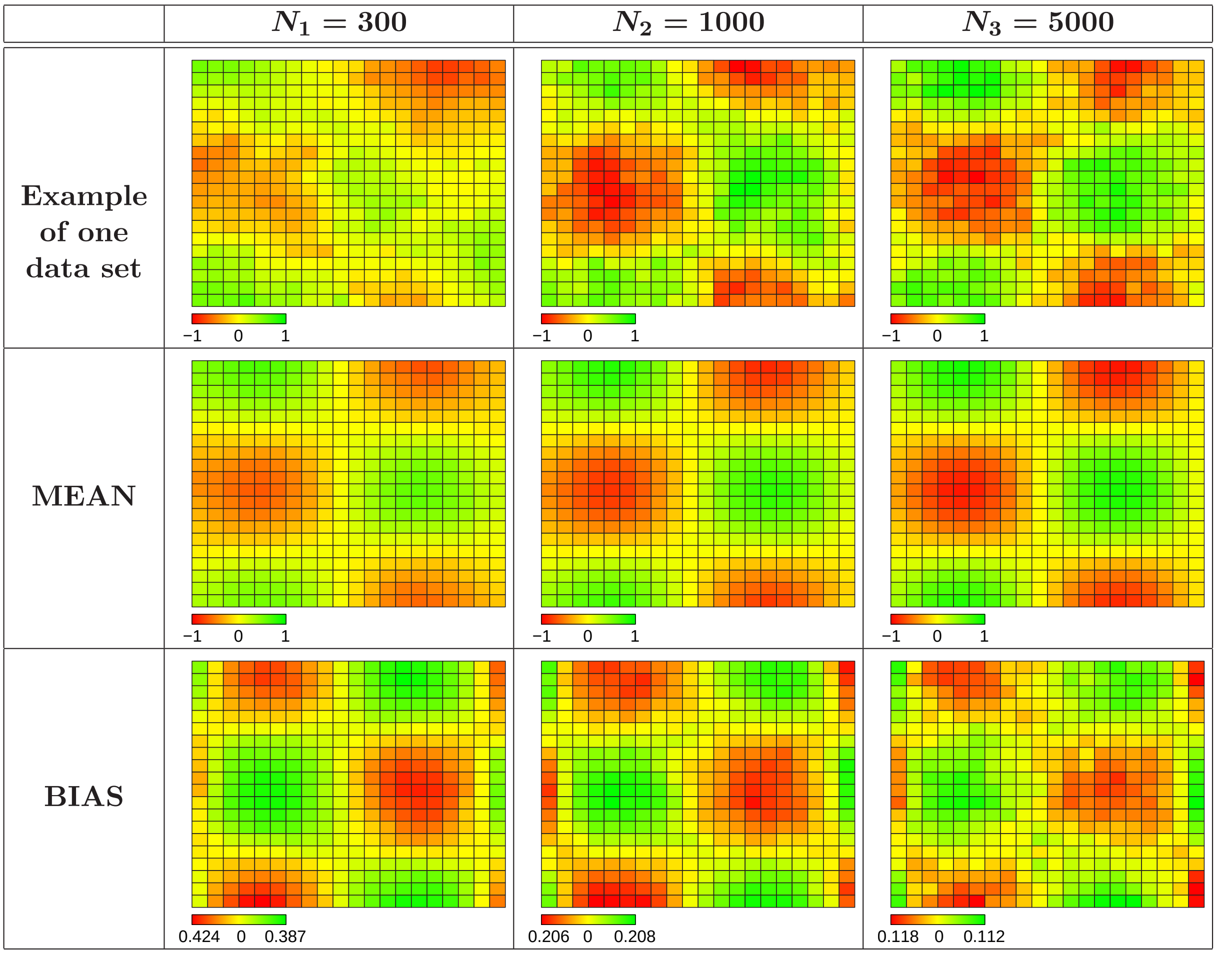


Figure 6.14: *Results of model S3. A simulation of one specific data set (top row), the MEAN (middle row), and the corresponding BIAS (bottom row) for different numbers of observations.*

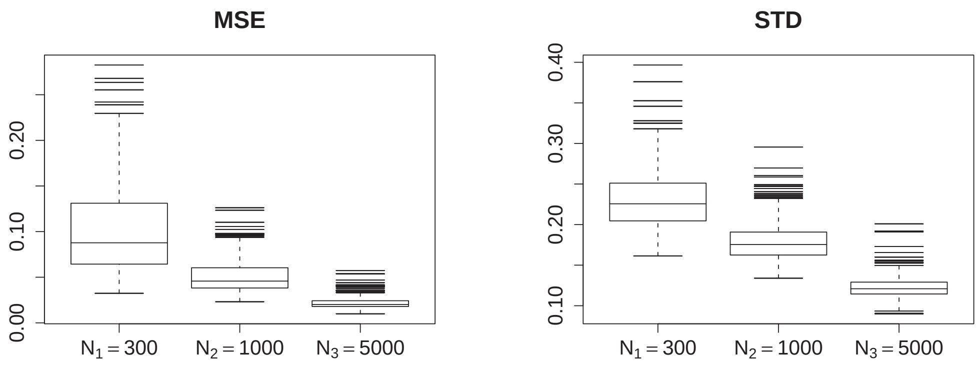


Figure 6.15: *Results of model S3. The MSE (left) and STD (right) of the 400 region parameters for different numbers of observations.*

illustrating the MEAN for $N_1 = 300$ reproduces the general form of the true underlying function, the graph of a specific example of one data set clearly does not show the true function. Other generated data sets reveal the basic structure of the true function in a better way but we wanted to show a typical mediocre example. The reason for that lies in the low number of observations; in this case for example, one region contains only very few 0.75 observations on average.

The BIAS graph shows that high and low function values are estimated too low and high, respectively. This property is also shared by the models with a higher number of observations, and thus is a general observation. The reason for this behaviour is that the estimated value of a single region is based upon the average function values of its neighboring regions due to the use of the Markov random field prior. Therefore, peak values are pulled down whereas very low values are increased by its neighbors. The function example f_{spat} used here contains very succinct peaks and valleys, and therefore might be not necessarily representative for real-life data that would typically produce a smoother and more flat spatial effect. However, the bias is very strong for $N_1 = 300$, and MSE and STD indicate a high fluctuation of the estimated parameters. For $N_2 = 1000$, which corresponds to an average 2.5 observations per region, the true function is already estimated pretty well. The BIAS dropped significantly (notice the different scale in the BIAS graphs), both MSE and STD have decreased considerably. For $N_3 = 5000$ (12.5 observations on average per region), quality of estimation improved even further, as expected. For a higher number of observations, the BIAS has the same functional form as for $N_1 = 300$, but the absolute effect is much smaller. The BIAS at the edges stems from the fact that the low number of neighboring regions which have a higher/lower value pull the function values of the edges up/down, respectively. Hence a spatial effect influencing the latent variables can be indeed estimated, although there should be a sufficient number of observations. For 400 different regions, a spatial effect could be estimated for 1000 observations, and estimates are probably better for real-life data when the spatial effect does not contain such succinct peaks and valleys as the function f_{spat} does in our example.

In order to check the effect of improper priors, all simulations are run again for the improper prior settings of the smoothing parameter κ with $a = -0.5$, $b = 0$, and $a = -1$, $b = 0$. MEAN and STD of the estimated values of κ are depicted in Table 6.14 – the effect of the different prior settings on the estimation quality of the spatial function is indicated by the MSE plotted in Figure 6.16. The effect of different prior settings in the case of a spatial covariate is even weaker than in the case of a metric covariate. MEAN and STD values of κ are very similar for all prior settings particularly for medium and high observations. The effect of the spatial function values on the MSE is almost negligible. We conclude that noninformative prior distributions of κ can also be employed for spatial covariates, and proper posteriors of the function estimates are obtained.

Estimates of smoothing parameters κ of spatial function						
Prior of κ	**$N_1 = 300$**		**$N_2 = 1000$**		**$N_3 = 5000$**	
	Mean	**STD**	**Mean**	**STD**	**Mean**	**STD**
$a = 0.001$, $b = 0.001$	0.624	0.2292	0.407	0.0590	0.203	0.0138
$a = -0.5$, $b = 0.0$	0.688	0.2312	0.418	0.0597	0.205	0.0139
$a = -1.0$, $b = 0.0$	0.746	0.2473	0.430	0.0602	0.207	0.0139

Table 6.14: *Results of model S3. MEAN and STD of the smoothing parameter κ for different prior settings of κ.*

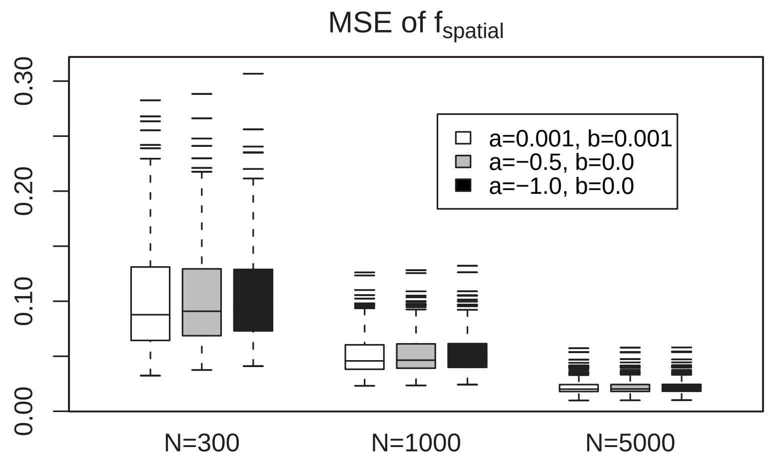


Figure 6.16: *Results of model S3. MSE of the regional estimated values of the spatial function for different prior settings of the smoothing parameter κ.*

6.3.4 Nonparametric effects of interacting covariates

To illustrate a VCM, we introduce an interacting variable v with two categories, so each observation i is assigned to one of the two categories of v. Furthermore, we want to estimate two nonparametric functions of the effect modifier x which is a continuous covariate. Thus we assume that the functional dependancy $f(x)$ for observations assigned to category 1 of v is different than for those observations assigned to category 2. We use the second sinusoidal function defined above in the subsection for metric covariates for the function of the first category of v. Table 6.15 shows the functional values for $f(x)$ of the two different categories of v.

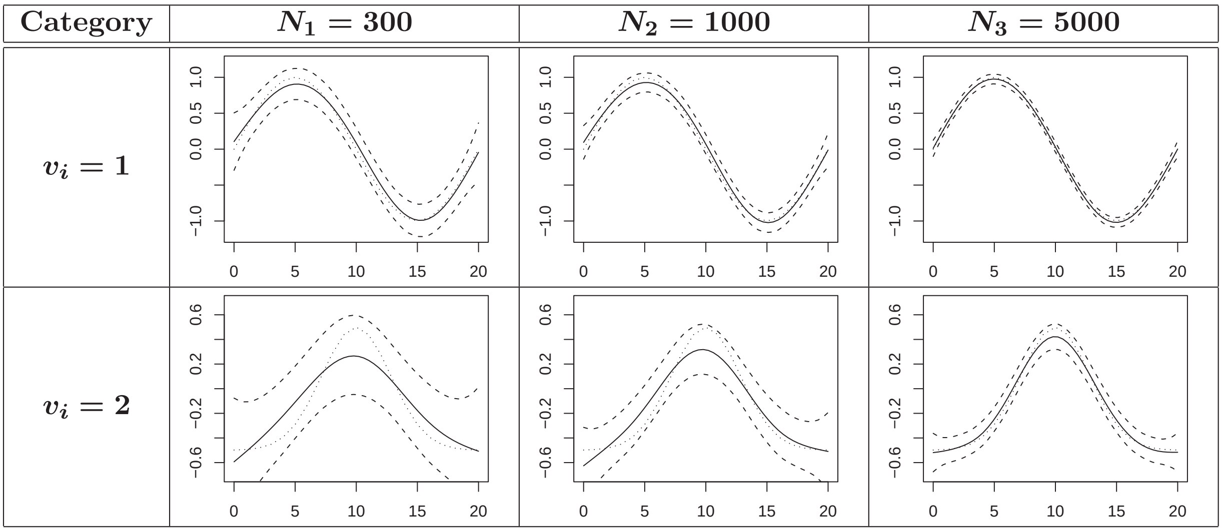
Category	**Functional value**
$v_i = 1$	$f_1(x_i) = \sin(2\pi x_i/20)\,, \quad x_i = 0, 1, \ldots, 20$
$v_i = 2$	$f_2(x_i) = f_1(x_i) + exp(-(x_i - 10)^2/16) - 0.5\,, \quad x_i = 0, 1, \ldots, 20$

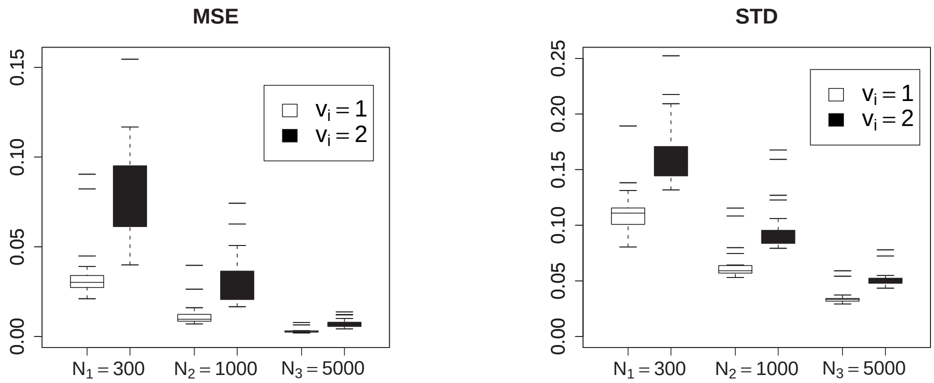
Table 6.15: *S4a: Functional values of $f(x)$ for the two categories of variable v.*

The function belonging to the second category of v results from the addition of the function related to the first category of v plus an additional curve. Hence, we expect our model to estimate both the sinusoidal function and the additional difference function between category $v_i = 1$ and $v_i = 2$. We use second-order random walk priors for modeling the nonparametric functions; first-order random walk and P-splines priors work equally well and are therefore ommitted in this paragraph. All simulations performed in this subsection use highly diffuse but proper priors for the smoothing parameters κ with $a = 0.001$ and $b = 0.001$. Noninformative priors might be used – their effect on the quality of the function estimate and on the estimated values of κ is similar to the results obtained for metric covariates in Section 6.3.2.

Let us have a look at the simulation results. The top of Figure 6.17 shows the estimated functions of categories $v_i = 1$ (reference category) and $v_i = 2$ for $N = \{300, 500, 1000\}$. The function estimates of the reference category $v_i = 1$ are satisfying, and are very similar to the results of Section 6.3.2 for the function estimates of a single metric covariate. The second estimated function for category $v_i = 2$, however, seems to show mediocre properties for a low number of observations. For small N, the estimated function is more flat than the true function; furthermore the 80% credible interval is rather broad. MSE and STD of the estimated parameters of both functions in the middle of Figure 6.17 confirm the difference in quality of estimation for both functions.

We suppose that the function for $v_i = 2$ is smoothed out too much. To investigate this behaviour further, the average estimate of the smoothing parameters including their standard deviation are given at the bottom of Figure 6.17. The smoothing parameters decrease for increasing N which is the normal expected behaviour. However, the smoothing parameters between the functions for $v_i = 1$ and $v_i = 2$ differ highly, especially for a low number of observations. The smoothing parameters for the function $v_i = 2$ are always lower than for





Smoothing parameters τ							
Category	**Parameter**	**$N_1 = 300$**		**$N_2 = 1000$**		**$N_3 = 5000$**	
		Mean	**STD**	**Mean**	**STD**	**Mean**	**STD**
$v_i = 1$	τ_{f_1}	0.431	0.112	0.321	0.059	0.214	0.018
$v_i = 2$	τ_{f_2}	0.196	0.145	0.157	0.069	0.148	0.039

Figure 6.17: *Results of model S4a.* ***Top****: The average mean (solid line), 10%- and 90%-quantiles (dashed lines) of the two estimated functions, and the true functions (dotted line).* ***Middle****: MSE and STD plots for the estimated parameters of both functions $f_1(x)$ and $f_2(x)$.* ***Bottom****: MEAN and STD of the estimated functions' smoothing parameters τ for categories $v_i = \{1, 2\}$.*

Category	Functional value
$\widetilde{v}_i = 1$	$\widetilde{f}_1(x_i) = -f_2(x_i)$, $x_i = 0, 1, \ldots, 20$
$\widetilde{v}_i = 2$	$\widetilde{f}_2(x_i) = f_1(x_i) + f_2(x_i) + 0.162$, $x_i = 0, 1, \ldots, 20$

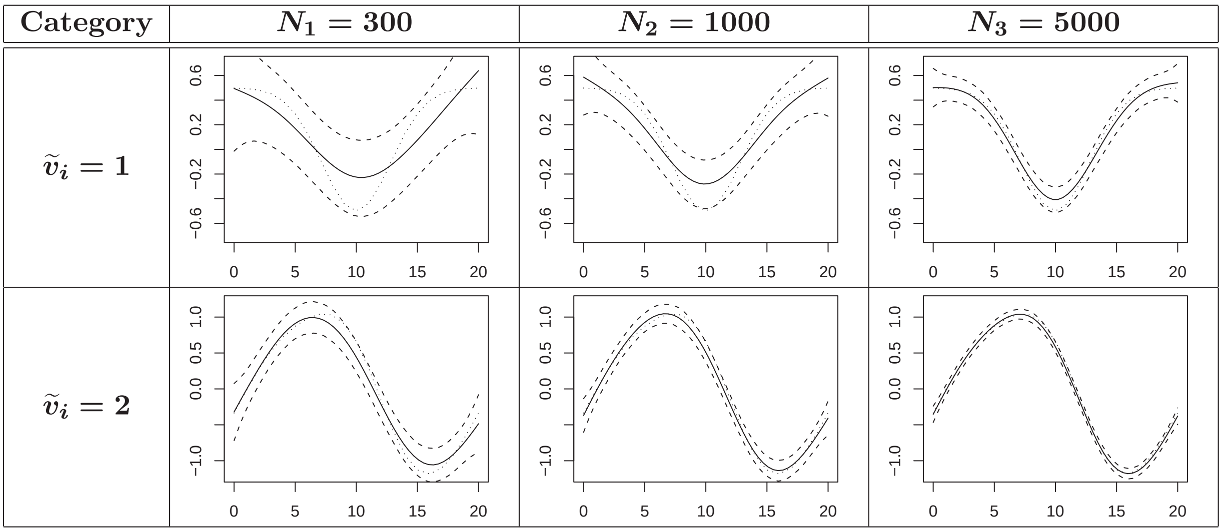
Table 6.16: *S4b: Functional values of $\widetilde{f}(x)$ for the two categories of interacting variable v.*

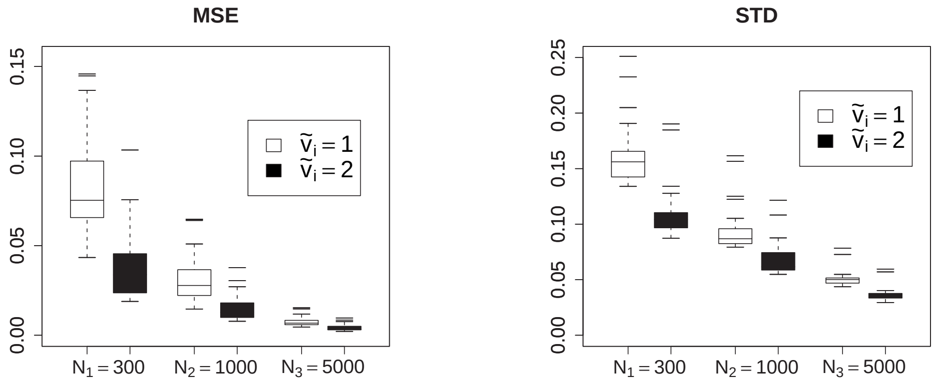
$v_i = 1$. One reason for that might be the different functional forms; nevertheless smoothing parameters of both functions align for a high number of observations.

In order to check whether this behaviour prevails if the reference category is set to category 2, the same simulations have been carried out with switched reference categories (simulation S4b); all functions, categories and results of this second set of simulations are denoted by a tilde. The resulting true functions are defined in Table 6.16, and the estimated functions are plotted at the top of Figure 6.18. Again the function estimates of the non-reference category $\widetilde{v}_i = 1$ are inferior to those of the reference category $\widetilde{v}_i = 2$. The values of MSE and STD (see the middle of Figure 6.18) are very similar to model S4a: MSE and STD of the function for the reference category are consistently lower than for the other category. The same is true for the smoothing parameters which are depicted at the bottom of Figure 6.18.

Consequently, there seems to be a general problem in estimating smooth functions for categories apart from the reference category; or possibly the functional form of our example promotes the poor function estimate. The results indicate that the smoothing parameter for non-reference categories is estimated too low, and therefore the estimated function is too flat due to a reason which requires further investigation. For a high number of observations, the estimation quality significantly improves and function estimates look satisfactory.

In the last paragraphs, we have shown that it is possible to estimate a semiparametric predictor in a LVM including indirect covariates. This allows a very detailed research how categorical, metric, and spatial covariates influence the latent variables. In order to demonstrate the applicability of our model in a real-life research setting, the next chapter deals with the employment of the model using the data set "Perspektive Deutschland 1".





Smoothing parameters $\boldsymbol{\tau}$							
Category	**Parameter**	$\boldsymbol{N_1 = 300}$		$\boldsymbol{N_2 = 1000}$		$\boldsymbol{N_3 = 5000}$	
		Mean	**STD**	**Mean**	**STD**	**Mean**	**STD**
$\widetilde{v}_i = 1$	$\widetilde{\tau}_{\widetilde{f}_1}$	0.181	0.153	0.127	0.058	0.139	0.039
$\widetilde{v}_i = 2$	$\widetilde{\tau}_{\widetilde{f}_2}$	0.466	0.114	0.400	0.080	0.289	0.031

Figure 6.18: *Results of model S4b.* ***Top**: The average mean (solid line), 10%- and 90%-quantiles (dashed lines) of the two estimated functions, and the true functions (dotted line).* ***Middle**: MSE and STD plots for the estimated parameters of both functions $\widetilde{f}_1(x)$ and $\widetilde{f}_2(x)$.* ***Bottom**: MEAN and STD of the estimated functions' smoothing parameters τ for categories $\widetilde{v}_i = \{1, 2\}$.*

Chapter 7

Application to an Internet survey

In this chapter, we apply the LVM model to a real-life research setting. The data set stems from an internet survey called "Perspektive Deutschland 1", abbreviated by PD1. In the first section, we provide some background information on PD1. In the second section, the latent constructs are defined, the manifest response variables are introduced, and the covariates modifying the latent variables are described. To gain a better understanding of the data structure, descriptive statistics are presented for the indicators and the indirect covariates. In the third section, the LVM with one latent variable is estimated for different compositions of the predictor of the structural equation. The fourth section gives one example of an analysis with two latent variables. In the fifth and last section, we discuss issues of interpretation and prediction.

7.1 Description of the data set "PD1"

The internet survey "Perspektive Deutschland 1" was initiated by the companies McKinsey&Company, stern.de, and T-Online. From October to December 2001, approximately 170,000 voluntary participants filled out the social survey on the internet, and thus PD1 was the largest internet survey ever realized in Europe at that time. The high number of participants ensured that opinions from German inhabitants of all ages and classes with different social and economic backgrounds were collected. The general goal of the survey was to receive answers from the population in which areas of life people are willing to bear responsibility, and in which areas they consider the state to fulfill the duty. For example, one typical question concerns the responsibility for higher education: should the state provide free university education, or should students be forced to give a monetary contribution to their own studies ? Another main focus of the study is to measure the happiness of the population with living conditions and infrastructure offerings of the state at their place of living. Results of the study have been presented in the media (e. g. in the Stern, the highest volume weekly magazine in Germany), and were discussed by many politicians to assess the willingness of the population to reform the country, develop the necessary political changes, and thus ensure the country's well-being in the future. Furthermore, local politicians could derive measures to improve local living conditions. Due to the tremendous success of the survey, the project "Perspektive Deutschland" was repeated three times in the three subsequent years with similar questions from different angles. The fourth survey conducted in late 2004 attracted over 500,000 online participants all over Germany, and thus is the largest sociopolitical survey worldwide.

By construction, an internet survey does not deliver a representative sample of the whole population. Two separate effects distort the representativeness: firstly, the probability of possessing an online access, and secondly the respondent's willingness to participate in the survey. In order to make the study representative, an additional offline survey with more than 2,000[1] persons was conducted simultaneously; this offline survey serves to generate weights for each response of the internet survey, so that finally representative statements about the online survey can be made. For our analyses, representativeness does not play a role – therefore we neglect the offline survey and the resulting weights for the internet survey. Our total population sample contains all people who filled out the survey on the internet.

The survey consists of four separate blocks of questions, covering 4 areas: main questions, employment, education, and savings. After having answered the main questions, which have to be filled out by all participants, one of three specialized question batteries is randomly chosen to be answered by the respondent. Finally, participants could voluntarily answer one or two of the remaining blocks. In the end, over 60% of the participants have filled out all questions. Totally, the survey contains about 70 questions with more than 240 different response variables. As common for social surveys, almost all of the variables are of binary or ordinal type. The only true pseudo-metric variable is given by age, and one spatial variable is provided by the license plate of the participants' living area.

7.2 Descriptive statistics of indicators and covariates

In this section, we provide general information about the data sample, descriptive statistics of the ten indicators measuring the two latent constructs, and the indirect covariates used in our analyses. There are no direct effects included in our models, but they could be included without greater effort if needed. We have also made experiments with direct effects which are generally estimated correctly, but in our analyses they do not allow for a useful interpretation or further insight into the data structure.

The total number of participants of the survey is about 170,000. We have generated a subsample which contains approximately 15% of the total data (about 25,000 participants).

[1] For "PD4", about 10,000 people have been interrogated for the offline survey.

Furthermore, those observations which have valid answers for all indicators and covariates are kept, which is about two thirds of the subsample for the LVM with one latent variable (Section 7.3), and about one fifth for the LVM with two latent variables (Section 7.4). After having deleted a small number of participants with an age lower than 20 or higher than 70 due to the low number of observations in those age ranges, we end up with 17,060 and 6,804 observations which are the basis for all future analyses with one and two latent variables, respectively.

The latent construct used in our analyses with one and two latent variables is supposed to reflect the participant's attitude if social coverage should be ensured by the citizens on one's own responsibility, or if the state should take care about the social coverage of its citizens. For example, social coverage includes social security, pensions and health care issues. A person with a high value of the latent construct would rather like citizens to take on more responsibility, to lay aside a certain amount of money in case of unemployment, to save privately for their retirement, and to possess a private health insurance policy according to their needs. Otherwise, a person with a low value of the latent construct would expect the state to take care of those issues. Since this latent construct cannot be measured or obtained directly, five indicators are used which measure the latent construct indirectly. The five indicator variables including the corresponding questions and response categories are summarized in Table 7.1. The response patterns of the five indicators are plotted in Figure 7.1.

The response pattern of the five indicators already reveals basic sentiments of the participants. Although people agree that they will have to show more initiative in saving money for their retirement (indicator 2), most people still expect the state to play a major role in providing provision for old age (indicator 3). The wish of the state being responsible for protecting citizens in difficult life situations (indicator 4) and providing the health care system (indicator 5) is even stronger. Finally, most people would still like a social system where everything is taken care of and they pay the costs with taxes and contributions (indicator 1). Although all five indicators seem to hint into the direction of the latent construct from different angles, we want to check the indicators' interrelationship before using them in the analyses. For that purpose, we calculated an improved version of the Spearman correlation coefficient between all indicator variables in Table 7.3. This rank correlation coefficient accounts for the fact that multiple observations may have the same response pattern (e. g. see Kendall, 1990). The results confirm an interrelationship between all of the five indicators, so we can be confident that all five indicators contribute to the measurement of the latent construct. The only negative correlations occur for indicator 2, therefore we expect the factor loading parameter of indicator 2 to be negative, while all other factor loading parameters are positive (remember that the first factor loading parameter is restricted to be greater than zero).

Furthermore, Section 7.4 deals with a two factor latent variable model. For this purpose five more indicators are used whose associated questions of the social survey and possible categories are given in Table 7.2. Since less people filled out all questions, the total number of observations is $N = 6804$. The corresponding response patterns of all ten indicators are depicted in Figure 7.2. This second set of five indicators forms a latent construct which is supposed to measure the respondents' ambition to achieve something in their job and in society. We refrain from a detailed discussion of the response patterns and the Spearman correlation coefficients because these five indicators are solely used in one analysis in Section 7.4 in order to demonstrate that a two factor LVM can also be estimated; furthermore this analysis shows that these five indicators indeed measure another latent construct.

Now let us move on to the indirect covariates used in our analyses. We employ four different covariates which are supposed to have an influence on the value of the latent constructs. There are 2 categorical covariates (*Sex*, *Inc* representing income), one metric covariate (*Age*) and one spacial covariate (*Reg* representing region), all summarized in Table 7.4. The covariate *Inc* is based on Germany's old currency, the "Deutsche Mark" (DM), and one EURO equals 1.95583 DM. The spatial covariate *Reg* is determined by the 402 regions (excluding the island of Rügen) with different license plate prefixes indicating the living area. Response patterns of the covariates *Sex*/*Inc*/*Age* and *Reg* can be found in Figures 7.3 and 7.5, respectively. We see that more than three times as many men as women have participated in the study, that most of the citizens' incomes fall almost equally into the two center income categories, that the ages 30–45 contributed most to the survey, and participation drops for ages lower than 25 or higher than 55 years. On the regional map, participation critically depends on the number of people living in that area. Rural areas (e. g. northeastern Germany, eastern Bavaria) have low response rates compared to cities and areas with a high population density (e. g. the Ruhrgebiet).

Finally, we want to illustrate how the covariate *Inc* depends on the covariates *Age* and *Reg*, respectively. In order to be able to calculate income averages for each year of age and region, we make the following assumptions about the individual income categories: a participant with income category 1 will have an expected income of 1500 DM, categories 2, 3 and 4 correspond to 3500 DM, 6500 DM and 9000 DM, respectively. Using these assumptions, the graphs plotted in Figure 7.6 indicate the dependence of the income from age and region. The average income increases steadily from low ages up to the age of 55, and remains on this level for higher ages. The regional map shows that people in eastern Germany still have lower average incomes than their neighbors in western Germany. It has to be noted that the age structure of respondents in all regions is very similar. These descriptive results on the dependence of income and age/region will be used for interpretation purposes in the analyses presented in the next two sections.

Indicators used in one and two latent variable models				
No.	**Indicator (type)**	**Question/statement**	**Response categories**	
1	*System* (binary)	Which type of social system would you prefer ?	1.	The state guarantees each citizen a sufficient social coverage. The associated costs are payed by all citizens in the form of taxes and contributions according to their income.
			2.	Citizens can decide by themselves if and to which extent they want to cover themselves and their families in the case of illness, unemployment, retirement and nursing. Everybody who is not insured in order to save contributions will have to bear the risks.
2	*Initiative* (ordinal)	I think it's correct that in future every individual must increasingly take care about his/her provision for old age than it's the case today.	1. 2, 3, 4, 5, 6.	Absolutely true. ... Absolutely wrong.
3	*Retirement* (ordinal)	To what extent should the state take care of the provision of old age ?	1. 2, 3, 4, 5, 6.	Completely. ... Not at all.
4	*Emergency* (ordinal)	To what extent should the state take care of the citizens' protection in difficult life circumstances and emergencies ?	1. 2, 3, 4, 5, 6.	Completely. ... Not at all.
5	*Health* (ordinal)	To what extent should the state make provisions for the citizens' health care ?	1. 2, 3, 4, 5, 6.	Completely. ... Not at all.

Table 7.1: *Variable names, variable types, questions/statements, and response categories of the first five indicators used in all models with one and two latent variables in this chapter.*

Indicators used in the two latent variable model				
No.	**Indicator (type)**	**Question/statement**	**Response categories**	
6	*Perform* (ordinal)	I consider it important to perform better than other people.	1. 2, 3, 4, 5, 6.	Absolutely true, ... Absolutely wrong.
7	*Society* (ordinal)	I want to achieve something in the society.	1. 2, 3, 4, 5, 6.	Absolutely true. ... Absolutely wrong.
8	*Reputation* (ordinal)	How important is the following regarding your job: To gain respect and a good reputation in the public.	1. 2, 3, 4, 5, 6.	Absolutely important. ... Not important at all.
9	*Salary* (ordinal)	How important is the following regarding your job: a high salary.	1. 2, 3, 4, 5, 6.	Absolutely important. ... Not important at all.
10	*Career* (ordinal)	How important is the following regarding your job: To make a career.	1. 2, 3, 4, 5, 6.	Absolutely important. ... Not important at all.

Table 7.2: *Variable names, variable types, questions/statements, and response categories of the second set of five indicators used in the LVM with two latent variables.*

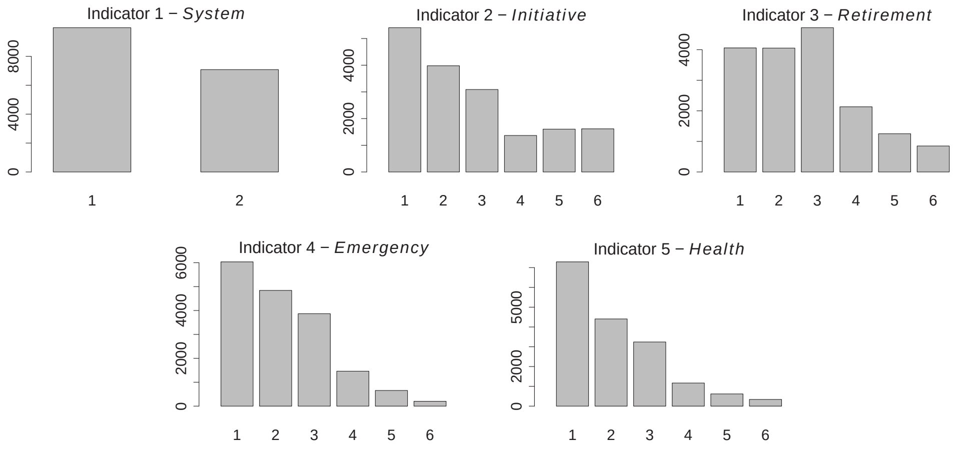


Figure 7.1: *Response patterns of the five indicators used in the LVM with one latent variable in Section 7.3 ($N = 17060$).*

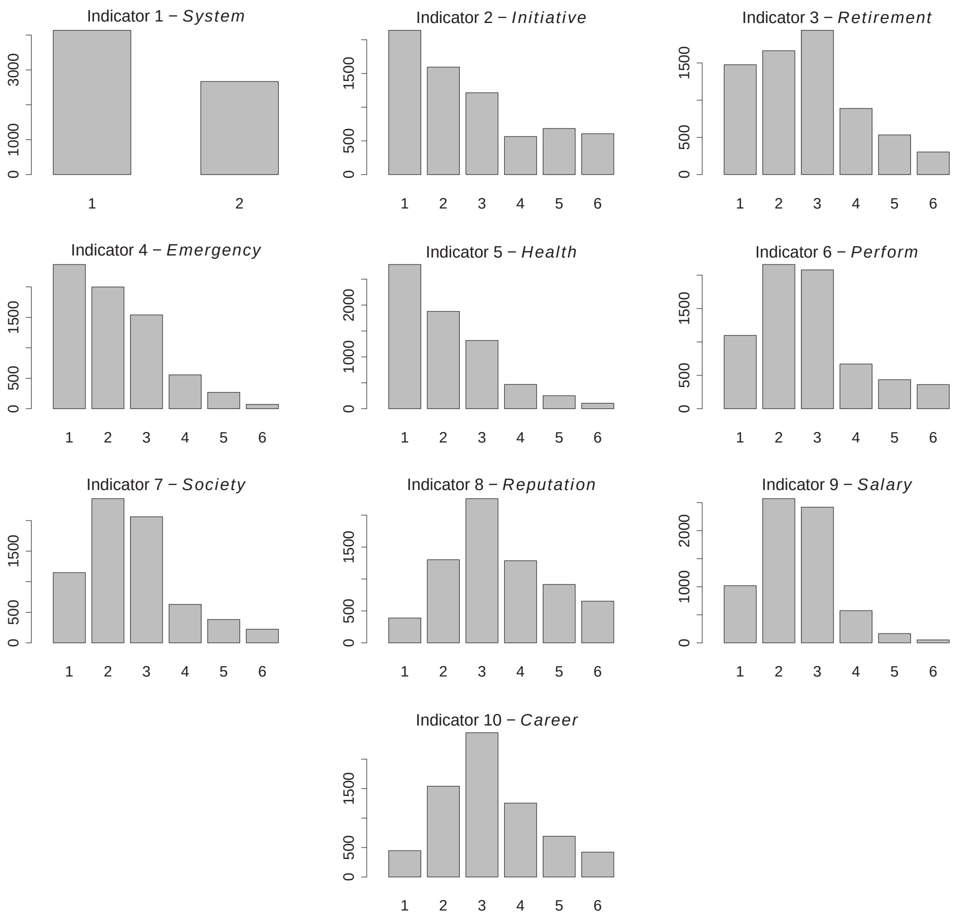


Figure 7.2: *Response patterns of all ten indicators used in the LVM with two latent variables in Section 7.4 ($N = 6804$).*

Indicator	**2.** *Initiative*	**3.** *Retirement*	**4.** *Emergency*	**5.** *Health*
1. *System*	-0.46	0.36	0.27	0.29
2. *Initiative*		-0.48	-0.27	-0.34
3. *Retirement*			0.43	0.52
4. *Emergency*				0.34

Table 7.3: *Corrected Spearman correlation coefficients between indicators 1, 2, 3, 4 and 5.*

Covariates used in the analyses		
Covariate name	**Covariate type**	**Response categories/range**
Sex	Categorical	1. Male 2. Female
Inc	Categorical	1. Less than 2500 DM net household income per month. 2. Between 2500 DM and 4500 DM net household income per month. 3. Between 4500 DM and 7500 DM net household income per month. 4. More than 7500 DM net household income per month.
Age	Metric	20, 21, ..., 70 years of age.
Reg	Spatial	1, 2, ..., 402 regions of Germany.

Table 7.4: *Variable names, variable types, and response categories of the four indirect covariates used in the analyses.*

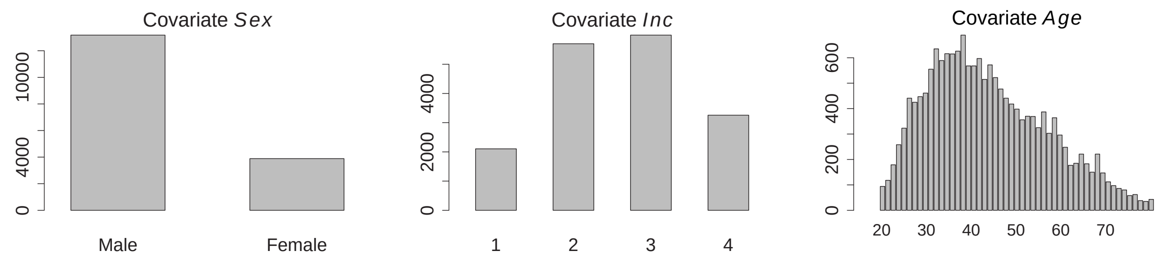


Figure 7.3: *Response patterns of covariates Sex, Inc, and Age in the analyses with one latent variable ($N = 17060$).*

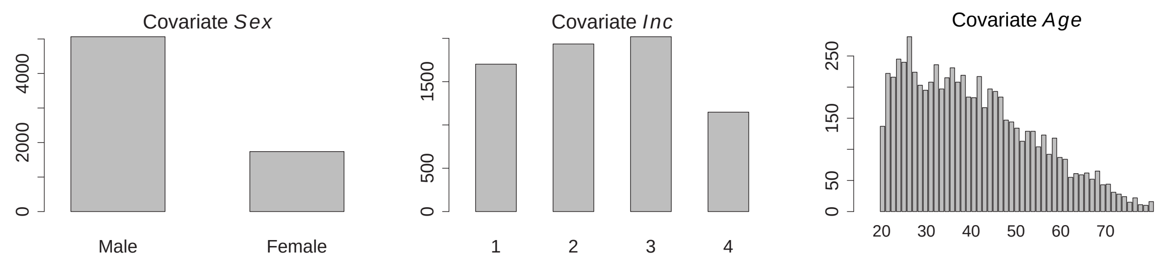


Figure 7.4: *Response patterns of covariates Sex, Inc, and Age in the analyses with two latent variables ($N = 6804$).*

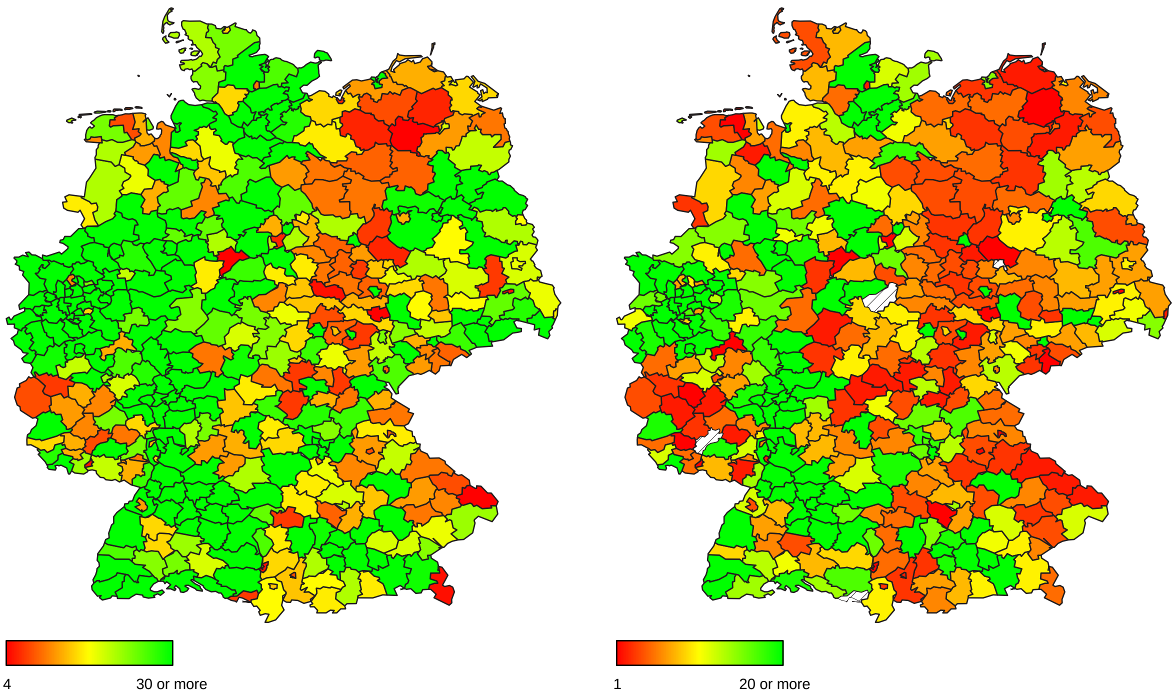


Figure 7.5: *Response patterns of covariate Reg for the analyses with one latent variable (left), and two latent variables (right).*

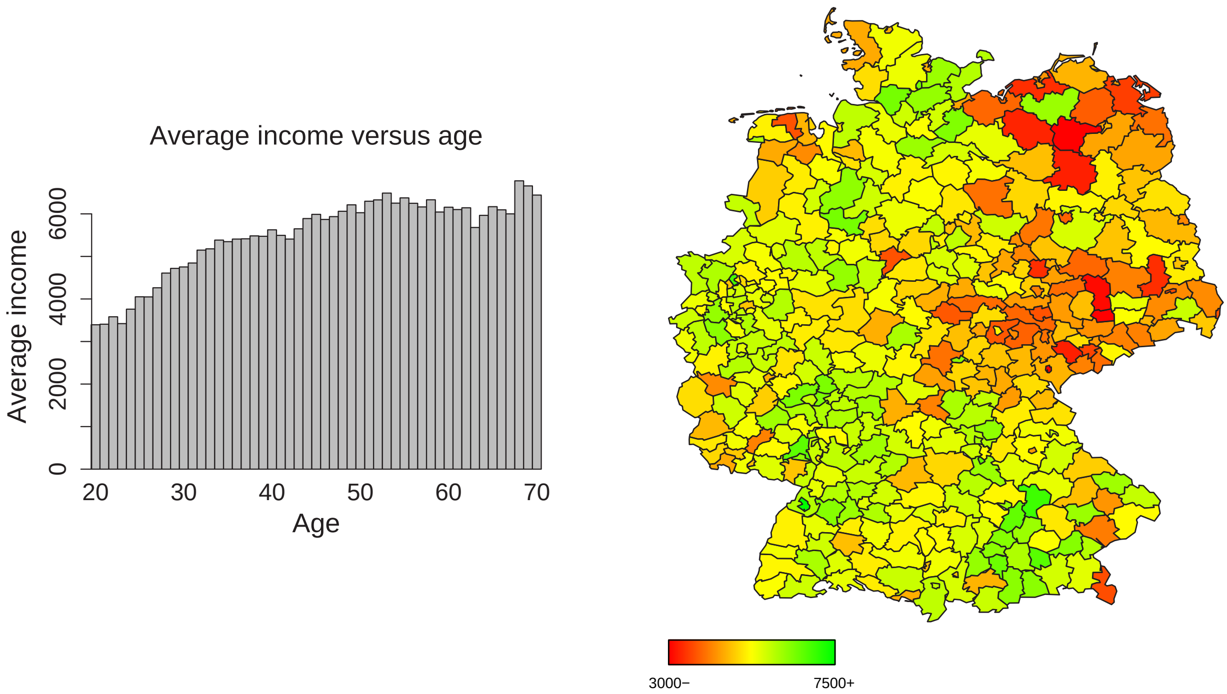


Figure 7.6: *Plots of average income in DM per years of age (left graph), and average income in DM per region (right graph) for the analyses with one latent variable ($N = 17060$).*

7.3 Model estimations with one latent variable

In this section, several models with one latent variable and various selections of covariates are estimated to show that it is possible to estimate the LVM for real-life data, and that it is useful for interpretation purposes to consider an extended predictor containing non-parametric effects influencing the latent variables. In the first subsection, a classic factor analysis model for ordinal variables without indirect effects is estimated. After that, two models with a predictor including indirect parametric effects *Sex*, *Inc*, *Age* and interactions are treated. The third subsection discusses the estimation of nonparametric effects, one analysis solely including the metric covariate *Age*, another analysis including the metric covariate *Age* and parametric covariates *Sex* and *Inc*. We proceed with the examination of the spatial covariate *Reg*. Again, two models are estimated – the first model only contains the spatial covariate *Reg*, the second model includes the estimation of the spatial covariate *Reg*, the metric covariate *Age*, and parametric effects *Sex* and *Inc*. We conclude with three models including interactions – the first model uses *Sex* as the interacting variable and *Age* as the effect modifier; the second model replaces *Sex* by *Inc*; the third model estimates a predictor with an interaction of *Sex* and *Age*, parametric effects *Inc* and the interaction of *Sex* and *Inc*, and a spatial effect based on *Reg*. The structural equations of all models estimated in this section are summarized in Table 7.5. The measurement model of all analyses in this section is

$$\begin{pmatrix} y_{i1}^* \\ y_{i2}^* \\ y_{i3}^* \\ y_{i4}^* \\ y_{i5}^* \end{pmatrix} = \begin{pmatrix} \lambda_{10} \\ \lambda_{20} \\ \lambda_{30} \\ \lambda_{40} \\ \lambda_{50} \end{pmatrix} + \begin{pmatrix} \lambda_{11} \\ \lambda_{21} \\ \lambda_{31} \\ \lambda_{41} \\ \lambda_{51} \end{pmatrix} \cdot \begin{pmatrix} z_{i1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \varepsilon_{i3} \\ \varepsilon_{i4} \\ \varepsilon_{i5} \end{pmatrix} \quad \text{, with} \quad \varepsilon_{ij} \sim N(0,1)\,.$$

Since a complete representation of all parameter estimates required too much space, only estimates of the factor loadings and indirect covariates are presented in the main text, whereas estimates of the less interesting parameters like intercepts and cutpoints can be found in appendix D.

7.3.1 Traditional factor analysis without covariates

We start with the easiest model possible, a classic factor analysis for binary and ordinal indicators. Hence, the predictor of the structural equation part of the model yields

$$\eta_i = 0\,. \tag{M1}$$

Estimates of factor loadings including the resulting communalities are depicted in Table 7.6. Looking at the estimated mean factor loadings and the respective communalities, it follows that the variation of all five indicators is based significantly on the latent construct. In

Section	Model	Predictor of indirect effects	Results in
7.3.1	M1	$\eta = 0$	Tab 7.6
7.3.2	M2a	$\eta = Sex + Inc + Age$	Tab 7.7
	M2b	$\eta = Sex + Inc + Age + Sex * Inc +$ $Sex * Age + Inc * Age$	Tab 7.8
7.3.3	M3a	$\eta = f(Age)$	Fig 7.7, Tab 7.9
	M3b	$\eta = Sex + Inc + f(Age)$	Fig 7.8, Tab 7.10
7.3.4	M4a	$\eta = f_{spatial}(Reg)$	Fig 7.9, Tab 7.11
	M4b	$\eta = Sex + Inc + f(Age) + f_{spatial}(Reg)$	Fig 7.10/7.11, Tab 7.12
7.3.5	M5a	$\eta = Sex * f(Age)$	Fig 7.12, Tab 7.13
	M5b	$\eta = Inc * f(Age)$	Fig 7.13, Tab 7.14
	M5c	$\eta = Inc + Sex * Inc + Sex * f(Age) +$ $f_{spatial}(Reg)$	Fig 7.14/7.15, Tab 7.15

Table 7.5: *Overview of the predictors in the structural equation of all estimated models with one latent variable in Section 7.3.*

this respect, parameters 3 and 4 stand out. Indicator 3 (*Retirement*) has the highest factor loading with a communality of 0.69 because this question regarding the old-age provision aims very closely at the idea of the latent construct. The question related to parameter 4 (*Emergency*), on the other hand, hints into a slightly different direction than the questions belonging to the other 4 indicators, thus its factor loading is the lowest.

Furthermore, the mean values almost perfectly agree with the mode values indicating a symmetric distribution of the factor loading samples. This fact, the narrow standard deviation, and 10%- and 90%-quantiles demonstrate the significance of the factor loading parameters, and strongly support the idea of a latent construct being responsible for the variation of the five indicators. As expected, the highest factor loading of indicator 3 shows the highest absolute standard deviation.

In order to check the basic validity of those estimates, a classic factor analysis was carried out. The results showed communalities which were 10%–35% below the estimates of our model. The reason for this inferior result lies in the simplification of treating the ordinal indicators as metric ones. In particular, the communality of binary indicator 1 with a value of 35% below the correct estimate confirms that assuming binary indicators to be metric produces very weak results.

7.3.2 LVM with parametric indirect effects

Now, the classic factor analysis model is extended by introducing indirect parametric covariates modifying the latent construct. Two models are analyzed – the predictor of the first model M2a contains the categorical covariates *Sex* and *Inc*, and the metric covariate *Age* which is treated as a parametric and hence linear effect; the second model M2b additionally includes some interactions between those variables. The predictors of the two models are defined as

$$\eta_i = \gamma_1 \cdot Sex2_i + \gamma_2 \cdot Inc2_i + \gamma_3 \cdot Inc3_i + \gamma_4 \cdot Inc4_i + \gamma_5 \cdot Age_i \,, \qquad \text{(M2a)}$$

and

$$\begin{aligned}\eta_i =& \gamma_1 \cdot Sex2_i + \gamma_2 \cdot Inc2_i + \gamma_3 \cdot Inc3_i + \gamma_4 \cdot Inc4_i + \gamma_5 \cdot Age_i + \\ & \gamma_6 \cdot Sex2_i Inc2_i + \gamma_7 \cdot Sex2_i Inc3_i + \gamma_8 \cdot Sex2_i Inc4_i + \gamma_9 \cdot Sex2_i Age_i \\ & \gamma_{10} \cdot Inc2_i Age_i + \gamma_{11} \cdot Inc3_i Age_i + \gamma_{12} \cdot Inc4_i Age_i \,.\end{aligned} \qquad \text{(M2b)}$$

Standard dummy coding is used – for example $Inc3_i$ is set to 1 if observation i belongs to the third income category, otherwise $Inc3_i$ is set to zero. Estimates of the factor loadings and parametric indirect effects are summarized in Table 7.7 for M2a, and in Table 7.8 for M2b.

Let us start with the discussion of model M2a. First of all, it is conspicuous that the estimates of all factor loadings are slightly lower than for the pure factor analysis without indirect covariates. However, this does not mean that this model is inferior to the traditional factor analysis. In a traditional factor analysis model, the indicators response solely determines the value of the latent variable. In any model with covariates however, the covariates additionally influence the value of the latent variable. Specifically, the normally distributed latent variable is not centered statically at zero like in a traditional factor analysis, but the mean of the distribution equals the value of the predictor in the structural part of the model. Thus, the latent variables cover a broader range of values and thus exert a greater influence on the variability of the indicators, even if the factor loadings are slightly lower. In a sense, the covariates explain some of the correlation among the indicators. Using covariates therefore allows more detailed statements on the dependence structure of the latent variable.

We proceed with the discussion of the estimates of the parametric indirect effects. First of all, females seem to have leaning towards a strong state because the mean of the latent factor is by 0.359 lower than for males. Furthermore, the covariate income exerts a very strong influence on the latent construct. With increasing income, the mean of the latent factor increases considerably, for example about 1.098 for a person in income category 4 compared to the reference category 1. This effect can be explained by the fact that people with high incomes generally show a higher inititiative of their own and a higher readiness to take risks than people with lower incomes. In addition, big earners make high monetary contributions to the social system without getting an adequate service in return. Finally, increasing age has a negative influence on the mean of the latent factor, hence older people tend to prefer a stronger state taking care of the social services. At this point, it is doubtful if the assumption of a parametric linear effect of age on the latent construct is justified. This question is further examined in the next section where the effect of age is modeled by a smooth nonparametric function.

The parameter estimates of model M2b including interactions between the covariates show slightly different results than model M2a. The effect for females ($Sex2$) is even more negative, and the pure income effect seems to be less pronounced than in the model excluding interactions. However, these effects are compensated by the positive effects of the interaction between females and age ($Sex2Age$), and the positive interaction between income and age ($Inc2Age$, $Inc3Age$, $Inc4Age$). The probably most useful interaction between females and income shows that for higher income categories, females have a bias towards a state provision system compared to males in the same income category.

7.3.3 LVM with nonparametric effect of a metric covariate

In this section, two models with nonparametric effects for metric covariates are calculated. The first model solely employs the covariate age as a nonparametric function. In order to demonstrate the effect of different prior settings, first- and second-order random walks and a P-splines prior of degree 3 with 10 intervals are used for both models. The second model additionally includes the parametric covariates *Sex* and *Inc* which were used in the last section. Here, only a P-splines prior is used. The predictors are defined as

$$\eta_i = f(Age_i)\,, \tag{M3a}$$

and

$$\eta_i = \gamma_1 \cdot Sex2_i + \gamma_2 \cdot Inc2_i + \gamma_3 \cdot Inc3_i + \gamma_4 \cdot Inc4_i + f(Age_i)\,. \tag{M3b}$$

Concerning model M3a, estimates of the factor loadings and smoothing parameters can be found in Table 7.9, plots of the smooth function estimates are depicted in Figure 7.7.

The prior choice of the smooth functions does not exert a big influence on the factor loadings which are very similar for all three prior settings. The estimated smoothing parameter κ, however, depends on the type of function employed. In general, the smoothing parameter of P-splines priors is substantially higher than for random walk priors due to its lower amount of parameters. Looking at the three nonparametric functions, we see that all three prior settings yield the same basic functional form. However, the smoothness of the three functions differs significantly. The first-order random walk prior produces a wiggly function and hence fits the actual data rather closely. The function of the second-order random walk is smoother, as expected. The smoothest function is obtained by the P-splines prior which reduces the danger of overfitting the analyzed data. All obtained functions basically vary about the zero horizontal line, and are therefore not very distinct; a weak trend could be detected for young people who tend to prefer a slightly stronger state as opposed to older people who would rather take care of themselves. The latter trend could be explained by the fact that old people participating at the survey are very energetic and progressive because they have online access and take part in the survey; furthermore older people's income is higher on average than those of young people as shown in the previous chapter.

Since the factor loading parameters are very similar to the case of the traditional factor analysis without covariates and the age effect is not very pronounced, the results implicate that a model solely containing the covariate *Age* is not very suited to sufficiently explain an influence on the latent variable.

In the second model M3b, the parametric covariates *Sex* and *Inc* are included on top of the metric covariate *Age*. We used a model with P-spline priors whose parameter estimates can be found in Figure 7.8 and Table 7.10, respectively. As in the last section, factor loading parameters are slightly reduced compared to the traditional factor analyis model, and regression coefficients for the parametric covariates are very similar to the pure parametric analysis. The smooth function representing the age effect now shows a distinctive sinusoidal shape. Given the same sex and income, young people prefer to take care of themselves, while medium aged people shortly before retirement seem to prefer a strong state. Moving to even older ages, respondents seem to become more progressive again. This result makes sense because young people suffer from the contributions to the social systems and do not get an appropriate service in return when they are old due to the ageing population. Additionally, young people are prepared to take bigger risks than older people. Medium-aged people around the age between 40 and 60 have to take care of their children, maybe pay mortgages on their property, save for their retirements, and hence would suffer tremendously from a loss of job. Therefore those people would rather avoid risks and prefer a stronger state. After those burdens have been endured, people older than 60 are more willing to take risks and save contributions again. This smooth function of the age effect also demonstrates that a big error is made if the age is only treated by a linear, parametric effect as in the last section. The parametric effect of the last section only reproduces the almost linear drop between the age of 30 and 50 which is the range of age where most of the observations are located.

7.3.4 LVM with nonparametric effect of a spatial covariate

In this section, models with the spatial covariate *Reg* are treated. Again, two models are analyzed. The first model solely employs the spatial covariate *Reg*. The second model finally treats all covariates, hence the predictor includes the parametric effects *Sex* and *Inc*, a function of the metric covariate *Age* (modeled with a P-splines prior of degree 3 and 10 intervals), and the spatial effect of covariate *Reg*. The predictors are

$$\eta_i = f_{spatial}(Reg_i)\,, \qquad \text{(M4a)}$$

and

$$\eta_i = \gamma_1 \cdot Sex2_i + \gamma_2 \cdot Inc2_i + \gamma_3 \cdot Inc3_i + \gamma_4 \cdot Inc4_i + f(Age_i) + f_{spatial}(Reg_i)\,. \qquad \text{(M4b)}$$

Starting with model M4a, parameter estimates for the factor loadings and smoothing parameter can be found in Table 7.11, the spatial plot of the function estimates of *Reg* is drawn in Figure 7.9. The estimates of the factor loadings are very similar to the estimates of the factor analysis model without covariates. Hence we expect the spatial effect not to be very strong. The two maps clearly show a north-south divide regarding the spatial influence on the latent variable. In the south and in some parts of Hessen – where the economic position is strong and governments are conservative – more people would like to take care of their own provisions than in eastern Germany and some parts of northern West Germany where economic conditions are comparably poor. Additionally, one might expect that people in eastern Germany would demand a strong state because they were used to an omnipotent state for several decades.

Model 4b includes all discussed covariates (*Sex*, *Inc*, *Age*, and *Region*). The estimates for the factor loadings, parametric regression coefficients, and smoothing parameters are summarized in Table 7.12, and the smooth function for age (modeled by a P-splines prior of degree 3 and 10 intervals) and the spatial function of covariate region are drawn in Figures 7.10 and 7.11, respectively. The estimates of the factor loadings, the parametric regression coefficients and the smooth function of covariate *Age* are very similar to the results of model M3b in the last section. However, the spatial effect changed with respect to model M4a. The south still prefers a system based on the inititiative of one's own, but now the people wishing a strong state come from north-west Germany, an area which was traditionally governed by the social-democratic party of Germany. Surprisingly in eastern Germany, the significant negative effect on the latent variable has disappeared; obviously in model M4b the covariate *Inc* accounts for the negative effect in eastern Germany in model M4a.

7.3.5 LVM with nonparametric effect of interacting covariates

Finally, the estimation of interacting covariates of "PD1" is carried out in three different models. The first model contains two smooth nonparametric functions of the age effect for males and females, i. e.

$$\eta = f_1(Age) + f_2(Age) \cdot Sex2\,; \qquad \text{(M5a)}$$

in the second model *Inc* replaces *Sex* as the interacting variable, hence

$$\eta = f_1(Age) + f_2(Age) \cdot Inc2 + f_3(Age) \cdot Inc3 + f_4(Age) \cdot Inc4\,; \qquad \text{(M5b)}$$

the third model entails all covariates of the previous analyses with

$$\begin{aligned}\eta =& \gamma_1 \cdot Inc2 + \gamma_2 \cdot Inc3 + \gamma_3 \cdot Inc4 + \gamma_4 \cdot Sex2Inc2 + \\ & \gamma_5 \cdot Sex2Inc3 + \gamma_6 \cdot Sex2Inc4 + \\ & f_1(Age) + f_2(Age) \cdot Sex2 + f_{spatial}(Reg)\,.\end{aligned} \qquad \text{(M5C)}$$

For all three models a P-splines prior of degree 3 with 10 intervals is used for the nonparametric functions of metric covariates.

The estimates of model M5a regarding the factor loadings and smoothing parameters can be found in Table 7.13, and the plots of the two nonparametric functions are drawn in Figure 7.12. The basic age effect $f_1(Age)$ for males is very similar to the nonparametric function estimate for the covariate *Age* in the predictor (see model M3a). For females, the added effect $f_2(Age)$ on the latent variables is negative as expected; however, a slight rise in the mean of the latent construct can be observed for increasing age. The credible interval of f_2 is rather broad for the VCM, especially for very high ages of females due to the low number of observations.

The results of model M5b are shown in Table 7.14 and Figure 7.13, respectively. The estimates for the factor loadings are very similar to model M5a. The age effect of the reference income category *Inc*1 indicates that young people rather tend to a provision system based on own initiative, and with rising age people would rather prefer provisions managed by the state. The three functions for the other income categories show that the higher the income, the higher the probability of people voting for more responsibility to manage their provisions. Furthermore, in higher income categories the negative effect of the basis function f_1 for people of higher age on the latent construct is somewhat balanced.

Finally the analysis of model M5c is summarized in Table 7.15 and Figures 7.14 and 7.15. We want to refrain from a discussion of the estimates because the results are analogous to previous analyses.

The analyses in this section clearly show that a predictor containing a variety of covariate types can be estimated and leads to useful results.

Parameter		Mean	Std. dev.	10% quantile	Mode	90% quantile	Communality
Factor loadings							
1. *System*	λ_{11}	0.847	0.019	0.823	0.847	0.871	0.42
2. *Initiative*	λ_{21}	-0.891	0.015	-0.911	-0.891	-0.872	0.44
3. *Retirement*	λ_{31}	1.479	0.028	1.444	1.478	1.515	0.69
4. *Emergency*	λ_{41}	0.693	0.012	0.678	0.693	0.709	0.32
5. *Health*	λ_{51}	0.971	0.016	0.949	0.971	0.992	0.49

Table 7.6: *Results of model M1 with $\eta = 0$ – estimates of factor loadings.*

Parameter		Mean	Std. dev.	10% quantile	Mode	90% quantile
Factor loadings						
1. *System*	λ_{11}	0.804	0.017	0.782	0.803	0.826
2. *Initiative*	λ_{21}	-0.850	0.015	-0.869	-0.850	-0.832
3. *Retirement*	λ_{31}	1.324	0.024	1.293	1.324	1.356
4. *Emergency*	λ_{41}	0.647	0.012	0.631	0.647	0.662
5. *Health*	λ_{51}	0.915	0.015	0.896	0.914	0.934
Parametric indirect effects						
*Sex*2	γ_1	-0.359	0.022	-0.387	-0.358	-0.331
*Inc*2	γ_2	0.281	0.030	0.243	0.280	0.319
*Inc*3	γ_3	0.610	0.030	0.572	0.609	0.649
*Inc*4	γ_4	1.098	0.034	1.055	1.097	1.142
Age	γ_5	-0.009	0.001	-0.010	-0.009	-0.008

Table 7.7: *Results of model M2a with $\eta = Sex + Inc + Age$ – estimates of factor loadings and parametric indirect effects.*

Figure 7.7: *Results of model M3a with $\eta = f(Age)$ – estimates of the smooth function $f(Age)$ modeled by three different prior settings. The mean values are connected by the solid line, 10%- and 90%-quantiles are connected by the dashed line.*

Parameter		Mean	Std. dev.	10% quantile	Mode	90% quantile
Factor loadings						
1. *System*	λ_{11}	0.799	0.017	0.777	0.799	0.821
2. *Initiative*	λ_{21}	-0.847	0.015	-0.866	-0.847	-0.828
3. *Retirement*	λ_{31}	1.321	0.025	1.290	1.320	1.353
4. *Emergency*	λ_{41}	0.645	0.012	0.630	0.645	0.660
5. *Health*	λ_{51}	0.915	0.015	0.895	0.915	0.934
Parametric indirect effects						
*Sex*2	γ_1	-0.699	0.092	-0.816	-0.700	-0.582
*Inc*2	γ_2	0.158	0.108	0.018	0.159	0.295
*Inc*3	γ_3	0.353	0.108	0.212	0.353	0.493
*Inc*4	γ_4	0.604	0.127	0.441	0.602	0.772
Age	γ_5	-0.018	0.002	-0.021	-0.018	-0.015
*Sex*2*Inc*2	γ_6	-0.085	0.063	-0.165	-0.085	-0.005
*Sex*2*Inc*3	γ_7	-0.150	0.066	-0.236	-0.152	-0.064
*Sex*2*Inc*4	γ_8	-0.154	0.080	-0.257	-0.156	-0.052
*Sex*2*Age*	γ_9	0.012	0.002	0.009	0.012	0.015
*Inc*2*Age*	γ_{10}	0.004	0.003	0.001	0.004	0.008
*Inc*3*Age*	γ_{11}	0.008	0.003	0.005	0.008	0.012
*Inc*4*Age*	γ_{12}	0.014	0.003	0.010	0.014	0.018

Table 7.8: *Results of model M2b with $\eta = Sex + Inc + Age + Sex * Inc + Sex * Age + Inc * Age$ – estimates of factor loadings and parametric indirect effects.*

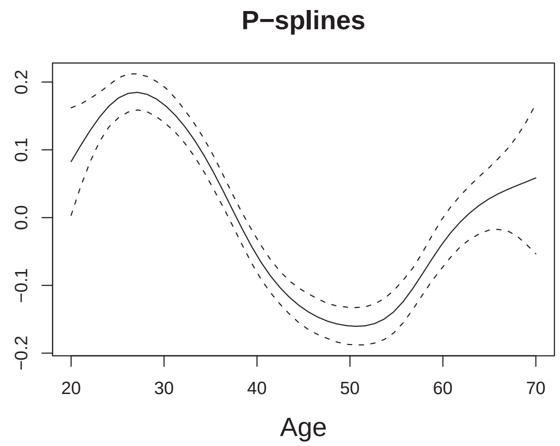


Figure 7.8: *Results of model M3b with $\eta = Sex + Inc + f(Age)$ – estimates of the smooth function $f(Age)$ modeled by a P-splines prior. The mean values are connected by the solid line, 10%- and 90%-quantiles are connected by the dashed line.*

First-order random walk prior						
Parameter		**Mean**	**Std. dev.**	**10% quantile**	**Mode**	**90% quantile**
1. *System*	λ_{11}	0.845	0.019	0.821	0.845	0.869
2. *Initiative*	λ_{21}	-0.892	0.015	-0.912	-0.892	-0.872
3. *Retirement*	λ_{31}	1.468	0.028	1.432	1.468	1.503
4. *Emergency*	λ_{41}	0.692	0.013	0.675	0.692	0.709
5. *Health*	λ_{51}	0.968	0.017	0.946	0.968	0.989
Smoothing par.	κ_1	$2.23 \cdot 10^{-3}$	$1.07 \cdot 10^{-3}$	$1.14 \cdot 10^{-3}$	$1.99 \cdot 10^{-3}$	$3.57 \cdot 10^{-3}$

Second-order random walk prior						
Parameter		**Mean**	**Std. dev.**	**10% quantile**	**Mode**	**90% quantile**
1. *System*	λ_{11}	0.844	0.019	0.820	0.844	0.868
2. *Initiative*	λ_{21}	-0.891	0.015	-0.911	-0.891	-0.872
3. *Retirement*	λ_{31}	1.472	0.030	1.435	1.472	1.511
4. *Emergency*	λ_{41}	0.691	0.012	0.676	0.691	0.707
5. *Health*	λ_{51}	0.966	0.017	0.945	0.965	0.987
Smoothing par.	κ_1	$5.43 \cdot 10^{-4}$	$3.28 \cdot 10^{-4}$	$2.46 \cdot 10^{-4}$	$4.56 \cdot 10^{-4}$	$9.35 \cdot 10^{-4}$

P-splines prior of degree 3 and 10 intervals						
Parameter		**Mean**	**Std. dev.**	**10% quantile**	**Mode**	**90% quantile**
1. *System*	λ_{11}	0.843	0.019	0.819	0.842	0.867
2. *Initiative*	λ_{21}	-0.890	0.016	-0.911	-0.891	-0.870
3. *Retirement*	λ_{31}	1.470	0.030	1.432	1.470	1.508
4. *Emergency*	λ_{41}	0.693	0.013	0.677	0.693	0.709
5. *Health*	λ_{51}	0.968	0.016	0.947	0.968	0.989
Smoothing par.	κ_1	0.025	0.032	0.007	0.017	0.049

Table 7.9: *Results of model M3a with $\eta = f(Age)$ – estimates of factor loadings and smoothing parameter of $f(Age)$ which is modeled by three different prior-settings first-order random walk (top), second-order random walk (middle), and P-splines (bottom).*

P-splines prior of degree 3 and 10 intervals						
Parameter		Mean	Std. dev.	10% quantile	Mode	90% quantile
Factor loadings						
1. *System*	λ_{11}	0.804	0.018	0.781	0.803	0.827
2. *Initiative*	λ_{21}	-0.852	0.015	-0.870	-0.852	-0.833
3. *Retirement*	λ_{31}	1.310	0.024	1.279	1.310	1.341
4. *Emergency*	λ_{41}	0.644	0.012	0.630	0.644	0.659
5. *Health*	λ_{51}	0.910	0.016	0.890	0.909	0.930
Parametric indirect effects						
*Sex*2	γ_1	-0.353	0.021	-0.381	-0.353	-0.326
*Inc*2	γ_2	0.278	0.030	0.241	0.278	0.316
*Inc*3	γ_3	0.619	0.030	0.580	0.619	0.657
*Inc*4	γ_4	1.117	0.034	1.073	1.117	1.161
Smoothing par.	κ_1	0.019	0.022	0.004	0.012	0.040

Table 7.10: *Results of model M3b with* $\eta = Sex + Inc + f(Age)$ *– estimates of factor loadings, parametric indirect effects, and smoothing parameter of* $f(Age)$ *which is modeled by a P-splines prior.*

Parameter		Mean	Std. dev.	10% quantile	Mode	90% quantile
Factor loadings						
1. *System*	λ_{11}	0.845	0.018	0.822	0.845	0.869
2. *Initiative*	λ_{21}	-0.889	0.015	-0.908	-0.889	-0.869
3. *Retirement*	λ_{31}	1.446	0.028	1.410	1.445	1.482
4. *Emergency*	λ_{41}	0.687	0.013	0.670	0.687	0.703
5. *Health*	λ_{51}	0.965	0.017	0.943	0.964	0.987
Smoothing parameters						
Smoothing par.	κ_{region}	0.030	0.008	0.020	0.029	0.041

Table 7.11: *Results of model M4a with* $\eta = f_{spatial}(Reg)$ *– estimates of factor loadings and smoothing parameter of* $f_{spatial}(Reg)$ *which is modeled by a Markov random field prior.*

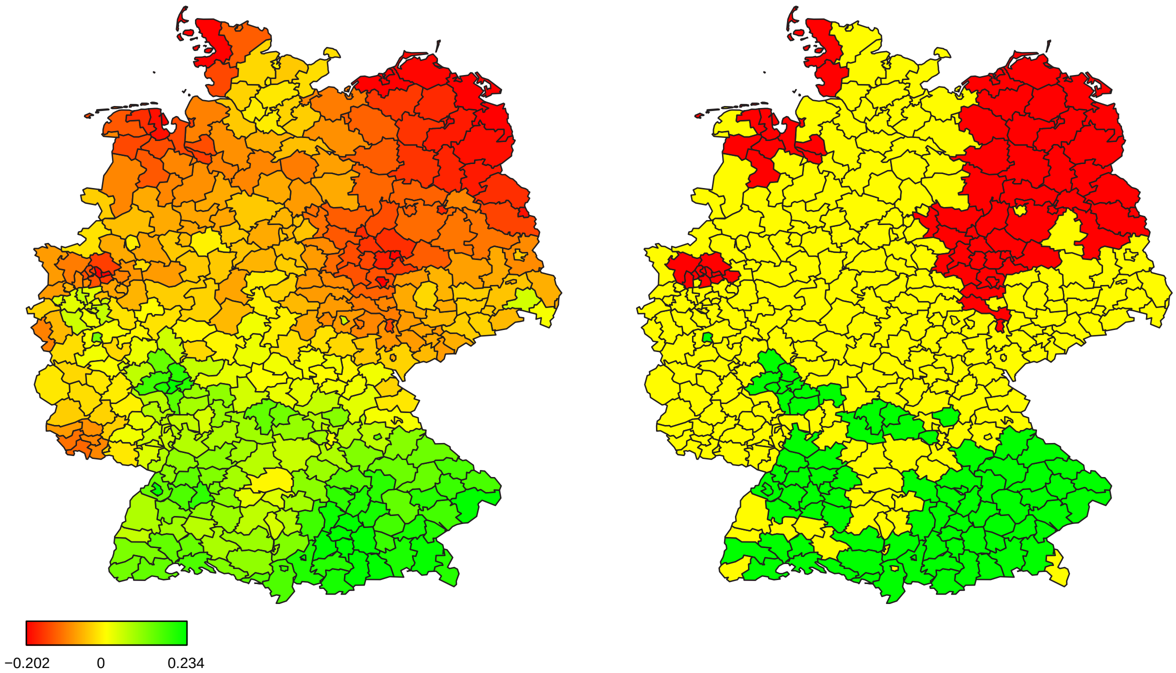


Figure 7.9: *Results of model M4a with* $\eta = f_{spatial}(Reg)$. ***Left:*** *Estimates of the spatial covariate Reg. Minimum and maximum values are set to 2.5% and 97.5% quantiles of the observed function estimates, respectively.* ***Right:*** *Plot of regions with a significant negative effect (red), a significant positive effect (green), or a non-significant effect (yellow).*

Parameter		Mean	Std. dev.	10% quantile	Mode	90% quantile
Factor loadings						
1. *System*	λ_{11}	0.800	0.017	0.778	0.799	0.822
2. *Initiative*	λ_{21}	-0.849	0.014	-0.867	-0.849	-0.831
3. *Retirement*	λ_{31}	1.300	0.024	1.270	1.300	1.331
4. *Emergency*	λ_{41}	0.642	0.012	0.627	0.642	0.657
5. *Health*	λ_{51}	0.905	0.016	0.885	0.904	0.925
Parametric indirect effects						
*Sex*2	γ_1	-0.359	0.022	-0.386	-0.359	-0.330
*Inc*2	γ_2	0.270	0.030	0.232	0.271	0.307
*Inc*3	γ_3	0.609	0.031	0.569	0.609	0.649
*Inc*4	γ_4	1.096	0.035	1.051	1.096	1.140
Smoothing parameters						
Smoothing par.	κ_{age}	0.020	0.025	0.005	0.013	0.041
Smoothing par.	κ_{region}	0.018	0.006	0.011	0.017	0.025

Table 7.12: *Results of model M4b with* $\eta = Sex + Inc + f(Age) + f_{spatial}(Reg)$ *– estimates of factor loadings, parametric indirect effects, and smoothing parameters of* $f(Age)$ *(modeled by a P-Splines prior) and* $f_{spatial}(Reg)$ *(modeled by a Markov random field prior).*

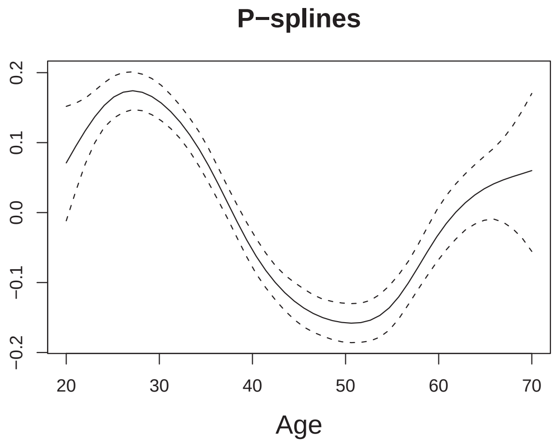


Figure 7.10: *Results of model M4b with* $\eta = Sex + Inc + f(Age) + f_{spatial}(Reg)$ *– estimates of the smooth function* $f(Age)$ *modeled by a P-splines prior. The mean values are connected by the solid line, 10%- and 90%-quantiles are connected by the dashed line.*

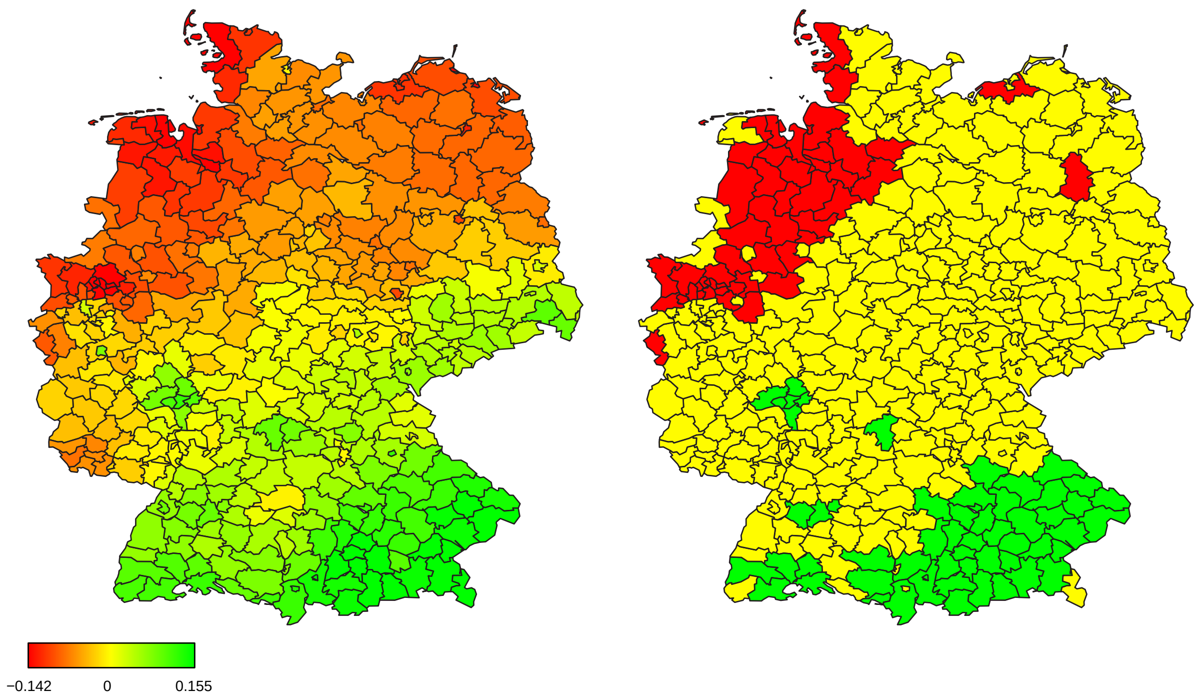


Figure 7.11: *Results of model M4b with* $\eta = Sex + Inc + f(Age) + f_{spatial}(Reg)$. ***Left****: Estimates of the spatial covariate Reg. Minimum and maximum values are set to 2.5% and 97.5% quantiles of the observed function estimates, respectively.* ***Right****: Plot of regions with a significant negative effect (red), a significant positive effect (green), or a non-significant effect (yellow).*

Parameter		Mean	Std. dev.	10% quantile	Mode	90% quantile
Factor loadings						
1. *System*	λ_{11}	0.822	0.018	0.799	0.822	0.845
2. *Initiative*	λ_{21}	-0.871	0.015	-0.891	-0.871	-0.852
3. *Retirement*	λ_{31}	1.455	0.028	1.419	1.455	1.491
4. *Emergency*	λ_{41}	0.681	0.013	0.665	0.681	0.697
5. *Health*	λ_{51}	0.957	0.016	0.936	0.957	0.978
Smoothing parameters						
Smoothing par.	κ_1	0.022	0.024	0.006	0.014	0.043
Smoothing par.	κ_2	0.007	0.016	0.001	0.003	0.015

Table 7.13: *Results of model M5a with $\eta = Sex * f(Age)$ – estimates of factor loadings and smoothing parameters of $f_1(Age)$ and $f_2(Age)$ which are modeled by P-splines priors.*

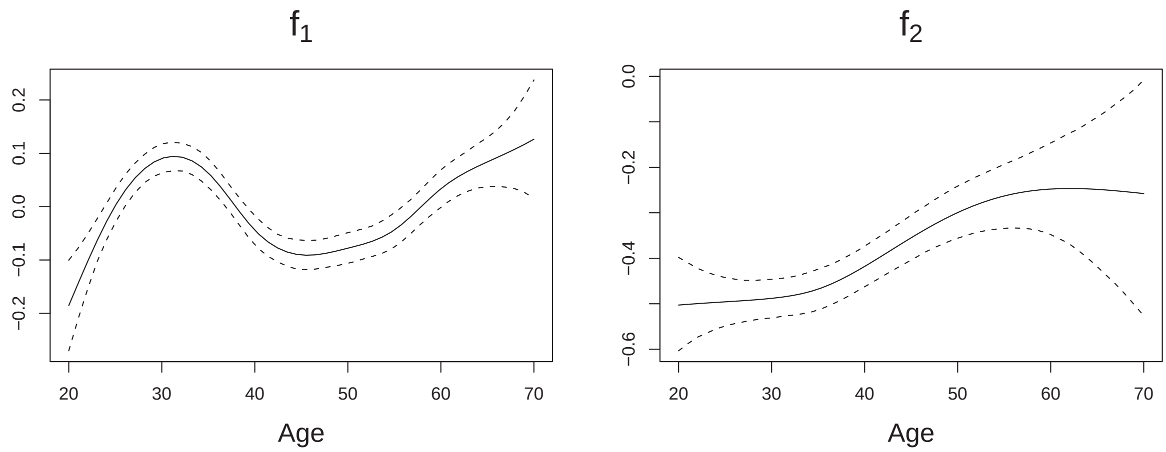


Figure 7.12: *Results of model M5a with $\eta = Sex * f(Age)$ – estimates of the smooth functions $f_1(Age)$ and $f_2(Age)$ modeled by P-splines priors. The mean values are connected by the solid line, 10%- and 90%-quantiles are connected by the dashed line.*

Parameter		Mean	Std. dev.	10% quantile	Mode	90% quantile
Factor loadings						
1. *System*	λ_{11}	0.815	0.018	0.792	0.815	0.838
2. *Initiative*	λ_{21}	-0.864	0.015	-0.882	-0.863	-0.845
3. *Retirement*	λ_{31}	1.317	0.024	1.286	1.317	1.347
4. *Emergency*	λ_{41}	0.652	0.012	0.636	0.652	0.667
5. *Health*	λ_{51}	0.913	0.015	0.894	0.913	0.933
Smoothing parameters						
Smoothing par.	κ_1	0.021	0.025	0.004	0.013	0.045
Smoothing par.	κ_2	0.025	0.041	0.002	0.013	0.059
Smoothing par.	κ_3	0.006	0.011	0.001	0.003	0.013
Smoothing par.	κ_4	0.019	0.025	0.003	0.012	0.040

Table 7.14: *Results of model M5b with $\eta = Inc * f(Age)$ – estimates of factor loadings and smoothing parameters of $f_1(Age)$, $f_2(Age)$, $f_3(Age)$, and $f_4(Age)$ which are modeled by P-splines priors.*

Parameter		Mean	Std. dev.	10% quantile	Mode	90% quantile
Factor loadings						
1. *System*	λ_{11}	0.797	0.017	0.775	0.797	0.819
2. *Initiative*	λ_{21}	-0.847	0.015	-0.866	-0.847	-0.828
3. *Retirement*	λ_{31}	1.299	0.024	1.269	1.299	1.331
4. *Emergency*	λ_{41}	0.639	0.012	0.624	0.639	0.654
5. *Health*	λ_{51}	0.903	0.015	0.884	0.904	0.923
Parametric indirect effects						
*Inc*2	γ_1	0.300	0.035	0.257	0.300	0.345
*Inc*3	γ_2	0.663	0.034	0.619	0.662	0.707
*Inc*4	γ_3	1.158	0.038	1.110	1.158	1.207
*Sex*2*Inc*2	γ_4	-0.085	0.064	-0.167	-0.086	-0.001
*Sex*2*Inc*3	γ_5	-0.163	0.066	-0.250	-0.162	-0.080
*Sex*2*Inc*4	γ_6	-0.205	0.080	-0.307	-0.203	-0.102
Smoothing parameters						
Smoothing par.	$\kappa_{agebasic}$	0.024	0.029	0.005	0.016	0.050
Smoothing par.	κ_{Sex2}	0.006	0.012	0.001	0.003	0.014
Smoothing par.	κ_{Reg}	0.017	0.006	0.010	0.016	0.024

Table 7.15: *Results of model M5c with $\eta = Inc + Sex * Inc + Sex * f(Age) + f_{spatial}(Reg)$ – estimates of factor loadings, parametric indirect effects, and smoothing parameters of $f_1(Age)$, $f_2(Age)$ (modeled by P-splines priors), and $f_{spatial}(Reg)$ (modeled by Markov random field priors).*

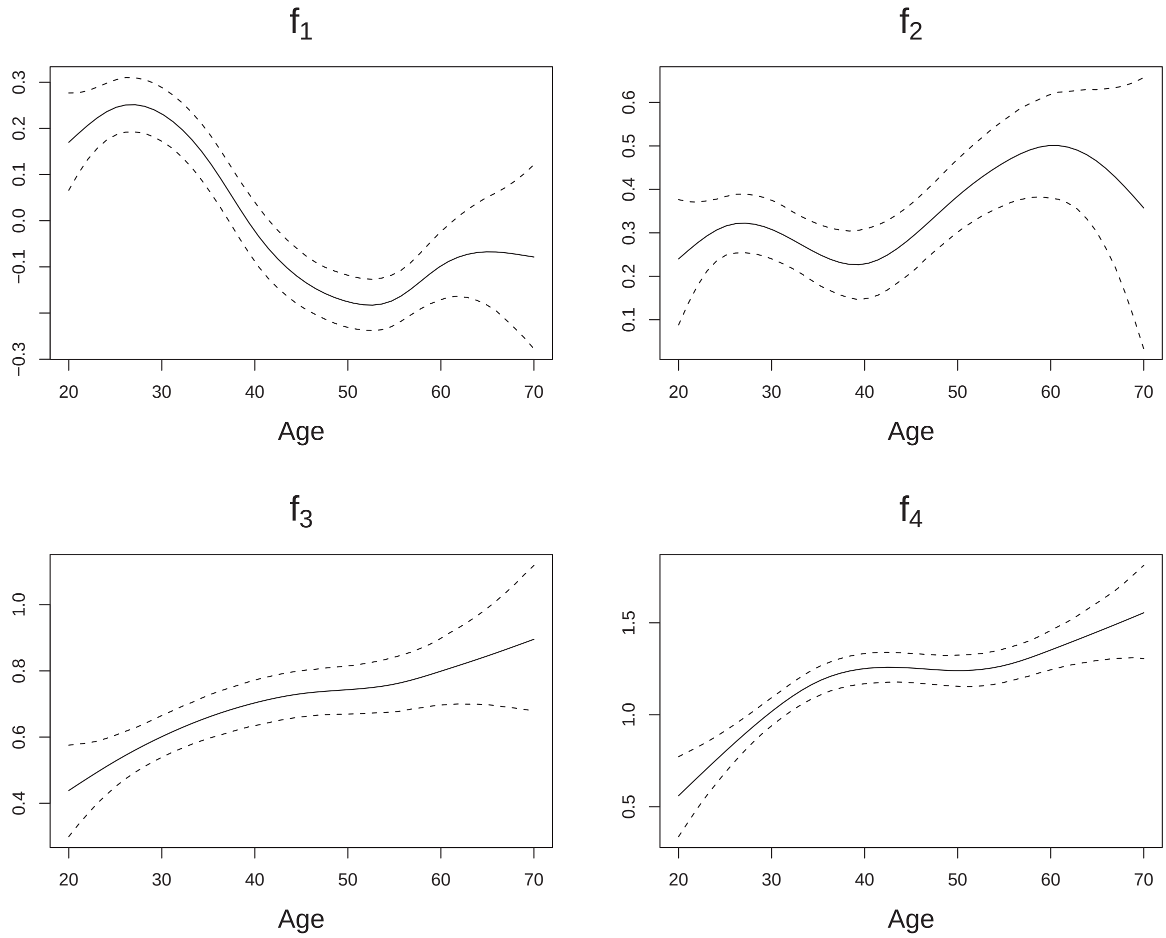


Figure 7.13: *Results of model M5b with* $\eta = Inc * f(Age)$ *– estimates of the smooth functions* $f_1(Age)$, $f_2(Age)$, $f_3(Age)$, *and* $f_4(Age)$ *modeled by P-splines priors. The mean values are connected by the solid line, 10%- and 90%-quantiles are connected by the dashed line.*

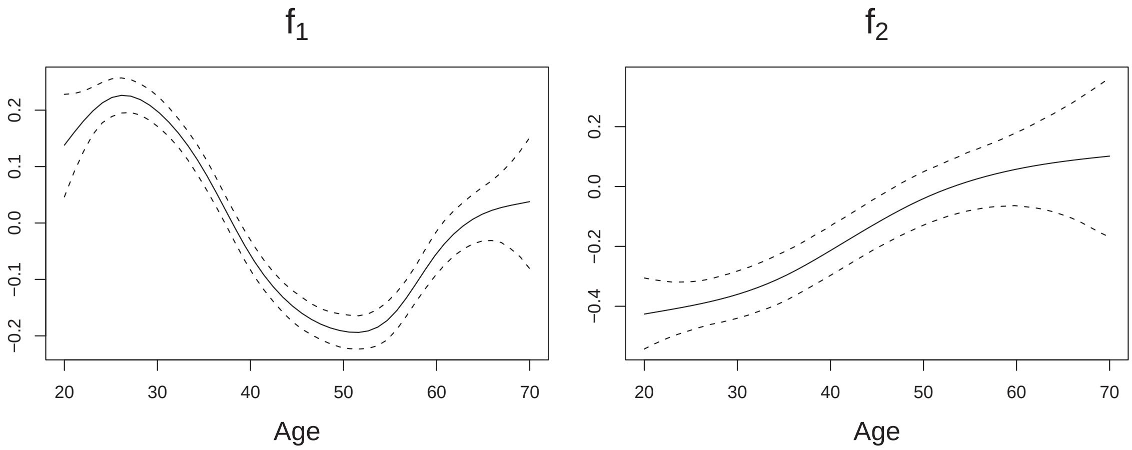


Figure 7.14: *Results of model M5c with* $\eta = Inc + Sex * Inc + Sex * f(Age) + f_{spatial}(Reg)$ – *estimates of the smooth functions* $f_1(Age)$ *and* $f_2(Age)$ *modeled by P-splines priors. The mean values are connected by the solid line, 10%- and 90%-quantiles are connected by the dashed line.*

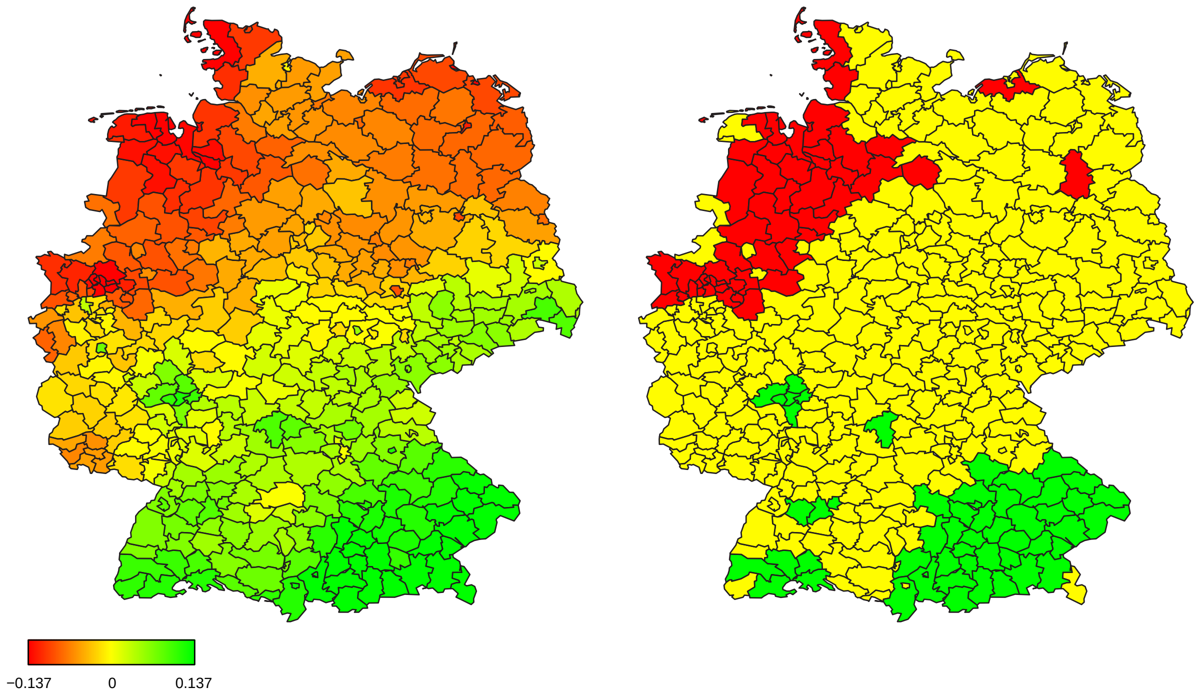


Figure 7.15: *Results of model M5c with* $\eta = Inc + Sex * Inc + Sex * f(Age) + f_{spatial}(Reg)$. ***Left****: Estimates of the spatial covariate Reg. Minimum and maximum values are set to 2.5% and 97.5% quantiles of the observed function estimates, respectively.* ***Right****: Plot of regions with a significant negative effect (red), a significant positive effect (green), or a non-significant effect (yellow).*

7.3.6 Model comparison

Now we want to find out which of the analyzed models can be considered to be the "best" model – the basic question is which covariates should be incorporated into the predictor of the structural equation, and in which form. For model comparison we use the two different versions of the DIC defined in Equations (5.7) and (5.10). The difference between the two versions is that DIC_1 uses the estimated latent scores $\boldsymbol{z}$ whereas the DIC_2 lacks the latent variables and uses the expected value of the structural equation $\boldsymbol{\eta}$ instead. The results of the DIC_1 for all models with one latent variable are summarized in Table 7.16; the obtained values for the DIC_2 can be found in Table 7.17.

Let us start with the discussion of the DIC_1. Model M2b with a pure parametric predictor and interactions, and model M4b with a combination of parametric covariates and smooth functions of metric and spatial covariates have the lowest DIC_1 values – hence those two models might be preferred to the other models. However, the behaviour of the values for $D(\overline{\boldsymbol{\theta}}, \overline{\boldsymbol{z}})$, $\overline{D(\boldsymbol{\theta}, \boldsymbol{z})}$, and p_D is rather unusual compared to generalized regression models. Using the results listed in Table 7.16 and additional DIC_1 analyses based on some simulation models of Section 6.3, the following observations can be made:

- The effective number of parameters p_D is about 5-20% below the number of observations for models without indirect effects, i. e. for classic factor analysis models.
- The values of the deviances $D(\overline{\boldsymbol{\theta}}, \overline{\boldsymbol{z}})$ and $\overline{D(\boldsymbol{\theta}, \boldsymbol{z})}$ are always lowest for models without indirect effects. If indirect covariates (both parametric and nonparametric) are included in the model, the values of both deviances increase. This behaviour contrasts sharply with standard regression models where the inclusion and addition of covariates always reduces the deviance values.
- If indirect covariates are added to a model, the number of effective parameters p_D decreases on a scale which does not depend on the number of parameters in the

Model	Predictor of indirect effects	Prior	DIC	$D(\overline{\boldsymbol{\theta}}, \overline{\boldsymbol{z}})$	$\overline{D(\boldsymbol{\theta}, \boldsymbol{z})}$	p_D
M1	$\eta = 0$	–	201793.11	174790.10	188291.61	13501.51
M2a	$\eta = Sex + Inc + Age$	–	201635.15	175543.42	188589.28	13045.87
M2b	$\eta = Sex + Inc + Age + Sex * Inc + Sex * Age + Inc + Age$	–	201619.25	175552.07	188585.66	13033.59
M3a	$\eta = f(Age)$	RW1	201810.75	174838.21	188324.48	13486.27
M3a	$\eta = f(Age)$	RW2	201782.62	174821.55	188302.09	13480.53
M3a	$\eta = f(Age)$	P3	201800.93	174830.61	188315.77	13485.16
M3b	$\eta = Sex + Inc + f(Age)$	P3	201649.44	175615.58	188632.51	13016.93
M4a	$\eta = f_{spatial}(Reg)$	–	201824.23	174919.97	188372.10	13452.13
M4b	$\eta = Sex + Inc + f(Age) + f_{spatial}(Reg)$	P3	201632.41	175650.60	188641.50	12990.90
M5a	$\eta = Sex * f(Age)$	P3	201697.75	174883.72	188290.73	13407.02
M5b	$\eta = Inc * f(Age)$	P3	201709.42	175560.65	188635.03	13074.39
M5c	$\eta = Inc + Sex * Inc + Sex * f(Age) + f_{spatial}(Reg)$	P3	201633.15	175658.22	188645.68	12987.46

Table 7.16: *Results of DIC_1 of estimated models based on PD1 dataset.*

Model	Predictor of indirect effects	Prior	DIC	$D(\overline{\boldsymbol{\theta}})$	$\overline{D(\boldsymbol{\theta})}$	p_D
M1	$\eta = 0$	–	248301.24	248222.39	248261.82	39.43
M2a	$\eta = Sex + Inc + Age$	–	240672.42	240341.57	240506.99	165.43
M2b	$\eta = Sex + Inc + Age + Sex * Inc +$ $Sex * Age + Inc + Age$	–	241429.44	240176.93	240803.19	626.25
M3a	$\eta = f(Age)$	RW1	248030.16	247763.60	247896.88	133.28
M3a	$\eta = f(Age)$	RW2	247945.64	247734.09	247839.87	105.77
M3a	$\eta = f(Age)$	P3	247946.91	247785.82	247866.36	80.55
M3b	$\eta = Sex + Inc + f(Age)$	P3	240146.98	239891.44	240019.21	127.77
M4a	$\eta = f_{spatial}(Reg)$	–	247328.65	246488.62	246908.64	420.02
M4b	$\eta = Sex + Inc + f(Age) +$ $+ f_{spatial}(Reg)$	P3	239860.19	239146.67	239503.43	356.76
M5a	$\eta = Sex * f(Age)$	P3	246441.99	246243.52	246342.76	99.24
M5b	$\eta = Inc * f(Age)$	P3	240844.42	240561.51	240702.97	141.46
M5c	$\eta = Inc + Sex * Inc + Sex * f(Age) +$ $f_{spatial}(Reg)$	P3	239731.68	238997.05	239364.37	367.31

Table 7.17: *Results of DIC_2 of estimated models based on PD1 dataset.*

predictor of the structural equation. Again this response is different from standard regression models where the number of effective parameters increases when further covariates are added to an analysis. In our LVM however, the inclusion of covariates reduces the already high number of effective parameters of the LVM without indirect effects – the covariates seem to explain the fluctuations of the latent scores and therefore decrease the number of effective parameters.

- The DIC_1 of a model including covariates can be lower – and hence indicate a better fitting model – than in a model with less covariates because the reduction in number of effective parameters is higher than the increase in deviance.
- The range of estimated DIC_1 values is rather narrow – the lowest value is 201,619.25 and the highest one is 201,824.23 which is not a big difference. The reason for this lies in the fact that the factor scores $\boldsymbol{z}$ are estimated in such a way for all models that they explain the actual response values in the best way possible, regardless of the used covariates in the predictor of the structural equation.

Looking at the values of the DIC_2 in Table 7.17, the properties of the DIC_2 are rather different:

- The values of the DIC_2 are clearly higher – thus a much smaller likelihood prevails – compared to the DIC_1. This behaviour can be expected since only the expected values of the latent scores $\boldsymbol{\eta}$ are considered in the DIC_2 instead of the estimated latent scores $\boldsymbol{z}$.
- The basic model M1 without any covariates has a much lower number of effective parameters for the DIC_2 than for the DIC_1 – it seems the number of latent scores $\boldsymbol{z}$ does not have an influence on p_D. When covariates are added to the model, the deviance values of $D(\overline{\boldsymbol{\theta}})$ and $\overline{D(\boldsymbol{\theta})}$ decrease, the number of effective parameters p_D

Model	**Predictor of indirect effects**	**Prior**	**Rank of** DIC_1	**Rank of** DIC_2
M1	$\eta = 0$	–	9	12
M2a	$\eta = Sex + Inc + Age$	–	4	4
M2b	$\eta = Sex + Inc + Age + Sex * Inc +$ $Sex * Age + Inc + Age$	–	1	6
M3a	$\eta = f(Age)$	RW1	11	11
M3a	$\eta = f(Age)$	RW2	8	9
M3a	$\eta = f(Age)$	P3	10	10
M3b	$\eta = Sex + Inc + f(Age)$	P3	5	3
M4a	$\eta = f_{spatial}(Reg)$	–	12	8
M4b	$\eta = Sex + Inc + f(Age) +$ $+ f_{spatial}(Reg)$	P3	2	2
M5a	$\eta = Sex * f(Age)$	P3	6	7
M5b	$\eta = Inc * f(Age)$	P3	7	5
M5c	$\eta = Inc + Sex * Inc + Sex * f(Age) +$ $f_{spatial}(Reg)$	P3	3	1

Table 7.18: *Comparison of the order of the best fitting models recommended by* DIC_1 *and* DIC_2.

increases; this behaviour corresponds with the DIC properties in standard regression settings.

- How the effective number of parameters is obtained is still a bit a mystery. As covariates are added to the model, p_D typically increases by a certain amount – however, the value of the increase cannot be wholly explained by the added number of parameters. Again, the effective number of parameters seem to be related to the amount of latent variable variance which can be attributed to the indirect covariates.
- The basic model without covariates almost always has the highest DIC_2 values – this is obvious because there are no covariates present which might explain the fluctuations of the expected values of the latent scores.
- The range of DIC_2 values is higher than in the case of DIC_1 with a spread from $239,731.68$ to $248,301.24$.

In order to compare the results of the two different DIC versions, the ranks of the "best" fitting models from low to high DIC values are depicted in Table 7.18. We recognize that both DIC versions favour different models; e.g. DIC_1 considers model M2b to be the best model – DIC_2, however, estimates model M2b to be the sixth best model. For the DIC_1 some results seem to be unlikely, for example the DIC_1 considers model M4a with a sole spatial effect as the worst fitting model although the spatial effect explains the differences in latent scores rather convincingly (see Section 7.3.4). Although the end results of the

DIC_1 seem to be somewhat reasonable, and therefore might be used for model comparison, the way how the DIC value is obtained by adding up deviance and the effective number of parameters is not fully understood and raises many questions. Furthermore, some results, e. g. for model M4a, seem to be implausible for the DIC_1. Due to these properties and the fact that we are interested in the amount of variation of the latent scores (and thus of the indicators) that can be explained by the covariates, we would rather recommend DIC_2 to be used for model comparison instead of DIC_1. Nevertheless, since the applicability of the DIC is not totally clear in LVM, and our results only allow preliminary conclusions, we recommend not to rely completely on the DIC for model selection, and see model comparison as one main area of future investigation regarding the LVM.

7.4 Model estimations with two latent variables

In this section we want to demonstrate that the LVM introduced in this thesis is able to estimate models with more than one latent variable. For that purpose all ten indicators described in Tables 7.1 and 7.2 are used, and two latent variables are included in the model. Due to identification restrictions explained in Section 3.1.2, the value of the second latent variable of indicator *Health* is fixed to zero. Accordingly the measurement model yields

$$\begin{pmatrix} y_{i1}^* \\ y_{i2}^* \\ y_{i3}^* \\ y_{i4}^* \\ y_{i5}^* \\ y_{i6}^* \\ y_{i7}^* \\ y_{i8}^* \\ y_{i9}^* \\ y_{i,10}^* \end{pmatrix} = \begin{pmatrix} \lambda_{10} \\ \lambda_{20} \\ \lambda_{30} \\ \lambda_{40} \\ \lambda_{50} \\ \lambda_{60} \\ \lambda_{70} \\ \lambda_{80} \\ \lambda_{90} \\ \lambda_{10,0} \end{pmatrix} + \begin{pmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \\ \lambda_{31} & \lambda_{32} \\ \lambda_{41} & \lambda_{42} \\ \lambda_{51} & 0 \\ \lambda_{61} & \lambda_{62} \\ \lambda_{71} & \lambda_{72} \\ \lambda_{81} & \lambda_{82} \\ \lambda_{91} & \lambda_{92} \\ \lambda_{10,1} & \lambda_{10,2} \end{pmatrix} \cdot \begin{pmatrix} z_{i1} \\ z_{i2} \end{pmatrix} + \begin{pmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \varepsilon_{i3} \\ \varepsilon_{i4} \\ \varepsilon_{i5} \\ \varepsilon_{i6} \\ \varepsilon_{i7} \\ \varepsilon_{i8} \\ \varepsilon_{i9} \\ \varepsilon_{i10} \end{pmatrix} \quad \text{, with} \quad \varepsilon_{ij} \sim N(0,1)\,.$$

The structural form of the predictor for the analysis stems from model M4b in Table 7.5, hence

$$\boldsymbol{\eta}_i = \begin{pmatrix} \eta_{i1} \\ \eta_{i2} \end{pmatrix} = \begin{pmatrix} f_{11}(Age_i) \\ f_{21}(Age_i) \end{pmatrix} + \begin{pmatrix} f_{12}(Reg_i) \\ f_{22}(Reg_i) \end{pmatrix} + \begin{pmatrix} \gamma_{11} & \gamma_{12} & \gamma_{13} & \gamma_{14} \\ \gamma_{21} & \gamma_{22} & \gamma_{23} & \gamma_{24} \end{pmatrix} \cdot \begin{pmatrix} Sex2_i \\ Inc2_i \\ Inc3_i \\ Inc4_i \end{pmatrix}.$$

As in the last section the prior distributions of the factor loadings are chosen to be medium, the hyperpriors of the smoothing parameters for all smooth functions are $a = 0.001$, $b = 0.001$.

The results of this analysis consist of parameter estimates of the factor loadings, parametric regression coefficients and smoothing parameters which can be found in Table 7.19;

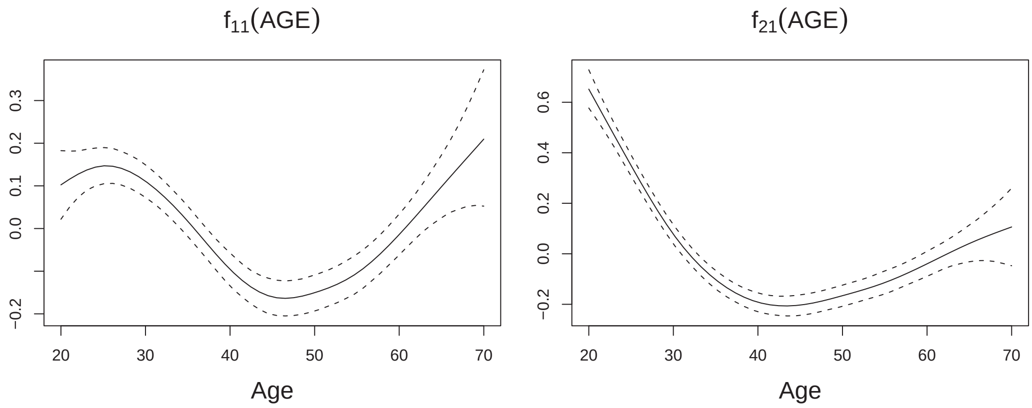


Figure 7.16: *Estimates of the nonparametric smooth functions for the covariate age for the first latent variable (left) and the second latent variable (right). The mean values are connected by the solid line, 10%- and 90%-quantiles are connected by the dashed line.*

function estimates for covariate *Age* are drawn in Figure 7.16, and function estimates of the spatial covariate *Reg* are plotted in Figure 7.17. Cutpoints and intercepts can be found in Table D.14 in the appendix.

The factor loadings' estimates clearly show that the first latent variable loads onto the first five indicators, and indicators 6 to 10 explain the second latent variable. This had to be expected because the two different sets of indicators are supposed to measure two different latent constructs. Both factor loadings and parametric regression coefficients of the first latent factor are very similar to the estimates of the single latent factor model given in Table 7.12. Regarding the factor loadings of the second latent variable, the indicator *Career* is noteworthy due to its very high factor loading of -1.579. The influence of the covariates *Sex* and *Inc* is not as pronounced for the second latent variable compared as for the first latent variable.

The function of covariate *Age* for the first latent variable resembles the function estimate of the model with one latent variable drawn in Figure 7.10. The functional form of the influence of covariate *Age* on the second latent variable looks fundamentally different. Apparently young people want to achieve more, want to have a career and advance in society. This effect rapidly drops with rising age and has its minimum at around the age of 42, probably because the family plays a more important role and children have to be raised in that age. Then the influence grows slowly with rising age.

The estimates of the spatial effect for the first latent variable (upper row of Figure 7.17) are similar but slightly different than the estimates for the one factor model in Figure 7.11. The reason for the slightly different parameter estimates and their significance lies in the fact that the number of observations in the two factor model is only about a third of the one factor model, and some areas have only a low number of observations, particularly in eastern Germany. The spatial effect of the second latent variable (bottom row of Figure

Parameter		Mean	Std. dev.	10% quantile	Mode	90% quantile
Factor loadings of first latent variable						
1. *System*	λ_{11}	0.856	0.030	0.818	0.856	0.895
2. *Initiative*	λ_{21}	-0.929	0.026	-0.962	-0.929	-0.896
3. *Retirement*	λ_{31}	1.278	0.036	1.234	1.278	1.323
4. *Emergency*	λ_{41}	0.692	0.019	0.668	0.693	0.717
5. *Health*	λ_{51}	0.907	0.024	0.877	0.907	0.938
6. *Perform*	λ_{61}	-0.179	0.019	-0.203	-0.178	-0.154
7. *Society*	λ_{71}	0.069	0.021	0.041	0.069	0.097
8. *Reputation*	λ_{81}	0.145	0.020	0.119	0.145	0.170
9. *Salary*	λ_{91}	0.068	0.019	0.044	0.068	0.093
10. *Career*	$\lambda_{10,1}$	0.051	0.036	0.005	0.050	0.097
Factor loadings of second latent variable						
1. *System*	λ_{12}	0.265	0.025	0.233	0.265	0.297
2. *Initiative*	λ_{22}	-0.265	0.022	-0.294	-0.265	-0.237
3. *Retirement*	λ_{32}	-0.025	0.026	-0.058	-0.025	0.009
4. *Emergency*	λ_{42}	-0.028	0.019	-0.052	-0.028	-0.003
5. *Health*	λ_{52}	0.000	0.000	0.000	0.000	0.000
6. *Perform*	λ_{62}	-0.629	0.019	-0.653	-0.629	-0.605
7. *Society*	λ_{72}	-0.789	0.021	-0.816	-0.789	-0.763
8. *Reputation*	λ_{82}	-0.679	0.018	-0.703	-0.679	-0.656
9. *Salary*	λ_{92}	-0.645	0.019	-0.669	-0.645	-0.621
10. *Career*	$\lambda_{10,2}$	-1.579	0.058	-1.655	-1.575	-1.507
Parametric indirect effects of first latent variable						
$Sex2$	γ_{11}	-0.373	0.033	-0.414	-0.373	-0.331
$Inc2$	γ_{12}	0.150	0.042	0.096	0.149	0.204
$Inc3$	γ_{13}	0.473	0.045	0.417	0.473	0.530
$Inc4$	γ_{14}	0.933	0.052	0.866	0.933	1.001
Parametric indirect effects of second latent variable						
$Sex2$	γ_{21}	-0.133	0.033	-0.176	-0.133	-0.090
$Inc2$	γ_{22}	0.155	0.041	0.103	0.156	0.206
$Inc3$	γ_{23}	0.299	0.044	0.242	0.299	0.355
$Inc4$	γ_{24}	0.587	0.054	0.518	0.588	0.655
Smoothing parameters of both latent variables						
Smoothing par.	κ_{11}	0.015	0.019	0.003	0.010	0.030
Smoothing par.	κ_{12}	0.008	0.005	0.003	0.008	0.015
Smoothing par.	κ_{21}	0.015	0.015	0.005	0.011	0.029
Smoothing par.	κ_{22}	0.003	0.002	0.001	0.002	0.005

Table 7.19: *Estimates of factor loadings, parametric indirect effects and smoothing parameters of the LVM with two latent variables.*

7.17) is very weak – therefore not a single region exhibits a significant effect on the latent variable. Obviously the attitude to pursue a career and to achieve something in the society and in the job is rather universal all over the country.

This section demonstrates that our model is able to estimate the influence of covariates on more than one latent variable. Accordingly the model can also be used in exploratory settings where typically a high number of indicators and latent variables are involved. Nevertheless we see the most useful applications in LVM involving one single latent variable and several covariates of various types that influence that latent variable, as elaborated in Section 3.1.5.

7.5 Approximation of smooth functions

In this section, a comparison is made between the results of two semiparametric models, and two parametric models which shall approximate the semiparametric models. For this purpose, we take the two semiparametric models M3b und M4a (see Table 7.5). The nonparametric function of the metric covariate *Age* of model M3b is approximated by a cubic, parametric predictor which yields

$$\begin{aligned} \eta_i &= \gamma_1 \cdot Sex2_i + \gamma_2 \cdot Inc2_i + \gamma_3 \cdot Inc3_i + \\ &\quad + \gamma_4 \cdot Inc4_i + \gamma_5 \cdot Age_i + \gamma_6 \cdot Age_i^2 + \gamma_7 \cdot Age_i^3 \,. \end{aligned}$$

For the effect of the spatial covariate *Reg* in model M4a, a classification of the regions into three areas is carried out: south, east, and northwest (see Figure 7.18). The reason for this segmentation stems from the hypothesis that people in the comparably wealthy south of Germany might have a different attitude towards state provisions than people in the northwest with a medium prosperity, and people in the east with a traditionally weaker economy. Consequently each observation is assigned one of those three geographical locations, and this new categorized variable is termed *RegPar*. Since the nonparametric spatial effect of M4a is centered around zero, an effect coding of *RegPar* has to be employed which delivers the difference effect of a certain region to the mean of all functional values. The predictor is defined as

$$\eta_i = \gamma_1 \cdot RegParSouth_i + \gamma_2 \cdot RegParEast_i \,.$$

Let us start with the discussion of model M3b with the metric variable *Age*. Estimates of the factor loadings and the parametric effects *Sex* and *Inc* given in Table 7.20 are virtually identical to the estimates of the nonparametric model estimates in Table 7.10. The estimated function of *Age* based on the parameters $\{\gamma_5, \gamma_6, \gamma_7\}$ of the cubic polynom of *Age* is plotted in Figure 7.19. The same figure shows a comparison of the cubic parametric estimate with the original nonparametric P-splines estimate. From the ages 20 to 60, the

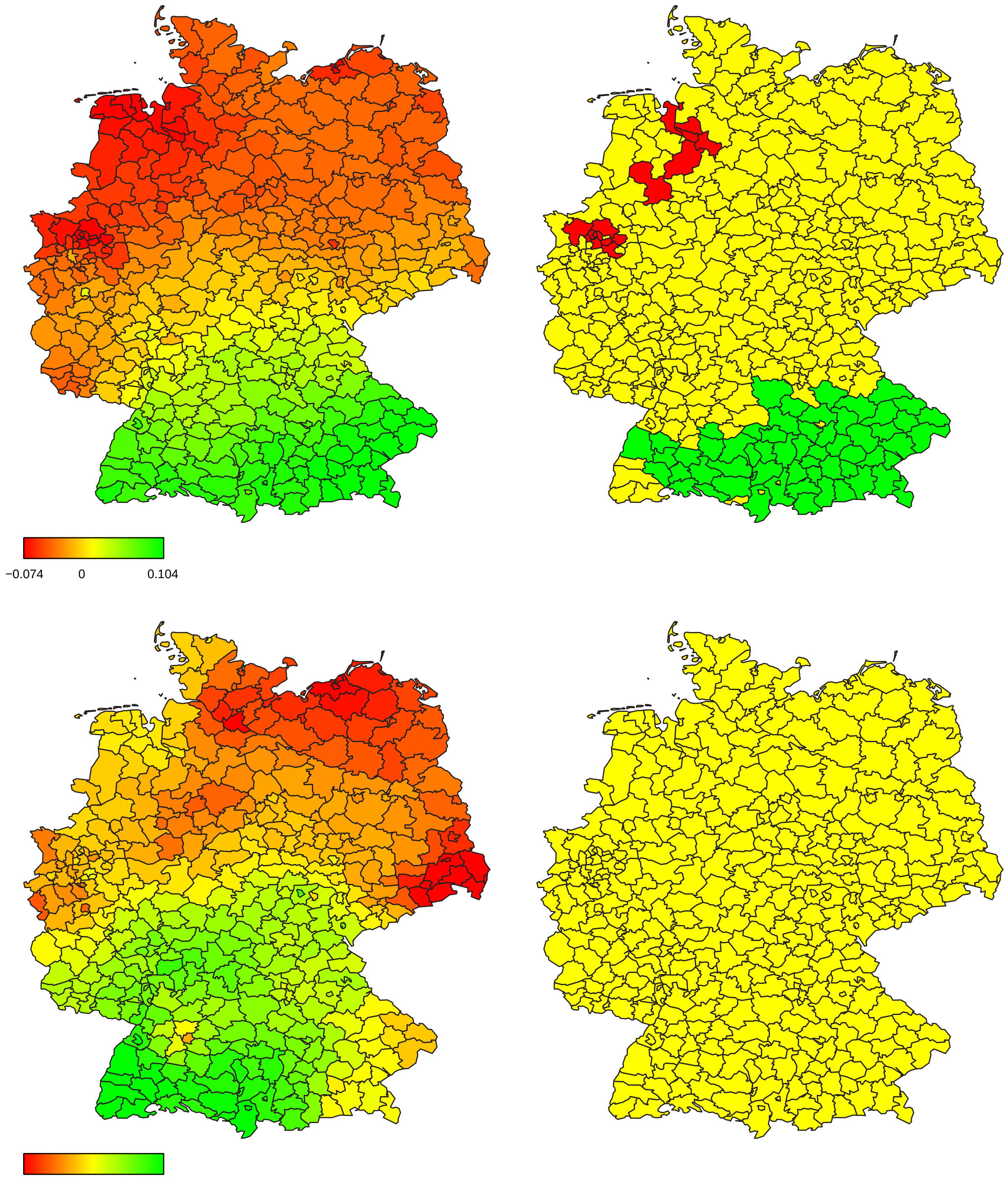


Figure 7.17: ***Left***: *Estimates of nonparametric effect of spatial covariate Reg for the first (top) and second (bottom) latent variable. Minimum and maximum values are set to 2.5% and 97.5% quantiles of observed function estimates, respectively.* ***Right***: *Plot of regions with a significant negative effect (red), significant positive effect (green), and non-significant effect (yellow) for the first (top) and second (bottom) latent variable.*

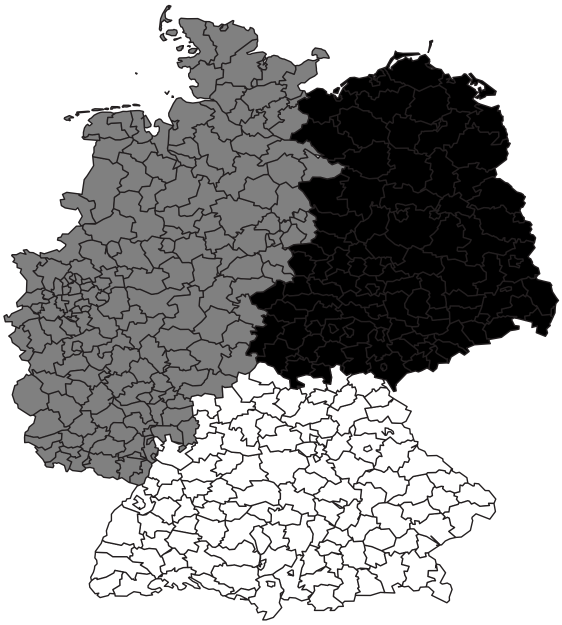


Figure 7.18: *Categorization of the regional landscape into three categories "Northwest" (grey), "South" (white), and "East" (black) of the new variable RegPar.*

cubic parametric estimate demonstrates a very close resemblance to the functional form of the nonparametric estimate; for ages higher than 60 however, the cubic estimate rises much to high up to the age of 70. Furthermore, the 80% credible region is particularly wide. The reason for this broad credible interval lies in the fact that the three cubic parameters $\{\gamma_5, \gamma_6, \gamma_7\}$ are not adjusted in each MCMC iteration in such a way that the resulting function estimate is centered around zero. Due to this lack of centering, a very broad and not meaningful credible region is obtained. Hence there are two advantages of the nonparametric approach; firstly, the nonparametric estimate follows the actual function in a better way than the parameterized functions; secondly, the centering of the estimated function estimate in each MCMC iteration leads to a narrow credible interval.

Results of the parametric estimates of the approximated model of M4a with three spatial regions are summarized in Table 7.21. The factor loadings are very close to the results of the parametric spatial model in Table 7.11. The two effects for the regions "South" and "East" are noticable and significant. Since effect coding is employed where the regression coefficients add up to zero, the parameter for the reference category "Northwest" is obtained by $\gamma_{Northwest} = -(\gamma_1 + \gamma_2) = -0.019$. The drawback of the categorized regions is that you need a hypothesis which regions might have a similar effect on the latent construct, and thus have to be merged before the conduct of the analysis; it is not reasonable to estimate a parametric analysis including 400 regions due to very high confidence regions of the individual parameters. Furthermore, the variation of the latent scores within a merged region is not visible in a parametric model.

Parameter		Mean	Std. dev.	10% quantile	Mode	90% quantile
Factor loadings						
1. *System*	λ_{11}	0.804	0.017	0.782	0.804	0.826
2. *Initiative*	λ_{21}	-0.851	0.014	-0.870	-0.851	-0.833
3. *Retirement*	λ_{31}	1.311	0.025	1.279	1.311	1.342
4. *Emergency*	λ_{41}	0.646	0.012	0.631	0.646	0.661
5. *Health*	λ_{51}	0.911	0.015	0.892	0.911	0.930
Parametric indirect effects						
*Sex*2	γ_1	-0.354	0.022	-0.382	-0.354	-0.327
*Inc*2	γ_2	0.283	0.030	0.244	0.283	0.322
*Inc*3	γ_3	0.622	0.030	0.583	0.622	0.661
*Inc*4	γ_4	1.121	0.034	1.076	1.120	1.165
Age	γ_5	0.102	0.027	0.068	0.102	0.137
Age^2	γ_6	$-3.14 \cdot 10^{-3}$	$6.45 \cdot 10^{-4}$	$-3.97 \cdot 10^{-3}$	$-3.13 \cdot 10^{-3}$	$-2.32 \cdot 10^{-3}$
Age^3	γ_7	$2.73 \cdot 10^{-5}$	$4.94 \cdot 10^{-6}$	$2.10 \cdot 10^{-5}$	$2.72 \cdot 10^{-5}$	$3.37 \cdot 10^{-5}$

Table 7.20: *Results of the approximated model of M3b – estimates of factor loadings and parametric effects for the cubic factor Age.*

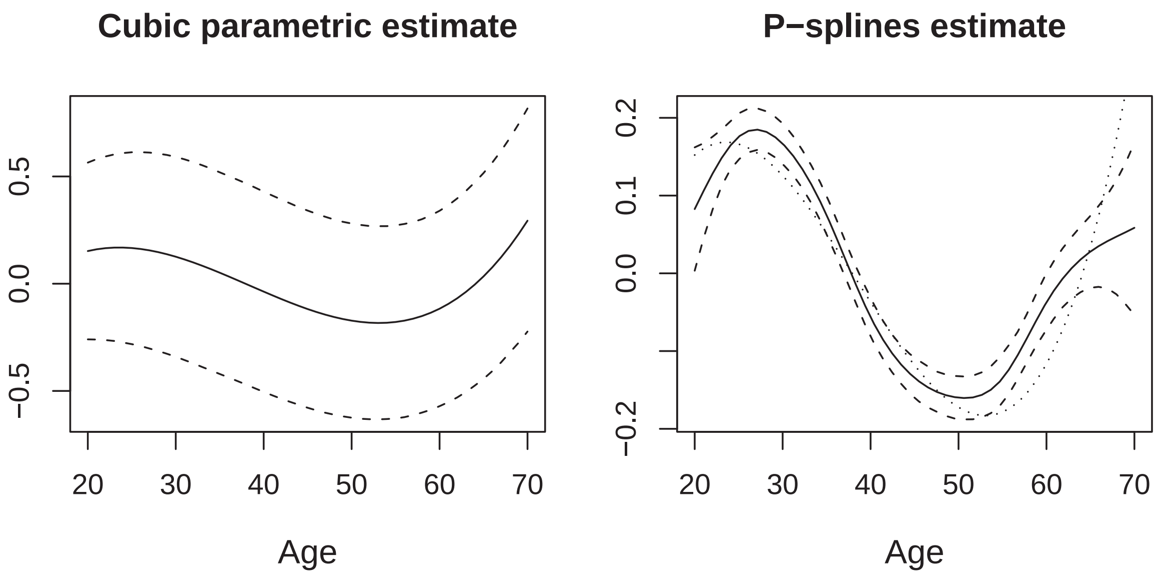


Figure 7.19: *Results of the approximated model of M3b.* ***Left****: Function of Age based on the cubic polynom of Age – the mean value is drawn as a straight line, and 10%- and 90%-quantiles are dashed lines.* ***Right****: The original nonparametric function estimate of Age from Figure 7.8 – the mean cubic parametric function estimate of Age (left picture) is drawn as a dotted line.*

Parameter		Mean	Std. dev.	10% quantile	Mode	90% quantile
Factor loadings						
1. *System*	λ_{11}	0.849	0.018	0.826	0.849	0.873
2. *Initiative*	λ_{21}	-0.893	0.015	-0.913	-0.893	-0.873
3. *Retirement*	λ_{31}	1.457	0.028	1.422	1.456	1.494
4. *Emergency*	λ_{41}	0.690	0.012	0.674	0.690	0.706
5. *Health*	λ_{51}	0.970	0.017	0.949	0.969	0.991
Parametric indirect effects						
RegPar-South	γ_1	0.134	0.013	0.117	0.134	0.150
RegPar-East	γ_2	-0.115	0.015	-0.134	-0.115	-0.097

Table 7.21: *Results of the approximated model of M4a – estimates of factor loadings and parametric effects for the parametric spatial analysis of covariate RegPar.*

7.6 Analysis of factor scores

In this section we discuss some properties of the latent scores z_i. The first part gives an introduction to factor scores and their estimation, and how statements can be made about the probability of one observation having a higher latent score than another. Secondly, we have a look at the distribution of all factor scores in two different models without and with covariates, and how the inclusion of covariates influences the actual latent scores. Finally, we present a method of predicting expected latent scores for new observations where the indicator response is missing but the indirect covariates are available.

7.6.1 Individual factor scores

So far, all presented posterior parameter estimates such as factor loadings and function evaluations provide information about the overall properties of the model. However, applied researchers are very often interested in the actual values of the latent variables for all observations i, also called latent factor scores or just latent scores. For this purpose, the Bayesian approach is especially suited because values of the latent scores $\boldsymbol{z}_i$ are treated as random variables, and therefore they are automatically estimated in the augmented MCMC sampler as a byproduct. Hence the sampling paths of the estimated factor scores are obtained in a Bayesian MCMC estimation process, and afterwards point estimates such as the mean, median or quantiles can be calculated, or the estimated density function can be plotted for all individual factor scores z_{ir}. This means that both point estimates and measures of uncertainty are obtained for the z_{ir}. Furthermore, statements can be made about the probability of the latent score of a single observation i_1 being higher than that of a second observation i_2.

In this respect, the Bayesian approach in LVM differs fundamentally from the frequentist

Observation	$i_1 = 5626$	$i_2 = 14540$
Indicator response		
1. *System*	2	1
2. *Initiative*	2	3
3. *Retirement*	3	2
4. *Emergency*	3	3
5. *Health*	1	2
Covariate values		
Sex	Male	Male
Inc	4500 – 7500	< 2500
Age	61	21

Table 7.22: *Response and covariate values for two selected observations.*

setting. In a frequentist model, latent scores are not considered to be random variables as is the case for the other unknown parameters, and are therefore not estimated during the model estimation process itself. The latent scores have to be calculated separately after model estimation, and several different procedures exist such as Bartletts' scores or regression scores (Bartholomew, 1987). A drawback of those methods is that typically the calculation of the factor scores does not take the uncertainty of other parameter estimates into account, and only rough statements about the uncertainty of the factor scores can be made; a Bayesian analysis solves these problems by providing the full posterior distribution of the factor scores. Finally, no frequentist models exist which could incorporate a semiparametric predictor in the structural equation which is included in our model.

Let us consider an example taken from the model M2 of Section 7.3.2 with a parametric predictor including the covariates *Sex*, *Inc*, and *Age*. This model was estimated again with 5,000 saved MCMC iterations after a burnin of 2,000 iterations; now the factor scores of every 10th iteration were stored, so that 500 sampled factor scores values are obtained for all observations. We pick out two observations $i_1 = 5,626$ and $i_2 = 14,540$ whose indicator response and covariates are shown in Table 7.22. The values of the first three manifest variables indicate that observation i_1 should have a higher latent score than i_2. This prediction is confirmed by the actual values of the factor scores which are plotted in Figure 7.20. Observation i_1 has a mean factor score of $\bar{z}_{5626} = 0.347$ which is higher than the latent score $\bar{z}_{14540} = -0.157$ of the second observation. Furthermore, we recognize that the latent scores of both observations are so close that their density overlaps which means that there is a certain probability of the true latent score z_{5626} being lower than z_{14540}. Unfortunately it is not possible to calculate the probability of $z_{5626} > z_{14540}$ analytically, because the factor scores samples of both observations in each iteration are mutually correlated and not independent. In order to obtain this probability, we determine the fraction of iterations where $z^{(m)}_{5626} > z^{(m)}_{14540}$. As a result, the probability of the latent score z_{5626} being higher than z_{14540} results in 0.796; hence with a probability of about 80%, observation $i_1 = 5626$ is more inclined to take care of his own provisions than observation i_2.

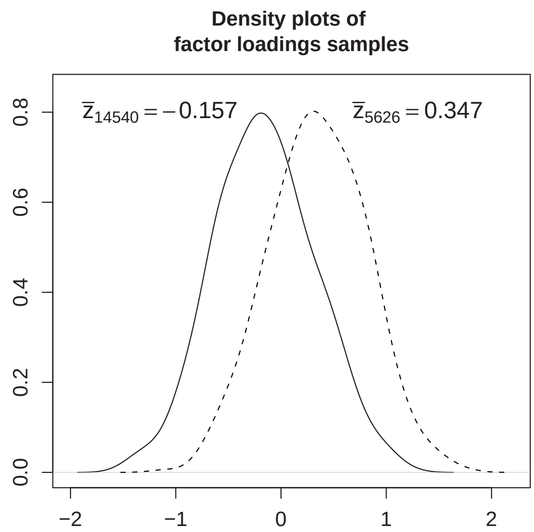


Figure 7.20: *Density plots of the estimated factor scores of observation* $i_1 = 5626$ *and* $i_2 = 14540$.

7.6.2 Distribution and location of factor scores

After having discussed the values of individual factor scores, we want to take a look at the whole distribution of factor scores. For this purpose, we choose two models; the classic factor analysis model M1 without covariates treated in Section 7.3.1, and the LVM with parametric covariates M2a of Section 7.3.2. For the traditional factor analysis model it is expected that the factor scores are distributed according to the standard normal distribution. For the model M2a including covariates, we would expect a distribution that looks similar to a normal distribution but with a wider standard devation due to the negative and positive effects of the predictor. Histograms of the latent scores for both models, including a true standard normal distribution, are drawn in Figure 7.21.

First we recognize that the distribution of the factor scores of model M1 is not entirely standard normal. The density around zero is higher than for a standard normal density, and the densities at the margins of the distribution are too low, especially at the negative margin where the density rapidly drops to zero for latent scores smaller than -2. The reason for this behaviour lies in the purely ordinal structure of the response. For example, the response vector belonging to the latent scores near -2 is $\boldsymbol{y}_i = (1, 6, 1, 1, 1)$ which is the response leading to the most negative latent score possible. Since 678 people have chosen this response pattern (this amounts to approximately 4% of the total sample), there is a sharp drop at the negative margin of the distribution. Furthermore, the ordinal structure causes a concentration of responses around zero which is indicated by the fact that the density around zero is substantially higher than the density of the standard normal distribution. Hence the classic factor analysis model for ordinal response does not lead

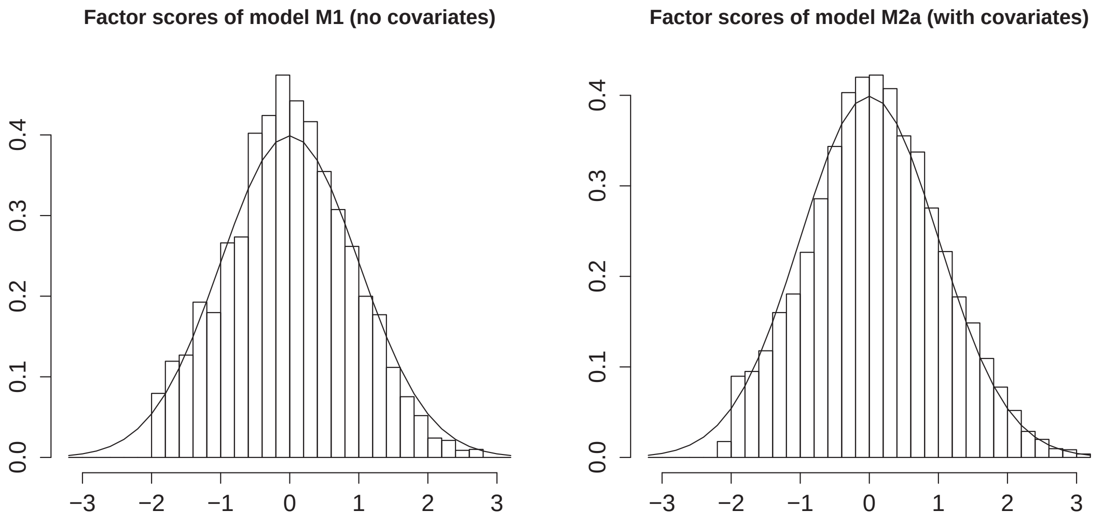


Figure 7.21: *Histograms of the factor scores' densities of model M1 (no covariates) and M2a (with covariates).*

to a strictly standard normal distribution of the factor scores as would be the case for normally distributed metric indicators. As expected, the density histogram of the model M2a with covariates widens up compared to the model M1 – the peak at the centre flattens out, and the latent scores reach out further along the right margin of the latent scale, so that the distribution resembles more a standard normal distribution now. The reason for that lies in the fact that indirect covariates influence the latent scores in the predictor. The distribution would look differently if we had chosen different reference categories; for example, if *Inc* category 4 had been set to the reference category, the effect of the *Inc* categories $\{1, 2, 3\}$ would have been negative so that the whole distribution would have been shifted to the left.

Another interesting question is how models with and without indirect effects cause different values of the latent scores. In order to examine the effect of covariates on the latent scores, we have built four groups with respect to the four income categories, and calculated the mean factor scores within those groups for both models M1 and M2a (see Table 7.23). The differences between the mean factor scores per income category are stretched out in model M2a compared to model M1; hence the effect of the covariate income is to separate the latent scores of the individual income categories in a more pronounced way. To study the effect of the covariates on the latent scores with respect to observations with the same indicator response, we have chosen the above mentioned extreme response vector $\boldsymbol{y}_i = (1, 6, 1, 1, 1)'$, and provide the average factor score per income category in Table 7.23. The mean values of the latent scores in model M1 are more or less the same for all income categories; this must be the case because the variable *Inc* is not included in the model at all. In model M2a however, the mean latent scores are different because the covariate *Inc*

	Average factor scores for			
	all $\boldsymbol{y}_i$		$\boldsymbol{y}_i = (\mathbf{1}, \mathbf{6}, \mathbf{1}, \mathbf{1}, \mathbf{1})'$	
Income category	**Model M1**	**Model M2a**	**Model M1**	**Model M2a**
1	-0.377	-0.428	-1.794	-1.950
2	-0.172	-0.141	-1.792	-1.836
3	0.067	0.185	-1.794	-1.698
4	0.424	0.665	-1.781	-1.493

Table 7.23: *Average factor scores per income category for all observations (column 2 and 3) and observations with response vector* $\boldsymbol{y}_i = (1, 6, 1, 1, 1)'$*, per model M1 (without indirect covariates) and M2a (including indirect covariates).*

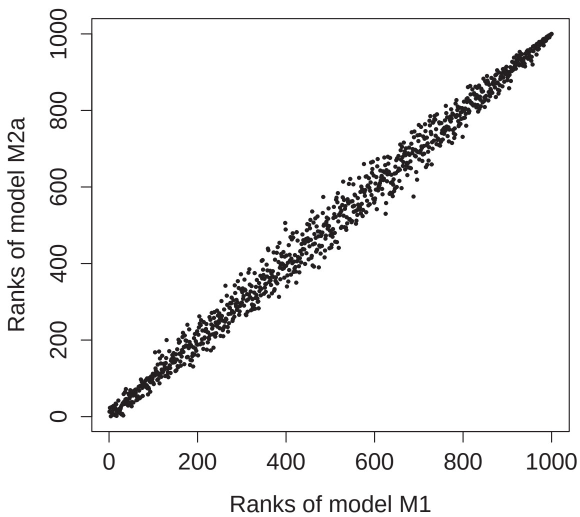


Figure 7.22: *Ranks of the latent scores for model M1 (without covariates) plotted along the x-axis, for model M2a (with covariates) plotted along the y-axis.*

modifies the latent scores – a higher income category leads to a higher average latent score although the response vector remains constant.

The last findings indicate that the inclusion of covariates modifying the latent scores changes the order of observations ranked along the latent scale. To show this effect, 1000 observations are randomly chosen from the whole sample, and their ranking along the latent scale is plotted in Figure 7.22 for both models. We recognize that the latent scores vary in such a way that the ranks of the latent scores of model M1 are not preserved in model M2a with covariates – the overall order remains intact but there are many observations that move some steps up or down in the ranks of latent scores, especially in the middle ranks. The maximum rank difference observed is approximately 120 steps which corresponds to a shift of about 12% in ranks.

7.6.3 Prediction of factor scores without indicator response

Let us consider the situation when a new additional observation belonging to the population is obtained, but only the covariates of this observation are available whereas the indicator response is missing. For example, it might be that we are interested in the latent score of a specific person who has not participated in the social survey answering the full questionnaire, but that person's sociodemographic variables are known.

If only a classic factor analysis without the inclusion of covariates has been estimated, no reasonable statement about the expected latent score of that new person can be made. The expected latent score is always 0, no matter whether the person is of low or high age, has a low or high income etc. Of course this result is of little value.

However, if a LVM including covariates has been estimated, more sophisticated statements about the expected latent score of a new observation can be made. An additional observation i has an expected latent score $E(z_i)$ which equals the value of its predictor value, i. e.

$$E(z_i) = \eta_{i_1} = \boldsymbol{f}_1(x_{i1}) + \ldots + \boldsymbol{f}_g(x_{ig}) + \boldsymbol{u}_i\boldsymbol{\gamma}\,.$$

Let us consider another example of model M2a from Section 7.3.2. We invent two fictitious persons $i = 1$ and $i = 2$, whose covariates and resulting expected latent scores are shown in Table 7.24. Nothing definite can be said about the real latent score of those two observations because the indicator response is not available.

The model assumes that an individual standard normally distributed error term ξ_i is added to the value of the predictor η_i. This allows us to make statements about the probability of the real latent score z_1 of observation $i = 1$ with expected value of $E(z_1) = \eta_1$ being higher than the real latent score z_2 of the second observation $i = 2$ with $E(z_2) = \eta_2$. The two not observed latent scores z_1 and z_2 can then be considered as random variables that are distributed according to $z_1 \sim N(\eta_1, 1)$ and $z_2 \sim N(\eta_2, 1)$; it follows that the difference $z_1 - z_2$ is distributed $z_1 - z_2 \sim N(\eta_1 - \eta_2, 2)$, using standard statistical calculus. Now we

Obser-vation	**Parametric effects**			**Expected latent score**
	Sex	*Inc*	*Age*	
$i = 1$	Female (-0.359)	2 (0.281)	23 (-0.207)	$E(z_1) = \eta_1 = -0.359 + 0.281 - 0.207 = -0.285$
$i = 2$	Male (0.000)	4 (1.098)	47 (-0.423)	$E(z_2) = \eta_2 = 1.098 - 0.423 = 0.675$

Table 7.24: *Calculation of expected factor scores of two fictitious observations $i = 1$ and $i = 2$. The effect values originate from the results of model M2a in Table 7.7.*

simply have to calculate the probability of the difference $z_1 - z_2$ being positive, i. e.

$$\begin{aligned} P(z_1 - z_2 > 0) &= \int_0^{\infty} \frac{1}{\sqrt{2\pi}\sqrt{2}} \cdot \exp\left(\frac{(x-(\eta_1-\eta_2))^2}{2\sqrt{2}^2}\right) dx \\ &= 1 - \int_{-\infty}^{0} \frac{1}{\sqrt{2\pi}\sqrt{2}} \cdot \exp\left(\frac{(x-(\eta_1-\eta_2))^2}{2\sqrt{2}^2}\right) dx \\ &= 1 - \Phi\left(\frac{\eta_2-\eta_1}{\sqrt{2}}\right), \end{aligned}$$

with the standard normal distribution function Φ. This expression gives the probability of the real latent score z_1 of observation 1 being higher than the latent score z_2 of observation 2. Taking the above example from Table 7.24 with expected latent scores of $\eta_1 = -0.285$ and $\eta_2 = 0.675$, it follows that z_1 is higher than z_2 with a probability of $1 - \Phi((0.675 + 0.285)/\sqrt{2}) = 0.249$.

Chapter 8

Summary and Outlook

In this thesis we introduce a latent variable model (LVM) for mixed metric, binary, and ordinal response including covariates that can influence both the manifest and the latent variables. In essence the model consists of two linked components: firstly, the measurement model which is based on traditional factor analysis models and explains the interrelationship between the manifest variables, latent variables, and direct covariates; secondly, the structural equation which describes the influence of indirect covariates on the latent variables or constructs. The basic structure of the whole model resembles the MIMIC model introduced by Jöreskog and Goldberger (1975). This thesis expands the models existing in the literature by using a semiparametric instead of a pure parametric predictor in the structural equation – this enables various types of covariates such as parametric effects, smooth functions of metric and spatial covariates, and interactions of metric and categorical variables (VCM) to influence the latent variables. This type of predictor is already being heavily and successfully used in generalized regression. Our model allows the applied statistician to conduct much more detailed analyses about the influence of covariates on the latent variables than currently available models.

The estimation of the model is based on a fully Bayesian approach including prior specifications of all involved parameters and functions, and all parameters are estimated by Markov chain Monte Carlo (MCMC) algorithms. Since most variables used in LVM are of ordinal nature (e. g. in social sciences) and the basic MCMC algorithm for ordinal data presented by Albert and Chib (1993) exhibits bad convergence properties, we have tested two other MCMC algorithms in order to improve convergence: the first algorithm based on the work of Cowles (1996) employs a Metropolis-Hastings step for the estimation of cutpoints, hence called the MH-sampler (MHS); the second algorithm inspired by the work of Liu and Sabatti (2000) uses a transformation of a set of parameters in each MCMC iteration, and is called the Generalized Gibbs sampler (GGS). The simulation studies indicate good convergence properties for the MHS, and even better convergence for the GGS which is also computationally much more efficient than the MHS. Unfortunately the GGS does not take full and proper account of the prior distributions of the factor loading parameters, and therefore the probability of obtaining a Heywood case (the fact that a latent variable fully loads onto one manifest variable) is higher than for the MHS. For that reason, we recommend the application of the MHS for a low number of observations whereas the GGS can be used for high numbers of observations. Whether the use of the GGS is appropriate and a Heywood case has not occurred can be easily checked by analyzing the sampling paths of the factor loading parameters. Here we see an opportunity for future research in order to identify a suitable and superior parameter transformation for the GGS which treats the prior distributions of the factor loadings properly, and hence does not increase the probability of Heywood cases for low numbers of observations.

Furthermore, simulation studies have shown that both parametric and nonparametric effects in the structural equation – i. e. smooth functions of metric and spatial covariates – can be estimated very accurately with narrow credible intervals. This is remarkable since the actual response in the structural equation are the latent variables which are not observed in the data but merely estimated in the MCMC algorithm. For interacting metric and categorical covariates however, the function estimates of the non-reference categories exhibit a strong bias for a low number of observations – the bias decreases as the number of observations increases. The good parameter estimation properties allow the applied researcher to deeply analyze the effect of covariates on the latent constructs, and narrow credible intervals of the parameters indicate if the influence of a covariate is significant.

In order to demonstrate the applicability of the model in a real research setting, we have used the data set "Perspektive Deutschland 1" which is a social survey carried out on the internet in 2001. Several analyses using different forms of the predictor in the structural equation have been conducted to measure the influence on one latent construct which is supposed to reflect the attitude of the citizens if they liked to make provisions on their own or preferred the state to be responsible for their social security. For example, a clear effect of the citizens' living area on their attitude is estimated. Another analysis clearly shows that a mistake can be made if a metric covariate is just incorporated into the model as a parametric effect: when the covariate *Age* is included as a smooth function into the model, a clear sinusoidal influence of the age on the latent construct is identified whereas the parametric estimate could not reflect this behaviour. The model results have proved to be very useful to explain the influence of covariates on the latent constructs.

For model comparison we have used two different versions of the DIC (deviance information criterion) as defined in Section 5.3. The first version DIC_1 includes the latent scores whereas they are lacking in the second version DIC_2. The discussion in Section 7.3.6 hints that the DIC_2 might be more useful in evaluating models than DIC_1. However, since the mechanisms of the DIC in LVM are not yet fully understood and the DIC results are not completely comprehensible, further model selection criteria should be identified and investigated in future research.

Having established our LVM, the model allows to be expanded along several different dimensions:

- Inclusion of other types of indicators, e. g. continuous variables which are exponentially, Gamma or Erlang distributed; or discrete variables which are binomial or Poisson distributed. The model presented in this section allows for mixed normally distributed metric, binary and ordinal response but there are certainly applications where the inclusion of other types of indicators might be useful. For example, counted data such as the number of events in a fixed period of time is typically Poisson distributed; the Gamma distribution might be used for continuous, non-negative random variables such as the amount of damage in an insurance context, or survival times. Since those new indicator types are not normally distributed, the Gibbs sampler with straightforward full conditionals cannot be used anymore but a Metropolis-Hastings algorithm has to be employed with a suitable proposal density (e. g. see Gamerman, 1997).

- Analysis of longitudinal data. If the researcher is interested in the fluctuation of the latent construct over time, a covariate indicating the time in the structural equation of the model should be included. The time trend can be easily integrated into our model by estimating a smooth function of the metric covariate time. For seasonal effects some minor modifications have to be integrated into the model, as explained in Fahrmeir and Lang (2001b). This allows the researcher to analyze the influences of a time trend and a seasonal effect on the latent constructs in the same model.

- Use of locally adaptive variances for the function estimation of metric covariates. If the nonparametric function is not smooth or exhibits a highly oscillating behaviour, the assumption of a fixed variance or smoothing parameter over the whole parameter range of the covariate is not appropriate. In such cases it is possible to let the smoothing parameter vary across the parameter range, and hence one smoothing parameter for each observed value of the metric covariate has to be estimated (e. g. see Lang et al., 2002). Of course, unsmooth or highly oscillating functions rarely occur in psychology or social sciences because influences are typically rather gradual but there might be useful applications in natural sciences such as biology or medicine.

Before extending the model by the suggested measures, the usefulness and applicability of this model should be verified by analyzing more real-life data sets covering various scientific fields.

Appendix A

Heywood cases

A Heywood case – which can be attributed to Heywood (1931) – occurs if the communality of one (or even more) indicator results in 1; this means that the latent variable fully loads onto this indicator, and thus the variance of the respective indicator is fully explained by the latent variable, and the individual error term ε_j does not contribute to the indicator's variance. It follows that the Heywood case manifests itself in different ways for continuous and ordinal indicators, respectively. For continuous indicators, the variance σ_j^2 of the individual error terms have to be estimated; in a Heywood case this variance yields zero[1] whereas the factor loading value of the respective indicator is estimated to be equal to the indicator's total variance. For ordinal indicators however, the variance σ_j^2 of the individual error terms is fixed to 1; in order to obtain a communality of 1 the corresponding factor loading has to converge towards plus or minus infinity in a Heywood case. To demonstrate the Heywood case for ordinal variables, Figure A.1 shows the example of an analysis of parameter set C (see Tables 6.1 and 6.2). The sampling paths of the indicators 3–5 look normal up to about iteration 4000 when the Heywood case occurs; the subsequent iterations show that the factor loading parameter of indicator 5 rises significantly, and tends to converge towards infinity; at the same time the factor loadings of the other indicators decrease considerably. For example, taking a factor loading value of only 5, the latent variable already accounts for $5^2/(5^2+1) \sim 96.2\%$ of the total variation of indicator 5. A quick glance at the density plots also reveals a failure in the estimation process.

The fact that the latent variable fully explains the variation of an indicator is not plausible at all; if this were actually true a factor analysis would not be needed in the first place, and the latent variable could be directly measured by the respective indicator. Heywood cases even occur in factor analyses where it is known that all indicators have a communality lower than 1 which made applied researchers worry about the validity of their model. Therefore several authors tried to identify the reasons why some factor analysis models are

[1] In maximum-likelihood analyses even negative variance values can be obtained depending on the estimation procedure.

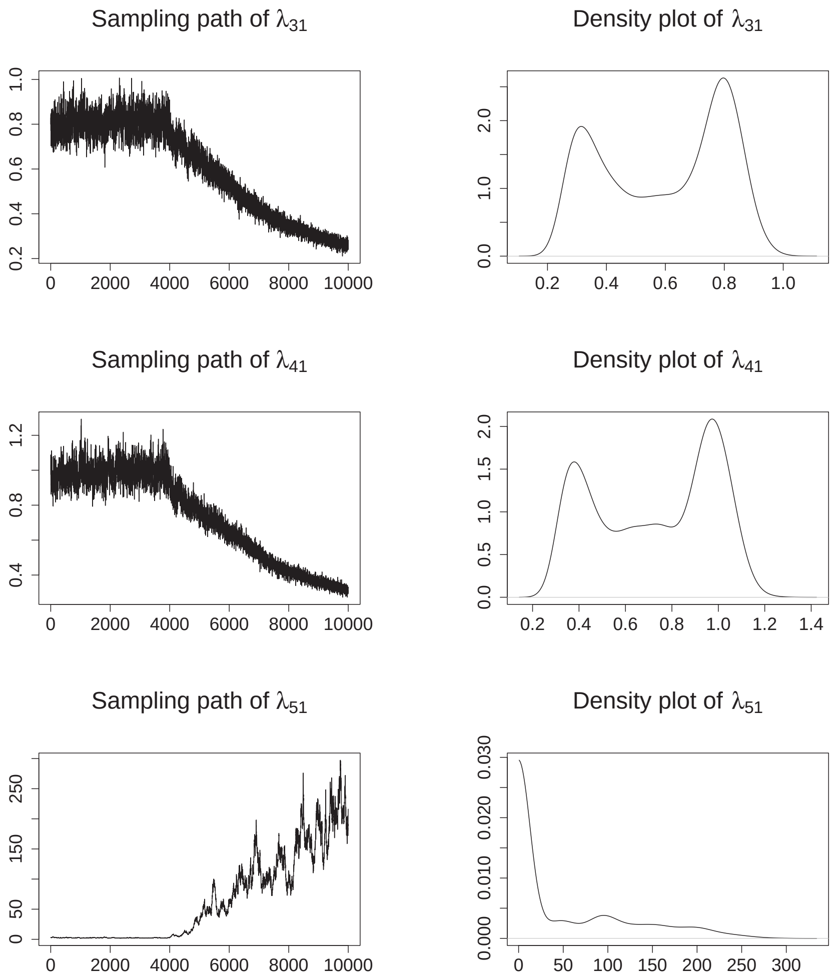


Figure A.1: *Sampling paths (left column) and density plots (right column) of 3 selected factor loadings for the GGS in a Heywood case setting. The simulation is based on parameter set C with* $N_2 = 1000$ *and a standard normal prior on all factor loadings.*

not unambiguously identified. In 1978 van Driel pointed out two main causes of improper solutions: there is no factor analysis model that fits the data, or sampling fluctuations. Since we assume that a proper factor analysis model exists, the almost single cause for Heywood cases in applied research is due to sampling fluctuations; hence the Heywood case represents a small sample phenomenon. Therefore the main remedy to prevent Heywood cases is a high sample size. For a model with all communalities smaller than 1, the probability of a Heywood case tends to zero as the sample size n approaches infinity. Summarizing some main results from van Driel (1978), Anderson (1984), Boomsma (1985), and adding the experiences of our own simulation studies, the following empirical statements about Heywood cases can be made:

- For a fixed number of indicators p, the probability of a Heywood case decreases as the number of observations n increases. This behaviour can be observed in all our simulation studies in Section 6.2.

- For a fixed number of observations n, the probability of a Heywood case decreases as the number of indicators p increases. This is confirmed by the simulation studies with the GGS (see Section 6.2) of parameter set E with ten indicators that show a much lower number of Heywood cases than the simulations of the parameter sets with five indicators each.

- For a fixed number of observations n and indicators p, the probability of a Heywood case decreases significantly for lower factor loadings of the indicator with the highest communality. This effect is very strong. For example, a model with the highest standardized factor loading in the area about 0.8 causes a Heywood case with a much lower probability than a model with the highest standardized factor loading of 0.95. This property can be observed for the simulations of parameter sets C and D which have a very similar structure of ordinal indicators but parameter set C has very different and some very high values of factor loadings, and parameter set D has all standardized factor loadings set to 0.7; therefore simulations with parameter set C cause more Heywood cases than those based on parameter set D.

- For a fixed number of observations n, indicators p and factor loadings $\boldsymbol{\lambda}_j$, the probability of a Heywood case increases if ordinal or even binary indicators are used instead of continuous indicators. Since the underlying variables of ordinal indicators have to be estimated, the sampling error increases which in turn promotes the occurrence of Heywood cases.

- Heywood cases tend to arise if the researcher attempts to extract more latent variables than the data contains. In this thesis this does not cause any problems because we include only one single latent variable in most of our analyses.

- The probability of a Heywood case increases if two highly correlated indicators are incorporated into the model. This problem can often be resolved by dropping one of those indicators in the analysis because one indicator contains all the necessary information.

In the Bayesian setting Heywood cases can be comfortably prevented by including suitable prior information for the factor loadings or error variances as proposed by Martin and McDonald (1981) for the case of continuous indicators. For continuous indicators the prior distribution of the variance σ_j^2 has to be of a form that the probability of σ_j^2 decays to zero at the origin. For ordinal indicators with a fixed error variance however, factor loadings have to be restricted to an area around zero so that they are being prevented to converge towards infinity. This can be easily done by introducing a normal prior with an expectation value of zero, and a fixed variance as set out in Table 5.1. In most cases the use of a standard normal prior works very fine – it is strong enough to prevent the occurrence of Heywood cases but is weak enough not to influence the parameter estimates in a way to cause bias.

Appendix B

Random walks for non-equidistant observations

Here we treat the first- and second-order random walk priors of a metric covariate x with non-equally spaced observations $x_{(t)}$, $t = 1, \ldots, d$, $d \leq n$. The unique observations $x_{(t)}$ are sorted according to $x_{(1)} < \ldots < x_{(t)} < \ldots < x_{(d)}$, and the distances between two successive observations are defined as follows

$$\delta_t = x_{(t)} - x_{(t-1)} \text{ , for } t = 2, \ldots, d \, .$$

As in the main text, let us set $f_t := f(x_{(t)})$ and let $f = (f_1, \ldots, f_t, \ldots, f_d)'$ denote the vector of function evaluations.

First-order random walk

The first-order random walk for non-equally spaced observations is defined as

$$f_t = f_{t-1} + u_t \text{ , with } u_t \sim N(0, \delta_t \tau^2) \, ,$$

for $t = 2, \ldots, d$ and a diffuse prior $f_1 \propto \textit{constant}$. The error variance u_t has been adjusted to $\delta_t \tau^2$ to account for the different distances between unique observations. For bigger differences between successive observations, a higher random error occurs. The entire prior distribution results to

$$\begin{aligned} p(f) = \prod_{t=2}^{d} p(f_t | f_{t-1}, \tau^2) &\propto \exp(-\sum_{t=2}^{d} \frac{1}{2\delta_t \tau^2} (f_t - f_{t-1})^2) \\ &= \exp(-\frac{1}{2\tau^2} f' K f) \, , \end{aligned}$$

with a modified penalty matrix

$$K = \begin{pmatrix} \delta_2^{-1} & -\delta_2^{-1} & & & & & \\ -\delta_2^{-1} & \delta_2^{-1}+\delta_3^{-1} & -\delta_3^{-1} & & & & \\ & -\delta_3^{-1} & \delta_3^{-1}+\delta_4^{-1} & -\delta_4^{-1} & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & -\delta_{d_f-2}^{-1} & \delta_{d_f-2}^{-1}+\delta_{d_f-1}^{-1} & -\delta_{d_f-1}^{-1} & \\ & & & & -\delta_{d_f-1}^{-1} & \delta_{d_f-1}^{-1}+\delta_{d_f}^{-1} & -\delta_{d_f}^{-1} \\ & & & & & -\delta_{d_f}^{-1} & \delta_{d_f}^{-1} \end{pmatrix}.$$

Second-order random walk

The second-order random walk for non-equally spaced observations is defined as

$$f_t = (1 + \frac{\delta_t}{\delta_{t-1}})f_{t-1} - \frac{\delta_t}{\delta_{t-1}}f_{t-2} + u_t \text{ , with } u_t \sim N(0, \rho_t\tau^2) \, ,$$

$t = 3, \ldots, d$ and diffuse priors $f_1 \propto \textit{constant}$ and $f_2 \propto \textit{constant}$. The error variance u_t has been adjusted to $\rho_t\tau^2$ with an appropriate weight ρ_t, $t = 2, \ldots, d$. Common choices for ρ_t include $\rho_t = \delta_t$ as for the first-order random-walk, or $\rho_t = \delta_t/(1 + \delta_t/\delta_{t-1})$, see also Fahrmeir and Lang (2001a). The entire prior distribution is given by

$$\begin{aligned} p(f) = \prod_{t=3}^{d} p(f_t|f_{t-1}, f_{t-2}, \tau^2) &\propto \exp(-\sum_{t=3}^{d} \frac{1}{2\rho_t\tau^2}(f_t - (1 + \frac{\delta_t}{\delta_{t-1}})f_{t-1} + \frac{\delta_t}{\delta_{t-1}}f_{t-2})^2) \\ &= \exp(-\frac{1}{2\tau^2}f'Kf) \, , \end{aligned}$$

with the penalty matrix K as shown in Table B.1.

$$K = \begin{pmatrix}
\left(\frac{\delta_3}{\delta_2}\right)^2 \frac{1}{\rho_3} & -\frac{\delta_3}{\delta_2}\frac{\bar{\delta}_3}{\rho_3} & \frac{\delta_3}{\delta_2}\frac{1}{\rho_3} & & & & \\
-\frac{\delta_3}{\delta_2}\frac{\bar{\delta}_3}{\rho_3} & \frac{\bar{\delta}_3^2}{\rho_3} + \left(\frac{\delta_4}{\delta_3}\right)^2 \frac{1}{\rho_4} & -\frac{\bar{\delta}_3}{\rho_3} - \frac{\delta_4}{\delta_3}\frac{\bar{\delta}_4}{\rho_4} & \frac{\delta_4}{\delta_3}\frac{1}{\rho_4} & & & \\
\frac{\delta_3}{\delta_2}\frac{1}{\rho_3} & -\frac{\bar{\delta}_3}{\rho_3} - \frac{\delta_4}{\delta_3}\frac{\bar{\delta}_4}{\rho_4} & \frac{1}{\rho_3} + \frac{\bar{\delta}_4^2}{\rho_4} + \left(\frac{\delta_5}{\delta_4}\right)^2 \frac{1}{\rho_5} & -\frac{\bar{\delta}_4}{\rho_4} - \frac{\delta_5}{\delta_4}\frac{\bar{\delta}_5}{\rho_5} & \frac{\delta_5}{\delta_4}\frac{1}{\rho_5} & & \\
 & \frac{\delta_4}{\delta_3}\frac{1}{\rho_4} & -\frac{\bar{\delta}_4}{\rho_4} - \frac{\delta_5}{\delta_4}\frac{\bar{\delta}_5}{\rho_5} & \frac{1}{\rho_4} + \frac{\bar{\delta}_5^2}{\rho_5} + \left(\frac{\delta_6}{\delta_5}\right)^2 \frac{1}{\rho_6} & -\frac{\bar{\delta}_5}{\rho_5} - \frac{\delta_6}{\delta_5}\frac{\bar{\delta}_6}{\rho_6} & \frac{\delta_6}{\delta_5}\frac{1}{\rho_6} & \\
 & & \ddots & \ddots & \ddots & \ddots & \ddots \\
 & & \frac{\delta_{d-2}}{\delta_{d-3}}\frac{1}{\rho_{d-2}} & -\frac{\bar{\delta}_{d-2}}{\rho_{d-2}} - \frac{\delta_{d-1}}{\delta_{d-2}}\frac{\bar{\delta}_{d-1}}{\rho_{d-1}} & \frac{1}{\rho_{d-2}} + \frac{\bar{\delta}_{d-1}^2}{\rho_{d-1}} + \left(\frac{\delta_d}{\delta_{d-1}}\right)^2 \frac{1}{\rho_d} & -\frac{\bar{\delta}_{d-1}}{\rho_{d-1}} - \frac{\delta_d}{\delta_{d-1}}\frac{\bar{\delta}_d}{\rho_d} & \frac{\delta_d}{\delta_{d-1}}\frac{1}{\rho_d} \\
 & & & \frac{\delta_{d-1}}{\delta_{d-2}}\frac{1}{\rho_{d-1}} & -\frac{\bar{\delta}_{d-1}}{\rho_{d-1}} - \frac{\delta_d}{\delta_{d-1}}\frac{\bar{\delta}_d}{\rho_d} & \frac{1}{\rho_{d-1}} + \frac{\bar{\delta}_d^2}{\rho_d} & -\frac{\bar{\delta}_d}{\rho_d} \\
 & & & & \frac{\delta_d}{\delta_{d-1}}\frac{1}{\rho_d} & -\frac{\bar{\delta}_d}{\rho_d} & \frac{1}{\rho_d}
\end{pmatrix}$$

Table B.1: *Penalty matrix K for the second-order random walk with non-equally spaced observations. To allow for the illustration of the matrix, the abbreviation $\bar{\delta}_t = 1 + \frac{\delta_t}{\delta_{t-1}}$ is used.*

Appendix C

Simulation results of the LVM

In this chapter, the remaining results of the simulation studies for the LVM excluding indirect effects of Section 6.2 are presented. The MEAN, STD, BIAS, and MSE (see Equations 6.2 and 6.3) of the simulation runs of parameter set B can be found in Tables C.1 and C.2; the results for parameter set C are outlined in Tables C.3 to C.7; Tables C.8 and C.9 show the estimated values of parameter set E. The discussion of the properties of the two sampler types MHS and GGS can be found in Section 6.2.

Simulations of parameter set B – standard sampler/MHS									
N	$N_1 = 300$			$N_2 = 1000$			$N_3 = 5000$		
Priori	**Weak**	**Std.**	**Strong**	**Weak**	**Std.**	**Strong**	**Weak**	**Std.**	**Strong**
	MEAN								
λ_{10}	-1.513	-1.475	-1.386	-1.437	-1.430	-1.402	-1.408	-1.407	-1.402
λ_{20}	-1.028	-1.008	-0.960	-0.992	-0.990	-0.979	-0.983	-0.983	-0.980
λ_{30}	-0.446	-0.440	-0.425	-0.408	-0.407	-0.404	-0.420	-0.420	-0.419
λ_{40}	0.739	0.733	0.706	0.715	0.714	0.706	0.700	0.700	0.699
λ_{50}	1.561	1.511	1.408	1.465	1.454	1.418	1.408	1.407	1.403
λ_{11}	1.047	0.962	0.857	1.034	1.023	0.975	0.987	0.986	0.977
λ_{21}	1.032	0.934	0.873	1.003	0.998	0.966	0.986	0.985	0.978
λ_{31}	1.017	0.958	0.909	0.960	0.956	0.935	0.979	0.979	0.974
λ_{41}	0.979	0.917	0.866	1.013	1.009	0.978	0.982	0.982	0.976
λ_{51}	1.068	0.964	0.851	1.049	1.034	0.976	0.982	0.982	0.973
	STD								
λ_{10}	0.263	0.212	0.151	0.134	0.121	0.104	0.052	0.052	0.051
λ_{20}	0.205	0.165	0.125	0.076	0.075	0.070	0.033	0.033	0.033
λ_{30}	0.121	0.112	0.103	0.062	0.062	0.061	0.022	0.022	0.022
λ_{40}	0.129	0.121	0.102	0.074	0.073	0.070	0.030	0.030	0.030
λ_{50}	0.270	0.212	0.147	0.163	0.147	0.114	0.053	0.053	0.052
λ_{11}	0.416	0.452	0.294	0.164	0.146	0.122	0.063	0.062	0.061
λ_{21}	0.416	0.501	0.316	0.124	0.119	0.109	0.048	0.047	0.046
λ_{31}	0.397	0.451	0.299	0.113	0.110	0.099	0.049	0.049	0.048
λ_{41}	0.369	0.447	0.293	0.121	0.118	0.103	0.056	0.056	0.055
λ_{51}	0.431	0.467	0.289	0.203	0.188	0.146	0.066	0.065	0.063
	BIAS								
λ_{10}	-0.113	-0.075	0.014	-0.037	-0.030	-0.001	-0.007	-0.007	-0.002
λ_{20}	-0.048	-0.028	0.021	-0.012	-0.010	0.001	-0.003	-0.002	0.000
λ_{30}	-0.026	-0.020	-0.004	0.013	0.013	0.016	0.000	0.000	0.001
λ_{40}	0.039	0.033	0.006	0.015	0.014	0.006	0.000	-0.000	-0.001
λ_{50}	0.161	0.110	0.008	0.064	0.054	0.018	0.007	0.007	0.002
λ_{11}	0.067	-0.018	-0.124	0.054	0.042	-0.006	0.007	0.006	-0.003
λ_{21}	0.051	-0.046	-0.107	0.023	0.017	-0.014	0.006	0.005	-0.002
λ_{31}	0.036	-0.022	-0.071	-0.020	-0.024	-0.045	-0.001	-0.001	-0.006
λ_{41}	-0.002	-0.063	-0.115	0.032	0.029	-0.002	0.002	0.001	-0.004
λ_{51}	0.088	-0.016	-0.129	0.069	0.054	-0.004	0.002	0.001	-0.007
	ln(MSE)								
λ_{10}	-2.508	-2.990	-3.778	-3.952	-4.176	-4.535	-5.896	-5.920	-5.957
λ_{20}	-3.130	-3.587	-4.137	-5.134	-5.179	-5.331	-6.800	-6.818	-6.831
λ_{30}	-4.184	-4.357	-4.553	-5.517	-5.521	-5.536	-7.643	-7.638	-7.647
λ_{40}	-4.017	-4.154	-4.568	-5.184	-5.206	-5.328	-7.021	-7.003	-7.044
λ_{50}	-2.325	-2.871	-3.836	-3.490	-3.717	-4.325	-5.856	-5.879	-5.924
λ_{11}	-1.736	-1.597	-2.294	-3.527	-3.770	-4.211	-5.530	-5.556	-5.593
λ_{21}	-1.749	-1.385	-2.205	-4.148	-4.240	-4.428	-6.083	-6.114	-6.176
λ_{31}	-1.849	-1.600	-2.371	-4.341	-4.380	-4.446	-6.047	-6.043	-6.058
λ_{41}	-2.004	-1.602	-2.322	-4.171	-4.223	-4.565	-5.764	-5.772	-5.814
λ_{51}	-1.654	-1.530	-2.310	-3.090	-3.276	-3.859	-5.448	-5.482	-5.521
	Coverage								
λ_{10}	94	96	96	90	93	94	92	93	91
λ_{20}	96	96	94	98	98	98	97	96	96
λ_{30}	96	95	97	94	95	94	100	100	100
λ_{40}	98	98	98	94	95	96	97	94	96
λ_{50}	97	97	96	91	93	93	91	92	92
λ_{11}	92	92	93	94	95	97	92	91	92
λ_{21}	93	93	94	94	96	97	97	98	96
λ_{31}	91	90	96	95	95	96	95	95	95
λ_{41}	93	91	93	93	96	97	94	93	93
λ_{51}	94	94	92	90	93	92	94	95	95

Table C.1: *Estimates of parameter set B simulated by the standard sampler and MHS – MEAN, STD, BIAS, ln(MSE), and coverage of estimated parameters obtained by simulations of 100 different data sets.*

Simulations of parameter set B – GGS									
N	$N_1 = 300$			$N_2 = 1000$			$N_3 = 5000$		
Priori	**Weak**	**Std.**	**Strong**	**Weak**	**Std.**	**Strong**	**Weak**	**Std.**	**Strong**
	MEAN								
λ_{10}	-712.923	-324.339	-1.526	-1.457	-1.450	-1.433	-1.410	-1.409	-1.408
λ_{20}	-318.611	-127.437	-1.028	-0.997	-0.995	-0.993	-0.984	-0.984	-0.983
λ_{30}	-46.316	-19.422	-0.449	-0.407	-0.406	-0.408	-0.420	-0.420	-0.420
λ_{40}	180.806	50.887	0.735	0.718	0.718	0.713	0.701	0.701	0.700
λ_{50}	726.186	478.324	1.559	11.236	7.997	1.462	1.410	1.410	1.407
λ_{11}	72.688	33.957	1.033	1.059	1.049	1.013	0.990	0.989	0.984
λ_{21}	43.498	15.739	1.006	1.011	1.006	0.988	0.987	0.987	0.983
λ_{31}	8.532	4.849	1.000	0.958	0.956	0.946	0.981	0.980	0.977
λ_{41}	27.530	8.466	0.964	1.016	1.012	0.997	0.984	0.984	0.979
λ_{51}	74.925	44.928	1.043	2.948	2.206	1.030	0.986	0.985	0.980
	STD								
λ_{10}	1149.385	604.248	0.288	0.151	0.150	0.128	0.053	0.053	0.052
λ_{20}	674.573	355.234	0.191	0.085	0.086	0.076	0.034	0.034	0.033
λ_{30}	218.826	83.657	0.121	0.068	0.069	0.062	0.022	0.022	0.022
λ_{40}	471.068	173.711	0.131	0.075	0.074	0.074	0.030	0.030	0.030
λ_{50}	1171.325	709.226	0.287	97.637	65.288	0.181	0.054	0.054	0.054
λ_{11}	108.506	56.262	0.289	0.183	0.181	0.152	0.063	0.063	0.062
λ_{21}	86.560	40.097	0.265	0.143	0.140	0.121	0.048	0.047	0.047
λ_{31}	37.360	16.299	0.236	0.128	0.129	0.108	0.050	0.049	0.049
λ_{41}	69.117	25.830	0.226	0.145	0.143	0.115	0.057	0.057	0.056
λ_{51}	110.548	61.951	0.291	18.870	11.529	0.211	0.066	0.067	0.065
	BIAS								
λ_{10}	-711.523	-322.939	-0.125	-0.056	-0.050	-0.033	-0.009	-0.009	-0.007
λ_{20}	-317.631	-126.457	-0.048	-0.017	-0.015	-0.013	-0.003	-0.003	-0.003
λ_{30}	-45.896	-19.002	-0.029	0.013	0.014	0.012	-0.000	-0.000	0.000
λ_{40}	180.106	50.187	0.035	0.018	0.018	0.013	0.001	0.001	-0.000
λ_{50}	724.786	476.924	0.159	9.835	6.597	0.061	0.010	0.010	0.007
λ_{11}	71.708	32.977	0.052	0.078	0.068	0.033	0.009	0.008	0.004
λ_{21}	42.518	14.759	0.025	0.031	0.026	0.008	0.007	0.007	0.003
λ_{31}	7.552	3.869	0.020	-0.022	-0.024	-0.035	0.001	-0.000	-0.004
λ_{41}	26.549	7.486	-0.017	0.035	0.032	0.016	0.004	0.004	-0.001
λ_{51}	73.945	43.948	0.063	1.967	1.226	0.049	0.006	0.005	-0.001
	ln(MSE)								
λ_{10}	14.411	13.051	-2.325	-3.658	-3.697	-4.064	-5.869	-5.870	-5.908
λ_{20}	13.220	11.856	-3.256	-4.907	-4.893	-5.145	-6.784	-6.790	-6.801
λ_{30}	10.810	8.894	-4.171	-5.364	-5.310	-5.520	-7.640	-7.630	-7.637
λ_{40}	12.438	10.386	-3.999	-5.141	-5.147	-5.193	-7.012	-7.010	-7.015
λ_{50}	14.449	13.495	-2.236	9.163	8.358	-3.323	-5.822	-5.824	-5.848
λ_{11}	9.729	8.348	-2.459	-3.234	-3.290	-3.735	-5.520	-5.514	-5.567
λ_{21}	9.130	7.501	-2.659	-3.859	-3.914	-4.237	-6.071	-6.092	-6.108
λ_{31}	7.272	5.628	-2.889	-4.088	-4.070	-4.369	-6.013	-6.028	-6.048
λ_{41}	8.600	6.574	-2.980	-3.811	-3.855	-4.316	-5.744	-5.731	-5.770
λ_{51}	9.774	8.654	-2.433	5.876	4.891	-3.064	-5.436	-5.424	-5.463
	Coverage								
λ_{10}	39	49	95	92	93	92	93	94	93
λ_{20}	41	56	94	97	97	98	95	96	97
λ_{30}	47	57	95	93	92	93	98	99	100
λ_{40}	41	60	99	94	95	95	96	97	96
λ_{50}	32	44	96	92	92	93	91	89	90
λ_{11}	17	35	95	94	95	94	93	92	93
λ_{21}	14	32	97	91	92	95	97	97	96
λ_{31}	13	28	93	95	94	95	96	95	95
λ_{41}	14	31	95	93	91	95	93	93	92
λ_{51}	18	35	96	90	88	92	92	94	95

Table C.2: *Estimates of parameter set B simulated by the GGS – MEAN, STD, BIAS, ln(MSE), and coverage of estimated parameters obtained by simulations of 100 different data sets.*

Simulations of parameter set C – MHS									
N	$N_1 = 300$			$N_2 = 1000$			$N_3 = 5000$		
Priori	**Weak**	**Std.**	**Strong**	**Weak**	**Std.**	**Strong**	**Weak**	**Std.**	**Strong**
	MEAN								
λ_{10}	0.960	0.960	0.962	0.942	0.942	0.943	0.945	0.946	0.946
λ_{20}	1.050	1.053	1.056	1.044	1.046	1.048	1.041	1.042	1.043
λ_{30}	1.123	1.130	1.134	1.136	1.138	1.143	1.123	1.124	1.125
λ_{40}	1.281	1.296	1.305	1.259	1.263	1.274	1.258	1.258	1.261
λ_{50}	2.379	2.099	1.772	2.178	2.096	1.912	2.100	2.088	2.040
λ_{11}	0.325	0.326	0.320	0.313	0.313	0.312	0.312	0.312	0.312
λ_{21}	0.581	0.585	0.574	0.582	0.583	0.582	0.576	0.576	0.577
λ_{31}	0.763	0.771	0.757	0.755	0.758	0.758	0.751	0.752	0.752
λ_{41}	0.990	1.005	0.990	0.992	0.997	1.003	0.981	0.981	0.984
λ_{51}	2.463	2.113	1.650	2.202	2.099	1.849	2.110	2.093	2.028
τ_{12}	0.652	0.652	0.653	0.624	0.624	0.625	0.629	0.629	0.629
τ_{13}	1.292	1.292	1.295	1.256	1.256	1.257	1.259	1.260	1.260
τ_{14}	1.931	1.932	1.935	1.888	1.887	1.890	1.889	1.889	1.889
τ_{22}	0.698	0.701	0.703	0.694	0.695	0.697	0.693	0.694	0.694
τ_{23}	1.398	1.403	1.407	1.393	1.394	1.397	1.389	1.390	1.391
τ_{24}	2.085	2.091	2.098	2.088	2.091	2.095	2.084	2.084	2.085
τ_{32}	0.749	0.754	0.756	0.760	0.760	0.764	0.752	0.752	0.753
τ_{33}	1.509	1.519	1.524	1.517	1.520	1.526	1.501	1.502	1.504
τ_{34}	2.266	2.281	2.289	2.269	2.273	2.282	2.251	2.251	2.254
τ_{42}	0.849	0.859	0.866	0.844	0.847	0.854	0.841	0.841	0.843
τ_{43}	1.701	1.721	1.733	1.689	1.695	1.709	1.682	1.682	1.687
τ_{44}	2.550	2.578	2.597	2.536	2.545	2.566	2.521	2.522	2.528
τ_{52}	1.562	1.383	1.169	1.457	1.402	1.279	1.405	1.397	1.365
τ_{53}	3.170	2.801	2.365	2.902	2.794	2.548	2.801	2.784	2.719
τ_{54}	4.762	4.205	3.548	4.359	4.197	3.827	4.212	4.187	4.090
	STD								
λ_{10}	0.095	0.094	0.095	0.045	0.046	0.046	0.022	0.022	0.023
λ_{20}	0.092	0.094	0.094	0.050	0.050	0.050	0.023	0.023	0.023
λ_{30}	0.114	0.117	0.116	0.063	0.063	0.064	0.026	0.026	0.026
λ_{40}	0.145	0.150	0.149	0.074	0.071	0.072	0.034	0.034	0.033
λ_{50}	0.487	0.320	0.193	0.279	0.229	0.140	0.119	0.115	0.094
λ_{11}	0.071	0.072	0.070	0.041	0.041	0.041	0.017	0.017	0.017
λ_{21}	0.080	0.080	0.078	0.046	0.046	0.045	0.018	0.018	0.018
λ_{31}	0.118	0.119	0.113	0.053	0.052	0.050	0.026	0.025	0.025
λ_{41}	0.119	0.117	0.110	0.070	0.067	0.063	0.027	0.028	0.027
λ_{51}	0.523	0.334	0.150	0.321	0.263	0.139	0.140	0.135	0.106
τ_{12}	0.075	0.074	0.075	0.042	0.042	0.043	0.019	0.019	0.019
τ_{13}	0.103	0.104	0.106	0.050	0.051	0.050	0.022	0.022	0.023
τ_{14}	0.129	0.128	0.130	0.053	0.055	0.054	0.029	0.029	0.029
τ_{22}	0.069	0.070	0.069	0.042	0.042	0.042	0.021	0.021	0.021
τ_{23}	0.107	0.109	0.109	0.056	0.056	0.057	0.026	0.026	0.026
τ_{24}	0.129	0.130	0.131	0.074	0.075	0.075	0.030	0.031	0.030
τ_{32}	0.077	0.079	0.077	0.054	0.055	0.055	0.022	0.022	0.022
τ_{33}	0.125	0.127	0.125	0.075	0.074	0.075	0.028	0.028	0.027
τ_{34}	0.173	0.177	0.175	0.081	0.081	0.081	0.035	0.034	0.033
τ_{42}	0.111	0.113	0.114	0.061	0.059	0.060	0.027	0.027	0.027
τ_{43}	0.156	0.159	0.160	0.076	0.074	0.074	0.037	0.038	0.037
τ_{44}	0.177	0.178	0.182	0.097	0.094	0.091	0.050	0.051	0.049
τ_{52}	0.307	0.235	0.153	0.184	0.152	0.099	0.082	0.080	0.067
τ_{53}	0.592	0.396	0.209	0.356	0.291	0.170	0.159	0.157	0.125
τ_{54}	0.888	0.569	0.277	0.526	0.423	0.227	0.228	0.219	0.175

Table C.3: *Estimates of parameter set C simulated by the MHS – MEAN and STD of estimated parameters obtained by simulations of 100 different data sets.*

Simulations of parameter set C – MHS									
N	**$N_1 = 300$**			**$N_2 = 1000$**			**$N_3 = 5000$**		
Priori	**Weak**	**Std.**	**Strong**	**Weak**	**Std.**	**Strong**	**Weak**	**Std.**	**Strong**
	BIAS								
λ_{10}	0.017	0.017	0.019	-0.001	-0.002	-0.000	0.002	0.002	0.002
λ_{20}	0.010	0.014	0.017	0.005	0.006	0.009	0.002	0.002	0.003
λ_{30}	-0.002	0.005	0.009	0.011	0.013	0.018	-0.002	-0.001	-0.000
λ_{40}	0.021	0.035	0.044	-0.001	0.003	0.014	-0.003	-0.002	0.001
λ_{50}	0.315	0.034	-0.293	0.113	0.032	-0.153	0.035	0.023	-0.025
λ_{11}	0.010	0.011	0.006	-0.002	-0.001	-0.002	-0.002	-0.002	-0.002
λ_{21}	0.003	0.007	-0.003	0.004	0.006	0.004	-0.001	-0.001	-0.001
λ_{31}	0.013	0.021	0.007	0.005	0.008	0.008	0.001	0.002	0.002
λ_{41}	0.010	0.025	0.010	0.012	0.017	0.023	0.001	0.001	0.004
λ_{51}	0.398	0.049	-0.415	0.137	0.034	-0.215	0.045	0.029	-0.037
τ_{12}	0.023	0.023	0.024	-0.005	-0.005	-0.004	-0.000	0.000	0.000
τ_{13}	0.034	0.034	0.037	-0.002	-0.002	-0.001	0.001	0.002	0.002
τ_{14}	0.044	0.045	0.048	0.001	0.000	0.003	0.002	0.002	0.003
τ_{22}	0.005	0.008	0.010	0.002	0.002	0.004	0.000	0.001	0.001
τ_{23}	0.012	0.017	0.021	0.007	0.009	0.012	0.004	0.004	0.005
τ_{24}	0.006	0.013	0.019	0.010	0.013	0.017	0.005	0.006	0.007
τ_{32}	-0.001	0.004	0.006	0.010	0.010	0.014	0.002	0.002	0.003
τ_{33}	0.009	0.019	0.024	0.017	0.020	0.026	0.001	0.002	0.004
τ_{34}	0.016	0.031	0.039	0.019	0.023	0.032	0.001	0.001	0.004
τ_{42}	0.009	0.019	0.026	0.004	0.007	0.014	0.001	0.001	0.003
τ_{43}	0.021	0.040	0.052	0.009	0.015	0.029	0.001	0.002	0.006
τ_{44}	0.029	0.058	0.077	0.016	0.025	0.046	0.000	0.002	0.008
τ_{52}	0.186	0.006	-0.207	0.081	0.026	-0.097	0.029	0.020	-0.012
τ_{53}	0.417	0.048	-0.388	0.149	0.041	-0.205	0.048	0.031	-0.034
τ_{54}	0.632	0.075	-0.581	0.229	0.067	-0.302	0.082	0.057	-0.040
	ln(MSE)								
λ_{10}	-4.695	-4.700	-4.677	-6.201	-6.182	-6.183	-7.594	-7.602	-7.578
λ_{20}	-4.762	-4.718	-4.711	-6.001	-6.003	-5.973	-7.569	-7.563	-7.571
λ_{30}	-4.349	-4.304	-4.315	-5.493	-5.490	-5.445	-7.281	-7.294	-7.313
λ_{40}	-3.852	-3.744	-3.734	-5.225	-5.288	-5.241	-6.794	-6.788	-6.822
λ_{50}	-1.097	-2.279	-2.099	-2.409	-2.941	-3.155	-4.184	-4.302	-4.670
λ_{11}	-5.271	-5.251	-5.315	-6.406	-6.400	-6.402	-8.100	-8.099	-8.095
λ_{21}	-5.053	-5.054	-5.115	-6.164	-6.159	-6.197	-8.016	-7.989	-8.023
λ_{31}	-4.266	-4.237	-4.360	-5.883	-5.921	-5.973	-7.327	-7.365	-7.372
λ_{41}	-4.264	-4.251	-4.408	-5.302	-5.339	-5.407	-7.199	-7.186	-7.234
λ_{51}	-0.845	-2.179	-1.639	-2.112	-2.662	-2.726	-3.849	-3.972	-4.391
τ_{12}	-5.105	-5.120	-5.092	-6.330	-6.314	-6.311	-7.961	-7.979	-7.977
τ_{13}	-4.448	-4.434	-4.388	-6.007	-5.968	-5.983	-7.607	-7.621	-7.579
τ_{14}	-3.994	-3.996	-3.953	-5.868	-5.826	-5.830	-7.100	-7.094	-7.064
τ_{22}	-5.363	-5.322	-5.335	-6.342	-6.342	-6.329	-7.756	-7.775	-7.753
τ_{23}	-4.463	-4.419	-4.407	-5.767	-5.747	-5.714	-7.263	-7.266	-7.236
τ_{24}	-4.101	-4.079	-4.055	-5.209	-5.168	-5.144	-6.973	-6.956	-6.953
τ_{32}	-5.146	-5.096	-5.135	-5.800	-5.792	-5.760	-7.609	-7.612	-7.617
τ_{33}	-4.164	-4.109	-4.128	-5.152	-5.151	-5.068	-7.161	-7.180	-7.195
τ_{34}	-3.509	-3.438	-3.444	-4.995	-4.967	-4.898	-6.743	-6.777	-6.806
τ_{42}	-4.393	-4.349	-4.298	-5.615	-5.650	-5.583	-7.209	-7.212	-7.226
τ_{43}	-3.702	-3.626	-3.573	-5.149	-5.175	-5.074	-6.585	-6.563	-6.576
τ_{44}	-3.445	-3.360	-3.255	-4.643	-4.666	-4.576	-6.009	-5.971	-6.003
τ_{52}	-2.056	-2.909	-2.714	-3.214	-3.747	-3.949	-4.903	-4.996	-5.386
τ_{53}	-0.653	-1.850	-1.639	-1.912	-2.460	-2.649	-3.605	-3.681	-4.099
τ_{54}	0.166	-1.119	-0.882	-1.121	-1.704	-1.951	-2.845	-2.981	-3.451

Table C.4: *Estimates of parameter set C simulated by the MHS – BIAS and ln(MSE) of estimated parameters obtained by simulations of 100 different data sets.*

Simulations of parameter set C – MHS									
N	$N_1 = 300$			$N_2 = 1000$			$N_3 = 5000$		
Priori	Weak	Std.	Strong	Weak	Std.	Strong	Weak	Std.	Strong
	Coverage								
λ_{10}	91	92	93	98	98	98	93	94	94
λ_{20}	98	96	98	95	96	96	96	95	96
λ_{30}	96	96	95	96	95	94	96	97	97
λ_{40}	90	90	91	92	93	95	94	96	96
λ_{50}	87	91	75	95	92	80	88	90	89
λ_{11}	96	96	95	90	90	92	95	95	95
λ_{21}	96	96	95	95	95	95	98	99	100
λ_{31}	89	89	89	93	93	92	90	94	91
λ_{41}	93	95	95	93	96	97	96	97	98
λ_{51}	87	94	61	92	90	77	87	90	88
τ_{12}	95	95	94	91	90	91	94	96	95
τ_{13}	92	92	92	96	96	95	96	97	96
τ_{14}	93	93	92	99	99	98	92	94	95
τ_{22}	100	99	100	94	95	96	93	94	93
τ_{23}	95	95	95	95	94	95	93	93	92
τ_{24}	93	94	95	95	94	95	92	93	93
τ_{32}	96	98	98	94	93	92	94	94	93
τ_{33}	94	94	95	92	92	91	97	96	96
τ_{34}	92	92	93	92	94	93	93	94	95
τ_{42}	90	92	92	95	95	94	95	96	95
τ_{43}	93	94	93	95	94	96	95	91	95
τ_{44}	94	95	95	94	94	94	91	89	91
τ_{52}	93	91	77	93	94	88	90	93	93
τ_{53}	87	90	72	92	93	78	85	90	90
τ_{54}	87	91	70	92	91	78	86	91	89

Simulations of parameter set C – GGS									
N	$N_1 = 300$			$N_2 = 1000$			$N_3 = 5000$		
Priori	Weak	Std.	Strong	Weak	Std.	Strong	Weak	Std.	Strong
	Coverage								
λ_{10}	83	81	86	90	91	97	94	93	94
λ_{20}	81	81	84	80	85	95	95	97	96
λ_{30}	85	86	84	86	82	93	97	96	97
λ_{40}	86	87	86	90	89	91	94	95	95
λ_{50}	4	3	17	57	64	90	91	91	92
λ_{11}	5	5	21	60	63	89	95	95	95
λ_{21}	5	5	18	62	69	93	99	99	97
λ_{31}	4	3	18	59	65	91	90	91	90
λ_{41}	5	5	18	59	65	89	98	98	97
λ_{51}	5	5	18	54	64	88	88	89	92
τ_{12}	96	98	97	91	90	91	88	85	88
τ_{13}	92	93	91	94	96	95	95	92	93
τ_{14}	91	92	91	99	97	98	96	94	95
τ_{22}	96	98	96	93	93	96	88	88	88
τ_{23}	87	87	87	90	90	93	91	90	92
τ_{24}	80	75	84	85	84	91	93	93	94
τ_{32}	89	89	95	89	92	89	87	87	88
τ_{33}	72	67	79	80	83	92	96	95	95
τ_{34}	53	51	70	75	78	95	94	95	95
τ_{42}	81	76	83	78	79	92	88	91	86
τ_{43}	53	47	64	69	68	91	92	92	94
τ_{44}	22	16	37	59	65	90	92	91	91
τ_{52}	4	4	17	55	66	89	92	94	93
τ_{53}	4	4	17	56	65	88	91	92	93
τ_{54}	5	4	17	56	66	89	89	91	92

Table C.5: *Estimates of parameter set C simulated by the MHS (upper table) and the GGS (lower table) – coverage of estimated parameters obtained by simulations of 100 different data sets.*

Simulations of parameter set C – GGS sampler									
N	$N_1 = 300$			$N_2 = 1000$			$N_3 = 5000$		
Priori	Weak	Std.	Strong	Weak	Std.	Strong	Weak	Std.	Strong
	MEAN								
λ_{10}	1.035	1.037	1.025	0.968	0.967	0.947	0.946	0.946	0.945
λ_{20}	1.141	1.142	1.135	1.078	1.079	1.051	1.042	1.042	1.042
λ_{30}	1.210	1.210	1.204	1.168	1.168	1.142	1.123	1.123	1.124
λ_{40}	1.326	1.323	1.334	1.281	1.284	1.263	1.257	1.257	1.258
λ_{50}	4301.223	3870.414	422.284	645.511	468.811	16.875	2.123	2.120	2.109
λ_{11}	0.068	0.052	0.106	0.239	0.240	0.296	0.312	0.312	0.311
λ_{21}	0.117	0.087	0.186	0.440	0.444	0.551	0.576	0.575	0.574
λ_{31}	0.156	0.120	0.248	0.568	0.572	0.715	0.750	0.750	0.749
λ_{41}	0.187	0.143	0.305	0.741	0.749	0.936	0.979	0.978	0.977
λ_{51}	375.629	251.613	38.394	135.777	73.579	4.554	2.136	2.130	2.108
τ_{12}	0.644	0.644	0.649	0.623	0.623	0.626	0.630	0.630	0.629
τ_{13}	1.274	1.273	1.281	1.251	1.252	1.259	1.261	1.260	1.259
τ_{14}	1.901	1.898	1.909	1.880	1.881	1.889	1.890	1.890	1.889
τ_{22}	0.666	0.662	0.677	0.684	0.687	0.698	0.694	0.694	0.695
τ_{23}	1.324	1.317	1.344	1.370	1.374	1.395	1.389	1.390	1.391
τ_{24}	1.963	1.953	1.992	2.052	2.056	2.087	2.084	2.084	2.084
τ_{32}	0.693	0.691	0.710	0.740	0.743	0.760	0.752	0.753	0.753
τ_{33}	1.383	1.374	1.414	1.475	1.480	1.515	1.501	1.501	1.502
τ_{34}	2.054	2.040	2.101	2.201	2.207	2.263	2.250	2.251	2.251
τ_{42}	0.750	0.741	0.776	0.813	0.818	0.845	0.841	0.841	0.842
τ_{43}	1.474	1.460	1.527	1.619	1.627	1.684	1.681	1.681	1.683
τ_{44}	2.172	2.150	2.253	2.417	2.429	2.523	2.519	2.519	2.521
τ_{52}	4134.335	3740.751	403.644	591.282	432.487	15.087	1.420	1.419	1.413
τ_{53}	4427.546	3951.603	438.148	690.154	493.887	18.720	2.831	2.828	2.813
τ_{54}	4680.427	4134.182	465.594	779.566	547.707	21.972	4.256	4.249	4.227
	STD								
λ_{10}	0.101	0.102	0.102	0.056	0.057	0.047	0.022	0.023	0.023
λ_{20}	0.096	0.097	0.098	0.066	0.064	0.051	0.024	0.023	0.023
λ_{30}	0.103	0.102	0.111	0.075	0.076	0.064	0.027	0.027	0.026
λ_{40}	0.121	0.124	0.128	0.070	0.072	0.074	0.034	0.033	0.034
λ_{50}	1105.413	949.564	275.066	810.563	686.585	70.041	0.124	0.121	0.119
λ_{11}	0.068	0.060	0.098	0.094	0.104	0.055	0.017	0.017	0.017
λ_{21}	0.109	0.090	0.172	0.171	0.186	0.090	0.019	0.019	0.018
λ_{31}	0.176	0.169	0.255	0.225	0.244	0.116	0.025	0.025	0.025
λ_{41}	0.194	0.179	0.300	0.310	0.333	0.157	0.028	0.028	0.027
λ_{51}	87.921	53.803	18.377	155.644	96.371	9.820	0.145	0.143	0.135
τ_{12}	0.074	0.072	0.074	0.044	0.044	0.044	0.020	0.020	0.021
τ_{13}	0.102	0.101	0.104	0.051	0.050	0.051	0.023	0.024	0.024
τ_{14}	0.126	0.125	0.129	0.056	0.056	0.055	0.029	0.029	0.029
τ_{22}	0.066	0.063	0.070	0.044	0.043	0.043	0.023	0.023	0.022
τ_{23}	0.102	0.095	0.113	0.061	0.063	0.061	0.027	0.027	0.026
τ_{24}	0.121	0.116	0.136	0.082	0.086	0.078	0.030	0.030	0.030
τ_{32}	0.075	0.070	0.079	0.058	0.058	0.055	0.024	0.024	0.024
τ_{33}	0.118	0.116	0.124	0.085	0.086	0.076	0.029	0.028	0.029
τ_{34}	0.163	0.157	0.172	0.110	0.110	0.087	0.034	0.034	0.034
τ_{42}	0.099	0.099	0.110	0.075	0.080	0.065	0.028	0.028	0.030
τ_{43}	0.139	0.142	0.170	0.121	0.127	0.088	0.039	0.037	0.039
τ_{44}	0.169	0.175	0.228	0.173	0.185	0.119	0.050	0.051	0.051
τ_{52}	1065.713	921.352	266.320	747.315	637.590	65.925	0.087	0.087	0.087
τ_{53}	1133.097	965.550	281.649	860.237	718.466	74.729	0.166	0.165	0.161
τ_{54}	1186.617	1000.951	293.793	961.639	789.040	82.045	0.236	0.232	0.223

Table C.6: *Estimates of parameter set C simulated by the GGS – MEAN and STD of estimated parameters obtained by simulations of 100 different data sets.*

Simulations of parameter set C – GGS sampler									
N	**$N_1 = 300$**			**$N_2 = 1000$**			**$N_3 = 5000$**		
Priori	**Weak**	**Std.**	**Strong**	**Weak**	**Std.**	**Strong**	**Weak**	**Std.**	**Strong**
	BIAS								
λ_{10}	0.091	0.093	0.082	0.025	0.024	0.004	0.003	0.003	0.002
λ_{20}	0.102	0.103	0.096	0.039	0.039	0.012	0.003	0.003	0.003
λ_{30}	0.085	0.085	0.079	0.043	0.043	0.017	-0.002	-0.002	-0.001
λ_{40}	0.066	0.063	0.073	0.021	0.024	0.003	-0.003	-0.003	-0.002
λ_{50}	4299.159	3868.349	420.219	643.446	466.746	14.810	0.058	0.055	0.044
λ_{11}	-0.246	-0.263	-0.209	-0.075	-0.074	-0.018	-0.002	-0.003	-0.003
λ_{21}	-0.461	-0.490	-0.392	-0.137	-0.134	-0.026	-0.002	-0.002	-0.003
λ_{31}	-0.594	-0.630	-0.502	-0.182	-0.178	-0.035	0.000	0.000	-0.001
λ_{41}	-0.793	-0.837	-0.675	-0.239	-0.231	-0.044	-0.001	-0.002	-0.003
λ_{51}	373.564	249.549	36.329	133.712	71.515	2.489	0.071	0.065	0.043
τ_{12}	0.015	0.015	0.020	-0.006	-0.006	-0.003	0.001	0.001	-0.000
τ_{13}	0.016	0.015	0.023	-0.007	-0.006	0.001	0.003	0.002	0.001
τ_{14}	0.014	0.011	0.023	-0.007	-0.006	0.002	0.003	0.003	0.002
τ_{22}	-0.027	-0.031	-0.015	-0.008	-0.006	0.005	0.001	0.002	0.002
τ_{23}	-0.062	-0.069	-0.042	-0.016	-0.012	0.009	0.004	0.004	0.006
τ_{24}	-0.116	-0.126	-0.086	-0.027	-0.023	0.009	0.006	0.005	0.005
τ_{32}	-0.057	-0.059	-0.040	-0.010	-0.007	0.010	0.002	0.003	0.003
τ_{33}	-0.117	-0.126	-0.086	-0.025	-0.020	0.015	0.001	0.001	0.002
τ_{34}	-0.196	-0.210	-0.149	-0.049	-0.043	0.013	0.000	0.001	0.001
τ_{42}	-0.090	-0.099	-0.064	-0.027	-0.022	0.005	0.001	0.000	0.002
τ_{43}	-0.206	-0.221	-0.153	-0.061	-0.054	0.004	0.001	0.000	0.002
τ_{44}	-0.348	-0.370	-0.268	-0.103	-0.092	0.002	-0.001	-0.001	0.001
τ_{52}	4132.958	3739.375	402.267	589.906	431.110	13.710	0.044	0.042	0.036
τ_{53}	4424.793	3948.850	435.395	687.402	491.134	15.967	0.078	0.075	0.060
τ_{54}	4676.298	4130.053	461.465	775.436	543.577	17.842	0.127	0.119	0.098
	ln(MSE)								
λ_{10}	-3.988	-3.966	-4.075	-5.587	-5.590	-6.114	-7.598	-7.549	-7.549
λ_{20}	-3.935	-3.916	-3.982	-5.153	-5.193	-5.907	-7.493	-7.567	-7.567
λ_{30}	-4.038	-4.043	-3.998	-4.913	-4.883	-5.430	-7.252	-7.260	-7.276
λ_{40}	-3.977	-3.962	-3.839	-5.244	-5.168	-5.220	-6.753	-6.817	-6.776
λ_{50}	16.796	16.579	12.435	13.878	13.436	8.532	-3.988	-4.038	-4.136
λ_{11}	-2.732	-2.622	-2.936	-4.241	-4.127	-5.702	-8.096	-8.094	-8.087
λ_{21}	-1.496	-1.394	-1.700	-3.043	-2.957	-4.743	-7.975	-7.965	-7.967
λ_{31}	-0.958	-0.856	-1.150	-2.487	-2.399	-4.228	-7.363	-7.377	-7.369
λ_{41}	-0.405	-0.311	-0.607	-1.880	-1.812	-3.642	-7.186	-7.180	-7.183
λ_{51}	11.900	11.084	7.411	10.642	9.569	4.622	-3.652	-3.713	-3.914
τ_{12}	-5.186	-5.218	-5.153	-6.246	-6.241	-6.272	-7.812	-7.785	-7.777
τ_{13}	-4.548	-4.578	-4.486	-5.936	-5.996	-5.950	-7.564	-7.496	-7.463
τ_{14}	-4.141	-4.168	-4.075	-5.777	-5.776	-5.816	-7.101	-7.097	-7.102
τ_{22}	-5.298	-5.327	-5.289	-6.229	-6.294	-6.282	-7.520	-7.582	-7.663
τ_{23}	-4.265	-4.289	-4.244	-5.538	-5.516	-5.594	-7.208	-7.188	-7.233
τ_{24}	-3.577	-3.532	-3.663	-4.910	-4.851	-5.090	-6.959	-6.977	-6.968
τ_{32}	-4.742	-4.785	-4.847	-5.690	-5.700	-5.779	-7.490	-7.460	-7.460
τ_{33}	-3.593	-3.537	-3.790	-4.863	-4.859	-5.122	-7.089	-7.134	-7.119
τ_{34}	-2.739	-2.682	-2.971	-4.245	-4.280	-4.871	-6.781	-6.773	-6.795
τ_{42}	-4.024	-3.941	-4.123	-5.058	-4.993	-5.476	-7.139	-7.130	-6.991
τ_{43}	-2.788	-2.679	-2.954	-4.009	-3.968	-4.871	-6.520	-6.587	-6.520
τ_{44}	-1.900	-1.787	-2.095	-3.211	-3.162	-4.268	-5.990	-5.977	-5.978
τ_{52}	16.717	16.512	12.355	13.711	13.285	8.410	-4.662	-4.684	-4.737
τ_{53}	16.853	16.620	12.499	14.002	13.531	8.663	-3.406	-3.422	-3.535
τ_{54}	16.962	16.709	12.606	14.232	13.723	8.851	-2.640	-2.697	-2.832

Table C.7: *Estimates of parameter set C simulated by the GGS – BIAS and ln(MSE) of estimated parameters obtained by simulations of 100 different data sets.*

Simulations of parameter set E – MHS									
N	**$N_1 = 300$**			**$N_2 = 1000$**			**$N_3 = 5000$**		
Priori	**Weak**	**Std.**	**Strong**	**Weak**	**Std.**	**Strong**	**Weak**	**Std.**	**Strong**
	MEAN								
λ_{10}	-0.537	-0.528	-0.489	-0.509	-0.507	-0.496	-0.499	-0.499	-0.497
λ_{20}	0.543	0.533	0.498	0.518	0.516	0.505	0.502	0.502	0.499
λ_{30}	1.235	1.225	1.180	1.169	1.166	1.153	1.170	1.170	1.168
λ_{40}	1.042	1.038	1.020	1.008	1.008	1.003	0.996	0.996	0.995
λ_{50}	1.246	1.235	1.179	1.174	1.172	1.159	1.171	1.171	1.167
λ_{60}	0.866	0.860	0.842	0.850	0.848	0.839	0.836	0.836	0.835
λ_{70}	0.746	0.732	0.739	0.711	0.713	0.708	0.709	0.710	0.709
λ_{80}	0.877	0.875	0.855	0.848	0.847	0.837	0.838	0.838	0.837
λ_{90}	2.019	2.015	2.018	1.991	1.991	1.992	2.004	2.005	2.004
$\lambda_{10,0}$	-4.945	-4.946	-4.948	-5.006	-5.005	-5.007	-4.992	-4.992	-4.992
a_{11}	0.911	0.891	0.827	0.852	0.848	0.829	0.845	0.845	0.840
a_{12}	-0.903	-0.882	-0.820	-0.859	-0.855	-0.837	-0.836	-0.836	-0.832
λ_{11}	1.489	1.427	1.232	1.386	1.372	1.311	1.341	1.340	1.327
λ_{12}	–	–	–	–	–	–	–	–	–
a_{21}	-0.927	-0.911	-0.846	-0.857	-0.854	-0.836	-0.837	-0.836	-0.833
a_{22}	0.917	0.901	0.837	0.869	0.866	0.846	0.834	0.833	0.830
λ_{21}	-0.011	-0.010	-0.014	-0.008	-0.008	-0.011	-0.002	-0.002	-0.002
λ_{22}	1.472	1.424	1.233	1.387	1.376	1.316	1.333	1.331	1.319
a_{31}	0.891	0.885	0.846	0.844	0.842	0.831	0.839	0.839	0.837
a_{32}	-0.881	-0.874	-0.838	-0.848	-0.846	-0.836	-0.835	-0.835	-0.833
λ_{31}	1.435	1.406	1.270	1.350	1.343	1.302	1.335	1.334	1.326
λ_{32}	-0.012	-0.014	-0.013	0.001	0.000	0.001	-0.003	-0.003	-0.004
a_{41}	-0.751	-0.748	-0.734	-0.720	-0.719	-0.716	-0.711	-0.711	-0.710
a_{42}	0.753	0.750	0.736	0.719	0.719	0.714	0.709	0.709	0.709
λ_{41}	0.735	0.725	0.673	0.723	0.721	0.704	0.707	0.706	0.703
λ_{42}	0.748	0.736	0.686	0.717	0.715	0.700	0.705	0.704	0.701
a_{51}	0.844	0.836	0.798	0.844	0.843	0.832	0.838	0.838	0.835
a_{52}	-0.862	-0.854	-0.815	-0.835	-0.833	-0.824	-0.834	-0.834	-0.832
λ_{51}	0.015	0.015	0.006	-0.001	-0.001	-0.005	0.002	0.002	0.002
λ_{52}	1.407	1.379	1.242	1.344	1.337	1.298	1.336	1.335	1.326
a_{61}	-0.882	-0.876	-0.852	-0.841	-0.839	-0.830	-0.835	-0.835	-0.834
a_{62}	0.866	0.861	0.838	0.856	0.855	0.845	0.841	0.840	0.839
λ_{61}	1.419	1.392	1.287	1.369	1.364	1.326	1.345	1.344	1.337
λ_{62}	-0.010	-0.013	-0.011	0.005	0.004	0.004	-0.006	-0.006	-0.007
a_{71}	0.718	0.711	0.708	0.712	0.713	0.709	0.711	0.711	0.710
a_{72}	-0.716	-0.709	-0.706	-0.710	-0.710	-0.707	-0.705	-0.705	-0.705
λ_{71}	0.761	0.745	0.707	0.719	0.718	0.702	0.709	0.709	0.707
λ_{72}	0.728	0.712	0.677	0.719	0.717	0.703	0.704	0.704	0.701
a_{81}	-0.894	-0.891	-0.863	-0.856	-0.853	-0.844	-0.835	-0.835	-0.833
a_{82}	0.881	0.878	0.850	0.854	0.852	0.842	0.836	0.835	0.834
λ_{81}	0.013	0.014	0.006	0.003	0.003	-0.001	-0.003	-0.002	-0.002
λ_{82}	1.409	1.388	1.277	1.358	1.351	1.314	1.331	1.330	1.324
a_{91}	0.499	0.503	0.501	0.504	0.505	0.504	0.508	0.508	0.508
a_{92}	-0.483	-0.483	-0.483	-0.511	-0.510	-0.512	-0.497	-0.497	-0.497
λ_{91}	2.969	2.955	2.899	3.010	3.006	2.987	3.002	3.001	2.997
λ_{92}	-0.017	-0.024	-0.026	-0.003	-0.005	-0.004	-0.003	-0.004	-0.006
$a_{10,1}$	-0.554	-0.555	-0.554	-0.501	-0.500	-0.501	-0.507	-0.508	-0.509
$a_{10,2}$	0.496	0.499	0.496	0.513	0.514	0.511	0.502	0.501	0.502
$\lambda_{10,1}$	0.035	0.037	0.003	0.002	0.002	-0.013	-0.001	0.001	0.001
$\lambda_{10,2}$	3.989	3.972	3.901	3.983	3.978	3.954	4.000	3.998	3.994
τ_{32}	2.471	2.456	2.360	2.347	2.342	2.314	2.342	2.341	2.336
τ_{42}	2.064	2.059	2.022	2.006	2.006	1.995	1.984	1.983	1.981
τ_{52}	2.454	2.434	2.325	2.356	2.351	2.325	2.341	2.341	2.334
τ_{62}	0.849	0.847	0.824	0.847	0.845	0.834	0.838	0.839	0.837
τ_{63}	1.898	1.891	1.838	1.867	1.863	1.842	1.841	1.841	1.838
τ_{64}	2.936	2.916	2.845	2.887	2.882	2.850	2.844	2.844	2.840
τ_{65}	3.991	3.959	3.862	3.901	3.896	3.855	3.847	3.847	3.841
τ_{72}	0.743	0.732	0.736	0.717	0.719	0.714	0.708	0.708	0.707
τ_{73}	1.612	1.591	1.595	1.579	1.582	1.572	1.559	1.560	1.559

τ_{74}	2.480	2.448	2.454	2.429	2.433	2.419	2.413	2.413	2.412
τ_{75}	3.363	3.322	3.321	3.287	3.293	3.273	3.259	3.260	3.258
τ_{82}	0.865	0.865	0.847	0.853	0.851	0.842	0.832	0.831	0.830
τ_{83}	1.881	1.877	1.829	1.866	1.861	1.842	1.834	1.832	1.830
τ_{84}	2.910	2.899	2.820	2.879	2.872	2.842	2.840	2.839	2.835
τ_{85}	3.967	3.954	3.839	3.911	3.901	3.858	3.843	3.842	3.836
					STD				
λ_{10}	0.152	0.144	0.131	0.074	0.073	0.072	0.030	0.030	0.030
λ_{20}	0.129	0.124	0.112	0.074	0.075	0.072	0.032	0.032	0.032
λ_{30}	0.181	0.169	0.158	0.083	0.081	0.080	0.034	0.034	0.034
λ_{40}	0.149	0.143	0.139	0.075	0.075	0.074	0.028	0.028	0.027
λ_{50}	0.183	0.175	0.159	0.094	0.095	0.090	0.040	0.039	0.039
λ_{60}	0.141	0.134	0.136	0.084	0.087	0.084	0.039	0.039	0.039
λ_{70}	0.116	0.117	0.117	0.066	0.069	0.068	0.027	0.027	0.027
λ_{80}	0.154	0.154	0.157	0.081	0.080	0.078	0.032	0.033	0.033
λ_{90}	0.234	0.234	0.233	0.109	0.108	0.109	0.045	0.045	0.045
$\lambda_{10,0}$	0.237	0.233	0.236	0.125	0.126	0.125	0.057	0.057	0.057
a_{11}	0.181	0.166	0.138	0.088	0.087	0.081	0.038	0.038	0.037
a_{12}	0.209	0.197	0.170	0.084	0.085	0.077	0.042	0.042	0.041
λ_{11}	0.276	0.233	0.154	0.134	0.133	0.114	0.056	0.055	0.053
λ_{12}	–	–	–	–	–	–	–	–	–
a_{21}	0.187	0.180	0.153	0.095	0.094	0.090	0.041	0.040	0.040
a_{22}	0.220	0.208	0.178	0.087	0.084	0.080	0.036	0.036	0.036
λ_{21}	0.176	0.176	0.150	0.097	0.095	0.091	0.049	0.050	0.049
λ_{22}	0.255	0.224	0.151	0.120	0.115	0.102	0.052	0.052	0.051
a_{31}	0.144	0.139	0.127	0.078	0.077	0.074	0.033	0.033	0.032
a_{32}	0.140	0.134	0.129	0.080	0.080	0.078	0.033	0.033	0.033
λ_{31}	0.194	0.180	0.131	0.093	0.091	0.084	0.042	0.042	0.042
λ_{32}	0.150	0.151	0.135	0.066	0.064	0.063	0.037	0.037	0.036
a_{41}	0.114	0.113	0.109	0.057	0.057	0.057	0.025	0.025	0.025
a_{42}	0.117	0.115	0.110	0.062	0.062	0.062	0.027	0.027	0.027
λ_{41}	0.132	0.130	0.116	0.068	0.069	0.066	0.032	0.033	0.032
λ_{42}	0.126	0.124	0.111	0.056	0.056	0.054	0.026	0.026	0.026
a_{51}	0.151	0.152	0.137	0.068	0.068	0.066	0.035	0.035	0.035
a_{52}	0.150	0.152	0.134	0.061	0.061	0.060	0.029	0.029	0.029
λ_{51}	0.186	0.183	0.163	0.083	0.083	0.078	0.042	0.043	0.042
λ_{52}	0.185	0.181	0.138	0.097	0.096	0.088	0.042	0.041	0.041
a_{61}	0.117	0.131	0.121	0.070	0.070	0.066	0.032	0.032	0.032
a_{62}	0.111	0.123	0.121	0.071	0.070	0.072	0.033	0.033	0.032
λ_{61}	0.169	0.164	0.137	0.095	0.094	0.089	0.045	0.045	0.044
λ_{62}	0.144	0.142	0.130	0.070	0.070	0.067	0.035	0.035	0.034
a_{71}	0.103	0.101	0.102	0.061	0.062	0.061	0.025	0.025	0.025
a_{72}	0.093	0.091	0.091	0.057	0.057	0.057	0.025	0.026	0.025
λ_{71}	0.121	0.119	0.115	0.064	0.066	0.064	0.030	0.030	0.030
λ_{72}	0.104	0.103	0.095	0.064	0.062	0.061	0.027	0.028	0.027
a_{81}	0.121	0.126	0.117	0.076	0.074	0.073	0.029	0.029	0.029
a_{82}	0.112	0.112	0.110	0.060	0.059	0.057	0.031	0.031	0.031
λ_{81}	0.146	0.143	0.130	0.084	0.082	0.080	0.044	0.045	0.044
λ_{82}	0.166	0.154	0.134	0.092	0.087	0.082	0.031	0.032	0.032
a_{91}	0.181	0.183	0.181	0.111	0.110	0.111	0.053	0.053	0.053
a_{92}	0.193	0.193	0.194	0.103	0.103	0.102	0.055	0.055	0.055
λ_{91}	0.187	0.181	0.180	0.102	0.101	0.100	0.047	0.047	0.047
λ_{92}	0.251	0.255	0.247	0.143	0.142	0.142	0.076	0.078	0.076
$a_{10,1}$	0.246	0.249	0.250	0.131	0.128	0.131	0.065	0.066	0.064
$a_{10,2}$	0.234	0.240	0.239	0.128	0.129	0.129	0.059	0.060	0.060
$\lambda_{10,1}$	0.389	0.391	0.379	0.238	0.236	0.230	0.119	0.122	0.120
$\lambda_{10,2}$	0.200	0.199	0.194	0.122	0.122	0.121	0.054	0.054	0.054
τ_{32}	0.260	0.242	0.210	0.130	0.127	0.123	0.053	0.054	0.053
τ_{42}	0.191	0.185	0.173	0.094	0.094	0.092	0.043	0.043	0.043
τ_{52}	0.257	0.253	0.210	0.132	0.131	0.125	0.063	0.061	0.062
τ_{62}	0.139	0.126	0.118	0.066	0.065	0.061	0.029	0.028	0.029
τ_{63}	0.203	0.200	0.195	0.115	0.116	0.113	0.050	0.049	0.050
τ_{64}	0.283	0.272	0.264	0.160	0.158	0.157	0.067	0.067	0.068
τ_{65}	0.362	0.356	0.348	0.204	0.205	0.199	0.085	0.088	0.088
τ_{72}	0.104	0.100	0.104	0.058	0.059	0.059	0.027	0.028	0.027

τ_{73}	0.155	0.159	0.162	0.072	0.074	0.072	0.033	0.035	0.034
τ_{74}	0.201	0.216	0.217	0.110	0.114	0.109	0.045	0.047	0.047
τ_{75}	0.271	0.266	0.273	0.132	0.135	0.130	0.053	0.056	0.056
τ_{82}	0.141	0.138	0.141	0.070	0.066	0.068	0.030	0.030	0.030
τ_{83}	0.209	0.196	0.193	0.109	0.102	0.100	0.036	0.036	0.037
τ_{84}	0.269	0.247	0.241	0.149	0.136	0.136	0.052	0.054	0.054
τ_{85}	0.354	0.355	0.346	0.193	0.175	0.176	0.065	0.067	0.066
	BIAS								
λ_{10}	-0.037	-0.028	0.011	-0.009	-0.007	0.004	0.001	0.001	0.003
λ_{20}	0.043	0.033	-0.002	0.018	0.016	0.005	0.002	0.002	-0.001
λ_{30}	0.068	0.059	0.013	0.002	-0.001	-0.014	0.003	0.004	0.001
λ_{40}	0.052	0.048	0.030	0.018	0.018	0.013	0.006	0.006	0.005
λ_{50}	0.080	0.069	0.013	0.007	0.006	-0.008	0.004	0.004	0.001
λ_{60}	0.033	0.027	0.009	0.016	0.015	0.006	0.002	0.003	0.002
λ_{70}	0.039	0.025	0.032	0.004	0.006	0.001	0.002	0.003	0.002
λ_{80}	0.043	0.041	0.022	0.015	0.013	0.004	0.005	0.005	0.004
λ_{90}	0.019	0.015	0.018	-0.009	-0.009	-0.008	0.004	0.005	0.004
$\lambda_{10,0}$	0.055	0.054	0.052	-0.006	-0.005	-0.007	0.008	0.008	0.008
a_{11}	0.078	0.058	-0.006	0.018	0.015	-0.005	0.011	0.011	0.007
a_{12}	-0.070	-0.049	0.013	-0.026	-0.021	-0.003	-0.002	-0.002	0.002
λ_{11}	0.156	0.094	-0.101	0.052	0.039	-0.022	0.008	0.007	-0.006
λ_{12}	–	–	–	–	–	–	–	–	–
a_{21}	-0.094	-0.077	-0.013	-0.024	-0.020	-0.002	-0.004	-0.003	0.000
a_{22}	0.084	0.068	0.004	0.036	0.033	0.013	0.001	-0.000	-0.003
λ_{21}	-0.011	-0.010	-0.014	-0.008	-0.008	-0.011	-0.002	-0.002	-0.002
λ_{22}	0.139	0.091	-0.101	0.054	0.043	-0.017	-0.000	-0.002	-0.014
a_{31}	0.057	0.052	0.013	0.010	0.009	-0.003	0.006	0.006	0.004
a_{32}	-0.048	-0.040	-0.004	-0.015	-0.012	-0.002	-0.001	-0.002	0.001
λ_{31}	0.102	0.073	-0.063	0.017	0.010	-0.031	0.001	0.000	-0.008
λ_{32}	-0.012	-0.014	-0.013	0.001	0.000	0.001	-0.003	-0.003	-0.004
a_{41}	-0.044	-0.040	-0.027	-0.013	-0.012	-0.009	-0.004	-0.004	-0.003
a_{42}	0.045	0.043	0.029	0.012	0.012	0.007	0.002	0.002	0.001
λ_{41}	0.028	0.018	-0.034	0.016	0.014	-0.003	-0.000	-0.001	-0.004
λ_{42}	0.041	0.029	-0.021	0.010	0.007	-0.007	-0.002	-0.003	-0.006
a_{51}	0.010	0.002	-0.035	0.011	0.009	-0.001	0.004	0.004	0.002
a_{52}	-0.029	-0.021	0.018	-0.002	0.000	0.009	-0.001	-0.001	0.001
λ_{51}	0.015	0.015	0.006	-0.001	-0.001	-0.005	0.002	0.002	0.002
λ_{52}	0.074	0.046	-0.091	0.010	0.003	-0.035	0.003	0.002	-0.007
a_{61}	-0.049	-0.043	-0.019	-0.007	-0.006	0.003	-0.002	-0.002	-0.001
a_{62}	0.033	0.028	0.004	0.023	0.022	0.011	0.007	0.007	0.006
λ_{61}	0.086	0.059	-0.046	0.036	0.030	-0.008	0.012	0.010	0.004
λ_{62}	-0.010	-0.013	-0.011	0.005	0.004	0.004	-0.006	-0.006	-0.007
a_{71}	0.011	0.004	0.001	0.005	0.006	0.002	0.003	0.004	0.003
a_{72}	-0.009	-0.002	0.001	-0.003	-0.003	-0.000	0.002	0.002	0.002
λ_{71}	0.054	0.038	0.000	0.012	0.011	-0.006	0.002	0.002	-0.000
λ_{72}	0.021	0.005	-0.031	0.012	0.010	-0.004	-0.003	-0.004	-0.006
a_{81}	-0.061	-0.058	-0.030	-0.022	-0.020	-0.011	-0.001	-0.001	-0.000
a_{82}	0.048	0.045	0.016	0.021	0.019	0.009	0.002	0.002	0.001
λ_{81}	0.013	0.014	0.006	0.003	0.003	-0.001	-0.003	-0.002	-0.002
λ_{82}	0.076	0.055	-0.056	0.025	0.017	-0.019	-0.002	-0.003	-0.010
a_{91}	-0.001	0.003	0.001	0.004	0.005	0.004	0.008	0.008	0.008
a_{92}	0.017	0.017	0.017	-0.011	-0.010	-0.012	0.003	0.003	0.003
λ_{91}	-0.031	-0.045	-0.101	0.010	0.006	-0.013	0.002	0.001	-0.003
λ_{92}	-0.017	-0.024	-0.026	-0.003	-0.005	-0.004	-0.003	-0.004	-0.006
$a_{10,1}$	-0.054	-0.055	-0.054	-0.001	-0.000	-0.001	-0.007	-0.008	-0.009
$a_{10,2}$	-0.004	-0.001	-0.004	0.013	0.014	0.011	0.002	0.001	0.002
$\lambda_{10,1}$	0.035	0.037	0.003	0.002	0.002	-0.013	-0.001	0.001	0.001
$\lambda_{10,2}$	-0.011	-0.028	-0.099	-0.017	-0.022	-0.046	-0.000	-0.002	-0.006
τ_{32}	0.137	0.122	0.027	0.014	0.009	-0.019	0.008	0.008	0.003
τ_{42}	0.084	0.079	0.042	0.026	0.026	0.015	0.004	0.003	0.001
τ_{52}	0.121	0.100	-0.008	0.022	0.018	-0.008	0.008	0.007	0.001
τ_{62}	0.015	0.014	-0.010	0.013	0.011	0.001	0.005	0.005	0.004
τ_{63}	0.065	0.058	0.005	0.033	0.030	0.009	0.008	0.008	0.005
τ_{64}	0.102	0.083	0.012	0.054	0.048	0.016	0.011	0.011	0.007
τ_{65}	0.158	0.126	0.028	0.068	0.062	0.021	0.014	0.014	0.008

τ_{72}	0.036	0.024	0.029	0.010	0.012	0.007	0.001	0.001	0.000
τ_{73}	0.057	0.035	0.039	0.023	0.026	0.017	0.004	0.004	0.003
τ_{74}	0.076	0.044	0.049	0.025	0.029	0.015	0.009	0.009	0.008
τ_{75}	0.111	0.069	0.068	0.035	0.040	0.021	0.007	0.007	0.005
τ_{82}	0.032	0.031	0.013	0.020	0.018	0.009	-0.001	-0.002	-0.003
τ_{83}	0.047	0.043	-0.005	0.033	0.028	0.008	0.001	-0.001	-0.003
τ_{84}	0.077	0.065	-0.013	0.046	0.039	0.008	0.007	0.005	0.002
τ_{85}	0.134	0.121	0.005	0.077	0.067	0.025	0.010	0.008	0.003
	ln(MSE)								
λ_{10}	-3.717	-3.854	-4.075	-5.191	-5.226	-5.274	-6.999	-7.014	-7.014
λ_{20}	-4.004	-4.123	-4.397	-5.156	-5.157	-5.279	-6.901	-6.903	-6.914
λ_{30}	-3.295	-3.447	-3.695	-4.976	-5.028	-5.027	-6.766	-6.767	-6.778
λ_{40}	-3.704	-3.795	-3.907	-5.122	-5.123	-5.200	-7.122	-7.105	-7.164
λ_{50}	-3.233	-3.353	-3.682	-4.733	-4.717	-4.816	-6.448	-6.479	-6.487
λ_{60}	-3.880	-3.993	-3.997	-4.917	-4.874	-4.962	-6.503	-6.511	-6.481
λ_{70}	-4.214	-4.249	-4.226	-5.427	-5.348	-5.394	-7.260	-7.193	-7.208
λ_{80}	-3.675	-3.683	-3.693	-4.998	-5.042	-5.103	-6.867	-6.827	-6.809
λ_{90}	-2.906	-2.912	-2.915	-4.439	-4.452	-4.433	-6.196	-6.184	-6.185
$\lambda_{10,0}$	-2.833	-2.867	-2.846	-4.159	-4.144	-4.171	-5.714	-5.707	-5.704
a_{11}	-3.256	-3.480	-3.964	-4.839	-4.865	-5.042	-6.463	-6.474	-6.552
a_{12}	-3.033	-3.194	-3.544	-4.876	-4.883	-5.135	-6.335	-6.369	-6.385
λ_{11}	-2.305	-2.774	-3.390	-3.889	-3.958	-4.317	-5.767	-5.785	-5.862
λ_{12}	–	–	–	–	–	–	–	–	–
a_{21}	-3.141	-3.268	-3.759	-4.646	-4.700	-4.820	-6.406	-6.418	-6.447
a_{22}	-2.906	-3.051	-3.460	-4.738	-4.814	-5.034	-6.637	-6.632	-6.648
λ_{21}	-3.476	-3.483	-3.801	-4.671	-4.701	-4.790	-6.022	-5.995	-6.046
λ_{22}	-2.481	-2.848	-3.426	-4.066	-4.206	-4.552	-5.905	-5.923	-5.874
a_{31}	-3.743	-3.825	-4.127	-5.105	-5.133	-5.204	-6.828	-6.828	-6.865
a_{32}	-3.824	-3.942	-4.103	-5.016	-5.041	-5.122	-6.859	-6.841	-6.860
λ_{31}	-3.047	-3.288	-3.861	-4.737	-4.801	-4.839	-6.372	-6.366	-6.333
λ_{32}	-3.795	-3.778	-4.004	-5.448	-5.509	-5.549	-6.623	-6.593	-6.625
a_{41}	-4.207	-4.252	-4.377	-5.678	-5.679	-5.723	-7.344	-7.351	-7.365
a_{42}	-4.164	-4.197	-4.349	-5.533	-5.538	-5.573	-7.236	-7.246	-7.255
λ_{41}	-4.010	-4.078	-4.236	-5.342	-5.318	-5.438	-6.863	-6.850	-6.895
λ_{42}	-4.052	-4.126	-4.371	-5.731	-5.769	-5.845	-7.280	-7.277	-7.256
a_{51}	-3.792	-3.780	-3.918	-5.353	-5.354	-5.444	-6.680	-6.683	-6.720
a_{52}	-3.773	-3.757	-4.015	-5.607	-5.610	-5.614	-7.103	-7.108	-7.104
λ_{51}	-3.371	-3.398	-3.635	-4.977	-4.992	-5.099	-6.331	-6.310	-6.333
λ_{52}	-3.235	-3.365	-3.611	-4.664	-4.690	-4.726	-6.340	-6.386	-6.388
a_{61}	-4.133	-3.978	-4.207	-5.328	-5.331	-5.434	-6.884	-6.880	-6.915
a_{62}	-4.327	-4.157	-4.237	-5.193	-5.224	-5.260	-6.767	-6.801	-6.843
λ_{61}	-3.332	-3.498	-3.877	-4.573	-4.640	-4.833	-6.147	-6.140	-6.229
λ_{62}	-3.887	-3.910	-4.089	-5.328	-5.336	-5.401	-6.714	-6.682	-6.715
a_{71}	-4.555	-4.593	-4.583	-5.594	-5.546	-5.596	-7.379	-7.365	-7.387
a_{72}	-4.751	-4.793	-4.797	-5.732	-5.726	-5.728	-7.347	-7.343	-7.363
λ_{71}	-4.048	-4.162	-4.340	-5.476	-5.421	-5.498	-7.035	-7.002	-7.052
λ_{72}	-4.489	-4.560	-4.627	-5.477	-5.550	-5.591	-7.195	-7.163	-7.152
a_{81}	-4.004	-3.954	-4.243	-5.073	-5.141	-5.212	-7.077	-7.064	-7.075
a_{82}	-4.219	-4.237	-4.404	-5.521	-5.566	-5.709	-6.966	-6.963	-6.976
λ_{81}	-3.851	-3.894	-4.087	-4.957	-5.005	-5.067	-6.259	-6.204	-6.235
λ_{82}	-3.409	-3.635	-3.867	-4.706	-4.855	-4.952	-6.929	-6.880	-6.808
a_{91}	-3.430	-3.404	-3.428	-4.409	-4.416	-4.402	-5.862	-5.867	-5.862
a_{92}	-3.292	-3.290	-3.279	-4.543	-4.553	-4.555	-5.791	-5.806	-5.825
λ_{91}	-3.341	-3.364	-3.160	-4.560	-4.587	-4.591	-6.125	-6.122	-6.120
λ_{92}	-2.772	-2.734	-2.794	-3.896	-3.915	-3.917	-5.157	-5.116	-5.148
$a_{10,1}$	-2.764	-2.742	-2.733	-4.081	-4.114	-4.082	-5.458	-5.443	-5.481
$a_{10,2}$	-2.913	-2.865	-2.876	-4.112	-4.090	-4.099	-5.661	-5.632	-5.637
$\lambda_{10,1}$	-1.892	-1.879	-1.951	-2.883	-2.899	-2.945	-4.262	-4.217	-4.251
$\lambda_{10,2}$	-3.223	-3.224	-3.054	-4.190	-4.190	-4.095	-5.842	-5.837	-5.824
τ_{32}	-2.454	-2.616	-3.113	-4.082	-4.127	-4.175	-5.856	-5.841	-5.882
τ_{42}	-3.145	-3.215	-3.455	-4.662	-4.668	-4.752	-6.303	-6.289	-6.316
τ_{52}	-2.523	-2.610	-3.133	-4.026	-4.061	-4.170	-5.521	-5.577	-5.582
τ_{62}	-3.940	-4.134	-4.281	-5.396	-5.458	-5.590	-7.080	-7.119	-7.052
τ_{63}	-3.100	-3.151	-3.280	-4.258	-4.249	-4.358	-5.990	-6.023	-5.980
τ_{64}	-2.413	-2.527	-2.672	-3.570	-3.612	-3.703	-5.379	-5.401	-5.365

τ_{65}	-1.868	-1.958	-2.112	-3.087	-3.087	-3.231	-4.923	-4.848	-4.869
τ_{72}	-4.428	-4.559	-4.466	-5.676	-5.629	-5.646	-7.222	-7.188	-7.238
τ_{73}	-3.617	-3.643	-3.595	-5.180	-5.093	-5.209	-6.789	-6.709	-6.759
τ_{74}	-3.083	-3.030	-3.019	-4.377	-4.285	-4.431	-6.188	-6.094	-6.101
τ_{75}	-2.467	-2.591	-2.542	-3.992	-3.927	-4.069	-5.859	-5.771	-5.776
τ_{82}	-3.873	-3.927	-3.925	-5.248	-5.368	-5.374	-7.035	-7.016	-6.984
τ_{83}	-3.089	-3.225	-3.300	-4.351	-4.510	-4.610	-6.684	-6.649	-6.615
τ_{84}	-2.555	-2.741	-2.850	-3.721	-3.923	-3.992	-5.887	-5.836	-5.843
τ_{85}	-1.955	-1.970	-2.133	-3.146	-3.357	-3.460	-5.442	-5.404	-5.451
	Coverage								
λ_{10}	96	96	97	97	97	95	95	96	96
λ_{20}	99	99	99	94	95	96	97	97	97
λ_{30}	92	91	95	95	94	93	99	99	100
λ_{40}	91	90	91	93	93	95	97	95	98
λ_{50}	91	90	95	90	90	93	90	94	91
λ_{60}	94	94	93	92	90	92	94	92	94
λ_{70}	91	90	90	91	90	89	95	95	94
λ_{80}	89	89	88	92	92	96	94	94	95
λ_{90}	91	91	92	97	97	97	99	98	99
$\lambda_{10,0}$	95	97	96	96	97	96	94	95	95
a_{11}	97	97	98	96	97	97	94	97	96
a_{12}	92	92	95	94	94	96	95	95	97
λ_{11}	93	97	91	92	93	95	92	93	93
λ_{12}	–	–	–	–	–	–	–	–	–
a_{21}	92	92	95	91	93	93	93	93	93
a_{22}	90	91	91	95	96	96	95	96	96
λ_{21}	97	97	96	93	92	95	89	89	91
λ_{22}	92	94	94	94	96	96	92	93	92
a_{31}	94	96	97	92	92	94	92	93	93
a_{32}	92	91	92	95	94	95	93	94	93
λ_{31}	95	95	93	98	98	93	95	97	97
λ_{32}	91	92	94	95	96	96	92	93	93
a_{41}	92	93	93	95	94	95	95	96	95
a_{42}	91	92	93	95	95	94	93	92	93
λ_{41}	96	96	93	97	97	98	94	94	97
λ_{42}	93	92	93	98	98	99	96	94	95
a_{51}	93	93	91	97	96	98	94	93	92
a_{52}	94	94	91	98	96	95	96	96	96
λ_{51}	93	94	95	96	94	97	92	90	90
λ_{52}	90	91	91	92	94	92	93	93	92
a_{61}	95	93	93	93	94	97	93	94	96
a_{62}	98	94	97	90	91	91	93	92	95
λ_{61}	87	90	94	87	90	90	87	87	90
λ_{62}	94	93	97	95	94	96	91	89	90
a_{71}	93	93	94	88	90	91	95	97	96
a_{72}	95	98	97	93	94	92	93	92	94
λ_{71}	91	93	95	95	95	97	92	91	92
λ_{72}	96	98	96	93	94	94	91	92	91
a_{81}	92	91	93	90	89	93	95	94	95
a_{82}	97	93	94	93	96	95	95	94	95
λ_{81}	97	97	97	94	97	95	91	87	90
λ_{82}	87	90	91	89	91	92	98	96	97
a_{91}	99	100	99	97	98	97	96	96	96
a_{92}	96	96	95	97	98	97	91	91	92
λ_{91}	94	93	92	96	97	95	95	96	96
λ_{92}	96	96	99	97	96	97	90	90	90
$a_{10,1}$	92	94	92	97	97	95	95	95	96
$a_{10,2}$	99	97	99	97	95	96	97	97	96
$\lambda_{10,1}$	95	96	93	96	94	95	89	87	88
$\lambda_{10,2}$	95	93	92	93	92	89	93	93	94
τ_{32}	90	93	98	95	95	94	98	97	97
τ_{42}	88	87	91	94	96	96	94	95	94
τ_{52}	91	90	94	92	92	93	89	91	88
τ_{62}	85	88	90	92	93	94	94	95	95
τ_{63}	80	80	84	87	84	90	87	92	86

τ_{64}	78	76	80	83	83	88	90	89	91
τ_{65}	77	82	75	81	81	87	90	91	92
τ_{72}	90	94	93	90	88	91	95	90	88
τ_{73}	86	86	87	91	93	89	95	93	95
τ_{74}	87	83	78	88	87	86	94	94	92
τ_{75}	84	80	79	89	90	92	95	94	93
τ_{82}	85	84	83	91	94	95	91	92	89
τ_{83}	79	83	86	82	90	89	98	98	95
τ_{84}	78	79	87	78	85	87	93	94	93
τ_{85}	76	76	75	85	84	88	94	95	94

Table C.8: *Estimates of parameter set E simulated by the MHS – MEAN, STD, BIAS, ln(MSE), and coverage of estimated parameters obtained by simulations of 100 different data sets.*

Simulations of parameter set E − GGS sampler									
N	$N_1 = 300$			$N_2 = 1000$			$N_3 = 5000$		
Priori	Weak	Std.	Strong	Weak	Std.	Strong	Weak	Std.	Strong
	MEAN								
λ_{10}	-16.984	-9.537	-0.546	-0.515	-0.514	-0.513	-0.500	-0.501	-0.500
λ_{20}	5.151	1.875	0.549	0.521	0.521	0.519	0.502	0.503	0.502
λ_{30}	1.515	1.252	1.267	1.173	1.174	1.176	1.171	1.172	1.172
λ_{40}	1.046	1.050	1.060	1.011	1.012	1.013	0.997	0.997	0.997
λ_{50}	1.266	1.269	1.275	1.181	1.182	1.185	1.172	1.172	1.172
λ_{60}	0.868	0.874	0.893	0.853	0.854	0.856	0.838	0.838	0.839
λ_{70}	0.744	0.746	0.759	0.713	0.714	0.716	0.710	0.711	0.711
λ_{80}	0.900	0.906	0.922	0.850	0.851	0.856	0.841	0.840	0.841
λ_{90}	2.006	2.006	2.008	1.990	1.990	1.986	2.004	2.004	2.004
$\lambda_{10,0}$	-4.940	-4.938	-4.929	-5.006	-5.005	-5.001	-4.992	-4.992	-4.992
a_{11}	36.664	18.686	0.931	0.860	0.860	0.855	0.846	0.847	0.846
a_{12}	-31.226	-14.531	-0.920	-0.868	-0.867	-0.862	-0.837	-0.838	-0.837
λ_{11}	8.257	4.093	1.384	1.401	1.395	1.355	1.344	1.344	1.336
λ_{12}	–	–	–	–	–	–	–	–	–
a_{21}	-10.002	-3.401	-0.934	-0.864	-0.862	-0.858	-0.838	-0.839	-0.838
a_{22}	11.839	4.248	0.925	0.874	0.874	0.870	0.835	0.835	0.834
λ_{21}	0.020	-0.031	-0.013	-0.009	-0.010	-0.013	-0.001	-0.002	-0.003
λ_{22}	3.752	2.147	1.360	1.397	1.391	1.354	1.335	1.334	1.327
a_{31}	1.133	0.913	0.907	0.847	0.848	0.847	0.840	0.840	0.840
a_{32}	-1.041	-0.902	-0.896	-0.852	-0.852	-0.851	-0.836	-0.836	-0.836
λ_{31}	1.591	1.427	1.361	1.357	1.353	1.329	1.336	1.335	1.331
λ_{32}	-0.032	-0.014	-0.021	0.001	0.002	0.001	-0.003	-0.003	-0.003
a_{41}	-0.748	-0.751	-0.751	-0.722	-0.721	-0.720	-0.711	-0.711	-0.711
a_{42}	0.751	0.753	0.754	0.720	0.720	0.720	0.710	0.709	0.709
λ_{41}	0.726	0.723	0.695	0.724	0.722	0.709	0.707	0.707	0.704
λ_{42}	0.746	0.743	0.706	0.719	0.717	0.706	0.705	0.705	0.703
a_{51}	0.849	0.846	0.836	0.847	0.847	0.843	0.838	0.838	0.837
a_{52}	-0.867	-0.865	-0.853	-0.838	-0.837	-0.833	-0.835	-0.835	-0.834
λ_{51}	0.016	0.014	0.010	-0.002	-0.003	-0.006	0.003	0.003	0.001
λ_{52}	1.422	1.409	1.329	1.351	1.346	1.322	1.337	1.337	1.332
a_{61}	-0.874	-0.876	-0.874	-0.843	-0.842	-0.838	-0.836	-0.836	-0.835
a_{62}	0.861	0.860	0.859	0.858	0.858	0.854	0.841	0.841	0.840
λ_{61}	1.405	1.400	1.349	1.376	1.372	1.350	1.346	1.345	1.341
λ_{62}	-0.012	-0.011	-0.018	0.005	0.005	0.005	-0.007	-0.007	-0.006
a_{71}	0.722	0.721	0.718	0.713	0.714	0.713	0.711	0.711	0.710
a_{72}	-0.718	-0.719	-0.715	-0.712	-0.711	-0.711	-0.706	-0.706	-0.705
λ_{71}	0.747	0.744	0.720	0.720	0.718	0.707	0.710	0.710	0.707
λ_{72}	0.723	0.720	0.687	0.721	0.719	0.709	0.704	0.704	0.702
a_{81}	-0.910	-0.913	-0.907	-0.857	-0.856	-0.856	-0.836	-0.836	-0.836
a_{82}	0.898	0.899	0.895	0.855	0.855	0.854	0.836	0.836	0.836
λ_{81}	0.018	0.015	0.010	0.002	0.001	-0.002	-0.002	-0.002	-0.004
λ_{82}	1.426	1.419	1.348	1.360	1.356	1.334	1.333	1.332	1.328
a_{91}	0.527	0.524	0.522	0.505	0.507	0.511	0.509	0.509	0.509
a_{92}	-0.504	-0.506	-0.503	-0.514	-0.513	-0.516	-0.497	-0.497	-0.498
λ_{91}	2.900	2.902	2.876	3.008	3.005	2.983	3.002	3.001	2.997
λ_{92}	-0.017	-0.016	-0.034	-0.001	-0.001	-0.003	-0.005	-0.005	-0.004
$a_{10,1}$	-0.567	-0.572	-0.578	-0.503	-0.501	-0.506	-0.509	-0.509	-0.510
$a_{10,2}$	0.511	0.514	0.524	0.514	0.515	0.520	0.502	0.502	0.503
$\lambda_{10,1}$	0.047	0.038	0.020	-0.000	-0.004	-0.014	0.003	0.002	-0.003
$\lambda_{10,2}$	3.952	3.947	3.867	3.982	3.976	3.947	3.999	3.999	3.993
τ_{32}	3.029	2.519	2.542	2.358	2.359	2.366	2.344	2.344	2.346
τ_{42}	2.078	2.083	2.097	2.012	2.013	2.016	1.985	1.985	1.985
τ_{52}	2.492	2.494	2.501	2.369	2.370	2.373	2.343	2.343	2.344
τ_{62}	0.856	0.862	0.877	0.851	0.851	0.853	0.843	0.843	0.843
τ_{63}	1.914	1.922	1.943	1.874	1.875	1.879	1.846	1.845	1.847
τ_{64}	2.952	2.962	2.983	2.900	2.899	2.901	2.847	2.846	2.847
τ_{65}	4.004	4.015	4.028	3.919	3.917	3.916	3.849	3.850	3.849
τ_{72}	0.744	0.744	0.754	0.720	0.720	0.721	0.709	0.711	0.711
τ_{73}	1.612	1.614	1.630	1.583	1.585	1.588	1.560	1.562	1.562

τ_{74}	2.478	2.482	2.496	2.435	2.436	2.439	2.413	2.414	2.416
τ_{75}	3.357	3.365	3.371	3.295	3.293	3.293	3.261	3.261	3.261
τ_{82}	0.887	0.893	0.903	0.854	0.855	0.859	0.836	0.835	0.836
τ_{83}	1.920	1.927	1.939	1.868	1.867	1.874	1.839	1.837	1.836
τ_{84}	2.962	2.969	2.973	2.883	2.883	2.885	2.844	2.843	2.844
τ_{85}	4.033	4.039	4.029	3.913	3.914	3.908	3.845	3.845	3.845
					STD				
λ_{10}	94.998	61.522	0.158	0.075	0.076	0.076	0.030	0.030	0.030
λ_{20}	44.434	13.175	0.130	0.074	0.075	0.074	0.032	0.032	0.032
λ_{30}	2.751	0.195	0.184	0.083	0.082	0.082	0.034	0.034	0.034
λ_{40}	0.148	0.150	0.146	0.075	0.075	0.075	0.028	0.028	0.027
λ_{50}	0.183	0.183	0.181	0.096	0.096	0.095	0.039	0.039	0.040
λ_{60}	0.140	0.140	0.134	0.082	0.082	0.084	0.038	0.040	0.038
λ_{70}	0.113	0.115	0.112	0.066	0.067	0.067	0.028	0.027	0.027
λ_{80}	0.159	0.160	0.158	0.078	0.079	0.079	0.033	0.033	0.033
λ_{90}	0.234	0.238	0.236	0.108	0.109	0.108	0.045	0.046	0.045
$\lambda_{10,0}$	0.234	0.239	0.237	0.124	0.125	0.126	0.057	0.057	0.057
a_{11}	204.976	123.531	0.206	0.090	0.089	0.088	0.038	0.038	0.038
a_{12}	181.084	88.485	0.220	0.086	0.086	0.084	0.042	0.042	0.042
λ_{11}	35.295	16.251	0.257	0.140	0.138	0.129	0.055	0.055	0.055
λ_{12}	–	–	–	–	–	–	–	–	–
a_{21}	87.015	24.474	0.194	0.097	0.098	0.096	0.040	0.041	0.041
a_{22}	104.415	33.073	0.222	0.089	0.089	0.086	0.036	0.036	0.036
λ_{21}	0.676	0.242	0.162	0.099	0.095	0.093	0.049	0.050	0.050
λ_{22}	19.681	6.536	0.222	0.123	0.121	0.115	0.052	0.052	0.052
a_{31}	2.258	0.152	0.147	0.078	0.078	0.078	0.032	0.032	0.033
a_{32}	1.433	0.149	0.143	0.081	0.081	0.081	0.033	0.033	0.033
λ_{31}	1.657	0.266	0.175	0.094	0.093	0.089	0.042	0.041	0.041
λ_{32}	0.233	0.151	0.148	0.067	0.065	0.064	0.037	0.037	0.037
a_{41}	0.121	0.120	0.114	0.058	0.058	0.058	0.026	0.025	0.025
a_{42}	0.117	0.120	0.116	0.062	0.062	0.062	0.027	0.027	0.026
λ_{41}	0.153	0.153	0.122	0.069	0.069	0.067	0.032	0.033	0.032
λ_{42}	0.140	0.129	0.119	0.057	0.055	0.055	0.026	0.026	0.027
a_{51}	0.164	0.160	0.154	0.069	0.069	0.069	0.035	0.035	0.035
a_{52}	0.162	0.160	0.154	0.062	0.063	0.061	0.029	0.029	0.029
λ_{51}	0.188	0.183	0.175	0.085	0.082	0.080	0.043	0.043	0.042
λ_{52}	0.224	0.206	0.172	0.099	0.098	0.095	0.042	0.041	0.041
a_{61}	0.146	0.141	0.125	0.067	0.068	0.068	0.032	0.032	0.032
a_{62}	0.132	0.132	0.120	0.071	0.071	0.071	0.033	0.033	0.033
λ_{61}	0.251	0.233	0.151	0.091	0.091	0.088	0.045	0.045	0.045
λ_{62}	0.143	0.140	0.136	0.072	0.069	0.069	0.035	0.035	0.034
a_{71}	0.105	0.102	0.103	0.062	0.062	0.061	0.025	0.025	0.025
a_{72}	0.098	0.097	0.094	0.057	0.057	0.057	0.025	0.025	0.025
λ_{71}	0.154	0.143	0.113	0.066	0.064	0.063	0.030	0.030	0.030
λ_{72}	0.114	0.106	0.096	0.063	0.061	0.061	0.027	0.028	0.027
a_{81}	0.122	0.124	0.123	0.074	0.075	0.074	0.029	0.029	0.029
a_{82}	0.114	0.113	0.110	0.059	0.059	0.058	0.031	0.030	0.031
λ_{81}	0.148	0.145	0.138	0.084	0.081	0.079	0.045	0.045	0.045
λ_{82}	0.193	0.174	0.148	0.089	0.088	0.086	0.032	0.032	0.031
a_{91}	0.205	0.199	0.186	0.112	0.111	0.113	0.052	0.053	0.053
a_{92}	0.209	0.206	0.198	0.103	0.103	0.104	0.055	0.055	0.055
λ_{91}	0.407	0.365	0.179	0.101	0.101	0.100	0.047	0.047	0.047
λ_{92}	0.258	0.250	0.248	0.146	0.142	0.140	0.077	0.078	0.077
$a_{10,1}$	0.254	0.250	0.251	0.130	0.131	0.131	0.065	0.064	0.065
$a_{10,2}$	0.247	0.239	0.240	0.129	0.129	0.128	0.061	0.059	0.060
$\lambda_{10,1}$	0.391	0.383	0.384	0.240	0.231	0.227	0.120	0.123	0.120
$\lambda_{10,2}$	0.344	0.274	0.191	0.122	0.123	0.122	0.055	0.055	0.054
τ_{32}	5.340	0.297	0.269	0.129	0.128	0.129	0.053	0.053	0.053
τ_{42}	0.196	0.194	0.190	0.094	0.094	0.095	0.043	0.043	0.043
τ_{52}	0.278	0.272	0.264	0.134	0.134	0.133	0.062	0.062	0.063
τ_{62}	0.134	0.131	0.131	0.065	0.063	0.065	0.029	0.031	0.028
τ_{63}	0.214	0.210	0.194	0.112	0.113	0.114	0.050	0.050	0.048
τ_{64}	0.296	0.289	0.263	0.151	0.153	0.154	0.066	0.067	0.066
τ_{65}	0.376	0.371	0.334	0.193	0.194	0.191	0.087	0.086	0.088
τ_{72}	0.100	0.103	0.104	0.058	0.059	0.059	0.030	0.030	0.028

τ_{73}	0.152	0.151	0.152	0.073	0.073	0.074	0.035	0.036	0.034
τ_{74}	0.205	0.206	0.199	0.108	0.106	0.107	0.046	0.045	0.046
τ_{75}	0.266	0.267	0.257	0.127	0.128	0.127	0.054	0.055	0.054
τ_{82}	0.145	0.146	0.143	0.070	0.070	0.072	0.034	0.033	0.031
τ_{83}	0.212	0.208	0.204	0.101	0.103	0.105	0.040	0.038	0.038
τ_{84}	0.261	0.259	0.256	0.137	0.136	0.135	0.055	0.053	0.054
τ_{85}	0.362	0.359	0.347	0.172	0.172	0.170	0.064	0.066	0.065
	BIAS								
λ_{10}	-16.484	-9.037	-0.046	-0.015	-0.014	-0.013	-0.000	-0.001	0.000
λ_{20}	4.651	1.375	0.049	0.021	0.021	0.019	0.002	0.003	0.002
λ_{30}	0.349	0.086	0.100	0.007	0.008	0.009	0.005	0.005	0.005
λ_{40}	0.056	0.060	0.070	0.021	0.022	0.023	0.007	0.007	0.007
λ_{50}	0.100	0.102	0.108	0.014	0.015	0.018	0.005	0.005	0.006
λ_{60}	0.034	0.041	0.060	0.020	0.021	0.023	0.005	0.004	0.005
λ_{70}	0.037	0.039	0.051	0.006	0.007	0.009	0.003	0.004	0.004
λ_{80}	0.067	0.073	0.089	0.017	0.017	0.023	0.007	0.007	0.008
λ_{90}	0.006	0.006	0.008	-0.010	-0.010	-0.014	0.004	0.004	0.004
$\lambda_{10,0}$	0.060	0.062	0.071	-0.006	-0.005	-0.001	0.008	0.008	0.008
a_{11}	35.831	17.853	0.098	0.026	0.027	0.022	0.013	0.013	0.012
a_{12}	-30.393	-13.698	-0.087	-0.035	-0.034	-0.028	-0.004	-0.004	-0.003
λ_{11}	6.923	2.759	0.051	0.068	0.062	0.022	0.011	0.010	0.003
λ_{12}	–	–	–	–	–	–	–	–	–
a_{21}	-9.168	-2.568	-0.101	-0.030	-0.029	-0.025	-0.005	-0.005	-0.004
a_{22}	11.005	3.415	0.092	0.041	0.040	0.037	0.002	0.002	0.001
λ_{21}	0.020	-0.031	-0.013	-0.009	-0.010	-0.013	-0.001	-0.002	-0.003
λ_{22}	2.419	0.813	0.027	0.064	0.057	0.021	0.001	0.001	-0.006
a_{31}	0.300	0.080	0.074	0.014	0.014	0.014	0.007	0.007	0.007
a_{32}	-0.208	-0.069	-0.063	-0.019	-0.018	-0.018	-0.002	-0.002	-0.002
λ_{31}	0.258	0.094	0.028	0.024	0.020	-0.004	0.003	0.002	-0.002
λ_{32}	-0.032	-0.014	-0.021	0.001	0.002	0.001	-0.003	-0.003	-0.003
a_{41}	-0.041	-0.044	-0.044	-0.015	-0.014	-0.013	-0.004	-0.004	-0.004
a_{42}	0.044	0.046	0.047	0.013	0.013	0.012	0.002	0.002	0.002
λ_{41}	0.019	0.016	-0.012	0.017	0.015	0.002	0.000	-0.000	-0.003
λ_{42}	0.039	0.036	-0.001	0.012	0.010	-0.001	-0.002	-0.002	-0.004
a_{51}	0.016	0.012	0.003	0.013	0.014	0.010	0.005	0.004	0.004
a_{52}	-0.034	-0.032	-0.020	-0.005	-0.004	0.000	-0.002	-0.002	-0.001
λ_{51}	0.016	0.014	0.010	-0.002	-0.003	-0.006	0.003	0.003	0.001
λ_{52}	0.089	0.076	-0.005	0.017	0.013	-0.011	0.004	0.003	-0.002
a_{61}	-0.041	-0.043	-0.041	-0.010	-0.009	-0.005	-0.003	-0.002	-0.002
a_{62}	0.027	0.027	0.026	0.024	0.024	0.021	0.007	0.007	0.007
λ_{61}	0.072	0.067	0.016	0.043	0.039	0.017	0.012	0.012	0.007
λ_{62}	-0.012	-0.011	-0.018	0.005	0.005	0.005	-0.007	-0.007	-0.006
a_{71}	0.015	0.014	0.011	0.006	0.007	0.006	0.004	0.003	0.003
a_{72}	-0.011	-0.012	-0.008	-0.005	-0.004	-0.004	0.002	0.001	0.002
λ_{71}	0.039	0.037	0.013	0.013	0.011	-0.001	0.003	0.003	-0.000
λ_{72}	0.016	0.013	-0.020	0.014	0.012	0.002	-0.003	-0.003	-0.005
a_{81}	-0.077	-0.080	-0.074	-0.024	-0.023	-0.022	-0.002	-0.002	-0.002
a_{82}	0.065	0.066	0.062	0.022	0.022	0.021	0.002	0.002	0.002
λ_{81}	0.018	0.015	0.010	0.002	0.001	-0.002	-0.002	-0.002	-0.004
λ_{82}	0.093	0.085	0.014	0.027	0.023	0.001	-0.001	-0.002	-0.006
a_{91}	0.027	0.024	0.022	0.005	0.007	0.011	0.009	0.009	0.009
a_{92}	-0.004	-0.006	-0.003	-0.014	-0.013	-0.016	0.003	0.003	0.002
λ_{91}	-0.100	-0.098	-0.124	0.008	0.005	-0.017	0.002	0.001	-0.003
λ_{92}	-0.017	-0.016	-0.034	-0.001	-0.001	-0.003	-0.005	-0.005	-0.004
$a_{10,1}$	-0.067	-0.072	-0.078	-0.003	-0.001	-0.006	-0.009	-0.009	-0.010
$a_{10,2}$	0.011	0.014	0.024	0.014	0.015	0.020	0.002	0.002	0.003
$\lambda_{10,1}$	0.047	0.038	0.020	-0.000	-0.004	-0.014	0.003	0.002	-0.003
$\lambda_{10,2}$	-0.048	-0.053	-0.133	-0.018	-0.024	-0.053	-0.001	-0.001	-0.007
τ_{32}	0.696	0.185	0.209	0.024	0.026	0.032	0.011	0.011	0.012
τ_{42}	0.098	0.103	0.118	0.032	0.033	0.036	0.005	0.005	0.005
τ_{52}	0.159	0.161	0.167	0.035	0.037	0.040	0.010	0.010	0.011
τ_{62}	0.023	0.029	0.044	0.018	0.018	0.020	0.010	0.009	0.010
τ_{63}	0.080	0.089	0.110	0.040	0.042	0.046	0.013	0.012	0.013
τ_{64}	0.119	0.129	0.149	0.066	0.066	0.067	0.014	0.013	0.014
τ_{65}	0.171	0.182	0.195	0.085	0.084	0.083	0.016	0.017	0.016

τ_{72}	0.037	0.037	0.047	0.013	0.013	0.014	0.002	0.004	0.003
τ_{73}	0.057	0.058	0.074	0.027	0.029	0.032	0.005	0.006	0.007
τ_{74}	0.074	0.078	0.092	0.031	0.032	0.035	0.009	0.010	0.011
τ_{75}	0.105	0.112	0.119	0.042	0.040	0.040	0.009	0.008	0.008
τ_{82}	0.054	0.059	0.070	0.021	0.022	0.026	0.003	0.002	0.002
τ_{83}	0.087	0.094	0.106	0.035	0.034	0.041	0.005	0.004	0.003
τ_{84}	0.128	0.136	0.140	0.050	0.050	0.051	0.010	0.009	0.010
τ_{85}	0.200	0.205	0.195	0.080	0.080	0.075	0.011	0.012	0.012
	ln(MSE)								
λ_{10}	9.128	8.250	-3.613	-5.137	-5.122	-5.138	-7.006	-7.007	-7.021
λ_{20}	7.589	5.158	-3.961	-5.129	-5.117	-5.146	-6.899	-6.883	-6.889
λ_{30}	2.030	-3.102	-3.130	-4.980	-5.000	-5.009	-6.747	-6.758	-6.758
λ_{40}	-3.699	-3.653	-3.651	-5.120	-5.101	-5.105	-7.106	-7.118	-7.133
λ_{50}	-3.143	-3.131	-3.119	-4.684	-4.669	-4.675	-6.458	-6.462	-6.445
λ_{60}	-3.885	-3.854	-3.853	-4.962	-4.940	-4.903	-6.514	-6.457	-6.538
λ_{70}	-4.269	-4.222	-4.188	-5.426	-5.415	-5.393	-7.184	-7.246	-7.224
λ_{80}	-3.528	-3.485	-3.423	-5.065	-5.051	-5.003	-6.756	-6.798	-6.794
λ_{90}	-2.913	-2.879	-2.896	-4.447	-4.430	-4.448	-6.188	-6.179	-6.211
$\lambda_{10,0}$	-2.848	-2.810	-2.804	-4.190	-4.162	-4.158	-5.725	-5.726	-5.716
a_{11}	10.666	9.644	-2.965	-4.752	-4.772	-4.816	-6.454	-6.456	-6.460
a_{12}	10.416	8.980	-2.892	-4.754	-4.777	-4.859	-6.350	-6.337	-6.346
λ_{11}	7.156	5.595	-2.687	-3.730	-3.786	-4.083	-5.768	-5.774	-5.819
λ_{12}	–	–	–	–	–	–	–	–	–
a_{21}	8.933	6.396	-3.049	-4.582	-4.582	-4.631	-6.422	-6.394	-6.398
a_{22}	9.298	6.998	-2.859	-4.662	-4.661	-4.741	-6.638	-6.650	-6.640
λ_{21}	-0.791	-2.830	-3.642	-4.631	-4.698	-4.731	-6.028	-5.995	-6.015
λ_{22}	5.964	3.760	-3.003	-3.958	-4.025	-4.310	-5.910	-5.906	-5.916
a_{31}	1.636	-3.536	-3.622	-5.072	-5.078	-5.073	-6.828	-6.829	-6.813
a_{32}	0.730	-3.627	-3.726	-4.980	-4.984	-4.991	-6.840	-6.846	-6.839
λ_{31}	1.024	-2.542	-3.476	-4.683	-4.713	-4.838	-6.366	-6.378	-6.388
λ_{32}	-2.907	-3.779	-3.815	-5.418	-5.478	-5.500	-6.584	-6.600	-6.611
a_{41}	-4.119	-4.126	-4.211	-5.631	-5.660	-5.659	-7.315	-7.328	-7.338
a_{42}	-4.169	-4.113	-4.162	-5.530	-5.516	-5.529	-7.242	-7.244	-7.265
λ_{41}	-3.751	-3.756	-4.202	-5.291	-5.326	-5.412	-6.889	-6.853	-6.871
λ_{42}	-3.874	-4.027	-4.275	-5.697	-5.783	-5.822	-7.288	-7.288	-7.234
a_{51}	-3.620	-3.668	-3.751	-5.311	-5.311	-5.346	-6.688	-6.683	-6.695
a_{52}	-3.607	-3.640	-3.735	-5.557	-5.548	-5.592	-7.086	-7.098	-7.081
λ_{51}	-3.344	-3.397	-3.490	-4.939	-5.008	-5.050	-6.317	-6.291	-6.332
λ_{52}	-2.857	-3.039	-3.530	-4.610	-4.636	-4.700	-6.364	-6.376	-6.375
a_{61}	-3.786	-3.844	-4.061	-5.380	-5.365	-5.394	-6.876	-6.890	-6.880
a_{62}	-4.023	-4.015	-4.211	-5.185	-5.184	-5.216	-6.790	-6.792	-6.797
λ_{61}	-2.695	-2.843	-3.774	-4.602	-4.625	-4.829	-6.145	-6.156	-6.203
λ_{62}	-3.888	-3.932	-3.983	-5.271	-5.341	-5.354	-6.694	-6.669	-6.715
a_{71}	-4.498	-4.560	-4.554	-5.573	-5.575	-5.577	-7.400	-7.389	-7.396
a_{72}	-4.636	-4.671	-4.721	-5.715	-5.721	-5.730	-7.356	-7.348	-7.356
λ_{71}	-3.688	-3.832	-4.356	-5.408	-5.474	-5.539	-7.024	-7.019	-7.034
λ_{72}	-4.336	-4.492	-4.651	-5.502	-5.562	-5.598	-7.193	-7.180	-7.172
a_{81}	-3.875	-3.840	-3.896	-5.116	-5.099	-5.130	-7.067	-7.074	-7.063
a_{82}	-4.069	-4.079	-4.148	-5.549	-5.550	-5.591	-6.970	-6.984	-6.980
λ_{81}	-3.812	-3.855	-3.962	-4.955	-5.047	-5.074	-6.227	-6.190	-6.211
λ_{82}	-3.087	-3.292	-3.826	-4.769	-4.803	-4.921	-6.908	-6.911	-6.907
a_{91}	-3.165	-3.229	-3.363	-4.393	-4.403	-4.369	-5.878	-5.869	-5.866
a_{92}	-3.141	-3.167	-3.248	-4.535	-4.537	-4.518	-5.807	-5.796	-5.798
λ_{91}	-1.750	-1.955	-3.058	-4.585	-4.587	-4.586	-6.132	-6.127	-6.108
λ_{92}	-2.716	-2.781	-2.776	-3.854	-3.918	-3.938	-5.140	-5.111	-5.145
$a_{10,1}$	-2.685	-2.699	-2.682	-4.090	-4.072	-4.080	-5.454	-5.475	-5.439
$a_{10,2}$	-2.806	-2.867	-2.856	-4.099	-4.099	-4.099	-5.607	-5.662	-5.632
$\lambda_{10,1}$	-1.875	-1.919	-1.921	-2.861	-2.940	-2.969	-4.254	-4.208	-4.245
$\lambda_{10,2}$	-2.125	-2.561	-2.922	-4.193	-4.169	-4.048	-5.825	-5.821	-5.821
τ_{32}	3.357	-2.108	-2.163	-4.078	-4.080	-4.049	-5.846	-5.839	-5.823
τ_{42}	-3.045	-3.035	-3.004	-4.631	-4.619	-4.588	-6.306	-6.296	-6.293
τ_{52}	-2.288	-2.313	-2.331	-3.968	-3.954	-3.956	-5.539	-5.539	-5.521
τ_{62}	-3.994	-4.021	-3.967	-5.396	-5.460	-5.396	-6.995	-6.883	-7.047
τ_{63}	-2.961	-2.967	-3.007	-4.269	-4.241	-4.199	-5.929	-5.953	-6.008
τ_{64}	-2.296	-2.312	-2.401	-3.613	-3.591	-3.578	-5.410	-5.386	-5.413

τ_{65}	-1.779	-1.777	-1.907	-3.118	-3.118	-3.149	-4.867	-4.886	-4.847
τ_{72}	-4.487	-4.429	-4.347	-5.644	-5.631	-5.604	-7.053	-7.004	-7.138
τ_{73}	-3.644	-3.655	-3.564	-5.123	-5.098	-5.043	-6.689	-6.623	-6.733
τ_{74}	-3.056	-3.040	-3.041	-4.377	-4.405	-4.379	-6.139	-6.177	-6.127
τ_{75}	-2.513	-2.490	-2.530	-4.038	-4.030	-4.046	-5.826	-5.804	-5.809
τ_{82}	-3.744	-3.699	-3.680	-5.238	-5.225	-5.154	-6.759	-6.831	-6.946
τ_{83}	-2.957	-2.960	-2.949	-4.486	-4.443	-4.370	-6.438	-6.555	-6.560
τ_{84}	-2.476	-2.465	-2.472	-3.865	-3.879	-3.880	-5.782	-5.846	-5.815
τ_{85}	-1.774	-1.773	-1.847	-3.332	-3.328	-3.372	-5.464	-5.403	-5.428
	Coverage								
λ_{10}	92	91	95	96	96	97	96	95	95
λ_{20}	98	97	99	94	95	96	97	97	97
λ_{30}	91	89	89	95	96	96	99	98	100
λ_{40}	93	90	93	94	93	94	96	96	97
λ_{50}	91	90	90	91	91	91	92	90	91
λ_{60}	95	97	97	94	93	93	94	91	92
λ_{70}	93	94	92	89	89	90	95	96	93
λ_{80}	91	94	93	93	94	93	95	95	95
λ_{90}	92	91	91	96	97	97	99	99	99
$\lambda_{10,0}$	96	96	95	98	96	96	95	95	95
a_{11}	91	93	97	97	97	96	97	96	96
a_{12}	87	86	89	94	92	93	94	95	94
λ_{11}	88	87	95	90	93	95	92	93	94
λ_{12}	–	–	–	–	–	–	–	–	–
a_{21}	89	87	90	92	92	92	92	94	92
a_{22}	90	90	90	96	96	96	94	96	94
λ_{21}	95	96	97	92	94	94	89	89	90
λ_{22}	89	91	94	95	95	96	92	93	95
a_{31}	94	94	95	91	91	90	92	93	92
a_{32}	93	93	92	96	96	95	95	95	95
λ_{31}	88	91	98	96	97	96	97	96	96
λ_{32}	91	92	89	96	96	95	92	94	95
a_{41}	91	91	93	95	94	95	94	95	95
a_{42}	93	91	91	94	95	95	92	92	92
λ_{41}	93	94	93	95	97	98	95	95	96
λ_{42}	91	92	92	96	99	99	96	95	95
a_{51}	92	91	91	96	97	98	93	93	93
a_{52}	91	92	90	97	97	97	96	96	96
λ_{51}	92	92	93	97	96	95	90	91	91
λ_{52}	88	91	95	92	92	91	94	93	93
a_{61}	90	91	94	96	94	94	94	96	94
a_{62}	96	96	97	90	90	90	92	93	92
λ_{61}	89	89	97	91	90	95	92	88	90
λ_{62}	88	94	92	93	94	96	90	87	90
a_{71}	93	95	94	90	88	89	96	97	96
a_{72}	95	97	97	93	93	93	94	95	93
λ_{71}	91	92	93	97	98	97	92	92	93
λ_{72}	96	98	96	95	94	94	94	93	91
a_{81}	90	91	90	90	91	90	94	93	95
a_{82}	95	95	96	94	94	95	96	96	96
λ_{81}	94	96	96	97	96	96	89	89	88
λ_{82}	89	90	93	91	90	92	97	97	97
a_{91}	96	99	98	97	97	97	96	96	96
a_{92}	94	95	96	97	98	98	91	91	91
λ_{91}	91	92	89	97	97	95	94	95	95
λ_{92}	94	94	95	97	97	97	90	90	90
$a_{10,1}$	93	94	92	97	96	97	94	95	95
$a_{10,2}$	98	99	97	97	95	96	97	97	94
$\lambda_{10,1}$	95	96	95	94	96	96	85	86	86
$\lambda_{10,2}$	94	94	91	92	92	87	93	92	92
τ_{32}	89	91	91	95	95	95	98	97	98
τ_{42}	88	87	90	96	95	94	94	94	93
τ_{52}	89	89	90	90	91	91	89	90	89
τ_{62}	91	92	92	95	94	96	90	86	89
τ_{63}	91	90	90	92	87	88	90	89	89

τ_{64}	93	92	94	89	89	88	90	89	90
τ_{65}	90	92	93	88	89	89	91	89	90
τ_{72}	95	91	91	88	90	87	81	79	85
τ_{73}	89	91	87	93	91	91	90	91	96
τ_{74}	92	92	93	91	91	92	94	94	91
τ_{75}	89	88	91	94	94	94	95	94	93
τ_{82}	89	88	88	91	92	90	85	87	88
τ_{83}	90	91	93	88	87	87	95	95	94
τ_{84}	90	89	90	92	92	92	94	94	93
τ_{85}	90	90	91	91	90	91	98	97	97

Table C.9: *Estimates of parameter set E simulated by the GGS – MEAN, STD, BIAS, ln(MSE), and coverage of estimated parameters obtained by simulations of 100 different data sets.*

Appendix D

Model estimates of PD1 data

In this chapter, the estimates of intercepts and cutpoints of "PD1" analyses are presented which were not included in the main text. Table D.1 summarizes the models calculated in Sections 7.3 (one latent variable) and 7.4 (two latent variables).

Section	Model	Predictor of indirect effects	Results in
7.3.1	M1	$\eta = 0$	D.2
7.3.2	M2a	$\eta = Sex + Inc + Age$	D.3
	M2b	$\eta = Sex + Inc + Age + Sex * Inc +$ $Sex * Age + Inc * Age$	D.4
7.3.3	M3a	$\eta = f(Age)$	D.5, D.6, D.7
	M3b	$\eta = Sex + Inc + f(Age)$	D.8
7.3.4	M4a	$\eta = f_{spatial}(Reg)$	D.9
	M4b	$\eta = Sex + Inc + f(Age) + f_{spatial}(Reg)$	D.10
7.3.5	M5a	$\eta = Sex * f(Age)$	D.11
	M5b	$\eta = Inc * f(Age)$	D.12
	M5c	$\eta = Inc + Sex * Inc + Sex * f(Age) +$ $f_{spatial}(Reg)$	D.13
7.4	–	$\eta = Sex + Inc + f(Age) + f_{spatial}(Reg)$	D.14

Table D.1: *Overview of all estimated models in Sections 7.3 and 7.4.*

Parameter	Mean	Std. dev.	10% quantile	Mode	90% quantile
Intercepts					
λ_{10}	-0.275	0.013	-0.292	-0.275	-0.259
λ_{20}	0.614	0.016	0.593	0.614	0.635
λ_{30}	1.257	0.022	1.229	1.257	1.286
λ_{40}	0.452	0.012	0.437	0.452	0.467
λ_{50}	0.249	0.013	0.233	0.249	0.266
Cutpoints					
τ_{22}	0.730	0.026	0.692	0.730	0.763
τ_{23}	1.416	0.026	1.380	1.417	1.447
τ_{24}	1.808	0.019	1.784	1.808	1.832
τ_{25}	2.408	0.023	2.379	2.407	2.437
τ_{32}	1.145	0.016	1.125	1.145	1.166
τ_{33}	2.453	0.040	2.400	2.454	2.505
τ_{34}	3.359	0.044	3.303	3.359	3.416
τ_{35}	4.203	0.057	4.129	4.203	4.276
τ_{42}	0.871	0.009	0.860	0.871	0.882
τ_{43}	1.785	0.016	1.764	1.784	1.806
τ_{44}	2.467	0.022	2.440	2.467	2.495
τ_{45}	3.228	0.037	3.180	3.230	3.274
τ_{52}	0.898	0.011	0.884	0.898	0.913
τ_{53}	1.854	0.019	1.830	1.854	1.879
τ_{54}	2.492	0.026	2.459	2.492	2.525
τ_{55}	3.158	0.033	3.116	3.159	3.200

Table D.2: *Results of model M1 with $\eta = 0$ – estimates of intercepts and cutpoints (see Section 7.3.1 and Table 7.6).*

Parameter	Mean	Std. dev.	10% quantile	Mode	90% quantile
Intercepts					
λ_{10}	-0.350	0.018	-0.373	-0.350	-0.327
λ_{20}	0.696	0.020	0.670	0.696	0.722
λ_{30}	1.096	0.030	1.058	1.097	1.134
λ_{40}	0.389	0.015	0.370	0.389	0.409
λ_{50}	0.167	0.019	0.142	0.167	0.191
Cutpoints					
τ_{22}	0.737	0.023	0.700	0.742	0.764
τ_{23}	1.426	0.027	1.389	1.427	1.460
τ_{24}	1.815	0.018	1.792	1.814	1.838
τ_{25}	2.435	0.025	2.403	2.435	2.466
τ_{32}	1.101	0.014	1.083	1.102	1.119
τ_{33}	2.376	0.032	2.335	2.378	2.416
τ_{34}	3.261	0.041	3.208	3.261	3.313
τ_{35}	4.097	0.052	4.031	4.098	4.164
τ_{42}	0.859	0.009	0.848	0.859	0.871
τ_{43}	1.786	0.015	1.767	1.786	1.806
τ_{44}	2.468	0.021	2.441	2.468	2.495
τ_{45}	3.231	0.031	3.192	3.231	3.271
τ_{52}	0.899	0.010	0.887	0.899	0.912
τ_{53}	1.859	0.019	1.834	1.859	1.884
τ_{54}	2.504	0.027	2.470	2.504	2.539
τ_{55}	3.190	0.034	3.147	3.191	3.234

Table D.3: *Results of model M2a with $\eta = Sex + Inc + Age$ – estimates of intercepts and cutpoints (see Section 7.3.2 and Table 7.7).*

Parameter	Mean	Std. dev.	10% quantile	Mode	90% quantile
Intercepts					
λ_{10}	-0.108	0.034	-0.151	-0.109	-0.065
λ_{20}	0.442	0.036	0.397	0.442	0.489
λ_{30}	1.500	0.058	1.427	1.501	1.573
λ_{40}	0.588	0.028	0.552	0.588	0.624
λ_{50}	0.443	0.038	0.395	0.442	0.493
Cutpoints					
τ_{22}	0.742	0.025	0.704	0.743	0.774
τ_{23}	1.432	0.030	1.391	1.434	1.471
τ_{24}	1.819	0.022	1.790	1.820	1.848
τ_{25}	2.428	0.025	2.396	2.428	2.461
τ_{32}	1.111	0.014	1.093	1.111	1.129
τ_{33}	2.387	0.034	2.344	2.387	2.430
τ_{34}	3.271	0.040	3.221	3.270	3.323
τ_{35}	4.080	0.052	4.016	4.080	4.149
τ_{42}	0.874	0.009	0.862	0.874	0.885
τ_{43}	1.788	0.015	1.769	1.788	1.808
τ_{44}	2.463	0.022	2.435	2.463	2.490
τ_{45}	3.214	0.033	3.171	3.215	3.256
τ_{52}	0.900	0.011	0.886	0.900	0.914
τ_{53}	1.870	0.020	1.844	1.871	1.895
τ_{54}	2.506	0.026	2.471	2.506	2.539
τ_{55}	3.178	0.038	3.129	3.179	3.226

Table D.4: *Results of model M2b with $\eta = Sex + Inc + Age + Sex * Inc + Sex * Age + Inc * Age$ – estimates of intercepts and cutpoints (see Section 7.3.2 and Table 7.8).*

Parameter	Mean	Std. dev.	10% quantile	Mode	90% quantile
λ_{10}	-0.262	0.013	-0.279	-0.262	-0.246
λ_{20}	0.602	0.017	0.579	0.602	0.623
λ_{30}	1.272	0.023	1.243	1.272	1.300
λ_{40}	0.461	0.012	0.446	0.461	0.476
λ_{50}	0.266	0.013	0.250	0.266	0.283
τ_{22}	0.738	0.029	0.696	0.741	0.773
τ_{23}	1.418	0.027	1.379	1.422	1.450
τ_{24}	1.802	0.019	1.778	1.802	1.826
τ_{25}	2.414	0.024	2.383	2.414	2.445
τ_{32}	1.131	0.016	1.110	1.131	1.150
τ_{33}	2.439	0.042	2.383	2.442	2.492
τ_{34}	3.350	0.046	3.292	3.350	3.408
τ_{35}	4.187	0.056	4.115	4.185	4.257
τ_{42}	0.866	0.008	0.856	0.866	0.876
τ_{43}	1.782	0.016	1.761	1.782	1.803
τ_{44}	2.474	0.022	2.446	2.474	2.502
τ_{45}	3.228	0.038	3.179	3.229	3.277
τ_{52}	0.904	0.013	0.887	0.904	0.920
τ_{53}	1.850	0.020	1.825	1.849	1.875
τ_{54}	2.490	0.026	2.456	2.490	2.524
τ_{55}	3.165	0.038	3.117	3.165	3.212

Table D.5: *Results of model M3a with $\eta = f(Age)$ – estimates of intercepts and cutpoints with a first-order random walk prior (see Section 7.3.3 and Table 7.9).*

Parameter	Mean	Std. dev.	10% quantile	Mode	90% quantile
λ_{10}	-0.262	0.013	-0.278	-0.262	-0.245
λ_{20}	0.602	0.016	0.581	0.602	0.622
λ_{30}	1.269	0.025	1.237	1.268	1.301
λ_{40}	0.463	0.012	0.448	0.463	0.478
λ_{50}	0.263	0.014	0.246	0.263	0.280
τ_{22}	0.737	0.024	0.701	0.741	0.767
τ_{23}	1.419	0.026	1.381	1.423	1.449
τ_{24}	1.809	0.019	1.786	1.809	1.834
τ_{25}	2.408	0.023	2.378	2.408	2.438
τ_{32}	1.124	0.021	1.097	1.125	1.151
τ_{33}	2.431	0.043	2.373	2.433	2.485
τ_{34}	3.347	0.047	3.287	3.344	3.407
τ_{35}	4.214	0.060	4.140	4.212	4.290
τ_{42}	0.874	0.008	0.864	0.874	0.884
τ_{43}	1.786	0.017	1.765	1.786	1.808
τ_{44}	2.460	0.021	2.433	2.460	2.488
τ_{45}	3.232	0.035	3.188	3.231	3.279
τ_{52}	0.892	0.011	0.879	0.892	0.907
τ_{53}	1.850	0.019	1.826	1.851	1.875
τ_{54}	2.488	0.025	2.456	2.488	2.520
τ_{55}	3.171	0.034	3.128	3.170	3.214

Table D.6: *Results of model M3a with $\eta = f(Age)$ – estimates of intercepts and cutpoints with a second-order random walk prior (see Section 7.3.3 and Table 7.9).*

Parameter	Mean	Std. dev.	10% quantile	Mode	90% quantile
λ_{10}	-0.259	0.013	-0.276	-0.259	-0.242
λ_{20}	0.598	0.016	0.576	0.598	0.618
λ_{30}	1.287	0.025	1.257	1.287	1.319
λ_{40}	0.465	0.012	0.449	0.465	0.481
λ_{50}	0.270	0.014	0.252	0.270	0.288
τ_{22}	0.735	0.025	0.693	0.742	0.761
τ_{23}	1.412	0.025	1.374	1.417	1.441
τ_{24}	1.808	0.018	1.786	1.808	1.831
τ_{25}	2.411	0.025	2.379	2.410	2.442
τ_{32}	1.154	0.018	1.131	1.154	1.177
τ_{33}	2.461	0.046	2.399	2.464	2.517
τ_{34}	3.342	0.047	3.282	3.341	3.401
τ_{35}	4.192	0.058	4.118	4.192	4.266
τ_{42}	0.874	0.010	0.861	0.873	0.887
τ_{43}	1.780	0.018	1.757	1.780	1.803
τ_{44}	2.472	0.022	2.443	2.473	2.500
τ_{45}	3.232	0.038	3.183	3.232	3.281
τ_{52}	0.910	0.015	0.889	0.912	0.928
τ_{53}	1.842	0.019	1.818	1.842	1.867
τ_{54}	2.496	0.026	2.462	2.496	2.530
τ_{55}	3.166	0.032	3.124	3.166	3.208

Table D.7: *Results of model M3a with $\eta = f(Age)$ – estimates of intercepts and cutpoints with a P-splines prior (see Section 7.3.3 and Table 7.9).*

Parameter	**Mean**	**Std. dev.**	**10% quantile**	**Mode**	**90% quantile**
λ_{10}	-0.643	0.019	-0.668	-0.643	-0.619
λ_{20}	1.011	0.021	0.984	1.011	1.038
λ_{30}	0.615	0.023	0.586	0.615	0.643
λ_{40}	0.160	0.014	0.141	0.160	0.178
λ_{50}	-0.162	0.017	-0.184	-0.162	-0.140
τ_{22}	0.751	0.025	0.715	0.752	0.783
τ_{23}	1.437	0.026	1.402	1.438	1.472
τ_{24}	1.823	0.018	1.799	1.823	1.846
τ_{25}	2.437	0.024	2.406	2.437	2.468
τ_{32}	1.101	0.016	1.081	1.100	1.121
τ_{33}	2.352	0.036	2.305	2.352	2.398
τ_{34}	3.246	0.040	3.195	3.245	3.297
τ_{35}	4.087	0.051	4.024	4.086	4.153
τ_{42}	0.877	0.008	0.867	0.877	0.888
τ_{43}	1.781	0.017	1.758	1.781	1.803
τ_{44}	2.456	0.023	2.427	2.456	2.485
τ_{45}	3.239	0.033	3.198	3.238	3.281
τ_{52}	0.906	0.011	0.891	0.906	0.920
τ_{53}	1.864	0.019	1.840	1.864	1.889
τ_{54}	2.505	0.025	2.474	2.505	2.538
τ_{55}	3.164	0.036	3.119	3.163	3.212

Table D.8: *Results of model M3b with $\eta = Sex + Inc + f(Age)$ – estimates of intercepts and cutpoints with a P-splines prior (see Section 7.3.3 and Table 7.10).*

Parameter	**Mean**	**Std. dev.**	**10% quantile**	**Mode**	**90% quantile**
λ_{10}	-0.298	0.013	-0.314	-0.298	-0.281
λ_{20}	0.640	0.016	0.618	0.640	0.661
λ_{30}	1.210	0.023	1.181	1.210	1.240
λ_{40}	0.432	0.012	0.416	0.432	0.447
λ_{50}	0.226	0.013	0.210	0.226	0.242
τ_{22}	0.737	0.028	0.700	0.735	0.774
τ_{23}	1.428	0.029	1.385	1.434	1.461
τ_{24}	1.812	0.019	1.787	1.812	1.835
τ_{25}	2.411	0.026	2.379	2.411	2.445
τ_{32}	1.137	0.016	1.116	1.137	1.157
τ_{33}	2.431	0.043	2.374	2.433	2.485
τ_{34}	3.321	0.046	3.263	3.320	3.381
τ_{35}	4.169	0.061	4.092	4.167	4.248
τ_{42}	0.864	0.010	0.852	0.864	0.877
τ_{43}	1.777	0.017	1.756	1.778	1.798
τ_{44}	2.469	0.021	2.442	2.469	2.496
τ_{45}	3.249	0.031	3.209	3.248	3.289
τ_{52}	0.901	0.010	0.888	0.901	0.913
τ_{53}	1.856	0.019	1.832	1.856	1.881
τ_{54}	2.489	0.025	2.457	2.488	2.521
τ_{55}	3.172	0.034	3.128	3.173	3.215

Table D.9: *Results of model M4a with $\eta = f_{spatial}(Reg)$ – estimates of intercepts and cutpoints (see Section 7.3.4 and Table 7.11).*

Parameter	**Mean**	**Std. dev.**	**10% quantile**	**Mode**	**90% quantile**
λ_{10}	-0.636	0.018	-0.660	-0.636	-0.614
λ_{20}	1.005	0.019	0.980	1.005	1.030
λ_{30}	0.629	0.024	0.599	0.629	0.660
λ_{40}	0.158	0.014	0.140	0.158	0.175
λ_{50}	-0.152	0.017	-0.174	-0.152	-0.130
τ_{22}	0.747	0.023	0.710	0.752	0.774
τ_{23}	1.439	0.020	1.412	1.441	1.463
τ_{24}	1.829	0.021	1.803	1.829	1.856
τ_{25}	2.438	0.023	2.409	2.438	2.467
τ_{32}	1.098	0.018	1.076	1.097	1.121
τ_{33}	2.373	0.040	2.321	2.375	2.423
τ_{34}	3.233	0.041	3.181	3.232	3.285
τ_{35}	4.069	0.055	4.000	4.068	4.142
τ_{42}	0.853	0.008	0.842	0.853	0.863
τ_{43}	1.782	0.016	1.762	1.783	1.803
τ_{44}	2.469	0.022	2.440	2.469	2.497
τ_{45}	3.235	0.035	3.189	3.237	3.280
τ_{52}	0.913	0.011	0.898	0.913	0.927
τ_{53}	1.866	0.020	1.840	1.866	1.891
τ_{54}	2.495	0.025	2.463	2.495	2.528
τ_{55}	3.163	0.037	3.116	3.162	3.212

Table D.10: *Results of model M4b with* $\eta = Sex + Inc + f(Age) + f_{spatial}(Reg)$ *– estimates of intercepts and cutpoints (see Section 7.3.4 and Table 7.12).*

Parameter	**Mean**	**Std. dev.**	**10% quantile**	**Mode**	**90% quantile**
λ_{10}	-0.193	0.013	-0.209	-0.193	-0.176
λ_{20}	0.525	0.016	0.505	0.525	0.546
λ_{30}	1.402	0.026	1.369	1.401	1.435
λ_{40}	0.519	0.013	0.502	0.519	0.535
λ_{50}	0.346	0.014	0.328	0.346	0.364
τ_{22}	0.727	0.025	0.693	0.726	0.760
τ_{23}	1.410	0.022	1.379	1.411	1.438
τ_{24}	1.801	0.019	1.776	1.801	1.825
τ_{25}	2.405	0.024	2.374	2.405	2.436
τ_{32}	1.143	0.021	1.118	1.142	1.171
τ_{33}	2.466	0.042	2.412	2.467	2.518
τ_{34}	3.343	0.045	3.285	3.344	3.400
τ_{35}	4.218	0.056	4.148	4.219	4.290
τ_{42}	0.870	0.008	0.859	0.870	0.880
τ_{43}	1.789	0.016	1.768	1.789	1.810
τ_{44}	2.466	0.022	2.438	2.466	2.494
τ_{45}	3.219	0.034	3.175	3.219	3.263
τ_{52}	0.907	0.012	0.892	0.907	0.924
τ_{53}	1.862	0.021	1.835	1.862	1.888
τ_{54}	2.496	0.025	2.464	2.496	2.528
τ_{55}	3.155	0.033	3.114	3.154	3.197

Table D.11: *Results of model M5a with* $\eta = Sex * f(Age)$ *– estimates of intercepts and cutpoints (see Section 7.3.5 and Table 7.13).*

Parameter	Mean	Std. dev.	10% quantile	Mode	90% quantile
λ_{10}	-0.778	0.030	-0.818	-0.778	-0.738
λ_{20}	1.156	0.030	1.116	1.158	1.195
λ_{30}	0.409	0.046	0.350	0.409	0.473
λ_{40}	0.050	0.024	0.018	0.049	0.081
λ_{50}	-0.308	0.032	-0.348	-0.310	-0.264
τ_{22}	0.754	0.026	0.714	0.757	0.784
τ_{23}	1.443	0.027	1.402	1.448	1.474
τ_{24}	1.831	0.020	1.806	1.831	1.855
τ_{25}	2.441	0.024	2.411	2.441	2.473
τ_{32}	1.117	0.017	1.095	1.117	1.139
τ_{33}	2.369	0.041	2.315	2.369	2.423
τ_{34}	3.239	0.039	3.190	3.239	3.289
τ_{35}	4.065	0.050	4.002	4.063	4.128
τ_{42}	0.865	0.009	0.854	0.865	0.877
τ_{43}	1.785	0.017	1.763	1.786	1.808
τ_{44}	2.467	0.025	2.435	2.466	2.500
τ_{45}	3.230	0.034	3.187	3.230	3.273
τ_{52}	0.906	0.010	0.893	0.906	0.919
τ_{53}	1.864	0.019	1.840	1.864	1.889
τ_{54}	2.488	0.027	2.454	2.488	2.524
τ_{55}	3.160	0.034	3.117	3.160	3.204

Table D.12: *Results of model M5b with $\eta = Inc * f(Age)$ – estimates of intercepts and cutpoints (see Section 7.3.5 and Table 7.14).*

Parameter	Mean	Std. dev.	10% quantile	Mode	90% quantile
λ_{10}	-0.672	0.021	-0.700	-0.672	-0.646
λ_{20}	1.043	0.023	1.013	1.043	1.072
λ_{30}	0.567	0.027	0.532	0.566	0.602
λ_{40}	0.134	0.016	0.113	0.134	0.155
λ_{50}	-0.193	0.021	-0.221	-0.193	-0.167
τ_{22}	0.748	0.023	0.711	0.754	0.773
τ_{23}	1.433	0.025	1.393	1.437	1.461
τ_{24}	1.821	0.018	1.798	1.821	1.845
τ_{25}	2.443	0.024	2.412	2.442	2.474
τ_{32}	1.094	0.017	1.073	1.093	1.118
τ_{33}	2.367	0.036	2.320	2.368	2.411
τ_{34}	3.239	0.042	3.185	3.238	3.293
τ_{35}	4.075	0.050	4.011	4.075	4.140
τ_{42}	0.872	0.011	0.858	0.872	0.886
τ_{43}	1.776	0.017	1.755	1.776	1.798
τ_{44}	2.471	0.023	2.441	2.472	2.501
τ_{45}	3.227	0.038	3.178	3.228	3.276
τ_{52}	0.917	0.012	0.901	0.917	0.932
τ_{53}	1.861	0.020	1.836	1.861	1.886
τ_{54}	2.500	0.026	2.466	2.500	2.533
τ_{55}	3.164	0.035	3.120	3.165	3.208

Table D.13: *Results of model M5c with $\eta = Inc + Sex * Inc + Sex * f(Age) + f_{spatial}(Reg)$ – estimates of intercepts and cutpoints (see Section 7.3.5 and Table 7.15).*

Parameter	Mean	Std. dev.	10% quantile	Mode	90% quantile
λ_{10}	0.086	0.027	0.052	0.086	0.121
λ_{20}	-0.650	0.031	-0.689	-0.650	-0.612
λ_{30}	0.986	0.031	0.947	0.985	1.027
λ_{40}	0.997	0.039	0.948	0.997	1.047
λ_{50}	0.309	0.023	0.280	0.308	0.339
λ_{60}	1.379	0.026	1.345	1.379	1.412
λ_{70}	1.391	0.029	1.354	1.391	1.428
λ_{80}	2.014	0.042	1.958	2.017	2.065
λ_{90}	1.374	0.026	1.341	1.373	1.408
$\lambda_{10,0}$	3.187	0.103	3.059	3.186	3.318
τ_{12}	0.966	0.019	0.942	0.966	0.991
τ_{13}	1.945	0.034	1.901	1.945	1.989
τ_{14}	2.599	0.047	2.539	2.596	2.661
τ_{15}	3.357	0.060	3.279	3.357	3.432
τ_{32}	0.847	0.018	0.825	0.846	0.872
τ_{33}	1.559	0.030	1.520	1.558	1.598
τ_{34}	1.963	0.034	1.919	1.963	2.008
τ_{35}	2.633	0.043	2.578	2.632	2.688
τ_{42}	1.152	0.027	1.118	1.151	1.189
τ_{43}	2.440	0.049	2.377	2.441	2.504
τ_{44}	3.299	0.061	3.221	3.297	3.378
τ_{45}	4.163	0.076	4.066	4.161	4.260
τ_{52}	0.913	0.015	0.895	0.913	0.932
τ_{53}	1.858	0.027	1.824	1.858	1.891
τ_{54}	2.537	0.036	2.491	2.538	2.583
τ_{55}	3.415	0.060	3.337	3.415	3.493
τ_{62}	1.107	0.018	1.086	1.109	1.129
τ_{63}	2.129	0.030	2.091	2.128	2.168
τ_{64}	2.633	0.033	2.591	2.632	2.675
τ_{65}	3.158	0.040	3.107	3.158	3.210
τ_{72}	1.255	0.024	1.223	1.257	1.284
τ_{73}	2.405	0.033	2.362	2.405	2.449
τ_{74}	2.971	0.040	2.920	2.970	3.021
τ_{75}	3.583	0.050	3.517	3.583	3.648
τ_{82}	1.044	0.042	0.979	1.058	1.088
τ_{83}	2.127	0.051	2.049	2.139	2.183
τ_{84}	2.802	0.039	2.751	2.803	2.851
τ_{85}	3.498	0.044	3.441	3.498	3.553
τ_{92}	1.321	0.023	1.292	1.320	1.351
τ_{93}	2.653	0.034	2.609	2.652	2.696
τ_{94}	3.478	0.049	3.415	3.478	3.539
τ_{95}	4.217	0.071	4.127	4.214	4.310
$\tau_{10,2}$	1.779	0.080	1.667	1.784	1.875
$\tau_{10,3}$	3.545	0.118	3.392	3.552	3.683
$\tau_{10,4}$	4.699	0.126	4.540	4.693	4.856
$\tau_{10,5}$	5.771	0.147	5.585	5.765	5.960

Table D.14: *Results of the model with two latent variables with $\eta = Sex + Inc + f(Age) + f_{spatial}(Reg)$ – estimates of intercepts and cutpoints (see Section 7.4 and Table 7.19).*

Appendix E

Computational details

In this chapter, we want to give a short overview of the programs and tools which were developed and used in the context of this thesis. Most programs are written in the statistical language R (2005) – furthermore the main simulation program contains a large amount of C++ code that processes performance critical calculations. All programs can be assigned to one of the temporal steps of the analysis: data editing and generation, simulation run and post processing. Table E.1 summarizes the employed programs and tools.

Data editing and generation

For the simulation studies in Sections 6.2 and 6.3, 10 different programs have been created to generate the corresponding data. The logic of all programs is identical: for each observation i, indirect and direct covariates, latent variable errors ξ_{ir}, and individual indicator errors ε_{ij} are drawn randomly. Afterwards, the underlying variables y^*_{ij} can be calculated, and the categories of all indicators are determined by the predefined cutpoints. The number of observations can be freely chosen.

For the PD1 dataset, data has to be edited and processed before it can be used by the MCMC algorithm. The function "ProcessPD1data" reads the raw PD1 data which is saved in the ASCII format. For each variable, the name and corresponding type (metric, ordinal, categorical) is set. Incomplete observations are deleted, and ordinal categories with a very low number of observations are merged with neighboring categories. The processed data is returned as a dataframe object which represents a standard data object in R for heterogenous variable types.

If a spatial analysis is to be conducted in the simulation run, the corresponding neighborhood structure has to be provided by the function "readgraphfile". This R function stems from the software package BayesX (2005) which is a multi-purpose software for Bayesian inference. The graph file containing the neighborhood structure can be easily generated in

Step	**Program**	**Task**	**Language**	**Source**
Data editing	simA, . . ., simE sim1, . . ., sim5	Returns the simulated data sets used in Sections 6.2 and 6.3.	R	Own
	ProcessPD1data	Loads the raw PD1 data and returns the processed PD1 data.	R	Own
	readgraphfile	Reads a graph file which contains the neighborhood structure of spatial covariates.	R	BayesX
Simulation run	MCMClvm	Performs the MCMC simulation and returns the sampling paths of all parameters as a MCMC object.	R, C++	Own, MCMCpack, Scythe
Post processing	summaryMCMC	Returns summary statistics such as means and quantiles of a MCMC object.	R	CODA
	plotMCMC, autocorr.plot	Plots sampling paths, parameter densities and autocorrelations of a MCMC object.	R	CODA
	plotnonp	Plots nonparametric estimated functions of metric covariates.	R	BayesX
	readbndfile, drawmap	Loads and draws spatial maps.	R	BayesX
	plotdata	Obtains a MCMClvm object and returns the data in such a way that nonparametric functions could be plotted by plotnonp.	R	Own
	standardize	Returns the standardized version of a MCMClvm object according to Table 3.2.	R	Own
	processSim	Returns summary statistics such as MSE and coverage for multiple simulation runs.	R	Own
	printlatex	Generates tables with results obtained by simulations.	R	Own

Table E.1: *Overview of programs and tools used in the context of this thesis.*

BayesX by providing the geographical coordinates of the regions in a so-called boundary file. We refer to the manual of BayesX for more details.

Simulation run

The function "MCMClvm" performs the calculations of the MCMC algorithm, and hence represents the centrepiece of all programs. The general structure of "MCMClvm" is based on a template for MCMC algorithms provided by the R-package MCMCpack (2005). The package MCMCpack contains several MCMC algorithms treating common statistical models such as the logit, probit and standard normal regression. It also contains two latent variable models, i. e. the standard factor analysis model for continuous and mixed responses. The function "MCMClvm" consists of two components: one written in R-Code, the other programmed in C++. In the R environment we invoke the R part of the function which performs the following tasks: processing of the data, for example variables not needed for the analysis and observations without a full response vector are deleted; the parameters and values of the function call are evaluated; design and penalty matrices are created; priori information is validated; starting values for the MCMC samplers are fixed and general parameters such as the number of MCMC iterations are set. After that, the R function calls the C++ code and hands over the relevant parameters and data matrices. The C++ code performs the actual MCMC simulation including all Gibbs, Metropolis Hastings and Grouped Move steps according to Section 5.4. The MCMC simulation has to be done by a fast programming language such as C++ because R code would be far too slow to perform those operations in a reasonable time. The standard matrix operations and the basic random number generators are provided by the Scythe Statistical Library (2005). However, special matrix operations for band matrices are programmed by ourselves in order to secure a good performance of the code. After the C++ code has calculated all MCMC iterations, the sampling paths of all parameters are returned to the R code which processes that information and finally returns a MCMC object according to the specifications of the R package CODA. The CODA package specifies how the data structure of a MCMC object has to look like, and provides tools for the analysis and visual illustration of MCMC objects.

To create a better understanding how a model has to be specified, we want to give an example of a "MCMClvm" function call. For that purpose, we choose the model M4b from Section 7.3.4. Recall that this PD1 model consisted of five indicators and the indirect effects *Sex* and *Inc* as parametric effects, and *Age* and *Region* as nonparametric metric and spatial covariates, respectively. The function call is given by:

```
result <- MCMClvm(ind.form =~ SYSTEM + INITIATIVE + RETIREMENT + EMERGENCY
          + HEALTH, direct.form = NA, indirect.form =~ SEX + INC +
          AGE(p,i=10,d=3,r=2) + REGION(s,map=map_g), data = PD1_data,
          burnin = 2000, mcmc = 5000, thin = 1, seed = 241137,
```

```
        loadings_prior = loadings_prior_medium, mh = FALSE, ggs = TRUE,
        verbose = TRUE, store.scores = FALSE, store.UV = FALSE)
```

The argument "ind.form" specifies all indicators used in the analysis, i. e. *System*, *Initiative*, *Retirement*, *Emergency*, and *Health*. "fixed.form" is set to NA, indicating that no direct effects are included in the model. The parameter "indirect.form" contains all indirect covariates employed in the model. We recognize that *Sex* and *Inc* are included as simple parametric effects. However, the inclusion of nonparametric effects such as *Age* and *Reg* requires additional specification of parameters. For the covariate *Age*, we have chosen P-splines with $i = 10$ intervals, $d = 3$ degrees and a second-order random walk prior on the regression coefficients. *Region* is declared as a spatial effect, and the spatial neighborhood structure is provided by the data matrix map_g which has to be loaded into the memory using the function "readgraphfile" before the analysis. "data" refers to the dataframe "PD1_data" which contains the whole social survey data needed for the analysis. Furthermore, several parameters for the MCMC simulation are handed over. "burnin" declares the number of burnin iterations, "mcmc" contains the actual number of saved iterations after the burnin phase, and the thinning parameter "thin" states that the sampled parameter values of each iteration are stored. The parameter "seed" defines the seed value of the underlying random number generator, and "loadings_prior" contains the prior information of the factor loadings. The parameters "mh" and "ggs" determine which sampling algorithm is used for the analysis – here the GGS is used. If "verbose" is set to TRUE, the C++ sampling code produces regular outputs to the screen containing relevant parameters during the sampling process. The two arguments "store.scores" and "store.UV" specify if the latent scores and the underlying variables of ordinal indicators are stored, respectively. Since this requires huge amounts of memory space, both options are switched off by default.

Post processing

After the MCMC simulation has been performed, several functions and tools are used to process the resulting sampling paths, and visualize density functions of parameters and estimated nonparametric functions of metric and spatial covariates. Standard procedures such as calculating means, medians and quantiles, and plotting sampling paths, autocorrelations and density functions of estimated parameters are provided by the CODA (2005) package. Regarding nonparametric functions, the authors of BayesX (2005) have written R functions that plot nonparametric function estimates of metric covariates, as well as maps containing the estimated functions of spatial covariates. For more detailed descriptions of those functions, we refer to the manuals of the CODA and BayesX packages, respectively.

Furthermore, additional tools have been programmed on our own which serve various purposes – the function "plotdata" transforms the data of an MCMC object in such a way that the estimated nonparametric functions could be drawn by the plotting functions provided by BayesX (2005); the tool ”standardize” takes a MCMC object of a ”MCMClvm” analysis and returns an MCMC object where all parameter values are standardized according to Table 3.2. For simulation studies, two more functions have been written: ”processSim” calculates summary statistics of multiple simulation runs, such as the coverage, the MEAN, or the MSE; the function ”printlatex” prints the summary statistics of a simulation study in Latex tables.

Bibliography

Aguilar, O., West, M. (2000). Bayesian dynamic factor models and variance matrix discounting for portfolio allocation. *Journal of Business and Economic Statistics* **18**, 338–357.

Albert, J. H., Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association* **88**, 669–679.

Anderson, T. W., Rubin, H. (1956). Statistical inference in factor analysis. In Neyman, J. (Ed.), *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability* **5**, 111–150. Berkeley, University of California Press.

Ansari, A., Jedidi, K. (2000). Bayesian factor analysis for multilevel binary observations. *Psychometrika* **65**, 475–496.

Arminger, G., Küsters, U. (1988). Latent trait models with indicators of mixed measurement level. In Langeheine, R., Rost, J. (Eds.) *Latent Trait and Latent Class models*, 51–73. Plenum, New York.

Arminger, G., Küsters, U. (1989). Construction principles for latent trait models. In Clogg, C. C. (Ed.), *Sociological Methodology 1989*, 369–393. Blackwell, Oxford.

Arminger, G., Clogg, C. C., Sobel, M. E. (Eds.) *Handbook of Statistical Modeling for the Social and Behavioral Sciences.* Plenum Press, New York.

Arminger, G., Wittenberg, J., Schepers, A. (1996). *MECOSA 3 – User Guide.* Fa. ADDITIVE, Friedrichsdorf.

Arminger, G., Muthén, B. O. (1998). A Bayesian approach to nonlinear latent variable models using the Gibbs sampler and the Metropolis-Hastings algorithm. *Psychometrika* **63**, 271–300.

Bartholomew, D. J. (1981). Posterior analysis of the factor model. *British Journal of Mathematical and Statistical Psychology* **34**, 93–99.

Bartholomew, D. J. (1987). *Latent variable models and factor analysis*. Charles Griffin & Company Ltd., London, Oxford University Press, New York. 2nd edition: Bartholomew, D. J., Knott, M. (1999). Arnold, London.

BayesX (2005). *Software for Bayesian inference* by Brezger, A., Kneib, T., Lang, S. URL: http://www.stat.uni-muenchen.de/~bayesx/bayesx.html

Bentler, P. M., Wu, E. J. C. (1995). *EQS for Windows User's Guide*. Multivariate Software, Inc., Encino.

Bernardo, J. M., Smith, A. F. M. (1994). *Bayesian Theory*. John Wiley, New York.

Berry, D. A. (1996). *Statistics: A Bayesian Perspective*. Duxbury, London.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of the Royal Statistical Society, Series B*, **36**, 192–236.

Besag, J. E., York, J., Mollie, A. (1991). Bayesian image restoration with two applications in spatial statistics (with discussion). *Annals of the Institute of Statistical Mathematics* **43**, 1–59.

Besag, J. E., Green, P., Higdon, D., Mengersen, K. (1995). Bayesian computation and stochastic systems (with discussion). *Statistical Science* **10**, 3–66.

Besag, J. E., Kooperberg, C. (1995). On conditional and intrinsic autoregressions. *Biometrika* **82**, 733–746.

Bollen, K. A. (1989). *Structural equations with latent variables*. John Wiley, New York.

Boomsma, A. (1985). Nonconvergence, improper solutions, and starting values in LISREL maximum likelihood estimation. *Psychometrika* **50**, 229–242.

Breiman, L. *Probability*. SIAM, Philadelphia.

Brezger, A., Lang, S. (2005). Generalized structured additive regression based on Bayesian P-splines. *Computational Statistics and Data Analysis*, to appear.

Brooks, S. P., Roberts, G. O. (1998). Assessing convergence of Markov chain Monte Carlo algorithms. *Statistics and Computing* **8**, 319–335.

Browne, M. W., Arminger, G. (1995). Specification and estimation of mean and covariance structure models. In Arminger, G., Clogg, C. C., Sobel, M. E. (Eds.) *Handbook of Statistical Modeling for the Social and Behavioral Sciences*, 185–249. Plenum Press, New York.

Chen, M.-H., Dey, D. K. (2000). Bayesian analysis for correlated ordinal data models. In: Dey, D. K., Ghosh, S. K., Mallick, B. K. (Eds.) *Generalized Linear Models: A Bayesian Perspective*. Marcel Dekker, New York.

CODA (2005). *Package for output analysis and diagnostics for MCMC* by Plummer, M., Best, N., Cowles, M. K., Vines, K. URL: http://www.r-project.org

Cowles, M. K. (1996). Accelerating Monte Carlo Markov chain convergence for cumulative-link generalized linear models. *Statistics and Computing* **6**, 101–111.

Cowles, M. K., Carlin, B. P. (1996). Markov chain Monte Carlo convergence diagnostics: a comparative review. *Journal of the American Statistical Association* **91**, 883–904.

Duane, S., Kennedy, A. D., Pendleton, B. J., Roweth, D. (1987). Hybrid Monte Carlo. *Physics Letters B* **195**, 216–222.

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

Eilers, P. H. C., Marx, B. D. (2004). Splines, knots and penalties. Submitted.

Everitt, B. S. (1984). *An introduction to latent variable models*. Chapman and Hall.

Fahrmeir, L., Tutz, G. (2001). *Multivariate statistical modelling based on generalized linear models*. Springer, New York.

Fahrmeir, L., Lang, S. (2001a). Bayesian inference for generalized additive mixed nodels based on Markov random field priors. *Journal of the Royal Statistical Society C* **50**, 201–220.

Fahrmeir, L., Lang, S. (2001b). Bayesian semiparametric regression analysis of multicategorical time-space data. *Annals of the Institute of Statistical Mathematics* **53**, 10–30.

Fahrmeir, L., Kneib, T., Lang, S. (2004). Penalized structured additive regression of space-time data: a Bayesian perspective. *Statistica Sinica* **14**, 731–761.

Gamerman, D. (1997). Efficient sampling from the posterior distribution in generalized linear models. *Statistics and Computing* **7**, 57–68.

Gelfand, A. E., Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association* **85**, 398–409.

Gelfand, A. E., Sahu, S. K., Carlin, B. P. (1995). Efficient parametrisations for normal linear mixed models. *Biometrika* **82**, 479-488.

Gelfand, A. E., Sahu, S. K., Carlin, B. P. (1996). Efficient parametrizations for generalized linear mixed models. *Bayesian Statistics* **5**, 165-180.

Gelman, A., Carlin, J. B., Stern, H. S., Rubin, D. B. (2004). *Bayesian data analysis* (2nd edition). Chapman and Hall.

George, A., Liu, J. W. (1981). *Computer solution of large sparse positive definite systems*. Prentice-Hall, New Jersey.

Geweke, J. (1991). Efficient simulation from the multivariate normal and Student-t distributions subject to linear constraints and the evaluation of constraint probabilities. In *Computing Science and Statistics: Proceedings of the Twenty-Third Symposium on the Interface*, 571–578.

Geweke, J., Zhou, G. (1996). Measuring the pricing error of the arbitrage pricing theory. *Review of Financial Studies* **9**, 557–587.

Gilks, W. R., Richardson, S., Spiegelhalter, D. J. (Eds.) *Markov chain Monte Carlo in practice*. Chapman and Hall.

Glas, C. A.,W. (2001). Differential item functioning depending on general covariates. In Boomsma, A., van Duijn, M. A. J., Snijders, T. A. B. (Eds.) *Essays on item response theory*, 131–148. Springer, New York.

Gould, S. J. (1981). *The mismeasure of man*. W. W. Norton, New York.

Hastie, T., Tibshirani, R. J. (1990). *Generalized additive models*. Chapman and Hall.

Hastie, T., Tibshirani, R. J. (1993). Varying-coefficient models. *Journal of the Royal Statistical Society B* **55**, 757–796.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**, 97–109.

Hauser, R. M., Goldberger, A. S. (1971). The treatment of unobservable variables in path analysis. In Costner, H. L. (Ed.) *Sociological Methodology 1971*, 81–117. Jossey-Bass, San Francisco.

Heywood, H. B. (1931). On finite sequences of real numbers. *Proceedings of the Royal Society A* **134**, 486–501.

Hobert, J. P., Casella, G. (1996). The effect of improper priors on Gibbs sampling in hierarchical linear mixed models. *Journal of the American Statistical Association* **91**, 1461–1473.

Holmes, L. B., Harvey, E. A., Kleiner, B. C., Leppig, K. A., Cann, C. I., Muñoz, A., Polk, B. F. (1987). Predictive value of minor anomalies: II. Use in cohort studies to identify teratogens. *Teratology* **36**, 291–297.

Jeffreys, H. (1961). *Theory of probability* (3rd edition). Oxford University Press, Oxford.

Johnson, V. E., Albert, J. H. (1999). *Ordinal data modeling*. Springer, New York.

Jöreskog, K. G. (1994). On the estimation of polychoric correlations and their asymptotic covariance matrix. *Psychometrika* **59**, 381–389.

Jöreskog, K. G., Goldberger, A. S. (1975). Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association* **70**, 631–639.

Jöreskog, K. G., Sörbom, D. (1996). *LISREL 8: User's Reference Guide.* Scientific Software International, Chicago.

Jöreskog, K. G., Moustaki, I. (2001). Factor Analysis of ordinal variables: a comparison of three approaches. *Multivariate Behavioral Research* **36**, 347–387.

Kammann, E. E., Wand, M. P. (2003). Geoadditive models. *Journal of the Royal Statistical Society C* **52**, 1–18.

Kendall, M., Gibbons, J. D. (1990). *Rank correlation methods.* Edward Arnold, London.

Lang, S., Fronk, E.-M., Fahrmeir, L. (2002). Function estimation with locally adaptive dynamic models. *Computational Statistics* **17**, 479–500.

Lang, S., Brezger, A. (2004). Bayesian P-splines. *Journal of Computational and Graphical Statistics* **13**, 183–212.

Lawley, D. N., Maxwell, A. E. (1963). *Factor analysis as a statistical method.* 2nd edition (1971). Butterworths, London.

Lazarsfeld, P. F., Henry, N. W. (1968). *Latent structure analysis.* Houghton Mifflin, Boston.

Lee, S.-Y. (1981). A Bayesian approach to confirmatory factor analysis. *Psychometrika* **46**, 153–160.

Lee, P. M. (2004). *Bayesian statistics: an introduction* (3rd edition). Arnold, London.

Lee, S.-Y., Poon, W.-Y., Bentler, P. M. (1992). Structural equation models with continuous and polytomous variables. *Psychometrika* **57**, 89–105.

Lee, S.-Y., Poon, W.-Y., Bentler, P. M. (1995). A two-stage estimation of structural equation models with continuous and polytomous variables. *British Journal of Mathematical and Statistical Psychology* **48**, 339–358.

van der Linden, W. J., Hambleton, R. K. (1997). *Handbook of modern item response theory.* Springer, New York.

Liu, J. S. (1994). The collapsed Gibbs sampler in Bayesian computations with applications to a gene regulation problem. *Journal of the American Statistical Association* **89**, 958–966.

Liu, J. S., Sabatti, C. (2000). Generalised Gibbs sampler and multigrid Monte Carlo for Bayesian computation. *Biometrika* **87**, 353–369.

Lopes, H. F., West, M. (2004). Bayesian model assessment in factor analysis. *Statistica Sinica* **14**, 41–67.

Martin, J. K., McDonald, R. P. (1981). Bayesian estimation in unrestricted factor analysis: a treatment for Heywood cases. *Psychometrika* **40**, 505–517.

MCMCPack (2005). *R-package for Markov chain Monte Carlo* by Martin, A. D., Quinn, K. M. URL: http://mcmcpack.wustl.edu/

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., Teller, E. (1953). Equations of state calculations by fast computing machines. *Journal of Chemical Physics* **21**, 1087–1092.

Moustaki, I. (2000). A latent variable model for ordinal variables. *Applied Psychological Measurement* **24**, 211–223.

Moustaki, I. (2003). A general class of latent variable models for ordinal manifest variables with covariate effects on the manifest and latent variables. *British Journal of Mathematical and Statistical Psychology* **56**, 337–357.

Moustaki, I., Jöreskog, K. G., Mavridis, D. (2004). Factor models for ordinal variables with covariate effects on the manifest and latent variables: a comparison of LISREL and IRT approaches. *Structural Equation Modeling* **11**, 487–513.

Muthén, B. O. (1984). A general structural equation model with dichotomous, ordered categorical and continuous latent variables indicators. *Psychometrika* **49**, 115–132.

Muthén, B. O. (1989). Latent variable modeling in heterogeneous populations. *Psychometrika* **54**, 557–585.

Muthén, B. O. (2002). Beyond SEM: General latent variable modeling. *Behaviormetrika* **29**, 81–117.

Muthén, L., Muthén, B. O. (1998-2001). *Mplus User's Guide*. Muthén and Muthén, Los Angeles.

Nandram, B., Chen, M.-H. (1996). Reparameterizing the generalized linear model to accelerate Gibbs sampler convergence. *Journal of Statistical Computation and Simulation* **54**, 129-144.

Neal, R. M. (1994). An improved acceptance procedure for the hybrid Monte Carlo algorithm. *Journal of Computational Physics* **111**, 194-203.

Neal, R. M. (2003). Slice sampling (with discussions). *Annals of Statistics* **31**, 705-767.

Press, S. J., Shigemasu, K. (1989). Bayesian inference in factor analysis. In Gleser, L. J., Perlman, M. D., Press, S. J., Sampson, A. R. (Eds.) *Contributions to probability and statistics*, 271–287. Springer, New York.

Quinn, K. M. (2004). Bayesian factor analysis for mixed ordinal and continuous responses. *Political Analysis* **12**, 338–353.

R – *The R Project for Statistical Computing* (2005). URL: http://www.r-project.org

Rao, M. M. (1987). *Measure theory and integration*. John Wiley, New York.

Reckase, M. D. (1997). The past and future of multidimensional item response theory. *Applied Psychological Measurement* **21**, 25–36.

Roberts, G. O., Tweedie, R. L. (1996). Exponential convergence of Langevin diffusions and their discrete approximations. *Bernoulli* **2**, 341-363.

Roberts, G. O., Sahu, S. K. (1997). Updating schemes, correlation structure, blocking and parameterization for the Gibbs sampler. *Journal of the Royal Statistical Society B* **59**, 291–317.

Robert, C. P., Casella, G. (2004). *Monte Carlo statistical methods* (2nd edition). Springer, New York.

Rue, H. (2001). Fast sampling of Gaussian Markov random fields. *Journal of the Royal Statistical Society B* **63**, 325–338.

Ruppert, D., Wand, M. P., Carroll, R. J. (2003). *Semiparametric regression*. Cambridge University Press.

Rupp, A. A., Dey, D. K., Zumbo, B. D. (2004). To Bayes or not to Bayes, from whether to when: applications of bayesian methodology to modeling. *Structural Equation Modeling* **11**, 424–451.

Sammel, M. D., Ryan, L. M. (1996). Latent variable models with fixed effects. *Biometrics* **52**, 650–663.

Sammel, M. D., Ryan, L. M., Legler, J. M. (1997). Latent variable models for mixed discrete and continuous outcomes. *Journal of the Royal Statistical Society B* **59**, 667–678.

Scheines, R., Hoijtink, H., Boomsma, A. (1999). Bayesian estimation and testing of structural equation models. *Psychometrika* **64**, 37–52.

Scythe Statistical Library (2005). *Open source C++ library for statistical computation* by Martin, A. D., Quinn, K. M., Pemstein, D. URL: http://scythe.wustl.edu

Seber, G. A. F. (1984). *Multivariate observations*. John Wiley, New York.

Skrondal, A., Rabe-Hesketh, S. (2004). *Generalized latent variable modeling.* Chapman and Hall.

Spearman, C. (1904). General intelligence objectively determined and measured. *American Journal of Psychology* **15**, 201–293.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society B* **64**, 583–639.

Shi, J. Q., Lee, S.-Y. (1997). A Bayesian estimation of factor score in confirmatory factor model with polytomous, censored or truncated data. *Psychometrika* **62**, 29–50.

Sun, D., Tsutakawa, R. K., He, Z. (2001). Propriety of posteriors with improper priors in hierarchical linear mixed models. *Statistica Sinica* **11**, 77–95.

Takane, Y., de Leeuw, J. (1987). On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika* **52**, 393–408.

Tanner, M. A., Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association* **82**, 528–540.

Tierney, L. (1994). Markov chains for exploring posterior distributions (with discussion). *Annals of Statistics* **22**, 1701–1786.

Tong, Y. L. (1990). *The multivariate normal distribution.* Springer, New York.

Verhelst, N. D., Glas, C. A. W., Verstralen, H. H. F. M. (1994). *OPLM: One-parameter logistic model. Computer program and manual.* CITO, Arnhem.

van Driel, O. P. (1978). On various causes of improper solutions in maximum likelihood factor analysis. *Psychometrika* **43**, 225–243.

Zellner, A. (1970). Estimation of regression relationships containing unobservable independent variables. *International Economic Review* **11**, 441–454.

Zhu, J., Eickhoff, J. C., Yan, P. (2005). Generalized linear latent variable models for repeated measures of spatially correlated multivariate data. *Biometrics* **61**, 674–683.

Zwinderman, A. (1997). Response models with manifest predictors. In van der Linden, W. J., Hambleton, R. K. (Eds.) *Handbook of modern item response theory*, 245–256. Springer, New York.