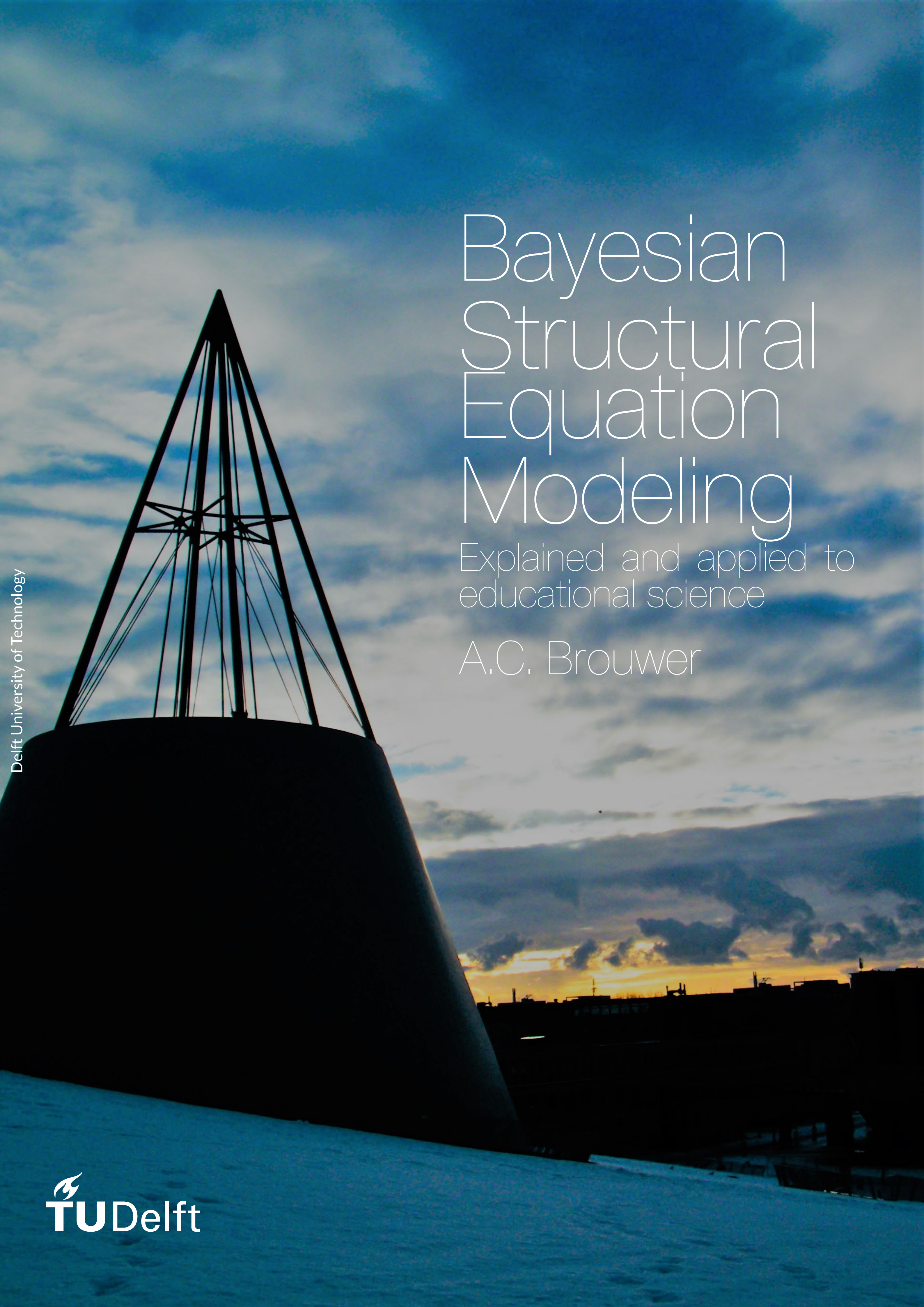
Bayesian Structural Equation Modeling

Explained and applied to educational science

A.C. Brouwer

Delft University of Technology



Bayesian Structural Equation Modeling

Explained and applied to educational science

by

A.C. Brouwer

to obtain the degree of Bachelor of Science
at the Delft University of Technology,
to be defended publicly on Thursday July 8, 2021 at 2:00 PM.

Student number:	4575687	
Project duration:	April 19, 2021 – July 8, 2021	
Thesis committee:	Dr. A.J. Cabo,	TU Delft, supervisor
	Dr. ir. N.J. van der Wal,	TU Delft, supervisor
	Dr. J.L.A. Dubbeldam,	TU Delft

An electronic version of this thesis is available at `http://repository.tudelft.nl/`.

“From a Bayesian perspective, estimation is less about deducing the values of population parameters and more about updating, sharpening, and refining our beliefs about the empirical world.”

Rick H. Hoyle

Abstract

Structural equation modeling (SEM) is frequently used in social sciences to analyze relations among observed and latent variables and test theoretical propositions regarding relations among these latent variables. Frequentist SEM relies on Maximum Likelihood Estimation, and although this method works well for many simple situations, its performance is unsatisfactory when dealing with complex models or small sample sizes. In search of a method that resolves those problems, Bayesian SEM has been developed recently. These models produce more accurate parameter estimates. The Bayesian approach to SEM offers the possibility of incorporating prior knowledge into SEM, allowing for model extension and improvement. In this research, the theory of both frequentist and Bayesian SEM is described. Subsequently, Bayesian SEM is illustrated with an application in educational sciences. A method is proposed to specify prior distributions that use correlation estimates found in previous research to reflect prior information and our confidence in that information. The results obtained by an informative prior model are analyzed and compared to the results of a noninformative, weakly informative, and frequentist model. It was found that the informative prior model produces more accurate estimates than the noninformative and weakly informative prior models, indicating the correctness of the specified priors.

Preface

This thesis has been written as the final requirement to obtain the degree of Bachelor of Science in Applied Mathematics at Delft University of Technology. The research was conducted in collaboration with the PRogramme of Innovation in Mathematics Education (PRIME) in the period April-July 2021, under the supervision of Dr. A.J. Cabo and Dr. N.J. van der Wal.

I want to thank Nathalie van der Wal for her support and guidance during my bachelor project, especially when my ideas became too ambitious for the limited time available. Secondly, I also want to thank Annoesjka Cabo for her guidance and advice. I have enjoyed our meetings and our shared enthusiasm for the applications and models we investigated.

During the past three months, I have been given the opportunity to get acquainted with the applications of statistics in educational science. Although she was not my direct supervisor, I want to thank Jacqueline Wong for her sincere interest and her help with the psychology- and educational parts of my research.

In addition, I want to thank both Prof. Ed Merkle and Dr. Mauricio Garnier Villareal for taking the time to help me with implementation questions and offering suggestions for improvements. Moreover, I want to thank Ed Merkle for creating the R package that I used in this thesis.

Finally, I want to thank Johan Dubbeldam for taking a seat in my thesis committee.

A.C. Brouwer
Rotterdam, July 2021

Contents

1

Introduction

Structural equation modeling (SEM) is a growing family of methods in the field of statistics that can be used to model and estimate relations between unobserved, or latent variables. In social sciences, it is common that certain variables cannot be measured in a direct way. Instead, these latent variables are estimated by multiple observable variables. Examples of latent variables are student motivation and academic performance, in which case the corresponding observed variables could be answers to questions relating to motivation and grades for different courses, respectively. SEM can be used to model the relations between latent variables using observed data and offers a flexible and increasingly popular approach to hypothesis testing in the social sciences. Although the SEM framework was first developed in the early 1980s [2], it is still expanding today through new developments. These emerging capabilities, along with improved software programs for implementation, have caused an enormous growth in the popularity of SEM. In the past two decades, SEM has become a widely used analysis method in various fields, such as psychology, sociology, and educational studies [13].
Most commonly used is frequentist SEM, in which the model parameters are estimated by fitting a maximum likelihood function. Since this method maximizes the likelihood function, it results in estimates for which the observed data are most probable given the model. Although this estimation method in SEM works well for many simple situations, its performance is unsatisfactory in dealing with complex problems that involve complicated model structures or small sample sizes [34].

1.1. Bayesian SEM

Bayesian structural equation modeling (Bayesian SEM, BSEM) has recently gained popularity because it can potentially resolve some of the difficulties with traditional frequentist SEM.
The primary difference between the frequentist and Bayesian approaches to SEM is that Bayesian SEM assumes that the unknown parameter vector is a random variable. In contrast, the frequentist SEM assumes the parameter vector to be a constant. The Bayesian approach offers several advantages, including the possibility of incorporating prior information based on expert knowledge and previous research. The prior information can be added to the model via prior distributions for the model parameters.
The estimation of Bayesian SEM often relies on Markov chain Monte Carlo methods. This approach resolves several difficulties of the maximum likelihood estimator and improves model outcomes [34]. Previous research has established that, generally, better results are realized using prior distributions and Markov chain Monte Carlo methods. In particular, Bayesian SEM generates more reliable results with complicated model structures or moderate sample sizes [13, 17].

To this date, research in many applications tended to focus on Bayesian structural equation models with a small number of latent variables. Furthermore, previous studies on Bayesian SEM in the field of psychometrics often lack proper mathematical formulation. This thesis gives an overview of Bayesian SEM to estimate relations between a large number of latent variables and explores its possibilities.

1.2. Application in educational science

This thesis provides an introduction to Bayesian SEM as an important alternative to frequentist approaches to SEM. To illustrate this, Bayesian SEM is used to test hypotheses from research in educational sciences by analyzing student data collected by PRIME (PRogramme of Innovation in Mathematics Education, research group at TU Delft). The hypotheses model of the PRIME research consists of a large number of latent variables. Therefore, the model structure can be categorized as complicated. Thus, for the PRIME research, the Bayesian approach to SEM is expected to be a better suited method than the frequentist approach. A comparison between Bayesian en frequentist SEM will be made to investigate if the Bayesian approach is indeed a better fit for the hypotheses model.

To improve mathematics education for engineering students at Delft University of Technology, further insight in the dynamics between academic success, engagement, and teacher behavior is needed. Therefore, the research question to be answered in the application of Bayesian SEM to the PRIME research is the following:

> What are the relationships between *perceived scaffolding and monitoring cues*, satisfaction and frustration of needs for *relatedness*, *autonomy* and *competence*, *autonomous*- and *controlled motivation*, and *academic performance*?

1.3. Thesis outline

This research starts with an introduction to the concept of latent and observed variables and an overview of the frequentist approach to SEM in Chapter 2. This chapter is intended to introduce readers to key ideas and the steps involved in the procedure of SEM. Subsequently, Chapter 3 elaborates on the Bayesian approach to SEM, including information on informative parameter priors, as well as noninformative and weakly informative priors. In Chapter 4, Bayesian SEM will be applied to the hypotheses formulated in the PRIME educational research. A method to specify prior distributions is proposed, that incorporates previously found correlation estimates to reflect both the prior information and our confidence in that information. To examine if these priors are correctly specified and if they improve the model, a comparison is made with a noninformative and a weakly informative prior model that both use the Bayesian approach. To investigate whether Bayesian SEM has offered more accurate results than the frequentist approach in the application to student motivation and performance, this chapter will compare the Bayesian results with results obtained by a frequentist SEM.

Following convention, in this thesis the abbreviation SEM will be used for both the method, structural equation modeling and the model, a structural equation model.

2
Structural Equation Modeling

The origins of structural equation modeling (SEM) stem from factor analysis and path analysis. The idea of factor analysis is to identify latent variables (factors) based on the correlation structure of the observed variables (indicators), where path analysis was designed to examine and compare relations between variables. The measurement approach (factor analysis) and structural approach (path analysis) were integrated to create SEM, a more generalized analytical framework that can be used to estimate relations between latent variables [36].

SEM is widely used in behavioral, educational, psychological, and social research. More recently, the method is gaining attention in biomedical research [17].

2.1. Observed and latent variables

Observed variables, sometimes called observable, manifest, or measurement variables, are those that can be measured in a direct way. Examples are test scores, the grade point average, gender, and Likert items (statement for which a respondent specifies their agreement level). Measurements from these variables provide observed data that can be used as the basic source of information for the statistical analysis of a structural equation model [17].

In psychological and educational research, it is common that certain variables cannot be measured in a direct way. These variables are called latent, or unobserved. Examples of such variables are student motivation and needs satisfaction. To gain insight in latent constructs, a combination of several observed variables is needed [17]. Hence, every latent variable is linked to a certain amount of observed variables through a coefficient called a factor loading that indicates the strength of the link. For instance, several research-based Likert items should be combined to evaluate student motivation and needs satisfaction.
We distinguish two types of latent variables: endogenous latent variables and exogenous latent variables. The former are influenced by other latent variables of the model, whilst the latter are not. Thus, the causes of exogenous latent variables lie outside the model, but these variables can influence other (endogenous) latent variables. The two types result from the specification of links between the latent variables in a SEM.

For practical research in social sciences, it is often necessary to examine the relations between the latent variables of interest. An example of such research can be found in Chapter 4. This gives motivation for the development of structural equation models: a statistical method that simultaneously groups highly correlated observed variables into latent variables and assesses relations among latent variables [17].

2.2. Structural equation model

Structural equation modeling can be summarized in five steps: model specification, model identification, model estimation, model evaluation, and model modification [3, 36]. This section will elaborate on these five steps, which are summarized in Figure 2.1.

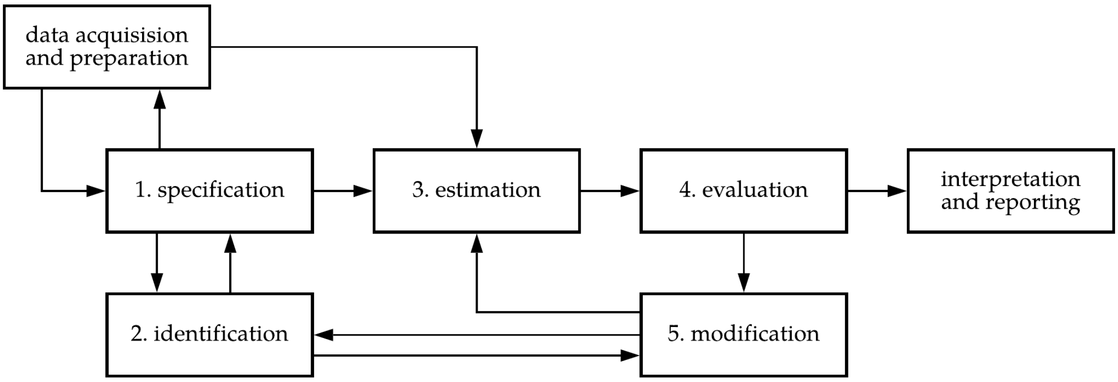


Figure 2.1: Steps in the implementation of a SEM. The arrows visualize the typical process of revisiting steps.

2.2.1. Model specification

The first step refers to correctly specifying and formulating the structural equation model for obtaining the correlation estimates. We will first introduce the variables and notation.

The endogenous latent variables of the model are represented by vector η, and the exogenous latent variables by vector ξ. The effects of the endogenous latent variables on other endogenous latent variables are represented by matrix B. The effects of the exogenous latent variables on the endogenous latent variables are represented by matrix Γ, and lastly, ζ represents the residual term. It is assumed that ζ is independent of η and ξ.

The endogenous and exogenous latent variables are defined by the corresponding observed variables, denoted X and Y, through two systems of linear equations with coefficient matrices Λ_y and Λ_x, respectively. In other words, Λ_y and Λ_x contain the factor loadings. For each latent variable, one of the factor loadings must be specified as 1, for the purpose of model identification and defining the scale of the latent variable. The measurement errors are denoted by ϵ and δ, it is assumed they are independent of each other and η, ξ and ζ [36]. An overview of the definitions of the variable vectors is given in Table 2.1 and an overview of the definitions of the parameter matrices is given in Table 2.2.

Table 2.1: Definitions of the variable vectors of the structural equation model.

Variable	Definition	Dimension
η	Endogenous latent variable	$m \times 1$
ξ	Exogenous latent variable	$n \times 1$
ζ	Residual term in equations	$m \times 1$
ϵ	Measurement errors of the endogenous factor loadings	$p \times 1$
δ	Measurement errors of the exogenous factor loadings	$p \times 1$

Note: m and n correspond to the number of endogenous and exogenous latent variables, respectively. p and q represent to the number of observed variables that correspond to the endogenous and exogenous latent variables, respectively.

Table 2.2: Definitions of the parameter matrices of the structural equation model.

Matrix	Definition	Dimension
B	Coefficient matrix for relations between endogenous latent variables	$m \times m$
Γ	Coefficient matrix relating exogenous latent variables to endogenous latent variables	$m \times n$
Y	Observed variables corresponding to endogenous latent variables	$p \times 1$
X	Observed variables corresponding to exogenous latent variables	$q \times 1$
Λ_y	Factor loadings relating to endogenous variables	$p \times m$
Λ_x	Factor loadings relating to exogenous latent variables	$q \times n$

Note: m and n correspond to the number of endogenous and exogenous latent variables, respectively. p and q represent to the number of observed variables that correspond to the endogenous and exogenous latent variables, respectively.

The general structural equation model can be expressed by three basic equations [36]:

Structural model

$$\eta = B\eta + \Gamma\xi + \zeta \tag{2.1}$$

Measurement model

$$Y = \Lambda_y \eta + \epsilon \tag{2.2}$$

$$X = \Lambda_x \xi + \delta \tag{2.3}$$

These equations are expressed in matrix format. Equation 2.1 represents the structural model which establishes the relations or structural equations among the latent variables. Equations 2.2 and 2.3 are measurement equations.

To further illustrate SEM, consider the following example.

Example 2.2.1. *Consider a SEM with one exogenous variable, ξ, that is defined by observed variables x_1, x_2, x_3 and x_4, and with two endogenous latent variables, ζ_1 and ζ_2. The effects of ξ on ζ_1 and ζ_2 are denoted by γ_{11} and γ_{12} respectively. Variable ζ_1 is defined by observed variables y_1 and y_2. Variable ζ_2 is influenced by ζ_1, specified by β_{12}, and is defined by y_3, y_4 and y_5. For all latent variables, the first factor loading is taken as 1 for scaling purposes. The other factor loadings are specified by λ's.*

The structural part of the model can be expressed as

$$\begin{bmatrix}\zeta_1\\ \zeta_2\end{bmatrix} = \begin{bmatrix}0 & 0\\ \beta_{12} & 0\end{bmatrix}\begin{bmatrix}\zeta_1\\ \zeta_2\end{bmatrix} + \begin{bmatrix}\gamma_{11}\\ \gamma_{21}\end{bmatrix} X + \begin{bmatrix}\zeta_1\\ \zeta_2\end{bmatrix}. \tag{2.4}$$

The first measurement equation can be expressed as

$$\begin{bmatrix}y_1\\ y_2\\ y_3\\ y_4\\ y_5\end{bmatrix} = \begin{bmatrix}1 & 0\\ \lambda_{y21} & 0\\ 0 & 1\\ 0 & \lambda y42\\ 0 & \lambda y52\end{bmatrix}\begin{bmatrix}\zeta_1\\ \zeta_2\end{bmatrix} + \begin{bmatrix}\epsilon_1\\ \epsilon_2\\ \epsilon_3\\ \epsilon_4\\ \epsilon_5\end{bmatrix}, \tag{2.5}$$

and the second measurement equation as

$$\begin{bmatrix}x_1\\ x_2\\ x_3\\ x_4\end{bmatrix} = \begin{bmatrix}1\\ \lambda_{x21}\\ \lambda_{x31}\\ \lambda_{x41}\end{bmatrix}\xi + \begin{bmatrix}\delta_1\\ \delta_2\\ \delta_3\\ \delta_4\end{bmatrix}. \tag{2.6}$$

A visualization of this example can be found in Figure 2.2.

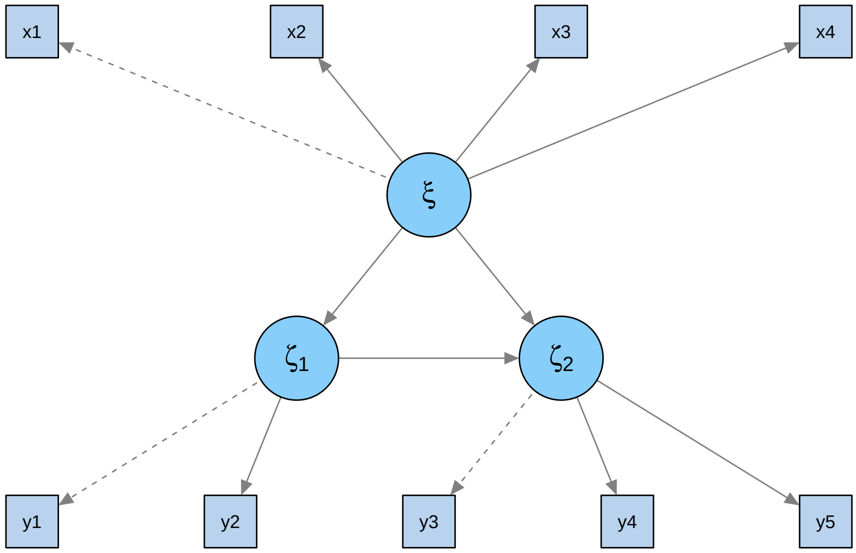


Figure 2.2: Path diagram of a structural equation model. The circles represent the latent variables and the squares represent the corresponding observed variables. The dotted lines represent the factor loadings that are specified to be 1, for scaling purposes.

2.2.2. Model identification

Identification is an important consideration when specifying a SEM. The model identification determines whether it is possible to obtain consistent and unique estimates for the unknown model parameters from the observed data.
The easiest test for identification is the t-rule [2]. Intuitively, this rule checks whether the amount of unknown information in the model is less than or equal to the amount of known information. The t-rule is given by the following equation:

$$t = 2p + z \leq \frac{1}{2}(p)(p+1), \tag{2.7}$$

where p is the number of observed variables and z is the number of relations between latent variables.

To illustrate this rule, we check the model specified in Example 2.2.1 for identification. In the example, there are 9 observed variables and 3 relations between latent variables. Hence, the t-rule holds:

$$t = 2(9) + 3 = 21 \leq 45 = \frac{1}{2}(9)(9+1). \tag{2.8}$$

2.2.3. Model estimation

The idea of SEM estimation is to minimize the difference between the sample covariances and the covariances predicted by the model. The key assumption is that if the model were correct and if we knew the parameters, the population covariance matrix would be exactly reproduced. Thus, the hypothesis that is tested by a SEM is

$$\Sigma = \Sigma(\theta)m \tag{2.9}$$

where Σ represents the population covariance matrix, which is the matrix containing the true covariances between the model parameters. The right part of the equation, $\Sigma(\theta)$, represents the model implied covariance matrix. The estimated model parameters are the result of minimizing the difference between these two matrices. Since we do not know Σ and $\Sigma(\theta)$, we minimize

$$(S - \Sigma(\hat{\theta})), \tag{2.10}$$

where S is the sample covariance matrix, which is an estimation of Σ based on the sample data, and $\Sigma(\hat{\theta})$ is the model estimated covariance matrix with $\hat{\theta}$ representing the estimated model parameters [2].

Thus, for model estimation, we are interested in those values for θ that minimize the difference between what we observe in the sample data and what the SEM implies. In general, this difference is minimized iteratively by Maximum Likelihood Estimation.

2.2.4. Model evaluation

Another important step in SEM is to conduct an overall model fit test on the basic hypothesis $S = \Sigma(\hat{\theta})$. If the model does not fit the data well, the estimates are not reliable. Therefore, the overall model fit evaluation should be done before interpreting the parameter estimates. To do this, many model fit indices have been developed [36]. In this section, we give the most commonly used model fit indices that are provided in 'blavaan', the R package used in this study (see section 4.1.4).

Measures of absolute fit assume that a perfect model has no deviation between the sample covariance matrix and the model implied covariance matrix. Thus, a measure of absolute fit compares the SEM with a theoretically perfect model. A low value of the measure of absolute fit implies a small difference between the covariance matrices, meaning that the SEM is a good fit. We give two measures of absolute fit [3, 36].

Definition 2.2.1. *Let N be the sample size, let df be the degrees of freedom of the model and let χ^2 be the χ^2-statistic*[1] *Then, the* ***root mean square error of approximation*** *(RMSEA) is defined by*

$$RMSEA = \sqrt{\frac{\chi^2 - df}{df(N-1)}}. \tag{2.11}$$

The RMSEA value ranges from 0 to 1, with smaller values indicating a better model fit. A value in the range of 0 and 0.05 is considered to indicate a good model fit, a value in the range of 0.05-0.08 a fair model fit and values above 0.08 a poor fit [36].

Definition 2.2.2. *Let p be the number of observed variables and let s and σ be the elements from the sample covariance matrix and the model estimated covariance matrix, respectively. The* ***standardized root mean squared residual*** *(SRMR) is defined by*

$$SRMR = \sqrt{\frac{\Sigma_j \Sigma_k \left(\frac{s_{jk}}{\sqrt{s_{jj} s_{kk}}} - \frac{\sigma_{jk}}{\sqrt{\sigma_{jj}\sigma_{kk}}} \right)^2}{p(p-1)/2}}. \tag{2.12}$$

The SRMR value ranges from 0 to 1, with smaller values indicating a better model fit. A value in the range of 0 and 0.08 is considered to indicate a good model fit, a value in the range of 0.08-0.10 an acceptable model fit and values above 0.10 a poor fit [36].

On the other hand, measures of relative fit compare the SEM to the null-model, which represents the model where all variables are uncorrelated. A high value of the measure of relative fit implies that the model is significantly better than the null-model, meaning that the SEM is a good fit. We give two measures of relative fit [3, 36].

[1] Or: chi-squared test. Test of overall model fit. The null hypothesis of the chi-square test is $H_0 : \Sigma = \Sigma(\theta)$ [2].

Definition 2.2.3. *Let $d = \chi^2 - df$. Then, the **comparative fit index** (CFI) is defined as*

$$CFI = \frac{d_{null} - d_{model}}{d_{null}}. \tag{2.13}$$

The CFI value ranges from 0 to 1; a value above 0.9 is considered to indicate a good fit.

Definition 2.2.4. *The **Tucker-Lewis index** (TLI) is defined by*

$$TLI = \frac{\frac{\chi^2_{null}}{d_{null}} - \frac{\chi^2_{model}}{d_{model}}}{\frac{\chi^2_{null}}{d_{null}}}. \tag{2.14}$$

The TLI value ranges from 0 to 1; a value above 0.9 is considered to indicate a good fit [36].

These measures combined offer acceptable insight in the goodness of fit. Hence, these fit indices are a good approach to model evaluation.

2.2.5. Model modification

The model specification in the application of SEM is generally based on theory or previous research. Often, the initial model does not fit the available data well. If the fit is not good, the model outcomes should be assessed to determine what is specifically wrong with the model specification. After that, the model specification has to be modified, taking into account identification, and re-tested using the same data.

Most software programs that can model SEMs offer suggestions of changes in the model structures in the form of modification indices. In the model modification step, variables or relations should not be added or removed solely for the purpose of model fit improvement if this does not make sense in the theoretical context of the application. The goal of SEM is to find a model that fits the data well from a statistical point of view, while the model parameters have a meaningful interpretation [36].

3

Bayesian Structural Equation Modeling

The structural equation model, as defined in chapter 2, can be categorized as a frequentist approach because the unknown model parameters are considered to be fixed. Parallel to expansions of the classical structural equation models, Bayesian structural equation models have been developed[13], where the model parameters are considered random variables.

Bayesian statistics, and in particular the Bayesian approach to structural equation modeling, has several advantages compared to its frequentist counterpart [17]. Firstly, prior knowledge and information can be incorporated into the analysis in order to produce better results. Secondly, more reliable results are obtained with moderate sample sizes or complex models. Lastly, the model comparison statistics, such as the Bayes factor and deviance information criterion, offer more reasonable and flexible tools for model comparison and model checking. Model comparison statistics can be used for hypothesis testing and for the assessment of the goodness of fit. These advantages make Bayesian structural equation modeling an attractive method to estimate links between latent variables.

3.1. Bayesian inference

Statistical inference can be defined as the problem of turning data into knowledge, where knowledge often is not directly present in the data but rather in models that one uses to interpret data [8]. In other words, statistical inference is concerned with drawing conclusions from numerical data about variables that are not directly observed.
There is no universal agreement on the best approach to statistical inference. Two commonly advocated views are the frequentist approach and the Bayesian approach [28]. To illustrate these different views, let θ be a vector of unknown parameters of an arbitrary SEM, and let $Y = (y_1, ..., y_n)$ be the observed data set with a sample size n. In a frequentist approach, this vector of variables θ is not considered random. It is assumed that there is an unknown true hypothesis about the underlying distribution of θ and that the observed data Y is sampled from that distribution. Analyses of the observed data are used to determine the underlying truth of an experiment. In particular, the probability that a hypothesis is true, regardless of the data, is not given or taken into account.

However, Bayesian inference models uncertainty by including the probability of the hypotheses in the model through prior distributions of unknown model parameters. The method depends on a subjective prior and the likelihood of observed data. In this approach, all unknowns are considered to be random variables. Thus, in a Bayesian approach, the vector of unknown variables θ is considered to be random, with prior distributions and associated prior density functions for every element. The mean of the estimated posterior distribution is often taken as estimate for the parameters of interest. Through this prior distribution, the model can be influenced by experiences of experts, thus giving room for discussion about what priors to specify. For this reason, Bayesian statistics are sometimes seen as more subjective than the frequentist approach. Nonetheless, the Bayesian approach is well recognized as an advantageous approach to analyze a wide variety of models [16].

The process of Bayesian inference can be described by the following three steps [11]:

1. Setting up a full probability model: specifying a probability distribution that incorporates previous beliefs about observable and unobservable variables. The model should be consistent with knowledge about the underlying theoretical framework and the data collection process.
2. Conditioning on the observed data: calculating and interpreting the appropriate posterior distribution.
3. Evaluating the model: how well does it fit the data, what are the mean and mode of the resulting posterior distribution? After analyzing the estimations, one can adjust the model and repeat the three steps.

3.2. Bayesian structural equation model

To introduce the Bayesian approach to structural equation modeling, we consider an arbitrary SEM with a vector of unknown parameters θ. This vector thus contains the parameters from Equations (2.1) to (2.3). We are interested in estimating the parameters of θ.
The posterior distribution of θ plays a critical role in the Bayesian estimations of the model parameters. For that reason, we determine the density function of the posterior distribution of θ, using Bayes' theorem.

Theorem 3.2.1 (Bayes' theorem). *Let (Ω, F, P) be a probability space, and $A, B \in F$ such that $p(B) > 0$. Then*

$$p(A|B) = \frac{p(B|A)p(A)}{p(B)}. \tag{3.1}$$

Bayes' theorem follows from the definition of conditional probability, the multiplicative law of probability, and the law of total probability [27].

To apply Theorem 3.2.1, we first define the posterior distribution of θ.

Definition 3.2.1. *Let M be an arbitrary SEM with a vector of model parameters θ, and let Y be the observed data with sample size n. The* ***posterior distribution of*** $\boldsymbol{\theta}$ *is defined as $p(\theta|Y, M)$, or in short $p(\theta|Y)$.*

By conditioning on the known value of the data Y and using Bayes' theorem, we obtain the following corollary regarding the posterior density:

Corollary 3.2.1.1.

$$p(\theta|Y) = \frac{p(Y|\theta)p(\theta)}{p(Y)}. \tag{3.2}$$

Here, $p(Y)$ does not depend on θ and, with fixed Y, can be considered to be a constant. Thus, an equivalent form of (3.2) is given by

$$p(\theta|Y) \propto p(Y|\theta)p(\theta), \quad or \tag{3.3}$$

$$\log\, p(\theta|Y) \propto \log\, p(Y|\theta) + \log\, p(\theta). \tag{3.4}$$

Note that, since θ is considered random in the Bayesian approach, $p(\theta)$, or $p(\theta|M)$, can be regarded as the density function of the prior distribution of θ. The second term, $p(Y|\theta)$, or $p(Y|\theta, M)$, is the likelihood function, because it is the probability density of Y conditional on the vector θ [16].

In case of a large sample size, $p(Y|\theta)$ could become very large and could thus decrease the influence of the prior distribution. This leads to the following lemma.

Lemma 3.2.1. *The Bayesian and frequentist approach to SEM are asymptotically equivalent*[1].

[1]Asymptotic equivalence occurs when functions eventually become essentially equal. More precisely, we say that two functions f and g are asymptotically equivalent if $\lim_{x\to\infty} \frac{f(x)}{g(x)}$ exists and is equal to 1 [11].

Proof. Let M be a structural equation model with an arbitrarily large sample size. The likelihood function of M, $p(Y|\theta)$, depends on the sample size, whereas the density function of the prior distribution, $p(\theta)$, does not. Thus, in the case of large sample size, $\log p(Y|\theta)$ becomes very large while $\log p(\theta)$ does not increase. From Equation 3.4, it follows that the term $\log p(\theta)$ becomes irrelevant in the equation. Hence, the logarithm of the posterior density function $p(\theta|Y, M)$ becomes arbitrarily close to the log-likelihood function $\log p(Y|\theta, M)$. In the frequentist approach to SEM, the likelihood function is maximized to best fit the data, thus resulting in the same results as the log-likelihood function [17]. □

It follows that, for large sample sizes, the performances of the Bayesian and frequentist approach to SEM are similar. However, in the cases of small or moderate sample sizes, prior information about the parameter vector θ plays an important role in the Bayesian approach. In general, its performance is proven to be better, meaning that the goodness of fit of the model is improved [17].

3.3. Prior distributions

The characteristic feature of Bayesian statistics is the input of prior information based on subjective expert knowledge, previous research, or closely related data. Therefore, in Bayesian SEM, prior distributions for all model parameters must be specified. We let θ be the vector of all unknown model parameters of a SEM. In a Bayesian approach, these model parameters are considered to be random variables with a prior distribution. These prior distributions of the elements of θ represent the distribution of possible parameter values from which the parameters in θ have been drawn [13, 17]. We can distinguish three types of priors for θ: noninformative, weakly informative, and informative priors. In general, the issue of choosing prior input should be approached on a problem-by-problem basis [17].

3.3.1. Noninformative and weakly informative priors

In some cases, sufficient prior information may not be available beyond the data included in the analysis. In those situations, it is still possible to apply Bayesian inference by using a noninformative prior. The prior distributions of noninformative priors play a minimal role in the posterior distribution of θ. In these cases, the accuracy of Bayesian estimates is similar to that of the Maximum Likelihood estimates obtained by frequentist SEM [17].
However, when using noninformative priors, the Bayesian approach still resolves some of the problems related to maximum likelihood estimation. For example, it has been shown that in the context of SEM, the statistical properties of the maximum likelihood approach are not robust to small sample sizes [17, 34, 38]. In contrast, the sampling-based Bayesian methods depend less on asymptotic theory and hence have the potential to produce reliable results even with small samples [17, 34].
The limitations of maximum likelihood in parameter estimations for SEM have led researchers to turn to alternative estimation methods, in particular Bayesian estimation. Since prior distributions are an important component of Bayesian analysis, 'default' priors were constructed. The superior performance of Bayesian SEM in the case of small sample sizes or complex models motivates the use of noninformative priors as oppose to frequentist SEM.

Most software packages have default noninformative priors implemented [34], making it very easy to produce a Bayesian SEM without defining prior distributions. These priors are constructed in an automatic fashion, since the software has default prior distributions specified for every type of model parameter. These distributions are chosen such that they are best fit for a variety of SEMs typically encountered in practice [22].
A common view is that noninformative priors are a good starting point for conducting SEM analysis. After estimating the model, the results can be analyzed and the SEM could be improved further. Additionally, by comparing the goodness of fit of the model with noninformative priors and the model with informative priors, the quality of the prior information can be evaluated. Often, the priors for at least a part of the model parameters are default priors.
Most commonly used noninformative prior distributions are the uniform distribution over some sensible range of values or a normal distribution with a large variance [13, 26].

A special case of noninformative priors are weakly informative priors, suggested by [10] and [17]. If the sample size is large, a portion of the data (one third or less) can be used to produce Bayesian estimates by conducting a Bayesian SEM using noninformative priors on the smaller sample. With the remainder data (two thirds or more), one can conduct a Bayesian analysis with the found estimates as prior input. Then, the second BSEM that is performed on the remaining data, is an informative prior model. Often, this construction gives more accurate estimates than a noninformative prior model. Therefore, when there is no prior information available, the weakly approach is sometimes conducted and compared to the noninformative approach, to compare which method produces more accurate estimates for the specified model. The weakly informative prior method is not commonly used.

3.3.2. Informative priors

In case there is expert knowledge or previous research, informative prior distributions for model parameters can be formulated.

Definition 3.3.1. *An **informative prior** is a prior distribution for a model parameter with hyperparameters (e.g., the mean and variance) that reflect most closely the expected value and the degree of confidence about the expected value, based on previous research or expert knowledge. A small variance corresponds to considerable confidence in the prior knowledge, while a large variance reflects large uncertainty in the parameter value.*

The use of informative priors can improve the model estimates, since it brings additional information to the problem. That is the case, if the prior information is correct. If the priors are not correct, the estimates will become less accurate. The use of informative priors is advantageous if there is insufficient information available to solve the model by noninformative Bayesian or frequentist methods [26], or if there is reliable prior knowledge available.
In a Bayesian SEM, a prior distribution can be chosen for every model parameter with prior information. This research will focus on prior distributions for the estimates of the correlations between latent variables.

One type of informative prior is the conjugate prior distribution:

Definition 3.3.2. *A **conjugate prior distribution** yields a posterior distribution that is in the same distributional family as the prior distribution [11, 13, 17].*

To illustrate this, we give the following example [17].

Example 3.3.1. *We consider an arbitrary univariate binomial model, with sample size n. The likelihood of an observation y as a function of θ, is of the form*

$$p(y|\theta) = \binom{n}{y}\theta^{y}(1-\theta)^{n-y}. \tag{3.5}$$

If the prior density is of the same form, it can be seen from 3.3 that the posterior density will also be of this form. For example, we consider the beta distribution with hyperparameters α and β as a prior density for θ:

$$p(\theta) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1}. \tag{3.6}$$

The posterior density then becomes

$$\begin{aligned} p(\theta|y) &\propto p(y|\theta)p(\theta) && (3.7)\\ &\propto \theta^{y}(1-\theta)^{n-y}\theta^{\alpha-1}(1-\theta)^{\beta-1} && (3.8)\\ &= \theta^{y+\alpha-1}(1-\theta)^{n-y+\beta-1}, && (3.9) \end{aligned}$$

which is a beta distribution with parameters $y+\alpha$ and $n-y+\beta$. We see that indeed $p(\theta)$ and $p(\theta|y)$ are of the same form.

In the case of conjugacy, the posterior distribution is more likely to be analytically simple to solve. It could be argued that, since numerical simulation methods such as MCMC sampling (Markov chain Monte Carlo, section 3.4) can work with nonconjugate priors, nonconjugacy is less of a problem [13]. Even so, conjugate prior distributions provide better manageable posterior distributions for developing the MCMC algorithm and are therefore very commonly used [17].

3.4. Markov chain Monte Carlo methods

The estimation of Bayesian SEM relies on Markov chain Monte Carlo (MCMC) methods, also called Markov chain simulation methods. The general idea of these methods is that instead of analytically solving for the posterior distribution, MCMC obtains samples from the posterior distribution [13]. The methods are particularly useful in Bayesian SEM because of the focus on posterior distributions that are difficult to work analytically [35].
MCMC methods combine the Markov chain property and the Monte Carlo algorithm.

Definition 3.4.1. *A **Markov chain** is a sequence of random variables $\theta^1, \theta^2, ...$, in which the current state θ^t depends only on its direct predecessor θ^{t-1} [11], i.e.,*

$$P(X_{n+1} = x | X_1 = x_1, ..., X_n = x_1) = P(X_{n+1} = x | X_n = x_1). \tag{3.10}$$

The Markov chain property of MCMC is the idea that the random samples are generated in a sequential process, where each random sample is used as a starting point to generate the next random sample. In this chain, new samples do not depend on any samples other than the previous one.
If a Markov chain has the property that each state can be reached from every other state, it is called ergodic. In the case of ergodicity, the chain will asymptotically converge to its invariant distribution [35]. If the Markov chains that are used in a sampling algorithm are ergodic, it will not be trapped in some subset of the state space, and hence will asymptotically converge to its unique invariant distribution [4].

Monte Carlo refers to a broad collection of computation algorithms that can be used to approximate quantities based on generated random samples. Following this approach, the method can be used to estimate properties of a distribution by examining random samples from the distribution. This is advantageous since it is often easier to draw a large random sample of a distribution than to calculate properties directly from a distribution's equation. With a sufficiently large sample size, this approximation can be made as good as required [27, 35].
To illustrate the Monte Carlo method, we consider Buffon's needle problem, where we are interested in the probability of a needle of 10 cm crossing a line when it is dropped on a floor with parallel lines that are 10 cm apart. The exact value of this probability can be obtained analytically as $2 / \pi = 0.63662$. In a Monte Carlo approach, one would drop a needle 1,000 times to estimate the probability. If it crosses a line 641 times, then the Monte Carlo estimate of the probability would be 641 / 1,000 = 0.641.

Combining ergodic Markov chains and Monte Carlo leads to the MCMC method.

Definition 3.4.2. *A **Markov chain Monte Carlo** method for the simulation of a target distribution π is any method producing an ergodic Markov chain whose stationary distribution is π [33]. In the case of BSEM, π represents the posterior distribution of θ [35].*

MCMC is particularly useful in Bayesian inference because it allows us to estimate mathematical properties of posterior distributions that are often difficult to work with analytically. In application to Bayesian SEM, an MCMC method can be used to draw specially constructed samples from the posterior distribution $p(\theta|Y)$ to estimate the model parameters. This sampling is done iteratively, such that at each step of the process, we can better estimate the model parameters, since the sample size of independent draws becomes larger. The prior distribution $p(\theta)$ can be taken as a starting point for the MCMC algorithm [35].

3.4.1. Hamiltonian Monte Carlo

There are various MCMC methods, the basic and most commonly used methods are the Metropolis-Hastings algorithm and the Gibbs sampler, a special version of that algorithm [11]. A disadvantage of these methods is that the sequence of samples drawn from the jumping distribution, or proposal density, is a random walk. This results in the tendency to move slowly around the sample space, rather than moving around quickly as desired [4, 31]. For that reason, in Stan, the probabilistic programming language that is used in the application to PRIME research, described in chapter 4, the Hamiltonian Monte Carlo (HMC) method is used [32]. Hence, HMC is the method that is further explained here.

The approach used in HMC stems from the union of MCMC and molecular dynamics approaches, originating in the field of physics [7], resulting in terms as 'momentum' and 'position' for the variables. Using several ideas from physics, HMC is able to suppress the local random walk behavior in the Metropolis algorithm. This allows the algorithm to move much more rapidly through the target distribution, thus resolving some inefficiencies in the Gibbs sampler and Metropolis-Hastings algorithm. For this reason, the method became more and more used in statistical applications [4].

The goal of sampling with HMC is to draw from the posterior density $p(\theta|Y)$. HMC introduces auxiliary momentum variables ϕ_j for every component θ_j in the target space. In the algorithm, the jumping distribution for θ is determined largely by ϕ, allowing it to move faster through the target distribution. HMC draws from a joint density $p(\theta,\phi|Y) = p(\phi)p(\theta|Y)$, to compute the posterior density $p(\theta|Y)$ to a multiplicative constant [11]. In Stan, the momentum distribution $p(\phi)$ is a multivariate normal distribution $\phi \sim N(0, M)$, where covariance matrix M is specified as the unit metric. As a starting point for the algorithm, the prior can be used as the vector of initial parameters θ [32].

Hamiltonian Monte Carlo algorithm

HMC is an iterative method. The four steps of an iteration can be described as follows [11, 32]:

1. The momentum variable ϕ is updated with a random draw from its posterior distribution $N(0, M)$.
2. Both θ and ϕ are updated in L steps, scaled by a factor ϵ, that represents the step size. This process is called leapfrog integration, because the momentum updates are split into half-steps. The following steps are repeated L times:
   (a) The gradient of the log-posterior density of θ is used to make a half-step of ϕ:
   $$\phi \leftarrow \frac{1}{2}\epsilon\frac{d\log p(\theta|Y)}{d\theta}. \tag{3.11}$$
   (b) The momentum vector ϕ is used to update the position vector θ:
   $$\theta \leftarrow \theta + \epsilon M^{-1}\phi. \tag{3.12}$$
   (c) The gradient of θ is then again used to half-update ϕ:
   $$\phi \leftarrow \frac{1}{2}\epsilon\frac{d\log p(\theta|Y)}{d\theta}. \tag{3.13}$$
3. The parameter and momentum vectors at the start of the leapfrog process are labeled θ^{t-1} and ϕ^{t-1} and the values after the L steps as θ^* and ϕ^*. In this step, the *accept-reject criterion*, r, is computed:
   $$r = \frac{p(\theta^*|Y)p(\phi^*)}{p(\theta^{t-1}|Y)p(\phi^{t-1})}. \tag{3.14}$$
4. In the final step, the new value for θ is either rejected or accepted:
   $$\theta^t = \begin{cases} \theta^* & \text{with probability } \min(r,1) \\ \theta^{t-1} & \text{otherwise.} \end{cases} \tag{3.15}$$

   As ϕ immediately gets updated at the beginning of the next iteration, there is no need to keep track of it after step 3.

These iterations are repeated until approximate convergence.[2]

[2]Convergence is monitored by estimating the factor by which the scale of the current distribution for each scalar estimand might be reduced if the simulations were continued in the limit $n \to \infty$. The potential scale reduction, denoted $\hat{R}$, declines to 1 as $n \to \infty$. A high potential scale reduction indicates that proceeding with further simulations may improve our inference about the target distribution. Approximate convergence is reached if $\hat{R}$ is near 1 and the effective sample size is large enough for all quantities of interest [11].

3.5. Bayesian model evaluation and comparison

One of the main goals of SEMs is the evaluation of hypotheses about the relations among observed variables and latent variables. Hence, model evaluation and hypotheses testing are important topics of interest in the field of (Bayesian) structural equation modeling. The Bayesian approach to SEM offers various model evaluation and comparison statistics, that will be discussed in this section.

3.5.1. Model evaluation statistics

There are several Bayesian methods to assess the goodness of fit of a SEM. The method that is most commonly used is posterior predictive model checking (PPMC) [11, 17, 19]. This method is based on the following principle: if the model fits, replicated data generated under the estimated model should look similar to the observed data. PPMC investigates discrepancies between generated data under the posterior predictive distribution and the observed data through a discrepancy function [11]. When MCMC estimation is used to estimate model parameters, the PPMC discrepancy function can be specified to assess the differences between the observed data and and the expected values based on the model estimation, for every iteration of the Markov chain [9].

In the R package 'blavaan' (Section 4.1.4), the method `ppmc()` allows users to conduct a posterior predictive model check to assess the global and local fit of a SEM using any discrepancy function that is available in the frequentist 'lavaan' package [29]. This includes the RMSEA, SRMR, CFI and TLI values that are discussed in Section 2.2.4. Therefore, these are the fit indices that will be used in the PRIME application in Chapter 4.
As described in [22], PPMC analysis in 'blavaan' becomes less reliable in cases of highly informative priors. Since the most important model in our application contains informative priors, we can not draw very strong conclusions from the results of the posterior predictive model checks. In future versions of 'blavaan', prior-predictive model checking may be implemented, This could resolve the problems relating to informative priors [22].

3.5.2. Model comparison statistics

The issue of hypothesis testing can be considered as model comparison, because a hypothesis can be represented via a specific SEM. Model comparison statistics are therefore an important part of the analysis of a SEM. The current state of the art Bayesian model fit measurement methods are not completely satisfying, because different methods fail in various examples [11]. In this section, we give three model comparison methods, that are most recommended by [9] and [22] as they give reasonable results in the applications of SEM. These measures can be used to compare models that use the (exact) same observed data [11].

The first comparison statistic we consider is the Bayes factor.

Definition 3.5.1. *The **Bayes factor** for comparing two structural equation models M_1 and M_0 is defined as*

$$B_{10} = \frac{p(Y|M_1)}{p(Y|M_0)}. \tag{3.16}$$

The Bayes factor is a summary of evidence provided by the data in favor of M_1 as opposed to M_0, or in favor of M_0 as opposed to M_1 [11, 17]. In Table 3.1, we give the interpretation of the log-Bayes factor, since that is the factor that is approximated in model comparison functions of the 'blavaan' package.

Table 3.1: Interpretation of the log-Bayes factor [17].

log B_{10}	**Evidence against** M_0
< 0	Negative (supports M_0)
0 to 2	Not worth more than a bare mention
2 to 6	Positive (supports M_1)
6 to 10	Strong
> 10	Decisive

Since the Bayes factor is sensitive to the choice of prior distributions, it can be used to compare different sets of priors. As the log-Bayes factor increases from 0, we gain increasing support for the prior odds of M_1, as opposed to M_0 [22].

The second comparison statistic we will discuss is the Watabe-Akaike (or 'widely applicable') information criterion (WAIC). This measure seeks to characterize a model's predictive accuracy through analysis of log-likelihoods associated with individual observations. The method estimates the model's expected log pointwise predictive density, which is a measure of predictive accuracy, and then adds a correction for effective number of parameters to adjust for overfitting. [11, 22]. Lower values of WAIC imply higher predictive accuracy, but a WAIC value of a model only has meaning in comparison to another model. There is not a conventional interpretation of the difference in WAIC, as this depends on the statistical model, the true distribution, the prior distribution, the Markov chain Monte Carlo method, and the experimental fluctuation [11, 37]. In general, a larger difference implies a stronger model preference.

Definition 3.5.2. *The **Watabe-Akaike information criterion** (WAIC) is defined as*

$$WAIC = -2lppd + 2efp_{WAIC}, \tag{3.17}$$

where the first term relates to log-likelihoods of the observed data and the second term involves an effective number of parameters.

The first term, the log pointwise predictive density (lppd) of the observed data, is estimated via

$$lppd = \sum_{i=1}^{n} \log\left(\frac{1}{S}\sum_{s=1}^{S} f(y_i|\theta^s)\right), \tag{3.18}$$

where S is the number of posterior draws and $f(y_i|\theta^s)$ is the density of observation i with respect to the parameters sampled at iteration s.

The second term, the effective number of parameters, is estimated via

$$efp_{WAIC} = \sum_{i=1}^{n} var_s(\log f(y_i|\theta^s)), \tag{3.19}$$

where we compute a separate variance for each observation i across the S posterior draws [22].

The third and last comparison statistic we consider, is the leave-one-out cross-validation statistic (LOO). The LOO measure estimates the predictive density of each individual observation from a cross-validation standpoint. Cross-validation means analyzing the predictive density when taking out one observation at a time and using the remaining observations to update the prior [22]. Similar to WAIC, lower values of the LOO measure imply higher predictive accuracy, but this only has meaning in comparison to another model. In general, a larger difference implies a stronger model preference [11].

Definition 3.5.3. *The **leave-one-out measure** (LOO) is defined by*

$$LOO = -2\sum_{i=1}^{n} \log\left(\frac{\sum_{s=1}^{S} w_i^s f(y_i|\theta^s)}{\sum_{i=1}^{S} w_i^s}\right), \tag{3.20}$$

where the w_i^s are importance sampling weights based on the relative magnitude of individual i's density function across the S posterior samples [22].

4

Application to PRIME research

In this chapter, Bayesian structural equation modeling will be applied to educational research to illustrate the use of BSEM. The method will be applied to test hypotheses on links between latent variables. Furthermore, a comparison with frequentist SEM will be made.

The PRogramme of Innovation in Mathematics Education (PRIME) is part of the Interfaculty Teaching from the department of Applied Mathematics at TU Delft, responsible for redesigning mathematics courses for engineers. The goals of PRIME are the following [5]:

1. *Academic success*: to improve study results.
2. *Transfer*: to improve the connection between mathematics and engineering.
3. *Engagement*: to increase students' active participation in class and motivation for the topic.

To improve mathematics education in line with these goals, further insight in the dynamics between academic success and engagement is needed. In a recent study, data has been collected to investigate the relations between academic success, engagement and teacher behavior. As these quantities are latent variables that form a complex model, Bayesian SEM is a suitable approach for estimating the links between the latent variables.

4.1. Methodology

4.1.1. Research question

The research question to be answered is the following:

> What are the relationships between *perceived scaffolding and monitoring cues*, satisfaction and frustration of needs for *relatedness*, *autonomy* and *competence*, *autonomous*- and *controlled motivation*, and *academic performance*?

For clarity, the psychological terms used in the research question are defined as follows [1]. For every concept, we give a relating statement from the survey.

Perceived scaffolding and monitoring cues	Teacher behavior in relation to scaffolding[1] and monitoring of student's progress, as perceived by the student. "The lecturer encourages me to reflect on how I can improve on my assignments"
Relatedness	The student's feeling of relation and connection with others. "I feel that the fellow students and lecturers I care about also care about me."

[1]The alignment of instruction in support of learning, i.e. in support of students managing their own learning process [5].

Autonomy	The student's tendency to regulate their behavior. "I feel a sense of choice and freedom in the things I undertake."
Competence	The student's feeling of effectiveness in reaching valuable and useful results. "I feel confident that I can do things well."
Autonomous motivation	The student's tendency to perform behavioral and psychological activities without any external controls and dependencies. "I am studying because I want to learn new things."
Controlled motivation	The student's tendency to perform behavioral and psychological activities with external controls and dependencies. "I am studying because that's what others (parents, friends, etc.) expect me to do."
Academic performance	The student's performance expressed with a 10-point scale grade system.

4.1.2. Hypotheses

Several hypotheses were formulated for the research question. For clarity, a path diagram of the hypotheses can be found in Figure 4.1.

1. *Perceived teacher scaffolding and monitoring cues* are positively related to *competence*- and *relatedness* satisfaction and negatively related to *competence*- and *relatedness* frustration.
2. *Competence*-, *autonomy*- and *relatedness* satisfaction are positively related to *autonomous motivation* and negatively related to *controlled motivation*.
3. *Competence*-, *autonomy*- and *relatedness* frustration are positively related to *controlled motivation*. *Autonomy* frustration is negatively related to *autonomous motivation*. *Competence*- and *relatedness* frustration are negatively, but less strongly, related to *autonomous motivation*.
4. *Autonomous motivation* is positively related to *academic performance*. *Controlled motivation* is negatively related to *academic performance*.

4.1.3. Research participants and data collection

The data has been collected in previous research of the PRIME group. The research population is comprised of 220 first year Computer Science students that followed Calculus, one of the PRIME mathematics courses. To measure *academic performance*, the final grade for the studied course was used. This includes resit grades, because the research interest lies in student performance for an entire course. To measure all other latent variables of the research question, a questionnaire was distributed among the students. Participants were asked to indicate on a Likert scale whether they agreed with the statements (1 = strongly disagree, 5 = strongly agree). In this questionnaire, 30 questions were related to *perceived scaffolding and monitoring cues*, 24 questions were related to *competence*, *relatedness*, and *autonomy*, and 16 questions were related to *autonomous*- and *controlled motivation*. Of the questions relating to *perceived scaffolding and monitoring cues*, 16 questions were omitted because they had a low factor loading, meaning that the observed variables did not fit the latent variable they were linked to [25]. Thus, the total number of observed variables is 55.

4.1.4. Implementation software: R package 'blavaan'

To implement a Bayesian SEM, several software programs are available. Programs that are often used are WinBUGS, Mplus, and various packages within the R software environment. In this study, the R package 'blavaan' [21, 23] is used because it is open source, thus offering a wide range of possibilities for extensions in further research. Additionally, the software is easy to use for researchers with R experience. The package 'blavaan' models Bayesian SEMs via Stan [32], a probabilistic programming language for Bayesian inference with HMC sampling, and summarizes the results. It is an extension of the package lavaan, which models frequentist SEMs.

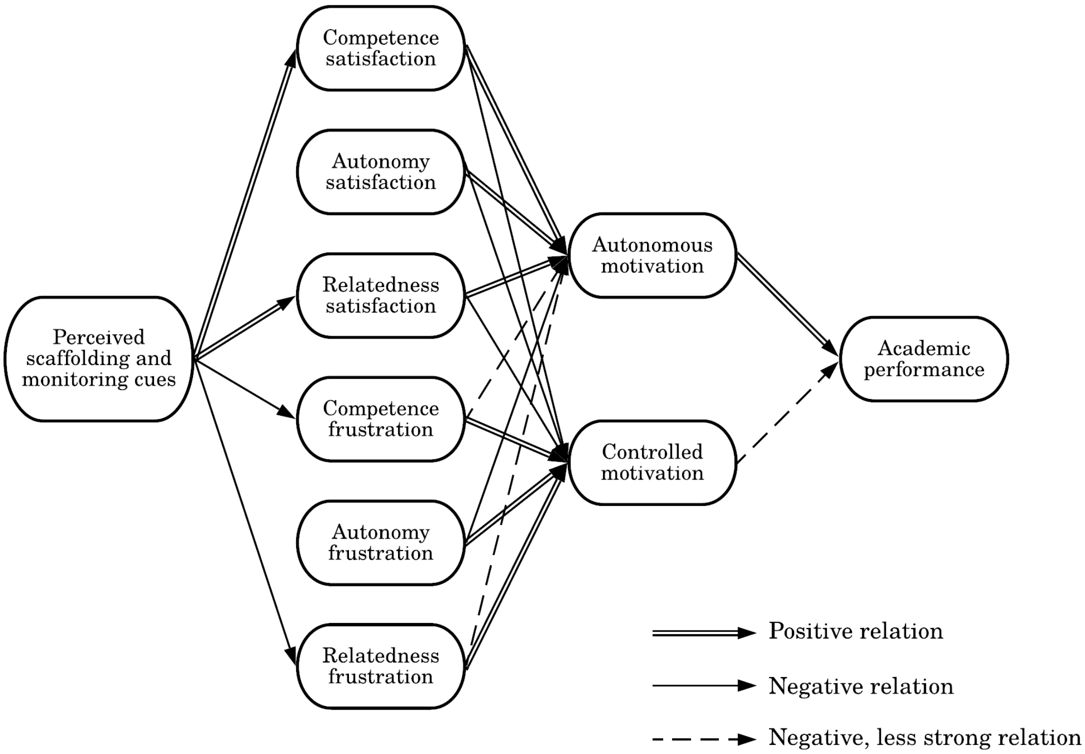


Figure 4.1: Path diagram corresponding to hypotheses.

4.2. Choosing informative prior distributions

In this study, previous psychological studies on the relations between *perceived scaffolding and monitoring cues*, basic psychological needs, motivation, and *academic performance*, were analyzed. Since there is information from previous research available, we have formulated informative priors.
In 'blavaan', the `prior()` modifier can be used to specify a prior distribution for a correlation. This is carried out by specifying a covariance parameter in the model syntax. As a result of some technical limitations of the package, the priors for correlations between latent variables should be beta distributions with support on (0, 1), which are then automatically converted to beta distributions with support on (−1, 1) [21]. In the Stan approach, the covariance matrix is consequently decomposed into standard deviations and correlations. The parameter matrix Θ is written as

$$\Theta = D_\Theta R_\Theta D_\Theta, \tag{4.1}$$

where D_Θ is the diagonal matrix of standard deviations and R_Θ is a correlation matrix. The prior distributions are placed individually on the free correlation parameters in the correlation matrix [23].

One difficulty lies in quantifying expert knowledge into well-founded priors; Literature often suggests a problem-to-problem approach. In this study, previous psychological studies, along with the opinion of a psychology researcher, have been used as prior information.
We use the correlation estimates from previous studies with the number of significance figures that was reported in the studies. If different research results were similar, meaning that the difference between all correlations is less than 0.15, we used the average of those correlations. We keep all significance figures that result from this computation, to ensure that the beta distributions are as precise as possible. If the difference between one correlation and the (average of the) other relation(s) is more than 0.15, we chose the correlations that align most with the hypotheses and discarded the rest. In these decisions, also the *p*-values of the correlations have been taken into account. A lower

p-value represents a more significant correlation estimate, meaning that the probability of the results being the outcome of random chance is lower [11]. This approach was constructed in collaboration with a psychology researcher.
For every correlation, we have computed a beta distribution supported on $(-1, 1)$ with a mean that reflects the outcome of previous research because we have taken the averages of previously found correlations. For simplicity, the variance of the beta distribution is set at 0.01 for all relations. This is a commonly used variance for a prior distribution that reflects relatively high confidence in the prior knowledge [13, 24]. Thus, the beta prior distributions corresponding to the specified mean and variance reflect the correlation we expect based on previous psychological studies and expert knowledge.

To compute the beta distributions, we analytically scale the mean and variance of the beta distribution supported on $(-1, 1)$ to correspond with a beta distribution supported on $(0, 1)$. The reason for this is that 'blavaan' only accepts beta distributions that are supported on (0,1), but we need positive and negative values to reflect the positive and negative correlations.
To scale the variables, we use the following property of the general beta distribution:

Theorem 4.2.1. *Suppose that Z has the standard beta distribution with left parameter $a \in (0, \infty)$ and right parameter $b \in (0,\infty)$. For $c \in \mathbb{R}$ and $d \in (0, \infty)$, the random variable $X = c + dZ$ has the beta distribution with left parameter a, right parameter b, location parameter c and scale parameter d. The mean and variance of X are [30]:*

$$\mathbb{E}(X) = c + d\frac{a}{a+b} \tag{4.2}$$

$$Var(X) = d^2\frac{ab}{(a+b)^2(a+b+1)}. \tag{4.3}$$

We see that to convert the scaled distribution supported on $(-1, 1)$ to the standard beta distribution supported on $(0, 1)$, we must have location parameter $c = \frac{1}{2}$ and scale parameter $d = \frac{1}{2}$. With Equations 4.2 and 4.3, we can update the mean and variance. If we then solve for a and b from the expressions for the mean and variance of the standard beta distributions [30]:

$$\mathbb{E}(X) = \frac{a}{a+b}, \tag{4.4}$$

$$\mathrm{Var}(X) = \frac{ab}{(a+b)^2(a+b+1)}, \tag{4.5}$$

we obtain the values for a and b that correspond to the required beta distribution that will be used as a prior. If we let μ be the expected value for X and σ^2 the variance of X, we have

$$a = \left(\frac{1-\mu}{\sigma^2}\right)\mu^2, \tag{4.6}$$

$$b = a\left(\frac{1}{\mu} - 1\right). \tag{4.7}$$

In our case, μ is the average of the correlations found in previous research, and $\sigma^2 = 0.01$. After the beta distribution corresponding to the specified a and b is given as input for `prior()`, function `bsem()` automatically converts this beta distribution into a beta distribution supported on $(-1, 1)$.

We have thus specified how to properly work with the means and variances, and will continue to specify the mean estimates that have been used in the model. The mean estimates and corresponding prior distributions can be found in Table 4.1.

4.2.1. Relation between perceived scaffolding and monitoring cues and needs satisfaction and -frustration

The links between the latent variable *perceived scaffolding and monitoring cues* and needs satisfaction and -frustration (combination of *competence*, *autonomy* and *relatedness*) are estimated by using [12] and [18]. In both studies, there is no distinction between the needs for *competence* and *relatedness*. Therefore, we have used the same estimates for the correlations with *competence*- and *relatedness* satisfaction and -frustration.
In [12], the relations between *perceived autonomy supported teaching* and *student needs satisfaction and -frustration* were studied. The correlations with the needs satisfaction and -frustration were .61 and −.26. In [18], the relations between *perceived need support in teacher behavior* and *student needs satisfaction and -frustration* were studied. These correlations were .64 and −.28. Since the results of both studies are very close, we have taken the average of the correlations as estimates for the prior distributions.

4.2.2. Relation between needs satisfaction and motivation

The relations between need satisfaction and *autonomous*- and *controlled motivation* have been estimated by using [1], [12], [14] and [18]. In [1], [12] and [18], the authors did not distinguish between different psychological needs. Therefore, we have used these results multiple times to compute the means for *competence*-, *autonomy*- and *relatedness* satisfaction. The correlations between needs satisfaction and *autonomous motivation* that these studies found, are .60, .70 and .59, respectively. For the relation between *competence* satisfaction and *autonomous motivation*, and *autonomy* satisfaction and *autonomous motivation*, [14] found the values −.274 and .461. Since both values differ more than .15 compared to the average of the previously mentioned values, we have discarded these correlations and have taken the average of .60, .70 and .59. For the relation between *relatedness* satisfaction and *autonomous motivation*, [14] found a value of .608. This is in line with the previously mentioned values, so we have taken the average of .60, .608, .70 and .59.

For the correlations between needs satisfaction and *controlled motivation*, we have used [12], [14] and [18]. The correlations between needs satisfaction and *controlled motivation* that [12] and [18] found, are −.02 and .29. For the relation between *competence* satisfaction and *controlled motivation*, and *relatedness* satisfaction and *controlled motivation*, [14] found the values −.452 and −.457. These values are more in line with our hypotheses and have lower *p*-values than the values mentioned previously. Therefore, these values are taken as mean estimates.
For the relation between *autonomy* satisfaction and *controlled motivation*, [14] found a value of .241. This is not in line with our hypothesis, but the value is similar to the result of [18] and both estimates have acceptable *p*-values. Therefore, we have taken the average of .241 and .29 for the mean estimate.

4.2.3. Relation between needs frustration and motivation

The relations between need frustration and *autonomous*- and *controlled motivation* are estimated by using [12] and [18]. The correlations between needs frustration and *autonomous motivation* that these studies found, are −.37 and −.25. We have taken the average of both values for the mean estimate. The correlations that these studies found for the relation between needs frustration and *controlled motivation* are .48 and .07. Because the former value is more in line with our hypothesis and has a significantly *p*-value, we have used only this value for the mean estimate.

4.2.4. Relation between motivation and academic performance

The relations between *autonomous*- and *controlled motivation* and *academic performance* are estimated by using [1], [15] and [20]. For the relation between *autonomous motivation* and *academic performance*, we have taken the average of the values that [1] and [15] found, which are .19 and .18. For the estimate of the correlation between *controlled motivation* and *academic performance*, we have used [15] and [20]. The resulting correlations of these studies are −.08 and −.12. We have taken the average of these values for the estimate of the mean.

All mean estimates and corresponding prior distributions can be found in Table 4.1.

Table 4.1: Prior distributions for the informative prior model. For every hypothesized relation, the average of correlations found in previous research was taken as a mean for a beta distribution with variance 0.01.

Prior	Causal direction	Previous correlations	Mean estimate	Prior distribution
1 2	*Perceived scaffolding and monitoring cues* ⟶ *Competence satisfaction* ⟶ *Relatedness satisfaction*	.61 ($p \leq .001$) .64 ($p < .01$)	.625	B(48.69922, 11.23828)
3 4	*Perceived scaffolding and monitoring cues* ⟶ *Competence frustration* ⟶ *Relatedness frustration*	−.26 ($p \leq .001$) −.28 ($p < .01$)	−.27	B(33.47415, 58.23585)
5	*Competence satisfaction* ⟶ *Autonomous motivation*	.60 ($p = .001$) −.274* ($p = .000$) .70 ($p \leq .001$) .59 ($p < .01$)	.63	B(48.33765, 10.97235)
6	*Autonomy satisfaction* ⟶ *Autonomous motivation*	.60 ($p = .001$) .461* ($p = .009$) .70 ($p \leq .001$) .59 ($p < .01$)	.63	B(48.33765, 10.97235)
7	*Relatedness satisfaction* ⟶ *Autonomous motivation*	.60 ($p = .001$) .608 ($p = .000$) .70 ($p \leq .001$) .59 ($p < .01$)	.6245	B(48.73498, 11.265)
8	*Competence satisfaction* ⟶ *Controlled motivation*	−.02* ($p > .05$) −.452 ($p = .000$) .29* ($p < .01$)	−.452	B(21.52807, 57.04153)
9	*Autonomy satisfaction* ⟶ *Controlled motivation*	−.02* ($p > .05$) .241 ($p = .014$) .29 ($p < .01$)	.2655	B(58.18198, 33.769)
10	*Relatedness satisfaction* ⟶ *Controlled motivation*	−.02* ($p > .05$) −.457 ($p = .000$) .29* ($p < .01$)	−.457	B(21.20825, 56.90685)
11 12 13	*Competence, autonomy and relatedness frustration* ⟶ *Autonomous motivation*	−.37 ($p \leq .001$) −.25 ($p < .01$)	−.31	B(30.83955, 58.55045)
14 15 16	*Competence, autonomy and relatedness frustration* ⟶ *Controlled motivation*	.48 ($p \leq .001$) .07* ($p > .05$)	.48	B(56.2104, 19.7496)
17	*Autonomous motivation* ⟶ *Academic performance*	.19 ($p = .003$) .18 ($p < .01$)	.185	B(56.62967, 38.94783)
18	*Controlled motivation* ⟶ *Academic performance*	−.08 ($p \geq .01$) −.12 ($p < .05$)	−.10	B(44.1,53.9)

*Note: these values were discarded in the computation of the corresponding mean estimate, because they differ more than 0.15 with the other values.

4.3. Implementation of the informative prior model

To specify the informative prior model, we have used the R package ‘blavaan’ (see section 4.1.4). The syntax of this package is shown in Table 4.2.

Table 4.2: ‘blavaan’ and ‘lavaan’ model syntax [29].

Formula type	Operator	Meaning
latent variable definition	=~	is measured by
regression	~	is regressed on
(residual) (co)variance	~~	is correlated with

We define the measurement part of the model by linking the latent variables to the observed variables. One of the endogenous latent variables, academic performance, is defined by the final grade of the studied course. The other endogenous latent variables are defined by questions from a survey (see Section 4.1.3), for example variables SM1 and Needs1 contain answers to the corresponding questions. The only exogenous latent variable, *perceived scaffolding and monitoring cues* (SMC), is also defined by items from a survey. These definitions reflect the relations that are of interest in this educational research. For every latent variable, the factor loading of the first observed variable is automatically set to 1 by the function `bsem()` in ‘blavaan’, for scaling purposes. In the structural part of the model, the relations between latent variables as given in Section 4.1.2 are defined. The measurement and the structural parts of the model in R code can be seen below.

```
#measurement part

SMC =~ SM1 + SM3 + SM4 + SM11 + SM12 + SM13 + SM14 + SM15 + SM16 + SM18 + SM19
       + SM20 + SM21 + SM26
autonomy_satisfaction  =~ Needs1 + Needs7 + Needs13 + Needs19
autonomy_frustration =~ Needs2 + Needs8 + Needs14 + Needs2
competence_satisfaction =~ Needs5 + Needs11 + Needs17 + Needs23
competence_frustration =~ Needs6 + Needs12 + Needs18 + Needs24
relatedness_satisfaction =~ Needs3 + Needs9 + Needs15 + Needs21
relatedness_frustration =~ Needs4 + Needs10 + Needs16 + Needs22
autonomous_motivation =~ Motivation2 + Motivation4 + Motivation7 + Motivation8
       + Motivation11 + Motivation13 + Motivation15 + Motivation16
controlled_motivation =~ Motivation1 + Motivation3 + Motivation5 + Motivation6
       + Motivation9 + Motivation10 + Motivation12 + Motivation14
academic_performance =~ Grade

#structural part

relatedness_satisfaction ~ SMC
relatedness_frustration ~ SMC
competence_satisfaction ~ SMC
competence_frustration ~ SMC
autonomous_motivation ~ competence_satisfaction + competence_frustration
       + autonomy_satisfaction + autonomy_frustration + relatedness_satisfaction
       + relatedness_frustration
controlled_motivation ~ competence_satisfaction + competence_frustration
       + autonomy_satisfaction + autonomy_frustration + relatedness_satisfaction
       + relatedness_frustration
academic_performance ~ autonomous_motivation + controlled_motivation
```

To specify the priors, we have used the ‘blavaan’ function `prior()` and the prior distributions as defined in Section 4.2. The prior specification of the model in R code can be found in Appendix A.1, and the used function to compute the beta distribution in Appendix C. For illustration, the beta distribution that corresponds to the first prior, that has a mean of .625, is visualized in Figure 4.2.

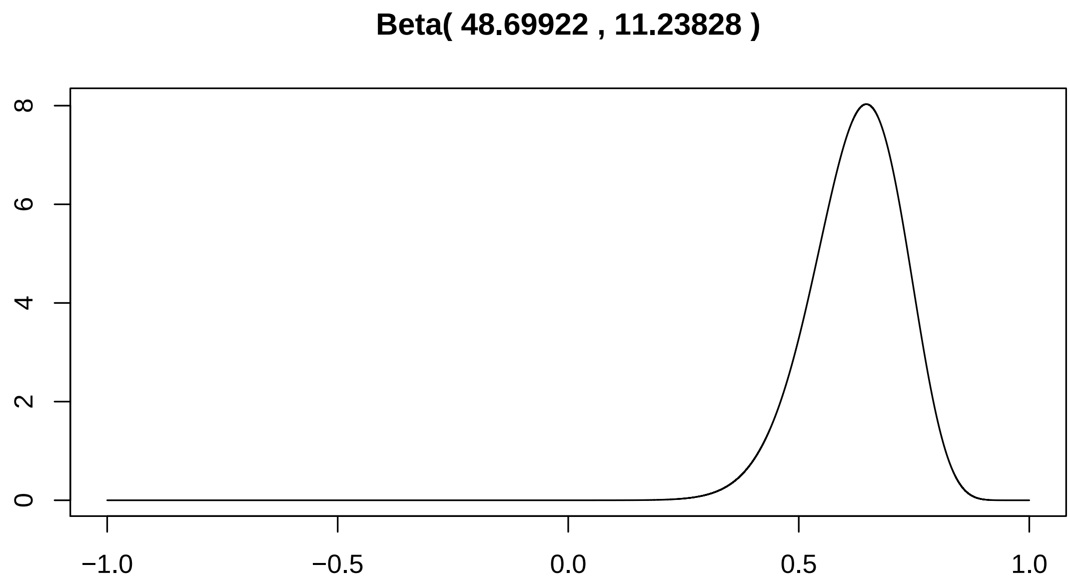


Figure 4.2: Beta distribution corresponding to the first prior, that has a mean of .625.

4.3.1. Model identification and estimation

For model identification, we check the t-rule, as defined in Equation 2.7. In the model, we have 55 observed variables (Section 4.1.3) and 18 relations (Section 4.1.2), hence $p = 55$ and $z = 18$. We see that the model passes the t-test:

$$t = 2(55) + (18) = 128 \leq 1540 = \frac{1}{2}(55)(55+1), \tag{4.8}$$

Since the model is identified, we can let 'blavaan' fit the model. A path diagram of the estimated links with the computed estimates can be seen in Figure 4.3. Before we can interpret and report these results, we have to evaluate the model to see if the fit is acceptable.

4.3.2. Model evaluation

The occurrence of standardized relation estimates greater than one, as is the case for the relation between *perceived scaffolding and monitoring cues* and *competence* satisfaction, and between *competence* frustration and *controlled motivation* (Figure 4.3), raises questions concerning the legitimacy of those estimates. However, it was demonstrated in [6] that these can legitimately occur. Here, it must first be noted that 'blavaan' uses standardized regression coefficients as estimates for the correlation coefficients. Standardized regression coefficients are analogous to correlation coefficients, but instead of correlations they are in fact rates of change. Consequently, they are not numerically bounded by ± 1 [6]. Furthermore, in [6], it is strongly advised not to modify models simply to reduce the presence of standardized coefficients greater than one, to prevent the biasing effects of such model modifications. Therefore, we interpret the estimates greater than one as very strong correlations. In addition, these estimates may indicate that the latent variables are specified in a too dependent way. Further analysis of the correlation estimates greater than 1 is outside the scope of this study.

To evaluate the goodness of fit of the model, we assess the fit indices as described in Section 3.5. To evaluate the prior input, we also estimate the noninformative prior model, which is the same model, without the priors, and assess its goodness of fit. Note that the model specification is equal for both the informative prior model and the noninformative model. The goodness of fit indices of the informative

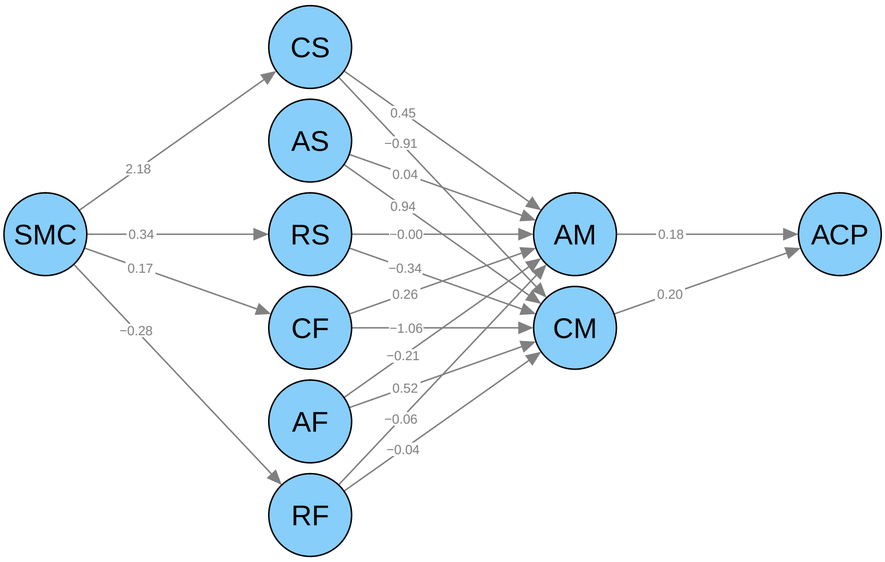


Figure 4.3: Path diagram corresponding to the informative prior model. The standardized relation estimates are shown on the edges. For simplicity, the diagram shows the latent variables without the observed variables.

prior model and the noninformative prior model can be found in Tables 4.3 and 4.4.
We see that, for the informative prior model, the RMSEA, SRMR, WAIC and LOO values are lower, and the CFI and TLI values are higher. This means that all fit measures indicate that the estimates of the informative prior model are more accurate, suggesting that the priors have indeed improved the model fit. However, the log-Bayes factor of the informative prior model compared to the noninformative prior model is −278.060, thus the Bayes factor strongly favors the noninformative prior model. Although the fit indices show conflicting preferences, the majority of the fit indices favor the informative prior model. Hence, we conclude that the priors have improved the model fit.

Upon further analysis of the fit indices, we see that the RMSEA value is 0.072, suggesting a fair model fit. The SRMR value, however, is 0.125, which is above the limit of acceptable model fit of 0.10 and thus suggests a poor fit. The CFI and TLI measures also suggest poor model fit: both values, 0.714 and 0.695, are below the lower limit of 0.9 that shows a good fit. Note that the WAIC and LOO indices are comparison indices, and do not indicate objective model fit.
The SRMR, CFI and TLI indices indicate that the model needs further improvement. Since ‘blavaan’ does not yet contain the functionality of giving modification indices, as is available in its frequentist counterpart ‘lavaan’, we decided not to proceed with the analysis of the informative prior model.

Table 4.3: Values of the fit indices of the informative prior model.

Fit index	Value
RMSEA	0.072
SRMR	0.123
CFI	0.713
TLI	0.693
WAIC	27453.275
LOO	27450.187

Table 4.4: Values of the fit indices of the noninformative prior model.

Fit index	Value
RMSEA	0.075
SRMR	0.132
CFI	0.687
TLI	0.670
WAIC	27575.458
LOO	27575.944

4.4. Implementation of the weakly informative prior model

If there is no prior information available, and the sample is large enough, one can formulate a weakly informative prior model (Section 3.3.1). In this approach, part of the sample is used to specify priors. The rest of the sample is used as input for the observed variables. We will apply this method to the PRIME research for illustration and comparison purposes.
We use one third of the (unordered) PRIME data as input for a Bayesian SEM with noninformative priors. Note that the model specification and identification is the same as the noninformative prior model of Section 4.3. The estimated model that we use for the priors can be found in Figure 4.4.

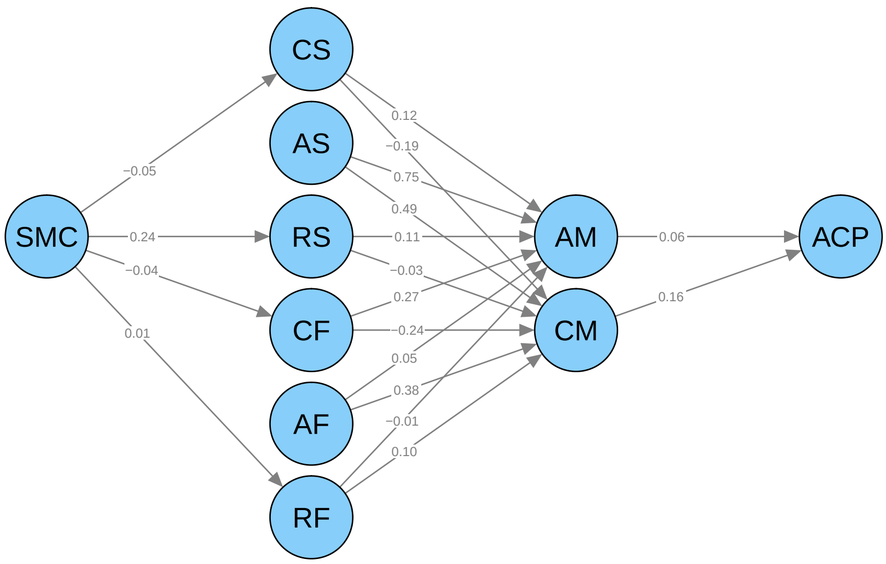


Figure 4.4: Path diagram corresponding to the noninformative prior model, using one third of the sample. The standardized relation estimates are shown on the edges. For simplicity, the diagram shows the latent variables without the observed variables.

The parameter estimates that result from this model, are taken as mean for the beta prior distributions, following the same approach as given in Section 4.2. The variance of the beta distributions will again be 0.01 for all priors. The computed weakly informative prior distributions can be found in Table 4.6. We update the model with the weakly informative priors and run the model using the remaining two thirds of the sample. We cannot directly compare the weakly informative prior model through the WAIC and LOO measures or the Bayes factor, because the sample sizes differ.
The fit indices of the informative prior model and the weakly informative prior model, given in Table 4.5 and Table 4.6, indicate that the fit of the weakly informative model is less good. This is as expected since the informative priors add new data to the model. If this new information is correct, the Bayesian inference improves since it can use more data [26]. Since the estimates are less reliable than the ones we found previously, they are not given here. The R code of the weakly informative approach can be found in Appendix A.3.2.

Table 4.5: Values of the fit indices of the informative prior model.

Fit index	Value
RMSEA	0.072
SRMR	0.123
CFI	0.713
TLI	0.693

Table 4.6: Values of the fit indices of the weakly informative prior model.

Fit index	Value
RMSEA	0.085
SRMR	0.130
CFI	0.635
TLI	0.610

Table 4.7: Prior distributions for the weakly informative prior model. For every hypothesized relation, the correlation that was found in estimating a noninformative prior model using one third of the sample was taken as a mean for a beta distribution with variance 0.01.

Prior	Causal direction	Mean estimate	Prior distribution
1	*Perceived scaffolding and monitoring cues* $\longrightarrow$ *Competence satisfaction*	−.05	B(46.90625, 51.84375)
2	*Perceived scaffolding and monitoring cues* $\longrightarrow$ *Relatedness satisfaction*	.24	B(57.8088, 35.431)
3	*Perceived scaffolding and monitoring cues* $\longrightarrow$ *Competence frustration*	−.04	B(47.4432, 51.3968)
4	*Perceived scaffolding and monitoring cues* $\longrightarrow$ *Relatedness frustration*	.01	B(49.98995, 49.00005)
5	*Competence satisfaction* $\longrightarrow$ *Autonomous motivation*	.12	B(54.6336, 42.9264)
6	*Autonomy satisfaction* $\longrightarrow$ *Autonomous motivation*	.75	B(37.40625, 5.34375)
7	*Relatedness satisfaction* $\longrightarrow$ *Autonomous motivation*	.11	B(54.27345, 43.51655)
8	*Competence satisfaction* $\longrightarrow$ *Controlled motivation*	−.19	B(38.63295, 56.75705)
9	*Autonomy satisfaction* $\longrightarrow$ *Controlled motivation*	.49	B(55.86755, 19.122451)
10	*Relatedness satisfaction* $\longrightarrow$ *Controlled motivation*	−.03	B(47.97135, 50.93865)
11	*Competence frustration* $\longrightarrow$ *Autonomous motivation*	.27	B(58.23585, 33.47415)
12	*Autonomy frustration* $\longrightarrow$ *Autonomous motivation*	.05	B(51.84375, 46.90625)
13	*Relatedness frustration* $\longrightarrow$ *Autonomous motivation*	−.01	B(49.00005, 49.98995)
14	*Competence frustration* $\longrightarrow$ *Controlled motivation*	−.24	B(57.8088, 35.4312)
15	*Autonomy frustration* $\longrightarrow$ *Controlled motivation*	.38	B(58.3464 , 26.2136)
16	*Relatedness frustration* $\longrightarrow$ *Controlled motivation*	.10	B(53.9, 44.1)
17	*Autonomous motivation* $\longrightarrow$ *Academic performance*	.06	B(52.2792, 46.3608)
18	*Controlled motivation* $\longrightarrow$ *Academic performance*	.16	B(55.9352, 40.5048)

4.5. Comparison to the frequentist approach

To investigate whether Bayesian SEM offers more reliable estimates in the PRIME application, we have also conducted a frequentist model analysis using the same model setup and sample. Note that, since the model specification is the same as the previously specified models, the frequentist model remains identified. To model the frequentist SEM, we have used R package 'lavaan'. The model correlation estimates can be found in Figure 4.5, and the fit indices can be found in Table 4.9. For comparison, we give the fit indices of the Bayesian approach of the informative prior model in Table 4.8. The R code for the frequentist model can be found in Appendix A.4.

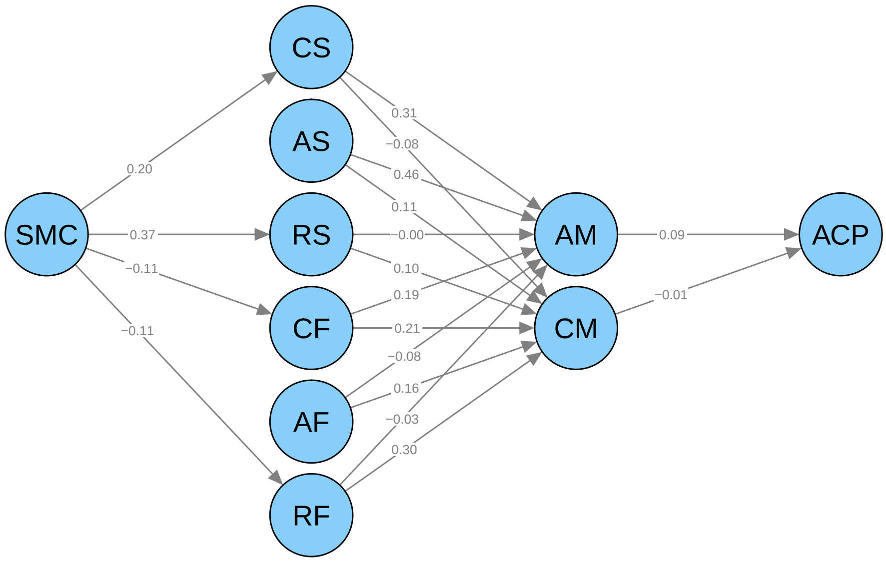


Figure 4.5: Path diagram corresponding to the frequentist prior model. The standardized relation estimates are shown on the edges. For simplicity, the diagram shows the latent variables without the observed variables.

Table 4.8: Values of the fit indices of the informative prior model.

Fit index	Value
RMSEA	0.072
SRMR	0.123
CFI	0.713
TLI	0.693

Table 4.9: Values of the fit indices of the frequentist model.

Fit index	Value
RMSEA	0.071
SRMR	0.120
CFI	0.719
TLI	0.703

We see that, unexpectedly, the frequentist SEM gives slightly more accurate parameter estimates according to the model evaluation measures. The frequentist RMSEA and SRMR values are lower and the CFI and TLI values are higher, indicating a better fit, although the differences are very small. This is not in line with previous research [17]. An explanation can be found in inadequacies in the goodness of fit indices. Further analysis is needed to fully account for the discovered performance differences between the Bayesian and frequentist approach to SEM.

4.6. Results

Although some of the fit indices indicate that the estimates obtained through the Bayesian informative prior approach are not reliable, we have seen that the informative priors have improved the noninformative model. We will therefore report and interpret the estimated correlations for the hypothesized relations that were obtained with the informative prior model. The estimates can be found in Figure 4.6 and Table 4.10. They are discussed further in Section 4.6.1.

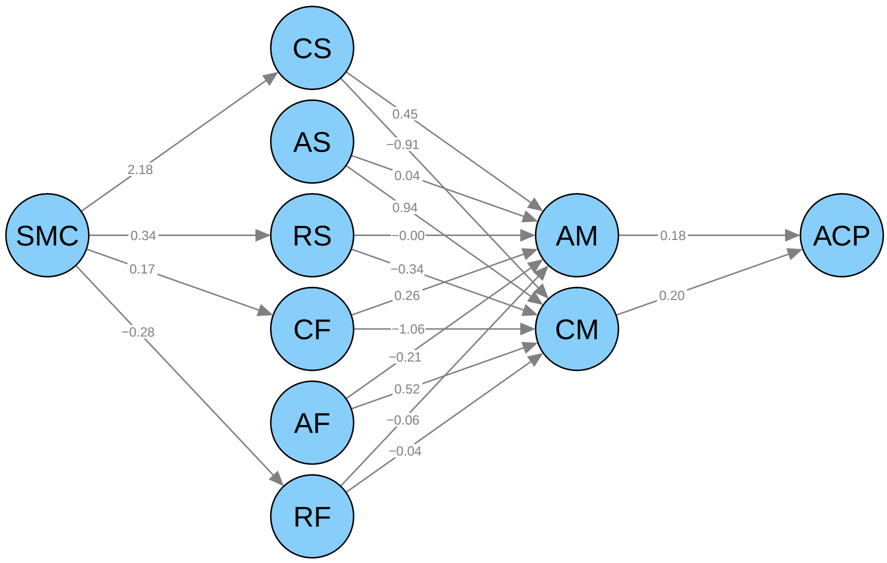


Figure 4.6: Path diagram corresponding to the informative prior model. The relation estimates are shown on the edges. For simplicity, the diagram shows the latent variables without the observed variables.

4.6.1. Interpretation of the estimates

To answer the research question formulated in Section 4.1.1, we will look at the hypotheses model and check whether the BSEM correlation estimates align with the hypotheses or not.

1. *Perceived teacher scaffolding and monitoring cues* are positively related to *competence*- and *relatedness* satisfaction and negatively related to *competence*- and *relatedness* frustration.

We see that *perceived scaffolding and monitoring cues* is indeed positively related to *competence*- and *relatedness* satisfaction and negatively to *relatedness* frustration. In particular, the relation between *perceived scaffolding and monitoring cues* and *competence* satisfaction was found to be very strong. However, the relation with *competence* frustration was found to be positive, where a negative relation was expected. Thus, not all results are in line with the first hypothesis.

2. *Competence*-, *autonomy*- and *relatedness* satisfaction are positively related to *autonomous motivation* and negatively related to *controlled motivation*.

For the second hypothesis, not all results are in line with the expected relations. *Competence*- and *autonomy* satisfaction are indeed positively related to *autonomous motivation*, although the found link between *autonomy* satisfaction and *autonomous motivation* is very weak. The relations between *competence*- and *relatedness* satisfaction and *controlled motivation* are negative, as expected.
However, for *relatedness* satisfaction and *autonomous motivation*, a relation was not detected. Furthermore, *autonomy* satisfaction is in fact quite strongly positively related to *autonomous motivation*, while a negative relation was expected.

Table 4.10: Standardized relation estimates of the informative prior model for all links investigated, alongside the mean estimates that were used to formulate the prior distributions.

Prior	**Relation**	**Mean estimate**	**Standardized relation estimate**
1	CS ~ SMC	.625	.36
2	RS ~ SMC	.625	.59
3	CF ~ SMC	−.27	−.07
4	RF ~ SMC	−.27	−.78
5	AM ~ CS	.63	.60
6	AM ~ AS	.63	−.76
7	AM ~ RS	.6245	.20
8	CM ~ CS	−.452	.10
9	CM ~ AS	.2655	.19
10	CM ~ RS	−.457	−.88
11	AM ~ CF	−.31	−.08
12	AM ~ AF	−.31	−.19
13	AM ~ RF	−.31	.85
14	CM ~ CF	.48	.10
15	CM ~ AF	.48	.00
16	CM ~ RF	.48	−.21
17	ACP ~ AM	.185	.05
18	ACP ~ CM	−.10	.19

3. *Competence*-, *autonomy*- and *relatedness* frustration are positively related to *controlled motivation*. *Autonomy* frustration is negatively related to *autonomous motivation*. *Competence*- and *relatedness* frustration are negatively, but less strongly, related to *autonomous motivation*.

For the third hypothesis, we also see that not all results are in line with the expected relations. For *autonomy*- and *relatedness* frustration and *autonomous motivation*, we indeed found positive, yet not very strong relations. The relation between *autonomy* satisfaction and *controlled motivation* was found to be positive, as hypothesized.
However, for *competence* frustration and *autonomous motivation*, we found a weakly positive relation where a negative one was expected. The relation between *competence* frustration and *controlled motivation* was found to be very strongly negative, and the relation between *relatedness* frustration and *controlled motivation* weakly negative, where two positive relations were expected. This is not in line with the third hypothesis.

4. *Autonomous motivation* is positively related to *academic performance*. *Controlled motivation* is negatively related to *academic performance*.

We see that *autonomous motivation* is positively related to *academic performance*, as expected, but the relation we found is not very strong. *Controlled motivation* and *academic performance* are positively related, where a negative relation was expected.

In conclusion, some, but not all of the relation estimates correspond to our hypotheses. Further psychological interpretation of the estimates is outside the scope of this study.

5
Conclusion

In this research, several statistical models for estimating relations between latent variables have been considered. We have given an overview of the statistical method of structural equation modeling, as well as the Bayesian approach to structural equation modeling. To illustrate the use of these methods and compare the methods, various variations of Bayesian and frequentist SEM were applied to a study in educational sciences.

The first method for estimating links between latent variables we discussed is frequentist structural equation modeling, which consists of five steps. We saw that the specification of a SEM consists of two parts: the relations among latent variables are defined in the structural model, and the latent variables are linked to the observed variables in the measurement model. For model identification, there is an identification condition that must be met in order to obtain reliable results. The model estimation is done by the Maximum Likelihood fitting function that minimizes the difference between the sample covariance matrix and the model estimated covariance matrix. After that, the model can be evaluated by measures of absolute or relative fit. If the model fit is unsatisfactory, one can modify the model to improve the model fit. These are the five steps that summarize the process of frequentist SEM.

The second method for estimating links between latent variables that was considered is Bayesian structural equation modeling. This method resolves some of the problems relating to Maximum Likelihood Estimation and is more accurate in the case of small samples or complex model structures. The characteristic feature of Bayesian analysis is the input of prior information. We saw that, if no prior information is available, we can perform Bayesian estimations using noninformative or weakly informative prior distributions. In cases where there is subjective expert knowledge, previous research or closely related data, one can formulate informative prior distributions. We saw that it is best to use conjugate prior distributions, because they provide better manageable posterior distributions for the MCMC algorithm, on which the estimation of Bayesian SEM relies. MCMC draws samples from the posterior distribution to estimate the model parameters. The implementation software used in this research makes use of Hamiltonian Monte Carlo, a MCMC method that resolves some disadvantages of the more commonly used Metropolis-Hastings algorithm and the Gibbs Sampler.

To illustrate the use of Bayesian structural equation modeling, we applied the method to a study in educational science to estimate relations among latent variables such as motivation and needs satisfaction. We used previous psychology research to choose prior distributions, in order to specify an informative prior model. Because of limitations in the used R package, we specified Beta distributions, with a mean that is the average of the previously found correlations and a variance of 0.01, to reflect rather high confidence in the priors.
The model evaluation indices suggest that the model does not fit the data well. In comparison to the noninformative prior model, the fit measures indicate that the informative prior model is a better fit to the data than the noninformative prior model. However, the Bayes factor shows a strong preference for the noninformative prior model as opposed to the informative prior model. Although the model evaluation and comparison statistics show conflicting results, most fit indices are in favor of the

informative model. Hence, we conclude that the priors are a good addition to the model, but that the hypothesized relations do not represent the data. The conjecture is that the number of latent variables that are specified in the model is too large for accurate estimations.
To illustrate the use of weakly informative priors, we applied this method to the same educational research model. We saw that, as would be expected, the fit indices of the weakly informative prior model imply that the model fit is less good than the fit of the informative prior model, again advocating the correctness of the informative priors.
In comparison to frequentist SEM, we saw that the model evaluation measures marginally favor frequentist SEM as opposed to the Bayesian model in terms of model accuracy. This is an unexpected result; an explanation could be found in inadequacies of the goodness of fit indices. Further analysis is needed to account for the discovered performance difference in this application.

6
Discussion

Structural equation modeling and Bayesian analysis are both substantially broad and well-known subjects, and this research investigates the recently developed combination of these two subjects. In this chapter, some of the complications and limitations of the subjects addressed in this thesis will be discussed. In addition, we will formulate some recommendations for future research.

6.1. Limitations of implementation software

The R package that is used in this research, 'blavaan', has some limitations as it is still in development. One of these limitations is that the prior distributions for the relations between latent variables are currently limited to Beta distributions, for computational reasons. Therefore, we have not investigated other options for prior distributions. For future research, in using different implementation software or as 'blavaan' further develops, it is recommended to examine different distributions to specify the priors more accurately.
In evaluating the informative prior model of our application, the model evaluation statistics indicated a poor model fit of the hypothesized relations to the data. However, a poor fit does not necessarily mean that the defined relation model was incorrect. That is, models with a poor model fit, but that are based on a solid theoretical framework, can generally be improved by analyzing the modification indices and re-specifying the model accordingly. However, this function is not yet available in 'blavaan'.
If these functionalities are not added within considerate time, we recommend looking into the commercial software Mplus, as it offers these and a number of other possibilities for model specification and evaluation.

Furthermore, in this research, we have used the standardized regression coefficients that 'blavaan' gives to estimate relation estimates. In 'blavaan' documentation, it is not entirely clear if that is indeed the correct interpretation of those coefficients. The 'blavaan' model output also gives standardized covariance coefficients, and some of the standardized regression coefficients we found are larger than one. Future research is recommended to look at the appropriate interpretation of the coefficients in the 'blavaan' output.

6.2. Model evaluation and comparison statistics

To test the measure of fit of the models, we used several model evaluation and comparison statistics. However, there are a number of situations in which these measures do not function appropriately. One of these situations is the case of highly informative prior distributions. Additionally, there exists a lot of controversy about measures of fit in general. There is no conventional approach as to which information criteria or fit indices provide the most reliable evaluation, as its performance is dependent on the statistical model, the true distribution, the prior distribution and the MCMC method, and possible experimental fluctuations. The model evaluation part of SEM is a subject that requires further research.

Furthermore, since hypotheses were tested through informative priors, it is recommended to estimate the model with different priors, to further evaluate the effect of the priors in general. Even though the priors in our application improved the accuracy according to the fit measures, a sensitivity analysis for the prior effects is recommended to confirm that the priors do not point the results in a desired (and biased) direction.

6.3. Hypothesized relations in PRIME

We concluded that the current specified informative prior model does not represent the observed data. Therefore, the correlation estimates that we found may not be reliable. An improvement in the accuracy of the model fit could be realized by altering the hypothesized relation model, or by reducing the number of latent variables to make the model less complex. We conjecture that a model that involves fewer latent variables could improve the model fit. For both suggestions, insight in the psychological concepts that are researched in the application is essential. These approaches are not further investigated in this thesis since it is out of scope.
Another approach to improving the model is to add mediator variables to account for a possible mediation effect, also known as the indirect effect. A latent variable could influence another latent variable indirectly, through an intermediate variable. These effects are not taken into account in this thesis, but could be added to make the model estimates more reliable.

In addition, through hierarchical and multisample data, the structural equation model could be further specified [17]. Often in substantive studies, as is in our application, the data is drawn from a number of different groups (clusters) with a known hierarchical structure. The PRIME data that is available for this research contained the following groups: gender, age, country of highest education, mathematics high school performance and lecturer. As individuals within a group share certain common influential factors, the random observations are correlated. The assumption of independent observed variables is in fact incorrect. Further research to combine models that analyze subpopulations, such as Latent Profile Analysis, and multisample SEM, could increase the accuracy of the results.

6.4. Comparison Bayesian and frequentist SEM

Unexpectedly, the results of the frequentist SEM appeared to be more accurate than the results of the Bayesian SEM, according to the fit indices used. We have already mentioned flaws in the evaluation statistics and suggested improvements for the (Bayesian) model specification. It is possible that after these alterations, the Bayesian approach indeed functions better than the frequentist approach, as would be expected based on literature. However, to explain the currently found differences in performance remains a subject where future research is needed.

6.5. Automated SEM strategies

Although the field of Bayesian structural equation modeling is still in development and has a lot of improvement potential, an alternative approach is offered by automated structural equation modeling. Following the trend of developing and applying automated strategies for the analysis of data sets, such as data mining and machine learning, research on automated SEM is currently exploring algorithms that (automatically) examine models that have some missing paths or parameters involved in unnecessarily restrictive constraints [13]. The combination of the Bayesian and automated approach to SEM could be a very interesting subject for future research.

A

R code: Model implementation

A.1. Informative prior model

```
library(blavaan)
library(readxl)

Data <- read_excel("Data corrected.xlsx")

model_ip<-'

#measurement part

SMC =~ SM1 + SM3 + SM4 + SM11 + SM12 + SM13 + SM14 + SM15 + SM16 + SM18 + SM19
       + SM20 + SM21 + SM26

autonomy_satisfaction  =~ Needs1 + Needs7 + Needs13 + Needs19
autonomy_frustration =~ Needs2 + Needs8 + Needs14 + Needs2
competence_satisfaction =~ Needs5 + Needs11 + Needs17 + Needs23
competence_frustration =~ Needs6 + Needs12 + Needs18 + Needs24
relatedness_satisfaction =~ Needs3 + Needs9 + Needs15 + Needs21
relatedness_frustration =~ Needs4 + Needs10 + Needs16 + Needs22

autonomous_motivation =~ Motivation2 + Motivation4 + Motivation7 + Motivation8
       + Motivation11 + Motivation13 + Motivation15 + Motivation16
controlled_motivation =~ Motivation1 + Motivation3 + Motivation5 + Motivation6
       + Motivation9 + Motivation10 + Motivation12 + Motivation14

academic_performance =~ Grade

#structural part

relatedness_satisfaction ~ SMC
relatedness_frustration ~ SMC
competence_satisfaction ~ SMC
competence_frustration ~ SMC

autonomous_motivation ~ competence_satisfaction + competence_frustration
       + autonomy_satisfaction + autonomy_frustration + relatedness_satisfaction
       + relatedness_frustration
controlled_motivation ~ competence_satisfaction + competence_frustration
       + autonomy_satisfaction + autonomy_frustration + relatedness_satisfaction
       + relatedness_frustration

academic_performance ~ autonomous_motivation + controlled_motivation
```

```

#priors

SMC ~~ prior("beta(48.69922,11.23828)") * competence_satisfaction #1
SMC ~~ prior("beta(48.69922,11.23828)") * relatedness_satisfaction #2
SMC ~~ prior("beta(33.47415,58.23585)") * competence_frustration #3
SMC ~~ prior("beta(33.47415,58.23585)") * relatedness_frustration #4

competence_satisfaction ~~
       prior("beta(48.33765,10.97235)") * autonomous_motivation #5
autonomy_satisfaction ~~
       prior("beta(48.33765,10.97235)") * autonomous_motivation #6
relatedness_satisfaction ~~
       prior("beta(48.73498,11.265)") * autonomous_motivation #7
competence_satisfaction ~~
       prior("beta(21.52807,57.04153)") * controlled_motivation #8
autonomy_satisfaction ~~
       prior("beta(58.18198,33.769)") * controlled_motivation #9
relatedness_satisfaction ~~
       prior("beta(21.20825,56.90685)") * controlled_motivation #10

competence_frustration ~~
       prior("beta(30.83955,58.55045)") * autonomous_motivation #11
autonomy_frustration ~~
       prior("beta(30.83955,58.55045)") * autonomous_motivation #12
relatedness_frustration ~~
       prior("beta(30.83955,58.55045") * autonomous_motivation #13
competence_frustration ~~
       prior("beta(56.2104,19.7496)") * controlled_motivation #14
autonomy_frustration ~~
       prior("beta(56.2104,19.7496)") * controlled_motivation #15
relatedness_frustration ~~
       prior("beta(56.2104,19.7496)") * controlled_motivation #16

autonomous_motivation
       ~~ prior("beta(56.62967,38.94783)") * academic_performance #17
controlled_motivation ~~ prior("beta(44.1,53.9)") * academic_performance #18
'

fit_ip<-bsem(model_ip, data=Data)
```

A.2. Noninformative prior model

```
library(blavaan)
library(readxl)

Data <- read_excel("Data corrected.xlsx")

model_np<-'

#measurement part

SMC =~ SM1 + SM3 + SM4 + SM11 + SM12 + SM13 + SM14 + SM15 + SM16 + SM18 + SM19
       + SM20 + SM21 + SM26

autonomy_satisfaction  =~ Needs1 + Needs7 + Needs13 + Needs19
autonomy_frustration =~ Needs2 + Needs8 + Needs14 + Needs2
competence_satisfaction =~ Needs5 + Needs11 + Needs17 + Needs23
competence_frustration =~ Needs6 + Needs12 + Needs18 + Needs24
relatedness_satisfaction =~ Needs3 + Needs9 + Needs15 + Needs21
```

```
relatedness_frustration =~ Needs4 + Needs10 + Needs16 + Needs22

autonomous_motivation =~ Motivation2 + Motivation4 + Motivation7 + Motivation8
       + Motivation11 + Motivation13 + Motivation15 + Motivation16
controlled_motivation =~ Motivation1 + Motivation3 + Motivation5 + Motivation6
       + Motivation9 + Motivation10 + Motivation12 + Motivation14

academic_performance =~ Grade

#structural part

relatedness_satisfaction ~ SMC
relatedness_frustration ~ SMC
competence_satisfaction ~ SMC
competence_frustration ~ SMC

autonomous_motivation ~ competence_satisfaction + competence_frustration
       + autonomy_satisfaction + autonomy_frustration + relatedness_satisfaction
       + relatedness_frustration
controlled_motivation ~ competence_satisfaction + competence_frustration
       + autonomy_satisfaction + autonomy_frustration + relatedness_satisfaction
       + relatedness_frustration

academic_performance ~ autonomous_motivation + controlled_motivation
'

fit_np<-bsem(model_np, data=Data)
```

A.3. Weakly informative prior model

A.3.1. Part 1: Noninformative priors

```
library(blavaan)
library(readxl)

Data <- read_excel("Data corrected part 1.xlsx")

model_wp1<-'

#define measurement part

SMC =~ SM1 + SM3 + SM4 + SM11 + SM12 + SM13 + SM14 + SM15 + SM16 + SM18 + SM19
       + SM20 + SM21 + SM26

autonomy_satisfaction  =~ Needs1 + Needs7 + Needs13 + Needs19
autonomy_frustration =~ Needs2 + Needs8 + Needs14 + Needs2
competence_satisfaction =~ Needs5 + Needs11 + Needs17 + Needs23
competence_frustration =~ Needs6 + Needs12 + Needs18 + Needs24
relatedness_satisfaction =~ Needs3 + Needs9 + Needs15 + Needs21
relatedness_frustration =~ Needs4 + Needs10 + Needs16 + Needs22

autonomous_motivation =~ Motivation2 + Motivation4 + Motivation7 + Motivation8
       + Motivation11 + Motivation13 + Motivation15 + Motivation16
controlled_motivation =~ Motivation1 + Motivation3 + Motivation5 + Motivation6
       + Motivation9 + Motivation10 + Motivation12 + Motivation14

academic_performance =~ Grade

#structural part
```

```
relatedness_satisfaction ~ SMC
relatedness_frustration ~ SMC
competence_satisfaction ~ SMC
competence_frustration ~ SMC

autonomous_motivation ~ competence_satisfaction + competence_frustration
       + autonomy_satisfaction + autonomy_frustration + relatedness_satisfaction
       + relatedness_frustration
controlled_motivation ~ competence_satisfaction + competence_frustration
       + autonomy_satisfaction + autonomy_frustration + relatedness_satisfaction
       + relatedness_frustration

academic_performance ~ autonomous_motivation + controlled_motivation
'

fit_wp1<-bsem(model_wp1, data=Data)
```

A.3.2. Part 2: Weakly informative priors

```
library(blavaan)
library(readxl)

Data <- read_excel("Data corrected part 2.xlsx")

model_wp2<-'

#define measurement part

SMC =~ SM1 + SM3 + SM4 + SM11 + SM12 + SM13 + SM14 + SM15 + SM16 + SM18 + SM19
       + SM20 + SM21 + SM26

autonomy_satisfaction  =~ Needs1 + Needs7 + Needs13 + Needs19
autonomy_frustration =~ Needs2 + Needs8 + Needs14 + Needs2
competence_satisfaction =~ Needs5 + Needs11 + Needs17 + Needs23
competence_frustration =~ Needs6 + Needs12 + Needs18 + Needs24
relatedness_satisfaction =~ Needs3 + Needs9 + Needs15 + Needs21
relatedness_frustration =~ Needs4 + Needs10 + Needs16 + Needs22

autonomous_motivation =~ Motivation2 + Motivation4 + Motivation7 + Motivation8
       + Motivation11 + Motivation13 + Motivation15 + Motivation16
controlled_motivation =~ Motivation1 + Motivation3 + Motivation5 + Motivation6
       + Motivation9 + Motivation10 + Motivation12 + Motivation14

academic_performance =~ Grade

#structural part

relatedness_satisfaction ~ SMC
relatedness_frustration ~ SMC
competence_satisfaction ~ SMC
competence_frustration ~ SMC

autonomous_motivation ~ competence_satisfaction + competence_frustration
       + autonomy_satisfaction + autonomy_frustration + relatedness_satisfaction
       + relatedness_frustration
controlled_motivation ~ competence_satisfaction + competence_frustration
       + autonomy_satisfaction + autonomy_frustration + relatedness_satisfaction
       + relatedness_frustration

academic_performance ~ autonomous_motivation + controlled_motivation
```

```

#priors

SMC ~~ prior("beta(46.90625, 51.84375)") * competence_satisfaction #1
SMC ~~ prior("beta(57.8088, 35.431)") * relatedness_satisfaction #2
SMC ~~ prior("beta(47.4432, 51.3968)") * competence_frustration #3
SMC ~~ prior("beta(49.98995, 49.00005)") * relatedness_frustration #4

competence_satisfaction ~~
       prior("beta(54.6336, 42.9264)") * autonomous_motivation #5
autonomy_satisfaction ~~
       prior("beta(37.40625, 5.34375)") * autonomous_motivation #6
relatedness_satisfaction ~~
       prior("beta(54.27345, 43.51655)") * autonomous_motivation #7
competence_satisfaction ~~
       prior("beta(38.63295, 56.75705)") * controlled_motivation #8
autonomy_satisfaction ~~
       prior("beta(55.86755, 19.122451)") * controlled_motivation #9
relatedness_satisfaction ~~
       prior("beta47.97135, 50.93865)") * controlled_motivation #10

competence_frustration ~~
       prior("beta(58.23585, 33.47415)") * autonomous_motivation #11
autonomy_frustration ~~
       prior("beta(51.84375, 46.90625)") * autonomous_motivation #12
relatedness_frustration ~~
       prior("beta(49.00005, 49.98995)") * autonomous_motivation #13
competence_frustration ~~
       prior("beta(57.8088, 35.4312)") * controlled_motivation #14
autonomy_frustration ~~
       prior("beta(58.3464 , 26.2136)") * controlled_motivation #15
relatedness_frustration ~~ prior("beta(53.9, 44.1)") * controlled_motivation #16

autonomous_motivation ~~
       prior("beta(52.2792, 46.3608)") * academic_performance #17
controlled_motivation ~~
       prior("beta(55.9352, 40.5048)") * academic_performance #18

'

fit_wp2<-bsem(model_wp2, data=Data)
```

A.4. Frequentist model

```
library(lavaan)
library(readxl)


my_data <- read_excel("~/3 Studie/BEP/Prime/R/Data/Data corrected.xlsx")

model_fsem<-'
#measurement part

SMC =~ SM1 + SM3 + SM4 + SM11 + SM12 + SM13 + SM14 + SM15 + SM16 + SM18 + SM19
       + SM20 + SM21 + SM26

autonomy_satisfaction  =~ Needs1 + Needs7 + Needs13 + Needs19
autonomy_frustration =~ Needs2 + Needs8 + Needs14 + Needs2
competence_satisfaction =~ Needs5 + Needs11 + Needs17 + Needs23
competence_frustration =~ Needs6 + Needs12 + Needs18 + Needs24
```

```
relatedness_satisfaction =~ Needs3 + Needs9 + Needs15 + Needs21
relatedness_frustration =~ Needs4 + Needs10 + Needs16 + Needs22

autonomous_motivation =~ Motivation2 + Motivation4 + Motivation7 + Motivation8
       + Motivation11 + Motivation13 + Motivation15 + Motivation16
controlled_motivation =~ Motivation1 + Motivation3 + Motivation5 + Motivation6
       + Motivation9 + Motivation10 + Motivation12 + Motivation14

academic_performance =~ Grade

#structural part

relatedness_satisfaction ~ SMC
relatedness_frustration ~ SMC
competence_satisfaction ~ SMC
competence_frustration ~ SMC

autonomous_motivation ~ competence_satisfaction + competence_frustration
       + autonomy_satisfaction + autonomy_frustration + relatedness_satisfaction
       + relatedness_frustration
controlled_motivation ~ competence_satisfaction + competence_frustration
       + autonomy_satisfaction + autonomy_frustration + relatedness_satisfaction
       + relatedness_frustration

academic_performance ~ autonomous_motivation + controlled_motivation
'

fit_fsem<-sem(model_fsem, as.data.frame(my_data))
```

B

R code: Model evaluation and comparison

B.1. Bayesian analysis

```
library(blavaan)

# Model evaluation (RMSEA, SRMR, CFI and TLI)

AFI_ip <- ppmc(fit_ip, fit.measures = c(”rmsea”, ”srmr”, ”cfi”, ”tli”))
summary(AFI_ip)

AFI_np <- ppmc(fit_np, fit.measures = c(”rmsea”, ”srmr”, ”cfi”, ”tli”))
summary(AFI_np)

AFI_wp2 <- ppmc(fit_wp2, fit.measures = c(”rmsea”, ”srmr”, ”cfi”, ”tli”))
summary(AFI_wp2)

# Model comparison (WAIC, LOO and Bayes Factor)

fitMeasures(fit_ip)
fitMeasures(fit_np)

blavCompare(fit_ip, fit_np)
```

B.2. Frequentist analysis

```
library(lavaan)

# Model evaluation (RMSEA, SRMR, CFI and TLI)

fitMeasures(fit_fsem, fit.measures = c(”rmsea”, ”srmr”, ”cfi”, ”tli”), output =
”matrix”)
```

B.3. Visualizations

```
library(semPlot)

m <- rbind(
c(0,7,0,0),
```

```
c(0,6,0,0),
c(1,5,9,10),
c(0,4,8,0),
c(0,3,0,0),
c(0,2,0,0))

semPaths(fit_ip,
structural = TRUE,
reorder = F,
what = ”path”,
whatLabels = ”std”,
latents = c(”SMC”, ”relatedness_frustration”,”autonomy_frustration”,
       ”competence_frustration”,”relatedness_satisfaction”, ”autonomy_
           satisfaction”,”competence_satisfaction”,
       ”controlled_motivation”,”autonomous_motivation”,”academic_performance”),
nodeLabels = c(”SMC”,”RF”,”AF”,”CF”,”RS”,”AS”,”CS”,”CM”,”AM”,”ACP”),
rotation = 2 ,
layout = m,
residuals = F,
intercepts = F,
thresholds = F,
exoCov = F,
edge.label.position = 0.3,
color = list(lat = ”lightskyblue”, man = ”lightsteelblue1”)
)
```

C

R code: Beta distributions

```
#displays the (0,1) beta converted to (-1,1)

beta <- function(a,b) {
       x <- seq(.00001, .99999, .00001)
       plot(-1 + 2*x, dbeta(x, a, b), type = "l", xlab = "", ylab = "",
       main = paste("Beta(",a,",",b,")"))
}

#converts mean and variance to (0,1), calculates a and b, gives beta(a,b)

prior <- function(mu_in) {
       sigma2_in = 0.01
       mu <- 0.5*mu_in + 0.5
       sigma2 <- 0.25*sigma2_in

       a = mu*(mu*(1-mu)/sigma2 - 1)
       b = a*(1-mu)/mu

       beta(a,b)
       cat("prior(\"beta(",a, ",",b,")\")")
}
```

Bibliography

[1] R. Badri, J. Amani-Saribaglou, G. Ahrari, N. Jahadi, and H. Mahmoudi. School culture, basic psychological needs, intrinsic motivation and academic achievement: Testing a casual model. *Mathematics Education Trends and Research*, 2014:1–13, 2014.

[2] K.A. Bollen. *Structural Equations with Latent Variables*. Wiley, 1989.

[3] K.A. Bollen and J.S. Long. *Testing Structural Equation Models*. Sage Publications, Inc., 1993.

[4] S. Brooks, A. Gelman, G.L. Jones, and X.-L. Meng. *Handbook of Markov Chain Monte Carlo*. CRC Press, 2011.

[5] A.J. Cabo and R.G Klaassen. Active learning in redesigning mathematics courses for engineering students. *14th International CDIO Conference*, 2018.

[6] J.R. Deegan. On the occurrence of standardized regression coefficients greater than one. *Educational and Psychological Measurement*, 4(38):873–888, 1978.

[7] S. Duane, A.D. Kennedy, B.J. Pendleton, and D. Roweth. Hybrid Monte Carlo. *Physics Letters B*, 195(2):216–222, 1987.

[8] National Research Council et al. *Frontiers in Massive Data Analysis*. The National Academies Press, 2013.

[9] M. Garnier Villarreal and T. Jorgensen. Adapting fit indices for Bayesian structural equation modeling: Comparison to Maximum Likelihood. *Psychological Methods*, 25(1):46–70, 2020.

[10] A. Gelman. Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, 3:515–534, 2006.

[11] A. Gelman, J.B. Carlin, H.S. Stern, D.B. Dunson, A. Vehtari, and D.B. Rubin. *Bayesian Data Analysis*. CRC Press, 2003.

[12] L. Haerens, N. Aelterman, M. Vansteenkiste, B. Soenens, and P. Van Petegem. Do perceived autonomy-supportive and controlling teaching relate to physical education students' motivational experiences through unique pathways? Distinguishing between the bright and dark side of motivation. *Psychology of Sport and Exercise*, 16:26–36, 2015.

[13] R.H. Hoyle. *Handbook of Structural Equation Modeling*. Guilford Press, 2012.

[14] P. Hu. The correlation between need satisfaction and learning motivation: A self-determination theory perspective. *International Journal of Learning and Teaching*, 9(1):319–329, 2017.

[15] L.M. Jeno, A.G. Danielsen, and A. Raaheim. A prospective investigation of students' academic achievement and dropout in higher education: a self-determination theory approach. *Educational Psychology*, 38(9):1163–1184, 2018.

[16] S.Y. Lee. *Structural Equation Modeling: A Bayesian Approach*. Wiley, 2007.

[17] S.Y. Lee and X.Y. Song. *Basic and Advanced Bayesian Structural Equation Modeling: With Applications in the Medical and Behavioral Sciences*. Wiley, 2012.

[18] F.M. Leo, A. Mouratidis, J.J. Pulido, M.A. López-Gajardo, and D. Sánchez-Oliva. Perceived teachers' behavior and students' engagement in physical education: the mediating role of basic psychological needs and self- determined motivation. *Physical Education and Sport Pedagogy*, 2020.

[19] R. Levy. Bayesian data-model fit assessment for structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 18(4):663–685, 2011.

[20] S. Manganelli, E. Cavicchiolo, L. Mallia, V. Biasi, F. Lucidi, and F. Alivernini. The interplay between self-determined motivation, self-regulated cognitive strategies, and prior achievement in predicting academic performance. *Educational Psychology*, 39(4):470–488, 2019.

[21] E.C. Merkle and Y. Rosseel. blavaan: Bayesian structural equation models via parameter expansion. *Journal of Statistical Software*, 2015.

[22] E.C. Merkle and T. Wang. Bayesian latent variable models for the analysis of experimental psychology data. *Psychonomic Bulletin and Review*, 25:256–270, 2018.

[23] E.C. Merkle, E. Fitzsimmons, J. Uanhoro, and B. Goodrich. Efficient Bayesian structural equation modeling in Stan. *Journal of Statistical Software*, 2021 (in press).

[24] B. Muthén and T. Asparouhov. Bayesian structural equation modeling: A more flexible representation of substantive theory bengt. *Psychological Methods*, 17(3):313–335, 2012.

[25] V.R. Plomp. *Estimating links between latent variables using Structural Equation Modeling in R*. Bachelor thesis, 2020.

[26] S.J. Press. *Subjective and Objective Bayesian Statistics: Principles, Models, and Applications*. Wiley, 2003.

[27] B. Puza. *Bayesian Methods for Statistical Analysis*. Australian National University Press, 2015.

[28] J.A. Rice. *Mathematical Statistics and Data Analysis*. Brooks/Cole, 2007.

[29] Y. Rosseel. lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2):1–36, 2012.

[30] K. Siegrist. *Probability, Mathematical Statistics, Stochastic Processes*. 2021.

[31] A.F.M. Smith and G.O. Roberts. Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society. Series B (Methodological)*, 55(1):3–23, 1993.

[32] Stan Development Team. *Stan Reference Manual, Version 2.26*. 2021. URL `http://mc-stan.org/`.

[33] F.H. Van der Meulen. *Lecture Notes Statistical Inference (WI4455)*. 2019.

[34] S.J. Van Erp. *Bayesian structural equation modeling: The power of the prior*. PhD thesis, 2020.

[35] D. Van Ravenzwaaij, P. Cassey, and S.D. Brown. A simple introduction to Markov chain Monte–Carlo sampling. *Psychonomic Bulletin & Review*, 25:143–154, 2018.

[36] J. Wang and X. Wang. *Structural Equation Modeling: Applications Using Mplus*. Wiley, 2012.

[37] S. Watanabe. Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, (11):3571–3594, 2010.

[38] M.A.J. Zonvervan-Zwijnenburg. *Standing on the Shoulders of Giants: Formalizing and Evaluating Prior Knowledge*. PhD thesis, 2019.