

PARTIAL LEAST SQUARES PATH MODELING
FOR INTERVAL-VALUED VARIABLES

by

NATALIA COSTA ARAUJO

(Under the Direction of Lynne Billard)

ABSTRACT

Symbolic data analysis was first introduced by Diday (1987) and presents an alternative approach to classical data when the data have a more complex formulation; assuming different forms, such as lists, intervals, histograms, etc. There is an increasing need to develop and improve techniques to deal with and to make inferences about symbolic data, while offering efficiency and interpretability to the results. Focusing solely on interval-valued data, where observations are represented by a lower and upper bound, an approach for partial least squares path modeling is proposed. Partial least squares path modeling (PLS-PM) is also known as a variance based method for structural equation modeling (SEM), and intends to quantify and estimate the directional relationship between latent and manifest variables. The PLS-PM uses a two-step iterative process to estimate the latent variables, followed by successive linear regressions to obtain the estimation of the parameters in structural and measurement models.

A partial least squares path modeling approach for interval-valued variables method is proposed, by using two of the currently available regression methods for interval-valued variables, the center method and the symbolic covariance method. The PLS-PM for interval-

valued data is illustrated with an example with data from the past ten years from the National Football League (NFL) games. Later, some simulation studies are conducted in order to understand the behavior of the estimation of the parameters in both the structural and measurement models when different aspects vary. Among other findings, the simulations show that, in general, the symbolic covariance method produces more variable estimates than does the center method, and is also more affected by any collinearity in the structural model. Furthermore, for both regression methods, wider interval-valued variables tend to produce estimates with larger absolute relative biases. The PLS-PM process and the algorithm to randomly generate interval-valued data with the SEM structure are thoroughly described in the dissertation.

INDEX WORDS: Symbolic data analysis, Interval data, Interval-valued data, SEM, Structural equation modeling, PLS, PLS-PM, Partial least squares, Partial least squares path modeling

PARTIAL LEAST SQUARES PATH MODELING
FOR INTERVAL-VALUED VARIABLES

by

NATALIA COSTA ARAUJO

B.S., Universidade Federal de Minas Gerais, Brazil, 2013

M.S., University of Georgia, 2017

A Dissertation Submitted to the Graduate Faculty
of The University of Georgia in Partial Fulfillment
of the

Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2019

©2019

Natalia Costa Araujo

All Rights Reserved

PARTIAL LEAST SQUARES PATH MODELING
FOR INTERVAL-VALUED VARIABLES

by

NATALIA COSTA ARAUJO

Major Professor: Lynne Billard

Committee: Pengsheng Ji
Nicole A. Lazar
T. N. Sriram

Electronic Version Approved:

Ron Walcott
Interim Dean of the Graduate School
The University of Georgia
December 2019

**Partial least squares path modeling
for interval-valued variables**

Natalia Costa Araujo

December 2019

For Mom, Dad and my brother André



Acknowledgments

I would like to start by thanking Dr. Lynne Billard for her guidance and support in these past years. It has been a long journey, but her encouragement and advice were always inspiring and helpful. She is not only an exceptional statistician, but also a great mentor, dedicated to help her students achieve their best.

Next, I would like to extend my appreciation to Dr. Pengsheng Ji, Dr. Nicole Lazar, Dr. Cheolwoo Park and Dr. T.N. Sriram for being part of my advisory committee, thank you for your suggestions and dedication. I am also thankful for the valuable lessons from the professors and all the help from the staff in the Department of Statistics. To my classmates and friends in the department, thank you for all the conversations and discussions that led to personal and professional growth.

I would like to thank the friends that I made in Athens, for being my family away from home, and for sharing so many important moments with me. To my friends and relatives in Brazil, thank you for the love and support during this time of absence.

Lastly, I would like to thank my family for their love and encouragement. To my loving brother, thank you for being an inspiration and always believing in me. To my boyfriend Lucas, thank you for the unconditional support, constant love and presence in my life despite the distance. I thank and dedicate this work to my parents, who are not only loving and supportive parents, but also my best friends. Their guidance and presence in every step of the way made it a much more delightful journey.

Contents

Acknowledgments	vi
1 Introduction	1
2 Literature Review	4
2.1 What are Symbolic Data	4
2.2 Types of Symbolic Data	5
2.3 Classical Linear Regression	17
2.4 Linear Regression for Interval Symbolic Data	19
2.5 Structural Equation Modeling	25
3 Partial Least Squares Path Modeling for Interval-valued data	45
3.1 Introduction	45
3.2 Interval-valued PLS-PM	50
3.3 NFL Example	63
4 Simulations for PLS-PM using interval-valued data	69
4.1 Simulated data	70
4.2 Simulation Studies	74
4.3 Conclusion	110

5 Future Work	114
5.1 Simulating data for misspecification study	115
References	119
Appendices	126
A Tables	127
B Figures	175
C R Code	196

Chapter 1

Introduction

In the past decades, there has been a constant increase in the amount of available data, with data formats evolving to represent and summarize all of the information available; raising the need to develop new approaches to deal with and to make inference about large and complex datasets. By presenting an alternative treatment to data, symbolic data analysis was first introduced by Diday (1987), and proposes that instead of points, the data could be treated as concepts. Each observation in symbolic data could refer to a single individual, or an aggregation of individuals, such as residents of a city, or a country, or players in a team. These observations can assume the form of lists, intervals or histograms. Some of the different types of variables in symbolic data analysis are described in Chapter 2; however, the subsequent chapters focus solely on interval-valued data.

Interval-valued data are the most commonly used symbolic data, where observations are represented by a lower and upper bound, instead of the classical single point observation. They can arise naturally, or due to the aggregation of large datasets. One example that may seem intuitive to use interval-valued data, is with blood pressure measures, where single observations can be highly affected by discomfort or anxiety at the moment of measurement. An interval-valued observation could then be expressed as the minimum and maximum

across many measurements throughout the day, forming a more trustworthy measure of the patient's blood pressure.

There currently are some approaches for interval-valued data for many techniques in statistics, such as regression analysis, clustering analysis, principal component analysis, among others; however, there are not much analyses for data in the format of structural equation modeling. Structural equation modeling (SEM) is a two component modeling structure intended to quantify and estimate the directional relationship between latent and manifest variables, the first component being the structural model, which is formed by equations with the relationship between the latent variables; while the second component, the measurement model, forms the relationships between latent and manifest variables. Latent variables are variables that can not be directly observed, whilst a manifest variable can be directly observed.

The most commonly used approaches to estimate the parameters for structural equation modeling are the covariance and variance based structures. The covariance based model is sometimes referred simply as SEM. This can be modeled with the maximum likelihood (ML) approach, that is established under the assumption that the manifest variables follow a multivariate Normal distribution. The variance based method is also called partial least squares path modeling (PLS-PM), and uses a two-step iterative process to estimate the latent variables, followed by successive linear regressions to obtain the estimation of the parameters.

The research shown in this dissertation presents the partial least squares path modeling for interval-valued data. Imakor (2007) has started dealing with partial least squares for interval-valued data, but using a different estimation structure than the one proposed here.

An interval-valued PLS-PM is proposed using two different methodologies to carry out the linear regression estimation, the center and symbolic covariance regression methods. The center regression method, proposed by Billard and Diday (2000), suggests that the interval-

valued data should be centered, in order to allow the use of classical linear regression for the estimation of the parameters, that are then used to obtain interval-valued predicted values. The symbolic covariance regression method, proposed by Xu (2010), uses the interval-valued structure to estimate the parameters using the covariance structure of the data. To illustrate the method proposed, an example using data from the past ten years of games of the National Football League (NFL) to construct interval-valued observations is analysed using PLS-PM, using both the center and symbolic covariance regression methods.

Simulation studies are then used to evaluate the behavior of the method proposed under different specifications. First, different organizations in the structural model are evaluated, followed by a study intended to understand how well the model estimates the parameters under different numbers of manifest variables associated with each latent variable. The third set of simulations compares the estimation of the parameters for interval-valued data generated with different ranges; and the last set of simulations evaluate the estimation under different sample sizes.

This dissertations consists of five chapters. Followed by the introduction in the first chapter, Chapter 2 has an extensive literature review of the types of symbolic data, linear regression methods for interval-valued data, and the classical approach to both the variance and covariance structural equation modeling. The partial least squares path modeling for interval-valued data is presented in Chapter 3, along with the real data example using information from the past ten years from NFL games to construct the interval-valued observations. In Chapter 4, the process to generate data containing the structural equation modeling organization is described, followed by the results and conclusions obtained with the simulations studies carried out. Finally, in Chapter 5, future research topics are briefly described.

Chapter 2

Literature Review

This chapter will present the description of different types of symbolic data in Section 2.1, and in Section 2.2, the necessary background in symbolic data and symbolic data analysis for the following topics. Classical linear regression will be briefly presented in Section 2.3, as well as many approaches for linear regression using symbolic data in Section 2.4, that later will be used in partial least squares estimation. The general format of structural equations, the covariance based structural equation modeling and partial least squares, which is a variance approach for the structural equation modeling, are introduced in the final section of this chapter, Section 2.5.

2.1 What are Symbolic Data

With the advance of technology in the past decades, large data sets have become more common, increasing the need to develop techniques to deal with and make inferences about them, while offering efficiency and interpretability to the results. Presenting an alternative treatment, Diday (1987) proposed that instead of points, the data could be treated as concepts, introducing the idea of the symbolic data analysis.

The observation values in symbolic data can assume different forms, such as lists, intervals, histograms, etc; Diday (1987) also calls these “observations” “concepts”. Each concept may contain information about a single individual, or an aggregation of individuals, such as people living in a city or in a country, or players in a team. An example of a symbolic measure in an individual level could consist of the extreme observations (minimum and maximum) throughout the day of some variable of interest for a single person forming his or her interval-valued observation.

Among other aspects, a key difference between symbolic data and classical data is due to its structure, which retains the internal variation of the information. The knowledge about the internal variation may provide a better understanding of the variability in the population of interest, leading to better inferences.

As described in Billard and Diday (2006a), the observations of symbolic data usually take the form of hypercubes in \mathfrak{R}^p , instead of being a single point in a p -dimensional space \mathfrak{R}^p , as usually assumed for classical data. For interval and histogram-valued variables, most of the methodologies available are derived with the assumption that observations across the interval and sub-intervals are uniformly distributed within that interval. This assumption is valid in some cases, but can be very restrictive in others, leading to inferences that could be doubted. The validity and limitations of this assumption will be addressed later, in Chapter 5.

To introduce symbolic data more formally, in the next section the main types of variables will be defined, and will be followed by some examples.

2.2 Types of Symbolic Data

In this section, we will present some of the main types of symbolic data, such as multi-valued, modal multi-valued, interval-valued and histogram data. Then, taxonomy variables

- variables in which there is a dependency structure in the data - will be briefly introduced.

2.2.1 Multi-valued Variable

In some situations, it might be of interest to collect information in the format of a list. One such situation is in the scenario where someone collects information regarding insurance companies that are present in each state of the country, where the observation for each state (which would be the concept) consists of a list of companies that are sold in that state. The response variable for the observations may vary in dimension, ranging from one object in the list, to the number of objects in the domain \mathcal{Y} . A formal definition and example of a multi-valued variable is presented in Billard and Diday (2006a) and shown below.

Definition 2.2.1 (Billard and Diday, 2006a) A multi-valued symbolic random variable Y is one whose possible value takes one or more values from the list of values in its domain \mathcal{Y} . The complete list of possible values in \mathcal{Y} is finite, and values may be well-defined categorical or quantitative values.

Example 2.2.1 *Suppose that a researcher is interested in evaluating the colors in the flags of different countries. Table 2.1 presents the observations for 6 countries, and each country is considered a concept. The response variable, Y , assumes a list of colors, with domain $\mathcal{Y} = \{blue, green, red, white, yellow\}$. For example, for the United States, the observation is $Y_1 = \{red, white, blue\}$.*

Table 2.1: Flag Colors - Multi-valued Variable

w	Country	Flag Colors
w_1	United States	{blue, red, white}
w_2	Brazil	{blue, green, white, yellow}
w_3	Italy	{blue, red, white}
w_4	Indonesia	{red, white}
w_5	Lybia	{green}
w_6	Ukraine	{blue, yellow}

2.2.2 Modal Multi-valued Variable

A modal multi-valued variable retains more information about the subjects/concepts than the regular multi-valued variable. In this type of variable, the information collected is stored in the format of a list, but it also contains information regarding the occurrence of each of the objects in that list, which are called weights, respecting that the sum of weights for each concept is always 1. As well as the multi-valued variable, the dimension of the list in a modal multi-valued variable can vary from one to the number of elements in the domain \mathcal{Y} .

Consider that a car manufacturer is interested in expanding the brand, and would like to start selling their cars in a new country. There are many ways to approach this question, but one possibility is to obtain the information of car brands sold in each country, as well as their popularity. This information can be kept in a modal multi-valued variable, where the popularity is measured as the proportion of cars for the different brands in each country. A formal definition and example of a modal multi-valued variable is presented in Billard and Diday (2006a) and shown below.

Definition 2.2.2 (Billard and Diday, 2006a) Let \mathcal{Y} be the domain of possible outcomes for a multi-valued random variable Y , with $\mathcal{Y} = \{\eta_1, \eta_2, \dots\}$. Then, a modal multi-valued variable is one whose observed outcome takes values that are a subset of \mathcal{Y} with a nonnegative

measure attached to each of the values in that subset. That is, a particular observation, for the category w_i , takes the form:

$$Y(w_i) = \{\eta_{i1}, p_{i1} \dots; \eta_{is_i}, p_{is_i}\}$$

where $\{\eta_{i1}, \dots, \eta_{is_i}\} \subseteq \mathcal{Y}$ and where the outcome η_{ik} occurs with weight p_{ik} , $k = 1, \dots, s_i$, $i = 1, \dots, n$, with $\sum_{k=1}^{s_i} p_{ik} = 1$.

Example 2.2.2 *Assuming the same context presented in the Example 2.2.1, suppose that now the researcher wants to evaluate the percentages of occurrence of each color in the flags. Therefore, the variable of interest would be modal multi-valued, and the colors would be followed by the extra information regarding its percentage in the countries' flag. To describe the variable, the domain remain the same, $\mathcal{Y} = \{\text{blue, green, red, white, yellow}\}$, but the observations should be presented as in Table 2.2. For the flag from the United States, we can see that red takes approximately 41.5% of the flag, followed by white, including the stars, that covers approximately 40.9% of the flag and blue covers only 17.6% of the American flag.*

Table 2.2: Flag Colors - Modal Multi-valued Variable

w	Country	Flag Colors
w_1	United States	{blue, 0.176; red, 0.415; white, 0.409}
w_2	Brazil	{blue, 0.125; green, 0.686; white, 0.012; yellow, 0.177}
w_3	Italy	{blue, $\frac{1}{3}$; red, $\frac{1}{3}$; white, $\frac{1}{3}$ }
w_4	Indonesia	{red, 0.5; white, 0.5}
w_5	Lybia	{green, 1}
w_6	Ukraine	{blue, 0.5; yellow, 0.5}

2.2.3 Interval-valued Variable

For classical data, analyses are limited to evaluating a single point for each individual or observation, but there are some occasions when there is more information available than a single point. In these cases, maintaining the internal variation of the data can lead to estimates that better describe the population variability.

Although not the only type of variable that can be used in such situations, an interval valued variable is frequently used to store information while maintaining this internal variation. It can be used to describe the relevant information from an individual or an aggregation of individuals.

Suppose that you want to evaluate the performance of different soccer teams in a tournament. Thinking in a classical approach, one could either calculate the mean or the total number of goals for each of these teams. Or, the analysis could be done using interval-valued variables, with each team being an observation, and the lower and upper bounds of the observation defined by the minimum and maximum number of goals scored in a game. The formal definition of an interval-valued variable is presented by Billard and Diday (2006a), and is outlined below.

Definition 2.2.3 (Billard and Diday, 2006a) An interval-valued symbolic random variable Y is one that takes values in an interval, i.e., $Y_i = [a_i, b_i] \subset \mathfrak{R}^1$, with $a \leq b, a, b \in \mathfrak{R}^1$, $i = 1, \dots, n$. The interval can be closed or open at either end, $(a_i, b_i), [a_i, b_i], [a_i, b_i)$ or $(a_i, b_i]$.

Under the assumption that observations across the intervals are uniformly spread, the following estimates for the sample mean and variance have been derived by Bertrand and Goupil (2000), for a sample of size n ,

$$\bar{Y} = \frac{1}{2n} \sum_{i=1}^n (b_i + a_i) \tag{2.1}$$

$$S^2 = \frac{1}{3n} \sum_{i=1}^n (b_i^2 + b_i a_i + a_i^2) - \frac{1}{4n^2} \left[\sum_{i=1}^n (b_i + a_i) \right]^2. \quad (2.2)$$

Similarly to the sample variance in Equation 2.1, the covariance between two interval-valued variables can be calculated using one of the two forms presented in Equations 2.3 and 2.4. In the subsequent chapters and sections the covariance approach shown in Equation 2.3 will be used.

Billard and Diday (2006a) proposed that the empirical covariance between two interval-valued variables should be calculated as,

$$\begin{aligned} Cov(\mathbf{Y}_1, \mathbf{Y}_2) &= Cov((\mathbf{a}_1, \mathbf{b}_1), (\mathbf{a}_2, \mathbf{b}_2)) \\ &= \frac{1}{3n} \sum_{i=1}^n G_{i1} G_{i2} (Q_{i1} Q_{i2})^{1/2} \end{aligned} \quad (2.3)$$

where, for $j = 1, 2$,

$$\begin{aligned} Q_{ij} &= (a_{ij} - \bar{Y}_j)^2 + (a_{ij} - \bar{Y}_j)(b_{ij} - \bar{Y}_j) + (b_{ij} - \bar{Y}_j)^2 \\ G_{ij} &= \begin{cases} -1, & \text{if } \bar{Y}_{ij} \leq \bar{Y}_j \\ 1, & \text{if } \bar{Y}_{ij} > \bar{Y}_j, \end{cases} \end{aligned}$$

where \bar{Y}_j is defined as Equation 2.1, and $\bar{Y}_{ij} = (a_{ij} + b_{ij})/2$.

Billard (2008) proposed that instead, the following form for the empirical covariance should be used,

$$\begin{aligned} Cov(\mathbf{Y}_1, \mathbf{Y}_2) &= Cov((\mathbf{a}_1, \mathbf{b}_1), (\mathbf{a}_2, \mathbf{b}_2)) \\ &= \frac{1}{6n} \sum_{i=1}^n [2(a_{i1} - \bar{Y}_1)(a_{i2} - \bar{Y}_2) + (a_{i1} - \bar{Y}_1)(b_{i2} - \bar{Y}_2) \\ &\quad + (b_{i1} - \bar{Y}_1)(a_{i2} - \bar{Y}_2) + 2(b_{i1} - \bar{Y}_1)(b_{i2} - \bar{Y}_2)] \end{aligned} \quad (2.4)$$

where \bar{Y}_1 and \bar{Y}_2 are defined as Equation 2.1. Once the empirical variance and covariance have been defined, the correlation coefficient between two interval-valued variables is,

$$Cor(\mathbf{Y}_1, \mathbf{Y}_2) = \frac{Cov(\mathbf{Y}_1, \mathbf{Y}_2)}{\sqrt{Var(\mathbf{Y}_1) \times Var(\mathbf{Y}_2)}}. \quad (2.5)$$

Example 2.2.3 *The organization Food Security Portal provides a broad range of datasets regarding worldwide food availability. One of the datasets available describes the Calorie Supply Per Capita available in each country throughout the years.*

Suppose that a researcher is interested in comparing the calorie supply across countries and decides to make intervals with the minimum and maximum calorie supply per capita availability from 2010 to 2013. Table 2.3 presents the information for 10 randomly selected countries from which the full information is available.

Using Equation 2.1 and Equation 2.2, we can see that the symbolic sample mean is 2467.95 calories and the symbolic sample variance is 112780.63 calories².

Table 2.3: Calorie Supply Per Capita

w	Country	Calorie Supply
w_1	Colombia	[2665, 2804]
w_2	Ethiopia	[2080, 2131]
w_3	India	[2435, 2459]
w_4	Mexico	[3028, 3072]
w_5	Mozambique	[2242, 2298]
w_6	Panama	[2579, 2733]
w_7	Philippines	[2570, 2584]
w_8	Tanzania	[2097, 2208]
w_9	Thailand	[2756, 2784]
w_{10}	Zambia	[1904, 1930]

2.2.4 Histogram-Valued Variable

A histogram valued variable presents a more complex format, where the observation for each concept or individual is represented by a histogram. This type of variable is defined by several sub-intervals, and each of the sub-intervals is associated with a non-negative weight that sums to 1. Due to its complexity, this type of variable usually arises from the aggregation of large datasets.

The interval-valued variable can also be seen as a special case of the histogram valued variable, where the histogram of each sub-interval is composed of a single interval, and the weight associated with this sub-interval is 1.

Definition 2.2.4 (Billard and Diday, 2006a) Let Y be a quantitative random variable that can take values on a finite number of non-overlapping intervals $\{[a_k, b_k), k = 1, 2, \dots\}$ with $a_k \leq b_k$. Then, an outcome for observation w_i for a histogram interval-valued random variable takes the form:

$$Y(w_i) = \{[a_{ik}, b_{ik}), p_{ik}; k = 1, \dots, s_i, i = 1, \dots, n\}$$

where $s_i < \infty$ is the finite number of intervals forming the support for the outcome $Y(w_i)$ for observation w_i , and where p_{ik} is the weight for the particular sub-interval $[a_{ik}, b_{ik}), k = 1, \dots, s_i$, with $\sum_{k=1}^{s_i} p_{ik} = 1$. The intervals $[a_k, b_k)$ can be open or closed at either end.

Under the assumption that observations across each of the sub-intervals are uniformly distributed, the following estimates for the mean and variance have been proposed by Billard and Diday (2006a):

$$\bar{Y} = \frac{1}{2n} \sum_{i=1}^n [\sum_{k=1}^{s_i} (b_{ik} + a_{ik}) p_{ik}] \quad (2.6)$$

$$S^2 = \frac{1}{3n} \sum_{i=1}^n [\sum_{k=1}^{s_i} (b_{ik}^2 + b_{ik} a_{ik} + a_{ik}^2)] - \frac{1}{4n^2} [\sum_{i=1}^n \sum_{k=1}^{s_i} (b_{ik} + a_{ik}) p_{ik}]^2 \quad (2.7)$$

Example 2.2.4 *The following example was extracted from Chapter 3 in Billard and Diday (2006a). Suppose that you want to analyze a large dataset containing the information about women's age and weight using histogram valued variables. Aggregating the subjects by age range, you can construct a histogram observation for each of the age ranges of interest. Table 2.4 presents a set of possible histogram valued observations, that are also described in Figure 2.1.*

Using the equations presented above, we can easily find that the symbolic sample mean is 143.91 pounds and the sample variance is 4471.51 pounds².

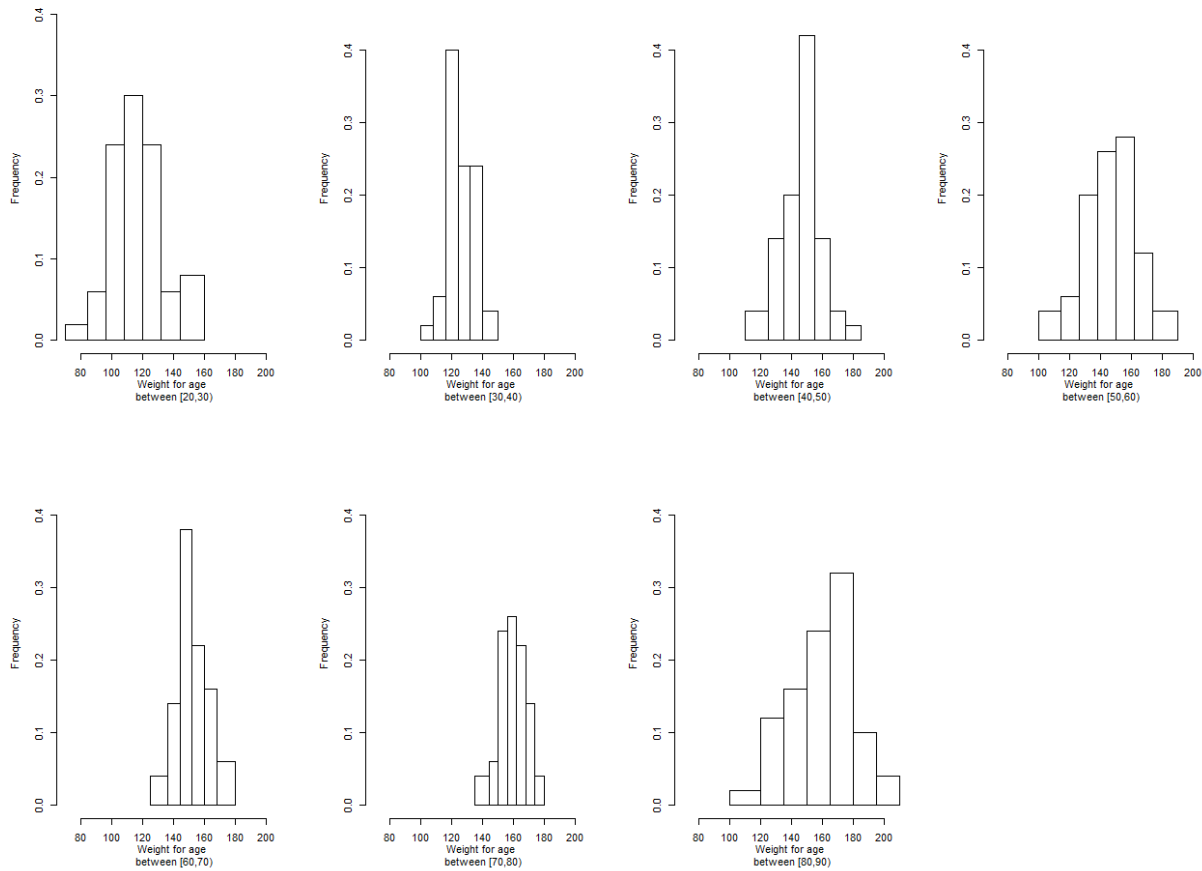


Figure 2.1: Histogram data of weight by age ranges

Table 2.4: Weight by age groups - Histogram data

w	Age	
w_1	[20,30)	{[70,84),0.02; [84,96),0.06; [96,108),0.24; [108,120),0.3; [120,132),0.24; [132,144),0.06; [144,160),0.08}
w_2	[30,40)	{[100,108),0.02; [108,116),0.06; [116,124),0.40; [124,132),0.24; [132,140),0.24; [140,150),0.04}
w_3	[40,50)	{[110,125),0.04; [125,135),0.14; [135,145),0.20; [145,155),0.42; [155,165),0.14; [165,175),0.04; [175,185),0.02}
w_4	[50,60)	{[100,114),0.04; [114,126),0.06; [126,138),0.20; [138,150),0.26; [150,162),0.28; [162,174),0.12; [174,190),0.04}
w_5	[60,70)	{[125,136),0.04; [136,144),0.14; [144,152),0.38; [152,160),0.22; [160,168),0.16; [168,180),0.06}
w_6	[70,80)	{[135,144),0.04; [144,150),0.06; [150,156),0.24; [156,162),0.26; [162,168),0.22; [168,174),0.14; [174,180),0.04}
w_7	[80,90)	{[100,120),0.02; [120,135),0.12; [135,150),0.16; [150,165),0.24; [165,180),0.32; [180,195),0.10; [195,210),0.04}

2.2.5 Dependent Variables

Dependency structures can be introduced in the data for many reasons, such as to avoid inconsistency in the data or maintain the structure of the information. A dependency structure can be imposed by using a set of rules (which are called logical dependencies), or by specifying the variables hierarchically (using taxonomy or hierarchy variables), Bock and Diday (2000).

Billard and Diday (2006b) thoroughly describe descriptive statistics for interval-valued observations when these variables are in the presence of rules. To exemplify the logical dependencies included via rules, Table 2.5 presents the age ($Y_{1,i}$) and number of years in school ($Y_{2,i}$) for 10 individuals, which can be aggregated to form interval-valued variables,

such as:

$$Y_1 = (Y_{1,1}, Y_{2,1}) = ([10, 26], [4, 16])$$

$$Y_2 = (Y_{1,2}, Y_{2,2}) = ([26, 42], [10, 22])$$

Table 2.5: Number of years in School and Age

Individual	Age	Years in School
1	35	15
2	19	12
3	21	12
4	16	10
5	10	4
6	25	16
7	40	18
8	28	12
9	33	10
10	42	22

By the nature of the variables, it is not possible for an individual to have $Y_{2,i} > Y_{1,i}$, since an individual that is 10 did not attend school for more than 10 years. In fact, considering that a child typically starts the first grade at 6 years old, we can define the following logical rule:

$$v = Y_{2,i} \leq (Y_{1,i} - 6).$$

The logical dependencies (such as the one described above) might be appropriate for some scenarios especially to help remove inconsistency in the data, but there are other situations that by using a taxonomy or hierarchical variable may offer some advantages, since they have the dependency rooted in its structure. Taxonomy and hierarchy rules can be used in classical or symbolic data, and are characterized by the dependency of the levels

of the variable. Both variables are defined and exemplified below, following the definitions presented in Billard and Diday (2006a).

Definition 2.2.5 (Billard and Diday, 2006a) Taxonomy variables are variables organized in a tree; as nested variables, with several levels of generality. The bottom of the tree for a given variable will be referenced as the first level, with the top of the tree corresponding to the total number of levels t .

Example 2.2.5 *A school might be interested in evaluating the types of snacks that the parents send to their kids. Then, it is possible to construct a taxonomy variable such as the one described in Figure 2.2. The second level of this taxonomy tree is $Y_1 = \text{'Type of Snack'}$ and the first level is $Y_2 = \text{'Description'}$.*

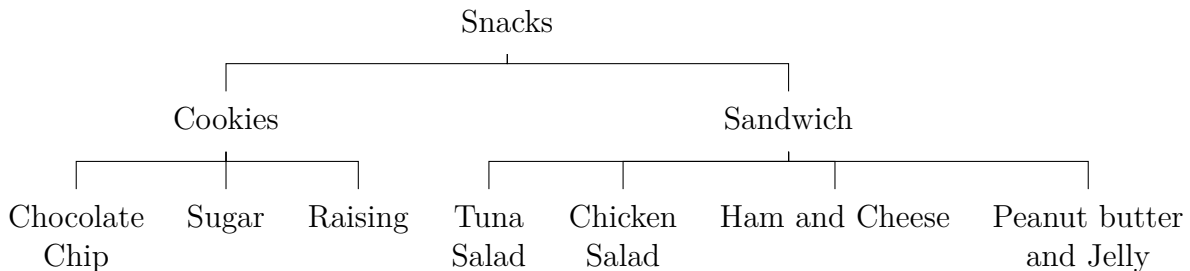


Figure 2.2: Taxonomy tree of types of snacks

Definition 2.2.6 (Billard and Diday, 2006a) A hierarchy tree (or a Mother-Daughter variable) is one in which whether or not a particular (Daughter) variable is operative at a given level of the tree is determined by the outcome of the (Mother) variable at the immediately preceding level of the tree.

Example 2.2.6 *Physical activities play a big role in a person's healthy life, and is a concern expressed by many doctors nowadays. To classify the activity level and types of activities in which a patient or individual engages, it is possible to use a hierarchy variable, such as*

described in Figure 2.3. This hierarchy tree is composed of two levels, where the second level is $Y_1 = \text{'Activity Level'}$ and the first level is $Y_2 = \text{'Physical Activity'}$. Note that if a person is classified as 'Sedentary', the second level ('Physical Activity') is not operative.

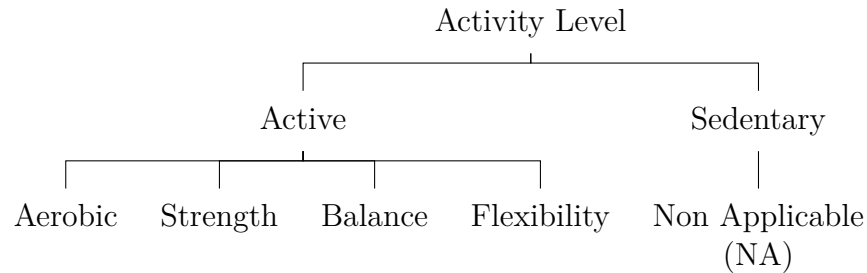


Figure 2.3: Hierarchy tree for activity level

2.3 Classical Linear Regression

There are many statistical methods available to quantify the relationship of a response variable and a set of explanatory variables. Regression analysis or linear models methodologies are broadly used for this purpose, and can be approached as a method to make predictions, or to evaluate the average behavior of the response variables in the presence of the explanatory variables (interpreting the estimated coefficients). The coefficients estimation can be done via maximum likelihood estimation (MLE) or least squares estimation (LSE). The latter is used more often, and will be briefly introduced in this section (for more details see, e.g., Draper and Smith, 1998).

Let \mathbf{Y} be a vector of length n , with the observations Y_i , for $i = 1, \dots, n$, where \mathbf{Y} is the vector of response variable. The design matrix of dimension $n \times (p + 1)$ is \mathbf{X} , with each of the p columns corresponding to an explanatory variable (first column is a vector of $\mathbf{1}$). The error term is defined as $\boldsymbol{\epsilon}_{n \times 1}$, and the vector of parameters (coefficients) is $\boldsymbol{\beta}_{(p+1) \times 1}$. By

using the notation described above, the matrix equation for the linear model is

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad (2.8)$$

where $E(\boldsymbol{\epsilon}) = \mathbf{0}$ and $Var(\boldsymbol{\epsilon}) = \mathbf{I}\sigma^2$. Therefore, Equation 2.8 yields $E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta}$ and $Var(\mathbf{Y}) = \mathbf{I}\sigma^2$.

The least squares estimation is derived under the principle of minimizing the sum of square of errors, $\boldsymbol{\epsilon}^T\boldsymbol{\epsilon}$, which can be written as:

$$\begin{aligned} \boldsymbol{\epsilon}^T\boldsymbol{\epsilon} &= (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \\ &= \mathbf{Y}^T\mathbf{Y} - \boldsymbol{\beta}^T\mathbf{X}^T\mathbf{Y} - \mathbf{Y}^T\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}^T\mathbf{X}^T\mathbf{X}\boldsymbol{\beta} \\ &= \mathbf{Y}^T\mathbf{Y} - 2\boldsymbol{\beta}^T\mathbf{X}^T\mathbf{Y} + \boldsymbol{\beta}^T\mathbf{X}^T\mathbf{X}\boldsymbol{\beta}. \end{aligned}$$

The least squares estimator of $\boldsymbol{\beta}$, denoted by $\hat{\boldsymbol{\beta}}$, can be found by setting $\frac{\partial \boldsymbol{\epsilon}^T\boldsymbol{\epsilon}}{\partial \boldsymbol{\beta}}|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} = \mathbf{0}$. Solving this derivative for $\hat{\boldsymbol{\beta}}$ yields the normal equations in Equation 2.9, that result in the least squares estimator of $\boldsymbol{\beta}$ presented in Equation 2.10 (provided that $\mathbf{X}^T\mathbf{X}$ is nonsingular),

$$(\mathbf{X}^T\mathbf{X})\hat{\boldsymbol{\beta}} = \mathbf{X}^T\mathbf{Y} \quad (2.9)$$

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}. \quad (2.10)$$

As Draper and Smith (1998) discuss in their book, the least squares estimation of the parameter does not require assumptions on the distribution of the errors; but assuming that the errors are normally distributed is necessary for further analysis in classical linear regression. Examples and applications of classical linear regression are available in Montgomery et al. (2001), Kleibbaum and Kupper (1978), Sheather (2009), etc.

2.4 Linear Regression for Interval Symbolic Data

Since the introduction of symbolic data analysis in 1987, many approaches have been proposed to fit a linear regression model for interval-valued data. The nature of these approaches differs substantially, and tends to lead to slightly different estimates.

Some of the methods available will be described in this section, such as the center method, the range and center method, the constrained range and center method and the symbolic covariance method. Despite all of the research done so far, there is no consensus as to what method seems more appropriate for general use. Therefore, the methods will be presented and followed by a brief summary of assumptions made in the model, as well as its drawbacks.

2.4.1 Center Method

A method to fit a linear regression method for interval-valued was first proposed by Billard and Diday (2000). This method suggests that the interval-valued data should be centered, such that the design matrix \mathbf{X}^c has dimension $n \times (p + 1)$, the response variable vector \mathbf{Y}^c has dimension $n \times 1$, and then the analysis should be conducted following the classical linear regression. This method retains the interpretation of the parameters of a classical linear model, and the estimators can be applied to the lower and upper bounds of intervals to produce predictions for the intervals.

Suppose that the interval-valued data are $X_{ij} = [a_{ij}, b_{ij}]$ and $Y_i = [c_i, d_i]$ (for $i = 1, \dots, n$, and $j = 1, \dots, p$), where $\mathbf{X}_1, \dots, \mathbf{X}_p$ are p independent interval-valued variables, and \mathbf{Y} is the interval-valued response variable. The centered matrices \mathbf{X}^c and \mathbf{Y}^c are calculated by finding the center of each of the intervals, such as, for $i = 1, \dots, n$, $j = 1, \dots, p$,

$$X_{ij}^c = (a_{ij} + b_{ij})/2 \quad (2.11)$$

$$Y_i^c = (c_i + d_i)/2. \quad (2.12)$$

Then, the regression model can be written as:

$$\mathbf{Y}^c = \mathbf{X}^c \boldsymbol{\beta}^c + \boldsymbol{\epsilon}^c; \quad (2.13)$$

and the parameters $\boldsymbol{\beta}^c$ can be estimated, for $(\mathbf{X}^c)^T \mathbf{X}^c$ nonsingular, using the least squares estimation presented in the previous section, to give

$$\hat{\boldsymbol{\beta}}^c = ((\mathbf{X}^c)^T \mathbf{X}^c)^{-1} (\mathbf{X}^c)^T \mathbf{Y}^c. \quad (2.14)$$

Thus, the predicted value for an observation $\mathbf{x}^* = (\mathbf{x}_L^*, \mathbf{x}_U^*)$, denoted by $\hat{y}^* = [\hat{y}_L^*, \hat{y}_U^*]$, is calculated as $\hat{y}_L^* = \mathbf{x}_L^{*T} \hat{\boldsymbol{\beta}}^c$ and $\hat{y}_U^* = \mathbf{x}_U^{*T} \hat{\boldsymbol{\beta}}^c$, where the capital letter subscript L indicates the interval lower bound of the interval-value variable and the capital letter subscript U indicates the interval upper bound of the variable. It is easy to see that this method does not guarantee that the lower value prediction is less than or equal to the upper value prediction. This issue is addressed in two of the other methods presented.

One of the many advantages of symbolic data is the presence of internal variation in the data. The centered method imposes the assumption that the intervals are uniformly distributed, and ignores the range of the data in the estimation of the linear regression parameters, but not for the predictions (as given in Billard and Diday, 2000). Even though this method has the advantage of being easy to compute and to interpret, we should consider the loss of information present in the symbolic data, before deciding which technique to use.

2.4.2 Center and Range Method

The center and range method, proposed by Lima Neto and De Carvalho (2008) suggests that the interval-valued data should be modeled by two independent regression equations, one for the center and another for the range of the intervals. Thus, their approach allows for the

range of the predicted response to differ from the original range. The parameters for both equations are such to minimize the sum of the square of both errors, the errors associated with the center squared, plus the errors associated with the range squared.

Suppose again that $X_{ij} = [a_{ij}, b_{ij}]$ and $Y_i = [c_i, d_i]$, for $i = 1, \dots, n$, $j = 1, \dots, p$. Then we can define the center and range of each observation as,

$$\begin{aligned} x_{ij}^c &= \frac{a_{ij} + b_{ij}}{2}, & y_i^c &= \frac{c_i + d_i}{2}, \\ x_{ij}^r &= \frac{b_{ij} - a_{ij}}{2}, & y_i^r &= \frac{d_i - c_i}{2}. \end{aligned}$$

This leads to the matrices \mathbf{X}^c , \mathbf{X}^r , \mathbf{Y}^c and \mathbf{Y}^r (such that $(\mathbf{X}^c)^T \mathbf{X}^c$ and $(\mathbf{X}^r)^T \mathbf{X}^r$ are both nonsingular). Hence, the regression equations for the center and range can be defined as:

$$\mathbf{Y}^c = \mathbf{X}^c \boldsymbol{\beta}^c + \boldsymbol{\epsilon}^c, \quad (2.15)$$

$$\mathbf{Y}^r = \mathbf{X}^r \boldsymbol{\beta}^r + \boldsymbol{\epsilon}^r \quad (2.16)$$

where $\boldsymbol{\epsilon}^c$ and $\boldsymbol{\epsilon}^r$ are the error vectors associated with the center and range, respectively.

Given that the new response and design matrices for the center and range have been calculated, the estimation of the vectors of parameters comes down to the least squares estimation. The estimation of $\boldsymbol{\beta}^c$ and $\boldsymbol{\beta}^r$ can be done independently since no relationship is assumed between the two equations, i.e.,

$$\hat{\boldsymbol{\beta}}^c = ((\mathbf{X}^c)^T \mathbf{X}^c)^{-1} (\mathbf{X}^c)^T \mathbf{Y}^c, \quad (2.17)$$

$$\hat{\boldsymbol{\beta}}^r = ((\mathbf{X}^r)^T \mathbf{X}^r)^{-1} (\mathbf{X}^r)^T \mathbf{Y}^r. \quad (2.18)$$

Using the center and range method, we have the predicted values for an interval-valued

data $\mathbf{x}^* = (\mathbf{x}_L^*, \mathbf{x}_U^*)$ as:

$$\hat{y}_L^* = (\mathbf{x}^{*c})^T \hat{\beta}^c - (\mathbf{x}^{*r})^T \hat{\beta}^r, \quad (2.19)$$

$$\hat{y}_U^* = (\mathbf{x}^{*c})^T \hat{\beta}^c + (\mathbf{x}^{*r})^T \hat{\beta}^r, \quad (2.20)$$

where the capital letter subscript L indicates the interval lower bound of the interval-value variable and the capital letter subscript U indicates the interval upper bound of the variable. Even though this approach presents some improvements to the center method, by estimating the range of the intervals, it assumes that the center and range of an interval-valued data are independent. This assumption is very strong and in practice, does not seem to be valid for most situations. The method also does not assess the possibility of a predicted lower value to be greater than the predicted upper bound.

In some special cases when the interval-valued observations of a variable are different, but they all share the same range or the same center, the center and range method will not perform well. In such situations, there arises a problem of singularity in $(\mathbf{X}^c)^T \mathbf{X}^c$, if all the centers are equal, or $(\mathbf{X}^r)^T \mathbf{X}^r$, if all ranges are equal. This singularity prevents the estimation of the parameters since the product of the matrices is non-invertible.

2.4.3 Constrained Center and Range Method

The constrained linear regression model for symbolic interval-valued variables was proposed by Lima Neto and De Carvalho (2010). This method suggests the use of the center and range method to estimate the parameters of the regression, but constraining the predicted coefficients for the equation modeling the range to be all greater than 0. This method guarantees that the predicted upper bound will always be greater than the predicted lower bound, and is recommended only when the regular center and range method fail, i.e., the predicted upper bound is lower than the predicted lower bound.

By following the same setup of the previous method, the equations to be estimated are:

$$\mathbf{Y}^c = \mathbf{X}^c \boldsymbol{\beta}^c + \boldsymbol{\epsilon}^c, \quad (2.21)$$

$$\mathbf{Y}^r = \mathbf{X}^r \boldsymbol{\beta}^r + \boldsymbol{\epsilon}^r \quad (2.22)$$

where $\beta_j^r \geq 0$ for $j = 0, 1, \dots, p$, and $\boldsymbol{\epsilon}^c$ and $\boldsymbol{\epsilon}^r$ are the error vectors associated with the center and range, respectively. The estimation of $\hat{\boldsymbol{\beta}}^c$ follows as Equation 2.17, the estimation of $\hat{\boldsymbol{\beta}}^r$ does not have a closed form, and needs to be done iteratively. The authors propose the use of Lawson and Hanson's algorithm (see Lawson and Hanson, 1974), that is thoroughly explained in Lima Neto and De Carvalho (2010).

This method presents an interesting solution to the unsolved problems of the center and range method, but this solution includes a restriction that may compromise the results of the estimation, since it forces all of the parameters associated with the range to be greater than 0. We should be careful when using this method, as this constraint may prevent the estimation to capture the real behavior of the data, and produce misleading estimators and predictions.

2.4.4 Symbolic Covariance Method

The symbolic covariance method proposed by Xu (2010) introduces the use of symbolic variance and covariance matrices to estimate the parameters to describe the linear relationship between a set of explanatory variables and a response variable. This method is based on the least squares estimation and therefore many of the results developed for this technique can be extended to the symbolic covariance method, which is one of the biggest advantages of this method; this is presented below.

For $\mathbf{Y} = (y_1, \dots, y_n)^T$ and regression design matrix \mathbf{X} , the classical linear regression can

be written as:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad (2.23)$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)^T$ is estimated - via least squares estimation - using Equation 2.10. Without loss of generality, an alternative formulation to this equation is obtained by centering it, which leads to the following centered classical estimators:

$$\hat{\boldsymbol{\beta}} = [(\mathbf{X} - \bar{\mathbf{X}})^T(\mathbf{X} - \bar{\mathbf{X}})]^{-1}(\mathbf{X} - \bar{\mathbf{X}})^T(\mathbf{Y} - \bar{\mathbf{Y}}), \quad (2.24)$$

$$\hat{\beta}_0 = \bar{Y} - (\hat{\beta}_1\bar{X}_1 + \dots + \hat{\beta}_p\bar{X}_p). \quad (2.25)$$

Given the nature of symbolic data, the functions for estimating the coefficients in Equations 2.24 and 2.25 cannot be directly applied, but Xu (2010) showed that we can rewrite $(\mathbf{X} - \bar{\mathbf{X}})^T(\mathbf{X} - \bar{\mathbf{X}})$ and $(\mathbf{X} - \bar{\mathbf{X}})^T(\mathbf{Y} - \bar{\mathbf{Y}})$, for $i = 1, \dots, n$, as:

$$(\mathbf{X} - \bar{\mathbf{X}})^T(\mathbf{X} - \bar{\mathbf{X}}) = (n \times \mathbf{Cov}(\mathbf{X}_{j_1}, \mathbf{X}_{j_2}))_{p \times p}, \quad j_1, j_2 = 1, \dots, p,$$

$$(\mathbf{X} - \bar{\mathbf{X}})^T(\mathbf{Y} - \bar{\mathbf{Y}}) = (n \times \mathbf{Cov}(\mathbf{X}_j, \mathbf{Y}))_{p \times 1}, \quad j = 1, \dots, p.$$

This leads to the following symbolic estimators:

$$\hat{\boldsymbol{\beta}} = (n \times \mathbf{Cov}(\mathbf{X}_{j_1}, \mathbf{X}_{j_2}))_{p \times p}^{-1} \times (n \times \mathbf{Cov}(\mathbf{X}_j, \mathbf{Y}))_{p \times 1} \quad (2.26)$$

$$\hat{\beta}_0 = \bar{Y} - (\hat{\beta}_1\bar{X}_1 + \dots + \hat{\beta}_p\bar{X}_p) \quad (2.27)$$

where the matrix of variance and covariance, can be obtained as in Equations 2.3 or 2.4 and,

$$\bar{X} = \frac{1}{2n} \sum_{i=1}^n (a_i + b_i). \quad (2.28)$$

The predicted value of $\mathbf{x}^* = (\mathbf{x}_L^*, \mathbf{x}_U^*)$ is calculated in a similar way as the center method, but providing a simple solution to the situations when the lower predicted value is greater than the upper predicted value:

$$\hat{Y}^* = (\hat{Y}_L^*, \hat{Y}_U^*) \quad (2.29)$$

where

$$\hat{Y}_{iL}^* = \min(\mathbf{x}_{iL}^* \times \hat{\beta}, \mathbf{x}_{iU}^* \times \hat{\beta}) \quad (2.30)$$

$$\hat{Y}_{iU}^* = \max(\mathbf{x}_{iL}^* \times \hat{\beta}, \mathbf{x}_{iU}^* \times \hat{\beta}), \quad (2.31)$$

and the capital letter subscript L indicates the interval lower bound of the interval-value variable and the capital letter subscript U indicates the interval upper bound of the variable.

2.5 Structural Equation Modeling

Structural equation modeling (SEM) is a two component modeling structure intended to quantify and estimate the directional relationship between latent and manifest variables. O'Rourke and Hatcher (2013) defined these components as: the measurement model “specifies relationships between latent and their indicator variables” and the structural model “specifies directional relationship between latent constructs”. Here a latent variable (LV) is a variable or hypothetical construct that can not be directly observed, while a manifest variable (MV) - sometimes also called an indicator variable - can be directly observed.

The SEM is considered a confirmatory method, such that previous exploratory work need to be available to indicate the structure for the model. Many aspects of the measurement model (and SEM in general) resemble a confirmatory factor analysis (CFA), but they differ mainly because CFA allows all the latent variables to correlate, while SEM specifies directional relationships to be modeled through its structural component.

There are some controversies about the usage of SEM, mainly because some researchers tend to misuse the method or because of its causality nature. Structural equation modeling is sometimes also called a causal model, because the estimation of the parameters is based on assumptions that a latent variable is a function of other variables. The main debate involving causality on SEM is related to the fact that the data available are usually non-experimental, and therefore the results could be misleading. This issue has been discussed extensively in the past years, and while there is no general consensus in the topic, some researchers suggest the careful use of SEM is acceptable, keeping in mind the non-experimental nature of the data, and therefore its limitations; see McDonald (2004) or Meehl and Waller (2002).

The full SEM is usually presented through a set of equations of matrices and a path diagram. In a path diagram, a single straight-headed arrow indicates that there is an unidirectional path between the variables, while a curved double-headed arrow indicates that the covariance between the variables is not fixed, and will be estimated. Latent variables are delimited by ovals or circles, while manifest variables are delimited by squares or rectangles (see, e.g., Figure 2.4). The structural equation modeling can distinguish between endogenous or exogenous variables. Exogenous variables are not influenced by other variables in the model, while endogenous variables act as dependent variables. In the path diagram, any variable that receives a straight single-headed arrow is considered an endogenous variable, such as x_h 's, y_j 's, η_1 and η_2 , for $h = 1, \dots, 6$ and $j = 1, \dots, 4$ in Figure 2.4.

There is a wide number of possible path diagrams that can be formed with different numbers of exogenous and endogenous variables, but to illustrate with a simple case, Figure

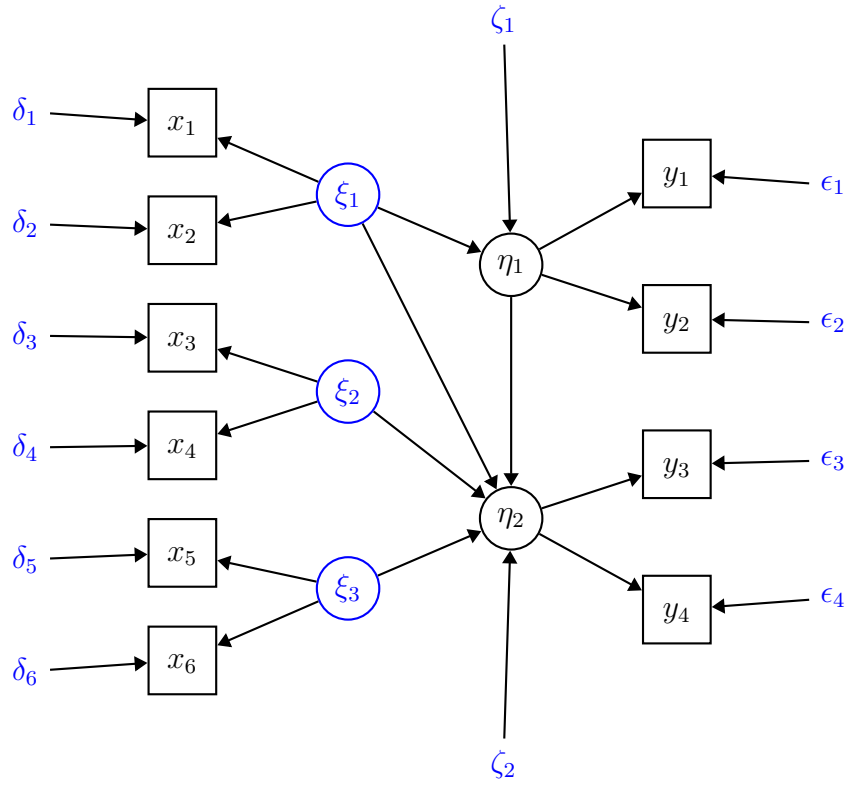


Figure 2.4: Example of a Path Diagram

2.4 shows the path diagram presented in Diamantopoulos (1994). The variables δ_h , ϵ_j , ξ_1 , ξ_2 , ξ_3 , ζ_1 and ζ_2 , in blue, indicate that they are exogenous variables, while the variables x_h , y_j , η_1 and η_2 , in black, are endogenous variables, for $h = 1, \dots, 6$ and $j = 1, \dots, 4$. The error terms δ_h and ϵ_j (for $h = 1, \dots, 6$ and $j = 1, \dots, 4$) are considered measurement errors associated with their manifest variable (x'_h 's and y'_j 's, respectively), which are assumed to be uncorrelated. The disturbance terms ζ_1 and ζ_2 , on the other hand, account for unspecified terms or misspecification in the model in their respective latent variables, η_1 and η_2 . The structural equation, measurement equation for the endogenous variables and the measurement equation for the exogenous variables in the path diagram (Figure 2.4) are presented in

Equation 2.32, Equation 2.33 and Equation 2.34, respectively, i.e,

$$\begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ \beta_{12} & 0 \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} + \begin{bmatrix} \gamma_{11} & 0 & 0 \\ \gamma_{21} & \gamma_{22} & \gamma_{23} \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{bmatrix} + \begin{bmatrix} \zeta_1 \\ \zeta_2 \end{bmatrix}, \quad (2.32)$$

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} \lambda_{y11} & 0 \\ \lambda_{y21} & 0 \\ 0 & \lambda_{y32} \\ 0 & \lambda_{y42} \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \end{bmatrix}, \quad (2.33)$$

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = \begin{bmatrix} \lambda_{x11} & 0 & 0 \\ \lambda_{x21} & 0 & 0 \\ 0 & \lambda_{x32} & 0 \\ 0 & \lambda_{x42} & 0 \\ 0 & 0 & \lambda_{x53} \\ 0 & 0 & \lambda_{x63} \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{bmatrix} + \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \\ \delta_6 \end{bmatrix}. \quad (2.34)$$

Further, in a more general format, these equations can be written in a matrix form, such as Equations 2.35, 2.35) and 2.35,

$$\boldsymbol{\eta} = \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta}, \quad (2.35)$$

$$\mathbf{y} = \boldsymbol{\Lambda}_y\boldsymbol{\eta} + \boldsymbol{\epsilon}, \quad (2.36)$$

$$\mathbf{x} = \boldsymbol{\Lambda}_x\boldsymbol{\xi} + \boldsymbol{\delta}. \quad (2.37)$$

The most commonly used approaches to estimate parameters for the structural equation modeling are the covariance and variance based structures. The covariance based model is sometimes referred simply as SEM, or by LISREL (Linear Structural Relationships) which is the name of the first software developed to use this technique. The variance based method, is also called partial least squares (PLS) and was first introduced by Wold (1973, 1975) as nonlinear iterative partial least squares (NIPALS). Both approaches are valid under different circumstances and require different sets of assumptions, which are described in the following sections.

Models can be defined as common factor or composite factor models. The common factor model sets the covariance among indicators within the same latent variable to 0, estimating only the variance (which is usually 1, since the manifest variables are set to have mean 0 and unit variance). The composite factor model, on the other hand, also estimates the covariance among indicators within the same latent variable.

2.5.1 Covariance Based Estimation - Maximum Likelihood

The covariance based structural equation modeling was initially implemented via LISREL (Linear Structural Relationships), a software package introduced by Joreskog and van Thillo (1972). Many updates for LISREL have been released since 1972, and this technique has also become available in other commonly used statistical analysis softwares. For more information regarding the historical foundations of SEM, refer to Kaplan (2009).

Modeling structural equations requires a high level of knowledge regarding which variables might affect the latent variable of interest, since the omission of important variables can lead to an issue called specification error (described in Diamantopoulos, 1994). Specification error occurs when the model being analyzed does not reflect the real population model. The SEM is a large sample technique and even though there is no general rule specifying the minimum necessary sample size, Boomsma (1985) suggested using at least 200 observations.

To ensure an appropriate estimation of the parameters, it is necessary to verify the *model identification*, meaning that the information in the empirical covariance matrix is sufficient to estimate all of the parameters in the model. Hence, “a necessary condition for identification is that the number of independent parameters being estimated is less than or equal to the number of non-redundant elements of \mathbf{S} , the sample matrix of covariances among the observed variables” (Long, 1983).

Consider the loadings associated with each latent variable in $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$. These latent variables have no defined unit of measurement, which causes the same issue well known in factor analysis, specifically the *scale determinancy*. To avoid this problem, O’Rourke and Hatcher (2013) suggests fixing one loading of each latent variable at 1, and they explain that “by fixing a path at 1 from the F variable to one of its manifest indicators, the unit of measurement for the F variable is set to the metric of measurement for that indicator variable (minus its error term)”, where F is a latent variable.

The estimation of the parameters of the structural equation modeling can be done using many approaches, such as maximum likelihood, unweighted least squares, generalized least squares, diagonally weighted least squares, etc. The maximum likelihood (ML) approach is the most commonly used, and is established under the assumption that the manifest variables follow a multivariate Normal distribution. This method is based on minimization of the discrepancy between the theoretical and the empirical covariance matrix. The discrepancy function can be obtained using the likelihood ratio, that compares the theoretical model and a perfect fitting model, see Crisci (2012).

Considering the model specifications presented in Equations 2.35, 2.36 and 2.37, it is necessary to assume that the expectation of the error and disturbance terms are $\mathbf{0}$, as well as the expectation of the exogenous latent variables, $\boldsymbol{\xi}$ are $\mathbf{0}$. The covariance matrix of the latent exogenous variables is the symmetrical matrix $\boldsymbol{\Phi}$, where the main diagonal is composed of 1’s, since the latent variables are standardized, and have unit variance. The

error and disturbance terms, on the other hand, are assumed to be uncorrelated, such that $Cov(\zeta) = \Psi$ is a diagonal matrix covariance matrix, and $Cov(\epsilon) = \Theta_\epsilon$ and $Cov(\delta) = \Theta_\delta$ are either diagonal or block diagonal covariance matrices. If the model is a common factor model, Θ_ϵ and Θ_δ are diagonal, whilst Θ_ϵ and Θ_δ are block diagonal if the model is a composite factor model.

Then, the i^{th} observation, $\mathbf{Y}_i = (y_{i1}, \dots, y_{ip})^T$ and $\mathbf{X}_i = (x_{i1}, \dots, x_{iq})^T$, where $i = 1, \dots, n$, and p is the number of endogenous manifest variables, and q is the number of exogenous manifest variables, assumes the following distribution:

$$\mathbf{Z}_i = \begin{pmatrix} \mathbf{Y}_i \\ \mathbf{X}_i \end{pmatrix} \sim N_{p+q} \left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \Sigma = \begin{pmatrix} \Sigma_{yy} & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_{xx} \end{pmatrix} \right) \quad (2.38)$$

where

$$\Sigma = \begin{pmatrix} \Lambda_y(I - B)^{-1}(\Gamma\Phi\Gamma' + \Psi)^{-1}(I - B)^{-1}\Lambda_y + \Theta_\epsilon & \Lambda_y(I - B)^{-1}\Gamma\Phi\Lambda'_x \\ \Lambda_x\Phi\Gamma'(I - B)^{-1}\Lambda'_y & \Lambda_x\Phi\Lambda'_x + \Theta_\delta \end{pmatrix}. \quad (2.39)$$

Since \mathbf{Z}_i follows a multivariate normal distribution with a mean $\mathbf{0}$, Wishart (1928) showed that the product of such multivariate normal distributions follows a Wishart distribution such as:

$$\mathbf{Z}^T \mathbf{Z} = \sum_{i=1}^n \mathbf{Z}_i^T \mathbf{Z}_i \sim \mathbf{W}_{p+q}(\Sigma, n).$$

Using the multivariate version of Cochran's Theorem which can be found in Wong et al. (1991), for \mathbf{C} idempotent, we have

$$\mathbf{Z}^T \mathbf{C} \mathbf{Z} \stackrel{d}{=} \mathbf{W}_{p+q}(\Sigma, \text{rank}(\mathbf{C}));$$

then,

$$(n-1)\mathbf{S} = \mathbf{Z}^T \left(\mathbf{I} - \frac{\mathbf{1}\mathbf{1}'}{n} \mathbf{J} \right) \mathbf{Z} \stackrel{d}{=} \mathbf{W}_{p+q}(\boldsymbol{\Sigma}, (n-1))$$

where \mathbf{J} is a square $(n \times n)$ matrix filled with 1's, \mathbf{I} is the identity matrix, and \mathbf{S} is the empirical covariance matrix. Hence, the probability density function of $(n-1)\mathbf{S}$ can be written as follows

$$\begin{aligned} f_{(n-1)\mathbf{S}} &= \frac{|(n-1)\mathbf{S}|^{\frac{(n-1)-(p+q)-1}{2}} e^{-\frac{1}{2}tr(\boldsymbol{\Sigma}^{-1}(n-1)\mathbf{S})}}{2^{\frac{(n-1)(p+q)}{2}} \pi^{\frac{(p+q)(p+q-1)}{4}} |\boldsymbol{\Sigma}|^{\frac{n-1}{2}} \prod_{i=1}^{p+q} \Gamma\left(\frac{(n-1)+1-i}{2}\right)} \quad (2.40) \\ &= \frac{e^{-\frac{n-1}{2}tr(\boldsymbol{\Sigma}^{-1}\mathbf{S})}}{|\boldsymbol{\Sigma}|^{\frac{n-1}{2}}} \times \frac{|(n-1)\mathbf{S}|^{\frac{(n-1)-(p+q)-1}{2}}}{2^{\frac{(n-1)(p+q)}{2}} \pi^{\frac{(p+q)(p+q-1)}{4}} \prod_{i=1}^{p+q} \Gamma\left(\frac{(n-1)+1-i}{2}\right)} \\ &= e^{-\frac{n-1}{2}tr(\boldsymbol{\Sigma}^{-1}\mathbf{S})} |\boldsymbol{\Sigma}|^{-\frac{n-1}{2}} \times k \end{aligned}$$

where k is a constant term.

The function to be minimized using the ML estimation is the discrepancy function, that can be obtained using the likelihood ratio of the likelihood of the hypothesized model and the likelihood of a perfect fit model,

$$\begin{aligned} L &= \frac{e^{-\frac{n-1}{2}tr(\boldsymbol{\Sigma}^{-1}\mathbf{S})} |\boldsymbol{\Sigma}|^{-\frac{n-1}{2}} \times k}{e^{-\frac{n-1}{2}tr(\mathbf{S}^{-1}\mathbf{S})} |\mathbf{S}|^{-\frac{n-1}{2}} \times k} \\ &= e^{-\frac{n-1}{2}tr(\boldsymbol{\Sigma}^{-1}\mathbf{S})} |\boldsymbol{\Sigma}|^{-\frac{n-1}{2}} |\mathbf{S}|^{\frac{n-1}{2}} e^{\frac{n-1}{2}tr(\mathbf{I})} \\ &= e^{-\frac{n-1}{2}tr(\boldsymbol{\Sigma}^{-1}\mathbf{S})} |\boldsymbol{\Sigma}|^{-\frac{n-1}{2}} |\mathbf{S}|^{\frac{n-1}{2}} e^{\frac{(n-1)(p+q)}{2}}. \end{aligned}$$

Finally, to estimate the parameters of the structural equation models, the log-likelihood

of the discrepancy function is

$$\begin{aligned} \log(L) &= -\frac{n-1}{2} \text{tr}(\boldsymbol{\Sigma}^{-1} \mathbf{S}) - \frac{n-1}{2} \log(|\boldsymbol{\Sigma}|) + \frac{n-1}{2} \log(|\mathbf{S}|) + \frac{(n-1)(p+q)}{2} \\ &\propto \text{tr}(\boldsymbol{\Sigma}^{-1} \mathbf{S}) + \log(|\boldsymbol{\Sigma}|) - \log(|\mathbf{S}|) - (p+q). \end{aligned} \quad (2.41)$$

The minimization of the log likelihood can be done using a numerical method, such as Newton Raphson.

2.5.2 Variance Based Estimation - Partial Least Squares Path Modeling

The partial least squares (PLS) estimation was initially introduced as a nonlinear iterative partial least squares (NIPALS) by Wold (1973, 1975) and is an alternative approach to the covariance based SEM. The PLS method is particularly of interest when the study has a very large number of manifest variables (and covariance SEM reaches its limits), as described by Haenlein and Kaplan (2004), and/or for analysis with small sample sizes. Simulation studies performed by Chin and Newsted (1999) have shown that PLS can be done with sample sizes as low as 50, depending on the effect size of interest.

The term partial least squares was first used by Herman Wold (Wold, 1979) and referred to the variance approach to model latent variables, used in scenarios similar to the covariance based SEM. Later in 1983, his son Svante Wold, published an article using the same terminology, the partial least squares method (Wold et al., 1983), while presenting a regression method to predict one or more dependent variables based on a set of independent variables. This method is known as the partial least squares regression (PLS-R) and is highly used in chemometrics, specially when the data present multicollinearity and/or the number of observations is less than the number of variables. An approach to partial least squares regression

using interval-valued variables was proposed by Imakor (2007). To avoid confusion between the two terms, “following a suggestion by H. Martens, we have decided to name ‘PLS Path Modeling’ the use of PLS for structural equation modeling”, Tenenhaus et al. (2005). The focus of this project is on partial least squares path modeling (PLS-PM).

The PLS-PM structure is composed of an extra component, the weight relations. The weight relations calculate the estimated latent variables via an iterative two-step procedure, that are used in the estimation of the parameter of both the structural and measurement models, using the ordinary least squares method. As a contrast to the covariance based SEM, the main goal of PLS-PM is the prediction of the latent variables, instead of the interpretation of the parameters, and since PLS-PM does not distinguish between endogenous and exogenous variables, the measurement equation is composed of only one matrix equation.

The PLS-PM allows two different types of manifest variables for the latent variables, viz., the formative and the reflective manifest variables (the reflective way is the only type of latent and manifest variables combination modeled by the covariance SEM approach). Formative variables are those such that the manifest variables are believed to form the latent variable, they are considered to be the cause of the latent variable. Figure 2.5 shows an example of formative manifest variables (also known as the formative way) related to the latent variable ‘quality of life’. In this example, ‘quality of life’ is assumed to be formed by ‘living environment’, ‘financial stability’ and ‘health’. The structural and measurement equations for the formative way are presented in Equations 2.42 and 2.43, as

$$\boldsymbol{\xi} = \mathbf{B}\boldsymbol{\xi} + \boldsymbol{\zeta}, \quad (2.42)$$

$$\boldsymbol{\xi} = \boldsymbol{\Lambda}\mathbf{x} + \boldsymbol{\delta}. \quad (2.43)$$

Reflective variables (or the reflective way) are used more commonly, and are believed

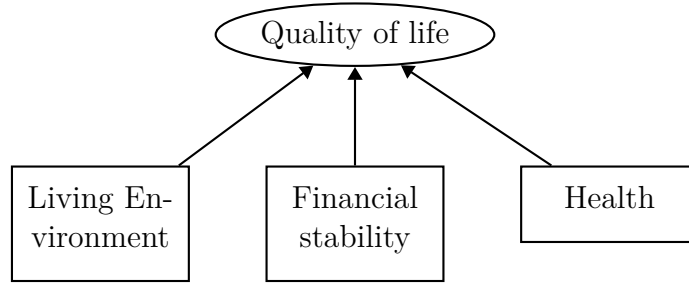


Figure 2.5: Example of Formative Manifest Variable

to be a reflection of the latent variable, such that the manifest variables depend on the latent variable. Figure 2.6 shows the same latent variable as the example of a formative way, but using reflective variables. In this example the latent variable ‘quality of life’ is assumed to depend on ‘happiness’, ‘productivity’ and ‘emotional stability’. The structural and measurement equations for the reflective way are presented in Equations 2.44 and 2.45, i.e.,

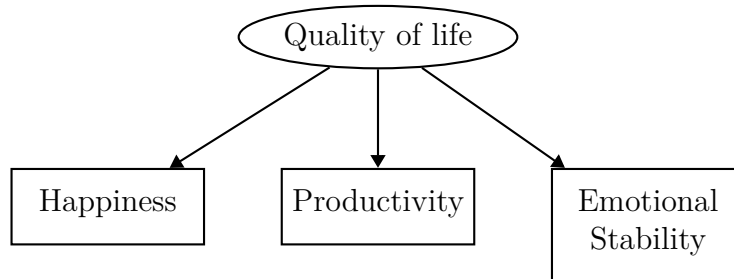


Figure 2.6: Example of Reflective Manifest Variable

$$\xi = B\xi + \zeta, \tag{2.44}$$

$$\mathbf{x} = \Lambda\xi + \epsilon. \tag{2.45}$$

It is possible to have both types of latent variables in the same analysis. This is called MIMIC and happens when it includes a mixture of formative and reflective latent variables.

Besides the direction of the causality assumed between latent and manifest variables, the reflective and formative ways are also expected to behave differently with respect to the relationship among its manifest variables. In a reflective way, it is necessary that the block presents unidimensionality, such that the manifest variables are measuring the same latent variable, and are correlated. Tenenhaus et al. (2004) suggests the use of Cronbach’s alpha (see Cronbach, 1951), Dillon-Goldstein’s ρ (see Dillon and Goldstein, 1984) or principal component analysis of a block to verify the construct unidimensionality. The formative way, on the other hand, does not require the manifest variables to be correlated. The construct unidimensionality ensures that there is internal consistency within a set of manifest variables related to a latent variable, such that the manifest variables are in fact measuring the same latent variable.

One of the advantages of PLS-PM is due to this approach not requiring the manifest variables to follow a multivariate normal distribution. Instead, it assumes the *predictor specification*: “This requirement states that the systematic part of the linear regression must be equal to the conditional expectation of the dependent variable”, Haenlein and Kaplan (2004). By rewriting the structural equation as $\xi_j = \sum_h \beta_{jh} \xi_h + \zeta_j$, for each j^{th} latent variable, the predictor specification assumption on this model is

$$E(\xi_j | \forall \xi_h) = \sum_h \beta_{jh} \xi_h$$

and $Cov(\zeta_j, \xi_h) = \mathbf{0}$, for all $h = 1, \dots, p$, where p is the total number of latent variables. The prediction specification for the measurement model in a formative way is $E(\xi | \mathbf{x}) = \mathbf{\Lambda} \mathbf{x}$, while the prediction specification for the measurement model in a reflective way is $E(\mathbf{x} | \xi) = \mathbf{\Lambda} \xi$.

The PLS-PM is sometimes referred to as a “soft model”, due to its flexibility in terms of assumptions. This flexibility introduces a problem that is called *consistency at large*. McDonald (1996) explains that the concept of consistency at large means “that the path

coefficients estimated through PLS converge on the parameters of the latent-variable model as both sample size and the number of indicators of each latent variable become infinite”.

The three components (weight relations, measurement model and structural model) of the PLS-PM as well as the estimation methods are presented below.

2.5.2.1 Weight Relations

The analysis of PLS-PM is done with a centered and scaled (to unit variance) design matrix, and starts by estimating the latent variables using the weight relations, ensuring that these estimated values have unit variance. The estimated value of each latent variable is obtained by using a weighted sum of its indicators, such as

$$\hat{\xi}_j = L_j = \sum_{k=1}^{k_j} w_{jk} \mathbf{x}_{jk},$$

for the j^{th} latent variable and the k_j indicators associated with the j^{th} latent variable, $j = 1, \dots, p$.

The estimation of the weight relations is done via an iterative process that consists of calculating outer and inner weights (based on the structural and measurement equations, respectively) and updating the latent variable estimation (scaled to unit variance) at each step, until convergence to a predetermined criteria. The weights can be initialized with an arbitrary number, and are usually set to 1. These initial weights are not technically outer weights, and to distinguish them from the actual outer weights, we use the notation \tilde{w}_{jk} .

The full iterative process to estimate the latent variable is described by Sanchez (2013), and is outlined below.

Step 1 - External Estimation

The external estimation of each iteration is obtained by using the weighted sum of the manifest variables, such as

$$\mathbf{L}_j \propto \pm \sum_k \tilde{w}_{jk} \mathbf{x}_{jk}, \quad j = 1, \dots, p,$$

where the \propto sign indicates that the latent variable needs to be set to unit variance, meaning that the j^{th} estimation of the latent variable should be divided by its standard deviation. The \pm sign, on the other hand, refers to the *sign ambiguity* that can happen when manifest variables do not have the same correlation sign with the estimated LV. It is suggested to “choose the sign so that the majority of the \mathbf{x}_{jk} is positively correlated with \mathbf{L}_j ” (Sanchez, 2013).

Then, for $w_{jk} = \tilde{w}_{jk}/sd(\mathbf{L}_j)$ and paying attention to the sign ambiguity, we have

$$\mathbf{L}_j = \sum_k w_{jk} \mathbf{x}_{jk}.$$

Step 2 - Inner Weights

The inner weights, e_{hj} , are calculated using the structural equation and the relations assumed between the latent variables (represented in the path diagram, e.g., see Figure 2.4),

$$\mathbf{Z}_j = \sum_{h \longleftrightarrow j} e_{hj} \mathbf{L}_h \quad (2.46)$$

where $h \longleftrightarrow j$ means that the path diagram indicates a connection between the h^{th} and j^{th} latent variables (it does not matter the direction of the connection). The e_{hj} is calculated using one of the three available weighting schemes, viz., the centroid (Inner Weight Scheme I), factor (Inner Weight Scheme II) or the path scheme (Inner Weight Scheme III). Studies

have shown (e.g., Noonan and Wold, 1982) that the choice of weighting scheme does not affect significantly the estimation, but the three schemes will be presented below.

Inner Weight Scheme I The centroid scheme, which is the scheme originally introduced by Wold (1982), considers only the direction of the correlation between the adjacent latent variables,

$$e_{hj} = \begin{cases} \text{sign}[\text{cor}(\mathbf{L}_h, \mathbf{L}_j)], & \text{if } L_h \text{ and } L_j \text{ are adjacents,} \\ 0, & \text{otherwise.} \end{cases}$$

Inner Weight Scheme II The factor scheme takes into consideration the strength of the correlation between the adjacent latent variables,

$$e_{hj} = \begin{cases} \text{cor}(\mathbf{L}_h, \mathbf{L}_j), & \text{if } L_h \text{ and } L_j \text{ are adjacents,} \\ 0, & \text{otherwise.} \end{cases}$$

Inner Weight Scheme III In the path scheme, the latent variables are divided into two categories, the followers and the predictors. A latent variable h is considered a *follower* of L_j if it is caused by the latent variable j . Its inner weight is obtained by finding the correlation between the latent variables, such as $e_{hj} = \text{cor}(\mathbf{L}_h, \mathbf{L}_j)$.

A latent variable h is considered a *predictor* of L_j if it is the reason why the adjacent latent variable j occurs. The inner weight (e_{hj}) is the coefficient estimation of the h^{th} latent variable obtained by regressing the j^{th} latent variable with all of its predictors L_h 's.

Note that in the path scheme $e_{hj} \neq e_{jh}$. From the path diagram, a latent variable h is considered a follower of the latent variable j if both h and j are directly connected, and the arrow is pointing to the h^{th} latent variable (see, e.g., in Figure 2.4, the latent variable η_2 is considered a follower of η_1). Similarly, a latent variable h is considered a predictor of the latent variable j if both h and j are directly connected, and the arrow is pointing to the j^{th}

latent variable (see, e.g., in Figure 2.4, the latent variable η_1 is considered a predictor of η_2).

Step 3 - Update Outer Weights

After the inner weights have been calculated and the latent variables estimated using Equation 2.46, it is necessary to update the outer weights. The updated outer weights can be calculated in two different ways, depending on the nature of the manifest variables at hand. If the manifest variables are constructed in a reflective way, Mode A (described below) should be used. If instead the manifest variables are constructed in a formative way, Mode B should be used.

Mode A

Mode A is preferred when the manifest variables relate to the latent variables in a reflective way, and estimates the outer weights by using a simple regression, such as,

$$\tilde{w}_{jk} = (\mathbf{Z}'_j \mathbf{Z}_j)^{-1} \mathbf{Z}'_j \mathbf{x}_{jk}. \quad (2.47)$$

Mode B

Mode B is preferred when the manifest variables relate to the latent variables in a formative way, and estimates the outer weights by using a multiple regression of \mathbf{x}_j on \mathbf{Z}_j , such as,

$$\tilde{\mathbf{w}}_j = (\mathbf{x}'_j \mathbf{x}_j)^{-1} \mathbf{x}'_j \mathbf{Z}_j. \quad (2.48)$$

Once Steps 1 - 3 have been completed, the latent estimation and outer weights can be obtained using Step 1. Repeat these steps until the weights w_{jk} have converged.

2.5.2.2 Structural Model Estimation

After the latent variables have been estimated using the weight relations, the estimation of the path coefficients comes down to an ordinary least squares estimation without intercept (since the latent variables are centered and have unit variance) of the model

$$\mathbf{L}_j = \sum_{h \rightarrow j} \beta_{jh} \mathbf{L}_h + \zeta_j$$

which provides the following solution to the estimated matrix of coefficients \mathbf{B} :

$$\hat{B}_{hj} = (\mathbf{L}'_h \mathbf{L}_h)^{-1} \mathbf{L}'_h \mathbf{L}_j.$$

2.5.2.3 Measurement Model Estimation

The estimation of the loadings associated with the measurement model can be done in a similar manner as the one presented for the structural component, but taking into consideration the nature of the relationship between the manifest and the latent variables, whether they are reflective or formative.

For the reflective way, the estimation is done via

$$\hat{\lambda}_{jk} = (\mathbf{L}'_j \mathbf{L}_j)^{-1} \mathbf{L}'_j \mathbf{x}_{jk}. \quad (2.49)$$

For the formative way, the estimation is done as

$$\hat{\lambda}_{jk} = (\mathbf{x}'_{jk} \mathbf{x}_{jk})^{-1} \mathbf{x}'_{jk} \mathbf{L}_j. \quad (2.50)$$

Since this model does not include an intercept, Equation 2.49 and 2.50 can be simplified

to the form presented in Equation 2.51

$$\hat{\lambda}_{jk} = \text{cor}(\mathbf{L}_j, \mathbf{x}_{jk}). \quad (2.51)$$

2.5.2.4 Unidimensionality of a construct

Unidimensionality of a block or of a construct is an assumption made in studies when the manifest variables are related to the latent variable in a reflective way. Since this type of relationship assumes that the manifest variables depend on the latent variable, it is expected that all of them are measuring the same factor, or latent variable. There are many ways to check for unidimensionality, and perhaps the most used is the Cronbach's alpha, α . The Cronbach's alpha was initially proposed by Cronbach (1951), and is calculated considering the correlation among the manifest variables, as long as they are positively correlated.

For standardized manifest variables,

$$\alpha_j = \frac{\sum_{k \neq k'} \text{cor}(\mathbf{x}_{jk}, \mathbf{x}_{jk'})}{k_j + \sum_{k \neq k'} \text{cor}(\mathbf{x}_{jk}, \mathbf{x}_{jk'})} \times \frac{k_j}{k_j - 1} \quad (2.52)$$

where $k = 1, \dots, k_j$, $k' = 1, \dots, k_j$, k_j is the number of manifest variables associated with the j^{th} latent variable, and $k \neq k'$. Keeping in mind that this estimate is affected by the number of manifest variables in a construct, it is desired that α presents large values.

Another measure to verify a block unidimensionality is the composite reliability, which is also known as the Dillon-Goldstein's ρ (Dillon and Goldstein, 1984). The composite reliability is usually preferred, since it does not assume that the indicators have the same weight over the latent variable, and therefore, "alpha tends to be a lower bound estimate of reliability, whereas ρ is a closer approximation under the assumption that the parameter estimates are accurate" (Chin, 1998).

The composite reliability is developed under the assumption that the manifest variables

are positively correlated among themselves, as well as with the latent variable of interest, and it can be estimated as,

$$\hat{\rho} = \frac{[\sum_{k=1}^{k_j} cor(\mathbf{x}_{jk}, \mathbf{t}_1)]^2}{[\sum_{k=1}^{k_j} cor(\mathbf{x}_{jk}, \mathbf{t}_1)]^2 + \sum_{k=1}^{k_j} (1 - cor^2(\mathbf{x}_{jk}, \mathbf{t}_1))} \quad (2.53)$$

where \mathbf{t}_1 is the first principal component of the construct, and k_j is the number of indicators associated with the j th latent variable.

2.5.2.5 Model Validation

Tenenhaus et al. (2005) showed that the validity of a partial least squares path model can be assessed through its individual parts using communality or redundancy, and globally, using a goodness of fit criteria proposed by Amato et al. (2004). The adjustment of the measurement model is measured by the communality index for each latent variable, that calculates the correlation between estimated latent variables and its manifest variables,

$$communality_j = \frac{1}{k_j} \sum_{k=1}^{k_j} cor^2(\mathbf{x}_{jk}, \mathbf{L}_j) \quad (2.54)$$

where k_j is the number of manifest variables associated with the j th latent variable.

It might also be of interest to calculate the average communality, since it is a measure of the overall variability shared by the latent variables with the manifest variables,

$$\overline{communality} = \frac{1}{(\sum_{j=1}^g k_j)} \sum_{j=1}^g k_j \times communality_j \quad (2.55)$$

where g is the total number of latent variables and k_j is the number of manifest variables associated with the j th latent variable. The structural component of the model can be assessed using the redundancy index, which combines the communality of the endogenous (or dependent) latent variable and the coefficient of determination estimated through the

endogenous and exogenous latent variables. Since the latent variables estimated in the PLS-PM method have been standardized, the coefficient of determination for each structural equation is

$$R_h^2 = \sum_{t \in t_h} \hat{\beta}_t \times cor(\mathbf{L}_h, \mathbf{L}_t) \quad (2.56)$$

where t_h is the set of latent variables that show a directional relationship with \mathbf{L}_h in the structural model, \mathbf{L}_h is the endogenous latent variable and \mathbf{L}_t 's are the exogenous latent variables.

Hence, the redundancy for the h^{th} equation in the structural model is such that

$$redundancy_h = communality_h \times R_h^2 \quad (2.57)$$

where $h = 1, \dots, s$, is the number of endogenous latent variables, or the number of equations in the structural component.

The last part of the model validation presented here was proposed by Amato et al. (2004) and sets a global criterion to measure the goodness of fit in a partial least squares path model. The goodness of fit, GoF, is a function of the average coefficient of determination and the average communality

$$GoF = \sqrt{communality \times \overline{R^2}} \quad (2.58)$$

where GoF can take values between 0 and 1.

Chapter 3

Partial Least Squares Path Modeling for Interval-valued data

3.1 Introduction

Structural equation modeling (SEM) is a modeling structure used to assess the directional relationship between latent and manifest variables. A latent variable (LV) is a hypothetical construct that cannot be directly observed, while a manifest variable (MV) is directly measured or observed, which is sometimes also called an indicator variable. For example, it might be difficult to find an unique variable that quantifies the health of a patient, but there exist many other variables that can provide an insight about it, such as lab results and daily habits; in this example, the lab results and daily habits can be seen as manifest variables, while the overall health of the patient is the latent variable.

Structural equation modeling presents a two component modeling structure for latent and manifest variables, with the first component being the measurement model, and the second component being the structural model. The overall goal of a SEM is to quantify the directional relationship (or structures) between the latent variables, while each of these

latent variables is measured by a set of manifest variables. This technique is of particular interest in behavioral sciences or other areas in which constructs (or latent variables) cannot be directly observed, but it is also important to model their relationship.

The nature of SEM is confirmatory; hence, the structure of the path model in the structural model expresses relationships obtained from extensive previous exploratory work. It is expected that manifest variables within the same latent variables present moderate to high correlation, whereas the variables between latent variables should not be highly correlated.

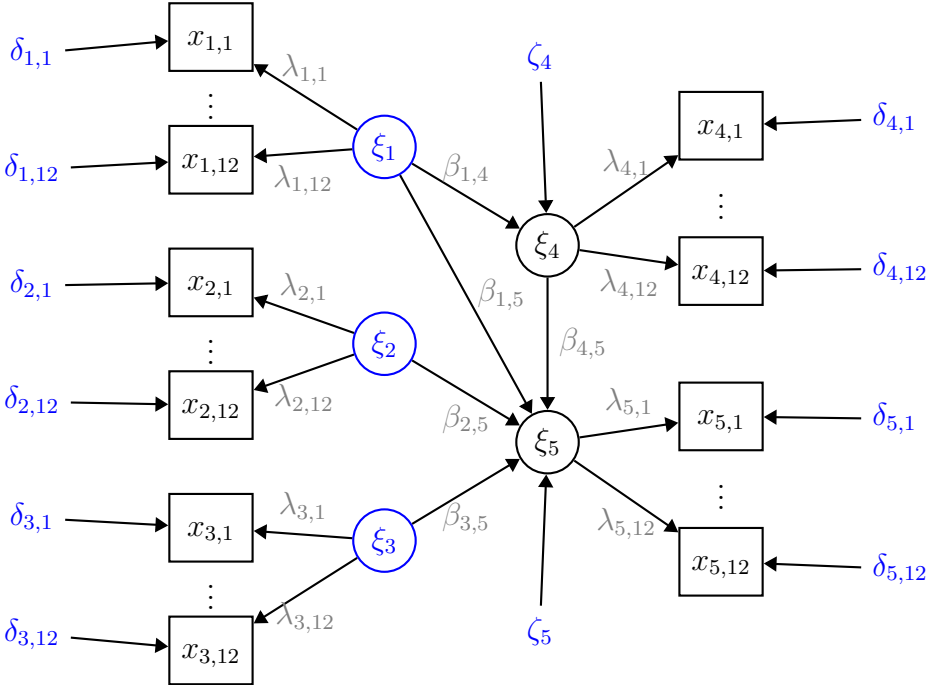


Figure 3.1: Path Diagram for PLS-PM, the reflective way

Figure 3.1 presents an example of a path diagram for a structural equation modeling where each manifest variable $x_{j,k}$ is associated with a latent variable ξ_j , forming the measurement model as $x_{j,k} = \lambda_{j,k} \times \xi_j + \delta_{j,k}$, for $j = 1, \dots, 5$ and $k = 1, \dots, 12$. Note that the term $\delta_{j,k}$ is a measurement error associated with each manifest variable. The structural model is set by the directional relationship between the latent variables, ξ_j , as $\xi_4 = \beta_{1,4} \times \xi_1 + \zeta_4$ and $\xi_5 = \beta_{1,5} \times \xi_1 + \beta_{2,5} \times \xi_2 + \beta_{3,5} \times \xi_3 + \beta_{4,5} \times \xi_4 + \zeta_5$, where ζ_4 and ζ_5 account for misspecification

in each model, respectively. The models are built without intercept because all variables are standardized, to avoid one or more variables having a higher impact in the analysis. Also, in this example, the latent variables ξ_1 , ξ_2 and ξ_3 act only as independent variables, whilst ξ_4 acts as both dependent and independent variables, and ξ_5 acts only as a dependent variable.

Referred generally as structural equation modeling, the relationships presented in Figure 3.1 are usually estimated using either covariance or variance based structures. The covariance based method (presented in Section 2.5.1) is developed based on the assumption of multivariate normality of the manifest variables ($x_{j,k}$'s), while the variance based method estimates the latent variables (ξ_j 's) through an iterative procedure, and uses ordinary least squares estimation for the parameters on both the structural and measurement components. The variance based structural equation modeling method, also called partial least squares path modeling (PLS-PM), was presented for classical data in Section 2.5.2; this chapter will study this method in depth for the interval-valued variables scenario.

The PLS-PM structure involves three components: the structural model, the measurement model and the weight relations. As a contrast to the covariance based SEM, the main goal of PLS-PM is the prediction of the latent variables. For the partial least squares estimation, the latent variables are predicted using the weight relations (calculated via an iterative two-step estimation), and the estimates for both models are obtained using ordinary least squares estimation. The equations for the structural and measurement models of the path diagram presented in Figure 3.1 is shown in Equations 3.1 and 3.2, respectively,

$$\begin{bmatrix} \xi_4 \\ \xi_5 \end{bmatrix} = \begin{bmatrix} \beta_{1,4} & 0 & 0 & 0 & 0 \\ \beta_{1,5} & \beta_{2,5} & \beta_{3,5} & \beta_{4,5} & 0 \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \\ \xi_4 \\ \xi_5 \end{bmatrix} + \begin{bmatrix} \zeta_4 \\ \zeta_5 \end{bmatrix} \quad (3.1)$$

$$\begin{bmatrix} x_{1,1} \\ \vdots \\ x_{1,12} \\ x_{2,1} \\ \vdots \\ x_{5,12} \end{bmatrix} = \begin{bmatrix} \lambda_{x_{1,1}} & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \lambda_{x_{1,12}} & 0 & 0 & 0 & 0 \\ 0 & \lambda_{x_{2,1}} & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \lambda_{x_{5,12}} \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \\ \xi_4 \\ \xi_5 \end{bmatrix} + \begin{bmatrix} \delta_{1,1} \\ \vdots \\ \delta_{1,12} \\ \delta_{2,1} \\ \vdots \\ \delta_{5,12} \end{bmatrix}. \quad (3.2)$$

For the partial least squares path modeling, the latent variables are related to the manifest variables either in a formative or a reflective way. The reflective way, which is portrayed in Figure 3.1 and Equations 3.2 and 2.32, is the most common, and assumes that the manifest variables are a reflection of the latent variables. When the variables are assumed to relate in a reflective way, it is assumed that there is unidimensionality within a construct, meaning that the manifest variables ($x_{j,k}$'s) associated with the same latent variables (ξ_j 's) are in fact measuring the same factor. While in the formative way the manifest variables ($x_{j,k}$'s) are believed to form the latent variable (ξ_j 's). Considering the same structure as in Figure 3.1, the measurement model, if the variables were related in a formative way, would be

$$\begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \\ \xi_4 \\ \xi_5 \end{bmatrix} = \begin{bmatrix} \lambda_{x_{1,1}} & \dots & \lambda_{x_{1,12}} & 0 & \dots & 0 & \dots & 0 & \dots & 0 \\ 0 & \dots & 0 & \lambda_{x_{2,1}} & \dots & \lambda_{x_{2,12}} & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & \dots & \dots & \dots & \dots & 0 & \lambda_{x_{5,1}} & \dots & \lambda_{x_{5,12}} \end{bmatrix} \begin{bmatrix} x_{1,1} \\ \vdots \\ x_{1,12} \\ x_{2,1} \\ \vdots \\ x_{5,12} \end{bmatrix} + \begin{bmatrix} \delta_1 \\ \delta_3 \\ \delta_3 \\ \delta_4 \\ \delta_5 \end{bmatrix}, \quad (3.3)$$

while the structural equation model would be the same as Equation 2.32. See Figures 2.5 and 2.6 for examples where the same latent variable is related in a formative and in a reflective way for different manifest variables.

Estimating the latent variables (ξ_j 's) is usually a hard task, since it is highly affected by the choice of the manifest variables, as well as the accuracy of these variables. Since it is a measure that can not be directly observed, it is also not easy to know how good the estimate is in reality. Given the nature of latent variables, it seems reasonable that the use of interval-valued variables might provide more trustworthiness to the estimation, since this takes into consideration the internal variability of the data. For example, suppose you have a set of manifest variables associated with the overall performance of schools for students in high school, using interval-valued variables to predict the latent variable 'overall performance in high school' might be a more sensitive approach, as the predicted value will include some of the variability that exists within the different students of the school.

Most of the symbolic approaches for interval-valued variables were constructed such that the parameter estimates match the estimates of the classical approaches when classical observations are transformed into interval-valued data. This attribute gives a broader usability to these techniques. By aiming to follow this same practice, the symbolic covariance and the center partial least squares path modeling (PLS-PM) methods presented in this chapter maintain the interval structure of the data as much as possible, while providing estimates with an easy interpretation.

This chapter will introduce the interval data PLS-PM approach using the symbolic covariance and the center regression methods in Section 3.2. For these methods, our approach will consist of the same three components presented in Chapter 2, i.e., the weight relations, structural model and the measurement model for interval-valued data. Then, we will present an example of application of the PLS-PM of interval-valued data for 32 football teams from NFL for the past 10 years, in Section 3.3.

3.2 Interval-valued PLS-PM

The partial least squares path modeling for interval-valued data method using both the symbolic covariance and center regression methods (described in Sections 2.4.1 and 2.4.4, respectively) are described in this section. Similarly to the classical approach, the interval-valued PLS-PM uses weight relations between the manifest and latent variables to estimate the latent variables, which are then used to estimate the parameters in the structural and measurement models. Note that this modelling approach does not receive an input of response variables, as the latent variables are not directly observed; instead, it estimates the latent variables using the weight relations calculated with the information in the structural and measurement models, and uses these estimates as response variables in the subsequent steps that use linear regression estimation.

From Section 2.5.2.1, we know that the weight relations are obtained by the iteration of the external and inner estimation of the latent variables, followed by the update of the outer weights using these estimates, repeated until convergence. The external estimation obtains an initial estimation of the latent variable by incorporating the relationship of the manifest variables as well as the weight associated with each of them; this weight is either an initial guess or the outer weights calculated in a previous iteration. Then, the inner estimation updates the estimation of the latent variable by including the relationship in the structural model. This relationship is usually also represented in a path diagram similar to the one presented in Figure 3.1.

For the interval-valued PLS-PM, the weights and parameter estimates do not differ for the lower and upper bounds of the interval, but are calculated using the internal structure of the data and through the desired regression approach, either the symbolic covariance or the center method. The estimation of each latent variable is obtained via a weighted sum of

its manifest variables, applied to the lower and upper bounds, such as

$$\mathbf{L}_{jL} = \sum_{k=1}^{k_j} w_{jk} \mathbf{x}_{jkL}, \quad (3.4)$$

$$\mathbf{L}_{jU} = \sum_{k=1}^{k_j} w_{jk} \mathbf{x}_{jkU} \quad (3.5)$$

where w_{jk} is the weight associated with the k^{th} manifest variable and the j^{th} latent variable, $\mathbf{x}_{jk} = [\mathbf{x}_{jkL}, \mathbf{x}_{jkU}]$ is the $(j, k)^{th}$ interval-valued manifest variable, for $j = 1, \dots, p$ and $k = 1, \dots, k_j$, where p is the number of latent variables and k_j is the number of manifest variables associated with the j^{th} latent variable, the capital letter subscript L indicates the interval lower bound of the interval-value variable and the capital letter subscript U indicates the interval upper bound of the variable.

The full process to obtain the weight relation between the latent and manifest variables, the parameter estimation of the loading and the structural coefficients is outlined below.

3.2.1 Weight Relations

Similarly to the classical approach, for the symbolic interval-valued PLS-PM, it is necessary to center and set the variance of the manifest variables to 1. This guarantees that all variables have the same impact in the analysis, and so they are not affected by a possible difference in unit or scale of variables. In fact, at all times that the latent variables are estimated, this normalization procedure is repeated.

The weights are calculated using the iterative process described below, which involves a combination of the information contained in the manifest variables as well as the relationship between the latent variables drawn in the structural model. The information in the manifest variables is taken into consideration in the external estimation, where the latent variables are initially estimated ($\mathbf{L}_j = [\mathbf{L}_{jL}, \mathbf{L}_{jU}]$) as the weighted average of the manifest variables

associated with it, using the outer weights obtained in the previous iteration. The information contained in the relationship between the latent variables is quantified in the inner estimation, by updating the estimated latent variables ($\mathbf{Z}_j = [\mathbf{Z}_{jL}, \mathbf{Z}_{jU}]$) as an weighted average of external estimation of the latent variables that present a directional relationship in its path diagram, using the inner weights (e_{jh}). To start this estimation process, the outer weights (\tilde{w}_{jk}) can be initialized by setting them to 1 or any other arbitrary value.

There are three step involved in calculating the weights, the external estimation of the latent variables, the inner estimation of the latent variables and the update of the outer weights, described as follows.

Step 1 - External Estimation

By using the weights (either the initial guess, or the update of the previous iteration), the external estimation of the latent variables consists of a weighted sum of the manifest variables (\mathbf{x}_{jk} 's) applied to its lower and upper bounds, for $j = 1, \dots, p$,

$$\begin{aligned}\mathbf{L}_{jL} &\propto \pm \sum_k \tilde{w}_{jk} \mathbf{x}_{jkL}, \\ \mathbf{L}_{jU} &\propto \pm \sum_k \tilde{w}_{jk} \mathbf{x}_{jkU}\end{aligned}$$

where $\mathbf{x}_{ik} = [\mathbf{x}_{jkL}, \mathbf{x}_{jkU}]$, \tilde{w}_{jk} 's are either the initial weights or the outer weights from the previous iteration, $k = 1, \dots, k_j$, ' \propto ' means that the right hand side of the equation should be standardized to have mean zero and standard deviation one, taking into consideration the interval-valued estimates produced ($\mathbf{L}_j = [\mathbf{L}_{jL}, \mathbf{L}_{jU}]$), and \pm indicates that there could be a *sign ambiguity*, that should be addressed to “choose the sign so that the majority of the \mathbf{x}_{jk} is positively correlated with \mathbf{L}_j ” (Sanchez, 2013).

In order to deal with the possibility of the sign ambiguity, after the interval-valued variables \mathbf{L}_j have been standardized, it is necessary to evaluate how many of the \mathbf{x}_{jk} are

positively or negatively correlated with \mathbf{L}_j . If the number of \mathbf{x}_{jk} that is positively correlated with \mathbf{L}_j is greater than the number that is negatively correlated, no correction is necessary, but if the number of \mathbf{x}_{jk} that is negatively correlated with \mathbf{L}_j is greater than the number that is positively correlated, there should be a sign change in \mathbf{L}_j .

Step 2 - Inner Estimation

The second step of the iterative process, that consists of estimating the inner weights (e_{jh} for $h, j = 1, \dots, p, h \neq j$), is responsible for incorporating the relationship between the latent variables described in the structural model, and is calculated as,

$$\mathbf{Z}_{jL} = \sum_{j \leftrightarrow h} e_{jh} \mathbf{L}_{jL}, \quad (3.6)$$

$$\mathbf{Z}_{jU} = \sum_{j \leftrightarrow h} e_{jh} \mathbf{L}_{jU} \quad (3.7)$$

where $j \leftrightarrow h$ means that the path diagram of interest (e.g., Figure 3.1) indicates a connection between the j^{th} and h^{th} latent variables (it does not matter the direction of the connection).

The inner estimation combines the external estimation of the latent variables from the previous step with their relationship introduced in the structural model (see, for example, Equation 3.1), by calculating a weighted average using the inner weights (e_{jh}). The inner weights carry information regarding the relationship held between the estimated latent variables. This information can be of the form of the sign of their correlation in the centroid scheme, in the form of the actual correlation in the factor scheme, or as the correlation or their coefficient estimation in a regression, depending on the direction of the relationship between the latent variables, for the path scheme. The direction of the relationship is expressed in its structural model and consequently in its path diagram.

The estimation of the inner weights (e_{jh}) does not differ from the classical way presented

in Chapter 2, Section 2.5.2.1, Inner Weight Schemes I, II and III, but the correlation is calculated using the interval-valued symbolic covariance (and consequently the interval-valued symbolic variance) presented in Equation 2.3 or in 2.4, and the linear regression coefficients (used in the path scheme) are obtained using the interval-valued symbolic formulas presented in Section 2.4.4, for the symbolic covariance method, or in Section 2.4.1, for the center method.

Step 3 - Update Outer Weights

Having calculated and standardized the inner estimate \mathbf{Z}_j of the latent variable for $j = 1, \dots, p$, the final weights for each iteration are then updated, to be used either as a starting point in the following iteration, or to estimate the latent variables to carry out the structural and measurement model estimations. These weights are called outer weights.

This step of the iterative process of the weight relations involves updating the outer weights, as well as the latent variables estimation, that should also be standardized. The method to update the outer weights depends on the way the manifest and latent variables are connected, whether they are connected in a reflective or formative way. If manifest variables are assumed to depend on the latent variables, relating in a reflective way, Mode A should be used. If on the other hand the manifest variables are considered to be the cause of the latent variable, relating in a formative way, Mode B should be used.

Note that \mathbf{L}_j , \mathbf{Z}_j and \mathbf{x}_{jk} are interval-valued variables, and when referred as such, this means that $\mathbf{x}_{jk} = [\mathbf{x}_{jkL}, \mathbf{x}_{jkU}]$, $\mathbf{L}_j = [\mathbf{L}_{jL}, \mathbf{L}_{jU}]$ and $\mathbf{Z}_j = [\mathbf{Z}_{jL}, \mathbf{Z}_{jU}]$, $j = 1, \dots, p$, $k = 1, \dots, k_j$.

Mode A - Reflective

Mode A is preferably used when the manifest variables are associated with the latent variables in a reflective way. By using the symbolic covariance method, the outer weights are calculated

as follows

$$\begin{aligned}\tilde{w}_{jk} &= (\mathbf{Z}'_j \mathbf{Z}_j)^{-1} \mathbf{Z}'_j \mathbf{x}_{jk} \\ &= (n \times \mathbf{Cov}(\mathbf{Z}_j, \mathbf{Z}_j))_{1 \times 1}^{-1} \times (n \times \mathbf{Cov}(\mathbf{Z}_j, \mathbf{x}_{jk}))_{1 \times 1}\end{aligned}\quad (3.8)$$

for $j = 1, \dots, p$, and $k = 1, \dots, k_j$, where p is the number of latent variables, k_j is the number of manifest variables associated with the j^{th} latent variable, n is the number of observations. This alternative form of the equation follows the same steps that were used to obtain Equation 2.26, given that both the manifest variables and the inner estimates are centered around zero.

When using the center method as the choice of regression method, the outer weights can be calculated as

$$\begin{aligned}\tilde{w}_{jk} &= (\mathbf{Z}'_j \mathbf{Z}_j)^{-1} \mathbf{Z}'_j \mathbf{x}_{jk} \\ &= ((\mathbf{Z}_j^c)' \mathbf{Z}_j^c)^{-1} (\mathbf{Z}_j^c)' \mathbf{x}_{jk}^c\end{aligned}\quad (3.9)$$

for $j = 1, \dots, p$, and $k = 1, \dots, k_j$, where \mathbf{Z}_j^c and \mathbf{x}_{jk}^c are the center interval-valued data such that $Z_{ij}^c = \frac{Z_{ijL} + Z_{ijU}}{2}$ and $x_{ijk}^c = \frac{x_{ijkL} + x_{ijkU}}{2}$ (see Equation 2.11 for interval-valued data centering equation), \mathbf{Z}_j is the estimated latent variable, p is the number of latent variables and n is the number of observations.

Mode B - Formative

Mode B is preferably used when the manifest variables are associated with the latent variables in a formative way. By using the same strategy presented in mode A, for the symbolic

covariance method, the outer weights can be written as

$$\begin{aligned}\tilde{\mathbf{w}}_j &= (\mathbf{x}'_j \mathbf{x}_j)^{-1} \mathbf{x}'_j \mathbf{Z}_j \\ &= (n \times \mathbf{Cov}(\mathbf{x}_{jk'}, \mathbf{x}_{jk''}))^{-1}_{k_j \times k_j} \times (n \times \mathbf{Cov}(\mathbf{x}_{jk}, \mathbf{Z}_j))_{k_j \times 1}\end{aligned}\quad (3.10)$$

where p is the number of latent variables, n is the number of observations, k_j is the number of manifest variables associated with the j^{th} latent variable, and $k, k', k'' = 1, \dots, k_j, j = 1, \dots, p$.

The outer weights for the center approach for the regression is

$$\begin{aligned}\tilde{\mathbf{w}}_j &= (\mathbf{x}'_j \mathbf{x}_j)^{-1} \mathbf{x}'_j \mathbf{Z}_j \\ &= ((\mathbf{x}_j^c)' \mathbf{x}_j^c)^{-1} (\mathbf{x}_j^c)' \mathbf{Z}_j^c\end{aligned}\quad (3.11)$$

where \mathbf{Z}_j^c and \mathbf{x}_{jk}^c are the center interval-valued data and estimated latent variable, respectively, $j = 1, \dots, p$.

This concludes the three iterative steps to obtain the weight relations. At the end of each iteration, the outer weights should be compared to the outer weights of the previous iteration to verify if it has converged, such as,

$$|\tilde{w}_{jk}^{s-1} - \tilde{w}_{jk}^s| < \varepsilon, \quad (3.12)$$

where \tilde{w}_{jk}^{s-1} is the outer weight of the previous iteration, \tilde{w}_{jk}^s is the outer weight of the current iteration and ε is the pre-determined tolerance.

If the outer weights have converged, Equations 3.13 and 3.14 are used to obtain a final estimation of the latent variables,

$$\mathbf{L}_{jL} \propto \sum_{k=1}^{k_j} \tilde{w}_{jk} \mathbf{x}_{jkL}, \quad (3.13)$$

$$\mathbf{L}_{jU} \propto \sum_{k=1}^{k_j} \tilde{w}_{jk} \mathbf{x}_{jkU} \quad (3.14)$$

where $j = 1, \dots, p$, $k = 1, \dots, k_j$, and \propto indicates that the estimate of the latent variable should be standardized to have mean zero and unit variance.

Otherwise, if the criteria in Equation 3.12 was not met, repeat steps 1-3 until the convergence condition of Equation 3.12 has been satisfied.

3.2.2 Structural Model Estimation

Once the weights have been estimated using the iterative process described above, it is possible to estimate the coefficients for the structural and measurement components by applying the ordinary least squares estimation method. For the structural model, the estimation of the parameters β (e.g., see Equation 3.1) is done using symbolic covariance linear regression, that was presented in Equation 2.26 or using the center method presented in Equation 2.14. Note that it is not necessary to estimate the intercept since all of the variables have been standardized and therefore the intercept is zero.

By using the relationship described in the structural model and also in the diagram path of interest, the equations to be estimated are:

$$\mathbf{L}_h = \sum_{j \rightarrow h} \beta_{jh} \mathbf{L}_j + \zeta_h,$$

for $h = p - s + 1, \dots, p$, where s is the number of independent latent variable, p is the overall number of latent variables, and $j \rightarrow h$ means that the path diagram of the model indicates a connection between the j^{th} and h^{th} latent variables.

3.2.3 Measurement Model Estimation

The last step of the partial least squares estimation procedure consists of obtaining the loadings associated with each of the manifest variables. Similarly to the classical approach presented in Chapter 2, this estimation depends on the way the manifest and latent variables are related. If the manifest and latent variables are related in a reflective way, the loadings associated with each manifest variable, when the center method is being used, are calculated as,

$$\hat{\lambda}_{jk} = ((\mathbf{L}_j^c)' \mathbf{L}_j^c)^{-1} (\mathbf{L}_j^c)' \mathbf{x}_{jk}^c \quad (3.15)$$

where \mathbf{L}_j^c is a vector containing the center (see Equation 2.11) of all interval-valued observations of the j^{th} latent variable, and \mathbf{x}_{jk}^c is a vector containing the center (see Equation 2.11) of the interval-valued observations of the k^{th} manifest variable associated with the j^{th} latent variable, for $j = 1, \dots, p$ and $k = 1, \dots, k_j$.

When the variables are related in a formative way, the loadings associated with each manifest variable, while using the center method can be calculated as

$$\hat{\lambda}_{jk} = ((\mathbf{x}_{jk}^c)' \mathbf{x}_{jk}^c)^{-1} (\mathbf{x}_{jk}^c)' \mathbf{L}_j^c. \quad (3.16)$$

When instead the covariance method is used, Equation 3.17 is used to obtain the loadings associated with each manifest variable, whether they are related in a reflective or formative way.

$$\begin{aligned} \hat{\lambda}_{jk} &= (m \times Cov(\mathbf{L}_j, \mathbf{L}_j))_{1 \times 1}^{-1} \times (m \times Cov(\mathbf{L}_j, \mathbf{x}_{jk}))_{1 \times 1} \\ &= cor(\mathbf{L}_j, \mathbf{x}_{jk}). \end{aligned} \quad (3.17)$$

Note that this simplification of the form to estimate the loadings in the symbolic covariance method happens because the model does not include an intercept, and the data have been standardized. Since the interval-valued data were standardized, and not their centers themselves, this simplification is not valid for the center method.

These steps complete the estimation process using the PLS-PM for interval-valued data. The estimates produced using this method are used mostly for prediction, but they can also provide insights of the relationship between the manifest variables and the latent variables, as well as the relationship between the latent variables. The quality of the predictions is highly affected by the information present in the structural and measurement models, since the omission of important manifest variables can lead to poor estimates. Sections 3.2.4 and 3.2.5 present ways to assess the unidimensionality of a construct, as well as the model validation methods and goodness of fit.

3.2.4 Unidimensionality of a construct

In the most common scenario, when manifest variables are related to latent variables in a reflective way, it is assumed that there is unidimensionality within a construct. A construct, also called a block sometimes, refers to each latent variable and the manifest variables associated with it; for example, in Figure 3.1, each latent variable ξ_j and the manifest variables $x_{j,1}, \dots, x_{j,12}$ compose a construct, for $j = 1, \dots, p$. The unidimensionality assumption ensures that the manifest variables are in fact quantifying the same factor, or the same latent variable. Manifest variables that are highly correlated will represent a construct with higher unidimensionality. Given that, in the formative way (described in Section 3.1) variables are believed to form the latent variable, they do not necessarily need to be correlated; therefore, this assumption is not made.

Among all of the methods available to verify the unidimensionality of a construct, Cronback's (1951) alpha α and Dillon-Goldstein's (1984) rho ρ (also known as composite reli-

ability) are the ones used more broadly, and are described in this section. Both of these methods require the manifest variables to be positively correlated, which usually can be done by adjusting the scale of the variables in the same direction.

The Cronbach's alpha, α , proposed by Cronback (1951), is calculated for each construct (i.e., each latent variable) as,

$$\alpha_j = \frac{\sum_{k \neq k'} \text{cor}(\mathbf{x}_{jk}, \mathbf{x}_{jk'})}{k_j + \sum_{k \neq k'} \text{cor}(\mathbf{x}_{jk}, \mathbf{x}_{jk'})} \times \frac{k_j}{k_j - 1} \quad (3.18)$$

where \mathbf{x}_{jk} are interval-valued variables centered at zero and set to have variance one, $\text{cor}(\mathbf{x}_{jk}, \mathbf{x}_{jk'})$ is calculated as in Equation 2.5, $k = 1, \dots, k_j$, $k' = 1, \dots, k_j$, k_j is the number of manifest variables associated with the j^{th} latent variable, and $k \neq k'$. Larger values of α_j indicate a higher unidimensionality of the construct. Tenenhaus et al. (2005) suggest that a block is unidimensional if its Cronbach's alpha is at least 0.7 for each $j = 1, \dots, p$.

As a measure of unidimensionality, Cronbach's alpha assumes that all manifest variables have the same weight over the latent variable, while the composite reliability takes into account the first principal component (from principal component analysis) considering all of the manifest variables. Principal component analysis uses an orthogonal transformation to create linearly independent variables from a set of correlated variables. The first principal component is of particular interest because it carries the most information and variability from the original data, when compared to the other principal components. For more information on classical principal component analysis, see e.g., Mardia et al. (1979) or Anderson (2003).

Besides the manifest variables being positively correlated, the composite reliability also assumes that the correlation between the manifest variables and the latent variable are

positive. It can be estimated as

$$\hat{\rho} = \frac{[\sum_{k=1}^{k_j} \text{cor}(\mathbf{x}_{jk}, \mathbf{t}_1)]^2}{[\sum_{k=1}^{k_j} \text{cor}(\mathbf{x}_{jk}, \mathbf{t}_1)]^2 + \sum_{k=1}^{k_j} (1 - \text{cor}^2(\mathbf{x}_{jk}, \mathbf{t}_1))} \quad (3.19)$$

where $j = 1, \dots, p$, $k = 1, \dots, k_j$, \mathbf{x}_{jk} are interval-valued variables centered at zero and set to have variance one, $\text{cor}(\mathbf{x}_{jk}, \mathbf{t}_1)$ is calculated as in Equation 2.5, \mathbf{t}_1 is the first principal component realization of the construct, and k_j is the number of indicators associated with the j^{th} latent variable. The first principal realization of the construct is obtained by applying the principal component analysis to the set of manifest variables associated with the j^{th} latent variable. There are currently a few different approaches for principal component analysis, which will not be described here; refer to Le-Rademacher (2008) and Le-Rademacher and Billard (2016) for more information.

3.2.5 Model validation

In every statistical method, it is important to develop techniques to evaluate its performance. For the partial least squares estimation, it is suggested (see Tenenhaus et al., 2005) that the model validation for each component is assessed individually - the structural and the measurement components - as well as the overall model. The fit of the measurement model can be evaluated using the communality of each construct; it is calculated as, for $j = 1, \dots, p$,

$$\text{communality}_j = \frac{1}{k_j} \sum_{k=1}^{k_j} \text{cor}^2(\mathbf{x}_{jk}, \mathbf{L}_j) \quad (3.20)$$

where the correlation for two interval-valued variables is calculated as in Equation 2.5, k_j is the number of manifest variables associated with the j^{th} latent variable. Furthermore, the average communality is calculated for the PLS-PM for interval-valued data in the same way it is done for the classical methods; see Equation 2.55.

The next component of the model to be evaluated is the structural model. Since the structural model has as many equations as there are latent variables that act as dependent variables, this is also the number of equations to be evaluated. For each equation of the dependent latent variable, a coefficient of determination can be calculated as

$$R_h^2 = \sum_{j \rightarrow h} \hat{\beta}_{jh} \times cor(\mathbf{L}_h, \mathbf{L}_j) \quad (3.21)$$

where $j \rightarrow h$ indicates that there is a directional relationship between the \mathbf{L}_j and \mathbf{L}_h in the structural model, \mathbf{L}_h is the dependent latent variable and \mathbf{L}_j 's are the independent latent variables, for $j \neq h$, $j = 1, \dots, p$, $h = p - s + 1, \dots, p$, where s is the number of equations in the structural model.

Hence, the redundancy for the h^{th} equation in the structural model is such that,

$$redundancy_h = communality_h \times R_h^2 \quad (3.22)$$

where $h = p - s + 1, \dots, p$ is the number of dependent latent variables, or the number of equations in the structural model.

The overall fit of the model is quantified using a goodness of fit measure (GoF) proposed by Amato et al. (2004). This index incorporates information about the indexes measuring the fit of both the structural and the measurement models, by using a function of the average of the coefficient of determination and the average communality,

$$GoF = \sqrt{communality \times \overline{R^2}}. \quad (3.23)$$

Note that GoF can take values between 0 and 1, and higher values indicate a better fit of the data to the model. This quantity should not be compared to any parametric measure, as no distributional assumption is made in the partial least squares path modeling estimation.

3.3 NFL Example

In recent years, there has been an increase in availability of data and information, including football data and statistics. In this section, to illustrate the partial least squares path modeling for interval-valued variables method presented in Section 3.2, interval-valued data from games for the 32 teams in the National Football League (NFL) for the past ten years are analyzed using a similar structure to the one presented in the help page of the R package ‘plsmp’, see Sanchez et al. (2017).

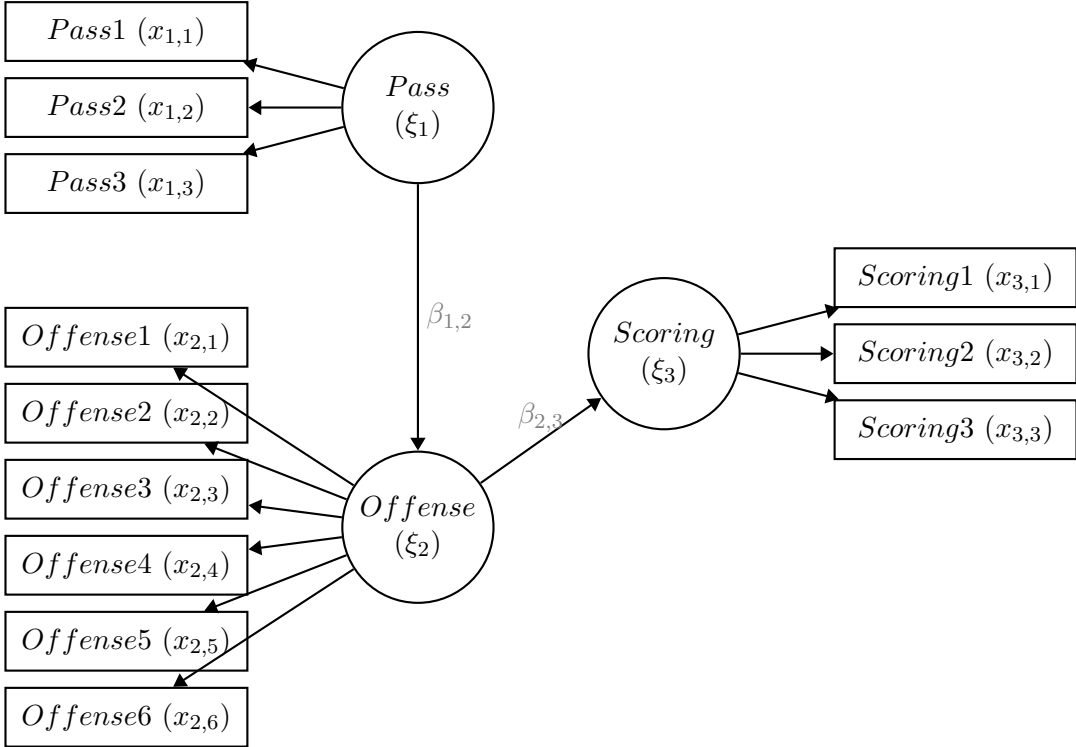


Figure 3.2: Path Diagram for NFL example

Figure 3.2 presents the path diagram for the latent and manifest variables in this example. The latent variable ‘Pass’ is believed to be a predictor of the latent variable ‘Offense’, which in turn is believed to be a predictor of ‘Scoring’. Each of these latent variables is related in a reflective way to a set of manifest variables. For example, the latent variable ‘Scoring’ is explained by the manifest variables average points per game, average offensive touchdowns

per game and the average overall touchdowns per game; the description of all manifest variables is presented in Table 3.1; thus, for example, the manifest variable ‘Pass1’ ($x_{1,1}$, see Figure 3.2) is the number of yards per completion of the pass.

The path diagram in Figure 3.2 can be translated into the measurement model equations presented in Equations 3.24, and also the structural model equations, presented in Equations 3.25 and 3.26, become,

$$\begin{aligned}
 \mathbf{x}_{1,1} &= \lambda_{1,1} \times \boldsymbol{\xi}_1 + \boldsymbol{\delta}_{1,1} & \mathbf{x}_{2,1} &= \lambda_{2,1} \times \boldsymbol{\xi}_2 + \boldsymbol{\delta}_{2,1} \\
 \mathbf{x}_{1,2} &= \lambda_{1,2} \times \boldsymbol{\xi}_1 + \boldsymbol{\delta}_{1,2} & \mathbf{x}_{2,2} &= \lambda_{2,2} \times \boldsymbol{\xi}_2 + \boldsymbol{\delta}_{2,2} \\
 \mathbf{x}_{1,3} &= \lambda_{1,3} \times \boldsymbol{\xi}_1 + \boldsymbol{\delta}_{1,3} & \mathbf{x}_{2,3} &= \lambda_{2,3} \times \boldsymbol{\xi}_2 + \boldsymbol{\delta}_{2,3} \\
 \mathbf{x}_{3,1} &= \lambda_{3,1} \times \boldsymbol{\xi}_3 + \boldsymbol{\delta}_{3,1} & \mathbf{x}_{2,4} &= \lambda_{2,4} \times \boldsymbol{\xi}_2 + \boldsymbol{\delta}_{2,4} \\
 \mathbf{x}_{3,2} &= \lambda_{3,2} \times \boldsymbol{\xi}_3 + \boldsymbol{\delta}_{3,2} & \mathbf{x}_{2,5} &= \lambda_{2,5} \times \boldsymbol{\xi}_2 + \boldsymbol{\delta}_{2,5} \\
 \mathbf{x}_{3,3} &= \lambda_{3,3} \times \boldsymbol{\xi}_3 + \boldsymbol{\delta}_{3,3} & \mathbf{x}_{2,6} &= \lambda_{2,6} \times \boldsymbol{\xi}_2 + \boldsymbol{\delta}_{2,6}
 \end{aligned} \tag{3.24}$$

$$\boldsymbol{\xi}_2 = \beta_{1,2} \times \boldsymbol{\xi}_1 + \boldsymbol{\zeta}_1 \tag{3.25}$$

$$\boldsymbol{\xi}_3 = \beta_{2,3} \times \boldsymbol{\xi}_2 + \boldsymbol{\zeta}_2. \tag{3.26}$$

Table 3.1: Manifest variables ($x_{..}$) description for NFL example

Variable	Name	Description	Variable	Name	Description
$x_{1,1}$	Pass1	Yards per completion	$x_{2,1}$	Offense1	Passing touchdowns per game
$x_{1,2}$	Pass2	Passed yards per game	$x_{2,2}$	Offense2	Rushing touchdowns per game
$x_{1,3}$	Pass3	Passing first downs per game	$x_{2,3}$	Offense3	Plays per game
$x_{3,1}$	Scoring1	Points per game	$x_{2,4}$	Offense4	Yards per play
$x_{3,2}$	Scoring2	Offensive touchdowns per game	$x_{2,5}$	Offense5	First downs per play
$x_{3,3}$	Scoring3	Touchdowns per game	$x_{2,6}$	Offense6	Time of possession percentage

The interval valued data consist of the aggregation of each of the variables described in Table 3.1 over a period of 10 years, from 2009 to 2018 (see Table 3.2 for full interval-valued

data). Information for the $n = 32$ football teams in the national league was obtained from the website “<https://www.teamrankings.com>”.

Table 3.2: NFL interval-valued data

Team	Pass1	Pass2	Pass3	Offense1	Offense2	Offense3	Offense4	Offense5	Offense6	Scoring1	Scoring2	Scoring3
Arizona	[8.9,12.9]	[157.7,287.3]	[8.6,14.8]	[0.6,2.2]	[0.4,1.2]	[56.4,67.9]	[4.1,6.2]	[0.242,0.356]	[0.4382,0.5242]	[14.1,29.4]	[1.2,3.1]	[1.6,3.5]
Atlanta	[9.5,12.8]	[218.6,298.9]	[12,15.8]	[1.3,2.5]	[0.7,1.3]	[61.9,67.4]	[4.9,6.7]	[0.312,0.385]	[0.4895,0.5366]	[21.2,34.1]	[2.3,7]	[2.2,4.1]
Baltimore	[8.3,11.9]	[189.4,266.9]	[10.5,13.2]	[1.2,1.8]	[0.4,1.4]	[62.8,70.2]	[4.5,5.7]	[0.271,0.341]	[0.4638,0.5349]	[20,26.1]	[1.6,2.8]	[1.8,2.9]
Buffalo	[9.6,11.3]	[157.2,231.6]	[7.9,11.6]	[0.8,1.5]	[0.4,1.8]	[56.9,69.8]	[4.7,5.7]	[0.256,0.324]	[0.4636,0.5218]	[16.1,24.9]	[1.4,2.9]	[1.6,3.1]
Carolina	[8.9,13]	[143.1,239.8]	[7.8,13.2]	[0.6,2]	[0.4,1.6]	[60.1,66.9]	[4.3,6.2]	[0.232,0.352]	[0.4568,0.5396]	[12.2,31.1]	[1.3,3]	[1.1,3.7]
Chicago	[9.6,11.5]	[175.7,268.1]	[9,13.4]	[0.8,2]	[0.4,0.9]	[58.4,64]	[4.9,6]	[0.27,0.34]	[0.4686,0.5304]	[16.5,27.8]	[1.6,2.8]	[1.9,3.2]
Cincinnati	[9.9,11.5]	[176.5,263.5]	[9.8,13.2]	[1.1,2]	[0.4,1.2]	[57.9,69.2]	[4.8,5.6]	[0.283,0.325]	[0.4526,0.5289]	[18.1,25.9]	[1.9,3]	[2.1,3.2]
Cleveland	[9.5,12.6]	[129.8,253.1]	[7.4,12.5]	[0.7,1.8]	[0.2,1.1]	[57.9,67.4]	[4.3,5.8]	[0.244,0.307]	[0.4591,0.5067]	[13.6,22.4]	[1.2,2.8]	[1.3,2.8]
Dallas	[10.1,11.9]	[196.3,295.6]	[10.1,14.8]	[1.2,3]	[0.3,1.4]	[59.8,65.6]	[5.3,6.1]	[0.312,0.357]	[0.4838,0.5406]	[17.2,28.4]	[1.5,3.3]	[1.6,3.4]
Denver	[10,12.1]	[158.8,334.4]	[8.1,17.8]	[1.1,3.2]	[0.5,1.1]	[63.4,71.6]	[4.8,6.2]	[0.276,0.372]	[0.4693,0.5245]	[18.1,34.9]	[1.7,4.1]	[1.9,4.3]
Detroit	[9.5,12.1]	[198,307.9]	[10.5,17]	[1.2,6]	[0.4,1.1]	[60.6,72.5]	[4.6,6.1]	[0.271,0.334]	[0.4824,0.5379]	[16.4,29.5]	[1.6,3.2]	[1.8,3.6]
GreenBay	[9,12.8]	[197.9,303.8]	[10.9,14.5]	[1.5,3.1]	[0.6,1.3]	[62.4,66.8]	[4.9,6.5]	[0.301,0.356]	[0.4769,0.5464]	[20,34.1]	[2.4,3.8]	[2.5,4.2]
Houston	[9.2,12.2]	[196.7,290.9]	[10.6,14.4]	[0.9,1.8]	[0.4,1.2]	[64.2,69.9]	[4.7,6]	[0.28,0.35]	[0.4836,0.5505]	[17.2,25.7]	[1.5,2.8]	[1.6,2.9]
Indianapolis	[9.9,12]	[180.8,297.5]	[9.2,15.5]	[0.8,2.4]	[0.4,0.9]	[59.4,70.4]	[4.6,5.9]	[0.269,0.348]	[0.437,0.527]	[15.2,27.1]	[1.4,3.1]	[1.6,3.2]
Jacksonville	[9.1,11.6]	[136.2,256.8]	[7.6,13.4]	[0.8,2.2]	[0.3,1.2]	[61.8,66.8]	[4.1,5.5]	[0.25,0.321]	[0.4532,0.5285]	[15.2,25.9]	[1.3,2.5]	[1.6,2.9]
KansasCity	[9.9,12.8]	[169.6,303.2]	[8.7,14.7]	[0.5,2.9]	[0.3,1.2]	[60.1,65.2]	[4.8,6.7]	[0.251,0.384]	[0.4615,0.5248]	[13.2,34.8]	[1.1,4.1]	[1.1,4.4]
LACHargers	[9.7,12.7]	[205.9,286.9]	[11.5,14.8]	[1.6,2.1]	[0.2,1.1]	[59.7,68.8]	[4.8,6.1]	[0.297,0.348]	[0.498,0.5508]	[20,27.6]	[1.9,3]	[2.2,3.2]
LARams	[8.6,12.1]	[167.9,273.3]	[7.9,14.3]	[0.6,1.7]	[0.2,1.4]	[57.5,66.5]	[4.4,6.2]	[0.255,0.367]	[0.4585,0.5149]	[10.9,30.8]	[1.3,2]	[1.1,3.4]
Miami	[9.5,11]	[181.2,238.2]	[9.6,14.1]	[0.8,1.7]	[0.2,1.4]	[54.9,68]	[4.9,5.8]	[0.278,0.347]	[0.4562,0.5322]	[17.1,24.2]	[1.6,2.5]	[1.6,2.7]
Minnesota	[9.3,11.1]	[171.1,260]	[9.4,13.8]	[0.8,2.2]	[0.6,1.4]	[60.5,66.4]	[5.5,7]	[0.294,0.326]	[0.4587,0.5464]	[17.6,29.6]	[1.8,3.4]	[2.1,3.6]
NewEngland	[10.6,12.5]	[241.5,313.2]	[12.6,16.4]	[1.4,2.5]	[0.9,1.5]	[62.6,74.3]	[5.4,6.3]	[0.329,0.373]	[0.4707,0.5461]	[25.9,33.9]	[2.9,3.6]	[3.1,4.1]
NewOrleans	[9.9,11.8]	[253.7,346.7]	[13.1,17.9]	[1.6,2.9]	[0.6,1.4]	[62.4,71]	[5.6,6.7]	[0.33,0.372]	[0.4803,0.5446]	[24.7,34.7]	[2.7,4]	[2.8,4.2]
NYGiants	[9.3,12.6]	[217.4,293.9]	[11,13.8]	[1.1,2.2]	[0.3,1.1]	[60.5,67.9]	[4.9,5.9]	[0.262,0.338]	[0.4691,0.5422]	[15.4,26.8]	[1.6,3]	[1.8,3]
NYJets	[9.9,11.7]	[155.6,253.6]	[8.3,13.1]	[0.8,2.1]	[0.6,1.3]	[60.7,67.1]	[4.6,5.5]	[0.266,0.306]	[0.4862,0.5269]	[17.2,24.2]	[1.6,2.8]	[1.7,2.8]
Oakland	[9,12.6]	[159.8,255.2]	[7.9,12.5]	[0.6,2.1]	[0.2,1.2]	[59,65.7]	[4.5,6]	[0.249,0.323]	[0.4716,0.5185]	[12.3,25.6]	[1.1,2.8]	[1.1,2.9]
Philadelphia	[9.4,12.9]	[224.1,272.2]	[11.5,13.8]	[1.2,4]	[0.6,1.2]	[60.6,70.7]	[5.6,2]	[0.297,0.344]	[0.431,0.5467]	[17.5,29.6]	[1.8,3.2]	[1.8,3.4]
Pittsburgh	[10.7,11.9]	[219.7,313]	[10.8,15.7]	[1.3,2.2]	[0.5,1.1]	[62.6,67.1]	[5.2,6.2]	[0.297,0.357]	[0.4905,0.5429]	[20.5,26.8]	[2.1,3.2]	[2.2,3.4]
SanFrancisco	[9.8,12.2]	[181.9,245.3]	[8.6,12.9]	[0.9,1.6]	[0.4,1.3]	[58.7,66.1]	[4.9,6.2]	[0.252,0.343]	[0.4398,0.5253]	[14.9,26.1]	[1.4,2.8]	[1.5,3]
Seattle	[9.4,12.1]	[194.1,252.3]	[9.5,12.5]	[0.9,2.1]	[0.2,1.2]	[60.1,65.3]	[4.8,5.9]	[0.264,0.326]	[0.4551,0.524]	[17.5,26.6]	[1.7,3.1]	[1.9,3.2]
TampaBay	[9.7,12.8]	[176.2,320.3]	[9.2,16.6]	[1.1,2.2]	[0.3,0.8]	[58.6,66.6]	[4.5,6.3]	[0.251,0.368]	[0.4544,0.5136]	[15.2,24.8]	[1.4,2.9]	[1.8,3.1]
Tennessee	[10.2,11.5]	[185.9,245.2]	[9.6,12.2]	[1.1,8]	[0.4,1.2]	[56.7,64.5]	[5.1,5.7]	[0.272,0.319]	[0.4322,0.5088]	[15.9,23.8]	[1.6,2.8]	[1.8,2.9]
Washington	[9.7,11.7]	[188.8,297.4]	[9.8,14.1]	[1.1,8]	[0.5,1.3]	[60.4,69.2]	[5.6,4]	[0.283,0.342]	[0.4655,0.5249]	[16.6,26.5]	[1.7,2.8]	[1.8,3.2]

As mentioned in Section 2.5.2, it has been shown that the choice of scheme for the inner weight calculation does not significantly affect the estimation of the parameters in the measurement and structural models, see Noonan and Wold (1982). Therefore, the NFL dataset will be analyzed using the centroid scheme, for both the center and the symbolic covariance regression methods.

Since the manifest and latent variables are related in a reflective way, the first step of the analysis is to evaluate the unidimensionality of the constructs. Table 3.3 presents the estimated Cronbach’s α and Dillon-Goldstein’s ρ for all three latent variables for both the center and symbolic covariance regression methods. The latent variable ‘Scoring’ (ξ_3) is the one with highest index of unidimensionality, with estimated Cronbach’s α of 0.9732 and estimated Dillon-Goldstein’s ρ of 0.9656, for both methods. Since all of the estimated values

for both α and ρ are large, there does not seem to exist a problem with unidimensionality within the three constructs - Pass (ξ_1), Offense (ξ_2) and Scoring (ξ_3).

Table 3.3: Unidimensionality of constructs for NFL example

Latent Variable	Center Method		Symbolic Cov. Method	
	α	ρ	α	ρ
ξ_1	0.7993	0.8491	0.7993	0.8491
ξ_2	0.7981	0.8129	0.7981	0.8129
ξ_3	0.9732	0.9656	0.9732	0.9656

Table 3.4: Weights and loadings associated with each manifest variable - NFL example

Manifest Variable	Center Method		Symbolic Cov. Method	
	Weights	Loadings	Weights	Loadings
$x_{1,1}$	0.1331	0.3403	0.1668	0.4530
$x_{1,2}$	0.4423	1.0444	0.4642	0.9960
$x_{1,3}$	0.4604	1.0702	0.4105	0.9050
$x_{2,1}$	0.2436	1.0203	0.2960	0.7780
$x_{2,2}$	0.0786	0.4181	0.0526	0.2894
$x_{2,3}$	0.1761	0.8211	0.1808	0.6819
$x_{2,4}$	0.2062	0.8928	0.2268	0.8017
$x_{2,5}$	0.2606	1.1449	0.2129	0.8202
$x_{2,6}$	0.1327	0.6896	0.1260	0.4418
$x_{3,1}$	0.3451	0.9942	0.3495	0.9285
$x_{3,2}$	0.3320	0.9796	0.3316	0.9236
$x_{3,3}$	0.3297	1.0060	0.3257	0.9982

The following steps of the analysis using partial least squares estimation for interval-valued variables is to estimate the weights associated with each manifest variables using the iterative procedure in Section 3.2, as well as the parameters of the measurement and structural models. The estimated weights and loadings associated with each manifest variable can be seen in Table 3.4. The manifest variable ‘Pass1’ ($x_{1,1}$) is predicted to weigh 0.1331 in the latent variable ‘Pass’ (ξ_1) for the center method, while ‘Pass2’ ($x_{1,2}$) and ‘Pass3’ ($x_{1,3}$) are predicted to weigh 0.4423 and 0.4604 on the same latent variable, respectively. The

symbolic covariance method predicts that the variable ‘Pass2’ ($x_{1,2}$) has a slightly higher weight on the latent variable ‘Pass’ (ξ_1) than ‘Pass3’ ($x_{1,3}$), which is the inverse of the center method, with weights of 0.4642 for ‘Pass2’ ($x_{1,2}$) and 0.4105 for ‘Pass3’ ($x_{1,3}$). For these same manifest variables, their loadings are 0.3403, 1.0444 and 1.0702 for the ‘Pass1’ ($x_{1,1}$), ‘Pass2’ ($x_{1,2}$) and ‘Pass3’ ($x_{1,3}$), for the center method; and 0.453, 0.996 and 0.905 for the symbolic covariance method, for ‘Pass1’ ($x_{1,1}$), ‘Pass2’ ($x_{1,2}$) and ‘Pass3’ ($x_{1,3}$), respectively. Despite the small differences in estimation of the weights and loadings for the center and symbolic covariance methods, both methods agree that passed yards per game and passing first downs per game have a higher weight in the latent variable passing score, than yards per completion. Considering the latent variable offense, which is associated with six manifest variables, the manifest variable rushing touchdowns per game ($x_{2,2}$) has a very small weight and loading, when compared to the other manifest variables (future analysis could consider removing this variable). All three manifest variables associated with the latent variable scoring have very similar weights (and consequently loadings). This result was expected since the variables points per game, offensive touchdowns per game and touchdowns per game are highly correlated, with correlation of 0.92 or more.

Tables 3.5 and 3.6 present the structural model and model validation statistics estimated for the analysis. The center method predicts the slope of the directional relationship between ‘Pass’ (ξ_1) and ‘Offense’ (ξ_2) to be $\hat{\beta}_{1,2} = 0.7657$, with $R^2 = 0.6255$; whilst, for this same method, the equation between ‘Offense’ (ξ_2) and ‘Scoring’ (ξ_3) has a larger coefficient of determination, $R^2 = 0.8904$, and the latent variables’ directional relationship is predicted to be $\hat{\beta}_{2,3} = 1.0914$. Although not shown on the table, the overall Goodness of Fit for the interval-valued PLS-PM for the center method is 0.6834.

The symbolic covariance method produces results very similar to the center method for the first equation in the structural model, the equation involving ‘Pass’ (ξ_1) and ‘Offense’ (ξ_2), but for the second equation, the results differ more. The coefficient of determination

for the equation involving ‘Offense’ (ξ_2) as an independent variable, and ‘Scoring’ (ξ_3) as a dependent variable have $R^2 = 0.6696$ in the symbolic covariance method, and the estimated coefficient is $\hat{\beta}_{2,3} = 0.8183$. The Goodness of Fit for this method is 0.6424, which is slightly lower than the goodness of fit of the center method, but it is important to remember that this method retains more of the variability within the intervals, and consequently adds some more uncertainty to the estimates.

Table 3.5: Structural model estimated for NFL example

Estimator	Center Method	Symbolic Cov. Method
$\hat{\beta}_{1,2}$	0.7657	0.8187
$\hat{\beta}_{2,3}$	1.0914	0.8183

Table 3.6: Model validation estimates for NFL example

	Communality		Redundancy		R^2	
Latent variable	Center Method	Symbolic Cov. Method	Center Method	Symbolic Cov. Method	Center Method	Symbolic Cov. Method
Pass	0.6724	0.6721	-	-	-	-
Offense	0.4442	0.4441	0.2778	0.2977	0.6255	0.6703
Scoring	0.9038	0.9038	0.8047	0.6052	0.8904	0.6696

This concludes Chapter 3 and the introductory example using PLS-PM for interval-valued data. In Chapter 4, a series of simulation studies aiming to understand how the method proposed in Section 3.2 behaves under different scenarios is conducted.

Chapter 4

Simulations for PLS-PM using interval-valued data

In this chapter, we aim to understand the behavior of the partial least squares path modeling for interval-valued data method proposed in Section 3.2. To assess the method, we conduct a series of simulation studies with different sample sizes, width of the intervals, relationships in a path diagram and number of manifest variables associated with each latent variable.

In order to evaluate the method proposed in Section 3.2 using simulation studies, it is necessary to obtain data in a way that represents the situations in which it will be used as realistically as possible; thus, the process to generate data with the PLS-PM structure is described in Section 4.1. Then in the following section, Section 4.2, the results for four different types of simulations are presented, when different structural models are used (Section 4.2.1), different number of manifest variables (Section 4.2.2), different ranges of the interval-valued variables (Section 4.2.3) and the last study (in Section 4.2.4) evaluates the effect of the sample size in the analysis.

4.1 Simulated data

In this section, the steps to generate randomly the data according to the structure of the partial least squares path modeling estimation is described in detail. Since the manifest and latent variables can relate to one another in two different ways, either the formative or the reflective ways, there are also two different approaches that can be used to generate the data for simulation. Given that the reflective way is more commonly used, this study will only consider it; therefore, all data generated will conform to the reflective way.

To include all the necessary relationships in the model, the data generation process consists of introducing the associations in the structural component through linear models, which starts by randomly generating the independent latent variables. This allows the dependent latent variables to be obtained, by introducing the linear relationship between them, as well as the disturbance terms, which are simply their error term. Once all of the latent variables have been produced, the manifest variables are calculated by considering its assumed relationship with the latent variable and the independently generated measurement error term. All of these steps are carried out using interval-valued variables. The interval-valued data will be generated according to the simulation method (Method IV) in Liu (2016), see also Liu and Billard (2019), while the PLS-PM structure will follow similar steps to Hulland et al. (2010).

For illustration purposes, Figure 3.1 shows a path diagram for an example of PLS-PM and Algorithm 1 presents detailed steps to generate the data as described above. Step 1 in Algorithm 1 starts by obtaining $p - s$ independent latent variables, using a multivariate normal distribution for the center and an exponential distribution for the range of the variables. The multivariate distribution will allow the variables to have different means and variances, while the exponential distribution for the ranges will guarantee that there is no negative range. The interval-valued independent latent variables are established by subtracting the

half range to the center for the lower bound, and adding the half range to the center for the upper bound, as $\xi_{ij} = [\xi_{ijL}, \xi_{ijU}] = [\xi_{ij}^{(c)} - 0.5\xi_{ij}^{(r)}, \xi_{ij}^{(c)} + 0.5\xi_{ij}^{(r)}]$ for all $i = 1, \dots, n$, and $j = 1, \dots, p - s$, where subscript L indicates the interval lower bound and U is the interval upper bound.

In the second step, the desired relationship between the latent variables is introduced and is done for each observation ($i = 1, \dots, n$) individually, via the aggregation of a large number of observations, m . For each of the $p - s$ independent latent variables, randomly generate m observations from a uniform distribution with parameters ξ_{ijL} and ξ_{ijU} . Then, for each dependent latent variable, randomly generate m observations from a normal distribution to compose the disturbance term, as $\zeta_{ih}^{(temp)} \sim N(0, \sigma^2)$, and calculate $\xi_{ih}^{(temp)} = \sum_{j \rightarrow h} \xi_{ij}^{(temp)} \beta_{jh} + \zeta_{ih}^{(temp)}$, for $i = 1, \dots, n$, $j = 1, \dots, p - s$, $h = p - s + 1, \dots, p$, where $j \rightarrow h$ means that the j^{th} latent variable acts as an independent variable to the h^{th} variable. Finally, ξ_{ihL} is obtained as the first quantile (Q1) of $\xi_{ih}^{(temp)}$ and ξ_{ihU} is the third quantile (Q3) of $\xi_{ih}^{(temp)}$; this is done to satisfy the assumption that the distribution within the interval is approximately uniformly distributed, see Xu (2010) and Cariou and Billard (2015).

Since the PLS-PM structure allows independent latent variables to act as dependent latent variables for each other, a new set of m observations for $\xi_{ih}^{(temp)}$ needs to be generated as $\xi_{ih}^{(temp)} \sim Unif[\xi_{ihL}, \xi_{ihU}]$. The third and final step starts by obtaining the manifest variables and is done similarly to step 2. For $j = 1, \dots, p$, and $i = 1, \dots, n$, generate m observations following a uniform distribution, such as $\xi_{ij}^{(temp2)} \sim Unif[\xi_{ijL}, \xi_{ijU}]$. Then, for each manifest variable, generate random measurement errors following a normal distribution with mean zero and variance σ^2 , and estimate the variables using the relationship presented in the path diagram, as $\mathbf{x}_{ijk}^{(temp2)} = \lambda_{jk} \xi_{ij}^{(temp2)} + \mathbf{e}_{ijk}^{(temp2)}$, for $i = 1, \dots, n$, $j = 1, \dots, p$, and $k = 1, \dots, k_j$, where k_j is the number of manifest variables associated with the j^{th} latent variable. The interval-valued manifest variables are then obtained by calculating the first (Q1) and third

(Q3) quantiles of $\mathbf{x}_{ijk}^{(temp2)}$, as $x_{ijk} = [x_{ijkL}, x_{ijkU}] = [Q1(\mathbf{x}_{ijk}^{(temp2)}), Q3(\mathbf{x}_{ijk}^{(temp2)})]$.

After every step presented above, the standard deviation of each variable (either latent or manifest variable) is stored and the variables are standardized such that each interval-valued variable has mean 0 and standard deviation 1. Note that when a variable is standardized, it reduces the effect size of the relationship introduced in the model by a factor equal to the inverse of its standard deviation. Thus, Hulland et al. (2010) suggests that the estimated coefficients should be weighted by the standard deviation before comparing the results with the true value from which the model is originated. Thus, the algorithm is,

Algorithm 1 Generating PLS-PM interval-valued data

Step 1: generate $p - s$ independent latent variables

- 1: Generate the center of the n observations as $\boldsymbol{\xi}^{(e)} \sim \mathbf{N}_{(p-s)}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$
- 2: Generate the range of the n observations as $\boldsymbol{\xi}_j^{(r)} \sim Exp(\lambda)$, for all $j = 1, \dots, p - s$
- 3: Calculate $\xi_{ij} = [\xi_{ijL}, \xi_{ijU}] = [\xi_{ij}^{(e)} - 0.5\xi_{ij}^{(r)}, \xi_{ij}^{(e)} + 0.5\xi_{ij}^{(r)}]$ for all $i = 1, \dots, n$
- 4: Store the standard deviation and standardize each $\boldsymbol{\xi}_j$, for $j = 1, \dots, p - s$

Step 2: generate s dependent latent variables

- 5: **for** $i = 1, \dots, n$, **do**
- 6: Generate m observations as $\boldsymbol{\xi}_{ij}^{(temp)} \sim Unif[\xi_{ijL}, \xi_{ijU}]$ for all $j = 1, \dots, p - s$
- 7: **for** $h = p - s + 1, \dots, p$, **do**
- 8: Generate m observations for the disturbance term, $\zeta_{ih}^{(temp)} \sim N(0, \sigma^2)$
- 9: Calculate $\boldsymbol{\xi}_{ih}^{(temp)} = \sum_{j \rightarrow h} \boldsymbol{\xi}_{ij}^{(temp)} \beta_{jh} + \zeta_{ih}^{(temp)}$, for $j = 1, \dots, p - s$
- 10: Obtain Q1 and Q3 of $\boldsymbol{\xi}_{ih}^{(temp)}$, $\xi_{ih} = [\xi_{ihL}, \xi_{ihU}] = [Q1(\boldsymbol{\xi}_{ih}^{(temp)}), Q3(\boldsymbol{\xi}_{ih}^{(temp)})]$
- 11: Generate m observations as $\boldsymbol{\xi}_{ih}^{(temp)} \sim Unif[\xi_{ihL}, \xi_{ihU}]$
- 12: Store the standard deviation and standardize each $\boldsymbol{\xi}_h$, for $h = p - s + 1, \dots, p$

Step 3: generate manifest variables

- 13: **for** $i = 1, \dots, n$, **do**

- 14: **for** $j = 1, \dots, p$, **do**
- 15: Generate m observations as $\boldsymbol{\xi}_{ij}^{(temp2)} \sim Unif[\xi_{ijL}, \xi_{ijU}]$
- 16: **for** $k = 1, \dots, k_j$, **do**
- 17: Generate m observations for the measurement error, $\mathbf{e}_{ijk}^{(temp2)} \sim N(0, \sigma^2)$
- 18: Calculate $\mathbf{x}_{ijk}^{(temp2)} = \lambda_{jk} \boldsymbol{\xi}_{ij}^{(temp2)} + \mathbf{e}_{ijk}^{(temp2)}$
- 19: Obtain Q1 and Q3 of $\mathbf{x}_{ijk}^{(temp2)}$, $x_{ijk} = [Q1(\mathbf{x}_{ijk}^{(temp2)}), Q3(\mathbf{x}_{ijk}^{(temp2)})]$
- 20: Store the standard deviation and standardize each \mathbf{x}_{jk} , for $j = 1, \dots, p$ and $k = 1, \dots, k_j$.
-

The simulation studies presented in Sections 4.2 apply the data generation process using the steps presented in Algorithm 1 followed by the partial least squares path modeling estimation using the method described in Section 3.2. This is repeated for a large number of times, say $s = 1, \dots, 1000$, and at each repetition the estimates are stored and compared to the true values that were used when generating the data. In order to assess the fit of the model, the average and standard error of the estimation will be calculated, as well as the average absolute relative bias and its standard error. The absolute relative bias for a parameter θ simulation is obtained as

$$RB_s = \frac{|\hat{\theta}_s - \theta|}{\theta} \quad (4.1)$$

where θ is the true value of the parameter and $\hat{\theta}_s$ is the parameter estimate of the s^{th} repetition. Along with the average absolute relative bias for the estimators, the root mean squared error of the latent variables is calculated as, for $j = 1, \dots, p$ and $i = 1, \dots, n$,

$$RMSE_L = \left[\frac{\sum_{i=1}^n (\xi_{ijL} - \hat{\xi}_{ijL})^2}{n} \right]^{\frac{1}{2}} \quad (4.2)$$

$$RMSE_U = \left[\frac{\sum_{i=1}^n (\xi_{ijU} - \hat{\xi}_{ijU})^2}{n} \right]^{\frac{1}{2}} \quad (4.3)$$

where $\xi_{ij} = [\xi_{ijL}, \xi_{ijU}]$ is the latent variable generated in Step 2 of Algorithm 1, $\hat{\xi}_{ij} = [\hat{\xi}_{ijL}, \hat{\xi}_{ijU}]$ is the estimated latent variable from the PLS-PM estimation.

4.2 Simulation Studies

In this section, the data generation and PLS-PM processes described in Section 4.1 are used to conduct a series of simulation studies. The first set of simulations, presented in Section 4.2.1, are intended to evaluate the behavior of the estimation of the parameters in the measurement and structural models under different relationships, including the presence of collinearity. The second set of simulations, in Section 4.2.2, evaluates the behavior of the estimation under different number of manifest variables associated with the latent variables, as well as using different values for the loadings when generating data. The third set of simulations, in Section 4.2.3, intends to assess the affect of the range of the intervals in the estimation of both the measurement and structural model parameters. Finally, the fourth and last set of simulations, in Section 4.2.4, examines the effect of different sample sizes in the estimation.

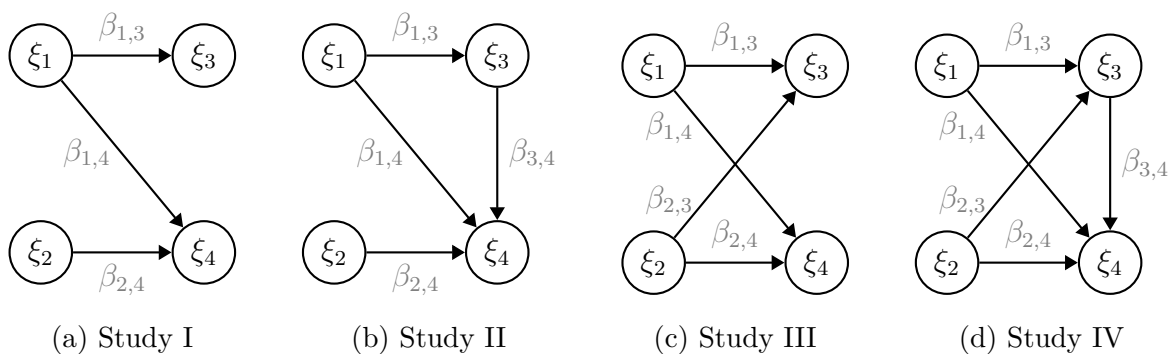


Figure 4.1: Structural model of simulation studies I-IV

The simulation studies conducted in this chapter will follow one of the ten possible structural models defined in Figures 4.1, 4.2 and 4.3. Studies I - X were chosen to represent some of the possible relationships that might affect the estimation of both the loadings and

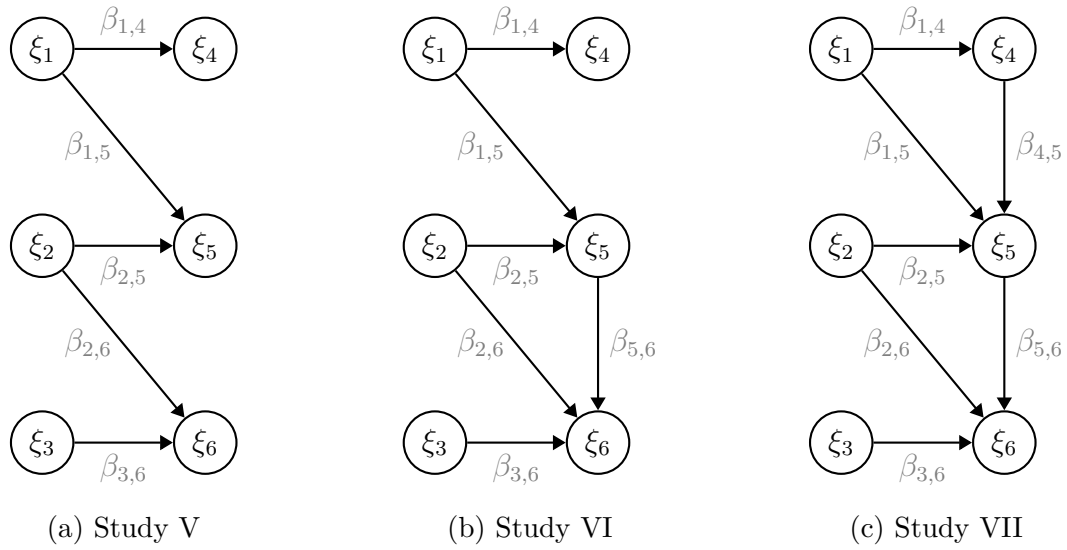


Figure 4.2: Structural model of simulation Studies V-VII

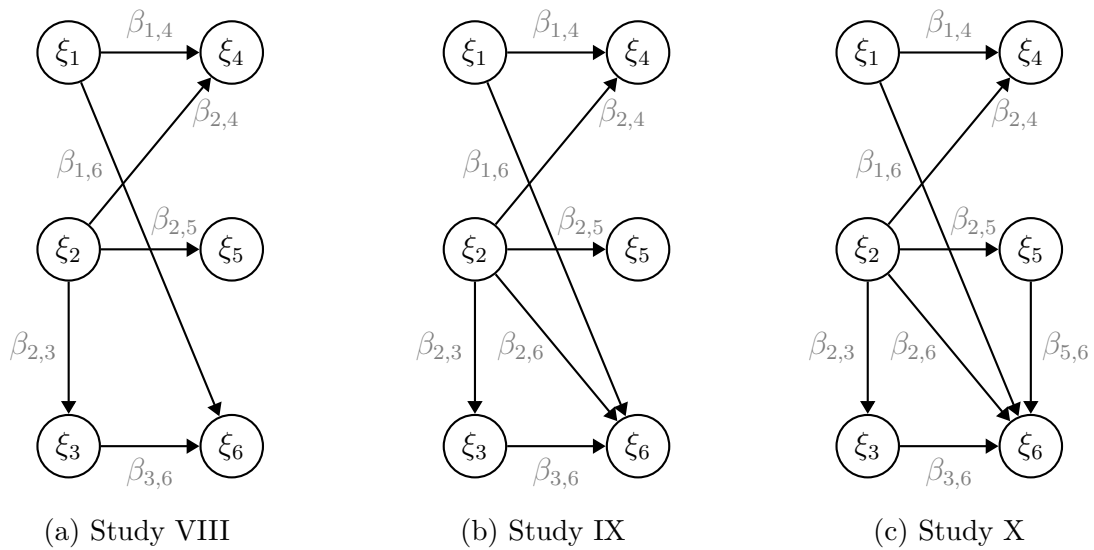


Figure 4.3: Structural model of simulation Studies VIII-X

the parameters associated with the structural model. Figures 4.1, 4.2 and 4.3 do not show the complete path diagram for the models, as depending on the simulation section, the number of manifest variables may vary; however, they will always be related to the latent variable in a reflective way.

The main relationships of each of the ten studies used in Chapter 3 and drawn in Figures 4.1, 4.2 and 4.3 are described below:

1. Study I portrays a simple example of the directional relationship between four latent variables, the latent variable ξ_1 acts as an independent variable for both ξ_3 and ξ_4 , and ξ_2 only acts as an independent variable for ξ_4 .
2. Study II has the same latent variables and relationships as Study I, but with an extra association between ξ_3 and ξ_4 . This extra association introduces some collinearity in the study, since the information contained in ξ_1 is brought to ξ_4 via ξ_1 and ξ_3 .
3. Study III presents the same structure as Study I, but with an extra association between ξ_2 and ξ_3 . There is no added collinearity.
4. Study IV adds collinearity to the structure of Study III, by adding a relationship between ξ_3 and ξ_4 . This extra relationship re-introduces information from both ξ_1 and ξ_2 in the equation where ξ_4 acts as a dependent variable. Note that this structural model contains four latent variables, and five parameters to be estimated.
5. Study V starts the studies with a larger number of latent variables. The variables ξ_1 , ξ_2 and ξ_3 act only as independent variables, and their information is available to the latent variables ξ_4 , ξ_5 and ξ_6 only once, meaning that no collinearity was introduced. Dependent latent variables ξ_5 and ξ_6 are associated with two independent latent variables, while ξ_4 is associated with only one independent latent variable.
6. Additional to the relationships in Study V, Study VI introduces some collinearity by adding an association between ξ_5 and ξ_6 , such that information about ξ_2 is passed to ξ_6 through ξ_2 itself and also through ξ_5 .
7. Study VII has two sources of collinearity, the one in Study VI and a added relationship between ξ_4 and ξ_5 .

8. Study VIII includes six latent variables, where four of them acts as dependent variables at some point of the structural model. It includes a larger number of association between the variables, but no collinearity is included.
9. Study IX has the same structure as Study VIII, but adding some collinearity through the association between ξ_2 and ξ_6 . Note that the dependent variable ξ_6 receives information about ξ_2 via the variable itself and ξ_3 .
10. Study X starts from the structure of Study IX, but include more collinearity regarding the latent variable ξ_2 , adding the association between ξ_5 and ξ_6 .

In all the simulations conducted in this chapter, the process described in Section 4.1 is done for 1000 repetitions. Since the data were generated to conform with the specifications of the model, the simulations in this chapter are done without correcting for sign ambiguity that was described in the weight relations of Section 3.2. The R code for the symbolic functions and simulations is presented in Appendix C.

The tables with results for the simulations in Section 4.2 are presented in Appendix A, for brevity in the text. For each study, the table presents the mean absolute relative bias (see Equation 4.1) for each of the parameters estimated in the structural model, considering the value of the parameter, either fixed or randomly chosen; for the simulations where the parameter was fixed, it was the same across all β 's. For each manifest variable, the mean absolute relative bias (MARB) of the loadings is calculated across the 1000 repetitions, and averaged for each latent variable. The average absolute relative bias for the loadings of manifest variables associated with each latent variables were averaged to present a more concise measure, and produce tables that are easier to read and to obtain information. Along with the mean absolute relative biases, their standard deviation across the repetitions is also presented in the tables. The last information that can be obtained in the tables in Appendix A is the root mean squared error for both the lower and upper bounds of the estimated

latent variables, see Equations 4.2 and 4.3.

4.2.1 Simulation 1

The simulation results presented in this section were intended to evaluate the effect of different relationships in the estimation, as well as the size of the parameter. Hence, all ten studies presented in Figures 4.1, 4.2 and 4.3 will be evaluated when the coefficients for the structural model (β 's) are set to 0.5, 0.7, 0.9 or when the value is randomly chosen out of 0.5, 0.7, 0.9 for each parameter at each simulation, for both the center and symbolic covariance regression methods. This is done to understand the impact of the size of the true parameter in the quality of the estimation. For all these simulations, the number of manifest variables associated with each latent variable was set to eight, and their true loading was randomly selected from the 0.5, 0.7, 0.9. All other variables are held constant in the simulations, to guarantee that they do not have a high impact in the estimations and results. Using the same notation as in Algorithm 1, the mean of all independent variables is $\mu_j = 5$ ($j = 1, \dots, p - s$), and the matrix of variance and covariance for the independent latent variables (Σ) is set to an identity matrix, $\lambda = 1$, $\sigma^2 = 1$, $n = 100$, and the number of aggregations for each observation is $m = 3000$.

Section A.1 includes all tables with results for the simulations described above. For each study there are three tables, one with the estimated mean absolute relative bias (MARB) of the coefficients in the structural model and its standard deviation, one with the average MARB for all manifest variables associated with the latent variables and its standard deviation, and the mean square root of the estimated latent variables, for both the center and symbolic covariance regressions methods. For example, Table A.1 shows that the MARB for the estimated coefficient $\beta_{1,3}$ obtained when setting the value in the data generation process to 0.5 and using the center method is 0.1156, with standard deviation of the absolute relative bias across the repetitions 0.0401, and MARB 0.2005 with standard deviation 0.0168 for the

symbolic covariance method. The MARB drops to 0.0542 when the coefficient is set to 0.7 in the center method, and 0.0804 with the symbolic covariance method.

From Table A.2, it is possible to note that the average across the eight manifest variables associated with ξ_1 yields a MARB of 0.0988 when the coefficients in the structural model are set to 0.5 in the center method, while the average of the standard deviation of these manifest variables is 0.0276. For the symbolic covariance method, the average MARB for the loadings associated with ξ_1 is 0.1048, with an average standard deviation of 0.0751.

Table A.3 presents the lower and upper bounds for the root mean squared error (RMSE) of the interval-valued latent variables for simulations with fixed structural model coefficients and when it is randomly selected. The lower bound RMSE for ξ_1 in the simulations where the coefficients were set to 0.5 is 0.6192 and the upper bound RMSE is 0.6193, for the center method. In addition to the tables presented in Appendix A, some of the most important results are also presented as figures in this section.

Figures 4.4, 4.5, 4.6, 4.7, 4.8, 4.9, 4.10, 4.11, 4.12 and 4.13 show the mean absolute relative bias for each of the estimated coefficient in the structural model for Studies I through X, with error bars indicating where 95% of the absolute relative bias for the 1000 repetition lies, avoiding distributional assumptions to build traditional 95% confidence intervals.

From Figure 4.4, it is possible to note that the average absolute relative bias for all three parameters being estimated in Study I is consistently low when any of the four possible coefficients are being used (0.5, 0.7, 0.9 or when it is randomly chosen), with 95% confidence bounds ranging from 0 to 0.2. The MARB is slightly higher for $\beta_{1,3}$ when the 0.5 is used as the coefficient of choice; and also when it is randomly chosen, it has a much larger upper bound in the 95% confidence interval. By comparing $\beta_{1,3}$ to $\beta_{1,4}$ and $\beta_{2,4}$, the structure in the structural model differs only in the fact that the latent variable ξ_3 is explained only by one independent latent variable, while ξ_4 is explained by two independent latent variables. In general, the center method seems to produce estimates with less variability than the symbolic

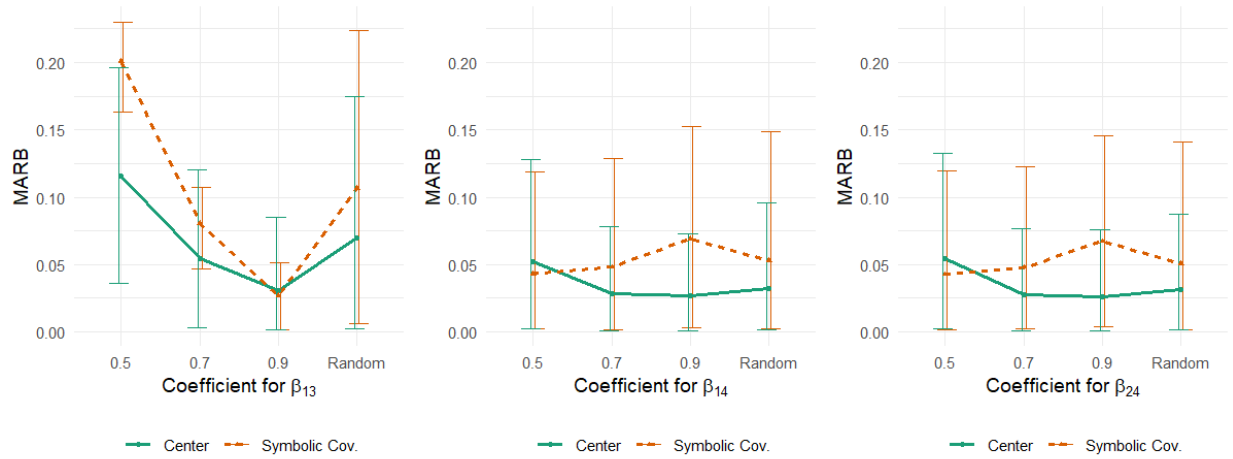


Figure 4.4: Mean absolute relative bias of coefficients in structural model for Study I

covariance method for regression.

In Study II (see Figure 4.5), when some collinearity is added to the model, the MARB is extremely large in the symbolic covariance method for the parameters that are affected by this collinearity, $\beta_{1,4}$ and $\beta_{3,4}$. The same does not happen in the center method, that produces estimates much closer to the real value. For example, from Table A.4, the MARB for the parameter $\beta_{3,4}$ using the center method is 0.4333, while it is 1.4944 for the symbolic covariance method when the parameters are fixed at 0.5. This suggests that the symbolic covariance method is much more affected by collinearity than the center method, as the symbolic covariance method retains the internal variability of the data, and the center method disregards this interval variability by using only the center of the interval-valued variable. Note that the scale of the plot of the estimation of coefficients not directly affected by the collinearity, $\beta_{1,3}$ and $\beta_{2,4}$, is different from the other two coefficients, to allow their pattern to be seen in the figure.

Figure 4.6 presents the MARB for Study III, where no collinearity is added. Study

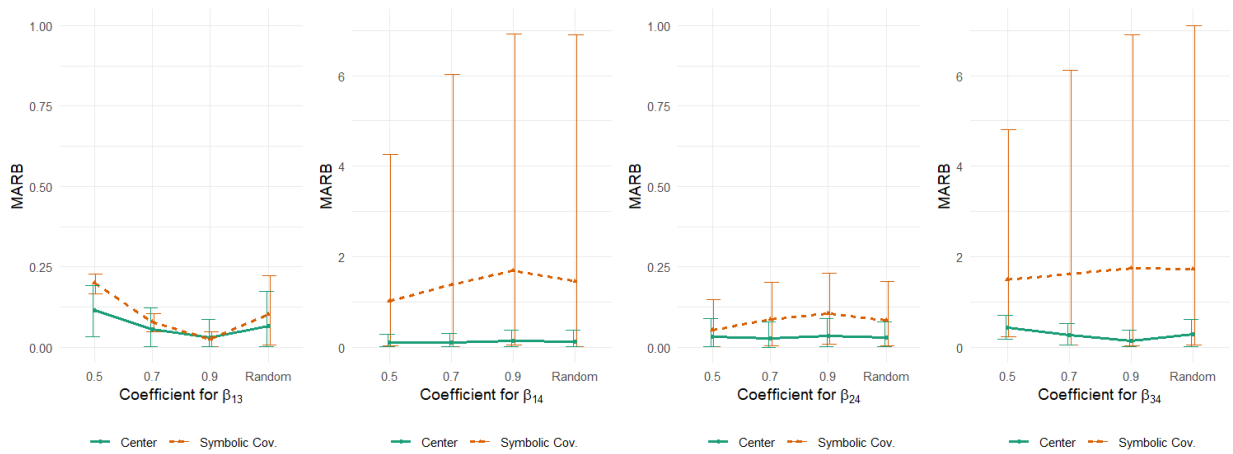


Figure 4.5: Mean absolute relative bias of coefficients in structural model for Study II

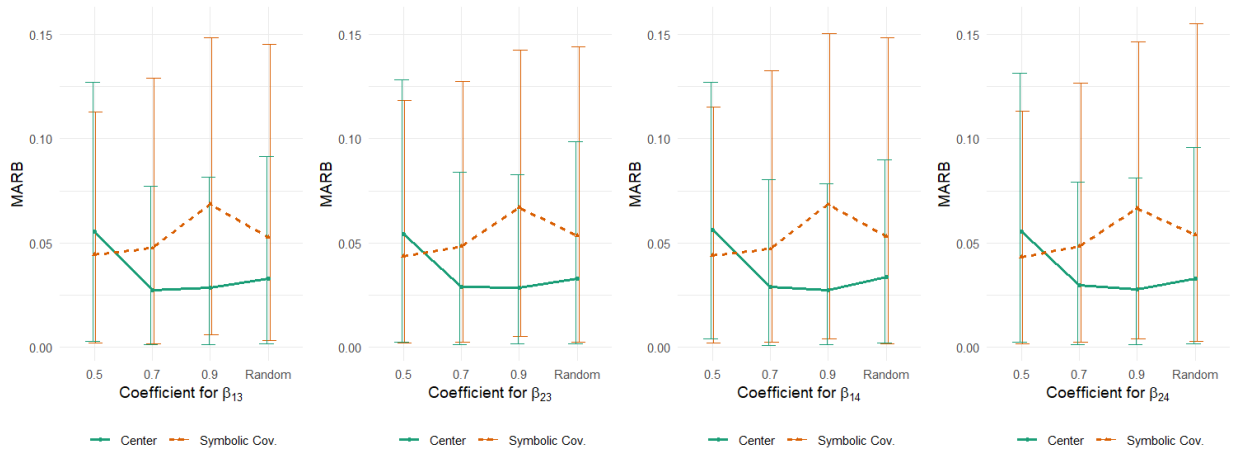


Figure 4.6: Mean absolute relative bias of coefficients in structural model for Study III

III is very similar to Study I, but in this study each of the dependent latent variables has two independent latent variables associated with it. Both methods present similar results and have low values of MARB, ranging 0 to 0.15. The range of the confidence bounds for the symbolic covariance method is larger than the bounds for the center method for all set coefficients of β , except for 0.5; the center method may be more unstable in the presence of smaller effect sizes.

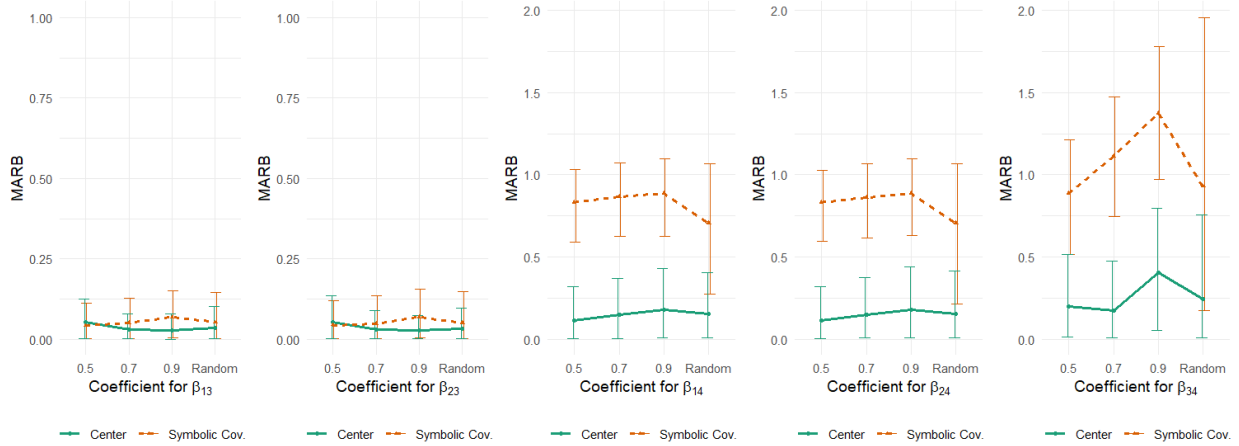


Figure 4.7: Mean absolute relative bias of coefficients in structural model for Study IV

Figure 4.7 summarizes the results for Study IV; the scale of the plots of the estimated MARB of $\beta_{1,4}$ and $\beta_{1,5}$ have been changed for clarity of the results. Remember that Study IV starts from Study III, but adds a source of collinearity with the parameter $\beta_{3,4}$. The latent variable ξ_4 receives information about ξ_1 and ξ_2 from themselves, and also from ξ_3 . From Figure 4.7, it becomes very clear that the all three parameters - $\beta_{1,4}$, $\beta_{2,4}$ and $\beta_{3,4}$ - associated with the latent variable ξ_4 are affected by the collinearity in the model. The symbolic covariance method is more affected than the center method, but the center method also presents higher confidence bounds for estimation these parameters. The estimation of $\beta_{3,4}$ is particularly higher for the symbolic covariance method; furthermore, the range for the coefficients randomly generated includes a broader number of mean absolute relative biases than when the coefficients are set to 0.5, 0.7 and 0.9.

Results for the structural model of Study V are presented in Figure 4.8. Study V includes six latent variables, and does not include collinearity. The relationships quantified by $\beta_{1,5}$, $\beta_{2,5}$, $\beta_{2,6}$ and $\beta_{3,6}$ behave very similarly, but $\beta_{1,4}$ presents larger upper bound for coefficients 0.5 and when it is randomly chosen. This coefficient has the same structure as $\beta_{1,3}$ of Study I,

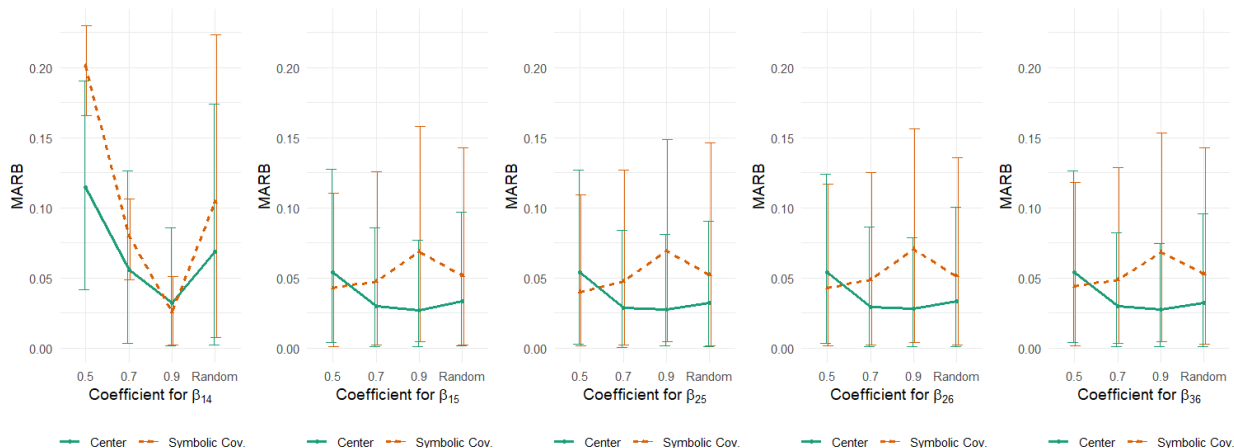


Figure 4.8: Mean absolute relative bias of coefficients in structural model for Study V

where only one independent latent variable is associated with its dependent latent variable. This suggests that when a latent variable is associated with more than one independent latent variable, but there is no collinearity, the estimated parameters might become closer to the true value, even for smaller true values.

Figure 4.9 presents the results for the structural model of Study VI. Study VI adds some collinearity to the structure in Study V, by adding the relationship between ξ_5 and ξ_6 . The MARB of the estimates of the added relationship $\beta_{5,6}$, is larger than for the other parameters, for both the center and symbolic covariance methods, specially when they are set to 0.5 or to 0.9 in the data generation process. However, the magnitude of the MARB in the parameters affected by collinearity in the symbolic covariance method is much smaller in Study VI than it was for Studies II and IV. Similarly to the other studies, the symbolic covariance method has a larger range of 95% confidence bounds than the center for almost all cases evaluated; this probably happens because the extra variability included in the models using the symbolic covariance method causes more instability in the estimation. Note that when another source of collinearity is added to the model, see Figure 4.10 for Study VII, the

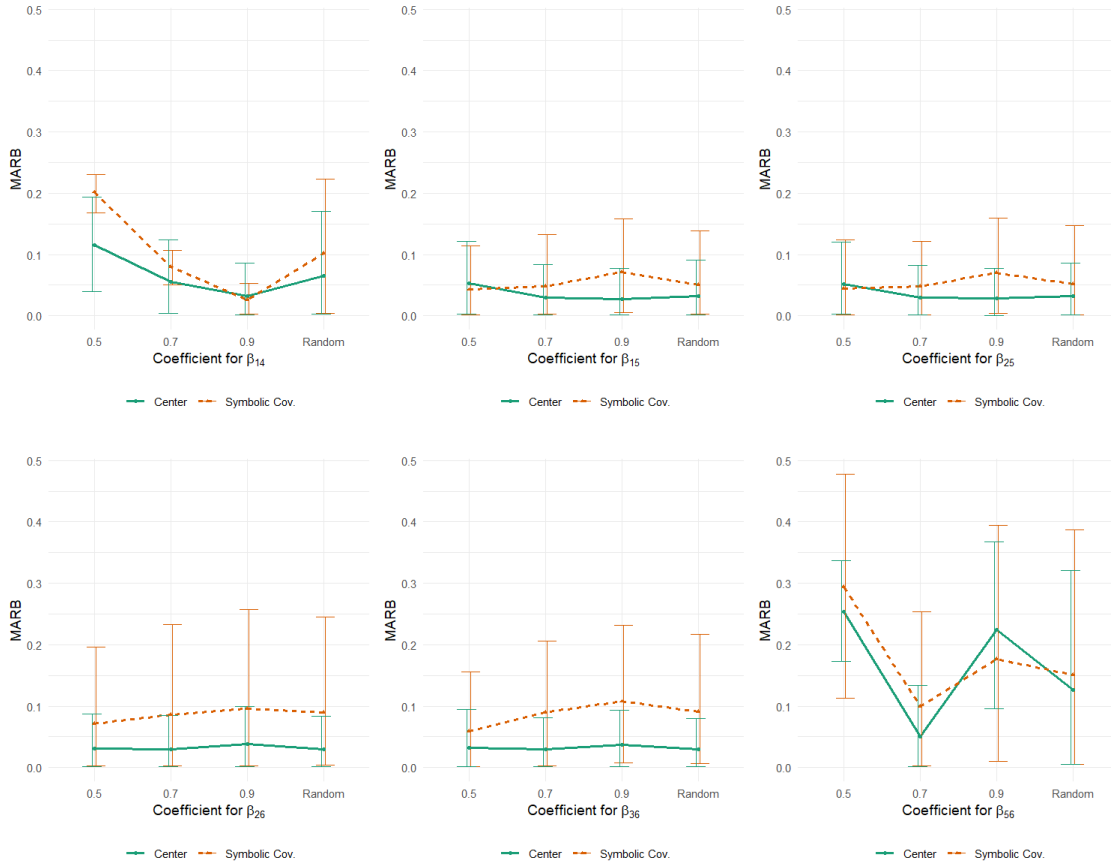


Figure 4.9: Mean absolute relative bias of coefficients in structural model for Study VI

symbolic covariance method has absolute relative biases ranging from 0 to 7 for $\beta_{1,5}$ and $\beta_{4,5}$; for this reason, the scale of the plots for these coefficients are different from the others. For Study VII, the estimation of the parameters of $\beta_{4,5}$ and $\beta_{5,6}$ are affected by the collinearity in both methods, while $\beta_{1,5}$ seems to be affected only in the symbolic covariance. Studies V, VI and VII start with the same structure and evolve in terms of complexity of collinearity, this increase of complexity is reflected in an increase in in MARB in the study and also in the average root mean squared error (RMSE). In Tables A.15, A.18 and A.21, it is possible to see that for Study VII the average root mean squared error for the latent variables ξ_5 and

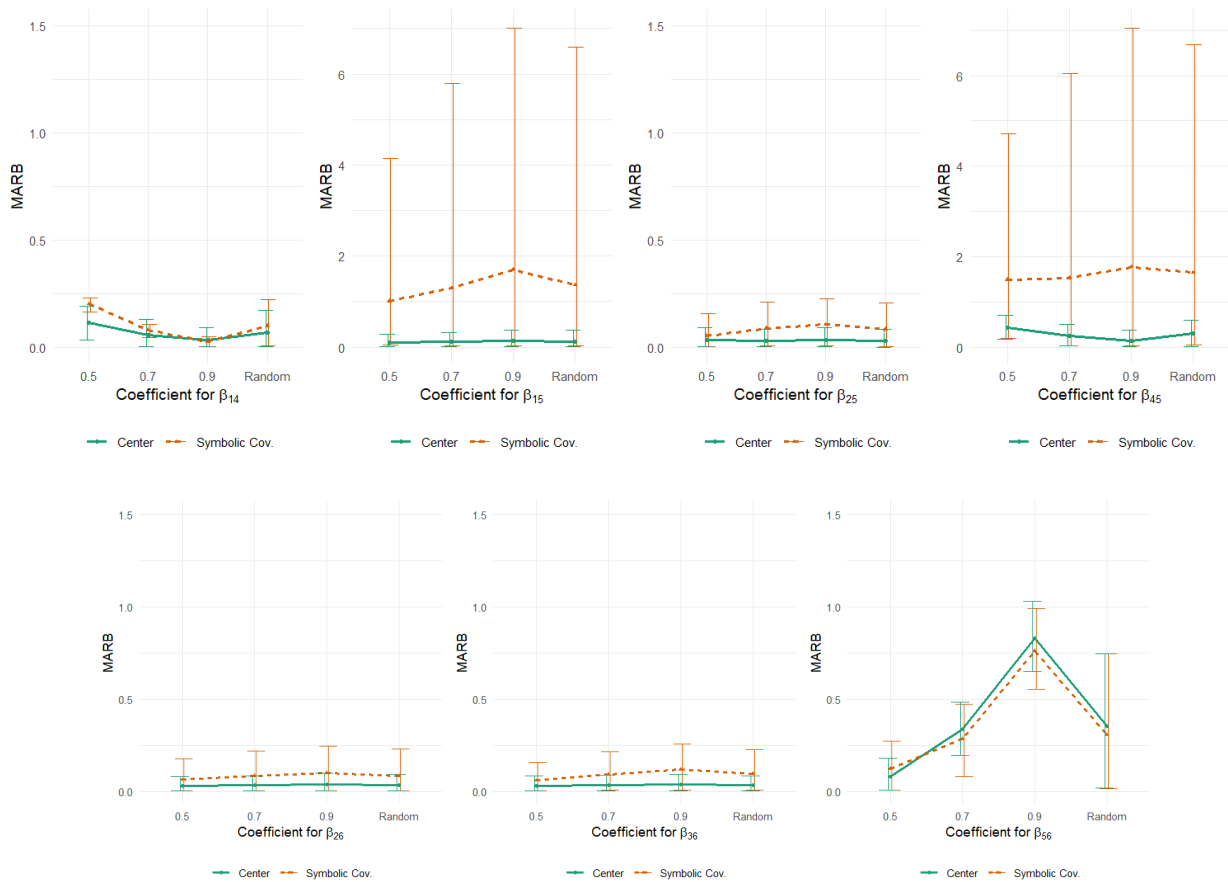


Figure 4.10: Mean absolute relative bias of coefficients in structural model for Study VII

ξ_6 show some increase when compared with the RMSE of Study V, in both methods. In contrast, for Study VI, this increase happens only in the latent variable ξ_6 , but also in both methods.

The structural model estimation of Study VIII shown in Figure 4.11 represents a model with four latent variables acting as dependent variables and two acting as independent latent variables, without an added collinearity. For the first five parameters - $\beta_{2,3}$, $\beta_{1,4}$, $\beta_{2,4}$, $\beta_{2,5}$ and $\beta_{1,6}$ - the MARB acts as expected, based on the studies analyzed previously. The estimation of $\beta_{3,6}$ is somehow unexpected, since these variables do not include a direct addition of

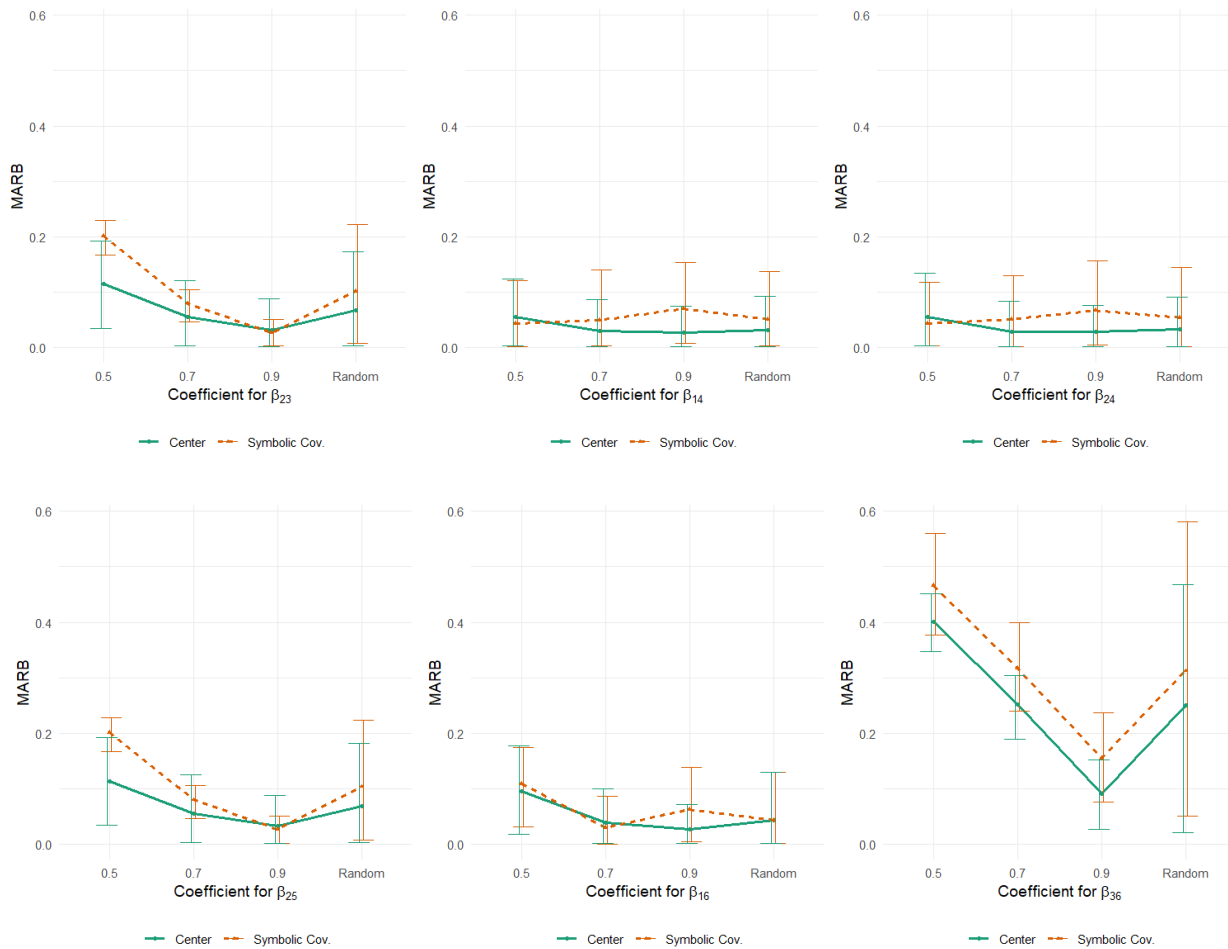


Figure 4.11: Mean absolute relative bias of coefficients in structural model for Study VIII

collinearity, but it is based on a latent variable that acts as a dependent variable in a different equation; so this could possibly be the reason why it behaves differently, as ξ_3 is not a true independent variable, though it is estimated with ξ_2 as its independent variable.

Study IX then adds a relationship between ξ_2 and ξ_6 to the initial structure of Study VIII, see Figure 4.12. As expected, for the symbolic covariance method the estimation of $\beta_{2,6}$ and $\beta_{3,6}$ performs poorly, since they are highly affected by the collinearity (note that the scale of the plot for these coefficients is different from the others). For the center method,

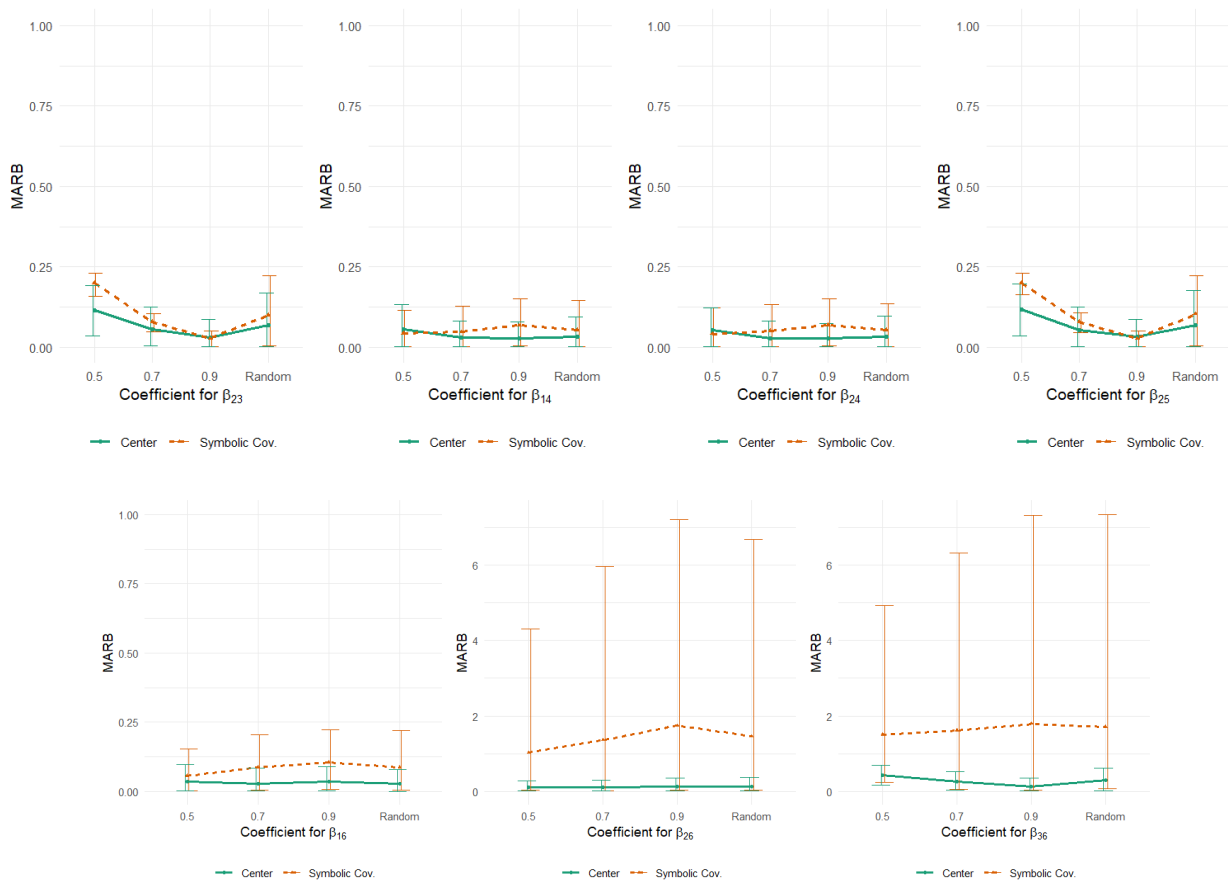


Figure 4.12: Mean absolute relative bias of coefficients in structural model for Study IX

even though it is not clear in the figure due to its scale, the MARB of the estimation of $\beta_{2,6}$ and $\beta_{3,6}$ are also higher than the other five parameters in the model; this can be seen in the results of Table A.25.

Study X is the most complex in terms of collinearity in this research. It adds to the structure of Study IX a term from ξ_5 to ξ_6 , meaning that information regarding ξ_2 is being passed to ξ_6 via $\beta_{2,6}$, $\beta_{3,6}$ and $\beta_{5,6}$. These are exactly the parameters that are being highly affected in the symbolic covariance method, with 95% confidence bounds of absolute relative biases reaching values of almost 10. Special attention should be paid in the scale of the plots

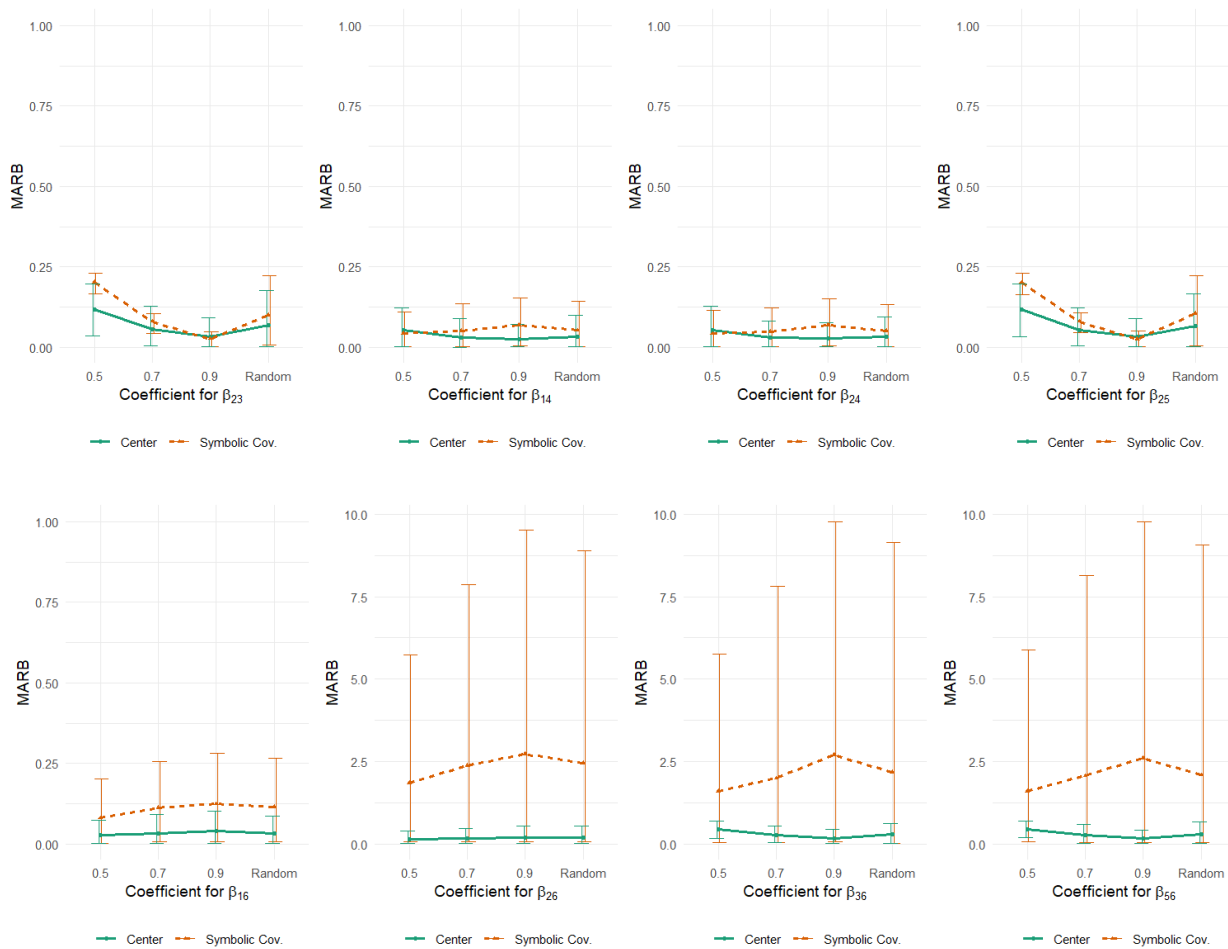


Figure 4.13: Mean absolute relative bias of coefficients in structural model for Study X

of Figure 4.13, as the scale of the last three parameters is from 0 to 10, and the plots of the parameters not directly affected by collinearity is from 0 to 1. Again, due to the scale of the plot, it is not very clear to see the behavior of the center method for these three parameters. Comparing the results from Table A.28, it is possible to notice that the estimation of $\beta_{3,6}$ and $\beta_{5,6}$ has a larger MARB than the other parameters, and also for $\beta_{2,6}$; however, this increase is very small when compared to the symbolic covariance method.

Plots with the mean absolute relative bias of the average loadings for each latent variable

for the ten studies are presented in Appendix B, Figure B.1 through Figure B.10. From these plots, it is possible to notice that, for the latent variables that act only as independent variables, the average MARB for the loadings lies around 0.1 for all studies, and remains steady for the simulations with all values for the coefficient, with larger 95% confidence bounds for the symbolic covariance method. For the latent variables that act as dependent variables in at least one in the structural model, the relationship is not as steady across the simulations with different coefficients, with lower MARB when the coefficient used for the structural model was 0.5, and larger when 0.9 is used. For all of the studies evaluated, the the MARB for the average loadings never approaches extremely large values, indicating that they are not affected (or minimally affected) by the presence of collinearity in the structural equation modeling.

Given all the results presented in this section, a few patterns seem to hold constant. The first behavior to be pointed out is the higher susceptibility of the symbolic covariance method to the presence of collinearity. Even though the center method is also affected by collinearity in the proposed model, it happens in a much smaller magnitude. The symbolic covariance method also seems to produce much more variable estimates, which results in wider confidence bounds. It seems beneficial to the estimation of the parameters if a dependent latent variable includes more than one independent latent variable, as long as it does not add collinearity to the model. By comparing the results for the four different values of coefficients used in these studies, either randomly generated from 0.5, 0.7 and 0.9, or when they are fixed at 0.5, 0.7 or 0.9, the estimation of the parameters in the structural model seems to benefit from using larger values to generate the data, and the opposite happens in the estimation of the loadings. The results for the simulations with randomly generated coefficients embrace the variability and general aspects of the other three values and are also more likely to represent what happens with real data. Therefore, these will be used in the simulations of Sections 4.2.2, 4.2.3 and 4.2.4.

4.2.2 Simulation 2

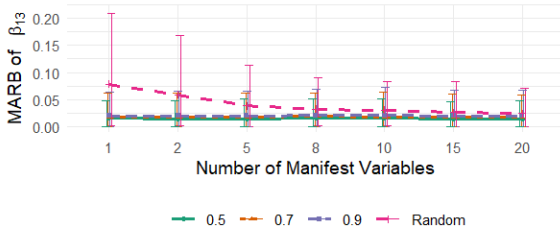
The set of simulations presented in this section intends to evaluate the behavior of the estimation under different number of manifest variables associated with each latent variable. Studies III, IV, V and VI (described in Section 4.2, above) will be evaluated when each latent variable is associated with 1, 2, 5, 8, 10, 15 or 20 manifest variables, and also when the loadings used to generate the data are 0.5, 0.7, 0.9 or when it is randomly generated from 0.5, 0.7 and 0.9. For the simulations when the loadings are randomly generated, this randomness is added at each manifest variable and at each repetition. Studies III, IV, V and VI were chosen to be carried out in this section because they exemplify distinct scenarios. For example, Study III shows a structural model with four latent variables, without collinearity, whilst Study IV has a relationship that adds collinearity to the model. Study V brings a structure including six latent variables, without collinearity, and Study VI adds collinearity to Study V. For this set of simulations, the coefficients in the structural model used to generate the data are randomly chosen from 0.5, 0.7 and 0.9 for each parameter, at each repetition and similarly to Section 4.2.1, using the same notation as in Algorithm 1, the mean of all independent variables is $\mu_j = 5$ ($j = 1, \dots, p - s$), and the matrix of variance and covariance for the independent latent variables (Σ) is set to an identity matrix, $\lambda = 1$, $\sigma^2 = 1$, $n = 100$ and the number of aggregations for each observation is $m = 3000$.

The tables with the mean absolute relative bias (MARB) of the estimated parameters of the structural and measurement models, as well as the lower and upper root mean squared errors bounds are presented in Appendix A, Section A.2. For example, Table A.31 shows the MARB and standard deviation of the absolute relative bias of each of the estimates for Studies III, IV, V and VI using the center and symbolic covariance methods for the simulations where the loadings were fixed to 0.5, 0.7 or 0.9, or when it was randomly generated from 0.5, 0.7 or 0.9, when there is only one manifest variable associated with each latent variable. Similar tables, Tables A.34, A.37, A.40, A.43, A.46 and A.49, respectively, show

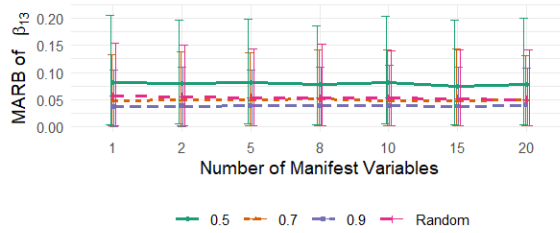
the results for simulations with 2, 5, 8, 10, 15 and 20 manifest variables associated with each latent variable.

Table A.34 shows the MARB of the estimated coefficient $\beta_{1,3}$ in Study III, for the simulation with two loadings associated with each manifest variable, which was set to 0.5, is 0.0147 (with standard deviation 0.0124) when the center method is used, and 0.0802 (with standard deviation 0.0509) when the symbolic covariance method is used; and so on for the other cells in the table. Table A.35 presents the average of the MARB for the two manifest variables associated with each latent variable used in this structural model, as well as the average of the standard deviation of their absolute relative biases. For the simulations in Study III, when using fixed loadings of 0.5, the average of the MARB for the loadings of ξ_1 is 0.2185 (average standard deviation of 0.012) for the center method, and 0.2088 (average standard deviation of 0.0159) for the symbolic covariance method. For this same number (2) of manifest variables, Table A.36 shows that for the center method the lower and upper bounds of the root mean squared error of ξ_1 are both 0.8036, while for the symbolic covariance method the lower bound RMSE is 0.8026 and the upper bound is 0.803. Some of the information from the tables in Section A.2 will be presented as plots in this section, in order to facilitate the visualization of patterns and behaviors of the different scenarios evaluated.

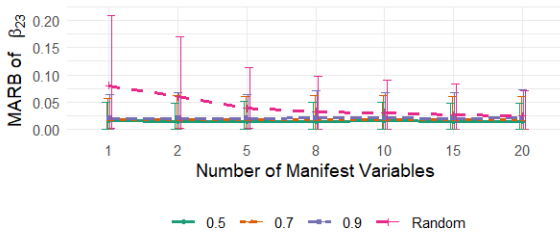
Figure 4.14 shows the MARB for the parameters and 95% confidence bounds in the structural model for Study III when different number of manifest variables is used, and also for different values of the loadings (0.5, 0.7, 0.9 or randomly chosen). Recall that the confidence bounds show where 95% of the absolute relative biases of the estimates of the 1000 repetitions lies. Study III does not include collinearity, and all four parameters are expected to behave similarly since they have the same structure - each dependent latent variable is associated with two different independent latent variables. For the center method, it is possible to notice that when the loadings are fixed, the MARB for all four parameters estimated remains constant and very close to 0; when the loadings are randomly chosen, the



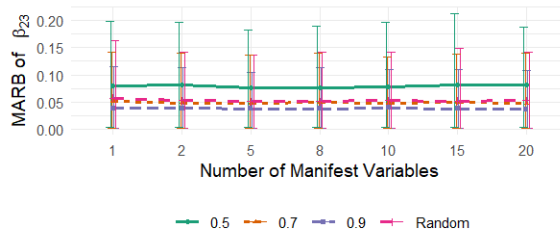
(a) Center Method - $\beta_{1,3}$



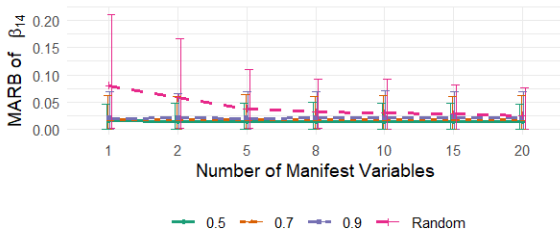
(b) Symbolic Covariance Method - $\beta_{1,3}$



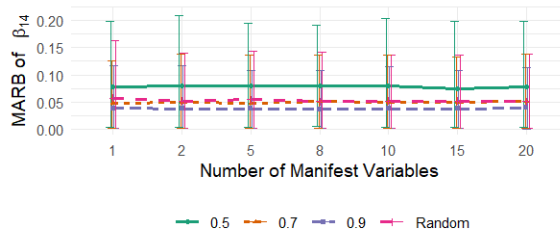
(c) Center Method - $\beta_{2,3}$



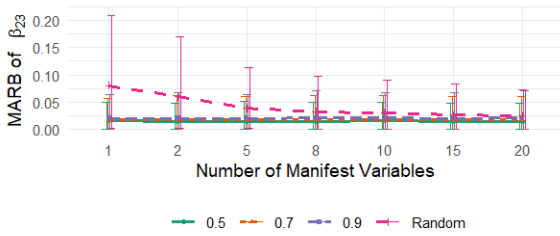
(d) Symbolic Covariance Method - $\beta_{2,3}$



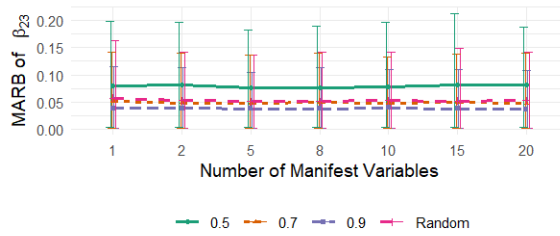
(e) Center Method - $\beta_{1,4}$



(f) Symbolic Covariance Method - $\beta_{1,4}$

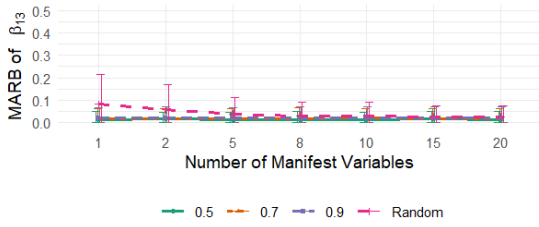


(g) Center Method - $\beta_{2,3}$

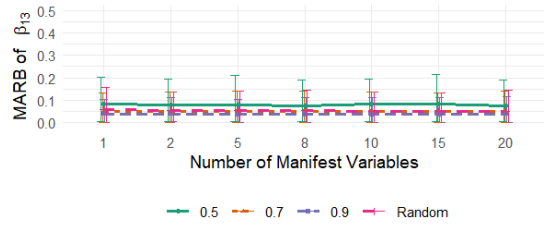


(h) Symbolic Covariance Method - $\beta_{2,3}$

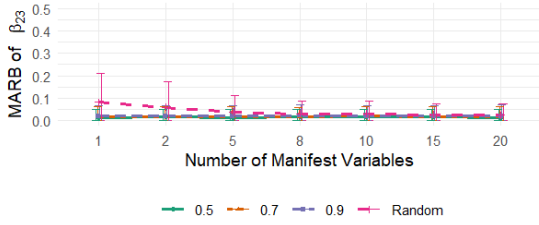
Figure 4.14: Mean absolute relative bias of coefficients in structural model for simulations with different loadings and number of manifest variables in Study III



(a) Center Method - $\beta_{1,3}$



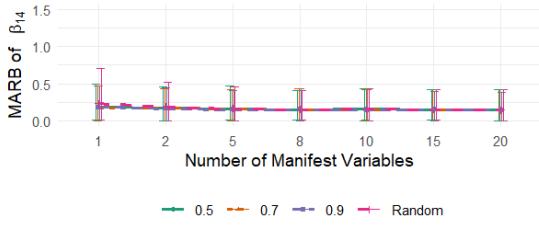
(b) Symbolic Covariance Method - $\beta_{1,3}$



(c) Center Method - $\beta_{2,3}$



(d) Symbolic Covariance Method - $\beta_{2,3}$



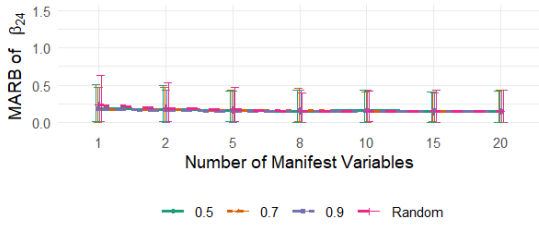
(e) Center Method - $\beta_{1,4}$



(f) Symbolic Covariance Method - $\beta_{1,4}$

Figure 4.15: Mean absolute relative bias of coefficients in structural model for simulations with different loadings and number of manifest variables in Study IV

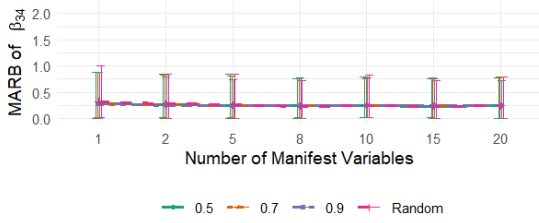
estimated parameters shows a larger variability for smaller numbers of manifest variables, but it reaches similar values to the fixed simulations when at least eight manifest variables are used. The symbolic covariance method, on the other hand, remains constant as the number of manifest variables increase, but when the loadings are set to 0.5, the MARB tends to be slightly bigger than for the other fixed and random loadings.



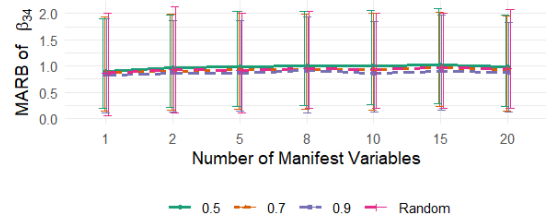
(a) Center Method - $\beta_{2,4}$



(b) Symbolic Covariance Method - $\beta_{2,4}$



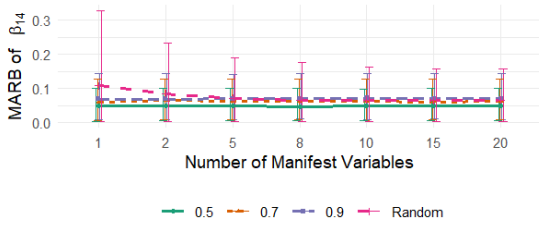
(c) Center Method - $\beta_{3,4}$



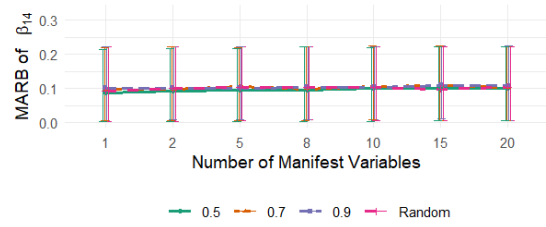
(d) Symbolic Covariance Method - $\beta_{3,4}$

Figure 4.16: Cont. Mean absolute relative bias of coefficients in structural model for simulations with different loadings and number of manifest variables in Study IV

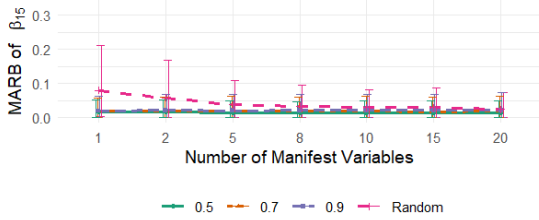
Differently from Study III, Study IV includes a relationship that adds collinearity in the structural model. This collinearity highly increases the magnitude of the MARB for the symbolic covariance method for the parameters associated with the dependent latent variable ξ_4 . From Figures 4.15 and 4.16, it is possible to see that, for the symbolic covariance regression method, the MARB remains mostly constant for the estimated coefficients for the simulations with 1 up to 20 manifest variables. Notice that the scale of Figures 4.15 and 4.16 vary depending on the parameter illustrated. The estimation of the parameters $\beta_{1,3}$ and $\beta_{2,3}$, which are not affected by the collinearity, have trends analogous to the ones in Study III, with larger variability for a smaller number of manifest variables in the center method when the loadings are randomly generated, and constant MARB for the symbolic covariance method with a small increase when the loadings are fixed to 0.5.



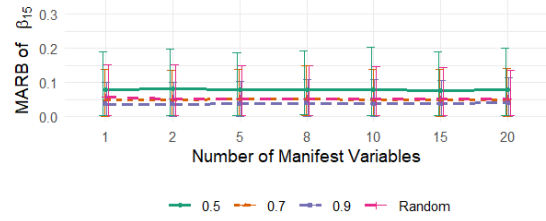
(a) Center Method - $\beta_{1,4}$



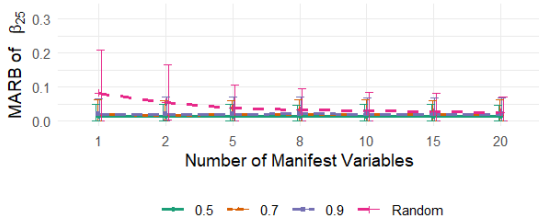
(b) Symbolic Covariance Method - $\beta_{1,4}$



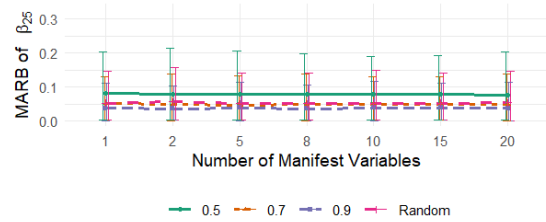
(c) Center Method - $\beta_{1,5}$



(d) Symbolic Covariance Method - $\beta_{1,5}$



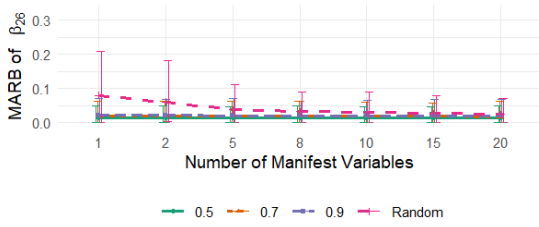
(e) Center Method - $\beta_{2,5}$



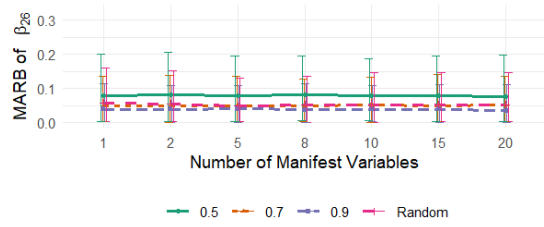
(f) Symbolic Covariance Method - $\beta_{2,5}$

Figure 4.17: Mean absolute relative bias of coefficients in structural model for simulations with different loadings and number of manifest variables in Study V

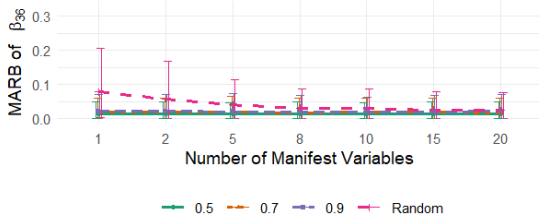
Figures 4.17 and 4.18 show results for the structural model of Study V, where six latent variables are included, but without collinearity. For all five parameters estimated, the same patterns noted for Study III can be observed. When the loadings are randomly chosen, there is a larger variability in the absolute relative biases for a small number of manifest variables in the center method, but it stabilizes and reaches confidence bounds similar to the



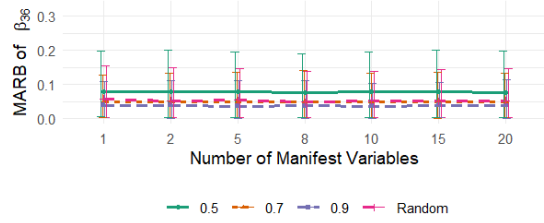
(a) Center Method - $\beta_{2,6}$



(b) Symbolic Covariance Method - $\beta_{2,6}$



(c) Center Method - $\beta_{3,6}$

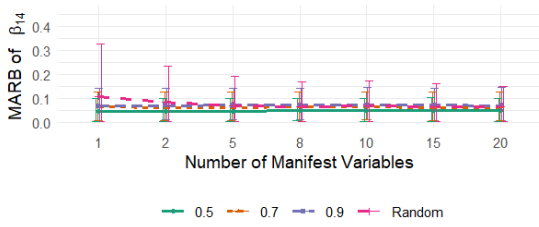


(d) Symbolic Covariance Method - $\beta_{3,6}$

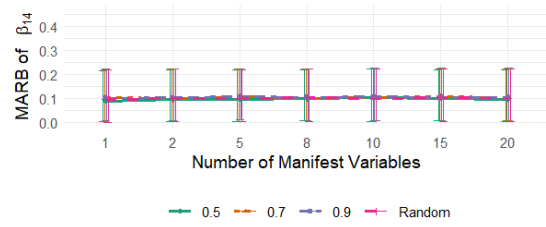
Figure 4.18: Cont. Mean absolute relative bias of coefficients in structural model for simulations with different loadings and number of manifest variables in Study V

other fixed loadings when there is at least 8 to 10 manifest variables. The MARB for the parameters estimated using the symbolic covariance method remains mostly constant with the change of the number of manifest variables associated with each latent variable; with the simulations using fixed loadings of 0.5 producing estimates with slightly bigger absolute relative biases than the other values for the loadings.

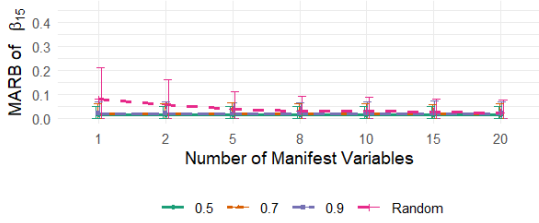
Figures 4.19 and 4.20 present the MARB for the structural model parameters in Study VI and their 95% confidence bounds, for different number of manifest variables and values of the loadings associated with the six latent variables. This study includes a relationship that adds collinearity, but as noted in Section 4.2.1, the magnitude of the MARB for the parameters that include collinearity is much smaller than for Study IV. Considering the changes as the number of manifest variables associated with each latent variables increase,



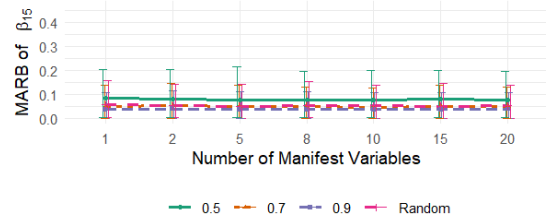
(a) Center Method - $\beta_{1,4}$



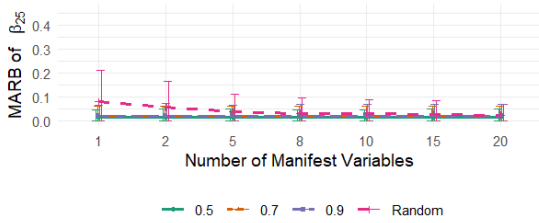
(b) Symbolic Covariance Method - $\beta_{1,4}$



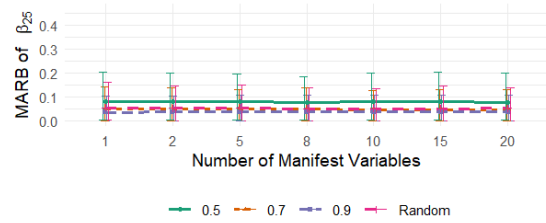
(c) Center Method - $\beta_{1,5}$



(d) Symbolic Covariance Method - $\beta_{1,5}$



(e) Center Method - $\beta_{2,5}$

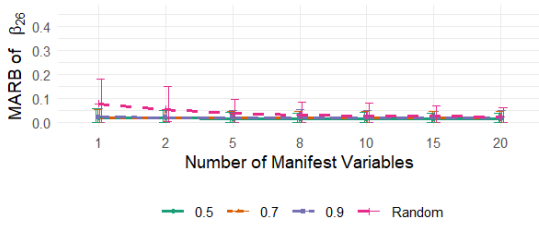


(f) Symbolic Covariance Method - $\beta_{2,5}$

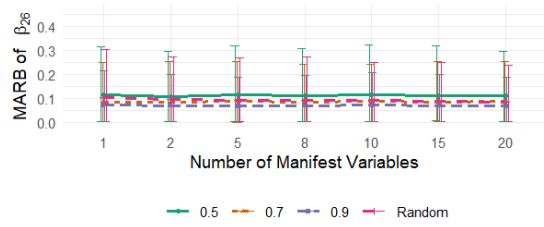
Figure 4.19: Mean absolute relative bias of coefficients in structural model for simulations with different loadings and number of manifest variables in Study VI

the same patterns mentioned above for Studies III, IV and V are observed.

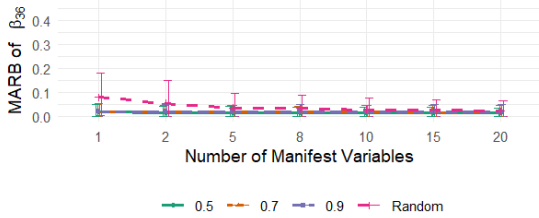
Figures B.11, B.12, B.13, B.14, B.15 and B.16 in Section B.2 of Appendix B show the average MARB for the loadings associated with each manifest variable, for Studies III, IV, V and VI. The same pattern seems to hold for all studies evaluated, as well as for both the center and symbolic covariance methods, i.e., the average MARB of the loadings remains



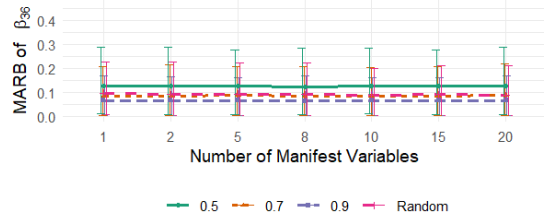
(a) Center Method - $\beta_{2,6}$



(b) Symbolic Covariance Method - $\beta_{2,6}$



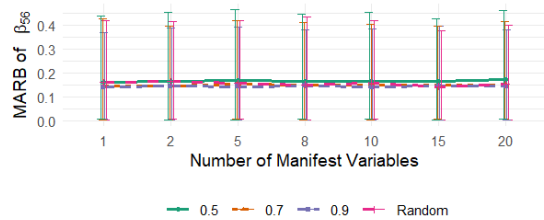
(c) Center Method - $\beta_{3,6}$



(d) Symbolic Covariance Method - $\beta_{3,6}$



(e) Center Method - $\beta_{5,6}$



(f) Symbolic Covariance Method - $\beta_{5,6}$

Figure 4.20: Cont. Mean absolute relative bias of coefficients in structural model for simulations with different loadings and number of manifest variables in Study VI

constant with the change of the number of manifest variables. Including a larger number of manifest variables does not seem to affect the estimation of the loadings, but using a fixed value of 0.5 as the loading to generate the data tends to produce estimates with higher MARB of the loadings. Overall, when the loadings are set to 0.7, the average MARB of the loadings lies around 0.1, while it is 0.2 when the loadings are set to 0.5. For the symbolic covariance

method, the 95% confidence bound for the simulations using randomly generated loadings tends to capture the confidence bounds for all simulations with fixed values of loadings. For the center method, when the loadings are randomly generated, having less manifest variables associated with each latent variables tends to produce more variable estimates.

By comparing the results of RMSE for the lower and upper bounds across different number of manifest variables (see Tables A.33, A.36, A.39, A.42, A.45, A.48 and A.51), it is possible to note that the root mean squared error does not have a significant change with the increase of the number of manifest variables. However, if the lower and upper bounds for the root mean squared errors are compared across the different values for the loadings for each of the number of manifest variables individually, it is possible to see that the lower the loadings, the higher is the root mean squared error, and that the loadings randomly generated follow a pattern similar to the fixed loading of 0.7.

Overall, evaluating all of the results presented in this section suggests that the symbolic covariance method performs similarly when the latent variables are associated with a few or many manifest variables. The center method, in the cases when the loadings were randomly generated, has wider 95% confidence bounds when fewer manifest variables are used, but these bounds approach the results for the other loadings in the estimation of the structural model with as many as eight manifest variables associated with each latent variable. The estimation of the loadings also does not seem to be very susceptible to the number of manifest variables, but to the true value of the loading. Overall it seems that using the randomly generated loadings is the method that approximated more to the reality, but future research should consider whether it is useful to use randomly generated values of 0.7 or more, as this is a minimum standard in some areas (see Chin et al., 2003), and also seems to hold as a good threshold for the interval-valued case.

4.2.3 Simulation 3

The following set of simulations is done in order to understand better the behavior of the estimation of parameters of both the structural and measurement models under the presence of different range parameters for the data generation process. From Algorithm 1, it is possible to note that the range of the interval-valued latent variables can be manipulated by changing the parameter λ used to generate the ranges of the independent latent variable, or by changing the standard deviation of the randomly generated disturbance terms σ , while the range of the manifest variable can be directly manipulated by changing the standard deviation of the error terms, σ , in Step 3 of Algorithm 1. Since the latent variables were generated solely to be used in the process of generating the manifest variables, and therefore are not used in the PLS-PM method, the effect of the range of these variables can not be directly analyzed. Thus, only the parameter σ , associated with the range of the interval-valued manifest variables will be evaluated for Studies III, IV, V and VI. For a fixed value of $\lambda = 1$, the estimation is carried out for $\sigma = 0.1, 0.5, 1, 1.5$ and 2 . Since σ is the standard deviation of the normally distributed error term, larger values of the σ tend to produce wider intervals. For all simulations presented in this section, the true value of the coefficients in the structural and measurement models were randomly generated from $0.5, 0.7$ and 0.9 ; the sample size is $n = 100$, the number of aggregations for each observation is $m = 3000$, the matrix of variance and covariance for the independent latent variables (Σ) is set to an identity matrix, and the mean for all independent latent variables is $\mu_j = 5$, for $j = 1, \dots, p - s$.

The mean absolute relative biases (MARB) for the parameters in the structural and measurement models in Studies III, IV, V and VI, for the simulations using different values of σ are presented in Appendix A in Tables A.52 and A.53, respectively. From Table A.52, in Study III, using the center method with $\sigma = 0.1$, the MARB of the estimate of the parameter $\beta_{1,3}$ in the structural model is 0.0426 , with standard deviation 0.0116 , and slightly higher for

the symbolic covariance method, with MARB 0.0612, and standard deviation 0.0197. From Table A.53, also in Study III using the center method with $\sigma = 0.1$, the MARB of the first latent variable, ξ_1 , is 0.0554 with standard deviation 0.0129, while it is 0.0558 with standard deviation 0.0127 for the symbolic covariance method; similarly for the other values of σ and parameters. Table A.54 shows the lower and upper bound for the root mean squared errors for the studies with different σ 's, using both the center and symbolic covariance methods, calculated using Equations 4.2 and 4.3, respectively.

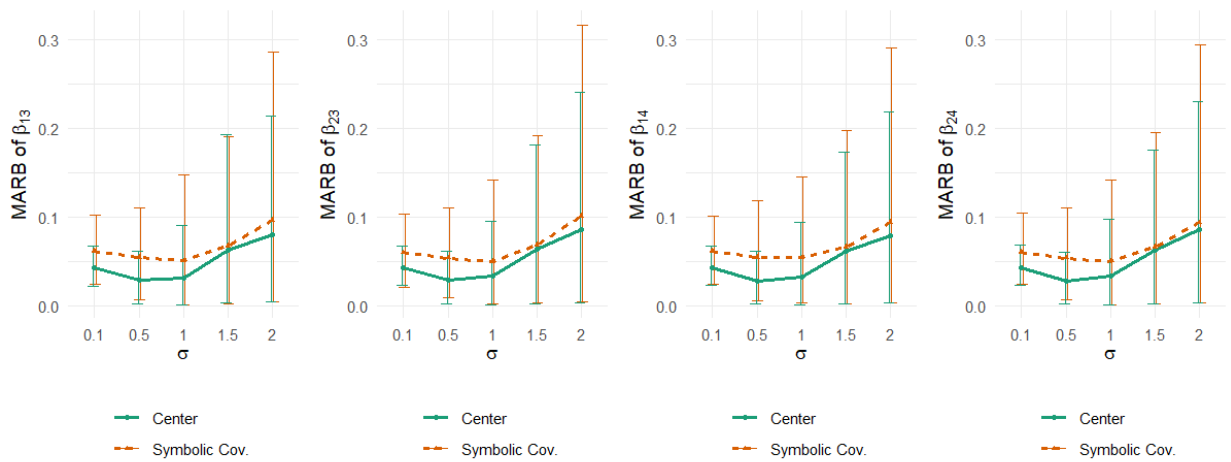


Figure 4.21: Mean absolute relative bias of coefficients in structural model for Study III with different σ 's

Figure 4.21 shows the mean absolute relative bias and its 95% confidence bounds for Study III for simulations with the different values of σ . Since σ is the standard deviation of the error term added to the manifest variables, it is expected to affect the quality of the estimation, since it is adding more variability to the only source of information used in the modeling process. From Figure 4.21, it is possible to see that the variability of the absolute relative biases associated with the parameters in Study III, where there is no collinearity, increases as σ increases for all parameters in the structural model, for both the center and symbolic covariance methods; the MARB itself also show some increase, but given the

confidence bounds there is no clear evidence of significant increase.

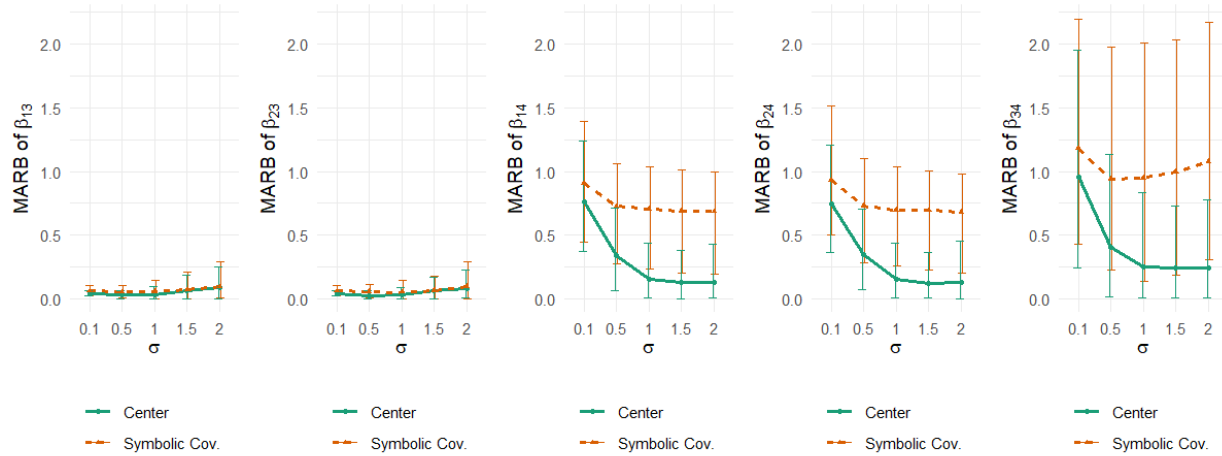


Figure 4.22: Mean absolute relative bias of coefficients in structural model for Study IV with different σ 's

By observing the results for Study IV in Figure 4.22, when collinearity is added to the initial structure in Study III, it is possible to note that due to the instability of estimation under collinearity between the variables, the patterns are not as clear as they were for Study III. For the parameters that are not directly affected by the collinearity as σ increases the bigger are the confidence bounds for the MARB. For the parameters in the structural model that are affected by the collinearity, $\beta_{1,4}$, $\beta_{2,4}$ and $\beta_{3,4}$, the MARB of the estimates tends to remain constant as σ increases, except when $\sigma = 0.1$, when it shows a pattern of increase in the MARB. This seems counter intuitive, as $\sigma = 0.1$ would generate narrower interval-valued variables, but it is possible that a narrower interval-valued variable would be more affected by the collinearity, while a wider interval-valued variable, with a larger error, would disperse some of the collinearity; or, since the confidence bounds for the estimates when $\sigma = 0.1$ and $\sigma = 0.5$ overlap for most of its width, it is possible that there is no actual difference. For the estimate of $\beta_{1,4}$, $\beta_{2,4}$ and $\beta_{3,4}$, the symbolic covariance method tends to have larger MARB, with wider 95% confidence bounds for the absolute relative biases than does the

center method.

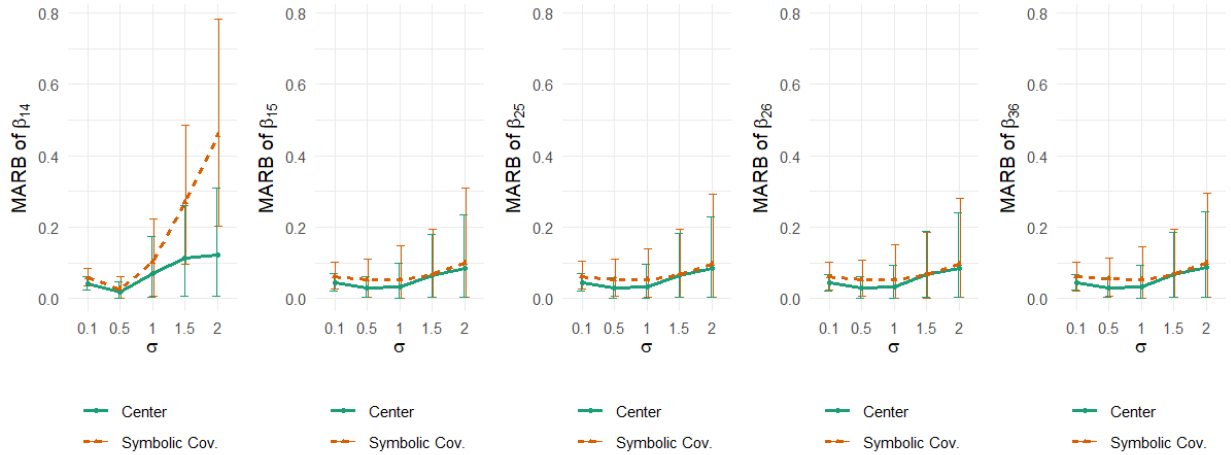


Figure 4.23: Mean absolute relative bias of coefficients in structural model for Study V with different σ 's

Figure 4.23 presents the mean absolute relative bias and the 95% confidence bounds from the 1000 repetitions for Study V, for the different values of σ . Recall that Study V includes six latent variables, without any direct collinearity added to the model. The latent variable ξ_4 acts as a dependent variable, but is associated with only one independent latent variable, while the other dependent latent variables, ξ_5 and ξ_6 , are each associated with two independent latent variables. The mean absolute relative bias of the estimate of $\beta_{1,4}$, the only parameter in the estimation of the latent variable ξ_4 , presents a larger increase in the MARB as σ increases than the estimation of the other four parameters. These results seem to suggest that when a dependent latent variable is explained by less independent latent variables, having wider intervals (with larger error terms) for the dependent variable leads to higher absolute relative biases than when the estimated parameters refer to a dependent latent variable associated with more independent variables.

The last study that we observe the behavior as σ increases is Study VI. Study VI starts from the structure of Study V, but adds a term including collinearity in the model. The

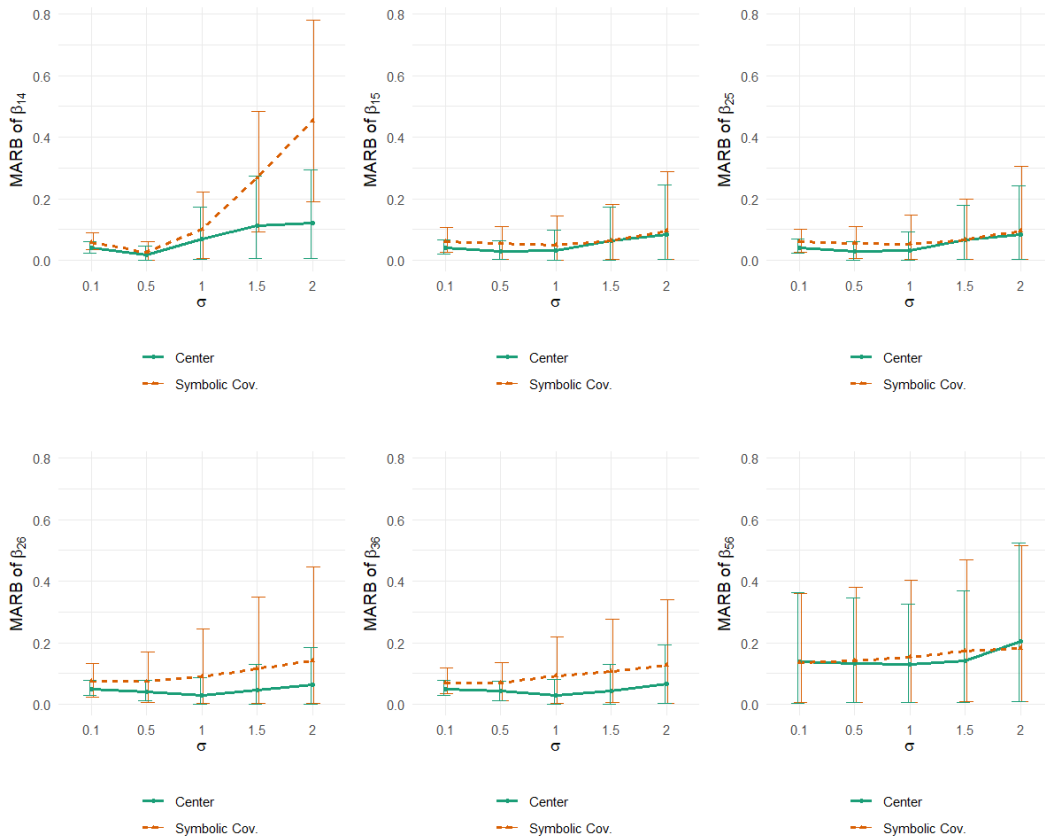


Figure 4.24: Mean absolute relative bias of coefficients in structural model for Study VI with different σ 's

MARB and its 95% confidence bounds for the simulations for changing σ in Study VI are presented in Figure 4.24. Even though this model has collinearity that was not included in Study V, the results found for the simulations that changing σ retains the same patterns as in Study V. By comparing the results as σ increases, in Figure 4.24, the MARB of the estimate of $\beta_{1,5}$, $\beta_{2,5}$, $\beta_{2,6}$ and $\beta_{3,6}$ presents wider confidence bounds. For the parameter estimating the dependent latent variable associated with only one independent latent variable, $\beta_{1,4}$, for the symbolic covariance method, the increase is steeper than for the center method, and also when compared to the other parameters. The estimation of the parameter that introduces

the collinearity, $\beta_{5,6}$, tends to have wider confidence bounds than for the other parameters estimated, particularly when σ is small.

Figures B.17, B.18, B.19 and B.20, in Appendix B, Section B.3, show the average of the MARB of the loadings for the eight manifest variables associated with each latent variable in Studies III, IV, V and VI, respectively. The pattern observed with the average loadings is the same that we observed in the estimation of the parameters in the structural model described above; larger values of σ , which produce wider interval-valued variables, result in a larger mean absolute relative bias in the estimation of the parameters of the measurement model. Similarly, from Table A.54, it is possible to see that for both the center and symbolic covariance methods as the standard deviation of the error term increases, the lower and upper bounds for the root mean squared error for the interval-valued latent variables also increase.

4.2.4 Simulation 4

The last set of simulations evaluated is intended to understand the patterns or behaviors of the estimates as the sample size increase. In general, it is expected that the method will be more stable and obtain better estimates with larger sample sizes. In order to observe if the PLS-PM also shares this behavior, Studies III, IV, V and VI were evaluated with samples of size $n = 10, 20, 30, 50, 75, 100, 150$ and 200 , for true values of the parameters in the structural and measurement models randomly generated from $0.5, 0.7$ and 0.9 . The values of the other parameters, using the same notation as in Algorithm 1, were fixed at $\sigma = 1$, $\lambda = 1$, $m = 3000$, the number of manifest variables associated with each latent variable is eight, the mean of all independent variables is $\mu_j = 5$ ($j = 1, \dots, p - s$), and the matrix of variance and covariance for the independent latent variables (Σ) is set to an identity matrix.

Tables A.55, A.58, A.61 and A.64, in Appendix A, Section A.4, show the mean absolute relative bias and standard deviation os parameters in the strucutural model for Studies III,

IV, V and VI, respectively, when the sample sizes are set to 10, 20, 30, 50, 75, 100, 150 and 200. For example, from Table A.55 it is possible to see that when $n = 10$, using the center method, the mean absolute relative bias of the estimate of $\beta_{1,3}$ is 0.0495 with standard deviation 0.0765; and when $n = 20$, the MARB is 0.0382 with standard deviation 0.0295.

Tables A.56, A.59, A.62 and A.65, in Appendix A, Section A.4, show the MARB of the average of the eight loadings associated with each manifest variable, along with the average of the standard deviation of the loadings. For example, from Table A.56, the average MARB of the eight manifest variables associated with ξ_1 , using the symbolic covariance, when $n = 10$ is 0.1071 with standard deviation 0.0764, while for $n = 20$, the average MARB is 0.1059 with standard deviation 0.076.

Tables A.57, A.60, A.63 and A.66, in Appendix A, Section A.4, show the average lower and upper bounds for the root mean squared error (RMSE) of the interval-valued latent variables. For example, from Table A.57, the lower bound of the RMSE of the latent variable ξ_1 is 0.6097 and the upper bound is 0.6103 for the center method, while for the symbolic covariance method the lower bound RMSE is 0.6195 and the upper bound is 0.6208.

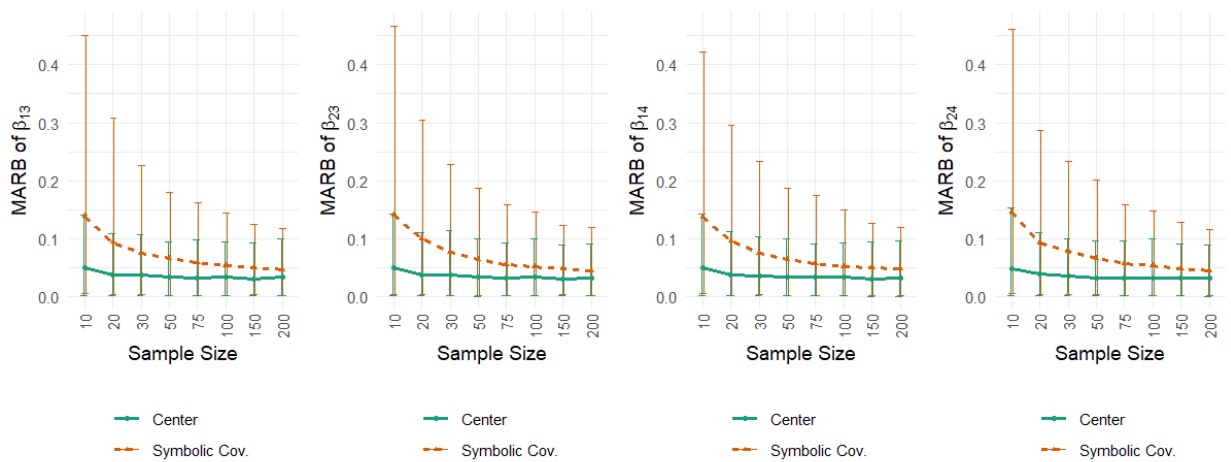


Figure 4.25: Mean absolute relative bias of coefficients in structural model for Study III with different sample sizes

Figure 4.25 shows the MARB of the parameters in the structural model for Study III as n increases, as well as the 95% confidence bounds obtained as the middle 95% of the absolute relative biases for the 1000 simulations. From this figure, it is possible to see that the symbolic covariance model is more affected by the change in sample size than the center method. The confidence bounds of the symbolic covariance method become clearly narrower as the sample size increases; the center method, on the other hand, remains approximately the same for sample sizes as low as 20 in this study, where no collinearity is directly introduced.

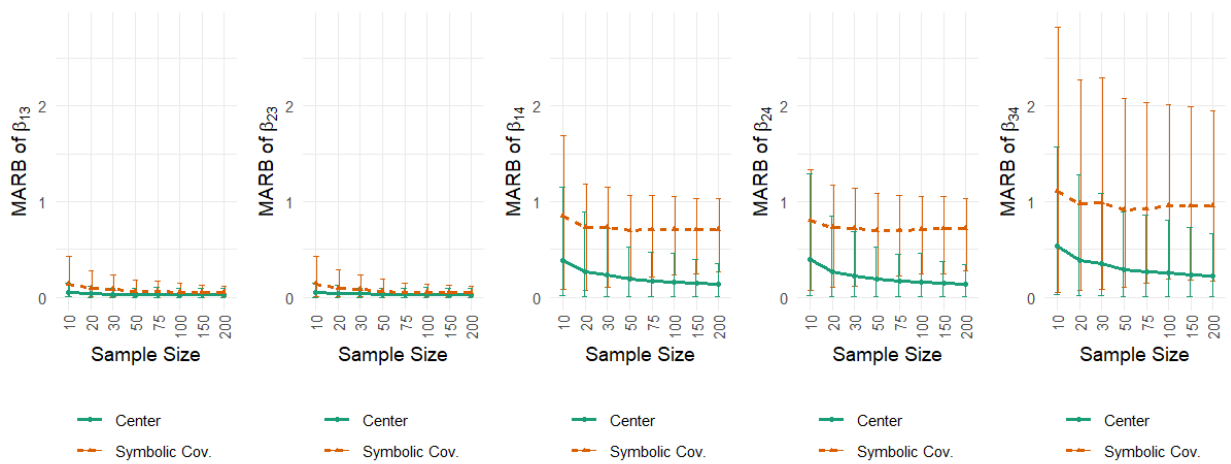


Figure 4.26: Mean absolute relative bias of coefficients in structural model for Study IV with different sample sizes

Figure 4.26 shows the MARB and 95% confidence bounds for the parameters in the structural model of Study IV. Recall that Study IV includes collinearity introduced by adding a relationship between the latent variables ξ_3 and ξ_4 ; additionally, this study has a structural model with four latent variables and five parameters. By considering the parameters that are affected by this collinearity, $\beta_{1,4}$, $\beta_{2,4}$ and $\beta_{3,4}$, it is possible to notice that the confidence bounds of the estimate for the center method tends to become narrower as the sample size increases. For the symbolic covariance method, for the parameters mentioned above, the confidence bounds of the estimate when $n = 10$ are wider than for the other sample sizes

evaluated, but the MARB is also extremely high for all other sample sizes. The estimated MARB of $\beta_{1,3}$ and $\beta_{2,3}$ behave similarly to the estimates in Study III, where the confidence bounds for the symbolic covariance method become narrower as the sample size increases, but there is not a big change for the center method.

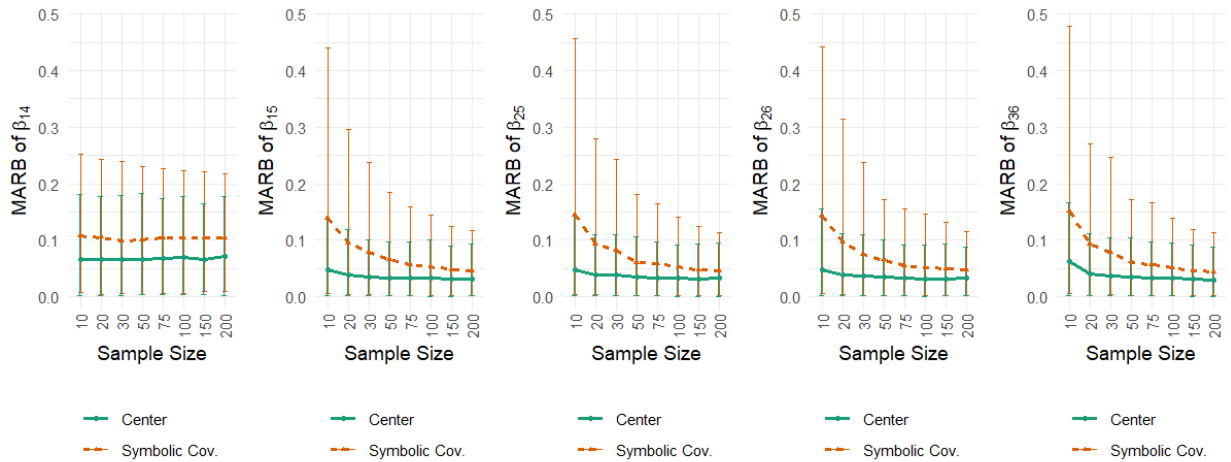


Figure 4.27: Mean absolute relative bias of coefficients in structural model for Study V with different sample sizes

Figure 4.27 show the MARB for the parameters in the structural model, and its 95% confidence bounds in Study V, which does not have a collinearity directly included. The estimates of the parameters $\beta_{1,5}$, $\beta_{2,5}$, $\beta_{2,6}$ and $\beta_{3,6}$ also behave in the same way as for the parameters in Study III, with the confidence bounds of the symbolic covariance method becoming narrower and close to zero as the sample size increase, while remaining mostly constant for the center method. However, the estimate of $\beta_{1,4}$ remains mainly constant for both the center and symbolic covariance methods; recall that this parameter is estimating the relationship between the dependent latent variable ξ_4 and its only independent latent variable ξ_1 .

Finally, the MARB and 95% confidence bounds for the parameters in the structural model of Study VI are presented Figure 4.28. The MARB of the estimate of $\beta_{1,4}$ shows

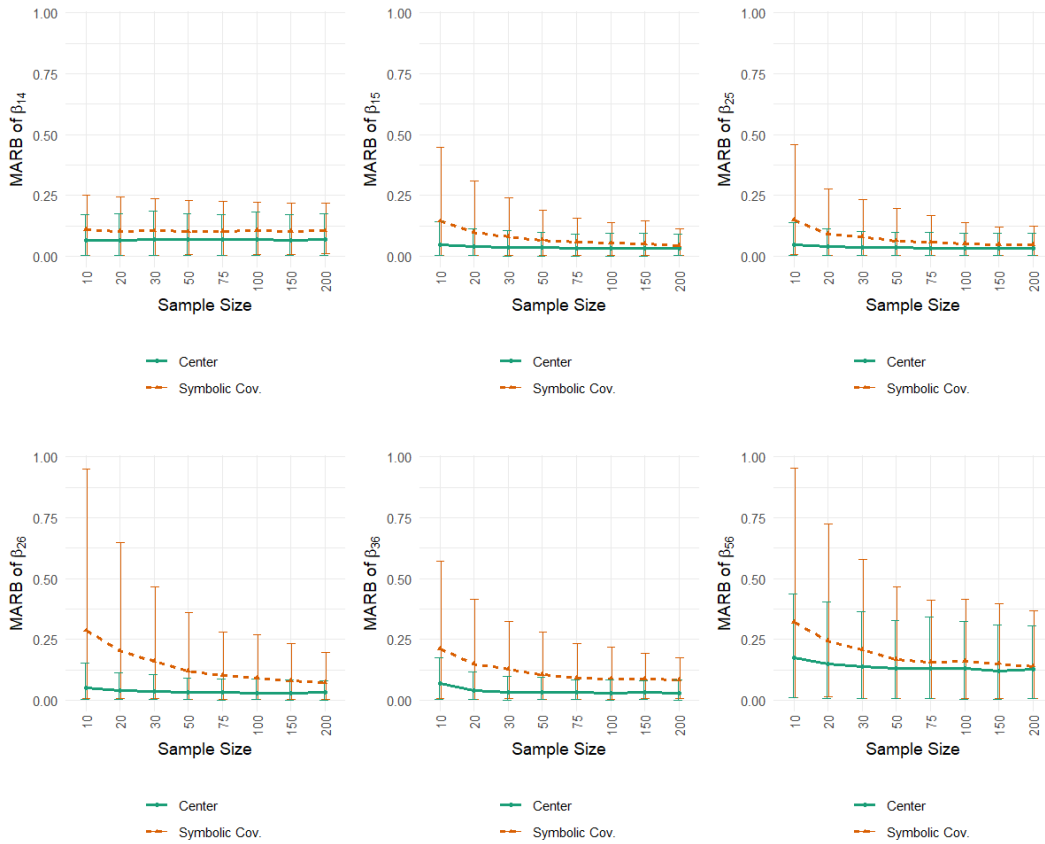


Figure 4.28: Mean absolute relative bias of coefficients in structural model for Study VI with different sample sizes

the same behavior as in Study V, with the confidence bounds remaining mostly the same for all sample sizes for both the center and symbolic covariance methods. The estimate of the parameters affected by the collinearity, $\beta_{2,6}$, $\beta_{3,6}$ and $\beta_{5,6}$, present higher MARB than the estimate of the parameters not affected by the collinearity. In general, the confidence bounds of the estimate of $\beta_{1,5}$, $\beta_{2,5}$, $\beta_{2,6}$, $\beta_{3,6}$ and $\beta_{5,6}$ in the symbolic covariance method becomes narrower for larger sample sizes and closer to zero, while the confidence bounds for the center method lie stay around the same values with the increase of the sample size.

Plots holding the same structure as the one presented for the structural model parameters,

but containing the average of the MARB for the eight manifest variables associated with each latent variable are presented in Appendix B, Section B.4, Figures B.21, B.22, B.23 and B.24. As observed in the other simulations described in this section, in general, the symbolic covariance method tends to produce wider confidence bounds for the absolute relative biases of the average loadings, this seems to be held constant across Studies III, IV, V and VI. The average MARB of the loadings of the latent variables that act solely as independent variables using the center method tend to have wider confidence bounds for smaller sample sizes, but given the scale of the average MARB, this difference is very small.

By comparing all of the results obtained in this section, it is possible to notice that the estimation with the symbolic covariance method is particularly affected by smaller sample sizes, when there is collinearity in the model and also when there is no collinearity. The center method presents more variable estimates with smaller sample sizes when the model has collinearity. The estimation of the parameters when $n = 10, 20$ and 30 seems to be more unstable for the center method when there is collinearity (see Figure 4.26). The symbolic covariance model seems to require a larger sample size to produce estimates closer to their real values. However, different studies suggest the need of different sample sizes for the estimates to have a narrower confidence bound in the symbolic covariance method. From Study III, where there are four latent variables and four relationships in the structural model, it seems necessary a sample of size at least 75 observations to reach better levels of MARB with the symbolic covariance regression method, but Studies V and VI might suggest that the confidence bounds of the MARB stabilizes when $n = 50$.

4.3 Conclusion

After all the simulation studies have been presented, it is possible to note that, overall, the estimation using the symbolic covariance regression method tends to produce estimates

ranging in a wider span than does the center regression method. This can be seen by comparing the 95% confidence intervals in the figures presented in Section 4.2; the confidence bounds show where the middle 95% of the absolute relative biases for the estimates for the 1000 simulations lie. Since the symbolic covariance method is calculated taking into consideration both the lower and upper limits of the intervals, this was expected as this method incorporates the interval variability of the data in the estimation.

If we compare the mean absolute relative biases (MARB) and 95% confidence bounds for the estimates in structural models with and without collinearity, a few patterns appear. The symbolic covariance method is very susceptible to the collinearity, with estimation of the parameters affected by the collinearity reaching levels of MARB 10 times greater (sometimes even more) than the estimation of parameters not directly affected by the collinearity. The center method, is also affected by collinearity, but in a much smaller scale, and produces narrower 95% confidence bounds than does the symbolic covariance method.

By looking closely at the results shown in the tables with the root mean squared error (RMSE), it is possible to note that when you compare two studies that have the same structure, but with one of them including collinearity (e.g., Studies I and II, Studies III and IV, Studies V and VI, and so on), the RMSE for the lower and upper bounds of the latent variables affected by collinearity is greater than it is for the model without it. This suggests that the collinearity affects negatively not only the precision of the parameters, but also the estimation of the latent variables. It seems beneficial to the estimation of the parameters in the structural model if a dependent latent variable is explained, or is associated with more than one independent latent variable, as long as it does not add collinearity to the model.

The results obtained when different numbers of manifest variables are used, in Section 4.2.2, suggest that for the symbolic covariance method, the estimation of the parameters in the structural model behaves similarly with few or many manifest variables associated with each latent variable; while the center method shows wider confidence bounds for smaller

number of manifest variables associated with each latent variable when the value of the loadings is randomly selected out of 0.5, 0.7 and 0.9. However, in the simulations presented in Section 4.2.2, it was noted that the estimation of the loadings is more affected by the true value of the loadings than the estimation in the structural model was. The smaller the value of the loading, the larger the average MARB for the loadings associated with each latent variable. Some research areas define the minimum standard for the loadings to be 0.7 (see Chin et al., 2003), and this also seems to be a good threshold for the interval-valued variables case.

Another aspect investigated was the effect of manipulating the range of the interval-valued variables, by changing the standard deviation of the error term of the manifest variables, σ . In general, larger σ results in estimations with larger mean absolute relative biases in the estimation of the parameters in the structural model as well as the parameters in the measurement model (i.e., the loadings). This holds for all studies evaluated, except for the study with more parameters estimated in the structural model than latent variables, while in the presence of collinearity (Study IV); in this specific study, there seems to be a larger instability in estimation of the parameters in the structural model.

After having evaluated the effect of the number of manifest variables and also the range of the interval-valued variables, the final simulation conducted intended to evaluate the estimation under different sample sizes. The results presented in Section 4.2.4 show that, as expected, the larger the sample size, the narrower the confidence bounds of the absolute relative biases. Simulations conducted with smaller sample sizes tend to have larger MARB and wider 95% confidence bounds. Different levels of complexity in the structural model require different sample sizes to reach a more stationary level for the confidence bounds of the absolute relative biases. In the center method, for a more complex model with collinearity, it may need 50 observations to reach a more stable level of the confidence bounds, compared to the results for larger sample sizes; while for simpler models, it is reached with samples

of size as low as $n = 30$. The symbolic covariance method required a larger sample size to produce estimates closer to the true value. It may seem necessary that a sample size be at least $n = 75$ to reach better levels of the upper bound of the confidence bounds of the absolute relative biases, but simpler models might reach that level with $n = 50$

The main trends and patterns observed in the results from Section 4.2 were brought up in this section, but many other findings can be made from the tables in Appendix A, and the figures in Section 4.2 and Appendix B. In conclusion, the partial least squares path modeling for interval-valued variables proposed in Chapter 4 seems to produce reliable estimates, but is highly affected by the choice of organization in the structural model and also the width of the interval-valued observations. Further studies should be carried out to continue to understand the accuracy and reliability of the method; some of the aspects that can be investigated are described in Chapter 5.

Chapter 5

Future Work

The work shown in this dissertation focuses mainly on the partial least squares path modeling estimation for interval-valued data, but there still are some aspects of the estimation that need to be studied further. By considering the function described to conduct the PLM-PM for interval-valued data, this function needs to be enhanced to allow the user to use bootstrapping resampling techniques to evaluate the significance and to build confidence intervals for the parameters estimated. This is necessary since no distributional assumption is made in the model, and parametric techniques should not be used. Future simulation studies should also assess the behavior of different structural model organizations, and also expand the simulations to include manifest variables that relate to the latent variables in a formative way.

Ever since symbolic data analysis was first introduced, most of the theory developed for interval and histogram-valued data is based on the assumption that the observations within the intervals or sub-intervals are uniformly distributed. The exceptions are a few studies proposing the estimation via maximum likelihood estimators under the assumption that the observations within the intervals or sub-intervals follow a triangular distribution, see Le-Rademacher and Billard (2011). These researches show great promise in expanding the

scenarios for which the symbolic data analysis will not only be appropriate, but also adequate, without violation of the strong assumption of the internal distribution of the intervals. Even though there has been some progress in the area, it still seems important to understand the effect on the estimation of parameters in the presence of misspecification on the internal distribution of the interval, when assumed to be uniformly distributed. Even though the results and simulations are not presented, Section 5.1 shows the way the simulated data should be generated in order to answer the questions raised above.

5.1 Simulating data for misspecification study

To understand the research questions proposed above, data need to be simulated under two separate studies. The first study is such that the response variable is uniformly distributed within the interval, while in the second study this variable is not uniformly distributed. The first study is set as a baseline to understand the effect of misspecification in evaluating the estimation; and will help understand how the methods behave under the correct assumptions. All other parameters and methodologies are held constant whenever possible. The misspecification is added only on the response variable. This will allow the same explanatory variables to be used under both studies.

Liu (2016) and Liu and Billard (2019) proposed four different methods to generate interval valued data, presenting the advantages and disadvantages of each method. The ‘Simulation Method IV’ proposed by Liu (2016) produces data that are uniformly distributed, but which on the other hand do not represent completely the way that usually happens in practice. Despite this disadvantage, the method IV seems to be the best way to conduct this simulation study, given that it ensures that the response variable does not violate the uniformity assumption, so most closely represent how data occur in practice.

The data to be generated will follow the relationship present in a linear model, as $\mathbf{Y} =$

$\mathbf{X}'\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where \mathbf{Y} is the vector of interval-valued response variable, \mathbf{X} is a matrix of p interval-valued explanatory variables and $\boldsymbol{\epsilon}$ is a vector of the error terms, also an interval-valued component.

This process starts by generating the explanatory variables that consist of two separate elements, the center and the range. The center of the p variables and n observations is set to follow a multivariate normal distribution with mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\Sigma}$, as $\mathbf{X}^{(c)} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Using the multivariate normal allows the model to have some covariance between different explanatory variables, or to use a diagonal matrix when no such relationship is assumed. The range of the p explanatory variables is set to follow an exponential distribution with parameter λ , as the range needs to be greater than zero, $\mathbf{X}_j^{(r)} \sim Exp(\lambda)$, for $j = 1, \dots, p$. The smaller the lambda the larger the mean of the ranges generated and therefore wider intervals are produced. Once the matrix $\mathbf{X}_{n \times p}$ has been created, by subtracting the half range to the mean of each observation to form the lower bound and adding the half range to form the upper bound of the observation, $x_{ij} = [x_{ijL}, x_{ijU}] = [x_{ij}^{(c)} - 0.5x_{ij}^{(r)}, x_{ij}^{(c)} + 0.5x_{ij}^{(r)}]$ for $i = 1, \dots, n$ and $j = 1, \dots, p$, the specified relationship of the parameters is added, as well as the error term, to generate the response variable.

The response variable is generated individually for each of n observations, and is done via aggregation of a large number of observations, m . For the i^{th} observation and the j^{th} variable, m random observations ($x_{ijk}^{(temp)}$) are generated from a uniform distribution with parameters x_{ijL} and x_{ijU} , where $i = 1, \dots, n$, $j = 1, \dots, p$, $k = 1, \dots, m$, and the capital letter subscript L indicates the interval lower bound of the interval-value variable and the capital letter subscript U indicates the interval upper bound of the variable. Similarly, m observations ($e_{ik}^{(temp)}$) are generated from a normal distribution with mean zero and variance σ^2 , to compose the error term in the model. Then, m classical observations of the response variable can be obtained as $\mathbf{y}_i^{(temp)} = \mathbf{x}_i^{(temp)} \boldsymbol{\beta} + \mathbf{e}_i^{(temp)}$. Finally, the interval-valued observation for the i^{th} observation is then found by calculating the first and third quantiles of $\mathbf{y}_i^{(temp)}$. Xu (2010)

and Cariou and Billard (2015) showed that for large m , the observations within the first (Q1) and third (Q3) quantiles will be approximately uniformly distributed. This procedure is repeated n times until all interval-valued observations are generated. In the study where the misspecification was added to the response variable, the data randomly generated from a uniform distribution were replaced by data randomly generated from the pert distribution - see U.S. Department of the Navy (1958). The pert distribution is a bounded distribution that can assume a symmetrical or asymmetrical shape, depending on the value of its mode. It is sometimes compared to a smoother version of the triangular distribution. The probability density function for $X \sim Pert(a, b, c)$ is given by

$$f_x(x) = \begin{cases} \frac{(x-a)^{(\alpha-1)}(c-x)^{(\beta-1)}}{B(\alpha,\beta)(c-a)^{(\alpha+\beta-1)}}, & \text{for } a, b, c \in \mathbb{R}, b > a, c > b, \\ \alpha = 1 + 4\frac{b-a}{c-a}, \beta = 1 + 4\frac{c-b}{c-a} \\ 0, & \text{otherwise.} \end{cases}$$

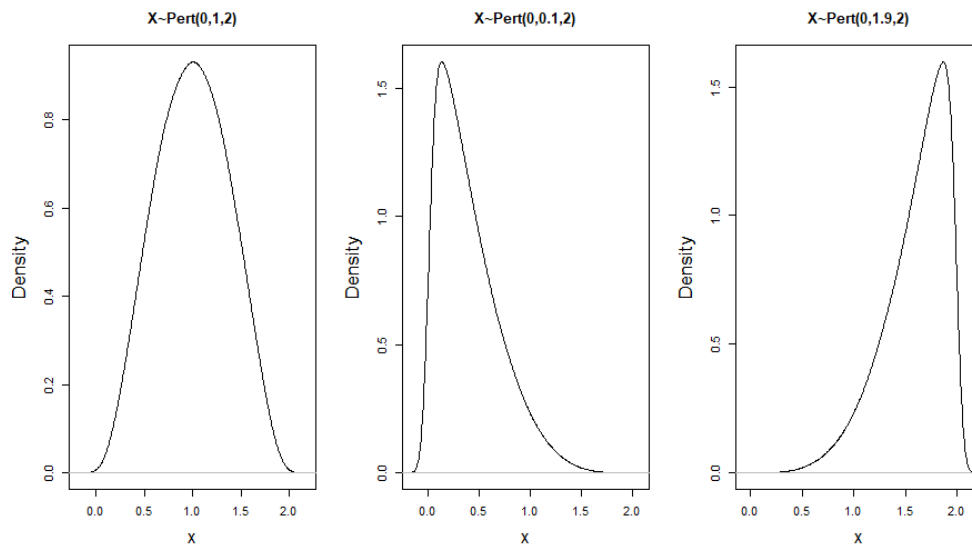


Figure 5.1: Pert Distribution with different modes

For illustration purposes, Figure 5.1 presents three different pert distributions. All of the distributions are bounded between zero and two, but have different modes. In this simulation

study, the parameters are set to be x_{ijL} and x_{ijU} for the minimum and the maximum, and $x_{ijL} + 0.1$ for the mode, $\mathbf{X}_{ij}^{(temp)} \sim Pert[x_{ijL}, x_{ijL} + 0.1, x_{ijU}]$, introducing a very large asymmetry.

The process to randomly generate data for this study described above is also presented in Algorithm 2.

Algorithm 2 Generating data

- 1: Generate the center of the n observations as $\mathbf{X}^{(c)} \sim \mathbf{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$
 - 2: Generate the range of the n observations as $\mathbf{X}_j^{(r)} \sim Exp(\lambda)$, for all $j = 1, \dots, p$
 - 3: Calculate $x_{ij} = [x_{ijL}, x_{ijU}] = [x_{ij}^{(c)} - 0.5x_{ij}^{(r)}, x_{ij}^{(c)} + 0.5x_{ij}^{(r)}]$ for all $i = 1, \dots, n$
 - 4: **for** $i = 1, \dots, n$, **do**
 - 5: **if** model is correctly specified **then**
 - 6: Generate m observations as $\mathbf{X}_{ij}^{(temp)} \sim Unif[x_{ijL}, x_{ijU}]$
 - 7: **else** (there is misspecification)
 - 8: Generate m observations as $\mathbf{X}_{ij}^{(temp)} \sim Pert[x_{ijL}, x_{ijL} + 0.1, x_{ijU}]$
 - 9: Generate m observations to compose the error term, $\mathbf{e}_i^{(temp)} \sim N(0, \sigma^2)$
 - 10: Calculate $\mathbf{y}_i^{(temp)} = \mathbf{x}_i^{(temp)} \boldsymbol{\beta} + \mathbf{e}_i^{(temp)}$, for a given vector $\boldsymbol{\beta}$ and design matrix $\mathbf{x}_i^{(temp)}$
 - 11: Obtain Q1 and Q3 of $\mathbf{y}_i^{(temp)}$, $y_i = [y_{iL}, y_{iU}] = [Q1(\mathbf{y}_i^{(temp)}), Q3(\mathbf{y}_i^{(temp)})]$
-

The R code to obtain the data described in Algorithm 2 is presented in Appendix C. Given that the data under the correct specification and misspecification are generated using Algorithm 2, simulation studies can be done to evaluate the behavior of the estimation of the parameters for different interval-valued variables regression methods at different sample sizes, and also different width of the interval-valued variables. This will provide an understanding of susceptibility of the regression methods, sample size and width of the interval to the misspecification of the assumption of uniformity within the interval-valued observations. This simulation study is left to be conducted in the future.

References

- [1] Anderson, T. W. (2003). Principal Components. In: *An Introduction to Multivariate Statistical Analysis* (pp. 451-479). New York: John Wiley and Sons.
- [2] Amato, S., Esposito Vinzi, V. and Tenenhaus, M. (2004). A Global Goodness-of-fit Index for PLS Structural Equation Modeling. *Oral Communication to PLS Club*, HEC School of Management, France, March 24.
- [3] Bertrand, P. and Goupil, F. (2000). Descriptive Statistics for Symbolic Data. *Analysis of Symbolic Data: Exploratory Methods for Extracting Statistical Information from Complex Data* (eds. H.-H. Bock and E. Diday). Springer-Verlag, Berlin, 103-124.
- [4] Billard, L. (2008). Sample Covariance Functions for Complex Quantitative Data. *Processing, World Conferences International Association of Statistical Computing 2008*, Yokohama, Japan.
- [5] Billard, L. and Diday, E. (2000). Regression Analysis for Interval-Valued Data. In: *Data Analysis, Classification, and Related Methods. Studies in Classification, Data Analysis, and Knowledge Organization* (eds. Kiers H. A. L., Rasson J. P., Groenen P. J. F. and Schader M.). Springer, Berlin, Heidelberg.
- [6] Billard, L. and Diday, E. (2006a). *Symbolic Data Analysis: Conceptual Statistics and Data Mining*. Wiley, New York.

- [7] Billard, L. and Diday, E. (2006b). Descriptive Statistics for Interval-valued Observations in the Presence of Rules. *Computational Statistics*, 21, 187-210.
- [8] Bock, H. H. (2000). The Classical Data Situation. In: *Analysis of Symbolic Data: Exploratory Methods for Extracting Statistical Information from Complex Data* (eds. H.-H. Bock and E. Diday). Springer-Verlag, Berlin, 24-38.
- [9] Boomsma, A. (1985). Nonconvergence, Improper Solutions, and Starting Values in LISREL Maximum Likelihood Estimation. *Psychometrika*, 50, 229–242.
- [10] Byrne, B. M. (1998). Structural Equation Modeling with LISREL, PRELIS and SIMPLIS: Basic Concepts, Application and Programming. *Psychology Press*, New York, NY.
- [11] Cariou, V. and Billard, L. (2015). Generalization Method when Manipulating Relational Database. *Revue des Nouvelles Technologies de l'Information*, 27, 59-86.
- [12] Chin, W. W. (1998). The Partial Least Squares Approach for Structural Equation Modeling. In: *Modern Methods for Business Research* (ed. G. A. Marcoulides). Hillsdale, NJ: Lawrence Erlbaum Associates.
- [13] Chin, W. and Newsted, P. R. (1999). Structural Equation Modeling Analysis with Small Samples Using Partial Least Squares. In: *Statistical Strategies for Small Sample Research* (ed. R. H. Hoyle). Thousands Oaks, Sage, 307-341.
- [14] Chin, W. W., Marcolin, B. L. and Newsted, P. R. (2003). A Partial Least Squares Latent Variable Modeling Approach for Measuring Interaction Effects: Results from a Monte Carlo Simulation Study and an Electronic-Mail Emotion/Adoption Study. *Information Systems Research*, 14, 189-217. 10.1287/isre.14.2.189.16018.

- [15] Crisci, A. (2012). Estimation Methods for the Structural Equation Models: Maximum Likelihood, Partial Least Squares and Generalized Maximum Entropy. *Journal of Applied Quantitative Methods*, 7, 3-17.
- [16] Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297–334.
- [17] Diday, E. (1987). Introduction à l'Approche Symbolique en Analyse des Données. *Premières Journées Symbolique - Numérique*. Ceremade, Université Paris, 21-56.
- [18] Dillon, W. and Goldstein, M. (1984). *Multivariate Analysis: Method and Applications*. John Wiley & Sons, Inc., New York, NY.
- [19] Draper, N. R. and Smith, H. (1998). *Applied Regression Analysis*. 3rd edition, John Wiley and Sons.
- [20] Haenlein, M. and Kaplan, A. M. (2004). A Beginner's Guide to PLS Analysis. *Understanding Statistics*, 3(4), 283–297.
- [21] Henseler, J., Dijkstra, T. K., Sarstedt, M., Ringle, C. M., Diamantopoulos, A., Straub, D. W., ... Calantone, R. J. (2014). Common Beliefs and Reality About PLS: Comments on Rönkkö and Evermann (2013). *Organizational Research Methods*, 17(2), 182–209.
- [22] Hulland, J., Ryan, M. J., and Rayner, R. K. (2010). Modeling Customer Satisfaction: A Comparative Performance Evaluation of Covariance Structure Analysis Versus Partial Least Squares. In: *Handbook of Partial Least Squares*, New York. Springer Handbooks of Computational Statistics.
- [23] Imakor, M. (2007). *Extension de la Régression PLE at de L'Approche PLS a des données de type intervalle*. Doctoral Dissertation, Université Paris Dauphine.

- [24] Joreskog, K. G. and Van Thillo, M. (1972). LISREL: A General Computer Program for Estimating a Linear Structural Equation System Involving Multiple Indicators of Unmeasured Variables. *TS Research Bulletin Series*, 1972: i-71.
- [25] Kaplan, D. (2009). *Structural Equation Modeling: Foundations and Extensions*, 2nd edition. Advanced Quantitative Techniques in the Social Sciences Series, 10.
- [26] Kleinbaum, D. G. and Kupper, L. L. (1978). *Applied Regression Analysis and Other Multivariate Methods*, 1st edition. Duxbury Press.
- [27] Kosmeji, K., Le-Rademacher, J. and Billard, L. (2014). Symbolic Covariance Matrix for Interval-valued Variables and its Application to Principal Component Analysis: a Case Study. *Metodoloski Zvezki*, 11, 1–20.
- [28] Lawson, C. L. and Hanson R. J. (1974). *Solving Least Squares Problem*. Prentice-Hall, New York.
- [29] Le-Rademacher, J. G. (2008). *Principal Component Analysis for Interval-valued and Histogram-valued Data and Likelihood Functions and Some Maximum Likelihood Estimators for Symbolic Data*. Doctoral Dissertation, University of Georgia.
- [30] Le-Rademacher, J. and Billard, L. (2011). Likelihood Functions and Some Maximum Likelihood Estimators for Symbolic Data. *Journal of Statistical Planning and Inference*, 141(4), 1593-1602.
- [31] Le-Rademacher, J. and Billard, L. (2017). Principal Component Analysis for Histogram-valued Data. *Advances in Data Analysis and Classification*, 11, 327-351.
- [32] Lima Neto, E. A. and De Carvalho, F. A. T. (2008). Centre and Range Method for Fitting a Linear Regression Model to Symbolic Interval Data. *Computational Statistics and Data Analysis*, 52, 1500-1515.

- [33] Lima Neto, E. A. and De Carvalho, F. A. T. (2010). Constrained Linear Regression Models for Symbolic Interval-valued Variables. *Computational Statistics and Data Analysis*, 54, 333-347.
- [34] Liu, F. (2016). *Cluster Analysis for Symbolic Interval Data Using Linear Regression Method*. Doctoral Dissertation, University of Georgia.
- [35] Liu, F. and Billard, L. (2019). Partitioning of Intervals Based on Regression. Manuscript submitted for publication.
- [36] Long, J. S. (1983). *Covariance Structure Models: An Introduction to LISREL*. Beverly Hills, Sage publications.
- [37] Mardia, K. V., Kent, J. T. and Bibby, J. M. (1979). Principal Component Analysis. In: *Multivariate Analysis* (pp. 213-254), Academic Press.
- [38] McDonald, R. P. (1996). Path Analysis with Composite Variables. *Multivariate Behavioral Research*, 31, 239-270.
- [39] McDonald, R. P. (2002). The Specific Analysis of Structural Equation Models. *Multivariate Behavioral Research*, 39, 687-713.
- [40] Meehl, P. E., Waller, N. G. (2002). The Path Analysis Controversy: a New Statistical Approach to Strong Appraisal of Verisimilitude. *Psychological Methods*, 2, 283-300.
- [41] Montgomery, D. C., Peck, E. A. and Vining G. G. (2001). *Introduction to Linear Regression Analysis*, 3rd edition, John Wiley and Sons.
- [42] Noonan, R. and Wold, H. (1982). PLS Path Modeling with Indirectly Observed Variables: A Comparison of Alternative Estimate for Latent Variables. In: *Systems Under Indirect Observation. Causality, Structure, Prediction* (eds. Jöreskog K. G, Wold, H). North-Holland, Amsterdam, 2, 75-94.

- [43] O'Rourke, N. and Hatcher, L. (2013). *A Step-by-Step Approach to Using SAS® for Factor Analysis and Structural Equation Modeling*, 2nd edition. SAS Institute Inc., Cary, North Carolina, USA.
- [44] Ronkko, M. and Evermann, J. (2013). A Critical Examination of Common Beliefs About Partial Least Squares Path Modeling. *Organizational Research Methods*, 16(3), 425-448.
- [45] Sanchez, G. (2013). *PLS Path Modeling with R*. Trowchez Editions. Berkeley.
- [46] Sanchez, G., Trinchera, L. and Rusolillo, G. (2017). plspm: Tools for Partial Least Squares Path Modeling (PLS-PM). R package version 0.4.9.
- [47] Sheather, S. J. (2009). *A Modern Approach to Regression with R*. Springer-Verlag, New York.
- [48] Tenenhaus, M., Esposito Vinzi, V., Chatelin, Y. and Lauro, C. (2005). PLS Path Modeling. *Computational Statistics and Data Analysis*, 48, 159-205.
- [49] U.S. Department of the Navy (1958). *PERT, Program Evaluation Research Task*, Phase I Summary Report. Washington, D.C.: Special Projects Office, Bureau of Ordnance.
- [50] Wishart, J. (1928). The Generalised Product Moment Distribution in Samples from a Normal Multivariate Population. *Biometrika*, 20A, (1-2): 32-52.
- [51] Wold, H. (1973). Nonlinear Iterative Partial Least Squares (NIPALS) Modelling: Some Current Developments. *Multivariate Analysis*, 3, 383-407.
- [52] Wold, H. (1975). *Path Models with Latent Variables: the NIPALS Approach*. New York, NY: Academic Press.

- [53] Wold, H. (1979). Model Construction and Evaluation when Theoretical Knowledge is Scarce: and Example of the Use of Partial Least Squares (Cahier 79.06 du département d'économétrie, faculté des sciences économiques et sociales). Genève: Université de Genève.
- [54] Wold, H. (1982). Soft Modeling: Intermediate between Traditional Model Building and Data Analysis. *Mathematical Statistics*, 6, 333-346.
- [55] Wold, S., Martens, H. and Wold, H. (1983). The Multivariate Calibration Problem in Chemistry Solved by the PLS Method. *Lectures Notes in Mathematics*, 973, 286-293.
- [56] Wong, C. S., Masaro, J., and Wang, T. (1991). Multivariate Versions of Cochran's Theorems. *Journal of Multivariate Analysis*, 39, 154-174.
- [57] Xu, W. (2010). *Symbolic Data Analysis: Interval-valued Data Regression*. Doctoral Dissertation, University of Georgia.

Appendices

Appendix A

Tables

A.1 Simulation 1 - Path Study

Table A.1: Absolute relative bias of structural model of Study I

Method	Estimator	Path coef. 0.5		Path coef. 0.7		Path coef. 0.9		Random Path coef.	
		Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
Center	$\hat{\beta}_{1,3}$	0.1156	0.0401	0.0542	0.0316	0.0311	0.0231	0.0701	0.0502
	$\hat{\beta}_{1,4}$	0.0526	0.0332	0.0284	0.0216	0.0272	0.0198	0.0324	0.0249
	$\hat{\beta}_{2,4}$	0.0541	0.0348	0.0279	0.0211	0.0265	0.0206	0.0317	0.0239
Symbolic Cov.	$\hat{\beta}_{1,3}$	0.2005	0.0168	0.0804	0.0156	0.0269	0.0132	0.1068	0.0755
	$\hat{\beta}_{1,4}$	0.0428	0.0324	0.0481	0.0349	0.0689	0.0397	0.053	0.0392
	$\hat{\beta}_{2,4}$	0.0426	0.0325	0.0479	0.0337	0.0672	0.0374	0.0508	0.0371

Table A.2: Average absolute relative bias of the loadings associated with each latent variable, for Study I

Method	Latent Variable	Path coef. 0.5		Path coef. 0.7		Path coef. 0.9		Random Path coef.	
		Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
Center	ξ_1	0.0988	0.0276	0.0995	0.0287	0.0984	0.0276	0.099	0.0287
	ξ_2	0.1011	0.0287	0.1006	0.0284	0.0985	0.0278	0.0999	0.0276
	ξ_3	0.0268	0.0182	0.0468	0.0267	0.0786	0.0264	0.0496	0.0318
	ξ_4	0.0482	0.0271	0.0899	0.0265	0.1069	0.0258	0.0853	0.0313
Symbolic Cov.	ξ_1	0.1048	0.0751	0.1073	0.0757	0.1048	0.0742	0.1061	0.0756
	ξ_2	0.1056	0.0746	0.1054	0.0748	0.1056	0.075	0.1066	0.0751
	ξ_3	0.075	0.0321	0.071	0.0596	0.0853	0.075	0.0756	0.0571
	ξ_4	0.0715	0.0607	0.0929	0.0748	0.1134	0.0744	0.0934	0.0739

Table A.3: Average root mean squared error for latent variable, for Study I

Method	Latent Variable	Path coef. 0.5		Path coef. 0.7		Path coef. 0.9		Random Path coef.	
		$RMSE_L$	$RMSE_U$	$RMSE_L$	$RMSE_U$	$RMSE_L$	$RMSE_U$	$RMSE_L$	$RMSE_U$
Center	ξ_1	0.6192	0.6193	0.6227	0.6228	0.6214	0.6216	0.6222	0.6225
	ξ_2	0.6252	0.6253	0.6259	0.626	0.6221	0.6223	0.6237	0.6242
	ξ_3	0.0674	0.0675	0.1232	0.1233	0.2212	0.2214	0.134	0.134
	ξ_4	0.1199	0.1199	0.2548	0.255	0.3418	0.342	0.2494	0.2496
Symbolic Cov.	ξ_1	0.6287	0.6288	0.6331	0.6331	0.6295	0.6295	0.6317	0.6319
	ξ_2	0.6299	0.6299	0.6276	0.6278	0.6297	0.6306	0.6317	0.6317
	ξ_3	0.0647	0.0647	0.1302	0.1303	0.2306	0.2306	0.1357	0.1359
	ξ_4	0.128	0.1281	0.259	0.2591	0.3531	0.3533	0.2554	0.2554

Table A.4: Absolute relative bias of structural model of Study II

Method	Estimator	Path coef. 0.5		Path coef. 0.7		Path coef. 0.9		Random Path coef.	
		Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
Center	$\hat{\beta}_{1,3}$	0.1152	0.04	0.0555	0.0329	0.0318	0.0241	0.0671	0.0489
	$\hat{\beta}_{1,4}$	0.0982	0.0767	0.11	0.0823	0.1332	0.101	0.1223	0.0999
	$\hat{\beta}_{2,4}$	0.0326	0.0246	0.0289	0.0209	0.0355	0.0244	0.03	0.0216
	$\hat{\beta}_{3,4}$	0.4333	0.1316	0.2569	0.1267	0.1321	0.1013	0.2856	0.1738
Symbolic Cov.	$\hat{\beta}_{1,3}$	0.2008	0.0166	0.0803	0.015	0.0265	0.013	0.1019	0.0753
	$\hat{\beta}_{1,4}$	1.0091	1.181	1.3696	1.73	1.6874	2.0601	1.4542	1.9378
	$\hat{\beta}_{2,4}$	0.0547	0.0418	0.0875	0.0533	0.1048	0.0589	0.0856	0.0548
	$\hat{\beta}_{3,4}$	1.4944	1.2167	1.6089	1.7716	1.7393	2.0654	1.7212	1.9561

Table A.5: Average absolute relative bias of the loadings associated with each latent variable, for Study II

Method	Latent Variable	Path coef. 0.5		Path coef. 0.7		Path coef. 0.9		Random Path coef.	
		Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
Center	ξ_1	0.1001	0.0284	0.0992	0.0288	0.0985	0.028	0.0996	0.0289
	ξ_2	0.1005	0.0285	0.1008	0.0288	0.0983	0.0282	0.0993	0.0284
	ξ_3	0.0256	0.0178	0.0453	0.0281	0.0772	0.027	0.0496	0.0316
	ξ_4	0.0791	0.0264	0.1153	0.0258	0.1313	0.0249	0.1129	0.0281
Symbolic Cov.	ξ_1	0.1051	0.075	0.1052	0.0748	0.1053	0.0754	0.1058	0.0753
	ξ_2	0.1056	0.0746	0.1049	0.0746	0.1052	0.075	0.1042	0.0749
	ξ_3	0.075	0.0323	0.0716	0.0597	0.0861	0.0749	0.0767	0.0581
	ξ_4	0.0844	0.0748	0.1204	0.0737	0.136	0.0727	0.1185	0.0744

Table A.6: Average root mean squared error for latent variable, for Study II

Method	Latent Variable	Path coef. 0.5		Path coef. 0.7		Path coef. 0.9		Random Path coef.	
		$RMSE_L$	$RMSE_U$	$RMSE_L$	$RMSE_U$	$RMSE_L$	$RMSE_U$	$RMSE_L$	$RMSE_U$
Center	ξ_1	0.6244	0.6249	0.623	0.6234	0.6208	0.6214	0.623	0.6233
	ξ_2	0.6228	0.6231	0.6255	0.6258	0.6202	0.6203	0.6214	0.6214
	ξ_3	0.0664	0.0665	0.1213	0.1215	0.2188	0.219	0.1353	0.1355
	ξ_4	0.2101	0.2102	0.3739	0.374	0.481	0.481	0.3685	0.3686
Symbolic Cov.	ξ_1	0.6291	0.6291	0.6287	0.6292	0.6295	0.6302	0.6303	0.6307
	ξ_2	0.6287	0.6289	0.627	0.6273	0.6269	0.6269	0.6253	0.6259
	ξ_3	0.0665	0.0667	0.1301	0.1302	0.2314	0.2318	0.1408	0.1409
	ξ_4	0.2164	0.2164	0.3821	0.3822	0.4867	0.4868	0.3748	0.3749

Table A.7: Absolute relative bias of structural model of Study III

Method	Estimator	Path coef. 0.5		Path coef. 0.7		Path coef. 0.9		Random Path coef.	
		Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
Center	$\hat{\beta}_{1,3}$	0.0555	0.0345	0.0274	0.0208	0.0284	0.0216	0.0326	0.0253
	$\hat{\beta}_{2,3}$	0.0544	0.034	0.0288	0.0217	0.0284	0.0217	0.0327	0.0261
	$\hat{\beta}_{1,4}$	0.056	0.0341	0.0289	0.0215	0.0274	0.0208	0.0334	0.0247
	$\hat{\beta}_{2,4}$	0.0553	0.0344	0.0296	0.0223	0.0277	0.0214	0.0327	0.025
Symbolic Cov.	$\hat{\beta}_{1,3}$	0.0443	0.0318	0.0473	0.0345	0.0685	0.0387	0.0527	0.0398
	$\hat{\beta}_{2,3}$	0.0434	0.032	0.0483	0.0338	0.0668	0.0373	0.0534	0.0391
	$\hat{\beta}_{1,4}$	0.044	0.0314	0.0469	0.0346	0.0683	0.0379	0.0529	0.0409
	$\hat{\beta}_{2,4}$	0.043	0.0319	0.0482	0.0341	0.0667	0.0381	0.054	0.0398

Table A.8: Average absolute relative bias of the loadings associated with each latent variable, for Study III

Method	Latent Variable	Path coef. 0.5		Path coef. 0.7		Path coef. 0.9		Random Path coef.	
		Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
Center	ξ_1	0.1013	0.0279	0.0969	0.0277	0.1021	0.0292	0.0983	0.0279
	ξ_2	0.0998	0.0286	0.1004	0.0278	0.1000	0.0293	0.0996	0.0288
	ξ_3	0.0476	0.0277	0.0885	0.0264	0.1085	0.027	0.0862	0.0329
	ξ_4	0.0468	0.0273	0.0868	0.0266	0.1076	0.0264	0.0853	0.032
Symbolic Cov.	ξ_1	0.104	0.0745	0.1057	0.0745	0.106	0.075	0.1061	0.0747
	ξ_2	0.1049	0.0747	0.1058	0.0751	0.1058	0.0751	0.1064	0.0747
	ξ_3	0.0718	0.0607	0.0944	0.0753	0.1133	0.074	0.0948	0.0745
	ξ_4	0.0711	0.0604	0.0931	0.075	0.1133	0.0741	0.0938	0.0736

Table A.9: Average root mean squared error for latent variable, for Study III

Method	Latent Variable	Path coef. 0.5		Path coef. 0.7		Path coef. 0.9		Random Path coef.	
		$RMSE_L$	$RMSE_U$	$RMSE_L$	$RMSE_U$	$RMSE_L$	$RMSE_U$	$RMSE_L$	$RMSE_U$
Center	ξ_1	0.6255	0.6262	0.6179	0.6183	0.6259	0.6262	0.6197	0.6200
	ξ_2	0.625	0.6251	0.6256	0.6257	0.623	0.6235	0.6225	0.6226
	ξ_3	0.1191	0.1192	0.2516	0.2516	0.3464	0.3465	0.2502	0.2502
	ξ_4	0.1175	0.1176	0.2483	0.2484	0.3447	0.3447	0.248	0.248
Symbolic Cov.	ξ_1	0.628	0.628	0.6316	0.6319	0.6288	0.6288	0.63	0.63
	ξ_2	0.6286	0.6288	0.6301	0.6304	0.6307	0.6308	0.6299	0.63
	ξ_3	0.1265	0.1265	0.261	0.261	0.3528	0.3528	0.2589	0.2589
	ξ_4	0.1243	0.1243	0.2585	0.2585	0.3531	0.3531	0.2595	0.2596

Table A.10: Absolute relative bias of structural model of Study IV

Method	Estimator	Path coef. 0.5		Path coef. 0.7		Path coef. 0.9		Random Path coef.	
		Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
Center	$\hat{\beta}_{1,3}$	0.0528	0.0332	0.0293	0.0215	0.0274	0.0204	0.0347	0.027
	$\hat{\beta}_{2,3}$	0.0544	0.0348	0.03	0.0232	0.0274	0.0198	0.0341	0.0265
	$\hat{\beta}_{1,4}$	0.1114	0.0868	0.1486	0.1022	0.1805	0.118	0.1522	0.1116
	$\hat{\beta}_{2,4}$	0.1116	0.0866	0.1484	0.1011	0.1814	0.1188	0.1552	0.1116
	$\hat{\beta}_{3,4}$	0.1994	0.1348	0.1718	0.1283	0.4032	0.1929	0.2432	0.1937
Symbolic Cov.	$\hat{\beta}_{1,3}$	0.0422	0.0311	0.0508	0.0351	0.0676	0.0392	0.0525	0.0387
	$\hat{\beta}_{2,3}$	0.0445	0.0327	0.0484	0.0359	0.0694	0.0404	0.0517	0.0392
	$\hat{\beta}_{1,4}$	0.8307	0.1137	0.8659	0.1137	0.8862	0.1227	0.7072	0.2171
	$\hat{\beta}_{2,4}$	0.8303	0.1145	0.8636	0.115	0.8867	0.1193	0.7064	0.2179
	$\hat{\beta}_{3,4}$	0.8882	0.179	1.1127	0.1874	1.372	0.2082	0.929	0.4729

Table A.11: Average absolute relative bias of the loadings associated with each latent variable, for Study IV

Method	Latent Variable	Path coef. 0.5		Path coef. 0.7		Path coef. 0.9		Random Path coef.	
		Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
Center	ξ_1	0.0994	0.029	0.0992	0.0279	0.0993	0.0285	0.1002	0.0289
	ξ_2	0.1013	0.0292	0.1001	0.029	0.0986	0.0286	0.0993	0.0285
	ξ_3	0.048	0.0265	0.0872	0.0268	0.1077	0.0263	0.0845	0.033
	ξ_4	0.0976	0.0268	0.1255	0.0255	0.1381	0.0237	0.1244	0.0255
Symbolic Cov.	ξ_1	0.105	0.0751	0.1063	0.0748	0.1064	0.0754	0.1057	0.0756
	ξ_2	0.107	0.075	0.1062	0.0748	0.1066	0.075	0.1057	0.0755
	ξ_3	0.0718	0.0606	0.0938	0.0754	0.1137	0.0741	0.0938	0.0739
	ξ_4	0.1028	0.0745	0.1326	0.0728	0.1441	0.0719	0.1319	0.0736

Table A.12: Average root mean squared error for latent variable, for Study IV

Method	Latent Variable	Path coef. 0.5		Path coef. 0.7		Path coef. 0.9		Random Path coef.	
		$RMSE_L$	$RMSE_U$	$RMSE_L$	$RMSE_U$	$RMSE_L$	$RMSE_U$	$RMSE_L$	$RMSE_U$
Center	ξ_1	0.6226	0.6227	0.6223	0.6224	0.621	0.6214	0.6246	0.6248
	ξ_2	0.6261	0.6264	0.6242	0.6246	0.6235	0.6236	0.6229	0.6233
	ξ_3	0.1197	0.1198	0.2489	0.249	0.3435	0.3435	0.2456	0.2456
	ξ_4	0.2807	0.2808	0.4438	0.4439	0.5453	0.5453	0.4385	0.4385
Symbolic Cov.	ξ_1	0.6285	0.6289	0.6318	0.632	0.6322	0.6322	0.6286	0.629
	ξ_2	0.6313	0.6315	0.6304	0.6306	0.6302	0.6307	0.6283	0.6283
	ξ_3	0.1281	0.1282	0.2597	0.2598	0.353	0.3533	0.2559	0.256
	ξ_4	0.2899	0.29	0.456	0.4561	0.5543	0.5544	0.4514	0.4514

Table A.13: Absolute relative bias of structural model of Study V

Method	Estimator	Path coef. 0.5		Path coef. 0.7		Path coef. 0.9		Random Path coef.	
		Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
Center	$\hat{\beta}_{1,4}$	0.1145	0.039	0.0555	0.0324	0.0325	0.0236	0.069	0.0487
	$\hat{\beta}_{1,5}$	0.0538	0.0334	0.0298	0.023	0.0272	0.0207	0.0335	0.0259
	$\hat{\beta}_{2,5}$	0.0538	0.033	0.0286	0.0225	0.0276	0.0212	0.0321	0.0249
	$\hat{\beta}_{2,6}$	0.0541	0.0337	0.029	0.0224	0.0279	0.0211	0.0332	0.0271
	$\hat{\beta}_{3,6}$	0.0542	0.0333	0.0301	0.0226	0.0277	0.0205	0.032	0.0257
Symbolic Cov.	$\hat{\beta}_{1,4}$	0.2013	0.0163	0.0799	0.0152	0.0256	0.0131	0.1039	0.0738
	$\hat{\beta}_{1,5}$	0.0425	0.0311	0.0477	0.0343	0.0689	0.0409	0.0516	0.0379
	$\hat{\beta}_{2,5}$	0.0401	0.0296	0.0477	0.0351	0.069	0.0394	0.0524	0.0399
	$\hat{\beta}_{2,6}$	0.0425	0.0316	0.0485	0.0342	0.0703	0.0401	0.0513	0.0382
	$\hat{\beta}_{3,6}$	0.0438	0.0327	0.0488	0.0349	0.0681	0.0388	0.053	0.0387

Table A.14: Average absolute relative bias of the loadings associated with each latent variable, for Study V

Method	Latent Variable	Path coef. 0.5		Path coef. 0.7		Path coef. 0.9		Random Path coef.	
		Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
Center	ξ_1	0.0982	0.0274	0.1005	0.0283	0.1008	0.0286	0.1011	0.0289
	ξ_2	0.0997	0.0277	0.099	0.0281	0.0966	0.0287	0.0996	0.0284
	ξ_3	0.0984	0.0289	0.099	0.0282	0.0994	0.0276	0.099	0.0282
	ξ_4	0.0259	0.018	0.0456	0.0256	0.0794	0.0266	0.0506	0.0334
	ξ_5	0.0468	0.0265	0.0887	0.0273	0.1063	0.0263	0.0853	0.0306
	ξ_6	0.0472	0.0268	0.0889	0.0276	0.1068	0.0258	0.0848	0.0318
Symbolic Cov.	ξ_1	0.107	0.075	0.1055	0.0754	0.1044	0.0748	0.1052	0.0744
	ξ_2	0.1063	0.0754	0.1046	0.0747	0.1051	0.0751	0.1064	0.0753
	ξ_3	0.1062	0.0751	0.106	0.0752	0.1055	0.0755	0.1054	0.0747
	ξ_4	0.0756	0.032	0.0718	0.0599	0.0845	0.0749	0.0765	0.0579
	ξ_5	0.0713	0.0604	0.0954	0.0749	0.1124	0.0744	0.0936	0.0741
	ξ_6	0.0714	0.0601	0.0923	0.0754	0.1136	0.074	0.096	0.0747

Table A.15: Average root mean squared error for latent variable, for Study V

Method	Latent Variable	Path coef. 0.5		Path coef. 0.7		Path coef. 0.9		Random Path coef.	
		$RMSE_L$	$RMSE_U$	$RMSE_L$	$RMSE_U$	$RMSE_L$	$RMSE_U$	$RMSE_L$	$RMSE_U$
Center	ξ_1	0.6215	0.6217	0.6234	0.6235	0.6234	0.6239	0.6236	0.6245
	ξ_2	0.624	0.6244	0.6235	0.6236	0.6177	0.6178	0.6245	0.6246
	ξ_3	0.6202	0.6202	0.6201	0.6203	0.6213	0.6214	0.6204	0.6205
	ξ_4	0.0668	0.0669	0.121	0.1211	0.2232	0.2235	0.1374	0.1376
	ξ_5	0.1174	0.1175	0.2524	0.2524	0.341	0.3411	0.2497	0.2497
	ξ_6	0.1183	0.1183	0.2527	0.2528	0.3414	0.3415	0.2475	0.2476
Symbolic Cov.	ξ_1	0.6329	0.6331	0.6297	0.6304	0.6293	0.6293	0.629	0.6295
	ξ_2	0.631	0.6311	0.6289	0.6293	0.6275	0.6278	0.63	0.6303
	ξ_3	0.632	0.6326	0.6304	0.6309	0.631	0.6311	0.6283	0.6285
	ξ_4	0.0646	0.0647	0.1312	0.1313	0.2287	0.2288	0.1392	0.1393
	ξ_5	0.1264	0.1264	0.2642	0.2643	0.3507	0.3508	0.2554	0.2554
	ξ_6	0.1262	0.1262	0.257	0.257	0.3536	0.3537	0.2624	0.2625

Table A.16: Absolute relative bias of structural model of Study VI

Method	Estimator	Path coef. 0.5		Path coef. 0.7		Path coef. 0.9		Random Path coef.	
		Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
Center	$\hat{\beta}_{1,4}$	0.1149	0.0403	0.0556	0.0319	0.0321	0.0235	0.0653	0.0465
	$\hat{\beta}_{1,5}$	0.0529	0.0329	0.0293	0.023	0.0277	0.0206	0.0327	0.0249
	$\hat{\beta}_{2,5}$	0.0515	0.0324	0.0294	0.0223	0.028	0.021	0.0326	0.0237
	$\hat{\beta}_{2,6}$	0.0306	0.0237	0.0301	0.0225	0.0379	0.0271	0.0294	0.0224
	$\hat{\beta}_{3,6}$	0.0317	0.0245	0.0299	0.0221	0.0368	0.0254	0.03	0.0217
	$\hat{\beta}_{5,6}$	0.2537	0.0419	0.05	0.0361	0.2245	0.0708	0.1267	0.0907
Symbolic Cov.	$\hat{\beta}_{1,4}$	0.2017	0.0161	0.0803	0.0146	0.0265	0.0133	0.1023	0.0745
	$\hat{\beta}_{1,5}$	0.0425	0.0307	0.0485	0.0348	0.0711	0.0419	0.0504	0.0375
	$\hat{\beta}_{2,5}$	0.0444	0.0326	0.0479	0.0332	0.07	0.0417	0.0522	0.0389
	$\hat{\beta}_{2,6}$	0.0713	0.0525	0.086	0.0623	0.0958	0.0706	0.0894	0.0662
	$\hat{\beta}_{3,6}$	0.0588	0.0434	0.0899	0.054	0.1075	0.0588	0.0912	0.0565
	$\hat{\beta}_{5,6}$	0.2936	0.0957	0.0993	0.0699	0.177	0.1036	0.1507	0.1075

Table A.17: Average absolute relative bias of the loadings associated with each latent variable, for Study VI

Method	Latent Variable	Path coef. 0.5		Path coef. 0.7		Path coef. 0.9		Random Path coef.	
		Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
Center	ξ_1	0.0993	0.0275	0.1003	0.0275	0.0999	0.0282	0.1000	0.0281
	ξ_2	0.0979	0.0274	0.1003	0.0287	0.0976	0.0292	0.1008	0.0276
	ξ_3	0.1002	0.0293	0.0984	0.0278	0.0989	0.0278	0.099	0.0281
	ξ_4	0.0263	0.018	0.0463	0.0273	0.0782	0.0264	0.0494	0.0312
	ξ_5	0.0482	0.0261	0.0876	0.0274	0.1071	0.0259	0.0864	0.0319
	ξ_6	0.0816	0.0261	0.1174	0.0253	0.1341	0.0255	0.1169	0.0265
Symbolic Cov.	ξ_1	0.1049	0.075	0.1059	0.0749	0.1058	0.0747	0.1043	0.0746
	ξ_2	0.1066	0.0752	0.1069	0.0753	0.1043	0.075	0.1066	0.0747
	ξ_3	0.1049	0.0749	0.1044	0.0744	0.1052	0.0752	0.1051	0.0752
	ξ_4	0.0756	0.032	0.0726	0.0601	0.0841	0.075	0.0771	0.0581
	ξ_5	0.0718	0.0606	0.0935	0.0755	0.1142	0.0744	0.0946	0.0742
	ξ_6	0.0886	0.0752	0.1243	0.0736	0.1393	0.0723	0.1234	0.0738

Table A.18: Average root mean squared error for latent variable, for Study VI

Method	Latent Variable	Path coef. 0.5		Path coef. 0.7		Path coef. 0.9		Random Path coef.	
		$RMSE_L$	$RMSE_U$	$RMSE_L$	$RMSE_U$	$RMSE_L$	$RMSE_U$	$RMSE_L$	$RMSE_U$
Center	ξ_1	0.6224	0.6227	0.6241	0.6245	0.624	0.6242	0.6228	0.6232
	ξ_2	0.6206	0.6212	0.624	0.6245	0.6203	0.6205	0.6251	0.6253
	ξ_3	0.6238	0.6239	0.6223	0.6228	0.6207	0.621	0.6236	0.6238
	ξ_4	0.0674	0.0674	0.122	0.1222	0.2203	0.2205	0.1361	0.1362
	ξ_5	0.1202	0.1203	0.2497	0.2498	0.343	0.3432	0.2505	0.2506
	ξ_6	0.222	0.222	0.3915	0.3916	0.5034	0.5034	0.3906	0.3906
Symbolic Cov.	ξ_1	0.6283	0.6285	0.63	0.6306	0.6292	0.6298	0.6279	0.6283
	ξ_2	0.6302	0.6303	0.6314	0.6317	0.6272	0.6276	0.6316	0.6317
	ξ_3	0.6274	0.6276	0.6267	0.6267	0.6297	0.6297	0.628	0.6281
	ξ_4	0.065	0.065	0.1328	0.1328	0.2279	0.228	0.1404	0.1404
	ξ_5	0.1254	0.1254	0.2585	0.2587	0.3543	0.3543	0.2600	0.2600
	ξ_6	0.2332	0.2333	0.404	0.404	0.5116	0.5117	0.4016	0.4017

Table A.19: Absolute relative bias of structural model of Study VII

Method	Estimator	Path coef. 0.5		Path coef. 0.7		Path coef. 0.9		Random Path coef.	
		Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
Center	$\hat{\beta}_{1,4}$	0.1136	0.0416	0.0555	0.0332	0.0324	0.0245	0.0667	0.0483
	$\hat{\beta}_{1,5}$	0.1008	0.0808	0.1122	0.0859	0.1336	0.1001	0.127	0.1034
	$\hat{\beta}_{2,5}$	0.0323	0.0243	0.0306	0.0226	0.0347	0.0243	0.0299	0.023
	$\hat{\beta}_{4,5}$	0.4379	0.1341	0.2554	0.1276	0.138	0.1017	0.2964	0.1657
	$\hat{\beta}_{2,6}$	0.0285	0.0222	0.0325	0.0227	0.039	0.0262	0.0324	0.0241
	$\hat{\beta}_{3,6}$	0.03	0.0223	0.0331	0.0239	0.0387	0.0254	0.0322	0.0227
	$\hat{\beta}_{5,6}$	0.0818	0.0464	0.3396	0.0736	0.8302	0.0954	0.3524	0.1984
Symbolic Cov.	$\hat{\beta}_{1,4}$	0.2011	0.0169	0.0795	0.0156	0.026	0.0133	0.1007	0.0737
	$\hat{\beta}_{1,5}$	1.0028	1.1281	1.29	1.6821	1.7046	2.0794	1.3631	1.837
	$\hat{\beta}_{2,5}$	0.0531	0.0425	0.0895	0.0561	0.1051	0.0586	0.0852	0.0547
	$\hat{\beta}_{4,5}$	1.4847	1.1691	1.5196	1.738	1.7704	2.1014	1.6451	1.8556
	$\hat{\beta}_{2,6}$	0.063	0.0465	0.0843	0.0608	0.0981	0.0673	0.0852	0.062
	$\hat{\beta}_{3,6}$	0.0592	0.043	0.0929	0.0581	0.1179	0.0692	0.094	0.0599
	$\hat{\beta}_{5,6}$	0.1219	0.0727	0.2845	0.0968	0.7612	0.1119	0.3071	0.1955

Table A.20: Average absolute relative bias of the loadings associated with each latent variable, for Study VII

Method	Latent Variable	Path coef. 0.5		Path coef. 0.7		Path coef. 0.9		Random Path coef.	
		Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
Center	ξ_1	0.0988	0.0287	0.0997	0.0292	0.0992	0.029	0.0991	0.0279
	ξ_2	0.0989	0.0275	0.0994	0.0283	0.0995	0.0287	0.0999	0.029
	ξ_3	0.0983	0.0284	0.0981	0.0288	0.1000	0.0292	0.0998	0.0277
	ξ_4	0.0264	0.0187	0.0446	0.0255	0.0768	0.0258	0.049	0.0325
	ξ_5	0.0792	0.027	0.1162	0.0271	0.1305	0.0257	0.1137	0.0272
	ξ_6	0.0882	0.0263	0.1246	0.0254	0.1382	0.0255	0.123	0.0267
Symbolic Cov.	ξ_1	0.1046	0.0746	0.1042	0.0746	0.1048	0.075	0.1066	0.0755
	ξ_2	0.1055	0.075	0.1076	0.0753	0.1056	0.0752	0.1064	0.0749
	ξ_3	0.105	0.075	0.1071	0.0753	0.1049	0.076	0.1064	0.0753
	ξ_4	0.0747	0.0327	0.0718	0.0598	0.0857	0.0753	0.0775	0.0593
	ξ_5	0.0844	0.0748	0.1213	0.0739	0.1356	0.0727	0.1187	0.0745
	ξ_6	0.0935	0.0756	0.1305	0.0731	0.1461	0.073	0.1292	0.0734

Table A.21: Average root mean squared error for latent variable, for Study VII

Method	Latent Variable	Path coef. 0.5		Path coef. 0.7		Path coef. 0.9		Random Path coef.	
		$RMSE_L$	$RMSE_U$	$RMSE_L$	$RMSE_U$	$RMSE_L$	$RMSE_U$	$RMSE_L$	$RMSE_U$
Center	ξ_1	0.6212	0.6213	0.6239	0.6243	0.6223	0.623	0.622	0.6221
	ξ_2	0.6216	0.6216	0.6221	0.6225	0.624	0.6243	0.6214	0.6216
	ξ_3	0.6207	0.6209	0.6216	0.6217	0.624	0.6241	0.6221	0.6224
	ξ_4	0.0669	0.067	0.1202	0.1205	0.2177	0.2179	0.1337	0.1338
	ξ_5	0.2097	0.2098	0.3762	0.3762	0.4777	0.4778	0.3694	0.3695
	ξ_6	0.2434	0.2434	0.4313	0.4314	0.5482	0.5483	0.4266	0.4267
Symbolic Cov.	ξ_1	0.6269	0.6271	0.6278	0.6278	0.6301	0.6305	0.632	0.6321
	ξ_2	0.6285	0.629	0.6323	0.6325	0.6302	0.6303	0.6283	0.6283
	ξ_3	0.6295	0.6297	0.6307	0.6310	0.6304	0.6311	0.6287	0.6289
	ξ_4	0.0637	0.0638	0.1314	0.1314	0.2313	0.2314	0.1434	0.1434
	ξ_5	0.216	0.2161	0.3830	0.3830	0.4847	0.4849	0.3761	0.3762
	ξ_6	0.2524	0.2525	0.4402	0.4402	0.5614	0.5615	0.4389	0.439

Table A.22: Absolute relative bias of structural model of Study VIII

Method	Estimator	Path coef. 0.5		Path coef. 0.7		Path coef. 0.9		Random Path coef.	
		Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
Center	$\hat{\beta}_{2,3}$	0.1152	0.0407	0.0549	0.0324	0.032	0.0235	0.0676	0.0497
	$\hat{\beta}_{1,4}$	0.0555	0.0337	0.0298	0.023	0.0275	0.0202	0.0319	0.0253
	$\hat{\beta}_{2,4}$	0.0549	0.0348	0.029	0.0218	0.0277	0.0203	0.0323	0.0244
	$\hat{\beta}_{2,5}$	0.1139	0.0416	0.0552	0.0328	0.0324	0.0235	0.0679	0.05
	$\hat{\beta}_{1,6}$	0.0959	0.0409	0.0384	0.0278	0.0262	0.0206	0.0439	0.0345
	$\hat{\beta}_{3,6}$	0.4012	0.0265	0.2518	0.0288	0.0907	0.0317	0.2503	0.1469
Symbolic Cov.	$\hat{\beta}_{2,3}$	0.2012	0.0165	0.0796	0.0154	0.0269	0.0128	0.1016	0.0736
	$\hat{\beta}_{1,4}$	0.0437	0.032	0.0497	0.0374	0.0706	0.0391	0.0502	0.0374
	$\hat{\beta}_{2,4}$	0.0435	0.0307	0.0507	0.036	0.0671	0.0396	0.0536	0.0386
	$\hat{\beta}_{2,5}$	0.2007	0.0161	0.08	0.0152	0.0268	0.0129	0.1037	0.0745
	$\hat{\beta}_{1,6}$	0.1091	0.0366	0.0302	0.0234	0.0632	0.0364	0.0433	0.0352
	$\hat{\beta}_{3,6}$	0.466	0.0465	0.3178	0.0417	0.1548	0.0422	0.3123	0.1644

Table A.23: Average absolute relative bias of the loadings associated with each latent variable, for Study VIII

Method	Latent Variable	Path coef. 0.5		Path coef. 0.7		Path coef. 0.9		Random Path coef.	
		Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
Center	ξ_1	0.0997	0.0283	0.0988	0.0293	0.1000	0.0289	0.1003	0.0277
	ξ_2	0.099	0.0286	0.0997	0.0285	0.0989	0.029	0.1001	0.0295
	ξ_3	0.0262	0.0176	0.0452	0.0266	0.079	0.0255	0.049	0.0312
	ξ_4	0.0465	0.0274	0.0866	0.0258	0.108	0.0254	0.0851	0.0311
	ξ_5	0.025	0.0175	0.0458	0.0262	0.0788	0.0268	0.0504	0.0319
	ξ_6	0.0227	0.018	0.0667	0.0264	0.1035	0.0258	0.0675	0.0344
Symbolic Cov.	ξ_1	0.1073	0.0762	0.1066	0.076	0.1063	0.0749	0.1055	0.0745
	ξ_2	0.104	0.0739	0.1071	0.075	0.1051	0.0746	0.1064	0.0754
	ξ_3	0.075	0.0324	0.0709	0.0594	0.0835	0.0752	0.0782	0.0594
	ξ_4	0.0727	0.0606	0.0942	0.076	0.1145	0.0743	0.0944	0.0738
	ξ_5	0.0755	0.0324	0.0693	0.0588	0.0851	0.0748	0.0771	0.0581
	ξ_6	0.0673	0.0434	0.0799	0.072	0.1092	0.0748	0.0834	0.0703

Table A.24: Average root mean squared error for latent variable, for Study VIII

Method	Latent Variable	Path coef. 0.5		Path coef. 0.7		Path coef. 0.9		Random Path coef.	
		$RMSE_L$	$RMSE_U$	$RMSE_L$	$RMSE_U$	$RMSE_L$	$RMSE_U$	$RMSE_L$	$RMSE_U$
Center	ξ_1	0.6232	0.6232	0.6212	0.6219	0.6222	0.6225	0.6245	0.6251
	ξ_2	0.6247	0.6251	0.6225	0.6225	0.6222	0.6223	0.6241	0.6245
	ξ_3	0.0672	0.0675	0.121	0.121	0.2224	0.2226	0.1349	0.135
	ξ_4	0.1163	0.1164	0.2477	0.2477	0.3453	0.3453	0.2489	0.249
	ξ_5	0.0656	0.0658	0.1216	0.1216	0.2219	0.222	0.1374	0.1375
	ξ_6	0.062	0.0622	0.176	0.1761	0.3158	0.3158	0.1829	0.183
Symbolic Cov.	ξ_1	0.6287	0.6291	0.6305	0.6309	0.6294	0.6298	0.6293	0.6299
	ξ_2	0.627	0.627	0.6339	0.6341	0.6285	0.6288	0.6314	0.6315
	ξ_3	0.0657	0.0658	0.1281	0.1281	0.2258	0.2259	0.1434	0.1435
	ξ_4	0.13	0.1301	0.2607	0.2608	0.3553	0.3554	0.2592	0.2592
	ξ_5	0.0649	0.065	0.1261	0.1261	0.2306	0.2307	0.1386	0.1386
	ξ_6	0.0643	0.0644	0.1918	0.1918	0.3241	0.3242	0.1973	0.1973

Table A.25: Absolute relative bias of structural model of Study IX

Method	Estimator	Path coef. 0.5		Path coef. 0.7		Path coef. 0.9		Random Path coef.	
		Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
Center	$\hat{\beta}_{2,3}$	0.1143	0.0397	0.0565	0.0338	0.0316	0.0233	0.0693	0.048
	$\hat{\beta}_{1,4}$	0.0555	0.035	0.0306	0.0232	0.0275	0.0204	0.0321	0.0251
	$\hat{\beta}_{2,4}$	0.0543	0.0326	0.029	0.022	0.0271	0.0199	0.0327	0.0259
	$\hat{\beta}_{2,5}$	0.1169	0.0406	0.0536	0.0321	0.0318	0.0244	0.0688	0.05
	$\hat{\beta}_{1,6}$	0.0344	0.0265	0.0291	0.0214	0.0353	0.0244	0.0285	0.0218
	$\hat{\beta}_{2,6}$	0.1001	0.0763	0.1116	0.0841	0.1305	0.0965	0.1251	0.0958
	$\hat{\beta}_{3,6}$	0.4311	0.1366	0.2577	0.1281	0.1328	0.0983	0.2847	0.1708
Symbolic Cov.	$\hat{\beta}_{2,3}$	0.2007	0.0172	0.0796	0.015	0.0269	0.0132	0.0991	0.0731
	$\hat{\beta}_{1,4}$	0.0442	0.0319	0.0475	0.0348	0.0677	0.038	0.054	0.0387
	$\hat{\beta}_{2,4}$	0.0419	0.032	0.0515	0.0358	0.0683	0.0395	0.0529	0.0376
	$\hat{\beta}_{2,5}$	0.2003	0.0169	0.0795	0.0149	0.0267	0.0129	0.1035	0.0761
	$\hat{\beta}_{1,6}$	0.0558	0.0417	0.0866	0.0548	0.1046	0.0566	0.0876	0.0576
	$\hat{\beta}_{2,6}$	1.0219	1.1947	1.3588	1.7485	1.7364	2.1617	1.4538	1.8979
	$\hat{\beta}_{3,6}$	1.5075	1.2354	1.61	1.7923	1.7913	2.169	1.7135	1.9608

Table A.26: Average absolute relative bias of the loadings associated with each latent variable, for Study IX

Method	Latent Variable	Path coef. 0.5		Path coef. 0.7		Path coef. 0.9		Random Path coef.	
		Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
Center	ξ_1	0.1008	0.0286	0.1004	0.0285	0.0989	0.028	0.1001	0.0275
	ξ_2	0.0992	0.0271	0.0994	0.0275	0.0987	0.0281	0.0987	0.0284
	ξ_3	0.025	0.0178	0.044	0.0269	0.0773	0.0262	0.0492	0.0319
	ξ_4	0.0468	0.0275	0.0856	0.0256	0.1071	0.0264	0.0852	0.0324
	ξ_5	0.0267	0.0185	0.046	0.0249	0.0782	0.0271	0.0479	0.0317
	ξ_6	0.0776	0.0282	0.1152	0.0262	0.1296	0.0251	0.1128	0.028
Symbolic Cov.	ξ_1	0.1049	0.0748	0.1052	0.0752	0.1044	0.0742	0.1048	0.0747
	ξ_2	0.1054	0.0746	0.1045	0.0747	0.1064	0.0751	0.1054	0.075
	ξ_3	0.0753	0.0325	0.0716	0.0597	0.0846	0.0745	0.0762	0.0588
	ξ_4	0.0721	0.0605	0.0936	0.0754	0.1128	0.074	0.0939	0.0743
	ξ_5	0.0746	0.0323	0.0712	0.0596	0.0839	0.0749	0.078	0.0585
	ξ_6	0.0834	0.0741	0.1213	0.0738	0.1368	0.0726	0.1196	0.0745

Table A.27: Average root mean squared error for latent variable, for Study IX

Method	Latent Variable	Path coef. 0.5		Path coef. 0.7		Path coef. 0.9		Random Path coef.	
		$RMSE_L$	$RMSE_U$	$RMSE_L$	$RMSE_U$	$RMSE_L$	$RMSE_U$	$RMSE_L$	$RMSE_U$
Center	ξ_1	0.6241	0.6243	0.6234	0.6241	0.6224	0.6226	0.6244	0.6245
	ξ_2	0.6231	0.6234	0.624	0.6242	0.6233	0.6234	0.622	0.6221
	ξ_3	0.0656	0.0658	0.1187	0.1188	0.219	0.2191	0.1344	0.1345
	ξ_4	0.1176	0.1176	0.2453	0.2454	0.3428	0.3428	0.2494	0.2494
	ξ_5	0.0678	0.068	0.1222	0.1223	0.2206	0.2206	0.1317	0.1317
	ξ_6	0.2065	0.2066	0.3731	0.3731	0.4755	0.4755	0.3671	0.3671
Symbolic Cov.	ξ_1	0.6289	0.629	0.6284	0.6284	0.6261	0.6263	0.6269	0.627
	ξ_2	0.6304	0.6309	0.6287	0.629	0.6308	0.6313	0.6307	0.6311
	ξ_3	0.0665	0.0665	0.129	0.129	0.229	0.2291	0.1415	0.1415
	ξ_4	0.1283	0.1283	0.2595	0.2595	0.3522	0.3523	0.2601	0.2602
	ξ_5	0.0667	0.0667	0.129	0.129	0.2273	0.2274	0.1422	0.1423
	ξ_6	0.2146	0.2146	0.3836	0.3837	0.4888	0.4889	0.3796	0.3797

Table A.28: Absolute relative bias of structural model of Study X

Method	Estimator	Path coef. 0.5		Path coef. 0.7		Path coef. 0.9		Random Path coef.	
		Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
Center	$\hat{\beta}_{2,3}$	0.1167	0.041	0.056	0.033	0.0324	0.024	0.0686	0.0489
	$\hat{\beta}_{1,4}$	0.0534	0.0328	0.0293	0.0231	0.0259	0.0202	0.0321	0.0259
	$\hat{\beta}_{2,4}$	0.0533	0.0342	0.0305	0.0222	0.027	0.0201	0.033	0.0257
	$\hat{\beta}_{2,5}$	0.1168	0.0425	0.0547	0.0321	0.0318	0.0239	0.0657	0.0474
	$\hat{\beta}_{1,6}$	0.0278	0.0205	0.0337	0.0248	0.0403	0.0265	0.0329	0.0234
	$\hat{\beta}_{2,6}$	0.1344	0.1042	0.1682	0.1231	0.1943	0.1474	0.18	0.1411
	$\hat{\beta}_{3,6}$	0.4427	0.1375	0.2719	0.1385	0.1578	0.1194	0.2988	0.1704
	$\hat{\beta}_{5,6}$	0.443	0.1354	0.2735	0.1492	0.1574	0.1113	0.2942	0.1816
Symbolic Cov.	$\hat{\beta}_{2,3}$	0.2016	0.0169	0.0804	0.0154	0.0266	0.0133	0.1005	0.0731
	$\hat{\beta}_{1,4}$	0.0427	0.0318	0.0508	0.0365	0.0682	0.04	0.0525	0.0389
	$\hat{\beta}_{2,4}$	0.042	0.0316	0.0479	0.0342	0.0691	0.0387	0.0506	0.0365
	$\hat{\beta}_{2,5}$	0.2013	0.0174	0.0804	0.0152	0.0263	0.0132	0.1043	0.0761
	$\hat{\beta}_{1,6}$	0.0806	0.0561	0.1136	0.0678	0.1242	0.076	0.1144	0.069
	$\hat{\beta}_{2,6}$	1.8472	1.6804	2.3775	2.3095	2.7206	2.6979	2.4388	2.5344
	$\hat{\beta}_{3,6}$	1.5908	1.5641	1.9841	2.2724	2.699	2.8488	2.1784	2.5083
	$\hat{\beta}_{5,6}$	1.602	1.5661	2.0621	2.3063	2.5912	2.8315	2.084	2.4659

Table A.29: Average absolute relative bias of the loadings associated with each latent variable, for Study X

Method	Latent Variable	Path coef. 0.5		Path coef. 0.7		Path coef. 0.9		Random Path coef.	
		Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
Center	ξ_1	0.0994	0.0274	0.0982	0.0284	0.0997	0.0281	0.1002	0.0284
	ξ_2	0.0998	0.0293	0.1000	0.0284	0.0996	0.0281	0.0992	0.0279
	ξ_3	0.0265	0.0178	0.0451	0.0264	0.078	0.0266	0.0499	0.0328
	ξ_4	0.0480	0.0276	0.0876	0.027	0.108	0.0258	0.085	0.0319
	ξ_5	0.0264	0.0183	0.0463	0.0262	0.0796	0.0259	0.0497	0.0306
	ξ_6	0.1014	0.0262	0.1278	0.0249	0.1407	0.0254	0.1274	0.025
Symbolic Cov.	ξ_1	0.1048	0.075	0.1048	0.0748	0.1049	0.075	0.1044	0.0749
	ξ_2	0.1051	0.0753	0.107	0.0754	0.1062	0.0746	0.1066	0.0753
	ξ_3	0.0754	0.0323	0.0715	0.0598	0.0843	0.0751	0.0776	0.059
	ξ_4	0.0724	0.0608	0.0946	0.0754	0.1127	0.0739	0.0948	0.0745
	ξ_5	0.0753	0.0324	0.0714	0.0596	0.0851	0.0752	0.0782	0.0586
	ξ_6	0.1054	0.0748	0.1356	0.0732	0.1473	0.0725	0.1328	0.0731

Table A.30: Average root mean squared error for latent variable, for Study X

Method	Latent Variable	Path coef. 0.5		Path coef. 0.7		Path coef. 0.9		Random Path coef.	
		$RMSE_L$	$RMSE_U$	$RMSE_L$	$RMSE_U$	$RMSE_L$	$RMSE_U$	$RMSE_L$	$RMSE_U$
Center	ξ_1	0.6250	0.6252	0.621	0.6211	0.6227	0.6229	0.6236	0.6241
	ξ_2	0.6229	0.6231	0.6226	0.6233	0.6235	0.6239	0.6221	0.6227
	ξ_3	0.0676	0.0677	0.1201	0.1203	0.2206	0.2206	0.1358	0.1358
	ξ_4	0.12	0.12	0.2493	0.2495	0.3445	0.3446	0.2483	0.2483
	ξ_5	0.0679	0.0679	0.1222	0.1224	0.2236	0.2237	0.1355	0.1357
	ξ_6	0.2976	0.2977	0.4611	0.4612	0.564	0.564	0.4567	0.4567
Symbolic Cov.	ξ_1	0.6264	0.6265	0.6278	0.6281	0.627	0.6271	0.6279	0.6287
	ξ_2	0.6293	0.6294	0.632	0.632	0.6304	0.6304	0.6319	0.6322
	ξ_3	0.0655	0.0657	0.13	0.1301	0.2285	0.2286	0.1413	0.1413
	ξ_4	0.128	0.1282	0.2608	0.2609	0.3522	0.3522	0.2576	0.2578
	ξ_5	0.065	0.0651	0.1305	0.1306	0.2294	0.2295	0.1425	0.1426
	ξ_6	0.3036	0.3036	0.4743	0.4744	0.5745	0.5745	0.4655	0.4656

A.2 Simulation 2 - Number of Manifest Variables

Table A.31: Average absolute relative bias for studies III-VI with 1 manifest variable

Study	Method	Estimator	Loadings 0.5		Loadings 0.7		Loadings 0.9		Random Loadings	
			Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
III	Center	$\hat{\beta}_{1,3}$	0.0156	0.0131	0.0179	0.0158	0.0204	0.0175	0.0789	0.0634
		$\hat{\beta}_{2,3}$	0.0155	0.0132	0.0178	0.0154	0.0203	0.0173	0.0805	0.0638
		$\hat{\beta}_{1,4}$	0.0157	0.0129	0.0186	0.0164	0.0205	0.0179	0.0799	0.0627
		$\hat{\beta}_{2,4}$	0.0155	0.0134	0.0182	0.0163	0.0197	0.0174	0.08	0.0641
	Symbolic Cov.	$\hat{\beta}_{1,3}$	0.0808	0.0528	0.0477	0.0365	0.0373	0.0285	0.0569	0.0417
		$\hat{\beta}_{2,3}$	0.0799	0.0521	0.0519	0.0377	0.0392	0.0303	0.0573	0.0439
		$\hat{\beta}_{1,4}$	0.0778	0.0514	0.0486	0.035	0.0385	0.0307	0.0577	0.0436
IV	Center	$\hat{\beta}_{1,3}$	0.0147	0.0128	0.0184	0.0161	0.0206	0.0185	0.0821	0.0659
		$\hat{\beta}_{2,3}$	0.0146	0.0124	0.0185	0.0164	0.0203	0.0176	0.0813	0.0651
		$\hat{\beta}_{1,4}$	0.1836	0.1327	0.1771	0.129	0.1814	0.126	0.2338	0.1786
		$\hat{\beta}_{2,4}$	0.1811	0.1294	0.1775	0.1279	0.177	0.1247	0.2303	0.1743
		$\hat{\beta}_{3,4}$	0.2822	0.24	0.2704	0.2303	0.274	0.2361	0.3233	0.2706
	Symbolic Cov.	$\hat{\beta}_{1,3}$	0.0826	0.0531	0.0515	0.0366	0.0387	0.0293	0.0575	0.0427
		$\hat{\beta}_{2,3}$	0.081	0.0551	0.0487	0.0361	0.0381	0.0306	0.0559	0.0421
V	Center	$\hat{\beta}_{1,4}$	0.0487	0.0318	0.06	0.0408	0.0685	0.0461	0.1085	0.0829
		$\hat{\beta}_{1,5}$	0.0157	0.0136	0.0182	0.0153	0.0195	0.017	0.0799	0.0645
		$\hat{\beta}_{2,5}$	0.0151	0.0129	0.0185	0.0157	0.0198	0.0173	0.0819	0.0644
		$\hat{\beta}_{2,6}$	0.0149	0.0127	0.0187	0.0166	0.0213	0.0184	0.08	0.0624
		$\hat{\beta}_{3,6}$	0.0151	0.0128	0.0187	0.0166	0.0214	0.019	0.079	0.0623
	Symbolic Cov.	$\hat{\beta}_{1,4}$	0.0868	0.0714	0.0987	0.0729	0.102	0.0759	0.0928	0.075
		$\hat{\beta}_{1,5}$	0.0784	0.0513	0.0487	0.0366	0.0365	0.0277	0.0568	0.0413
VI	Center	$\hat{\beta}_{2,5}$	0.0816	0.054	0.0507	0.0361	0.0382	0.0295	0.0529	0.0402
		$\hat{\beta}_{2,6}$	0.0793	0.0517	0.0489	0.0371	0.039	0.0303	0.0579	0.0423
		$\hat{\beta}_{3,6}$	0.0785	0.0517	0.0492	0.0348	0.0388	0.0297	0.0577	0.0416
		$\hat{\beta}_{1,4}$	0.0472	0.0312	0.064	0.0406	0.0689	0.0457	0.1078	0.0793
		$\hat{\beta}_{1,5}$	0.0152	0.0132	0.0189	0.0161	0.0205	0.0182	0.0803	0.0646
		$\hat{\beta}_{2,5}$	0.0152	0.0127	0.0189	0.0161	0.0205	0.0177	0.0806	0.0621
	Symbolic Cov.	$\hat{\beta}_{2,6}$	0.0215	0.0162	0.0205	0.0149	0.0225	0.0154	0.0788	0.057
Symbolic Cov.	$\hat{\beta}_{3,6}$	0.0188	0.0137	0.0196	0.0133	0.0207	0.014	0.0791	0.0582	
	$\hat{\beta}_{5,6}$	0.126	0.0882	0.1261	0.092	0.1272	0.0936	0.1486	0.1038	
	$\hat{\beta}_{1,4}$	0.0889	0.072	0.103	0.0756	0.1004	0.0751	0.0972	0.0745	
	$\hat{\beta}_{1,5}$	0.0834	0.0534	0.0493	0.037	0.0377	0.0285	0.0572	0.043	
	$\hat{\beta}_{2,5}$	0.0798	0.054	0.0503	0.0375	0.0364	0.028	0.0546	0.0414	
	$\hat{\beta}_{2,6}$	0.1142	0.0898	0.0855	0.067	0.0726	0.0572	0.1022	0.08	
Symbolic Cov.	$\hat{\beta}_{3,6}$	0.1278	0.0726	0.0854	0.0552	0.0672	0.0431	0.0965	0.0618	
	$\hat{\beta}_{5,6}$	0.1594	0.1183	0.1475	0.114	0.1409	0.1029	0.1612	0.1178	

Table A.32: Average absolute relative bias of the loadings associated with each latent variable, study with 1 manifest variable

Study	Method	Latent Variable	Loadings 0.5		Loadings 0.7		Loadings 0.9		Random Loadings	
			Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
III	Center	ξ_1	0.2182	0.0121	0.0906	0.0126	0.0338	0.0123	0.1102	0.0777
		ξ_2	0.2187	0.0118	0.0902	0.0128	0.0334	0.0125	0.1157	0.0778
		ξ_3	0.2064	0.0176	0.0773	0.0188	0.0233	0.0126	0.1013	0.0787
		ξ_4	0.2062	0.0176	0.0762	0.0192	0.024	0.0123	0.101	0.0781
	Symbolic Cov.	ξ_1	0.2189	0.0116	0.0903	0.0123	0.0337	0.0123	0.1152	0.0786
		ξ_2	0.219	0.0118	0.09	0.0123	0.0329	0.0125	0.1141	0.0792
		ξ_3	0.2062	0.0171	0.0759	0.0185	0.0235	0.0126	0.1033	0.0791
		ξ_4	0.2067	0.0179	0.0769	0.0188	0.0243	0.0127	0.0996	0.0789
IV	Center	ξ_1	0.219	0.0114	0.0905	0.0126	0.0332	0.0126	0.1173	0.0811
		ξ_2	0.2193	0.0115	0.0902	0.0128	0.0336	0.0125	0.1172	0.079
		ξ_3	0.208	0.0164	0.0765	0.0193	0.0237	0.0125	0.0979	0.0779
		ξ_4	0.2432	0.0075	0.1162	0.0085	0.0598	0.0084	0.1376	0.0757
	Symbolic Cov.	ξ_1	0.2182	0.0116	0.09	0.0127	0.0334	0.0123	0.1141	0.0789
		ξ_2	0.2187	0.0117	0.0902	0.0131	0.0327	0.0129	0.1138	0.0787
		ξ_3	0.2067	0.0167	0.0755	0.0197	0.0241	0.0124	0.1044	0.0794
		ξ_4	0.2425	0.0079	0.116	0.0084	0.0595	0.0089	0.137	0.0772
V	Center	ξ_1	0.2189	0.0117	0.0902	0.0129	0.0328	0.0126	0.1176	0.0802
		ξ_2	0.2189	0.0114	0.0906	0.0124	0.0332	0.0125	0.1126	0.078
		ξ_3	0.2187	0.0113	0.0906	0.0121	0.0333	0.0127	0.1086	0.0776
		ξ_4	0.1608	0.0346	0.0434	0.0198	0.0391	0.0335	0.0836	0.0657
		ξ_5	0.2066	0.0172	0.0763	0.0188	0.0242	0.0126	0.1015	0.078
		ξ_6	0.2067	0.0166	0.0759	0.0191	0.0233	0.0127	0.1011	0.0773
	Symbolic Cov.	ξ_1	0.2195	0.0113	0.0898	0.0126	0.0331	0.0121	0.1146	0.0784
		ξ_2	0.2194	0.0114	0.0903	0.0136	0.0327	0.0127	0.1125	0.0802
		ξ_3	0.2192	0.0115	0.0908	0.0128	0.0328	0.0127	0.117	0.0799
		ξ_4	0.1648	0.0345	0.0419	0.0195	0.0402	0.034	0.0815	0.0648
		ξ_5	0.2067	0.0173	0.0763	0.0189	0.0238	0.0126	0.1000	0.0775
		ξ_6	0.2067	0.0172	0.0765	0.019	0.0234	0.0123	0.1022	0.0781
VI	Center	ξ_1	0.2188	0.0119	0.0905	0.0127	0.0336	0.0123	0.1177	0.0792
		ξ_2	0.2181	0.0113	0.0908	0.0129	0.0326	0.0127	0.1137	0.0774
		ξ_3	0.2184	0.0118	0.0906	0.0126	0.0333	0.0128	0.1112	0.0778
		ξ_4	0.1624	0.0344	0.0417	0.0198	0.0397	0.0334	0.0829	0.0659
		ξ_5	0.207	0.0168	0.0765	0.0191	0.0239	0.0123	0.1071	0.0776
		ξ_6	0.2349	0.0093	0.1083	0.0102	0.0515	0.0103	0.1324	0.0781
	Symbolic Cov.	ξ_1	0.2187	0.0119	0.0904	0.0124	0.0332	0.0121	0.1119	0.0782
		ξ_2	0.2189	0.0117	0.0897	0.0126	0.0328	0.0127	0.1166	0.0794
		ξ_3	0.2189	0.0113	0.0899	0.0126	0.0329	0.0127	0.1119	0.0779
		ξ_4	0.163	0.0341	0.0422	0.02	0.0393	0.0335	0.085	0.0654
		ξ_5	0.2069	0.0173	0.0763	0.0186	0.0233	0.0123	0.1056	0.0791
		ξ_6	0.2357	0.0094	0.1082	0.0101	0.0517	0.0102	0.1286	0.0772

Table A.33: Average root mean squared error for latent variable, study with 1 manifest variable

Study	Method	Latent Variable	Loadings 0.5		Loadings 0.7		Loadings 0.9		Random Loadings	
			$RMSE_L$	$RMSE_U$	$RMSE_L$	$RMSE_U$	$RMSE_L$	$RMSE_U$	$RMSE_L$	$RMSE_U$
III	Center	ξ_1	0.8043	0.8044	0.6076	0.6084	0.4901	0.4903	0.6259	0.6262
		ξ_2	0.8020	0.8027	0.6091	0.6095	0.4893	0.4893	0.6368	0.6372
		ξ_3	0.4551	0.4552	0.2374	0.2374	0.1105	0.1107	0.2666	0.2667
		ξ_4	0.4529	0.4531	0.2328	0.2329	0.1124	0.1125	0.2627	0.2628
	Symbolic Cov.	ξ_1	0.8032	0.8036	0.6092	0.6093	0.4908	0.491	0.6337	0.6341
		ξ_2	0.8036	0.8037	0.6102	0.6107	0.4923	0.4923	0.6323	0.6327
		ξ_3	0.4536	0.4536	0.2313	0.2314	0.1109	0.1111	0.2673	0.2674
		ξ_4	0.4553	0.4555	0.2359	0.236	0.1136	0.1136	0.2624	0.2626
IV	Center	ξ_1	0.8034	0.8035	0.6089	0.6091	0.4901	0.4901	0.6399	0.6401
		ξ_2	0.803	0.804	0.6089	0.6095	0.4909	0.4912	0.6387	0.6394
		ξ_3	0.4606	0.4608	0.2354	0.2356	0.1113	0.1114	0.2585	0.2586
		ξ_4	0.6607	0.6609	0.4246	0.4248	0.2688	0.2688	0.4468	0.4469
	Symbolic Cov.	ξ_1	0.8022	0.8022	0.6098	0.61	0.4907	0.491	0.6354	0.6357
		ξ_2	0.8037	0.8042	0.6102	0.6103	0.4918	0.492	0.6345	0.6347
		ξ_3	0.4545	0.4549	0.2323	0.2324	0.1124	0.1126	0.2722	0.2723
		ξ_4	0.6568	0.6569	0.4232	0.4232	0.2666	0.2667	0.4428	0.4429
V	Center	ξ_1	0.8037	0.804	0.6091	0.6091	0.4901	0.4905	0.6387	0.6389
		ξ_2	0.8036	0.8038	0.6094	0.6096	0.4913	0.4915	0.6314	0.6322
		ξ_3	0.8039	0.8045	0.6083	0.6086	0.4904	0.4911	0.6267	0.6271
		ξ_4	0.2981	0.2981	0.1302	0.1303	0.1182	0.1184	0.187	0.187
		ξ_5	0.4542	0.4542	0.2341	0.2342	0.1138	0.1139	0.2681	0.2681
		ξ_6	0.4561	0.4564	0.2321	0.2322	0.1102	0.1103	0.266	0.2662
	Symbolic Cov.	ξ_1	0.8033	0.8037	0.609	0.6096	0.4899	0.4901	0.6337	0.6341
		ξ_2	0.8038	0.8041	0.6104	0.6106	0.4911	0.4912	0.6329	0.6337
		ξ_3	0.8024	0.8027	0.6103	0.6107	0.4913	0.4914	0.6403	0.6405
		ξ_4	0.3101	0.3101	0.1262	0.1262	0.1199	0.1201	0.1844	0.1844
		ξ_5	0.456	0.4562	0.2343	0.2343	0.1124	0.1125	0.2626	0.2626
		ξ_6	0.4541	0.4542	0.2347	0.2347	0.1111	0.1111	0.2673	0.2674
VI	Center	ξ_1	0.8035	0.8044	0.609	0.609	0.4899	0.4901	0.6394	0.6395
		ξ_2	0.8028	0.803	0.6094	0.6094	0.4911	0.4913	0.6353	0.6358
		ξ_3	0.8021	0.8023	0.6078	0.6081	0.4913	0.4914	0.629	0.6291
		ξ_4	0.303	0.3033	0.1245	0.1247	0.1186	0.1186	0.1855	0.1857
		ξ_5	0.4568	0.457	0.2347	0.2348	0.1122	0.1124	0.2763	0.2763
		ξ_6	0.6028	0.6028	0.3758	0.3758	0.2211	0.2212	0.402	0.4021
	Symbolic Cov.	ξ_1	0.8042	0.8052	0.609	0.6091	0.4909	0.4912	0.63	0.6304
		ξ_2	0.804	0.8048	0.6086	0.6089	0.4918	0.4919	0.637	0.6371
		ξ_3	0.8024	0.8024	0.6118	0.6122	0.4911	0.4911	0.6306	0.6308
		ξ_4	0.3049	0.3052	0.1262	0.1263	0.1185	0.1186	0.188	0.1881
		ξ_5	0.4568	0.4568	0.234	0.2341	0.1104	0.1104	0.2735	0.2736
		ξ_6	0.6066	0.6068	0.3748	0.3748	0.223	0.223	0.3942	0.3943

Table A.34: Average absolute relative bias for studies III-VI with 2 manifest variables

Study	Method	Estimator	Loadings 0.5		Loadings 0.7		Loadings 0.9		Random Loadings	
			Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
III	Center	$\hat{\beta}_{1,3}$	0.0147	0.0124	0.0187	0.0169	0.0201	0.018	0.0591	0.0453
		$\hat{\beta}_{2,3}$	0.0147	0.0127	0.0187	0.0164	0.0202	0.0177	0.0598	0.0461
		$\hat{\beta}_{1,4}$	0.0149	0.0126	0.0184	0.0165	0.021	0.0179	0.0585	0.0452
		$\hat{\beta}_{2,4}$	0.0152	0.013	0.0187	0.0168	0.0203	0.0186	0.0581	0.0455
	Symbolic Cov.	$\hat{\beta}_{1,3}$	0.0802	0.0509	0.0501	0.037	0.0376	0.0299	0.0553	0.0401
		$\hat{\beta}_{2,3}$	0.0817	0.051	0.0489	0.0371	0.0387	0.0299	0.0528	0.0398
		$\hat{\beta}_{1,4}$	0.0801	0.0538	0.0503	0.0375	0.0383	0.0305	0.0524	0.0385
		$\hat{\beta}_{2,4}$	0.08	0.0532	0.0487	0.0363	0.0367	0.0295	0.0546	0.0374
IV	Center	$\hat{\beta}_{1,3}$	0.0156	0.0124	0.0184	0.0159	0.0204	0.0187	0.0565	0.0445
		$\hat{\beta}_{2,3}$	0.0152	0.0126	0.0183	0.0163	0.0204	0.0179	0.0587	0.0458
		$\hat{\beta}_{1,4}$	0.1763	0.1274	0.1702	0.1213	0.1641	0.1192	0.1866	0.1387
		$\hat{\beta}_{2,4}$	0.1797	0.1339	0.1735	0.1273	0.1636	0.1156	0.1911	0.1411
		$\hat{\beta}_{3,4}$	0.2718	0.223	0.2627	0.2243	0.2491	0.2109	0.281	0.2318
	Symbolic Cov.	$\hat{\beta}_{1,3}$	0.0801	0.0526	0.0513	0.0372	0.0385	0.0295	0.0542	0.0382
		$\hat{\beta}_{2,3}$	0.0789	0.0493	0.0479	0.035	0.0389	0.0304	0.0523	0.0376
		$\hat{\beta}_{1,4}$	0.7286	0.191	0.6856	0.2014	0.6492	0.2093	0.6954	0.26
		$\hat{\beta}_{2,4}$	0.7173	0.1961	0.6831	0.2001	0.6538	0.2086	0.6947	0.2578
		$\hat{\beta}_{3,4}$	0.9699	0.4597	0.9072	0.4669	0.8686	0.4507	0.9288	0.5087
V	Center	$\hat{\beta}_{1,4}$	0.0487	0.0319	0.0646	0.0407	0.0685	0.0461	0.0853	0.0647
		$\hat{\beta}_{1,5}$	0.0154	0.0132	0.0183	0.0159	0.0209	0.0183	0.0575	0.0461
		$\hat{\beta}_{2,5}$	0.0154	0.0131	0.0181	0.0158	0.0207	0.0189	0.0556	0.0447
		$\hat{\beta}_{2,6}$	0.015	0.0128	0.019	0.0167	0.0209	0.0181	0.0607	0.048
		$\hat{\beta}_{3,6}$	0.0147	0.0129	0.0186	0.0163	0.0211	0.0183	0.0577	0.047
	Symbolic Cov.	$\hat{\beta}_{1,4}$	0.0932	0.0729	0.0985	0.0738	0.1018	0.0744	0.0996	0.0744
		$\hat{\beta}_{1,5}$	0.0831	0.0548	0.0501	0.037	0.0371	0.0271	0.0529	0.0407
		$\hat{\beta}_{2,5}$	0.0789	0.0548	0.0495	0.037	0.0368	0.0275	0.0564	0.0405
		$\hat{\beta}_{2,6}$	0.0805	0.0541	0.0504	0.0376	0.0388	0.03	0.0551	0.0404
		$\hat{\beta}_{3,6}$	0.0792	0.0545	0.0494	0.035	0.039	0.0305	0.0525	0.0408
VI	Center	$\hat{\beta}_{1,4}$	0.0478	0.0311	0.0632	0.0408	0.0698	0.0461	0.0841	0.0636
		$\hat{\beta}_{1,5}$	0.0149	0.0131	0.0195	0.0164	0.0204	0.0188	0.0574	0.0445
		$\hat{\beta}_{2,5}$	0.0153	0.0129	0.0194	0.0169	0.0203	0.0184	0.0563	0.0443
		$\hat{\beta}_{2,6}$	0.0178	0.0137	0.0198	0.0135	0.0201	0.0142	0.0557	0.0402
		$\hat{\beta}_{3,6}$	0.0166	0.0116	0.0184	0.0125	0.0207	0.0138	0.0547	0.0395
		$\hat{\beta}_{5,6}$	0.1247	0.0897	0.1278	0.0913	0.1284	0.0926	0.141	0.1011
	Symbolic Cov.	$\hat{\beta}_{1,4}$	0.0948	0.0732	0.0993	0.0746	0.1027	0.0745	0.0983	0.0745
		$\hat{\beta}_{1,5}$	0.0798	0.0529	0.0527	0.0383	0.0402	0.0306	0.053	0.0382
		$\hat{\beta}_{2,5}$	0.0812	0.0536	0.049	0.0383	0.0375	0.0283	0.0533	0.0407
		$\hat{\beta}_{2,6}$	0.1093	0.0807	0.0853	0.0659	0.0693	0.056	0.0945	0.073
		$\hat{\beta}_{3,6}$	0.126	0.0731	0.0855	0.0559	0.0659	0.0438	0.0916	0.0586
		$\hat{\beta}_{5,6}$	0.1651	0.1247	0.1448	0.107	0.1448	0.1025	0.1634	0.1163

Table A.35: Average absolute relative bias of the loadings associated with each latent variable, study with 2 manifest variable

Study	Method	Latent Variable	Loadings 0.5		Loadings 0.7		Loadings 0.9		Random Loadings	
			Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
III	Center	ξ_1	0.2185	0.012	0.0909	0.0121	0.0329	0.0126	0.1057	0.0543
		ξ_2	0.219	0.0118	0.0902	0.0128	0.0326	0.0126	0.1052	0.0548
		ξ_3	0.2071	0.0167	0.0757	0.0193	0.0239	0.0122	0.0913	0.056
		ξ_4	0.2067	0.0167	0.0765	0.0192	0.0233	0.0123	0.0947	0.0577
	Symbolic Cov.	ξ_1	0.2088	0.0159	0.0852	0.0142	0.0307	0.0127	0.1061	0.0755
		ξ_2	0.2086	0.0155	0.0852	0.0148	0.0309	0.0128	0.1075	0.076
		ξ_3	0.1969	0.0203	0.0709	0.0202	0.0225	0.0126	0.0966	0.0758
IV	Center	ξ_1	0.2181	0.0124	0.0897	0.0129	0.0331	0.0124	0.1047	0.0557
		ξ_2	0.2184	0.0121	0.0897	0.0132	0.0334	0.0126	0.1065	0.0548
		ξ_3	0.2056	0.017	0.0762	0.0186	0.0239	0.0126	0.0929	0.0549
		ξ_4	0.2418	0.0081	0.1155	0.0082	0.0594	0.0083	0.1316	0.0541
	Symbolic Cov.	ξ_1	0.2087	0.0156	0.0855	0.0138	0.0314	0.0123	0.1078	0.0766
		ξ_2	0.208	0.0155	0.0856	0.0141	0.0312	0.0129	0.1062	0.0749
		ξ_3	0.1966	0.0202	0.0714	0.0195	0.0227	0.0125	0.0988	0.0759
V	Center	ξ_4	0.2328	0.0127	0.1123	0.0098	0.0573	0.0094	0.1361	0.0746
		ξ_1	0.2192	0.012	0.0907	0.0132	0.0334	0.0127	0.1065	0.0558
		ξ_2	0.2187	0.0122	0.0902	0.0129	0.033	0.0124	0.109	0.0557
		ξ_3	0.2188	0.0119	0.089	0.0125	0.0329	0.0125	0.1062	0.0547
		ξ_4	0.1625	0.0345	0.0418	0.0197	0.0392	0.0335	0.0641	0.049
		ξ_5	0.2057	0.0175	0.0768	0.0188	0.0238	0.0126	0.0932	0.0552
	Symbolic Cov.	ξ_6	0.2072	0.0172	0.0754	0.0192	0.0233	0.0126	0.0907	0.0554
		ξ_1	0.208	0.016	0.0861	0.0139	0.0307	0.0128	0.108	0.0753
		ξ_2	0.2086	0.0154	0.0856	0.0142	0.031	0.0129	0.1087	0.0758
		ξ_3	0.2087	0.0156	0.0859	0.0144	0.0311	0.0129	0.1085	0.0759
		ξ_4	0.15	0.0378	0.0417	0.0178	0.0409	0.036	0.0785	0.0603
		ξ_5	0.1959	0.0207	0.0721	0.0197	0.0226	0.0125	0.0962	0.0746
		ξ_6	0.1954	0.0209	0.0714	0.0206	0.0227	0.0121	0.0968	0.0752
VI	Center	ξ_1	0.2179	0.0122	0.0896	0.0131	0.0327	0.0126	0.105	0.0545
		ξ_2	0.2184	0.0126	0.0901	0.0131	0.0338	0.0121	0.1023	0.0528
		ξ_3	0.2178	0.0116	0.0899	0.0127	0.0327	0.013	0.1048	0.053
		ξ_4	0.1622	0.0343	0.0423	0.0198	0.0405	0.0339	0.0611	0.0476
		ξ_5	0.2061	0.0178	0.0751	0.0196	0.0241	0.0126	0.0918	0.0556
		ξ_6	0.2345	0.0102	0.1076	0.0105	0.0515	0.011	0.1245	0.0535
	Symbolic Cov.	ξ_1	0.2082	0.0162	0.0854	0.0145	0.0307	0.013	0.107	0.0766
		ξ_2	0.208	0.0158	0.0853	0.0146	0.031	0.0129	0.1069	0.0759
		ξ_3	0.2079	0.0166	0.0854	0.0135	0.0309	0.0134	0.1082	0.0747
		ξ_4	0.15	0.0377	0.0423	0.018	0.0418	0.0358	0.0769	0.0594
		ξ_5	0.1965	0.0197	0.0721	0.0194	0.0224	0.0124	0.098	0.0765
		ξ_6	0.2257	0.0133	0.1039	0.0113	0.0494	0.0114	0.1276	0.0756

Table A.36: Average root mean squared error for latent variable, study with 2 manifest variables

Study	Method	Latent Variable	Loadings 0.5		Loadings 0.7		Loadings 0.9		Random Loadings	
			$RMSE_L$	$RMSE_U$	$RMSE_L$	$RMSE_U$	$RMSE_L$	$RMSE_U$	$RMSE_L$	$RMSE_U$
III	Center	ξ_1	0.8036	0.8036	0.6094	0.6094	0.4904	0.4906	0.6285	0.6289
		ξ_2	0.8029	0.803	0.6102	0.6105	0.4907	0.4909	0.6279	0.6288
		ξ_3	0.4576	0.4577	0.2313	0.2315	0.111	0.1112	0.2531	0.2532
		ξ_4	0.4559	0.456	0.2343	0.2345	0.1091	0.1091	0.2613	0.2613
	Symbolic Cov.	ξ_1	0.8026	0.803	0.608	0.6083	0.4912	0.4916	0.6291	0.6292
		ξ_2	0.8031	0.8036	0.6097	0.6098	0.4895	0.4896	0.6327	0.6329
		ξ_3	0.458	0.4581	0.2307	0.2308	0.1094	0.1094	0.2611	0.2612
IV	Center	ξ_1	0.8027	0.8031	0.6085	0.6085	0.4893	0.4893	0.625	0.6253
		ξ_2	0.8033	0.8035	0.608	0.6081	0.4896	0.49	0.6265	0.6269
		ξ_3	0.4512	0.4514	0.2331	0.2332	0.1109	0.111	0.2585	0.2586
		ξ_4	0.6535	0.6536	0.4205	0.4205	0.2665	0.2665	0.4444	0.4444
	Symbolic Cov.	ξ_1	0.8038	0.804	0.6083	0.6084	0.4915	0.4917	0.6314	0.6322
		ξ_2	0.8016	0.8022	0.6094	0.6095	0.4893	0.4894	0.6281	0.6281
		ξ_3	0.4584	0.4585	0.232	0.232	0.1109	0.1109	0.2683	0.2684
V	Center	ξ_4	0.6566	0.6567	0.4246	0.4247	0.2648	0.2649	0.4569	0.4569
		ξ_1	0.8034	0.8038	0.6076	0.6078	0.4896	0.4898	0.6292	0.6298
		ξ_2	0.8021	0.8021	0.6086	0.6089	0.4894	0.4896	0.6318	0.6324
		ξ_3	0.803	0.8034	0.6091	0.6093	0.4902	0.4905	0.6293	0.6299
		ξ_4	0.3025	0.3026	0.1213	0.1214	0.1164	0.1164	0.1582	0.1582
		ξ_5	0.4512	0.4512	0.2346	0.2346	0.1105	0.1106	0.2591	0.2591
	ξ_6	0.4588	0.4589	0.2303	0.2304	0.1084	0.1084	0.2545	0.2548	
Symbolic Cov.	ξ_1	0.8034	0.8037	0.6085	0.6086	0.4907	0.4907	0.6346	0.6347	
	ξ_2	0.8041	0.8043	0.6096	0.6098	0.4905	0.4907	0.6332	0.6335	
	ξ_3	0.803	0.8037	0.609	0.6095	0.4902	0.4904	0.63	0.6306	
	ξ_4	0.3027	0.3027	0.1263	0.1263	0.1161	0.1162	0.1574	0.1577	
	ξ_5	0.4547	0.4548	0.2333	0.2334	0.1102	0.1102	0.2627	0.2628	
	ξ_6	0.4512	0.4512	0.2321	0.2322	0.1092	0.1094	0.2615	0.2616	
VI	Center	ξ_1	0.8019	0.803	0.6086	0.6093	0.4921	0.4924	0.6283	0.6286
		ξ_2	0.8039	0.804	0.6093	0.6095	0.4907	0.4908	0.6248	0.6248
		ξ_3	0.8023	0.8026	0.6103	0.6104	0.491	0.491	0.6262	0.6264
		ξ_4	0.3015	0.3019	0.1239	0.1241	0.1181	0.1182	0.1527	0.153
		ξ_5	0.4535	0.4537	0.2289	0.229	0.1111	0.1112	0.2576	0.2577
		ξ_6	0.602	0.6021	0.3721	0.3721	0.2217	0.222	0.3968	0.3969
	Symbolic Cov.	ξ_1	0.8025	0.8034	0.6093	0.6094	0.4902	0.4903	0.6323	0.6325
		ξ_2	0.8037	0.8043	0.6096	0.6096	0.4915	0.4916	0.629	0.6291
		ξ_3	0.8027	0.8033	0.6087	0.6089	0.4897	0.4899	0.6318	0.6319
		ξ_4	0.301	0.3012	0.1267	0.1267	0.1178	0.1179	0.158	0.1581
		ξ_5	0.4558	0.4558	0.2338	0.2338	0.1101	0.1101	0.2646	0.2646
		ξ_6	0.6052	0.6053	0.3722	0.3723	0.2219	0.2219	0.4058	0.4059

Table A.37: Average absolute relative bias for studies III-VI with 5 manifest variables

Study	Method	Estimator	Loadings 0.5		Loadings 0.7		Loadings 0.9		Random Loadings	
			Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
III	Center	$\hat{\beta}_{1,3}$	0.0155	0.0135	0.0186	0.0163	0.0206	0.0176	0.0386	0.0302
		$\hat{\beta}_{2,3}$	0.0149	0.0134	0.0185	0.0162	0.0204	0.0177	0.0389	0.031
		$\hat{\beta}_{1,4}$	0.0146	0.0126	0.0188	0.0168	0.0207	0.0184	0.0383	0.0294
		$\hat{\beta}_{2,4}$	0.015	0.0133	0.0189	0.0167	0.0203	0.0174	0.0394	0.0302
	Symbolic Cov.	$\hat{\beta}_{1,3}$	0.0812	0.053	0.05	0.0363	0.039	0.0294	0.054	0.0389
		$\hat{\beta}_{2,3}$	0.0764	0.0497	0.0486	0.0359	0.0377	0.0293	0.0514	0.0378
		$\hat{\beta}_{1,4}$	0.0793	0.0526	0.0485	0.0361	0.0381	0.0296	0.0551	0.0402
		$\hat{\beta}_{2,4}$	0.0777	0.0506	0.0497	0.0374	0.0374	0.0289	0.0508	0.0382
IV	Center	$\hat{\beta}_{1,3}$	0.0143	0.0125	0.0182	0.0162	0.02	0.0179	0.0393	0.0305
		$\hat{\beta}_{2,3}$	0.0142	0.0123	0.0187	0.0162	0.0201	0.0181	0.0389	0.0307
		$\hat{\beta}_{1,4}$	0.1631	0.1197	0.1617	0.1146	0.1518	0.1101	0.1662	0.1227
		$\hat{\beta}_{2,4}$	0.1623	0.1158	0.161	0.1171	0.1542	0.1133	0.1718	0.1274
		$\hat{\beta}_{3,4}$	0.2559	0.2216	0.254	0.2121	0.2462	0.2015	0.2612	0.2193
	Symbolic Cov.	$\hat{\beta}_{1,3}$	0.0804	0.0548	0.05	0.0368	0.0368	0.0285	0.0523	0.0393
		$\hat{\beta}_{2,3}$	0.0792	0.0518	0.0468	0.0339	0.0367	0.0303	0.0536	0.0386
		$\hat{\beta}_{1,4}$	0.7354	0.1962	0.6899	0.2062	0.6556	0.2125	0.6969	0.2239
		$\hat{\beta}_{2,4}$	0.7457	0.1901	0.7035	0.2029	0.6625	0.212	0.6976	0.234
		$\hat{\beta}_{3,4}$	0.9786	0.4616	0.9304	0.4864	0.87	0.4565	0.9113	0.4858
V	Center	$\hat{\beta}_{1,4}$	0.0487	0.0312	0.0629	0.0407	0.0698	0.0463	0.0723	0.0536
		$\hat{\beta}_{1,5}$	0.0149	0.0129	0.0186	0.0166	0.0206	0.0186	0.0373	0.0293
		$\hat{\beta}_{2,5}$	0.0149	0.0127	0.0181	0.0162	0.0196	0.0179	0.0387	0.0294
		$\hat{\beta}_{2,6}$	0.0148	0.0125	0.0185	0.0165	0.0204	0.0182	0.0377	0.0292
		$\hat{\beta}_{3,6}$	0.0145	0.0125	0.0182	0.0169	0.0204	0.0192	0.0404	0.0312
	Symbolic Cov.	$\hat{\beta}_{1,4}$	0.0966	0.0737	0.105	0.0743	0.1019	0.0735	0.1031	0.0754
		$\hat{\beta}_{1,5}$	0.0781	0.05	0.0511	0.038	0.0371	0.0287	0.0528	0.0398
		$\hat{\beta}_{2,5}$	0.079	0.0544	0.047	0.0351	0.0389	0.0306	0.0513	0.0381
		$\hat{\beta}_{2,6}$	0.0797	0.053	0.0489	0.0352	0.0402	0.0299	0.0492	0.0353
		$\hat{\beta}_{3,6}$	0.0783	0.0537	0.0492	0.0357	0.037	0.0293	0.0539	0.0404
VI	Center	$\hat{\beta}_{1,4}$	0.0472	0.0312	0.0613	0.0396	0.0723	0.0455	0.0708	0.0527
		$\hat{\beta}_{1,5}$	0.0148	0.0129	0.0195	0.0174	0.0198	0.0174	0.0383	0.0308
		$\hat{\beta}_{2,5}$	0.0153	0.0134	0.0192	0.0166	0.0198	0.018	0.0385	0.0308
		$\hat{\beta}_{2,6}$	0.0166	0.0113	0.0187	0.0124	0.0206	0.0136	0.0369	0.0276
		$\hat{\beta}_{3,6}$	0.0161	0.0108	0.0182	0.0124	0.0196	0.013	0.0362	0.0264
	Symbolic Cov.	$\hat{\beta}_{5,6}$	0.126	0.0887	0.1263	0.0945	0.1236	0.0899	0.1331	0.0947
		$\hat{\beta}_{1,4}$	0.0978	0.073	0.1087	0.0759	0.1067	0.0748	0.1015	0.0751
		$\hat{\beta}_{1,5}$	0.0779	0.0545	0.0503	0.0382	0.0382	0.0295	0.052	0.0402
		$\hat{\beta}_{2,5}$	0.0792	0.0529	0.049	0.035	0.0383	0.0287	0.0541	0.0407
		$\hat{\beta}_{2,6}$	0.1136	0.0845	0.0874	0.0674	0.0689	0.0518	0.0937	0.0686
	$\hat{\beta}_{3,6}$	0.1253	0.0706	0.0881	0.0547	0.0669	0.0424	0.0938	0.0567	
	$\hat{\beta}_{5,6}$	0.1691	0.1256	0.1542	0.1177	0.1423	0.1048	0.1573	0.1142	

Table A.38: Absolute relative bias of structural model of the loadings associated with each latent variable, study with 5 manifest variable

Study	Method	Latent Variable	Loadings 0.5		Loadings 0.7		Loadings 0.9		Random Loadings	
			Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
III	Center	ξ_1	0.2188	0.0124	0.0904	0.0129	0.033	0.0123	0.1013	0.0341
		ξ_2	0.2188	0.0124	0.0899	0.0129	0.033	0.0126	0.1021	0.0352
		ξ_3	0.2057	0.0178	0.0759	0.0194	0.0233	0.013	0.0878	0.039
		ξ_4	0.2062	0.0176	0.0756	0.0196	0.0237	0.0126	0.0888	0.0384
	Symbolic Cov.	ξ_1	0.2061	0.0168	0.0853	0.0134	0.0301	0.0126	0.1061	0.0751
		ξ_2	0.2057	0.0165	0.0846	0.0138	0.0304	0.0126	0.1047	0.0751
		ξ_3	0.1929	0.0212	0.0696	0.0211	0.022	0.0123	0.0925	0.0743
IV	Center	ξ_1	0.2172	0.0123	0.0895	0.0131	0.0327	0.0124	0.1026	0.036
		ξ_2	0.2184	0.0124	0.0899	0.0126	0.0331	0.0127	0.1003	0.0351
		ξ_3	0.2066	0.0171	0.0761	0.019	0.0242	0.0127	0.0865	0.0377
		ξ_4	0.2416	0.0087	0.1156	0.0086	0.0598	0.0086	0.128	0.0338
	Symbolic Cov.	ξ_1	0.2051	0.0169	0.0849	0.0145	0.0302	0.0128	0.1083	0.0753
		ξ_2	0.206	0.0166	0.0854	0.0137	0.0299	0.0129	0.1061	0.0753
		ξ_3	0.193	0.0214	0.0704	0.0209	0.0222	0.0126	0.0956	0.0749
V	Center	ξ_4	0.2306	0.0132	0.1108	0.0101	0.0573	0.0091	0.1306	0.0734
		ξ_1	0.2181	0.012	0.0897	0.0124	0.0332	0.0128	0.1017	0.0343
		ξ_2	0.2179	0.0122	0.0891	0.0131	0.0323	0.0123	0.1000	0.0349
		ξ_3	0.2181	0.0124	0.0896	0.013	0.0331	0.0126	0.1023	0.0362
		ξ_4	0.162	0.0341	0.0426	0.0196	0.04	0.0338	0.0514	0.0352
		ξ_5	0.2057	0.0175	0.0757	0.0197	0.024	0.0127	0.0882	0.0375
	Symbolic Cov.	ξ_6	0.2061	0.0173	0.0755	0.0196	0.0239	0.0125	0.0855	0.0376
VI	Center	ξ_1	0.2047	0.017	0.0841	0.014	0.0302	0.0129	0.1065	0.0757
		ξ_2	0.2055	0.0169	0.085	0.0136	0.0304	0.0128	0.1076	0.0752
		ξ_3	0.2058	0.017	0.0847	0.0139	0.0306	0.0127	0.1065	0.0747
		ξ_4	0.1464	0.0391	0.0412	0.0172	0.0408	0.0358	0.078	0.0591
		ξ_5	0.1926	0.0209	0.0708	0.0201	0.0219	0.0123	0.095	0.0743
		ξ_6	0.193	0.0219	0.0703	0.0202	0.0225	0.0127	0.0939	0.0743
	Symbolic Cov.	ξ_1	0.2179	0.0126	0.0899	0.0126	0.0328	0.0124	0.0979	0.0344
Center	ξ_2	0.2181	0.0122	0.0893	0.013	0.032	0.0128	0.1015	0.0349	
	ξ_3	0.2173	0.0128	0.0895	0.0127	0.033	0.0123	0.1005	0.0351	
	ξ_4	0.1636	0.034	0.043	0.0195	0.0415	0.0338	0.0528	0.037	
	ξ_5	0.2058	0.0179	0.0744	0.0201	0.0237	0.0123	0.0865	0.0379	
	ξ_6	0.2341	0.0104	0.107	0.0106	0.0508	0.0106	0.1202	0.0342	
	Symbolic Cov.	ξ_1	0.2063	0.0164	0.0844	0.0144	0.0303	0.0122	0.1065	0.0761
ξ_2		0.2048	0.0169	0.0851	0.0141	0.0299	0.0132	0.1061	0.0756	
ξ_3		0.2054	0.0166	0.0841	0.0141	0.0298	0.0126	0.1068	0.0757	
ξ_4		0.1474	0.039	0.0395	0.0174	0.0436	0.0369	0.0765	0.0584	
ξ_5		0.1928	0.0217	0.0702	0.0209	0.0229	0.012	0.0934	0.074	
ξ_6		0.2228	0.0148	0.1033	0.0118	0.0492	0.0108	0.125	0.075	

Table A.39: Average root mean squared error for latent variable, study with 5 manifest variables

Study	Method	Latent Variable	Loadings 0.5		Loadings 0.7		Loadings 0.9		Random Loadings	
			$RMSE_L$	$RMSE_U$	$RMSE_L$	$RMSE_U$	$RMSE_L$	$RMSE_U$	$RMSE_L$	$RMSE_U$
III	Center	ξ_1	0.8021	0.8025	0.6073	0.6076	0.4904	0.4904	0.6257	0.6262
		ξ_2	0.8018	0.8019	0.6083	0.6084	0.4892	0.4892	0.6242	0.6243
		ξ_3	0.4527	0.4528	0.2322	0.2323	0.1082	0.1083	0.2535	0.2535
		ξ_4	0.4541	0.4543	0.2311	0.2312	0.1088	0.1089	0.2556	0.2557
	Symbolic Cov.	ξ_1	0.8026	0.8028	0.6083	0.6085	0.4903	0.4905	0.6311	0.6315
		ξ_2	0.8031	0.8038	0.6093	0.6099	0.4905	0.4906	0.6252	0.6259
		ξ_3	0.452	0.452	0.2293	0.2295	0.1089	0.109	0.2515	0.2515
IV	Center	ξ_1	0.8011	0.8013	0.6109	0.611	0.4911	0.4911	0.6275	0.6278
		ξ_2	0.804	0.804	0.6096	0.6096	0.4896	0.4896	0.6259	0.6259
		ξ_3	0.4562	0.4563	0.2324	0.2325	0.1107	0.1107	0.251	0.2512
		ξ_4	0.6538	0.6539	0.4219	0.422	0.269	0.269	0.4462	0.4462
	Symbolic Cov.	ξ_1	0.8033	0.8047	0.6082	0.6082	0.4897	0.4899	0.6334	0.6336
		ξ_2	0.8028	0.8035	0.6085	0.6086	0.4905	0.4906	0.6295	0.6295
		ξ_3	0.4539	0.4539	0.2317	0.2318	0.1089	0.109	0.26	0.2601
V	Center	ξ_4	0.6551	0.6551	0.4218	0.4219	0.2677	0.2678	0.4484	0.4485
		ξ_1	0.8024	0.8025	0.6081	0.6082	0.4901	0.4902	0.6249	0.6249
		ξ_2	0.8024	0.8027	0.6092	0.6095	0.4911	0.4912	0.6227	0.6231
		ξ_3	0.8039	0.8043	0.6091	0.6091	0.4909	0.491	0.6264	0.6265
		ξ_4	0.3003	0.3005	0.1225	0.1226	0.1159	0.116	0.1381	0.1381
		ξ_5	0.4522	0.4522	0.2317	0.2318	0.1104	0.1106	0.2531	0.2532
	ξ_6	0.4533	0.4534	0.2318	0.232	0.11	0.1101	0.2477	0.2478	
Symbolic Cov.	ξ_1	0.8034	0.8042	0.6083	0.6085	0.4895	0.4896	0.6308	0.6311	
	ξ_2	0.8033	0.8034	0.6084	0.609	0.4909	0.4911	0.6306	0.6307	
	ξ_3	0.8032	0.8035	0.6089	0.6091	0.4909	0.4909	0.6303	0.6303	
	ξ_4	0.3011	0.3014	0.1211	0.1212	0.1137	0.1139	0.1451	0.1452	
	ξ_5	0.4524	0.4525	0.233	0.2331	0.1088	0.1089	0.2612	0.2613	
	ξ_6	0.4547	0.455	0.2312	0.2314	0.1104	0.1105	0.2562	0.2564	
VI	Center	ξ_1	0.8034	0.8035	0.6095	0.6098	0.4903	0.4905	0.619	0.619
		ξ_2	0.8032	0.8032	0.6089	0.609	0.4916	0.4917	0.6264	0.6264
		ξ_3	0.8038	0.8042	0.6089	0.609	0.4908	0.491	0.6239	0.6242
		ξ_4	0.3055	0.3056	0.1237	0.124	0.1178	0.1179	0.1403	0.1405
		ξ_5	0.453	0.453	0.2267	0.2269	0.1102	0.1104	0.2497	0.2497
		ξ_6	0.6011	0.6011	0.3688	0.3688	0.2172	0.2173	0.3959	0.3959
	Symbolic Cov.	ξ_1	0.8033	0.804	0.6088	0.6092	0.4907	0.491	0.6319	0.6319
		ξ_2	0.8039	0.8041	0.6097	0.6098	0.4892	0.4898	0.6296	0.6301
		ξ_3	0.8025	0.8031	0.6093	0.6095	0.4918	0.492	0.6297	0.63
		ξ_4	0.3026	0.3028	0.117	0.117	0.1179	0.1181	0.1446	0.1447
		ξ_5	0.4526	0.4527	0.2318	0.2318	0.1113	0.1113	0.2597	0.2599
		ξ_6	0.604	0.6043	0.3761	0.3762	0.2227	0.2229	0.4045	0.4046

Table A.40: Average absolute relative bias for studies III-VI with 8 manifest variables

Study	Method	Estimator	Loadings 0.5		Loadings 0.7		Loadings 0.9		Random Loadings			
			Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.		
III	Center	$\hat{\beta}_{1,3}$	0.0157	0.0137	0.0192	0.0168	0.0211	0.0186	0.0331	0.0249		
		$\hat{\beta}_{2,3}$	0.0154	0.0133	0.0187	0.0166	0.0214	0.0189	0.0327	0.0263		
		$\hat{\beta}_{1,4}$	0.0148	0.0127	0.0186	0.0162	0.0209	0.018	0.0319	0.0251		
		$\hat{\beta}_{2,4}$	0.0153	0.0134	0.0185	0.016	0.0208	0.0177	0.0311	0.024		
	Symbolic Cov.	$\hat{\beta}_{1,3}$	0.0788	0.0503	0.0509	0.0387	0.0385	0.0299	0.0527	0.0395		
		$\hat{\beta}_{2,3}$	0.0756	0.0507	0.0491	0.0375	0.0382	0.0299	0.0516	0.038		
		$\hat{\beta}_{1,4}$	0.0797	0.0501	0.0508	0.0374	0.0368	0.0284	0.0514	0.0377		
		$\hat{\beta}_{2,4}$	0.0759	0.0504	0.0493	0.037	0.0401	0.0313	0.0501	0.0378		
IV	Center	$\hat{\beta}_{1,3}$	0.0146	0.0127	0.0185	0.0164	0.0208	0.0192	0.0313	0.0254		
		$\hat{\beta}_{2,3}$	0.0154	0.0133	0.0183	0.016	0.021	0.0187	0.0313	0.0243		
		$\hat{\beta}_{1,4}$	0.1501	0.1112	0.1538	0.1141	0.1506	0.1154	0.1506	0.108		
		$\hat{\beta}_{2,4}$	0.1502	0.1121	0.1605	0.1196	0.151	0.1147	0.1509	0.107		
		$\hat{\beta}_{3,4}$	0.2433	0.204	0.2447	0.2003	0.2429	0.2036	0.2366	0.1893		
	Symbolic Cov.	$\hat{\beta}_{1,3}$	0.0751	0.0503	0.0503	0.037	0.0371	0.0292	0.0531	0.0379		
		$\hat{\beta}_{2,3}$	0.0831	0.0546	0.0485	0.0354	0.0412	0.0315	0.0524	0.0389		
		$\hat{\beta}_{1,4}$	0.7452	0.1959	0.6931	0.2006	0.6733	0.2064	0.7128	0.2187		
		$\hat{\beta}_{2,4}$	0.7572	0.1824	0.6937	0.2056	0.6578	0.2095	0.7045	0.2228		
		$\hat{\beta}_{3,4}$	1.005	0.4606	0.9291	0.4763	0.9152	0.4679	0.9539	0.4845		
		V	Center	$\hat{\beta}_{1,4}$	0.0478	0.0313	0.0633	0.0413	0.071	0.0459	0.0663	0.0489
				$\hat{\beta}_{1,5}$	0.0145	0.0121	0.0189	0.0164	0.0215	0.0183	0.0334	0.0256
$\hat{\beta}_{2,5}$	0.0143			0.0123	0.0189	0.0166	0.0215	0.0187	0.032	0.0251		
$\hat{\beta}_{2,6}$	0.0148			0.0131	0.0184	0.0161	0.0197	0.0171	0.032	0.0248		
$\hat{\beta}_{3,6}$	0.0153			0.013	0.0179	0.016	0.0195	0.0173	0.0307	0.0242		
Symbolic Cov.	$\hat{\beta}_{1,4}$		0.0957	0.0742	0.0993	0.0737	0.1044	0.0743	0.1027	0.0741		
	$\hat{\beta}_{1,5}$		0.0785	0.0509	0.0516	0.0391	0.0392	0.0296	0.0524	0.0392		
	$\hat{\beta}_{2,5}$		0.0789	0.0527	0.0491	0.0382	0.0366	0.0281	0.0513	0.0382		
VI	Center	$\hat{\beta}_{2,6}$	0.081	0.051	0.0482	0.0348	0.0381	0.0306	0.0522	0.0385		
		$\hat{\beta}_{3,6}$	0.077	0.0504	0.0498	0.0373	0.0372	0.03	0.049	0.0367		
		$\hat{\beta}_{1,4}$	0.0493	0.0318	0.0646	0.0407	0.0729	0.0469	0.0674	0.0484		
		$\hat{\beta}_{1,5}$	0.0147	0.0125	0.0189	0.0167	0.0199	0.0175	0.0324	0.0251		
		$\hat{\beta}_{2,5}$	0.0148	0.0131	0.019	0.0167	0.0202	0.017	0.032	0.0256		
	Symbolic Cov.	$\hat{\beta}_{2,6}$	0.0155	0.0108	0.0181	0.0125	0.0207	0.0138	0.0314	0.0232		
		$\hat{\beta}_{3,6}$	0.0154	0.0108	0.0178	0.0119	0.0201	0.0131	0.0337	0.069		
		$\hat{\beta}_{5,6}$	0.1245	0.0899	0.129	0.0937	0.1215	0.0896	0.1279	0.0923		
		$\hat{\beta}_{1,4}$	0.0994	0.0721	0.1012	0.0749	0.1048	0.075	0.102	0.0736		
		$\hat{\beta}_{1,5}$	0.0787	0.0513	0.0488	0.0356	0.0385	0.0295	0.0545	0.0401		
Symbolic Cov.	$\hat{\beta}_{2,5}$	0.0772	0.0508	0.0488	0.0374	0.0376	0.0294	0.051	0.0372			
	$\hat{\beta}_{2,6}$	0.1126	0.0813	0.0849	0.0654	0.0694	0.0535	0.0907	0.072			
	$\hat{\beta}_{3,6}$	0.1249	0.0716	0.0833	0.0516	0.0645	0.0437	0.0921	0.079			
	$\hat{\beta}_{5,6}$	0.1647	0.1214	0.1495	0.1124	0.1467	0.1049	0.1505	0.1171			

Table A.41: Average absolute relative bias of the loadings associated with each latent variable, study with 8 manifest variable

Study	Method	Latent Variable	Loadings 0.5		Loadings 0.7		Loadings 0.9		Random Loadings	
			Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
III	Center	ξ_1	0.218	0.0132	0.09	0.013	0.0329	0.0128	0.0992	0.0285
		ξ_2	0.2183	0.012	0.0895	0.0129	0.0338	0.0128	0.1000	0.0275
		ξ_3	0.2049	0.0177	0.0753	0.0195	0.0234	0.0125	0.0847	0.0324
		ξ_4	0.2059	0.0182	0.0752	0.0192	0.0237	0.0126	0.0876	0.0315
	Symbolic Cov.	ξ_1	0.2045	0.0169	0.0846	0.0146	0.0303	0.0126	0.1038	0.0748
		ξ_2	0.2055	0.0166	0.084	0.0146	0.0303	0.0124	0.1039	0.0749
		ξ_3	0.1917	0.0222	0.07	0.0198	0.0228	0.0122	0.0949	0.074
IV	Center	ξ_1	0.2177	0.0121	0.09	0.0129	0.0329	0.0119	0.0996	0.0276
		ξ_2	0.2193	0.0125	0.0903	0.0125	0.0331	0.0124	0.0982	0.0284
		ξ_3	0.2063	0.018	0.0758	0.0188	0.0238	0.0124	0.0841	0.0302
		ξ_4	0.2417	0.0088	0.1154	0.0087	0.0594	0.0087	0.1252	0.0262
	Symbolic Cov.	ξ_1	0.2064	0.0167	0.0835	0.0147	0.0305	0.0129	0.1045	0.0745
		ξ_2	0.2049	0.017	0.0848	0.0142	0.0304	0.0127	0.1055	0.0748
		ξ_3	0.1927	0.0218	0.0692	0.0201	0.0223	0.0123	0.0928	0.0739
V	Center	ξ_4	0.2	0.0137	0.1105	0.0098	0.0572	0.009	0.131	0.0729
		ξ_1	0.2185	0.0124	0.0906	0.0126	0.0329	0.0127	0.0989	0.0291
		ξ_2	0.2183	0.0117	0.0901	0.0125	0.0331	0.0121	0.0988	0.0289
		ξ_3	0.2183	0.0121	0.0901	0.0131	0.0331	0.0121	0.0981	0.0281
		ξ_4	0.1636	0.0341	0.0428	0.0201	0.0405	0.0339	0.0505	0.0314
		ξ_5	0.2069	0.017	0.0753	0.0193	0.0229	0.0124	0.0857	0.0308
	Symbolic Cov.	ξ_6	0.2053	0.0175	0.0769	0.0191	0.0232	0.0127	0.0872	0.0317
		ξ_1	0.2046	0.017	0.0837	0.0146	0.0299	0.0129	0.1064	0.0752
		ξ_2	0.2039	0.0172	0.0845	0.0137	0.0302	0.013	0.1045	0.0745
		ξ_3	0.2047	0.0174	0.0841	0.014	0.0303	0.0132	0.105	0.0751
		ξ_4	0.1474	0.0382	0.0414	0.0177	0.0429	0.0369	0.0778	0.059
		ξ_5	0.1916	0.0217	0.0708	0.0198	0.0223	0.0124	0.0941	0.074
		ξ_6	0.1925	0.0208	0.0704	0.0203	0.0218	0.0126	0.0937	0.0739
VI	Center	ξ_1	0.218	0.0124	0.0894	0.0131	0.0329	0.0125	0.0997	0.029
		ξ_2	0.2185	0.0127	0.0901	0.0136	0.0327	0.0127	0.0989	0.0287
		ξ_3	0.2181	0.0126	0.0899	0.0125	0.0331	0.013	0.1014	0.0726
		ξ_4	0.1616	0.0347	0.0415	0.0195	0.0424	0.0345	0.0495	0.0317
		ξ_5	0.2061	0.0177	0.075	0.0197	0.0237	0.0122	0.086	0.0319
		ξ_6	0.2346	0.0102	0.1073	0.0108	0.052	0.0103	0.1169	0.0276
	Symbolic Cov.	ξ_1	0.2057	0.0166	0.0848	0.0137	0.0299	0.0129	0.1049	0.075
		ξ_2	0.2055	0.0168	0.0836	0.0143	0.0301	0.0124	0.106	0.0754
		ξ_3	0.2056	0.0166	0.0839	0.0141	0.0313	0.0132	0.1077	0.0979
		ξ_4	0.1465	0.0383	0.0414	0.0178	0.0428	0.0367	0.0773	0.0584
		ξ_5	0.1923	0.022	0.0713	0.0204	0.0221	0.0124	0.0943	0.0744
		ξ_6	0.2219	0.0149	0.103	0.012	0.0491	0.0108	0.123	0.0741

Table A.42: Average root mean squared error for latent variable, study with 8 manifest variables

Study	Method	Latent Variable	Loadings 0.5		Loadings 0.7		Loadings 0.9		Random Loadings	
			$RMSE_L$	$RMSE_U$	$RMSE_L$	$RMSE_U$	$RMSE_L$	$RMSE_U$	$RMSE_L$	$RMSE_U$
III	Center	ξ_1	0.8017	0.8025	0.6077	0.6082	0.4902	0.4904	0.6216	0.622
		ξ_2	0.8017	0.8019	0.6098	0.6103	0.4892	0.4896	0.6221	0.6222
		ξ_3	0.4485	0.4485	0.2299	0.2299	0.1071	0.1072	0.2469	0.2469
		ξ_4	0.4536	0.4537	0.2294	0.2294	0.1082	0.1083	0.2547	0.2547
	Symbolic Cov.	ξ_1	0.8031	0.8035	0.6108	0.6112	0.4901	0.4902	0.6265	0.6272
		ξ_2	0.8016	0.8018	0.6093	0.61	0.4896	0.4897	0.6276	0.6277
		ξ_3	0.4514	0.4515	0.2321	0.2321	0.1103	0.1103	0.2603	0.2603
		ξ_4	0.4543	0.4543	0.234	0.2341	0.108	0.1081	0.2568	0.2569
IV	Center	ξ_1	0.802	0.8029	0.6091	0.6092	0.4886	0.4886	0.6231	0.6234
		ξ_2	0.8025	0.8033	0.607	0.6075	0.4906	0.4907	0.6222	0.6223
		ξ_3	0.4552	0.4553	0.2309	0.231	0.1086	0.1086	0.2468	0.247
		ξ_4	0.6542	0.6544	0.4206	0.4207	0.2665	0.2665	0.4413	0.4413
	Symbolic Cov.	ξ_1	0.8028	0.8033	0.6095	0.6098	0.4908	0.491	0.6283	0.6284
		ξ_2	0.802	0.8023	0.6085	0.6087	0.4897	0.4899	0.6303	0.6307
		ξ_3	0.4552	0.4554	0.228	0.228	0.1095	0.1095	0.2562	0.2562
		ξ_4	0.6554	0.6555	0.4205	0.4205	0.2688	0.2689	0.4497	0.4498
V	Center	ξ_1	0.8021	0.8024	0.609	0.6092	0.4903	0.4903	0.621	0.6212
		ξ_2	0.8025	0.8029	0.6094	0.6095	0.4896	0.4899	0.6223	0.6226
		ξ_3	0.8025	0.8025	0.6088	0.6091	0.4902	0.4904	0.6227	0.6229
		ξ_4	0.3052	0.3053	0.1227	0.1227	0.1161	0.1162	0.1375	0.1376
		ξ_5	0.457	0.4571	0.2293	0.2293	0.1055	0.1056	0.2497	0.25
		ξ_6	0.4505	0.4505	0.2355	0.2356	0.1079	0.1079	0.2552	0.2553
	Symbolic Cov.	ξ_1	0.8024	0.8028	0.6099	0.61	0.4919	0.4919	0.6303	0.6308
		ξ_2	0.8031	0.8037	0.6093	0.6095	0.4911	0.4911	0.6274	0.6276
		ξ_3	0.8039	0.805	0.6094	0.6097	0.4902	0.4903	0.6292	0.6294
		ξ_4	0.3051	0.3053	0.1246	0.1248	0.1163	0.1164	0.1431	0.1432
		ξ_5	0.4514	0.4517	0.2344	0.2344	0.1094	0.1096	0.2576	0.2578
		ξ_6	0.4531	0.4533	0.2333	0.2333	0.1082	0.1083	0.2557	0.2558
VI	Center	ξ_1	0.8028	0.8033	0.6084	0.6087	0.49	0.4904	0.6238	0.6243
		ξ_2	0.8042	0.8048	0.61	0.6102	0.4917	0.4917	0.622	0.6222
		ξ_3	0.8021	0.8025	0.6083	0.6083	0.4908	0.491	0.6235	0.6236
		ξ_4	0.2998	0.2998	0.119	0.1191	0.1189	0.1189	0.1362	0.1362
		ξ_5	0.4543	0.4544	0.2286	0.2287	0.1086	0.1086	0.2512	0.2512
		ξ_6	0.6033	0.6034	0.3713	0.3714	0.2238	0.2238	0.3906	0.3906
	Symbolic Cov.	ξ_1	0.8028	0.8032	0.609	0.6093	0.4891	0.4892	0.6299	0.6301
		ξ_2	0.8027	0.8029	0.6075	0.6078	0.4904	0.4907	0.6272	0.6278
		ξ_3	0.8031	0.804	0.6094	0.6098	0.4893	0.4894	0.6314	0.6316
		ξ_4	0.3005	0.3008	0.1242	0.1242	0.1162	0.1162	0.1407	0.1407
		ξ_5	0.4538	0.4539	0.236	0.236	0.1092	0.1093	0.2583	0.2584
		ξ_6	0.6025	0.6026	0.3758	0.3758	0.222	0.2221	0.4006	0.4006

Table A.43: Average absolute relative bias for studies III-VI with 10 manifest variables

Study	Method	Estimator	Loadings 0.5		Loadings 0.7		Loadings 0.9		Random Loadings	
			Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
III	Center	$\hat{\beta}_{1,3}$	0.0156	0.0139	0.0185	0.0163	0.0212	0.0191	0.0307	0.023
		$\hat{\beta}_{2,3}$	0.0157	0.0131	0.0188	0.0164	0.021	0.0182	0.0305	0.0236
		$\hat{\beta}_{1,4}$	0.0146	0.0126	0.0185	0.0164	0.021	0.0188	0.0297	0.0243
		$\hat{\beta}_{2,4}$	0.0147	0.0128	0.0189	0.0166	0.0205	0.0182	0.0303	0.0252
	Symbolic Cov.	$\hat{\beta}_{1,3}$	0.0808	0.0531	0.0489	0.0371	0.0392	0.0304	0.0529	0.038
		$\hat{\beta}_{2,3}$	0.0781	0.052	0.0483	0.0361	0.0395	0.0306	0.053	0.0398
		$\hat{\beta}_{1,4}$	0.0804	0.0547	0.0493	0.0368	0.0378	0.0302	0.0522	0.0367
		$\hat{\beta}_{2,4}$	0.0769	0.0526	0.0479	0.0355	0.0385	0.0296	0.0539	0.0393
IV	Center	$\hat{\beta}_{1,3}$	0.0148	0.0128	0.0198	0.017	0.0209	0.0188	0.0304	0.0241
		$\hat{\beta}_{2,3}$	0.0151	0.0128	0.0193	0.0169	0.0206	0.018	0.0304	0.0232
		$\hat{\beta}_{1,4}$	0.1577	0.1158	0.1518	0.1107	0.1457	0.1135	0.1555	0.1149
		$\hat{\beta}_{2,4}$	0.158	0.1132	0.1521	0.1096	0.1503	0.1177	0.153	0.1132
		$\hat{\beta}_{3,4}$	0.2536	0.2091	0.2435	0.1961	0.2471	0.2079	0.253	0.2069
	Symbolic Cov.	$\hat{\beta}_{1,3}$	0.0826	0.052	0.0493	0.0366	0.0369	0.0294	0.0498	0.0371
		$\hat{\beta}_{2,3}$	0.0757	0.0517	0.0507	0.038	0.0381	0.0292	0.052	0.0378
		$\hat{\beta}_{1,4}$	0.7421	0.1947	0.6998	0.1958	0.6494	0.2125	0.7119	0.2074
		$\hat{\beta}_{2,4}$	0.7527	0.1936	0.7087	0.2	0.6512	0.2112	0.7005	0.2237
		$\hat{\beta}_{3,4}$	1.013	0.4798	0.9317	0.4695	0.8558	0.4618	0.9424	0.4793
V	Center	$\hat{\beta}_{1,4}$	0.0496	0.0311	0.0618	0.0408	0.071	0.0459	0.0657	0.0475
		$\hat{\beta}_{1,5}$	0.0149	0.0128	0.019	0.0162	0.0209	0.0182	0.0296	0.0233
		$\hat{\beta}_{2,5}$	0.0151	0.0131	0.0183	0.0168	0.0205	0.0184	0.0299	0.0236
		$\hat{\beta}_{2,6}$	0.0141	0.0121	0.0189	0.0163	0.0197	0.0171	0.0311	0.0239
		$\hat{\beta}_{3,6}$	0.0142	0.0123	0.019	0.0164	0.0196	0.0174	0.0313	0.024
	Symbolic Cov.	$\hat{\beta}_{1,4}$	0.0994	0.0731	0.1065	0.0739	0.1047	0.0749	0.1034	0.0751
		$\hat{\beta}_{1,5}$	0.0796	0.0533	0.0483	0.037	0.0382	0.0299	0.0527	0.0389
		$\hat{\beta}_{2,5}$	0.0782	0.0503	0.0505	0.037	0.0393	0.031	0.0532	0.0413
		$\hat{\beta}_{2,6}$	0.0792	0.0507	0.0513	0.0373	0.0385	0.0288	0.0533	0.0388
		$\hat{\beta}_{3,6}$	0.0792	0.0517	0.0483	0.0356	0.0363	0.0287	0.0526	0.0381
VI	Center	$\hat{\beta}_{1,4}$	0.0492	0.0323	0.0638	0.0409	0.0713	0.0459	0.0693	0.0489
		$\hat{\beta}_{1,5}$	0.0148	0.0131	0.0188	0.0167	0.0207	0.0185	0.0311	0.0238
		$\hat{\beta}_{2,5}$	0.0147	0.0126	0.0185	0.016	0.0203	0.0179	0.0314	0.0241
		$\hat{\beta}_{2,6}$	0.0159	0.0113	0.0188	0.0125	0.0205	0.0135	0.0287	0.0209
		$\hat{\beta}_{3,6}$	0.0155	0.0102	0.018	0.0121	0.0196	0.0128	0.0285	0.0206
	Symbolic Cov.	$\hat{\beta}_{5,6}$	0.1193	0.0872	0.1261	0.09	0.1267	0.0942	0.1282	0.0939
		$\hat{\beta}_{1,4}$	0.1026	0.0754	0.1026	0.0746	0.1076	0.0765	0.1015	0.0754
		$\hat{\beta}_{1,5}$	0.077	0.0507	0.0474	0.0344	0.0376	0.0294	0.0523	0.0373
		$\hat{\beta}_{2,5}$	0.0794	0.0519	0.0477	0.0341	0.0374	0.0297	0.051	0.0363
		$\hat{\beta}_{2,6}$	0.1143	0.0874	0.088	0.0668	0.073	0.0562	0.0921	0.0703
$\hat{\beta}_{3,6}$	0.1281	0.071	0.0853	0.0538	0.0664	0.0423	0.0897	0.0559		
$\hat{\beta}_{5,6}$	0.1658	0.1233	0.1515	0.1104	0.1413	0.1048	0.1586	0.1169		

Table A.44: Average absolute relative bias of the loadings associated with each latent variable, study with 10 manifest variable

Study	Method	Latent Variable	Loadings 0.5		Loadings 0.7		Loadings 0.9		Random Loadings	
			Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
III	Center	ξ_1	0.2182	0.0123	0.0898	0.0122	0.0332	0.0127	0.097	0.0257
		ξ_2	0.2182	0.0123	0.0897	0.0127	0.0333	0.0127	0.0987	0.0264
		ξ_3	0.2053	0.0182	0.0751	0.0197	0.0233	0.0123	0.0845	0.0296
		ξ_4	0.206	0.0174	0.075	0.0196	0.0237	0.0122	0.0846	0.0301
	Symbolic Cov.	ξ_1	0.2048	0.0166	0.0842	0.0144	0.0298	0.0127	0.1046	0.0747
		ξ_2	0.2055	0.0172	0.0845	0.0139	0.0294	0.0129	0.1044	0.075
		ξ_3	0.1929	0.0214	0.0705	0.0195	0.0219	0.0123	0.0949	0.0744
IV	Center	ξ_1	0.2182	0.0121	0.0904	0.0126	0.0327	0.0124	0.0984	0.0257
		ξ_2	0.218	0.0127	0.0903	0.013	0.0326	0.0125	0.1003	0.0258
		ξ_3	0.2055	0.0176	0.0749	0.0202	0.0234	0.0124	0.0841	0.0301
		ξ_4	0.2414	0.0088	0.1154	0.0085	0.0598	0.0084	0.1235	0.0238
	Symbolic Cov.	ξ_1	0.2041	0.0172	0.0848	0.0142	0.0304	0.013	0.1056	0.0747
		ξ_2	0.205	0.0169	0.0839	0.0143	0.0304	0.0125	0.1057	0.0746
		ξ_3	0.1928	0.0216	0.0698	0.0206	0.0222	0.0123	0.0939	0.074
V	Center	ξ_4	0.2298	0.0137	0.1107	0.0101	0.0571	0.0093	0.1314	0.0736
		ξ_1	0.218	0.0122	0.0899	0.0129	0.0332	0.0128	0.0991	0.0252
		ξ_2	0.2181	0.012	0.0893	0.0127	0.0326	0.0127	0.0989	0.0257
		ξ_3	0.2184	0.0124	0.0902	0.0128	0.0324	0.0128	0.1002	0.0256
		ξ_4	0.1613	0.0347	0.043	0.0199	0.0404	0.0339	0.0491	0.0307
		ξ_6	0.2066	0.0168	0.0755	0.0198	0.0238	0.0123	0.0854	0.0308
	Symbolic Cov.	ξ_1	0.2048	0.0173	0.0838	0.0145	0.0295	0.0127	0.105	0.0749
		ξ_2	0.2048	0.0171	0.0843	0.0145	0.0303	0.0126	0.1067	0.0753
		ξ_3	0.2047	0.017	0.0841	0.0146	0.0302	0.0127	0.1046	0.075
		ξ_4	0.1457	0.0385	0.0399	0.0171	0.0435	0.0372	0.0771	0.0577
		ξ_5	0.1914	0.0217	0.0707	0.0197	0.0218	0.0121	0.0942	0.0742
		ξ_6	0.1932	0.0214	0.0696	0.0208	0.0218	0.0126	0.093	0.073
		VI	Center	ξ_1	0.2174	0.0128	0.0906	0.0129	0.0336	0.0124
ξ_2	0.2175			0.0127	0.0897	0.0131	0.0324	0.0128	0.0995	0.0257
ξ_3	0.2174			0.0124	0.0901	0.0125	0.0332	0.0125	0.0998	0.0256
ξ_4	0.1613			0.0349	0.0427	0.0201	0.0405	0.0336	0.0479	0.0307
ξ_5	0.2055			0.0172	0.076	0.0191	0.0241	0.0125	0.0845	0.0295
ξ_6	0.2343			0.0104	0.1078	0.0107	0.0511	0.0109	0.1177	0.0243
Symbolic Cov.	ξ_1		0.2047	0.0173	0.0834	0.0145	0.0305	0.0128	0.1048	0.0747
	ξ_2		0.205	0.0173	0.0847	0.0139	0.0297	0.0127	0.1061	0.0753
	ξ_3		0.2047	0.017	0.0841	0.0146	0.03	0.0132	0.1052	0.0749
	ξ_4		0.1442	0.0397	0.0407	0.0173	0.0436	0.0369	0.0765	0.0587
	ξ_5		0.1924	0.022	0.0703	0.0202	0.022	0.0122	0.0942	0.074
	ξ_6		0.2219	0.0148	0.1022	0.0116	0.0488	0.0113	0.1226	0.0739

Table A.45: Average root mean squared error for latent variable, study with 10 manifest variables

Study	Method	Latent Variable	Loadings 0.5		Loadings 0.7		Loadings 0.9		Random Loadings	
			$RMSE_L$	$RMSE_U$	$RMSE_L$	$RMSE_U$	$RMSE_L$	$RMSE_U$	$RMSE_L$	$RMSE_U$
III	Center	ξ_1	0.8021	0.8023	0.6104	0.6105	0.4895	0.4896	0.6196	0.6198
		ξ_2	0.8031	0.8036	0.6073	0.6078	0.4898	0.4901	0.6216	0.6217
		ξ_3	0.4508	0.4509	0.2289	0.2291	0.1067	0.1067	0.2473	0.2474
		ξ_4	0.4539	0.4539	0.2283	0.2286	0.108	0.108	0.2475	0.2476
	Symbolic Cov.	ξ_1	0.8035	0.8035	0.6084	0.6092	0.4906	0.4906	0.6296	0.6298
		ξ_2	0.8012	0.8021	0.6082	0.6083	0.4908	0.4909	0.6266	0.6267
		ξ_3	0.4568	0.457	0.2327	0.2328	0.1073	0.1074	0.2609	0.261
		ξ_4	0.4543	0.4545	0.2347	0.2347	0.1085	0.1086	0.2591	0.2592
IV	Center	ξ_1	0.8025	0.8026	0.6077	0.608	0.4912	0.4915	0.6204	0.6206
		ξ_2	0.802	0.8029	0.6077	0.6077	0.4908	0.491	0.6215	0.6219
		ξ_3	0.4519	0.4521	0.2284	0.2284	0.1075	0.1076	0.247	0.2472
		ξ_4	0.6523	0.6523	0.421	0.4211	0.2695	0.2695	0.4377	0.4378
	Symbolic Cov.	ξ_1	0.8011	0.8014	0.6091	0.6092	0.4887	0.4888	0.6284	0.6285
		ξ_2	0.8023	0.8023	0.6104	0.6109	0.4889	0.489	0.6296	0.6296
		ξ_3	0.4572	0.4572	0.231	0.2311	0.1084	0.1084	0.2567	0.2568
		ξ_4	0.6569	0.6569	0.423	0.423	0.2685	0.2686	0.4518	0.4518
V	Center	ξ_1	0.8021	0.8023	0.6086	0.6087	0.4906	0.4906	0.6219	0.6226
		ξ_2	0.8016	0.8022	0.6096	0.6097	0.4905	0.4907	0.6214	0.6216
		ξ_3	0.8019	0.8024	0.609	0.6095	0.4908	0.491	0.6243	0.6245
		ξ_4	0.2981	0.2983	0.1235	0.1236	0.1158	0.1159	0.1349	0.135
		ξ_5	0.4511	0.4511	0.2311	0.2311	0.1074	0.1075	0.2483	0.2483
		ξ_6	0.4563	0.4563	0.2303	0.2305	0.1094	0.1095	0.2504	0.2505
	Symbolic Cov.	ξ_1	0.8028	0.8032	0.6087	0.6088	0.4907	0.4911	0.6313	0.632
		ξ_2	0.8017	0.8023	0.6082	0.6089	0.489	0.489	0.6319	0.6323
		ξ_3	0.8014	0.8022	0.6082	0.6082	0.4909	0.4911	0.6283	0.6286
		ξ_4	0.3002	0.3003	0.1181	0.1181	0.1171	0.1172	0.1385	0.1385
		ξ_5	0.4514	0.4514	0.2347	0.2349	0.1081	0.1082	0.2604	0.2604
		ξ_6	0.4581	0.4581	0.2306	0.2308	0.1091	0.1092	0.2569	0.2569
VI	Center	ξ_1	0.8021	0.8023	0.6086	0.6086	0.4892	0.4893	0.6251	0.6255
		ξ_2	0.803	0.803	0.6089	0.6096	0.4902	0.4905	0.6221	0.6223
		ξ_3	0.8028	0.803	0.6088	0.6089	0.4903	0.4905	0.6238	0.6241
		ξ_4	0.2992	0.2993	0.1223	0.1223	0.1152	0.1153	0.1315	0.1317
		ξ_5	0.4519	0.4521	0.2324	0.2325	0.1099	0.11	0.2476	0.2478
		ξ_6	0.6021	0.6021	0.3738	0.3739	0.2192	0.2192	0.397	0.3971
	Symbolic Cov.	ξ_1	0.8042	0.8049	0.608	0.6081	0.4906	0.4909	0.6282	0.6287
		ξ_2	0.8025	0.8026	0.6088	0.6088	0.4908	0.4909	0.6286	0.6286
		ξ_3	0.8032	0.8037	0.6084	0.6087	0.49	0.4902	0.6283	0.6288
		ξ_4	0.2974	0.2974	0.1211	0.1212	0.1178	0.1178	0.1405	0.1406
		ξ_5	0.4538	0.4541	0.2319	0.2321	0.1096	0.1096	0.2558	0.2559
		ξ_6	0.6022	0.6022	0.3724	0.3725	0.2212	0.2213	0.3987	0.3987

Table A.46: Average absolute relative bias for studies III-VI with 15 manifest variables

Study	Method	Estimator	Loadings 0.5		Loadings 0.7		Loadings 0.9		Random Loadings	
			Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
III	Center	$\hat{\beta}_{1,3}$	0.0153	0.013	0.0183	0.0162	0.0207	0.0178	0.0269	0.0216
		$\hat{\beta}_{2,3}$	0.0149	0.0129	0.0179	0.0155	0.0201	0.018	0.0267	0.0212
		$\hat{\beta}_{1,4}$	0.0148	0.0132	0.0187	0.0159	0.0218	0.0187	0.0285	0.0219
		$\hat{\beta}_{2,4}$	0.015	0.0132	0.018	0.0158	0.0203	0.0178	0.0279	0.0219
	Symbolic Cov.	$\hat{\beta}_{1,3}$	0.0747	0.0508	0.0483	0.0378	0.0375	0.0296	0.0524	0.0384
		$\hat{\beta}_{2,3}$	0.0816	0.0539	0.0496	0.0369	0.0383	0.0295	0.0517	0.0393
		$\hat{\beta}_{1,4}$	0.0755	0.052	0.0494	0.0368	0.0379	0.0296	0.052	0.0366
		$\hat{\beta}_{2,4}$	0.0786	0.0536	0.0506	0.0369	0.0394	0.0291	0.0512	0.038
IV	Center	$\hat{\beta}_{1,3}$	0.0157	0.0136	0.019	0.0167	0.0209	0.0187	0.0266	0.021
		$\hat{\beta}_{2,3}$	0.0156	0.0133	0.0188	0.016	0.0206	0.0184	0.0267	0.0199
		$\hat{\beta}_{1,4}$	0.1497	0.1132	0.1498	0.1166	0.1513	0.1083	0.1492	0.1116
		$\hat{\beta}_{2,4}$	0.1487	0.112	0.1474	0.1086	0.1498	0.1111	0.1488	0.1133
		$\hat{\beta}_{3,4}$	0.2386	0.2001	0.2487	0.2071	0.2443	0.1958	0.2386	0.1938
	Symbolic Cov.	$\hat{\beta}_{1,3}$	0.0837	0.054	0.0491	0.0367	0.0371	0.0305	0.0504	0.0376
		$\hat{\beta}_{2,3}$	0.0772	0.0517	0.0488	0.0346	0.0386	0.0299	0.052	0.0384
		$\hat{\beta}_{1,4}$	0.7551	0.1881	0.7046	0.2007	0.659	0.2036	0.7092	0.2036
		$\hat{\beta}_{2,4}$	0.755	0.1915	0.7072	0.1987	0.6595	0.2113	0.7153	0.2039
		$\hat{\beta}_{3,4}$	1.0242	0.4803	0.9612	0.4627	0.9019	0.4736	0.9638	0.4712
V	Center	$\hat{\beta}_{1,4}$	0.0485	0.0315	0.0614	0.0406	0.0703	0.0464	0.0648	0.0454
		$\hat{\beta}_{1,5}$	0.0152	0.0133	0.018	0.0159	0.0211	0.0183	0.029	0.0229
		$\hat{\beta}_{2,5}$	0.0149	0.013	0.0184	0.0157	0.0207	0.0176	0.0268	0.0218
		$\hat{\beta}_{2,6}$	0.0152	0.0129	0.0183	0.0159	0.0204	0.0185	0.0264	0.0208
		$\hat{\beta}_{3,6}$	0.0152	0.0128	0.0186	0.0161	0.0197	0.0178	0.0258	0.0217
	Symbolic Cov.	$\hat{\beta}_{1,4}$	0.1001	0.0733	0.1093	0.0768	0.1097	0.0765	0.0992	0.0744
		$\hat{\beta}_{1,5}$	0.0764	0.0509	0.0479	0.0372	0.0374	0.0279	0.0519	0.0379
		$\hat{\beta}_{2,5}$	0.0801	0.0534	0.0494	0.0363	0.0388	0.0295	0.0521	0.0399
		$\hat{\beta}_{2,6}$	0.0781	0.0521	0.0498	0.0379	0.0388	0.0292	0.0524	0.0396
		$\hat{\beta}_{3,6}$	0.0794	0.053	0.0491	0.0374	0.0388	0.0291	0.0507	0.0378
VI	Center	$\hat{\beta}_{1,4}$	0.0505	0.0319	0.0628	0.0399	0.0728	0.0462	0.0645	0.0455
		$\hat{\beta}_{1,5}$	0.0157	0.0132	0.0183	0.0164	0.0215	0.019	0.028	0.0216
		$\hat{\beta}_{2,5}$	0.0158	0.0136	0.0185	0.0163	0.0206	0.0183	0.0274	0.0221
		$\hat{\beta}_{2,6}$	0.0155	0.0107	0.0189	0.012	0.0198	0.0128	0.0255	0.0186
		$\hat{\beta}_{3,6}$	0.0149	0.0099	0.0176	0.0119	0.0197	0.0128	0.0265	0.066
	Symbolic Cov.	$\hat{\beta}_{5,6}$	0.1272	0.0897	0.1288	0.091	0.1267	0.0908	0.1293	0.0928
		$\hat{\beta}_{1,4}$	0.0986	0.0729	0.1067	0.0751	0.1052	0.0759	0.1041	0.0749
		$\hat{\beta}_{1,5}$	0.0792	0.0526	0.0503	0.0367	0.0371	0.0287	0.0517	0.0395
		$\hat{\beta}_{2,5}$	0.0798	0.0532	0.0479	0.0359	0.0381	0.0298	0.0519	0.0384
		$\hat{\beta}_{2,6}$	0.1106	0.0858	0.0859	0.066	0.0712	0.0546	0.09	0.0666
	$\hat{\beta}_{3,6}$	0.1267	0.0716	0.0856	0.0538	0.0661	0.0418	0.091	0.0566	
	$\hat{\beta}_{5,6}$	0.1649	0.1199	0.148	0.1071	0.1459	0.1063	0.1423	0.106	

Table A.47: Average absolute relative bias of the loadings associated with each latent variable, study with 15 manifest variable

Study	Method	Latent Variable	Loadings 0.5		Loadings 0.7		Loadings 0.9		Random Loadings	
			Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
III	Center	ξ_1	0.2182	0.0126	0.0899	0.0132	0.0333	0.0131	0.0989	0.0226
		ξ_2	0.2179	0.0127	0.089	0.0132	0.0325	0.0127	0.0991	0.0222
		ξ_3	0.2051	0.018	0.0767	0.0191	0.0237	0.0129	0.0847	0.0261
		ξ_4	0.2055	0.0181	0.0763	0.019	0.0236	0.0125	0.0838	0.0271
	Symbolic Cov.	ξ_1	0.2044	0.0169	0.0837	0.0145	0.0302	0.0131	0.1045	0.0742
		ξ_2	0.2049	0.0173	0.084	0.0143	0.0304	0.0128	0.105	0.0749
		ξ_3	0.1919	0.0221	0.071	0.0202	0.022	0.0123	0.0926	0.0736
IV	Center	ξ_1	0.2183	0.0125	0.0895	0.0122	0.0333	0.0128	0.0992	0.0225
		ξ_2	0.2182	0.0121	0.0894	0.013	0.0324	0.0127	0.0982	0.0217
		ξ_3	0.2048	0.0185	0.0748	0.0197	0.0238	0.0127	0.0849	0.0264
		ξ_4	0.2414	0.0092	0.1154	0.0086	0.0593	0.0089	0.1254	0.0202
	Symbolic Cov.	ξ_1	0.2042	0.0171	0.0843	0.0142	0.0305	0.0129	0.1049	0.0747
		ξ_2	0.2042	0.0171	0.084	0.0144	0.0302	0.0125	0.1054	0.0747
		ξ_3	0.1924	0.0222	0.0708	0.0193	0.0217	0.0121	0.0923	0.0732
V	Center	ξ_4	0.2295	0.014	0.1106	0.01	0.0567	0.0092	0.1306	0.0731
		ξ_1	0.2182	0.0125	0.0901	0.0128	0.0338	0.0123	0.0988	0.023
		ξ_2	0.218	0.0123	0.0901	0.0124	0.0327	0.0128	0.0978	0.0223
		ξ_3	0.218	0.0122	0.0901	0.0127	0.0328	0.0123	0.0978	0.0223
		ξ_4	0.1624	0.035	0.0435	0.0202	0.0398	0.034	0.0485	0.0297
		ξ_5	0.2057	0.0179	0.0762	0.0187	0.0234	0.0126	0.0844	0.028
	Symbolic Cov.	ξ_6	0.2056	0.0175	0.0758	0.019	0.0238	0.0127	0.0847	0.0268
		ξ_1	0.205	0.0168	0.0841	0.0141	0.0308	0.0127	0.1049	0.0749
		ξ_2	0.2046	0.0168	0.084	0.0141	0.0301	0.0128	0.1056	0.0751
		ξ_3	0.2046	0.0175	0.0843	0.0142	0.0299	0.0129	0.1049	0.0749
		ξ_4	0.1455	0.0386	0.0411	0.0173	0.0442	0.0374	0.0771	0.0591
		ξ_5	0.1921	0.0216	0.0705	0.0196	0.0218	0.012	0.0936	0.0738
		ξ_6	0.191	0.0221	0.0699	0.0208	0.022	0.0122	0.0927	0.0737
VI	Center	ξ_1	0.2183	0.0129	0.0895	0.0129	0.0327	0.013	0.099	0.0228
		ξ_2	0.2182	0.0121	0.0898	0.0126	0.0327	0.0124	0.0982	0.0234
		ξ_3	0.2183	0.0125	0.0906	0.0128	0.0328	0.0125	0.1017	0.0668
		ξ_4	0.1602	0.0346	0.0425	0.0195	0.042	0.0346	0.0474	0.0282
		ξ_5	0.2049	0.0181	0.0754	0.0199	0.023	0.0125	0.0843	0.027
		ξ_6	0.2345	0.0103	0.1077	0.0106	0.0508	0.0107	0.1153	0.0209
	Symbolic Cov.	ξ_1	0.2052	0.0166	0.0841	0.0144	0.0299	0.0129	0.106	0.075
		ξ_2	0.2042	0.0173	0.0842	0.0144	0.029	0.0129	0.1052	0.0746
		ξ_3	0.2043	0.0172	0.0839	0.0146	0.0298	0.0126	0.1037	0.0748
		ξ_4	0.1464	0.0392	0.0409	0.0171	0.0428	0.0364	0.0772	0.0579
		ξ_5	0.1916	0.022	0.0698	0.0201	0.0218	0.0122	0.0941	0.0743
		ξ_6	0.2218	0.015	0.1027	0.0117	0.0484	0.0113	0.123	0.0738

Table A.48: Average root mean squared error for latent variable, study with 15 manifest variables

Study	Method	Latent Variable	Loadings 0.5		Loadings 0.7		Loadings 0.9		Random Loadings	
			$RMSE_L$	$RMSE_U$	$RMSE_L$	$RMSE_U$	$RMSE_L$	$RMSE_U$	$RMSE_L$	$RMSE_U$
III	Center	ξ_1	0.804	0.8046	0.6079	0.6082	0.4905	0.4906	0.6218	0.622
		ξ_2	0.8024	0.8024	0.6099	0.61	0.4899	0.49	0.6229	0.623
		ξ_3	0.4506	0.4508	0.2349	0.2349	0.1086	0.1087	0.2484	0.2485
		ξ_4	0.4521	0.4522	0.233	0.2331	0.1078	0.1078	0.2455	0.2455
	Symbolic Cov.	ξ_1	0.8022	0.8029	0.609	0.6093	0.4911	0.4911	0.6284	0.6284
		ξ_2	0.8025	0.8027	0.6096	0.61	0.4903	0.4904	0.63	0.6302
		ξ_3	0.4543	0.4544	0.2361	0.2362	0.1091	0.1092	0.2566	0.2567
IV	Center	ξ_1	0.8017	0.802	0.6086	0.609	0.4902	0.4904	0.6246	0.6248
		ξ_2	0.8023	0.8028	0.6099	0.6106	0.49	0.4901	0.6208	0.621
		ξ_3	0.449	0.4491	0.228	0.2282	0.1086	0.1087	0.2498	0.2498
		ξ_4	0.6536	0.6536	0.4208	0.4209	0.2657	0.2657	0.445	0.4451
	Symbolic Cov.	ξ_1	0.8031	0.8038	0.609	0.609	0.4898	0.4901	0.628	0.628
		ξ_2	0.8032	0.8033	0.6085	0.6088	0.4894	0.4896	0.6283	0.6288
		ξ_3	0.4568	0.4568	0.2349	0.2349	0.1063	0.1063	0.2531	0.2533
V	Center	ξ_4	0.6576	0.6576	0.4238	0.4238	0.2654	0.2655	0.4486	0.4487
		ξ_1	0.8033	0.8038	0.6095	0.6096	0.4898	0.4898	0.6225	0.6227
		ξ_2	0.8019	0.8029	0.6095	0.6096	0.4895	0.4896	0.6223	0.6225
		ξ_3	0.8016	0.8022	0.6079	0.6083	0.4904	0.4905	0.622	0.6222
		ξ_4	0.3024	0.3026	0.1249	0.1251	0.1144	0.1146	0.1347	0.1348
		ξ_5	0.4534	0.4536	0.2326	0.2326	0.107	0.1072	0.249	0.249
	ξ_6	0.452	0.452	0.231	0.231	0.1097	0.1097	0.2477	0.2477	
Symbolic Cov.	ξ_1	0.8027	0.8032	0.6089	0.6092	0.4896	0.4897	0.6284	0.6285	
	ξ_2	0.8026	0.8026	0.6091	0.6092	0.4909	0.491	0.6298	0.6299	
	ξ_3	0.8027	0.8033	0.6095	0.6096	0.4917	0.492	0.6312	0.6317	
	ξ_4	0.3003	0.3005	0.1186	0.1186	0.1178	0.1178	0.1408	0.1409	
	ξ_5	0.4539	0.4539	0.2332	0.2333	0.1076	0.1077	0.2584	0.2584	
	ξ_6	0.4509	0.451	0.2321	0.2321	0.1078	0.1079	0.2565	0.2566	
VI	Center	ξ_1	0.8014	0.8017	0.6093	0.6097	0.4907	0.4909	0.6231	0.6234
		ξ_2	0.8027	0.803	0.6078	0.6084	0.4904	0.4905	0.6213	0.6215
		ξ_3	0.8041	0.8046	0.6097	0.6098	0.4909	0.4909	0.6231	0.6237
		ξ_4	0.2951	0.2954	0.1213	0.1214	0.1181	0.1182	0.1329	0.1331
		ξ_5	0.4491	0.4491	0.2303	0.2305	0.1063	0.1063	0.247	0.2471
		ξ_6	0.6037	0.6038	0.3738	0.3739	0.2174	0.2175	0.3885	0.3886
	Symbolic Cov.	ξ_1	0.8029	0.8035	0.6089	0.6092	0.49	0.49	0.6317	0.6318
ξ_2		0.8021	0.8024	0.6086	0.6087	0.4925	0.4925	0.6276	0.6277	
ξ_3		0.8029	0.803	0.6091	0.6096	0.4903	0.4903	0.627	0.6272	
ξ_4		0.3036	0.3037	0.1198	0.1198	0.1161	0.1161	0.1375	0.1376	
ξ_5		0.4528	0.4529	0.231	0.2311	0.1069	0.107	0.2609	0.2609	
ξ_6		0.6045	0.6047	0.3747	0.3748	0.2194	0.2195	0.4014	0.4015	

Table A.49: Average absolute relative bias for studies III-VI with 20 manifest variables

Study	Method	Estimator	Loadings 0.5		Loadings 0.7		Loadings 0.9		Random Loadings	
			Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
III	Center	$\hat{\beta}_{1,3}$	0.0141	0.0126	0.0184	0.0163	0.0204	0.0184	0.025	0.0195
		$\hat{\beta}_{2,3}$	0.0146	0.0127	0.0187	0.0164	0.021	0.0193	0.0251	0.0197
		$\hat{\beta}_{1,4}$	0.0144	0.0125	0.0183	0.0165	0.0211	0.0189	0.0252	0.02
		$\hat{\beta}_{2,4}$	0.0144	0.0126	0.0185	0.0162	0.0215	0.0194	0.0259	0.02
	Symbolic Cov.	$\hat{\beta}_{1,3}$	0.0789	0.0521	0.05	0.0363	0.0394	0.0292	0.0504	0.0376
		$\hat{\beta}_{2,3}$	0.0821	0.0523	0.0475	0.0358	0.0374	0.0294	0.0529	0.0394
		$\hat{\beta}_{1,4}$	0.079	0.0513	0.0519	0.0377	0.0399	0.0306	0.0508	0.0381
		$\hat{\beta}_{2,4}$	0.0784	0.0528	0.0479	0.0357	0.0378	0.0291	0.053	0.0385
IV	Center	$\hat{\beta}_{1,3}$	0.0147	0.0126	0.0185	0.0163	0.0203	0.018	0.0255	0.0191
		$\hat{\beta}_{2,3}$	0.0144	0.0125	0.0184	0.0163	0.0201	0.0186	0.0251	0.0202
		$\hat{\beta}_{1,4}$	0.1523	0.117	0.1474	0.117	0.1433	0.1022	0.1448	0.1119
		$\hat{\beta}_{2,4}$	0.1492	0.1099	0.1488	0.1137	0.1481	0.1131	0.1494	0.117
	Symbolic Cov.	$\hat{\beta}_{3,4}$	0.2563	0.2117	0.2494	0.2098	0.2392	0.1966	0.247	0.2063
		$\hat{\beta}_{1,3}$	0.0766	0.0507	0.048	0.0374	0.0389	0.0299	0.052	0.0385
		$\hat{\beta}_{2,3}$	0.0797	0.0546	0.05	0.0379	0.0386	0.0292	0.0527	0.0374
		$\hat{\beta}_{1,4}$	0.7489	0.1891	0.7118	0.203	0.6573	0.2105	0.7087	0.2058
V	Center	$\hat{\beta}_{2,4}$	0.7454	0.1919	0.6954	0.2165	0.6651	0.2088	0.7145	0.2112
		$\hat{\beta}_{3,4}$	0.9881	0.4513	0.933	0.4659	0.8807	0.4566	0.9463	0.4807
		$\hat{\beta}_{1,4}$	0.0492	0.0319	0.0637	0.0408	0.0707	0.0461	0.0661	0.0455
		$\hat{\beta}_{1,5}$	0.015	0.0129	0.0189	0.0165	0.0212	0.0185	0.0244	0.0199
	Symbolic Cov.	$\hat{\beta}_{2,5}$	0.0149	0.0124	0.019	0.0166	0.0205	0.018	0.025	0.0196
		$\hat{\beta}_{2,6}$	0.0147	0.0128	0.0178	0.0156	0.0207	0.0182	0.0247	0.0192
		$\hat{\beta}_{3,6}$	0.0143	0.013	0.0177	0.0157	0.0212	0.0188	0.0248	0.0197
		$\hat{\beta}_{1,4}$	0.1011	0.0748	0.1043	0.074	0.1079	0.0759	0.1036	0.0756
VI	Center	$\hat{\beta}_{1,5}$	0.0799	0.0541	0.0498	0.0372	0.0398	0.03	0.0525	0.0378
		$\hat{\beta}_{2,5}$	0.077	0.0533	0.0491	0.0369	0.0383	0.031	0.0544	0.0394
		$\hat{\beta}_{2,6}$	0.0777	0.0525	0.0506	0.0383	0.0356	0.0294	0.0531	0.0399
		$\hat{\beta}_{3,6}$	0.0768	0.0532	0.0481	0.0354	0.0384	0.0308	0.0531	0.0395
	Symbolic Cov.	$\hat{\beta}_{1,4}$	0.0505	0.0326	0.0635	0.0406	0.0702	0.0462	0.0643	0.0439
		$\hat{\beta}_{1,5}$	0.015	0.0134	0.0181	0.016	0.0204	0.0184	0.0251	0.0198
		$\hat{\beta}_{2,5}$	0.0147	0.013	0.0179	0.0158	0.0205	0.0181	0.0237	0.019
		$\hat{\beta}_{2,6}$	0.0153	0.0102	0.0183	0.0124	0.0196	0.0134	0.0232	0.0169
Symbolic Cov.	$\hat{\beta}_{3,6}$	0.0152	0.0099	0.0179	0.0118	0.0197	0.0131	0.0229	0.0167	
	$\hat{\beta}_{5,6}$	0.1234	0.088	0.1225	0.0898	0.1282	0.0952	0.1263	0.0912	
	$\hat{\beta}_{1,4}$	0.0969	0.0726	0.1053	0.0751	0.103	0.0744	0.1002	0.0733	
	$\hat{\beta}_{1,5}$	0.0779	0.0512	0.0491	0.0353	0.0389	0.0289	0.053	0.0378	
Symbolic Cov.	$\hat{\beta}_{2,5}$	0.0774	0.0522	0.0483	0.0358	0.037	0.0291	0.0532	0.0377	
	$\hat{\beta}_{2,6}$	0.112	0.0804	0.0869	0.0685	0.0692	0.0535	0.0854	0.0662	
	$\hat{\beta}_{3,6}$	0.1269	0.0738	0.0877	0.0555	0.0658	0.0435	0.0881	0.0557	
	$\hat{\beta}_{5,6}$	0.1724	0.1253	0.1505	0.1123	0.1444	0.1061	0.1519	0.1111	

Table A.50: Average absolute relative bias of the loadings associated with each latent variable, study with 20 manifest variable

Study	Method	Latent Variable	Loadings 0.5		Loadings 0.7		Loadings 0.9		Random Loadings	
			Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
III	Center	ξ_1	0.2177	0.0123	0.0896	0.0131	0.0331	0.0126	0.0978	0.0207
		ξ_2	0.2179	0.0124	0.0907	0.0126	0.033	0.0122	0.0985	0.0202
		ξ_3	0.2066	0.0176	0.0757	0.0193	0.0239	0.0125	0.0845	0.0256
		ξ_4	0.206	0.0174	0.0763	0.019	0.0234	0.0126	0.0835	0.0252
	Symbolic Cov.	ξ_1	0.2039	0.0168	0.0833	0.0139	0.0299	0.0131	0.1044	0.0746
		ξ_2	0.2045	0.0173	0.0844	0.014	0.0302	0.0128	0.1041	0.0745
		ξ_3	0.1918	0.022	0.0697	0.0207	0.0221	0.0123	0.0938	0.0732
IV	Center	ξ_1	0.2181	0.0123	0.0894	0.0128	0.0329	0.0124	0.0987	0.0198
		ξ_2	0.2183	0.0123	0.0898	0.0131	0.033	0.0123	0.0979	0.0204
		ξ_3	0.2063	0.0173	0.0761	0.0192	0.0239	0.0124	0.0844	0.0252
		ξ_4	0.2417	0.0086	0.1157	0.0082	0.0596	0.0085	0.1241	0.0173
	Symbolic Cov.	ξ_1	0.2043	0.0171	0.0842	0.0138	0.0297	0.0127	0.1044	0.0749
		ξ_2	0.2044	0.0171	0.0847	0.0145	0.0298	0.0128	0.1055	0.0751
		ξ_3	0.192	0.0215	0.07	0.0203	0.0219	0.0123	0.0935	0.0738
V	Center	ξ_4	0.2291	0.0138	0.1103	0.01	0.0568	0.0092	0.131	0.0732
		ξ_1	0.2183	0.0122	0.0901	0.0124	0.0336	0.0128	0.0987	0.021
		ξ_2	0.2186	0.0125	0.09	0.013	0.0331	0.0128	0.0993	0.0205
		ξ_3	0.219	0.0123	0.09	0.013	0.0331	0.0126	0.0999	0.0204
		ξ_4	0.162	0.0352	0.0424	0.0196	0.0402	0.034	0.0465	0.0281
		ξ_6	0.2067	0.0175	0.0763	0.018	0.0236	0.0129	0.085	0.0243
	Symbolic Cov.	ξ_1	0.2045	0.017	0.0848	0.0138	0.0299	0.013	0.1047	0.0748
		ξ_2	0.2048	0.0168	0.0845	0.0145	0.0303	0.0127	0.1043	0.0749
		ξ_3	0.2048	0.0169	0.0841	0.0141	0.03	0.0128	0.1045	0.0747
		ξ_4	0.1454	0.0387	0.0411	0.0175	0.0443	0.0373	0.0771	0.0582
		ξ_5	0.1913	0.0219	0.0692	0.021	0.0217	0.0124	0.0932	0.0738
		ξ_6	0.1919	0.0217	0.0699	0.0196	0.0217	0.0118	0.0938	0.0738
		VI	Center	ξ_1	0.2183	0.0123	0.0896	0.0134	0.0329	0.0127
ξ_2	0.2181			0.0123	0.0899	0.013	0.0331	0.0126	0.0976	0.0196
ξ_3	0.2179			0.0128	0.09	0.0128	0.0334	0.0125	0.0993	0.0211
ξ_4	0.1608			0.0353	0.042	0.0195	0.0401	0.0336	0.0464	0.0285
ξ_5	0.2052			0.0179	0.0761	0.0188	0.0245	0.0122	0.0852	0.0247
ξ_6	0.2341			0.0102	0.1075	0.0106	0.0517	0.0104	0.1153	0.0193
Symbolic Cov.	ξ_1		0.2043	0.0172	0.085	0.0142	0.0298	0.0128	0.105	0.0751
	ξ_2		0.2041	0.017	0.0839	0.0144	0.0306	0.0127	0.1048	0.0749
	ξ_3		0.2042	0.017	0.0836	0.0144	0.0306	0.0127	0.1056	0.0745
	ξ_4		0.1465	0.039	0.0412	0.0178	0.0416	0.0362	0.0772	0.0591
	ξ_5		0.1914	0.0217	0.07	0.0201	0.0219	0.0121	0.0936	0.0739
	ξ_6		0.2216	0.0148	0.1019	0.0119	0.0484	0.0112	0.123	0.0739

Table A.51: Average root mean squared error for latent variable, study with 20 manifest variables

Study	Method	Latent Variable	Loadings 0.5		Loadings 0.7		Loadings 0.9		Random Loadings	
			$RMSE_L$	$RMSE_U$	$RMSE_L$	$RMSE_U$	$RMSE_L$	$RMSE_U$	$RMSE_L$	$RMSE_U$
III	Center	ξ_1	0.8015	0.8022	0.6103	0.6107	0.4905	0.4908	0.6234	0.6234
		ξ_2	0.8016	0.8021	0.6083	0.6085	0.4894	0.4895	0.6222	0.6226
		ξ_3	0.4567	0.4568	0.2307	0.2308	0.1082	0.1083	0.2493	0.2494
		ξ_4	0.4539	0.4541	0.233	0.2331	0.1068	0.1069	0.2459	0.246
	Symbolic Cov.	ξ_1	0.8029	0.803	0.6097	0.6098	0.4902	0.4902	0.6267	0.627
		ξ_2	0.8031	0.8033	0.6096	0.61	0.4888	0.489	0.6272	0.6272
		ξ_3	0.4543	0.4545	0.2308	0.2308	0.1071	0.1071	0.2588	0.259
		ξ_4	0.4544	0.4545	0.2311	0.2312	0.1079	0.108	0.2595	0.2596
IV	Center	ξ_1	0.8028	0.8034	0.609	0.6092	0.4902	0.4905	0.6223	0.6228
		ξ_2	0.8032	0.8036	0.609	0.6097	0.4897	0.4899	0.6226	0.6229
		ξ_3	0.4552	0.4553	0.232	0.2322	0.1092	0.1093	0.2486	0.2486
		ξ_4	0.6547	0.6548	0.4231	0.4232	0.2677	0.2678	0.4426	0.4427
	Symbolic Cov.	ξ_1	0.804	0.8041	0.6097	0.6101	0.491	0.4912	0.6296	0.6299
		ξ_2	0.8031	0.8036	0.6089	0.6095	0.4902	0.4903	0.6289	0.629
		ξ_3	0.4531	0.4531	0.2322	0.2324	0.1079	0.108	0.2581	0.2581
		ξ_4	0.6542	0.6542	0.4214	0.4215	0.2669	0.2669	0.4529	0.453
V	Center	ξ_1	0.8033	0.8037	0.6079	0.6083	0.4895	0.4897	0.6212	0.6212
		ξ_2	0.8031	0.8034	0.6091	0.6094	0.4906	0.4906	0.6243	0.6245
		ξ_3	0.8029	0.8035	0.6093	0.6094	0.4906	0.4906	0.6244	0.6246
		ξ_4	0.3012	0.3012	0.1203	0.1206	0.1148	0.1149	0.1293	0.1294
		ξ_5	0.454	0.454	0.2282	0.2283	0.1065	0.1066	0.2516	0.2516
		ξ_6	0.4571	0.4571	0.2327	0.2327	0.1083	0.1083	0.2488	0.2489
	Symbolic Cov.	ξ_1	0.803	0.8035	0.608	0.6087	0.4913	0.4913	0.6291	0.6295
		ξ_2	0.802	0.8027	0.6087	0.6089	0.4896	0.49	0.6299	0.63
		ξ_3	0.8025	0.8027	0.608	0.6083	0.4915	0.4916	0.6284	0.6286
		ξ_4	0.3006	0.3007	0.1209	0.121	0.1183	0.1184	0.1361	0.1362
		ξ_5	0.4528	0.4528	0.2288	0.2289	0.1062	0.1063	0.2577	0.2577
		ξ_6	0.4535	0.4537	0.2305	0.2306	0.1065	0.1067	0.2607	0.2607
VI	Center	ξ_1	0.8038	0.8039	0.609	0.6095	0.4897	0.49	0.6214	0.6214
		ξ_2	0.8013	0.8029	0.6087	0.6097	0.4896	0.4897	0.6214	0.6216
		ξ_3	0.803	0.8034	0.6069	0.607	0.4899	0.49	0.623	0.6235
		ξ_4	0.2974	0.2975	0.1202	0.1203	0.1155	0.1155	0.1303	0.1303
		ξ_5	0.4503	0.4505	0.2322	0.2324	0.1112	0.1112	0.2516	0.2517
		ξ_6	0.6012	0.6012	0.3716	0.3717	0.2223	0.2224	0.3904	0.3905
	Symbolic Cov.	ξ_1	0.8024	0.8031	0.6089	0.6093	0.4906	0.4909	0.6302	0.6305
		ξ_2	0.8015	0.8016	0.6084	0.6089	0.4889	0.4889	0.629	0.629
		ξ_3	0.802	0.8028	0.608	0.6082	0.4892	0.4892	0.6288	0.6288
		ξ_4	0.305	0.3052	0.1213	0.1214	0.1151	0.1151	0.1403	0.1403
		ξ_5	0.4508	0.451	0.2313	0.2313	0.1082	0.1083	0.2574	0.2574
		ξ_6	0.6005	0.6006	0.3707	0.3708	0.2188	0.2189	0.4014	0.4015

A.3 Simulation 3 - Range

Table A.52: Average absolute relative bias for studies III-VI with different values of σ

Study	Method	Estimator	$\sigma = 0.1$		$\sigma = 0.5$		$\sigma = 1$		$\sigma = 1.5$		$\sigma = 2$	
			Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
III	Center	$\hat{\beta}_{1,3}$	0.0426	0.0116	0.0289	0.0156	0.0319	0.0244	0.0622	0.0484	0.0803	0.0593
		$\hat{\beta}_{2,3}$	0.0429	0.0117	0.029	0.0157	0.0332	0.0259	0.0641	0.0475	0.0865	0.0669
		$\hat{\beta}_{1,4}$	0.0426	0.0116	0.0278	0.0156	0.0328	0.0247	0.0617	0.0475	0.0795	0.0585
		$\hat{\beta}_{2,4}$	0.0428	0.0118	0.0277	0.0154	0.0333	0.0264	0.0626	0.0468	0.0857	0.0631
	Symbolic Cov.	$\hat{\beta}_{1,3}$	0.0612	0.0197	0.0545	0.0275	0.0506	0.0387	0.067	0.0518	0.0965	0.0776
		$\hat{\beta}_{2,3}$	0.06	0.0195	0.054	0.0257	0.0503	0.0382	0.0687	0.0517	0.1013	0.082
		$\hat{\beta}_{1,4}$	0.0614	0.0197	0.055	0.0275	0.0542	0.0388	0.066	0.0532	0.0941	0.0776
		$\hat{\beta}_{2,4}$	0.0601	0.0193	0.054	0.0262	0.0504	0.0386	0.0664	0.054	0.0937	0.0796
IV	Center	$\hat{\beta}_{1,3}$	0.0428	0.0116	0.0289	0.0158	0.0321	0.0256	0.064	0.0493	0.086	0.0637
		$\hat{\beta}_{2,3}$	0.0425	0.0117	0.0281	0.0157	0.0306	0.0245	0.0631	0.0472	0.0836	0.0581
		$\hat{\beta}_{1,4}$	0.7595	0.2231	0.3428	0.1597	0.154	0.1148	0.1284	0.1028	0.1309	0.1178
		$\hat{\beta}_{2,4}$	0.7426	0.2173	0.3453	0.1624	0.1581	0.1154	0.1225	0.1013	0.1329	0.1179
		$\hat{\beta}_{3,4}$	0.957	0.4414	0.4027	0.3054	0.2555	0.2191	0.2397	0.2012	0.2426	0.2046
	Symbolic Cov.	$\hat{\beta}_{1,3}$	0.0621	0.0204	0.0536	0.0267	0.0548	0.0402	0.0701	0.056	0.0948	0.0757
		$\hat{\beta}_{2,3}$	0.0616	0.0201	0.0554	0.0271	0.0528	0.0397	0.0634	0.05	0.0968	0.0807
		$\hat{\beta}_{1,4}$	0.9104	0.2353	0.7254	0.2091	0.7039	0.213	0.6869	0.2131	0.6848	0.207
		$\hat{\beta}_{2,4}$	0.9338	0.2472	0.7271	0.2195	0.6968	0.2127	0.6934	0.2083	0.6782	0.2042
		$\hat{\beta}_{3,4}$	1.1824	0.467	0.941	0.4595	0.9474	0.4823	0.9937	0.4917	1.0748	0.4817
V	Center	$\hat{\beta}_{1,4}$	0.0412	0.0097	0.019	0.0125	0.0687	0.05	0.1136	0.0703	0.1223	0.0803
		$\hat{\beta}_{1,5}$	0.043	0.0122	0.0287	0.0156	0.0324	0.0262	0.0636	0.0469	0.0847	0.0644
		$\hat{\beta}_{2,5}$	0.0431	0.0118	0.029	0.0154	0.0322	0.0251	0.0656	0.049	0.0856	0.0615
		$\hat{\beta}_{2,6}$	0.0429	0.0118	0.029	0.0154	0.0321	0.0253	0.0674	0.0512	0.0854	0.065
		$\hat{\beta}_{3,6}$	0.0427	0.0115	0.0283	0.015	0.0317	0.0249	0.0668	0.0505	0.0879	0.0661
	Symbolic Cov.	$\hat{\beta}_{1,4}$	0.0574	0.0133	0.0267	0.0158	0.101	0.073	0.2682	0.1372	0.4565	0.1997
		$\hat{\beta}_{1,5}$	0.0616	0.0192	0.0539	0.0278	0.0517	0.0392	0.0658	0.052	0.0998	0.0814
		$\hat{\beta}_{2,5}$	0.0612	0.0198	0.0534	0.0262	0.0533	0.038	0.0658	0.0528	0.0966	0.079
		$\hat{\beta}_{2,6}$	0.0606	0.0198	0.0535	0.0263	0.0515	0.0399	0.0674	0.0531	0.0963	0.0757
		$\hat{\beta}_{3,6}$	0.0608	0.0205	0.0562	0.0269	0.0526	0.0383	0.0684	0.0521	0.0989	0.0791
VI	Center	$\hat{\beta}_{1,4}$	0.0405	0.0091	0.0192	0.0128	0.0686	0.0496	0.1127	0.0714	0.1212	0.0804
		$\hat{\beta}_{1,5}$	0.0422	0.0114	0.029	0.0156	0.0324	0.0267	0.0652	0.0487	0.0853	0.063
		$\hat{\beta}_{2,5}$	0.0424	0.0116	0.0283	0.0158	0.0317	0.0247	0.0667	0.0486	0.0855	0.0657
		$\hat{\beta}_{2,6}$	0.0496	0.0127	0.0419	0.0163	0.0297	0.0225	0.0461	0.0361	0.0633	0.0486
		$\hat{\beta}_{3,6}$	0.0502	0.0126	0.0424	0.0162	0.0297	0.0217	0.0444	0.035	0.0669	0.1021
	Symbolic Cov.	$\hat{\beta}_{5,6}$	0.1388	0.1021	0.1325	0.1003	0.1311	0.0922	0.1413	0.0998	0.2056	0.1419
		$\hat{\beta}_{1,4}$	0.0574	0.0142	0.0254	0.0155	0.1003	0.0736	0.2668	0.1362	0.4524	0.2041
		$\hat{\beta}_{1,5}$	0.0615	0.0205	0.0539	0.0267	0.0508	0.0386	0.0648	0.0526	0.0951	0.0759
		$\hat{\beta}_{2,5}$	0.0605	0.0196	0.0545	0.0274	0.0538	0.039	0.0669	0.053	0.0948	0.0808
		$\hat{\beta}_{2,6}$	0.0765	0.0277	0.0742	0.0447	0.0893	0.0659	0.1156	0.0933	0.1409	0.1199
		$\hat{\beta}_{3,6}$	0.0709	0.0219	0.0706	0.0314	0.092	0.0582	0.1083	0.0757	0.1272	0.0915
		$\hat{\beta}_{5,6}$	0.135	0.1033	0.1408	0.1046	0.1545	0.1128	0.173	0.1296	0.1823	0.1361

Table A.53: Average absolute relative bias of the loadings associated with each latent variable, study with different σ 's

Study	Method	Latent Variable	$\sigma = 0.1$		$\sigma = 0.5$		$\sigma = 1$		$\sigma = 1.5$		$\sigma = 2$	
			Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
III	Center	ξ_1	0.0554	0.0129	0.0182	0.0132	0.0987	0.0281	0.2607	0.0483	0.4473	0.0672
		ξ_2	0.0555	0.0131	0.018	0.0135	0.1002	0.0281	0.2614	0.0469	0.4552	0.0748
		ξ_3	0.0133	0.0033	0.0136	0.0086	0.0848	0.0316	0.2036	0.0556	0.3619	0.0779
		ξ_4	0.0134	0.0033	0.0125	0.008	0.0854	0.0324	0.2048	0.0545	0.3611	0.0752
	Symbolic Cov.	ξ_1	0.0558	0.0127	0.0254	0.0154	0.1064	0.0751	0.2751	0.1406	0.4755	0.2104
		ξ_2	0.055	0.0129	0.0252	0.0153	0.1067	0.0754	0.2803	0.1419	0.4765	0.2093
		ξ_3	0.0134	0.0033	0.0197	0.0164	0.0918	0.074	0.2171	0.1468	0.3836	0.2209
		ξ_4	0.0134	0.0034	0.0197	0.0168	0.0942	0.0742	0.2197	0.147	0.3799	0.218
IV	Center	ξ_1	0.0554	0.0128	0.0183	0.0135	0.099	0.0287	0.2612	0.0469	0.4551	0.0706
		ξ_2	0.0553	0.0128	0.0179	0.0129	0.0978	0.0284	0.2617	0.0484	0.4511	0.0703
		ξ_3	0.0133	0.0032	0.0131	0.0083	0.0853	0.0312	0.2032	0.0558	0.3604	0.0761
		ξ_4	0.0041	0.0019	0.0302	0.0077	0.1254	0.0262	0.2632	0.0477	0.4288	0.0728
	Symbolic Cov.	ξ_1	0.0559	0.0133	0.0252	0.0155	0.1043	0.075	0.2731	0.1398	0.4762	0.209
		ξ_2	0.0556	0.0134	0.0249	0.015	0.1049	0.0754	0.2739	0.1409	0.4754	0.2093
		ξ_3	0.0133	0.0033	0.0198	0.0167	0.0957	0.0741	0.2154	0.1472	0.3787	0.2188
		ξ_4	0.0041	0.0021	0.0313	0.0208	0.1301	0.0734	0.2785	0.1419	0.45	0.2129
V	Center	ξ_1	0.0558	0.0135	0.0183	0.0129	0.1005	0.0284	0.2612	0.0492	0.4514	0.0694
		ξ_2	0.0558	0.0129	0.0185	0.013	0.0989	0.0273	0.2637	0.0507	0.452	0.0722
		ξ_3	0.0555	0.0129	0.0176	0.0123	0.099	0.0283	0.2612	0.0483	0.4561	0.0721
		ξ_4	0.0153	0.0042	0.015	0.0117	0.0508	0.0329	0.1402	0.0668	0.3058	0.0876
		ξ_5	0.0133	0.0032	0.0132	0.0085	0.0857	0.032	0.2041	0.0563	0.3622	0.0808
		ξ_6	0.0135	0.0032	0.0135	0.0087	0.0858	0.0318	0.2001	0.0561	0.3601	0.0777
	Symbolic Cov.	ξ_1	0.0558	0.0127	0.0253	0.0153	0.1061	0.0755	0.2765	0.1406	0.4736	0.2076
		ξ_2	0.0549	0.0131	0.0251	0.0149	0.106	0.0752	0.2762	0.1404	0.4727	0.2102
		ξ_3	0.0554	0.0132	0.0251	0.0155	0.1049	0.0751	0.2781	0.1413	0.4719	0.2076
		ξ_4	0.0152	0.004	0.0226	0.0161	0.078	0.0592	0.1686	0.1387	0.3125	0.2229
		ξ_5	0.0132	0.0032	0.0195	0.0166	0.093	0.0737	0.2197	0.1473	0.3796	0.2184
		ξ_6	0.0133	0.0033	0.0196	0.0169	0.093	0.0737	0.2169	0.1471	0.3784	0.2186
VI	Center	ξ_1	0.0548	0.0126	0.0178	0.0129	0.0993	0.0287	0.262	0.0475	0.4511	0.0724
		ξ_2	0.0549	0.0129	0.0173	0.0131	0.0989	0.0282	0.2639	0.0491	0.4516	0.0704
		ξ_3	0.0557	0.0129	0.0181	0.0131	0.1000	0.0271	0.2607	0.0482	0.4542	0.1124
		ξ_4	0.015	0.004	0.0145	0.0118	0.0496	0.0329	0.1426	0.0682	0.3066	0.0868
		ξ_5	0.0132	0.0032	0.0138	0.0086	0.0858	0.0317	0.202	0.0558	0.3606	0.0817
		ξ_6	0.0059	0.0023	0.0273	0.0084	0.1171	0.0267	0.2497	0.0506	0.4074	0.0729
	Symbolic Cov.	ξ_1	0.0558	0.0135	0.0247	0.015	0.1038	0.0749	0.2756	0.1396	0.4739	0.2081
		ξ_2	0.0545	0.0126	0.0252	0.0153	0.1053	0.0754	0.2788	0.1411	0.4772	0.2087
		ξ_3	0.0557	0.0137	0.0253	0.0155	0.1055	0.075	0.2771	0.141	0.4761	0.2091
		ξ_4	0.0153	0.0043	0.0224	0.0161	0.0767	0.059	0.1704	0.1392	0.3094	0.2226
		ξ_5	0.0132	0.0032	0.0202	0.0169	0.0931	0.0736	0.2187	0.1472	0.3792	0.2188
		ξ_6	0.0057	0.0024	0.0287	0.021	0.1225	0.0741	0.2629	0.1425	0.4324	0.211

Table A.54: Average root mean squared error for the latent variables, study with different σ 's

Study	Method	Latent Variable	$\sigma = 0.1$		$\sigma = 0.5$		$\sigma = 1$		$\sigma = 1.5$		$\sigma = 2$	
			<i>RMSE_L</i>	<i>RMSE_U</i>	<i>RMSE_L</i>	<i>RMSE_U</i>	<i>RMSE_L</i>	<i>RMSE_U</i>	<i>RMSE_L</i>	<i>RMSE_U</i>	<i>RMSE_L</i>	<i>RMSE_U</i>
III	Center	ξ_1	0.3114	0.3115	0.3689	0.3691	0.6215	0.6215	0.8474	0.8487	1.0093	1.0096
		ξ_2	0.3114	0.3122	0.3705	0.3706	0.6237	0.6239	0.8494	0.8496	1.0132	1.0133
		ξ_3	0.1552	0.1553	0.1404	0.1406	0.2473	0.2473	0.332	0.3321	0.3712	0.3713
		ξ_4	0.1554	0.1555	0.1345	0.1346	0.2484	0.2485	0.3342	0.3342	0.3687	0.3688
	Symbolic Cov.	ξ_1	0.3122	0.3123	0.3705	0.3708	0.6307	0.6308	0.8584	0.8591	1.026	1.0274
		ξ_2	0.3102	0.3107	0.3700	0.3700	0.6303	0.6303	0.8656	0.8663	1.0274	1.0282
		ξ_3	0.1556	0.1556	0.1383	0.1384	0.2507	0.2507	0.3484	0.3485	0.3867	0.3868
		ξ_4	0.1555	0.1555	0.1403	0.1403	0.2595	0.2596	0.349	0.3491	0.39	0.3901
IV	Center	ξ_1	0.3114	0.3116	0.3682	0.3682	0.6221	0.6223	0.849	0.8492	1.0142	1.0142
		ξ_2	0.3112	0.3113	0.3697	0.3699	0.6212	0.6214	0.8486	0.8494	1.0118	1.0128
		ξ_3	0.155	0.1551	0.1366	0.1366	0.2502	0.2502	0.3302	0.3303	0.3682	0.3685
		ξ_4	0.0881	0.0881	0.2336	0.2336	0.441	0.441	0.5578	0.5578	0.6078	0.6079
	Symbolic Cov.	ξ_1	0.3121	0.3125	0.3714	0.3714	0.6295	0.6298	0.857	0.8576	1.0278	1.0285
		ξ_2	0.3118	0.312	0.3696	0.3697	0.6281	0.6286	0.8576	0.8577	1.0258	1.0258
		ξ_3	0.1554	0.1555	0.1401	0.1403	0.2613	0.2614	0.3416	0.3417	0.3864	0.3864
		ξ_4	0.0886	0.0886	0.2353	0.2353	0.4483	0.4484	0.5742	0.5742	0.6263	0.6263
V	Center	ξ_1	0.3121	0.3125	0.3682	0.3685	0.6242	0.6244	0.8495	0.8499	1.0116	1.0118
		ξ_2	0.3125	0.3125	0.3706	0.3706	0.6222	0.6223	0.8506	0.8506	1.0097	1.0118
		ξ_3	0.3118	0.312	0.3668	0.3669	0.6228	0.6231	0.849	0.8492	1.0144	1.0145
		ξ_4	0.1663	0.1664	0.1243	0.1243	0.1376	0.1376	0.1861	0.1862	0.2365	0.2366
		ξ_5	0.1553	0.1554	0.1375	0.1376	0.2514	0.2514	0.3304	0.3305	0.3723	0.3724
		ξ_6	0.1562	0.1562	0.1372	0.1373	0.2501	0.2501	0.3275	0.3279	0.3679	0.3679
	Symbolic Cov.	ξ_1	0.3121	0.3124	0.3698	0.3698	0.6314	0.6321	0.8617	0.8622	1.0267	1.0278
		ξ_2	0.31	0.3101	0.3719	0.3719	0.6319	0.632	0.8613	0.863	1.0268	1.027
		ξ_3	0.3112	0.3113	0.3698	0.37	0.6291	0.6293	0.8636	0.8645	1.024	1.0244
		ξ_4	0.1661	0.1662	0.1246	0.1247	0.1428	0.1429	0.1995	0.1995	0.2472	0.2474
		ξ_5	0.1547	0.1548	0.1386	0.1386	0.2543	0.2544	0.3493	0.3494	0.3861	0.3863
		ξ_6	0.1549	0.155	0.1417	0.1417	0.2554	0.2554	0.3445	0.3446	0.3858	0.3861
VI	Center	ξ_1	0.31	0.3102	0.3686	0.3686	0.6221	0.6225	0.8492	0.8492	1.0112	1.0119
		ξ_2	0.3102	0.3105	0.3676	0.3677	0.6232	0.6234	0.8522	0.8524	1.011	1.0112
		ξ_3	0.3121	0.3124	0.3698	0.3698	0.6227	0.6232	0.8479	0.8484	1.0139	1.0149
		ξ_4	0.1648	0.1648	0.1236	0.1237	0.1348	0.1351	0.1884	0.1885	0.237	0.237
		ξ_5	0.1546	0.1546	0.1383	0.1385	0.249	0.2493	0.3263	0.3264	0.3683	0.3684
		ξ_6	0.1022	0.1023	0.2071	0.2071	0.3908	0.3909	0.4944	0.4944	0.5318	0.5318
	Symbolic Cov.	ξ_1	0.312	0.3121	0.3695	0.3699	0.6264	0.6264	0.8603	0.8616	1.028	1.0284
		ξ_2	0.3091	0.3092	0.3698	0.3699	0.6278	0.6287	0.8633	0.8633	1.0277	1.0285
		ξ_3	0.3115	0.3119	0.3701	0.3701	0.6293	0.6298	0.8603	0.8605	1.0259	1.026
		ξ_4	0.1661	0.1661	0.1235	0.1236	0.1423	0.1423	0.2028	0.203	0.2476	0.2478
		ξ_5	0.1547	0.1547	0.1404	0.1404	0.2556	0.2556	0.3488	0.349	0.3861	0.3861
		ξ_6	0.1011	0.1011	0.209	0.2091	0.4002	0.4003	0.5079	0.508	0.5564	0.5564

A.4 Simulation 4 - Sample Size

Table A.55: Average absolute relative bias for Study III, for different sample sizes

Method	Estimator	$n = 10$		$n = 20$		$n = 30$		$n = 50$	
		Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
Center	$\hat{\beta}_{1,3}$	0.0495	0.0765	0.0382	0.0295	0.0374	0.0286	0.0342	0.0256
	$\hat{\beta}_{2,3}$	0.0493	0.0721	0.0384	0.029	0.0384	0.0309	0.0343	0.0262
	$\hat{\beta}_{1,4}$	0.0507	0.0783	0.0376	0.0298	0.0361	0.0287	0.0333	0.0266
	$\hat{\beta}_{2,4}$	0.0485	0.0720	0.0393	0.0308	0.035	0.0282	0.033	0.0259
Symbolic Cov.	$\hat{\beta}_{1,3}$	0.1395	0.1212	0.0936	0.0779	0.0749	0.0609	0.0655	0.0482
	$\hat{\beta}_{2,3}$	0.1403	0.1227	0.0993	0.0808	0.0762	0.0616	0.0641	0.0509
	$\hat{\beta}_{1,4}$	0.1374	0.1163	0.0955	0.0787	0.0753	0.0607	0.0637	0.0491
	$\hat{\beta}_{2,4}$	0.1467	0.1238	0.0933	0.0783	0.0786	0.0642	0.0666	0.0521
Method	Estimator	$n = 75$		$n = 100$		$n = 150$		$n = 200$	
		Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
Center	$\hat{\beta}_{1,3}$	0.032	0.0247	0.0338	0.0267	0.0306	0.0256	0.0333	0.0261
	$\hat{\beta}_{2,3}$	0.0313	0.0248	0.0336	0.0262	0.0309	0.0235	0.0314	0.0248
	$\hat{\beta}_{1,4}$	0.0333	0.0255	0.0333	0.0252	0.0311	0.0251	0.0329	0.026
	$\hat{\beta}_{2,4}$	0.0324	0.0258	0.0328	0.0261	0.0314	0.024	0.0321	0.0246
Symbolic Cov.	$\hat{\beta}_{1,3}$	0.0580	0.0431	0.054	0.0386	0.0498	0.0338	0.0464	0.0322
	$\hat{\beta}_{2,3}$	0.0553	0.042	0.0515	0.0402	0.0487	0.0339	0.045	0.0312
	$\hat{\beta}_{1,4}$	0.0569	0.0452	0.0517	0.0397	0.0498	0.0341	0.0489	0.0326
	$\hat{\beta}_{2,4}$	0.0577	0.0443	0.0531	0.0391	0.0477	0.0336	0.0441	0.0301

Table A.56: Average absolute relative bias of the loadings associated with each latent variable for Study III, for different sample sizes

Method	Latent Variable	$n = 10$		$n = 20$		$n = 30$		$n = 50$	
		Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
Center	ξ_1	0.0981	0.0789	0.0972	0.0372	0.098	0.0349	0.0996	0.0299
	ξ_2	0.0965	0.0759	0.0962	0.0382	0.0971	0.0342	0.0972	0.0313
	ξ_3	0.0778	0.0373	0.0838	0.0355	0.0814	0.0344	0.0847	0.0335
	ξ_4	0.0784	0.0383	0.0829	0.0346	0.084	0.0336	0.086	0.0322
Symbolic Cov.	ξ_1	0.1071	0.0764	0.1059	0.076	0.1061	0.0767	0.1052	0.0752
	ξ_2	0.1069	0.0763	0.1066	0.0764	0.1041	0.0753	0.1037	0.0757
	ξ_3	0.0927	0.0743	0.0922	0.0742	0.0936	0.0744	0.0928	0.0747
	ξ_4	0.0933	0.0741	0.0927	0.0742	0.0945	0.0746	0.095	0.0742
Method	Latent Variable	$n = 75$		$n = 100$		$n = 150$		$n = 200$	
		Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
Center	ξ_1	0.0988	0.0285	0.0991	0.029	0.0997	0.0275	0.101	0.0264
	ξ_2	0.1000	0.0287	0.1013	0.0285	0.099	0.0275	0.0987	0.027
	ξ_3	0.085	0.0306	0.0845	0.0316	0.0856	0.0310	0.0862	0.0316
	ξ_4	0.0848	0.0325	0.0872	0.0327	0.0864	0.0318	0.0857	0.0326
Symbolic Cov.	ξ_1	0.1066	0.0756	0.1049	0.0752	0.1065	0.0754	0.1053	0.0746
	ξ_2	0.1042	0.0751	0.1062	0.0751	0.1047	0.0749	0.105	0.0740
	ξ_3	0.0945	0.0745	0.0928	0.0743	0.0951	0.0748	0.095	0.0742
	ξ_4	0.0937	0.0736	0.0934	0.0738	0.0942	0.0739	0.0946	0.0740

Table A.57: Average root mean squared error for the latent variables in Study III, for different sample sizes

Method	Latent Variable	$n = 10$		$n = 20$		$n = 30$		$n = 50$	
		$RMSE_L$	$RMSE_U$	$RMSE_L$	$RMSE_U$	$RMSE_L$	$RMSE_U$	$RMSE_L$	$RMSE_U$
Center	ξ_1	0.6097	0.6103	0.6181	0.6186	0.6206	0.6223	0.6206	0.6206
	ξ_2	0.6131	0.6166	0.6146	0.6159	0.6153	0.6156	0.6218	0.6221
	ξ_3	0.2322	0.2323	0.2459	0.2459	0.24	0.2402	0.2463	0.2464
	ξ_4	0.2316	0.2317	0.243	0.2432	0.2481	0.2482	0.2507	0.2508
Symbolic Cov.	ξ_1	0.6195	0.6208	0.6241	0.6252	0.6256	0.6262	0.6255	0.6259
	ξ_2	0.6178	0.6189	0.6284	0.6284	0.6221	0.6228	0.6264	0.6269
	ξ_3	0.2417	0.2419	0.2489	0.2492	0.2528	0.2529	0.2555	0.2557
	ξ_4	0.2422	0.2424	0.2495	0.2498	0.257	0.257	0.2607	0.2608
Method	Latent Variable	$n = 75$		$n = 100$		$n = 150$		$n = 200$	
		$RMSE_L$	$RMSE_U$	$RMSE_L$	$RMSE_U$	$RMSE_L$	$RMSE_U$	$RMSE_L$	$RMSE_U$
Center	ξ_1	0.6222	0.6225	0.6215	0.6215	0.6244	0.6248	0.6265	0.6271
	ξ_2	0.6227	0.6227	0.6259	0.6261	0.6214	0.6214	0.6208	0.6212
	ξ_3	0.2495	0.2496	0.2463	0.2464	0.2506	0.2508	0.2506	0.2506
	ξ_4	0.2488	0.2489	0.2524	0.2525	0.2526	0.2527	0.2502	0.2503
Symbolic Cov.	ξ_2	0.6309	0.6311	0.6299	0.6301	0.6308	0.6309	0.6295	0.6298
	ξ_3	0.6282	0.6287	0.6315	0.6317	0.6275	0.6275	0.6294	0.6298
	ξ_4	0.2588	0.2588	0.2556	0.2558	0.26	0.26	0.2617	0.2618
	ξ_3	0.259	0.259	0.2578	0.2581	0.2579	0.2579	0.2597	0.2597

Table A.58: Average absolute relative bias for Study IV, for different sample sizes

Method	Estimator	$n = 10$		$n = 20$		$n = 30$		$n = 50$	
		Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
Center	$\hat{\beta}_{1,3}$	0.0508	0.0747	0.041	0.0696	0.0343	0.0271	0.0333	0.0263
	$\hat{\beta}_{2,3}$	0.0501	0.0774	0.0407	0.0322	0.0347	0.0267	0.0334	0.0263
	$\hat{\beta}_{1,4}$	0.3878	0.3118	0.2712	0.2396	0.2302	0.1843	0.1866	0.1457
	$\hat{\beta}_{2,4}$	0.3969	0.3435	0.2636	0.2213	0.2272	0.1819	0.1864	0.1424
	$\hat{\beta}_{3,4}$	0.5389	0.4212	0.3826	0.3339	0.35	0.2879	0.2896	0.2384
Symbolic Cov.	$\hat{\beta}_{1,3}$	0.1391	0.1174	0.0916	0.076	0.0819	0.0636	0.063	0.0496
	$\hat{\beta}_{2,3}$	0.142	0.1211	0.0925	0.0758	0.0813	0.0644	0.0638	0.0514
	$\hat{\beta}_{1,4}$	0.8497	0.584	0.7334	0.2983	0.7316	0.2722	0.6936	0.2357
	$\hat{\beta}_{2,4}$	0.7997	0.4194	0.7317	0.2916	0.7201	0.2766	0.6976	0.2432
	$\hat{\beta}_{3,4}$	1.1105	0.796	0.9771	0.5878	0.9836	0.5722	0.9163	0.5107
Method	Estimator	$n = 75$		$n = 100$		$n = 150$		$n = 200$	
		Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
Center	$\hat{\beta}_{1,3}$	0.0334	0.0271	0.0329	0.0256	0.0314	0.0242	0.0319	0.0248
	$\hat{\beta}_{2,3}$	0.0333	0.0262	0.032	0.0264	0.0331	0.0252	0.0317	0.0249
	$\hat{\beta}_{1,4}$	0.1669	0.1265	0.1578	0.1228	0.1455	0.1023	0.1318	0.0927
	$\hat{\beta}_{2,4}$	0.1650	0.1224	0.1606	0.1202	0.1443	0.1026	0.1316	0.0925
	$\hat{\beta}_{3,4}$	0.2613	0.2198	0.2519	0.2237	0.2374	0.191	0.2201	0.1786
Symbolic Cov.	$\hat{\beta}_{1,3}$	0.0582	0.0449	0.0542	0.0399	0.0476	0.0331	0.0471	0.0305
	$\hat{\beta}_{2,3}$	0.0552	0.0414	0.0509	0.0377	0.0491	0.0346	0.0458	0.0303
	$\hat{\beta}_{1,4}$	0.7045	0.2242	0.7052	0.2165	0.7091	0.2119	0.7125	0.2031
	$\hat{\beta}_{2,4}$	0.702	0.2296	0.707	0.2152	0.7177	0.213	0.7164	0.2034
	$\hat{\beta}_{3,4}$	0.9208	0.4847	0.9588	0.4818	0.9568	0.4663	0.956	0.473

Table A.59: Average absolute relative bias of the loadings associated with each latent variable for Study IV, for different sample sizes

Method	Latent Variable	$n = 10$		$n = 20$		$n = 30$		$n = 50$	
		Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
Center	ξ_1	0.0951	0.0751	0.0979	0.0728	0.0976	0.0349	0.0979	0.0297
	ξ_2	0.0942	0.0785	0.0939	0.0401	0.0995	0.0332	0.098	0.0311
	ξ_3	0.0765	0.0390	0.0814	0.0351	0.0857	0.0332	0.0853	0.0328
	ξ_4	0.1177	0.0329	0.1226	0.0282	0.1242	0.0272	0.1248	0.0275
Symbolic Cov.	ξ_1	0.1072	0.0774	0.1032	0.0755	0.1063	0.0762	0.1041	0.0748
	ξ_2	0.1093	0.077	0.1052	0.0757	0.1045	0.0754	0.1044	0.0753
	ξ_3	0.0939	0.0747	0.094	0.0746	0.0928	0.0742	0.0942	0.0742
	ξ_4	0.1274	0.0773	0.1301	0.0755	0.1305	0.0744	0.1301	0.0742
Method	Latent Variable	$n = 75$		$n = 100$		$n = 150$		$n = 200$	
		Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
Center	ξ_1	0.0999	0.0296	0.1008	0.029	0.0987	0.0276	0.1005	0.0267
	ξ_2	0.1004	0.0313	0.0987	0.0293	0.1012	0.0263	0.1001	0.0273
	ξ_3	0.0862	0.0328	0.0848	0.0313	0.0867	0.0322	0.0865	0.0311
	ξ_4	0.1253	0.0259	0.125	0.0248	0.1248	0.026	0.1264	0.0263
Symbolic Cov.	ξ_1	0.1068	0.075	0.1049	0.0748	0.1048	0.0744	0.1061	0.0747
	ξ_2	0.1051	0.0753	0.1052	0.0744	0.1058	0.0748	0.1069	0.0749
	ξ_3	0.0936	0.074	0.0958	0.0744	0.0932	0.0736	0.093	0.0737
	ξ_4	0.1311	0.0738	0.1314	0.0737	0.1312	0.073	0.1323	0.0738

Table A.60: Average root mean squared error for the latent variables in Study IV, for different sample sizes

Method	Latent Variable	$n = 10$		$n = 20$		$n = 30$		$n = 50$	
		$RMSE_L$	$RMSE_U$	$RMSE_L$	$RMSE_U$	$RMSE_L$	$RMSE_U$	$RMSE_L$	$RMSE_U$
Center	ξ_1	0.6105	0.6111	0.6193	0.6197	0.6204	0.6205	0.6177	0.6187
	ξ_2	0.6119	0.6129	0.6166	0.6175	0.6205	0.6213	0.6226	0.6238
	ξ_3	0.2305	0.2312	0.2417	0.2417	0.2521	0.2522	0.248	0.248
	ξ_4	0.4105	0.4107	0.4297	0.4299	0.4368	0.4368	0.4417	0.4417
Symbolic Cov.	ξ_1	0.6191	0.6200	0.6217	0.6218	0.6289	0.6295	0.6243	0.6247
	ξ_2	0.6206	0.6216	0.6212	0.6224	0.6212	0.6215	0.6276	0.6279
	ξ_3	0.2452	0.2456	0.2538	0.2539	0.2516	0.2517	0.257	0.257
	ξ_4	0.4294	0.4294	0.4432	0.4433	0.4453	0.4453	0.4457	0.4458
Method	Latent Variable	$n = 75$		$n = 100$		$n = 150$		$n = 200$	
		$RMSE_L$	$RMSE_U$	$RMSE_L$	$RMSE_U$	$RMSE_L$	$RMSE_U$	$RMSE_L$	$RMSE_U$
Center	ξ_1	0.6217	0.622	0.624	0.6241	0.6196	0.6198	0.6232	0.6233
	ξ_2	0.6265	0.627	0.6212	0.6215	0.6242	0.6242	0.6231	0.6234
	ξ_3	0.2515	0.2515	0.2472	0.2473	0.2524	0.2524	0.2515	0.2515
	ξ_4	0.4425	0.4425	0.4409	0.4409	0.4421	0.4421	0.444	0.4441
Symbolic Cov.	ξ_2	0.6318	0.632	0.6276	0.6281	0.6287	0.6291	0.631	0.6311
	ξ_3	0.6263	0.6266	0.6271	0.6272	0.6303	0.6307	0.6309	0.631
	ξ_4	0.2561	0.2563	0.2627	0.2628	0.258	0.2582	0.2578	0.2578
	ξ_3	0.45	0.4501	0.4504	0.4505	0.451	0.451	0.4538	0.4538

Table A.61: Average absolute relative bias for Study V, for different sample sizes

Method	Estimator	$n = 10$		$n = 20$		$n = 30$		$n = 50$	
		Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
Center	$\hat{\beta}_{1,4}$	0.0668	0.0497	0.0662	0.0483	0.0667	0.0484	0.0667	0.0492
	$\hat{\beta}_{1,5}$	0.0485	0.0388	0.038	0.0308	0.0352	0.0265	0.0336	0.0254
	$\hat{\beta}_{2,5}$	0.0472	0.0384	0.0382	0.0301	0.0377	0.0293	0.0353	0.0276
	$\hat{\beta}_{2,6}$	0.0472	0.0395	0.0379	0.0303	0.0363	0.0293	0.0344	0.0277
	$\hat{\beta}_{3,6}$	0.0628	0.1706	0.0412	0.0721	0.037	0.0294	0.0342	0.0275
Symbolic Cov.	$\hat{\beta}_{1,4}$	0.1081	0.0751	0.1051	0.0761	0.098	0.0736	0.1003	0.0751
	$\hat{\beta}_{1,5}$	0.1391	0.1167	0.0956	0.0802	0.0788	0.0633	0.0663	0.0512
	$\hat{\beta}_{2,5}$	0.144	0.1266	0.0934	0.0743	0.083	0.0664	0.0612	0.0486
	$\hat{\beta}_{2,6}$	0.1432	0.1189	0.0963	0.0795	0.0752	0.062	0.0649	0.0486
	$\hat{\beta}_{3,6}$	0.1509	0.133	0.0928	0.0745	0.0779	0.0635	0.0613	0.0466
Method	Estimator	$n = 75$		$n = 100$		$n = 150$		$n = 200$	
		Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
Center	$\hat{\beta}_{1,4}$	0.068	0.0489	0.0695	0.0497	0.0652	0.0459	0.0706	0.05
	$\hat{\beta}_{1,5}$	0.033	0.0256	0.0335	0.0258	0.0316	0.0244	0.0321	0.0255
	$\hat{\beta}_{2,5}$	0.0337	0.0252	0.0322	0.0247	0.0314	0.0252	0.0328	0.0251
	$\hat{\beta}_{2,6}$	0.0326	0.0246	0.031	0.0245	0.0309	0.0242	0.0323	0.025
	$\hat{\beta}_{3,6}$	0.033	0.0256	0.0336	0.0254	0.0307	0.0242	0.03	0.0236
Symbolic Cov.	$\hat{\beta}_{1,4}$	0.1049	0.0753	0.1046	0.0765	0.1036	0.0738	0.1043	0.0743
	$\hat{\beta}_{1,5}$	0.0559	0.0414	0.0535	0.0388	0.0482	0.0327	0.0463	0.0303
	$\hat{\beta}_{2,5}$	0.0579	0.044	0.0524	0.0389	0.0474	0.0329	0.0452	0.0308
	$\hat{\beta}_{2,6}$	0.0557	0.042	0.052	0.0384	0.0489	0.0353	0.0471	0.0316
	$\hat{\beta}_{3,6}$	0.0564	0.0433	0.0511	0.0372	0.0465	0.0315	0.0443	0.0297

Table A.62: Average absolute relative bias of the loadings associated with each latent variable for Study V, for different sample sizes

Method	Latent Variable	$n = 10$		$n = 20$		$n = 30$		$n = 50$	
		Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
Center	ξ_1	0.0926	0.0456	0.0952	0.0373	0.0981	0.0344	0.0993	0.0305
	ξ_2	0.0937	0.0421	0.0961	0.0374	0.0967	0.0355	0.0996	0.0311
	ξ_3	0.1066	0.1697	0.0972	0.0736	0.098	0.0346	0.0987	0.0314
	ξ_4	0.0515	0.0336	0.0494	0.0326	0.0498	0.0315	0.0512	0.0322
	ξ_5	0.0787	0.0378	0.0842	0.0348	0.0842	0.0338	0.0848	0.0327
	ξ_6	0.0774	0.037	0.0834	0.0345	0.0848	0.0347	0.0836	0.0324
Symbolic Cov.	ξ_1	0.1087	0.0771	0.1066	0.0763	0.1056	0.0759	0.1057	0.0756
	ξ_2	0.1078	0.0769	0.1054	0.0755	0.1048	0.0759	0.105	0.0754
	ξ_3	0.1096	0.0969	0.1055	0.0763	0.1063	0.076	0.1067	0.076
	ξ_4	0.0796	0.0611	0.0781	0.0597	0.0774	0.0602	0.0778	0.0598
	ξ_5	0.0936	0.0745	0.0945	0.0752	0.0957	0.0749	0.0943	0.0741
	ξ_6	0.0969	0.0741	0.0942	0.0747	0.0931	0.0745	0.094	0.0744
Method	Latent Variable	$n = 75$		$n = 100$		$n = 150$		$n = 200$	
		Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
Center	ξ_1	0.0991	0.0298	0.0999	0.0278	0.0989	0.0271	0.0996	0.0271
	ξ_2	0.0992	0.0299	0.0974	0.0277	0.0973	0.0272	0.1002	0.0281
	ξ_3	0.0992	0.0299	0.1012	0.029	0.0994	0.0268	0.0988	0.026
	ξ_4	0.0512	0.0322	0.0489	0.0322	0.0495	0.0322	0.0482	0.0308
	ξ_5	0.0847	0.0314	0.0852	0.0317	0.0858	0.0313	0.0853	0.0317
	ξ_6	0.0865	0.0323	0.0858	0.0323	0.0871	0.0322	0.0863	0.0306
Symbolic Cov.	ξ_1	0.1063	0.0749	0.1061	0.075	0.1055	0.075	0.1053	0.0747
	ξ_2	0.1036	0.0747	0.1062	0.0749	0.1062	0.0749	0.1066	0.0748
	ξ_3	0.1043	0.0749	0.1047	0.0752	0.1052	0.0747	0.1052	0.0743
	ξ_4	0.0778	0.0587	0.0769	0.058	0.0768	0.0585	0.0781	0.0582
	ξ_5	0.0944	0.0744	0.0941	0.0742	0.0925	0.0731	0.0942	0.0738
	ξ_6	0.0936	0.0741	0.0949	0.0743	0.0927	0.0735	0.0945	0.0745

Table A.63: Average root mean squared error for the latent variables in Study V, for different sample sizes

Method	Latent Variable	$n = 10$		$n = 20$		$n = 30$		$n = 50$	
		$RMSE_L$	$RMSE_U$	$RMSE_L$	$RMSE_U$	$RMSE_L$	$RMSE_U$	$RMSE_L$	$RMSE_U$
Center	ξ_1	0.6145	0.616	0.6174	0.6176	0.6226	0.6226	0.6213	0.6214
	ξ_2	0.6108	0.6125	0.6163	0.617	0.6234	0.6234	0.6219	0.6226
	ξ_3	0.6246	0.6269	0.6188	0.6192	0.6237	0.6245	0.6208	0.621
	ξ_4	0.1389	0.1393	0.1345	0.135	0.1355	0.1356	0.1385	0.1385
	ξ_5	0.2372	0.2375	0.2474	0.2475	0.2469	0.2469	0.248	0.2481
	ξ_6	0.2304	0.2305	0.246	0.246	0.2487	0.249	0.2444	0.2445
Symbolic Cov.	ξ_1	0.6181	0.6186	0.6271	0.6282	0.6278	0.628	0.6299	0.6299
	ξ_2	0.6173	0.6181	0.6239	0.6246	0.625	0.6259	0.6297	0.6301
	ξ_3	0.6181	0.6189	0.625	0.6252	0.6298	0.6304	0.6301	0.6305
	ξ_4	0.14	0.14	0.1399	0.1406	0.1442	0.1442	0.1455	0.1456
	ξ_5	0.2455	0.2458	0.2532	0.2532	0.2604	0.261	0.2575	0.2577
	ξ_6	0.2525	0.2532	0.2528	0.2528	0.2523	0.2526	0.2576	0.2577
Method	Latent Variable	$n = 75$		$n = 100$		$n = 150$		$n = 200$	
		$RMSE_L$	$RMSE_U$	$RMSE_L$	$RMSE_U$	$RMSE_L$	$RMSE_U$	$RMSE_L$	$RMSE_U$
Center	ξ_1	0.6231	0.6235	0.6246	0.6248	0.6201	0.6204	0.6224	0.6228
	ξ_2	0.6224	0.6227	0.6181	0.6184	0.6211	0.6212	0.6247	0.6248
	ξ_3	0.6218	0.6226	0.6252	0.6253	0.622	0.6224	0.6216	0.6219
	ξ_4	0.1378	0.1379	0.1329	0.133	0.1347	0.1347	0.1315	0.1316
	ξ_5	0.2473	0.2475	0.248	0.2481	0.2504	0.2505	0.2503	0.2503
	ξ_6	0.2555	0.2556	0.2496	0.2499	0.2551	0.2552	0.2506	0.2508
Symbolic Cov.	ξ_1	0.6312	0.6314	0.6323	0.6331	0.6285	0.629	0.6301	0.6303
	ξ_2	0.6259	0.6259	0.6303	0.6303	0.6309	0.6313	0.6326	0.6328
	ξ_3	0.6277	0.628	0.6282	0.6282	0.6285	0.6286	0.6285	0.6289
	ξ_4	0.141	0.1411	0.1392	0.1395	0.1392	0.1393	0.1417	0.1417
	ξ_5	0.2606	0.2608	0.2598	0.26	0.2545	0.2545	0.2583	0.2584
	ξ_6	0.2593	0.2596	0.2592	0.2593	0.2517	0.2517	0.2588	0.2589

Table A.64: Average absolute relative bias for Study VI, for different sample sizes

Method	Estimator	$n = 10$		$n = 20$		$n = 30$		$n = 50$	
		Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
Center	$\hat{\beta}_{1,4}$	0.0643	0.0471	0.0655	0.0492	0.0689	0.0479	0.068	0.0484
	$\hat{\beta}_{1,5}$	0.048	0.0395	0.0381	0.031	0.0356	0.0284	0.035	0.0267
	$\hat{\beta}_{2,5}$	0.0476	0.0375	0.0384	0.0297	0.0353	0.0277	0.0343	0.0261
	$\hat{\beta}_{2,6}$	0.0496	0.0415	0.038	0.031	0.0351	0.027	0.0321	0.0239
	$\hat{\beta}_{3,6}$	0.0695	0.2098	0.0395	0.0669	0.034	0.0259	0.0321	0.0246
	$\hat{\beta}_{5,6}$	0.1749	0.1179	0.1499	0.111	0.1378	0.0983	0.132	0.0927
Symbolic Cov.	$\hat{\beta}_{1,4}$	0.1087	0.0783	0.1022	0.0768	0.1035	0.0757	0.1001	0.0741
	$\hat{\beta}_{1,5}$	0.1435	0.1229	0.0980	0.0835	0.0796	0.0631	0.0666	0.0533
	$\hat{\beta}_{2,5}$	0.1485	0.1224	0.0910	0.0746	0.0792	0.0649	0.063	0.0509
	$\hat{\beta}_{2,6}$	0.2855	0.2582	0.2017	0.1724	0.1606	0.1257	0.121	0.0956
	$\hat{\beta}_{3,6}$	0.2108	0.1758	0.1492	0.1111	0.1271	0.1246	0.1065	0.0938
	$\hat{\beta}_{5,6}$	0.3202	0.2643	0.2441	0.1907	0.2062	0.1562	0.1683	0.1285
Method	Estimator	$n = 75$		$n = 100$		$n = 150$		$n = 200$	
		Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
Center	$\hat{\beta}_{1,4}$	0.0671	0.0485	0.0702	0.0494	0.0638	0.0474	0.0675	0.049
	$\hat{\beta}_{1,5}$	0.0335	0.0255	0.0316	0.0247	0.0321	0.0251	0.0322	0.0246
	$\hat{\beta}_{2,5}$	0.0336	0.0265	0.0324	0.0249	0.0319	0.0254	0.0308	0.0251
	$\hat{\beta}_{2,6}$	0.0315	0.0237	0.0301	0.0227	0.0302	0.0222	0.0309	0.0221
	$\hat{\beta}_{3,6}$	0.0307	0.0225	0.0299	0.0229	0.0304	0.0217	0.0303	0.0222
	$\hat{\beta}_{5,6}$	0.131	0.0947	0.1296	0.0937	0.1192	0.0906	0.126	0.0894
Symbolic Cov.	$\hat{\beta}_{1,4}$	0.1026	0.0764	0.1061	0.0751	0.1026	0.0749	0.1034	0.0744
	$\hat{\beta}_{1,5}$	0.0579	0.0435	0.0524	0.0378	0.0497	0.0362	0.0435	0.0299
	$\hat{\beta}_{2,5}$	0.0564	0.0435	0.0519	0.0375	0.0476	0.0337	0.0463	0.0324
	$\hat{\beta}_{2,6}$	0.1017	0.0759	0.0897	0.0732	0.0789	0.0631	0.0712	0.0503
	$\hat{\beta}_{3,6}$	0.0897	0.0617	0.0886	0.0572	0.0862	0.048	0.0829	0.0432
	$\hat{\beta}_{5,6}$	0.1562	0.1121	0.1588	0.1166	0.1505	0.1105	0.1396	0.1023

Table A.65: Average absolute relative bias of the loadings associated with each latent variable for Study VI, for different sample sizes

Method	Latent Variable	$n = 10$		$n = 20$		$n = 30$		$n = 50$	
		Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
Center	ξ_1	0.0946	0.0429	0.0924	0.0383	0.0968	0.0345	0.098	0.0308
	ξ_2	0.0947	0.0438	0.0964	0.0386	0.0959	0.034	0.0983	0.031
	ξ_3	0.1165	0.2184	0.0989	0.0737	0.0978	0.0345	0.0971	0.0306
	ξ_4	0.0519	0.0345	0.0506	0.0322	0.0485	0.0324	0.0513	0.0331
	ξ_5	0.0792	0.0378	0.0829	0.0349	0.0852	0.0338	0.0843	0.0325
	ξ_6	0.115	0.0334	0.115	0.0299	0.1155	0.0293	0.116	0.0282
Symbolic Cov.	ξ_1	0.1100	0.0774	0.1054	0.0765	0.1053	0.0761	0.1054	0.0752
	ξ_2	0.1051	0.0763	0.1046	0.0763	0.1053	0.0761	0.1063	0.0758
	ξ_3	0.1113	0.1292	0.1049	0.0759	0.1116	0.1307	0.1076	0.0972
	ξ_4	0.0803	0.0616	0.0789	0.0596	0.079	0.0601	0.077	0.0587
	ξ_5	0.0949	0.0744	0.0926	0.0742	0.0939	0.0746	0.0925	0.0738
	ξ_6	0.1195	0.0777	0.1215	0.0759	0.1231	0.0751	0.1234	0.0751
Method	Latent Variable	$n = 75$		$n = 100$		$n = 150$		$n = 200$	
		Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
Center	ξ_1	0.1003	0.0291	0.1004	0.0291	0.1003	0.0266	0.1008	0.0273
	ξ_2	0.1000	0.0296	0.1001	0.0269	0.1006	0.0281	0.1006	0.0268
	ξ_3	0.0988	0.0274	0.1002	0.0287	0.0997	0.0271	0.099	0.0265
	ξ_4	0.0496	0.0322	0.05	0.0335	0.0516	0.0332	0.05	0.0318
	ξ_5	0.0872	0.0332	0.0858	0.0307	0.086	0.0311	0.0852	0.0308
	ξ_6	0.1176	0.0279	0.1173	0.0275	0.117	0.0276	0.1175	0.0273
Symbolic Cov.	ξ_1	0.1036	0.075	0.1057	0.0751	0.1051	0.0746	0.1043	0.0739
	ξ_2	0.1062	0.0751	0.1059	0.0752	0.1066	0.0748	0.1064	0.0743
	ξ_3	0.1057	0.0757	0.1044	0.0742	0.1061	0.0752	0.1041	0.0737
	ξ_4	0.077	0.0586	0.0768	0.0577	0.0774	0.0584	0.0766	0.0578
	ξ_5	0.0939	0.0744	0.0936	0.0735	0.0938	0.0741	0.0933	0.0737
	ξ_6	0.1236	0.0742	0.1246	0.0749	0.1249	0.0743	0.1245	0.0736

Table A.66: Average root mean squared error for the latent variables in Study VI, for different sample sizes

Method	Latent Variable	$n = 10$		$n = 20$		$n = 30$		$n = 50$	
		$RMSE_L$	$RMSE_U$	$RMSE_L$	$RMSE_U$	$RMSE_L$	$RMSE_U$	$RMSE_L$	$RMSE_U$
Center	ξ_1	0.6095	0.6096	0.6162	0.6168	0.6166	0.6167	0.6189	0.6191
	ξ_2	0.6106	0.6126	0.6161	0.6171	0.6169	0.6173	0.6216	0.6217
	ξ_3	0.6677	0.6709	0.6394	0.6401	0.63	0.6307	0.6202	0.6207
	ξ_4	0.139	0.1393	0.1378	0.1379	0.1313	0.1313	0.1373	0.1374
	ξ_5	0.2368	0.2372	0.2453	0.2458	0.2506	0.2507	0.2474	0.2474
	ξ_6	0.3796	0.3798	0.3816	0.3816	0.3874	0.3876	0.3857	0.3857
Symbolic Cov.	ξ_1	0.6223	0.6226	0.6298	0.6304	0.6286	0.6287	0.6308	0.6309
	ξ_2	0.6188	0.619	0.6258	0.6264	0.6271	0.6282	0.6306	0.6306
	ξ_3	0.6243	0.6246	0.6268	0.6281	0.6318	0.6322	0.6297	0.6307
	ξ_4	0.1381	0.1382	0.1429	0.1431	0.1434	0.1436	0.1407	0.1407
	ξ_5	0.2465	0.2469	0.2513	0.2516	0.2566	0.2567	0.2533	0.2535
	ξ_6	0.3803	0.3806	0.3913	0.3915	0.3985	0.3985	0.4003	0.4004
Method	Latent Variable	$n = 75$		$n = 100$		$n = 150$		$n = 200$	
		$RMSE_L$	$RMSE_U$	$RMSE_L$	$RMSE_U$	$RMSE_L$	$RMSE_U$	$RMSE_L$	$RMSE_U$
Center	ξ_1	0.6264	0.6265	0.6241	0.6243	0.6233	0.6233	0.6239	0.6239
	ξ_2	0.6238	0.6238	0.6235	0.6235	0.6247	0.6253	0.6229	0.6231
	ξ_3	0.622	0.6222	0.6251	0.6252	0.6224	0.6226	0.6214	0.6216
	ξ_4	0.1367	0.1369	0.1371	0.1372	0.1404	0.1405	0.1366	0.1367
	ξ_5	0.2544	0.2544	0.2509	0.251	0.2503	0.2505	0.2495	0.2496
	ξ_6	0.3942	0.3943	0.3921	0.3921	0.3907	0.3907	0.3921	0.3921
Symbolic Cov.	ξ_1	0.6273	0.6276	0.6282	0.6286	0.6289	0.6293	0.6269	0.6275
	ξ_2	0.6294	0.6295	0.6297	0.6299	0.6299	0.6301	0.6313	0.6317
	ξ_3	0.6266	0.6266	0.6281	0.6288	0.6297	0.6298	0.6262	0.6265
	ξ_4	0.1415	0.1416	0.1381	0.1381	0.1411	0.1412	0.1401	0.1402
	ξ_5	0.258	0.2582	0.255	0.255	0.2573	0.2573	0.2565	0.2565
	ξ_6	0.4009	0.4009	0.4035	0.4036	0.4053	0.4054	0.4044	0.4044

Appendix B

Figures

B.1 Simulation 1 - Path Study

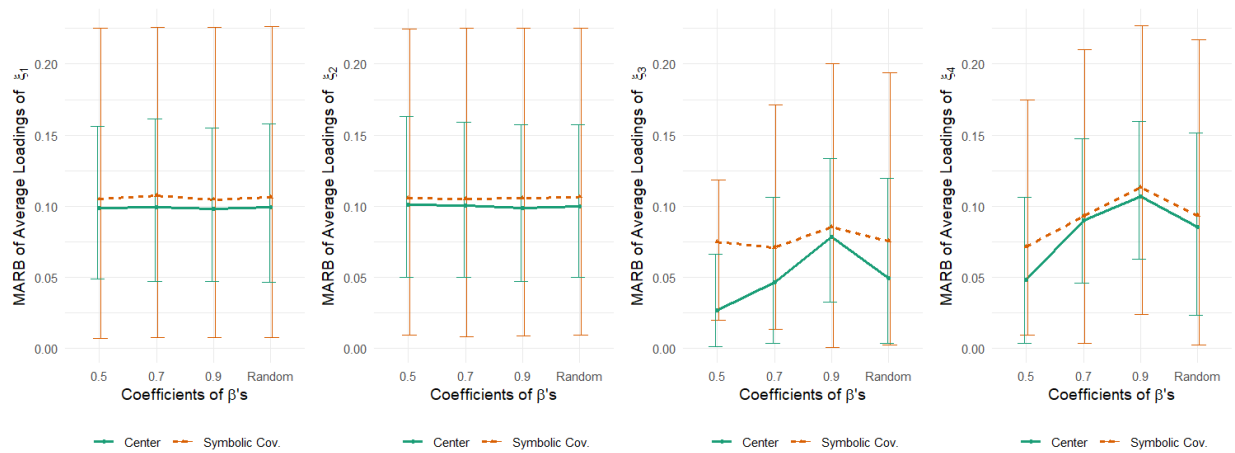


Figure B.1: Mean absolute relative bias of average loadings for each latent variable in Study I

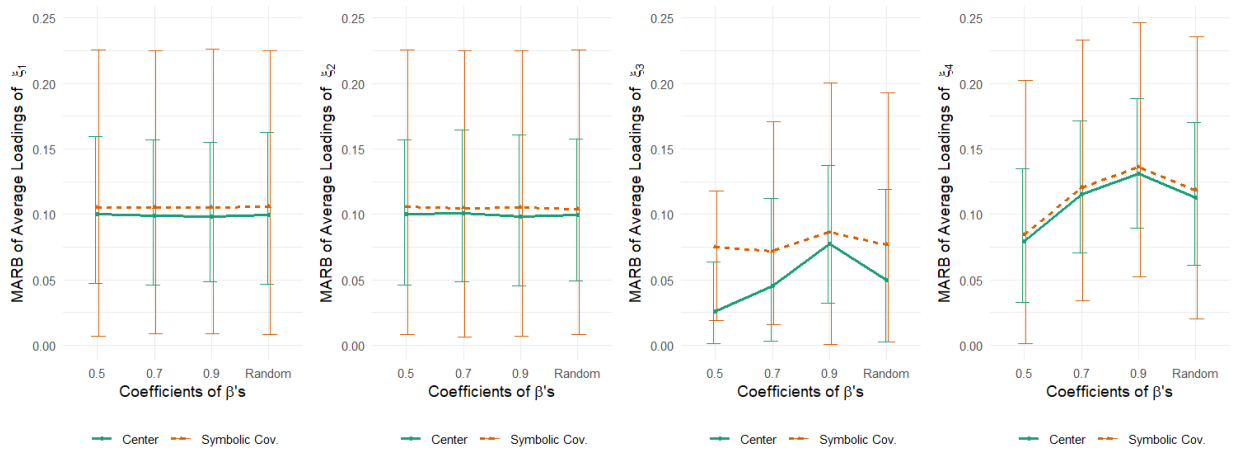


Figure B.2: Mean absolute relative bias of average loadings for each latent variable in Study II

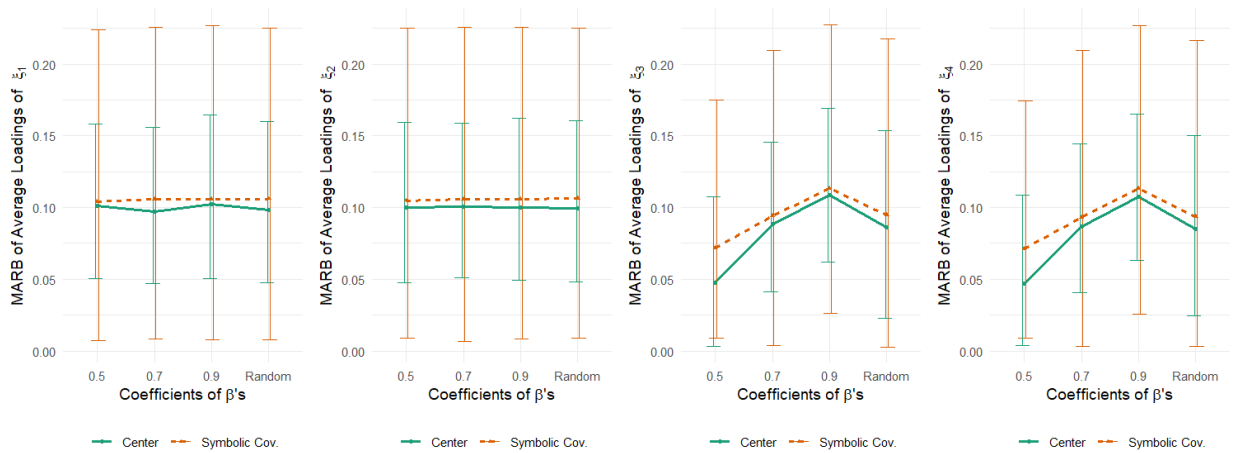


Figure B.3: Mean absolute relative bias of average loadings for each latent variable in Study III

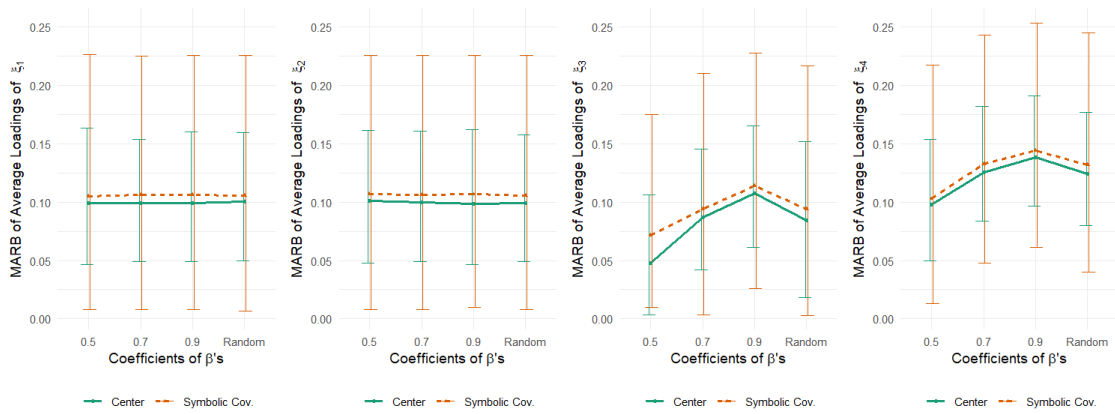


Figure B.4: Mean absolute relative bias of average loadings for each latent variable in Study IV

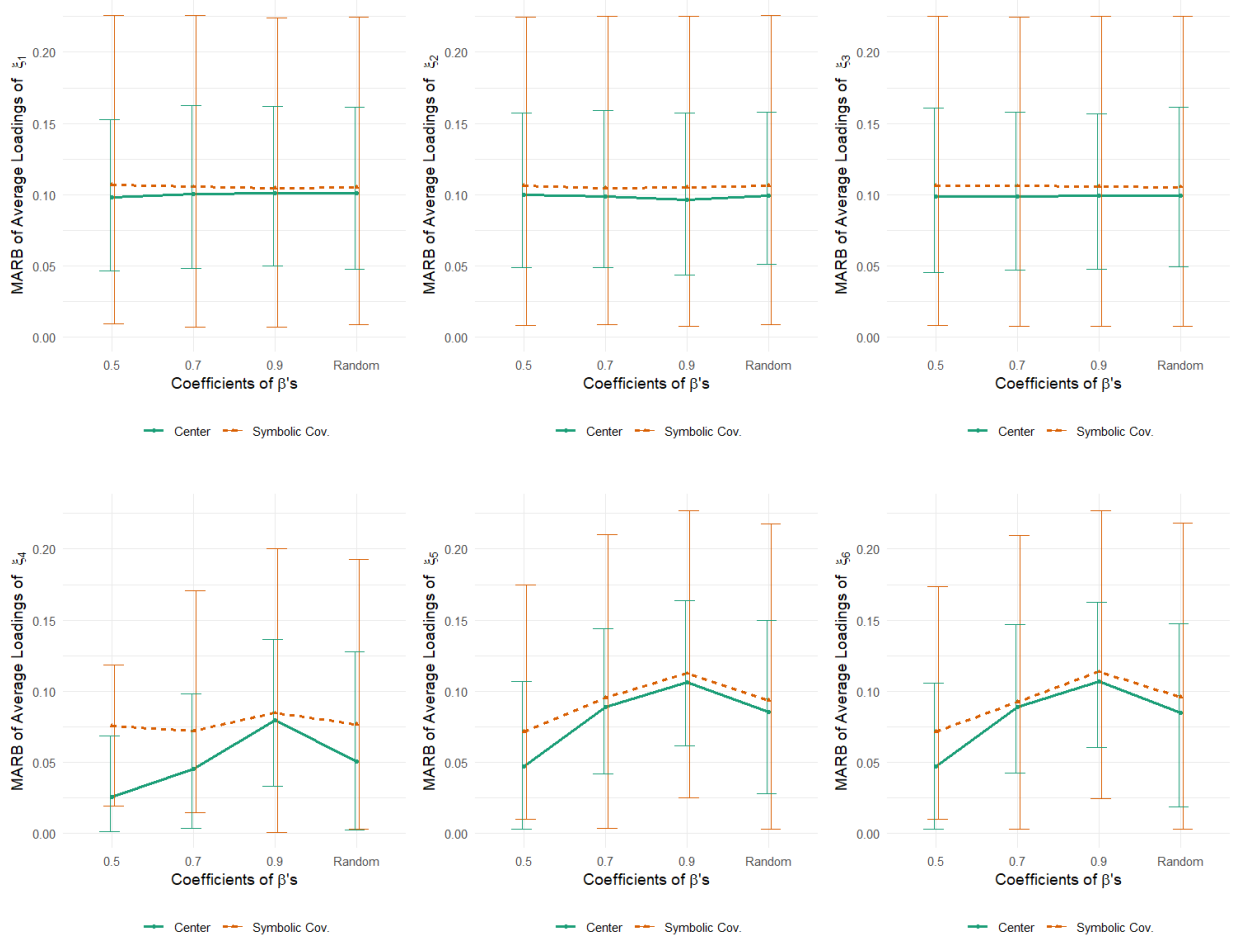


Figure B.5: Mean absolute relative bias of average loadings for each latent variable in Study V

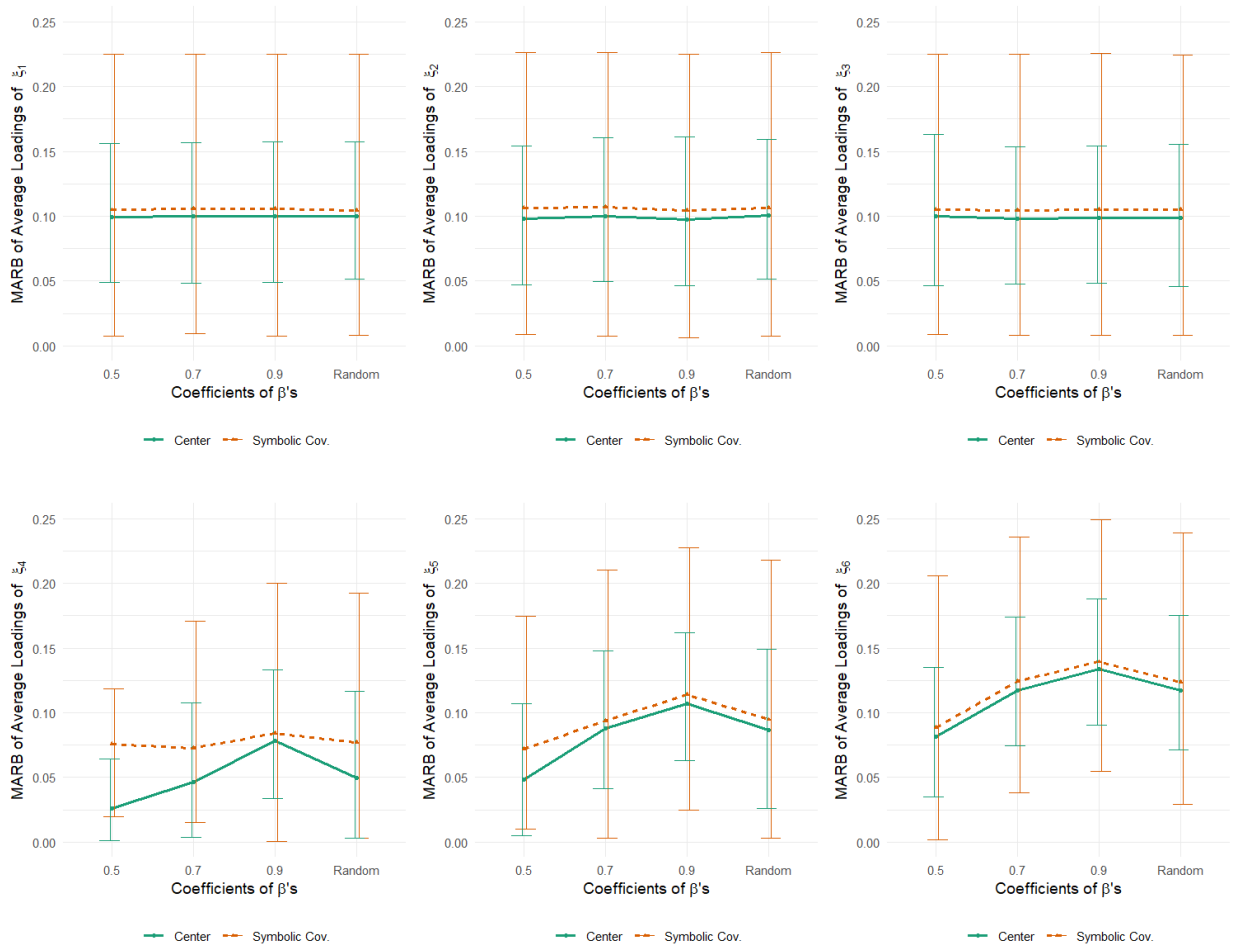


Figure B.6: Mean absolute relative bias of average loadings for each latent variable in Study VI

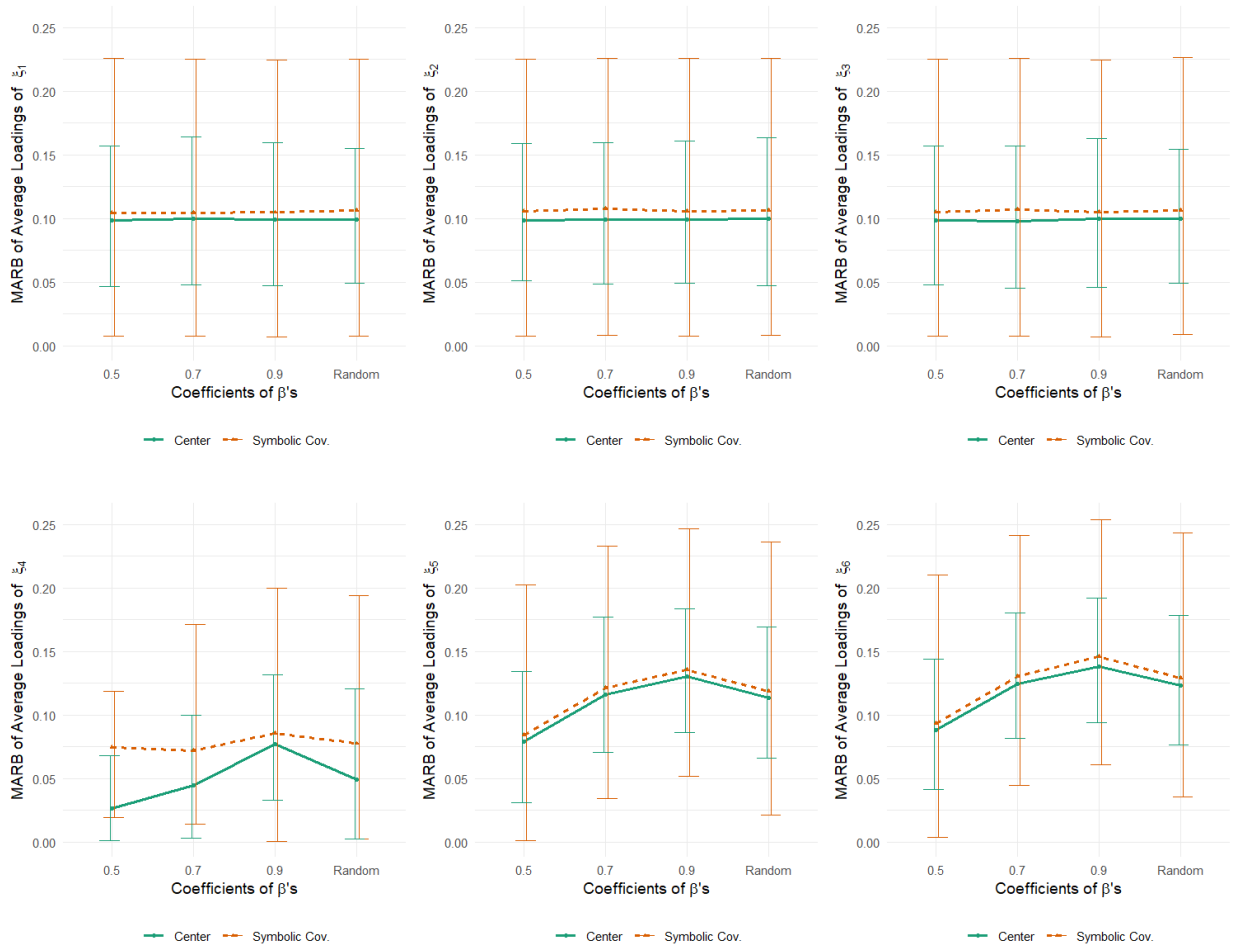


Figure B.7: Mean absolute relative bias of average loadings for each latent variable in Study VII

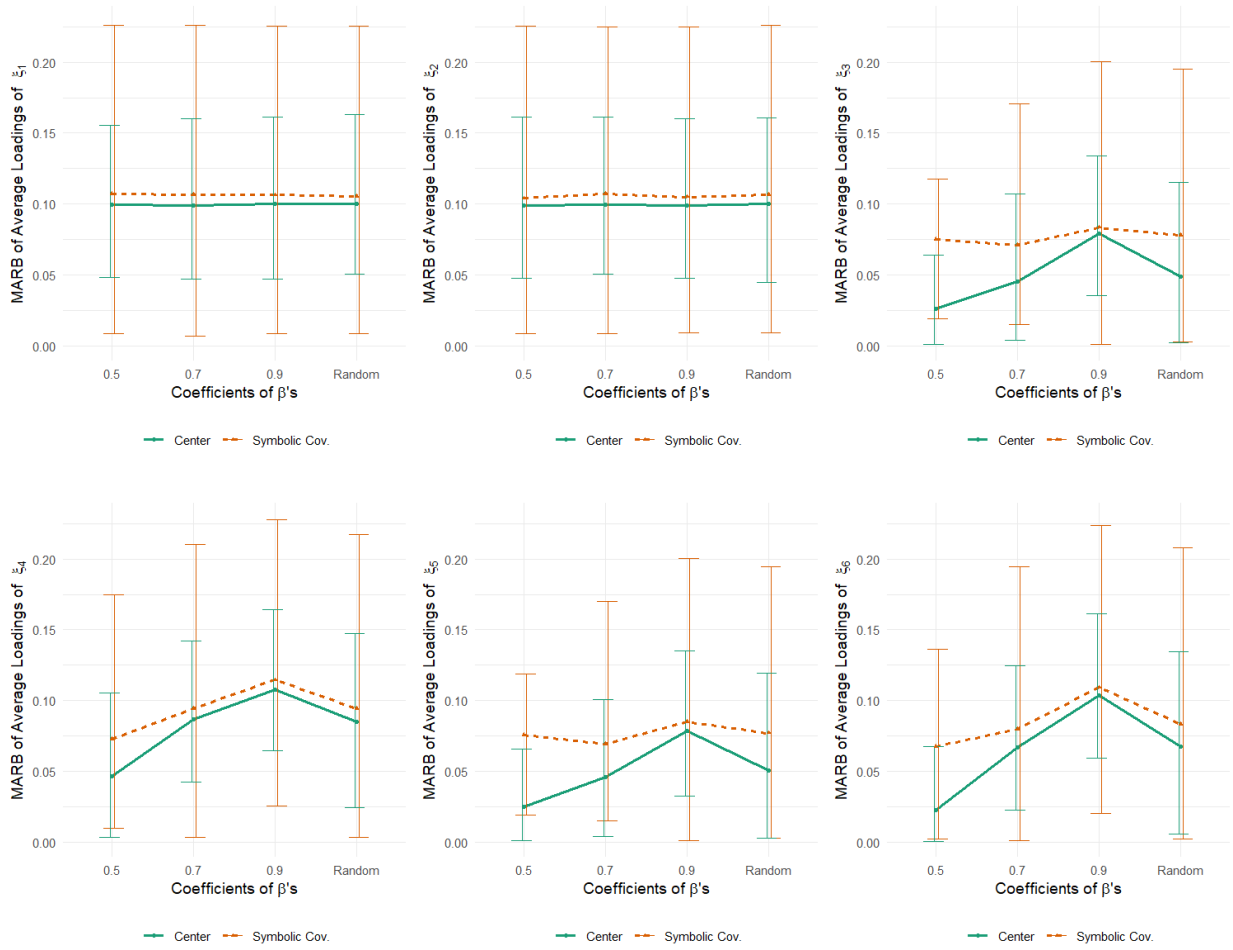


Figure B.8: Mean absolute relative bias of average loadings for each latent variable in Study VIII

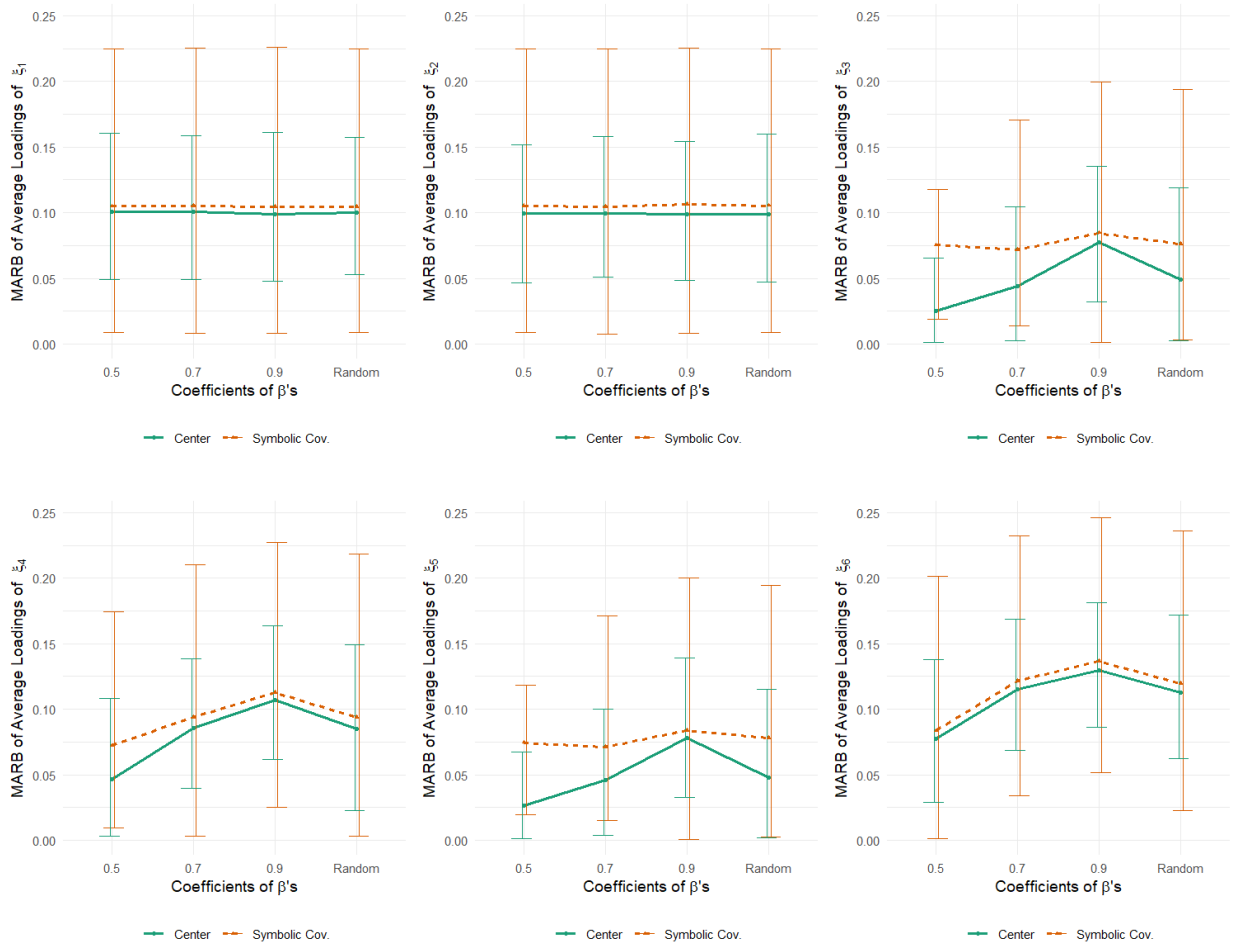


Figure B.9: Mean absolute relative bias of average loadings for each latent variable in Study IX

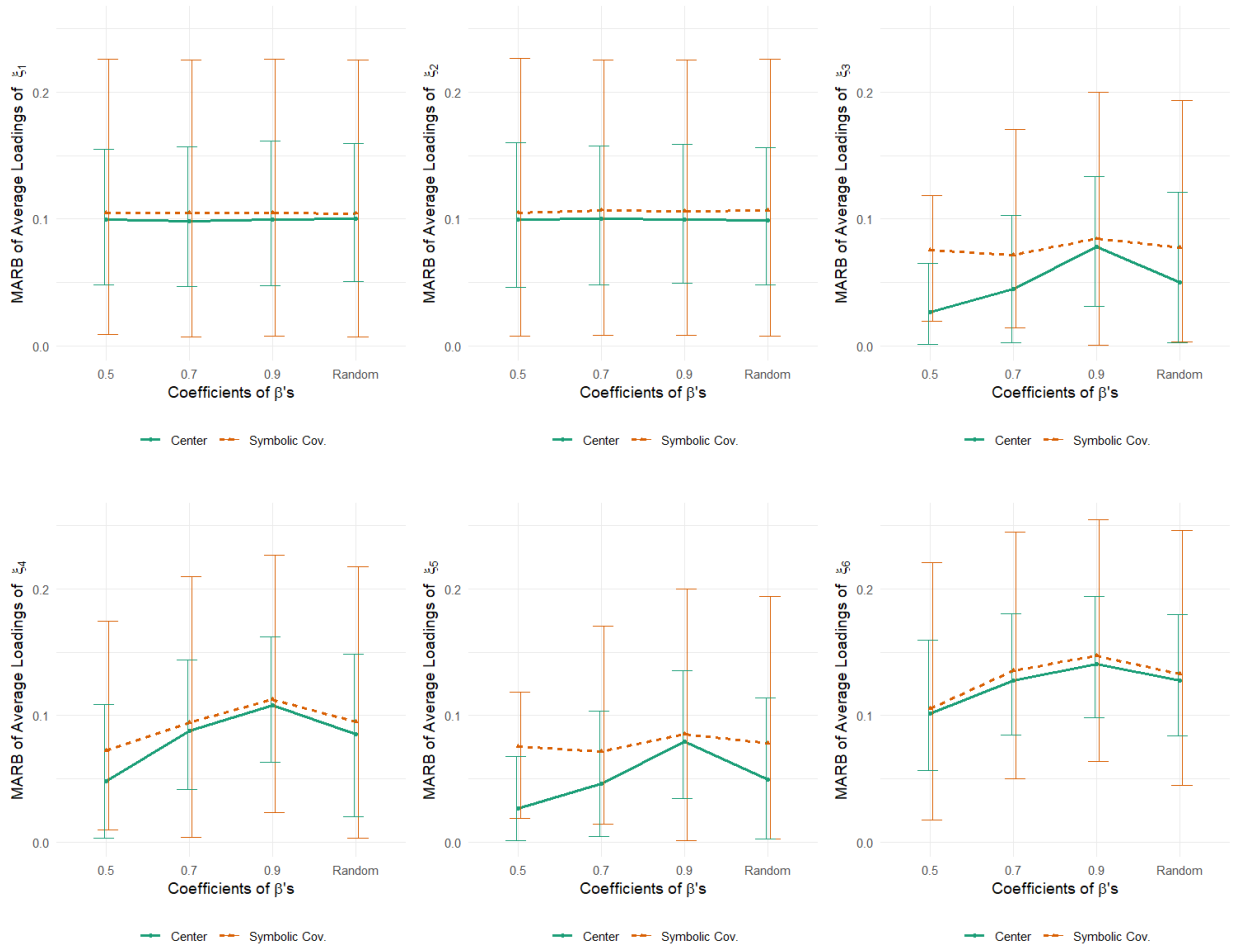


Figure B.10: Mean absolute relative bias of average loadings for each latent variable in Study X

B.2 Simulation 2 - Number of Manifest Variables

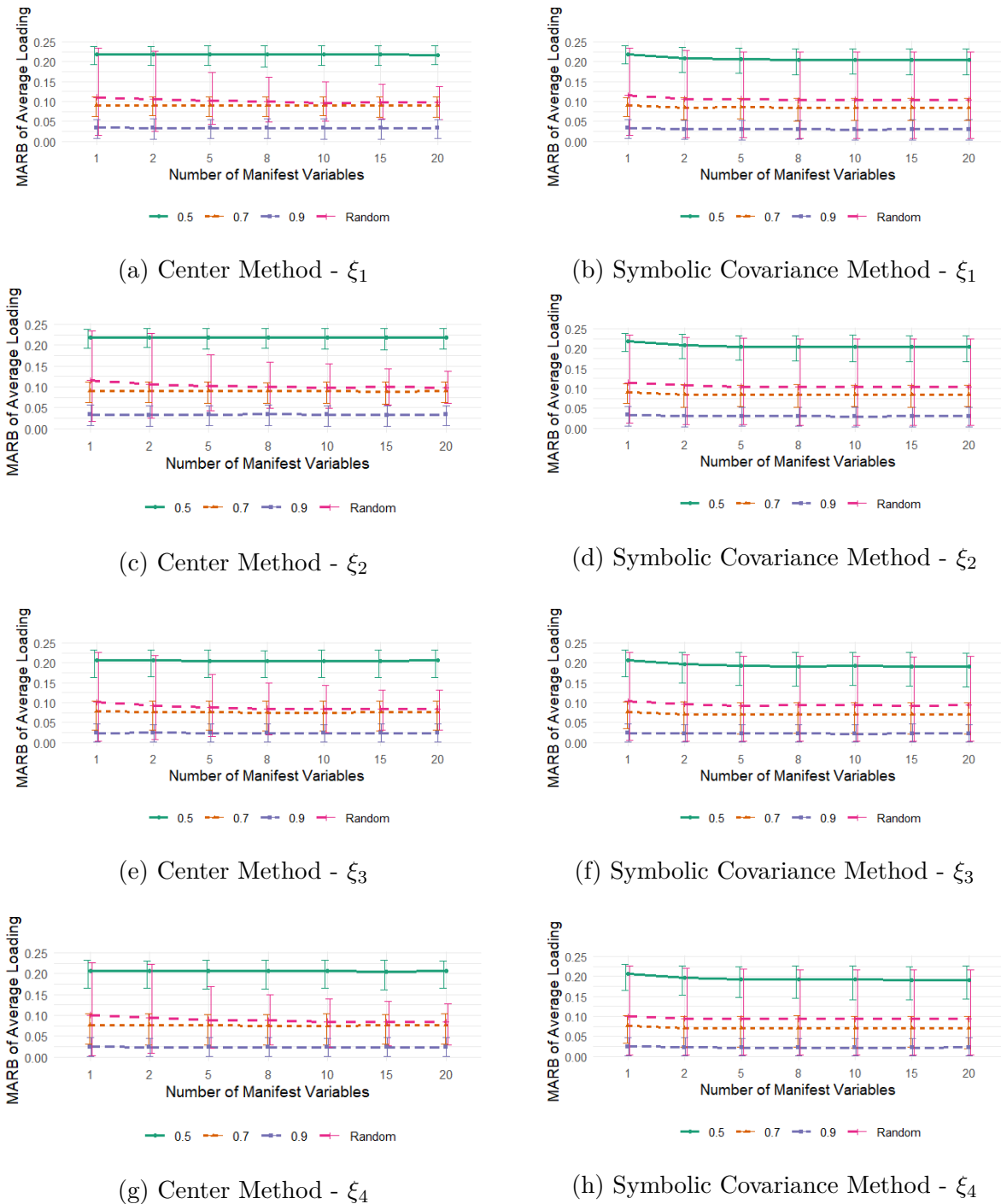
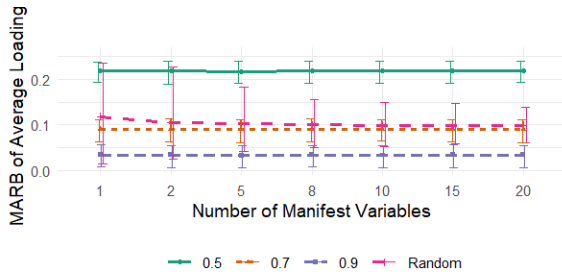
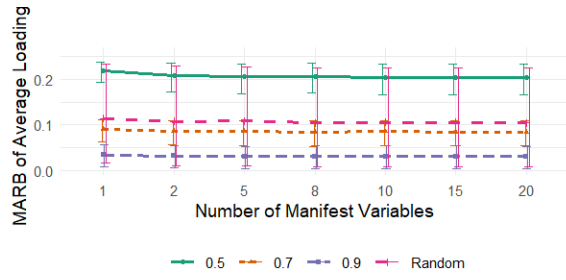


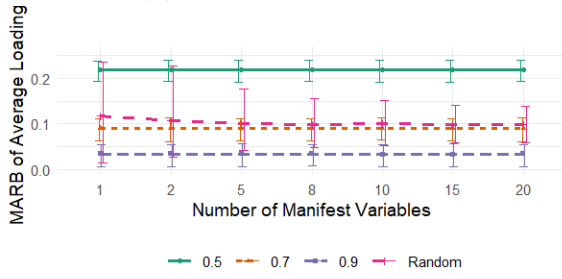
Figure B.11: Mean absolute relative bias of average loadings for each latent variable in Study III, for different loadings and number of manifest variables



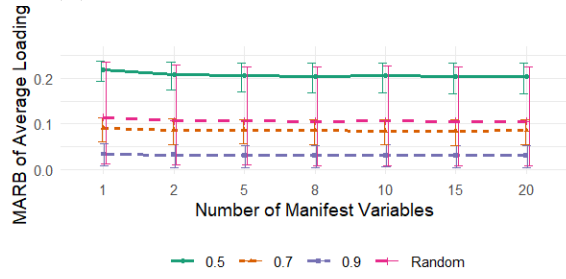
(a) Center Method - ξ_1



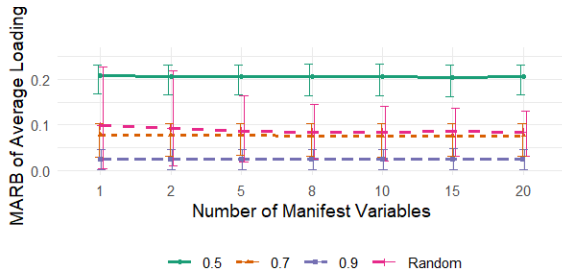
(b) Symbolic Covariance Method - ξ_1



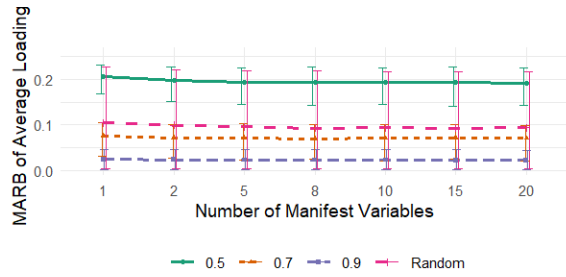
(c) Center Method - ξ_2



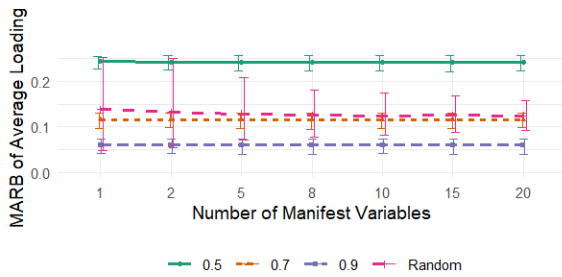
(d) Symbolic Covariance Method - ξ_2



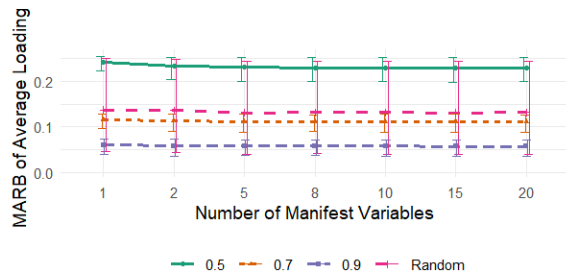
(e) Center Method - ξ_3



(f) Symbolic Covariance Method - ξ_3

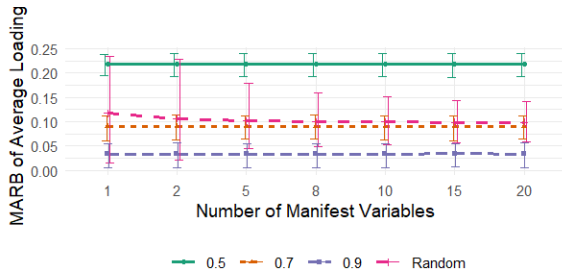


(g) Center Method - ξ_4

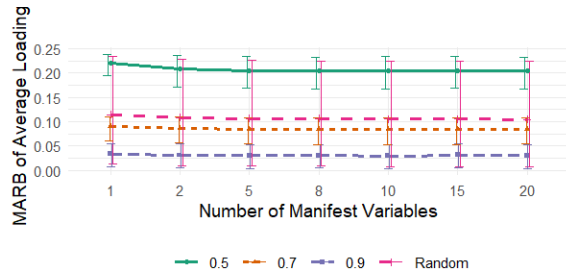


(h) Symbolic Covariance Method - ξ_4

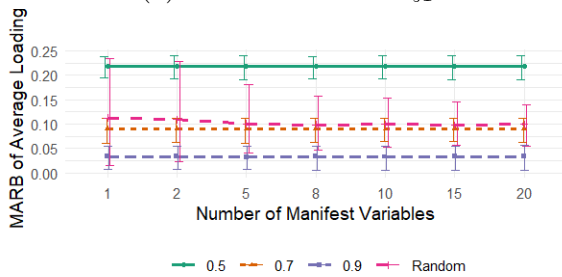
Figure B.12: Mean absolute relative bias of average loadings for each latent variable in Study IV, for different loadings and number of manifest variables



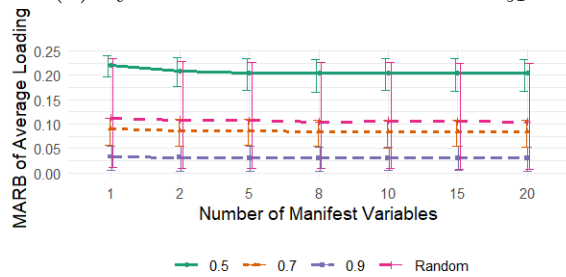
(a) Center Method - ξ_1



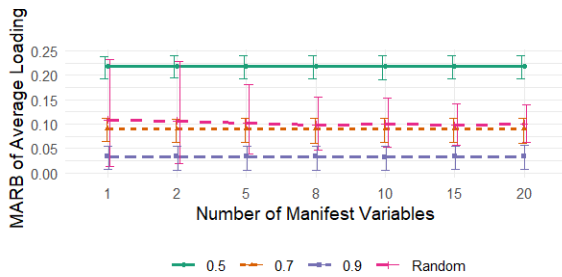
(b) Symbolic Covariance Method - ξ_1



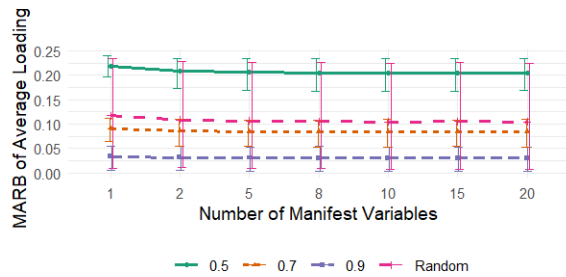
(c) Center Method - ξ_2



(d) Symbolic Covariance Method - ξ_2

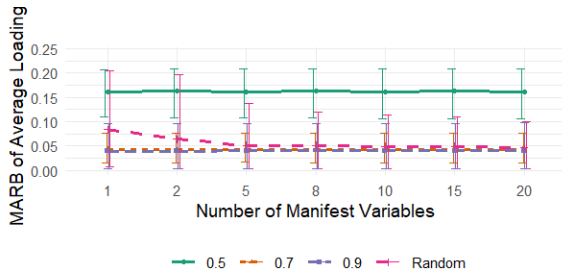


(e) Center Method - ξ_3

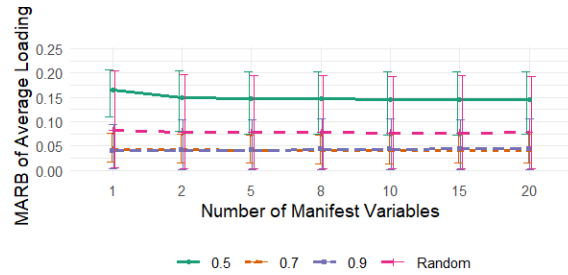


(f) Symbolic Covariance Method - ξ_3

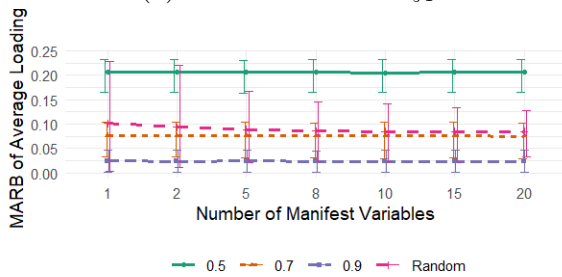
Figure B.13: Mean absolute relative bias of average loadings for each latent variable in Study V, for different loadings and number of manifest variables



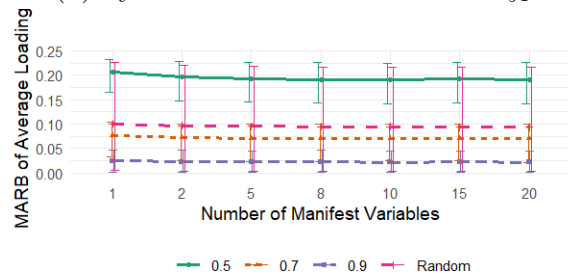
(a) Center Method - ξ_4



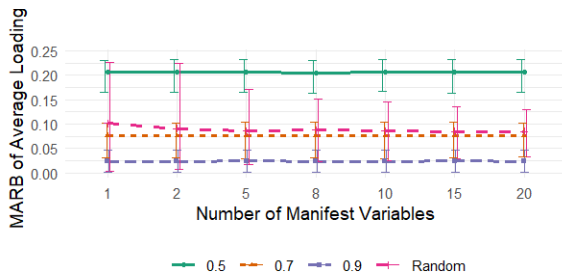
(b) Symbolic Covariance Method - ξ_4



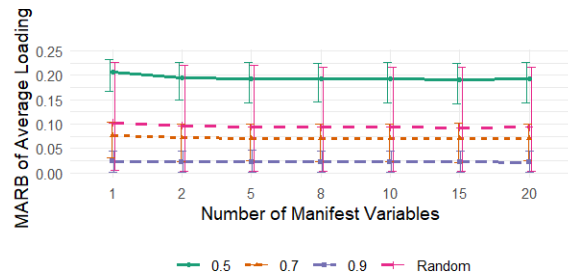
(c) Center Method - ξ_5



(d) Symbolic Covariance Method - ξ_5

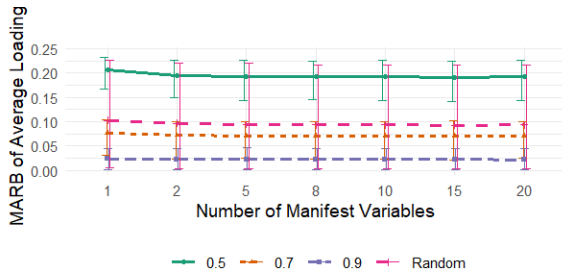


(e) Center Method - ξ_6

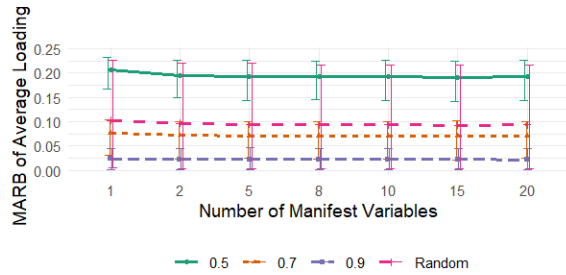


(f) Symbolic Covariance Method - ξ_6

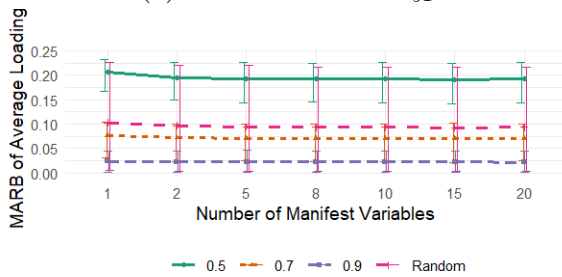
Figure B.14: Cont. Mean absolute relative bias of average loadings for each latent variable in Study V, for different loadings and number of manifest variables



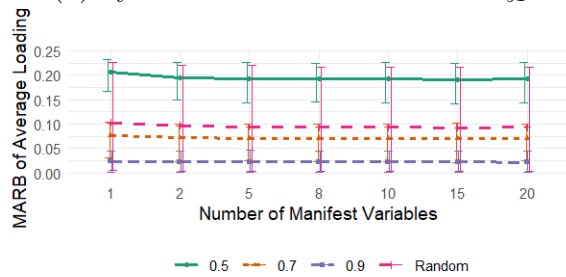
(a) Center Method - ξ_1



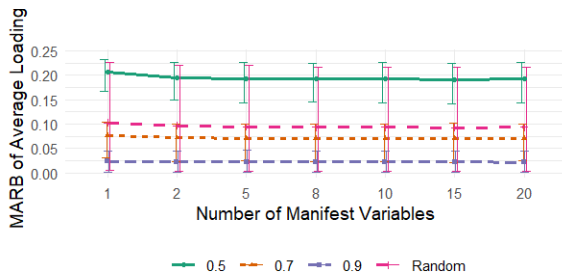
(b) Symbolic Covariance Method - ξ_1



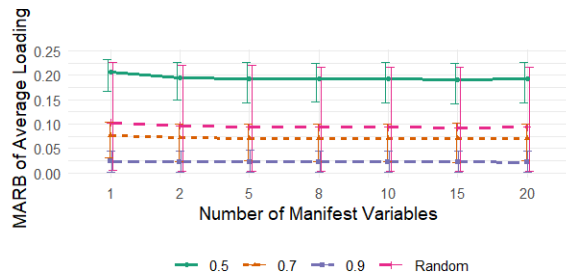
(c) Center Method - ξ_2



(d) Symbolic Covariance Method - ξ_2

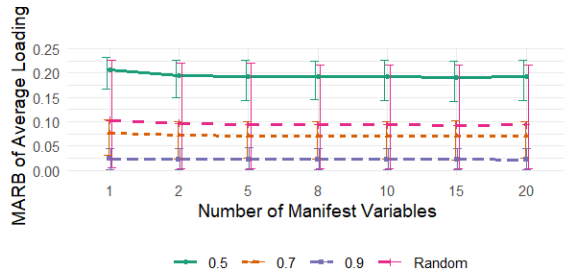


(e) Center Method - ξ_3

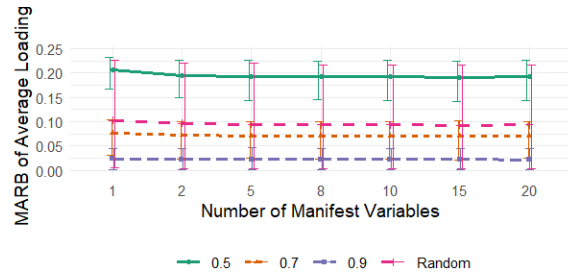


(f) Symbolic Covariance Method - ξ_3

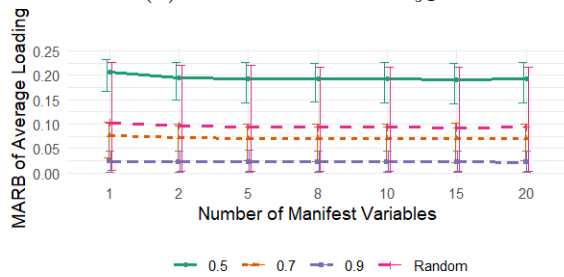
Figure B.15: Mean absolute relative bias of average loadings for each latent variable in Study VI, for different loadings and number of manifest variables



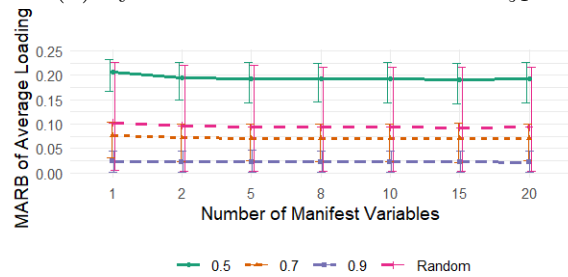
(a) Center Method - ξ_4



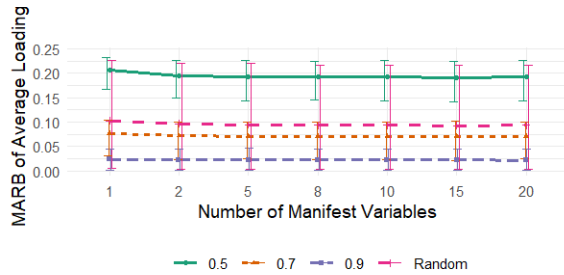
(b) Symbolic Covariance Method - ξ_4



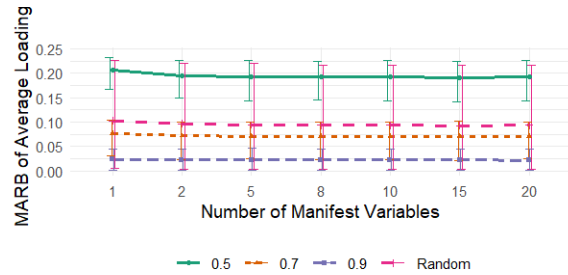
(c) Center Method - ξ_5



(d) Symbolic Covariance Method - ξ_5



(e) Center Method - ξ_6



(f) Symbolic Covariance Method - ξ_6

Figure B.16: Cont. Mean absolute relative bias of average loadings for each latent variable in Study VI, for different loadings and number of manifest variables

B.3 Simulation 3 - Range

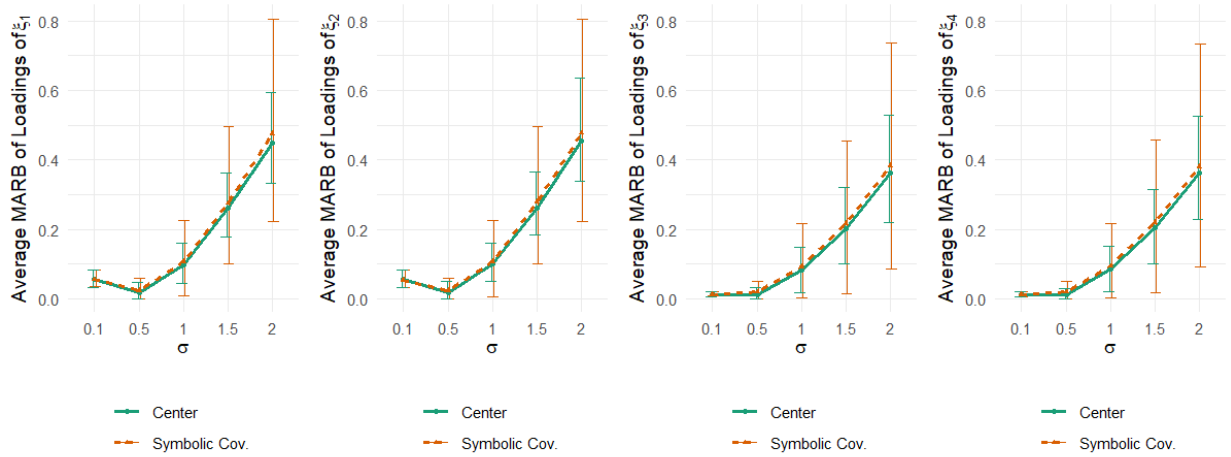


Figure B.17: Mean absolute relative bias of average loadings for each latent variable in Study III, for different σ 's

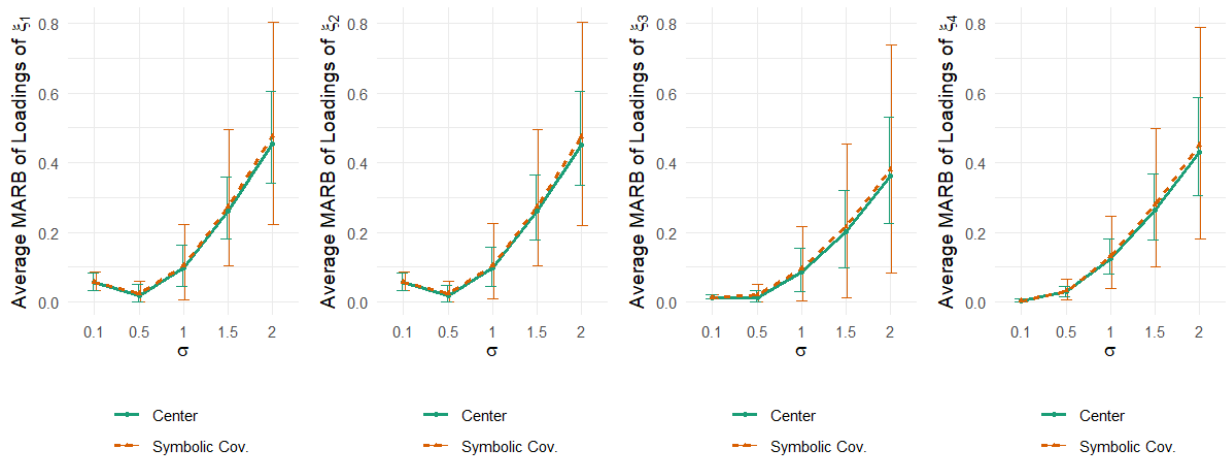


Figure B.18: Mean absolute relative bias of average loadings for each latent variable in Study IV, for different σ 's

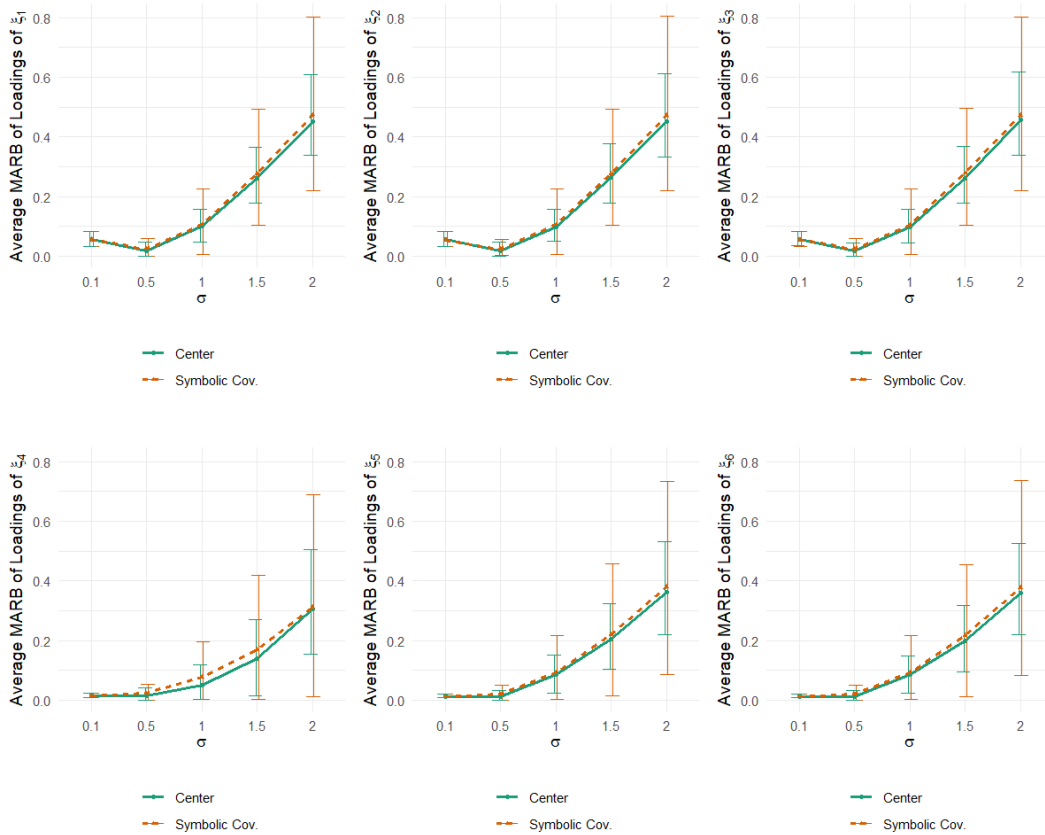


Figure B.19: Mean absolute relative bias of average loadings for each latent variable in Study V, for different σ 's

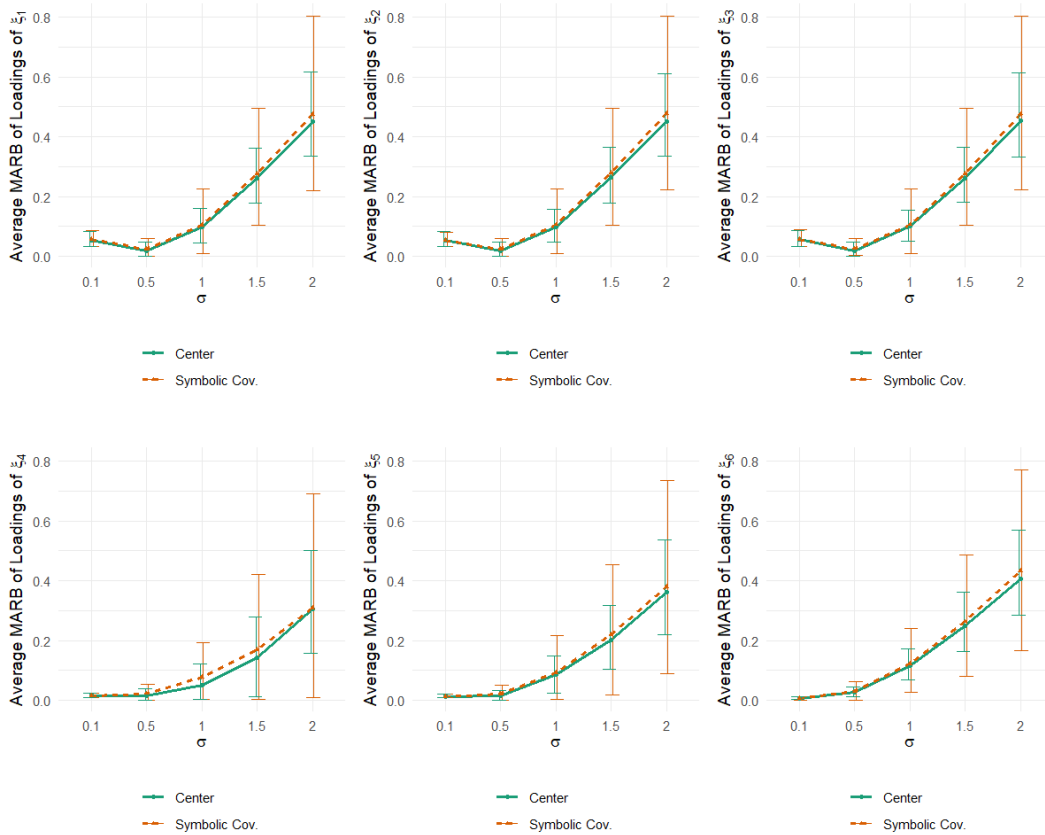


Figure B.20: Mean absolute relative bias of average loadings for each latent variable in Study VI, for different σ 's

B.4 Simulation 4 - Sample Size

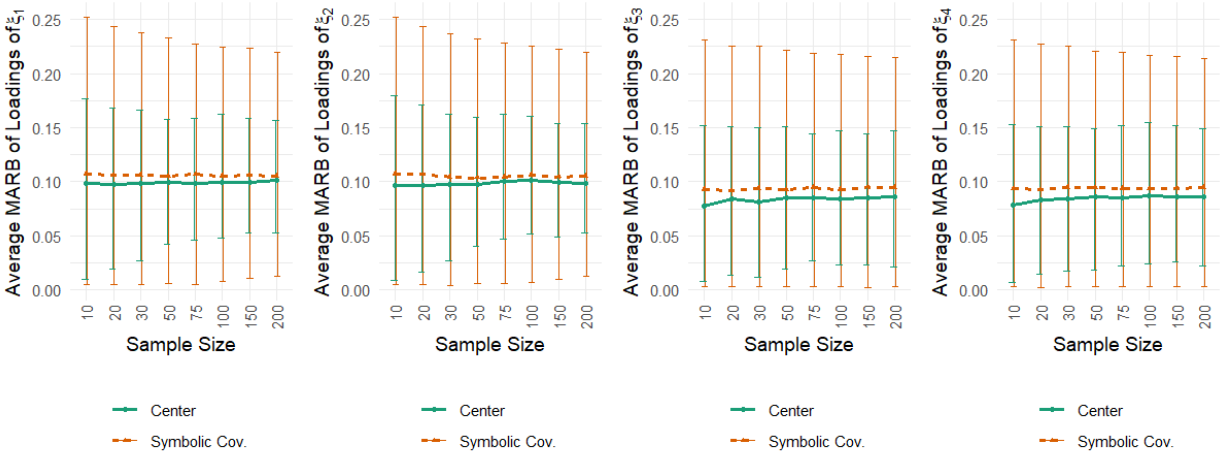


Figure B.21: Mean absolute relative bias of average loadings for each latent variable in Study III, for different sample sizes

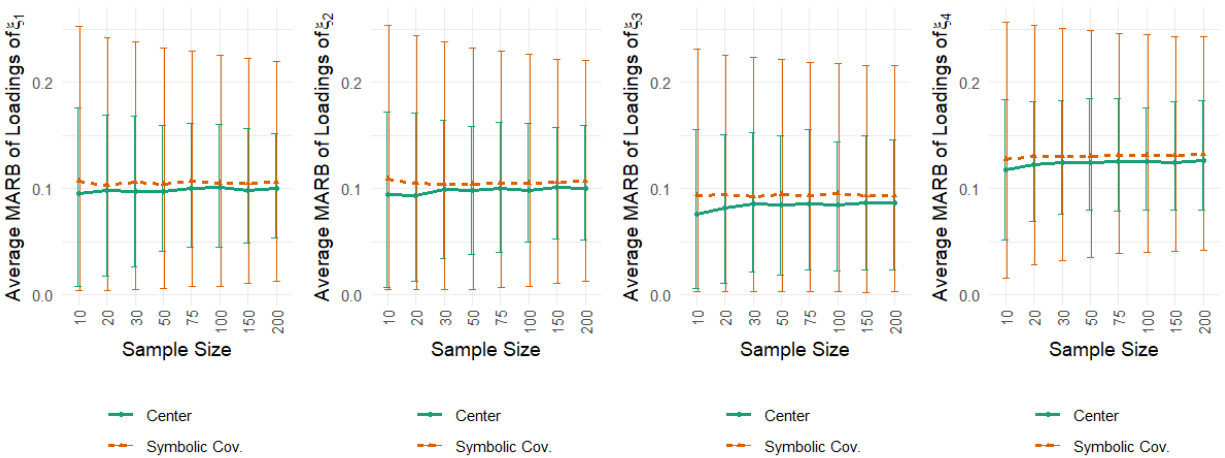


Figure B.22: Mean absolute relative bias of average loadings for each latent variable in Study IV, for different sample sizes

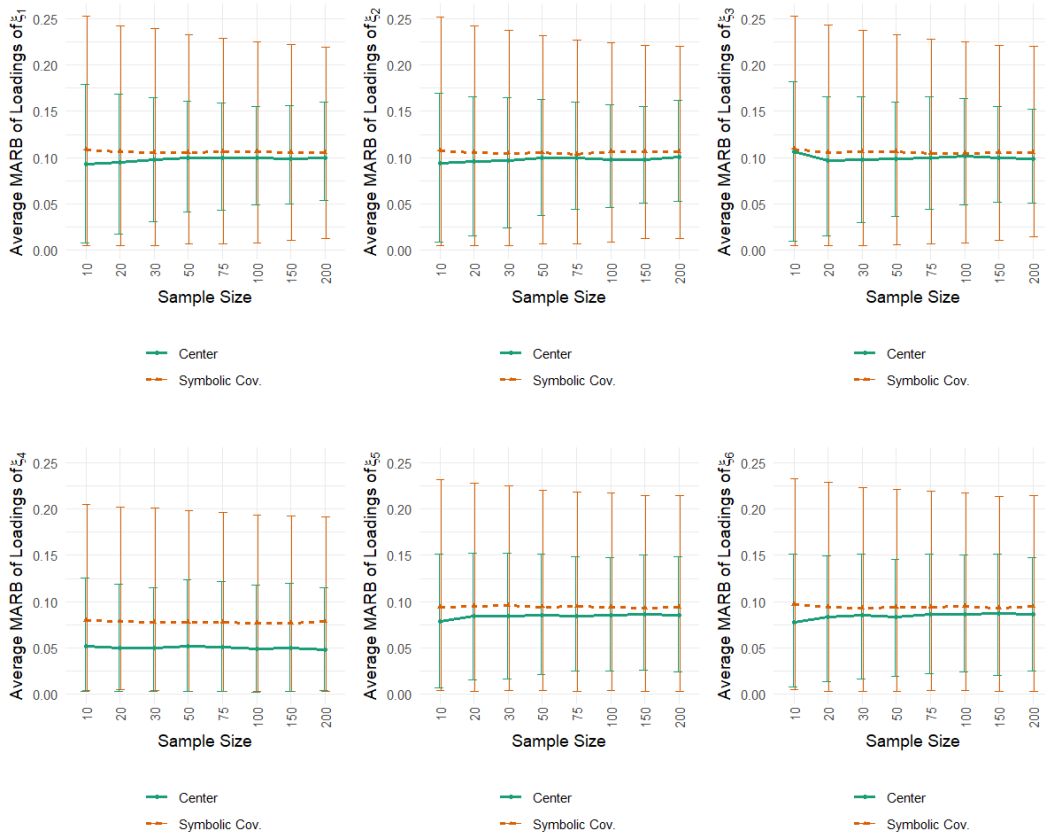


Figure B.23: Mean absolute relative bias of average loadings for each latent variable in Study V, for different sample sizes

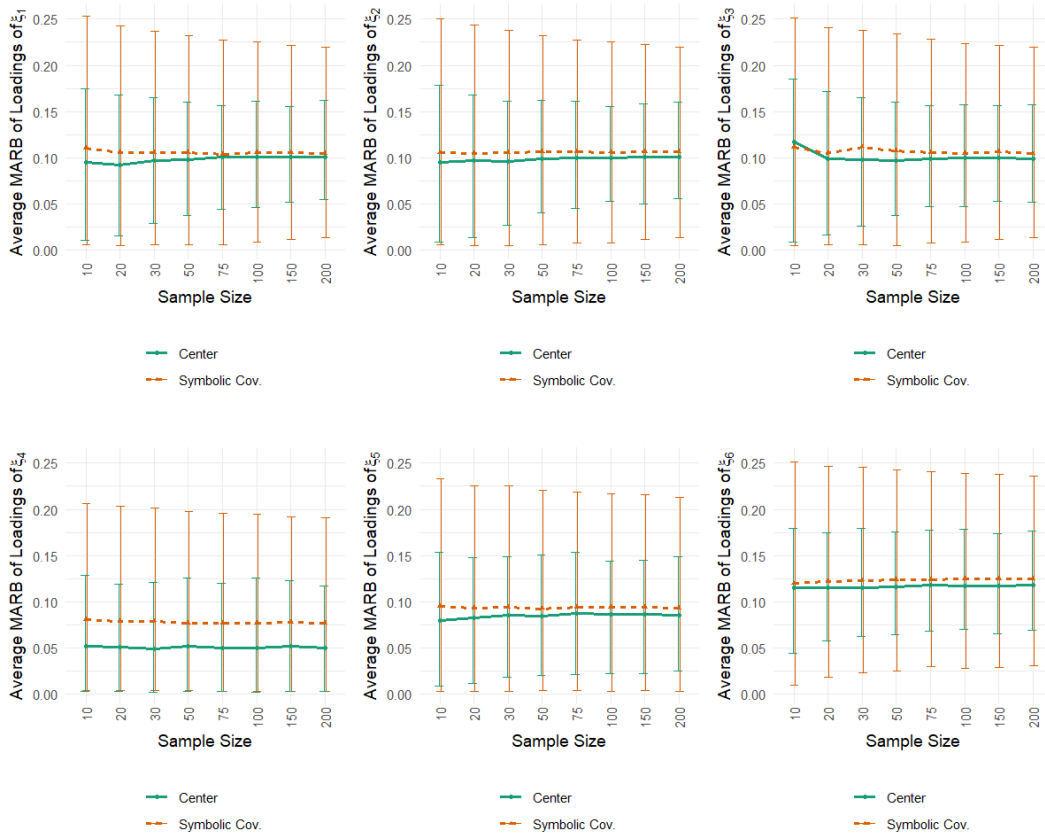


Figure B.24: Mean absolute relative bias of average loadings for each latent variable in Study VI, for different sample sizes

Appendix C

R Code

C.1 Functions for interval-valued data

Mean

```
mean_symb <- function(X){
  ## This function obtains the mean of interval-valued data,
  ## it receives the matrix X containing the interval-valued variables
  ## such that the lower and upper limits of each variable are side by side
  ## and returns a vector (or single value) of the symbolic mean.

  ## If there is an odd number of columns, it will stop
  if (ncol(X)%2 != 0 || is.null(ncol(X)) )
    stop("There should be an even number of columns, since an interval
         variable has a column of lower and another of upper limits")

  p <- ncol(X)/2 ## number of interval variables
  ## vector containing the indexes of the lower bound of the p variables
  index_Xlower <- seq(from=1,to=(2*p),by=2)
  ## vector containing the indexes of the upper bound of the p variables
  index_Xupper <- seq(from=2,to=(2*p),by=2)

  mean_vec <- c() ## this empty vector will receive the symbolic means
  for (j in 1:p){
    mean_vec[j] <- mean((X[,index_Xlower[j]]+X[,index_Xupper[j]])/2)
    ## calculates the symbolic mean for the jth variable
  }
}
```

```

}
return(mean_vec)
}

```

Variance

```

var_symb <- function(X){
  ## This function obtains the variance of interval-valued data,
  ## it receives the matrix X containing the interval variables
  ## such that the lower and upper limits of each variable are side by side
  ## and returns the symbolic vector of variance

  ## If there is an odd number of columns, it will stop
  ## the following message.
  if (ncol(X)%2 != 0 || is.null(ncol(X)) )
    stop("There should be an even number of columns, since an interval
         variable has a column of lower and another of upper limits")

  p <- ncol(X)/2 ## number of interval variables
  n <- nrow(X)   ## number of observations
  ## vector containing the indexes of the lower bound of the p variables
  index_Xlower <- seq(from=1,to=(2*p),by=2)
  ## vector containing the indexes of the upper bound of the p variables
  index_Xupper <- seq(from=2,to=(2*p),by=2)

  varS <- c() ## initializes the vector that will receive the
             ## variance of each of the p interval variables
  for (j in 1:p){
    ## calculates the first term of the variance
    term1 <- sum(X[,index_Xlower[j]]^2+X[,index_Xlower[j]]*
                X[,index_Xupper[j]]+X[,index_Xupper[j]]^2)/(3*n)
    ## calculates the second term of the variance
    term2 <- (sum(X[,index_Xlower[j]]+X[,index_Xupper[j]])^2)/(2*n)^2
    varS[j] <- term1-term2 ## calculates the variance of jth variable
  }
  return(varS)
}

```

Covariance

```
cov_symb <- function(X, method='old'){
  ## This function obtains the matrix of variance and covariance of
  ## interval-valued data, it receives the matrix X containing the data
  ## such that the lower and upper limits of each variable are side by side
  ## and returns the matrix of symbolic variance and covariance.
  ## Can be done using any of the two covariance formulas for interval data
  ## method=old is the method presented in Billard and Diday (2006a)
  ## method=new if the method presented in Billard (2008)

  ## If there is an odd number of columns, it will stop
  if (ncol(X)%2 != 0 || is.null(ncol(X)) )
    stop("There should be an even number of columns, since an interval variable
         has a column of lower and another of upper limits")

  ## Let's check if the method was correctly identified, or it will stop
  if (method=='n' || method=='ne' || method=='new' || method=='N' ||
      method=='Ne' || method=='NE' || method=='New' || method=='NEW') {method='new'
  } else {
    if (method=='o' || method=='ol' || method=='old' || method=='O' ||
        method=='Ol' || method=='OL' || method=='Old' || method=='OLD')
      {method='old'
    } else {stop(" 'method' should be defined either as 'old' or 'new'." )}}

  X <- as.matrix(X) ## makes sure the data is in matrix form
  p <- ncol(X)/2 ## number of interval variables
  n <- nrow(X) ## number of observations

  ## vector containing the indexes of the lower bound of the p variables
  index_Xlower <- seq(from=1,to=(2*p),by=2)
  ## vector containing the indexes of the upper bound of the p variables
  index_Xupper <- seq(from=2,to=(2*p),by=2)

  ## initializes the matrix that will receive the variance and covariances
  Cov_matrix <- matrix(NA, ncol=p,nrow=p)
  variances <- var_symb(X) ## calculates variance of the p variables

  for (j in 1:p){
    ## assigns the variances to the main diagonal
    Cov_matrix[j,j] <- variances[j]
```

```

}

## If p=1, it will return just the variance, otherwise it will continue
if (p>1){
  ## since the matrix is symmetric and we already did the diagonal,
  ## we will calculate the upper triangle and assign it to lower triangle
  for (k in 1:(p-1)){
    for (j in 2:p){
      ## defines what is the first variable to covariate
      Y1 <- matrix(X[,c(index_Xlower[j],index_Xupper[j])],ncol=2)
      ## defines what is the second variable to covariate
      Y2 <- matrix(X[,c(index_Xlower[k],index_Xupper[k])],ncol=2)
      mean1 <- mean_symb(Y1) ## finds the symbolic mean of the first var.
      mean2 <- mean_symb(Y2) ## finds the symbolic mean of the second var.

      ## covariance is calculated and assigned to the lower and upper triangles
      if (method=="new"){
        ## calculates the upper triangle
        Cov_matrix[k,j] <- sum(2*(Y1[,1]-mean1)*(Y2[,1]-mean2)+(Y1[,1]-mean1)*
                              (Y2[,2]-mean2)+(Y1[,2]-mean1)*(Y2[,1]-mean2)+
                              2*(Y1[,2]-mean1)*(Y2[,2]-mean2))/(6*n)
        Cov_matrix[j,k] <- Cov_matrix[k,j] ## assigns it to the lower triangle
      }
      if (method=="old"){
        ## using the same notation as the formula, calculates the covariance
        G1 <- rep(1,times=n); G2 <- rep(1,times=n);
        for (i in 1:n){
          if ((Y1[i,1]+Y1[i,2])/2 <=mean1) G1[i] <- -1
          if ((Y2[i,1]+Y2[i,2])/2 <=mean2) G2[i] <- -1
        }
        Q1 <- (Y1[,1]-mean1)^2+(Y1[,1]-mean1)*(Y1[,2]-mean1)+(Y1[,2]-mean1)^2
        Q2 <- (Y2[,1]-mean2)^2+(Y2[,1]-mean2)*(Y2[,2]-mean2)+(Y2[,2]-mean2)^2

        ## Calculates the covariance and assigns it to lower and upper triangles
        Cov_matrix[k,j] <- Cov_matrix[j,k] <- sum(G1*G2*sqrt(Q1*Q2))/(3*n)
      }
    }
  }
  return(Cov_matrix)
}

```

Correlation

```
cor_symb <- function(X,method='old'){
  ## This function obtains the correlation matrix of interval-valued data,
  ## it receives the matrix X containing the interval variables
  ## such that the lower and upper limits of each variable are side by side
  ## and returns the matrix of symbolic correlation.
  ## Can be done using any of the two covariance formulas for interval data
  ## method=old is the method presented in Billard and Diday (2006a)
  ## method=new if the method presented in Billard (2008)

  ## If there is an odd number of columns, it will stop
  if (ncol(X)%2 != 0 || is.null(ncol(X)) )
    stop("There should be an even number of columns, since an interval variable
         has a column of lower and another of upper limits")

  p <- ncol(X)/2 ## number of interval variables
  ## vector containing the indexes of the lower bound of the p variables
  index_Xlower <- seq(from=1,to=(2*p),by=2)
  ## vector containing the indexes of the upper bound of the p variables
  index_Xupper <- seq(from=2,to=(2*p),by=2)

  sd_symb <- sqrt(var_symb(X)) ## calculates standard deviation of variables
  Cov_matrix <- cov_symb(X,method) ## calculates the covariance

  ## initializes the matrix that will store the correlations
  Cor_matrix <- matrix(NA, ncol=p,nrow=p)
  for (k in 1:p)
    for (j in 1:p){
      ## for each cell of the matrix it divides the covariance by
      ## the product of the standard deviations
      Cor_matrix[k,j] <- Cov_matrix[k,j]/(sd_symb[k]*sd_symb[j])
    }
  return(Cor_matrix)
}
```

Standardizing variables

```
stand_symb <- function(X,center=T,scale=T){
  ## This function obtains standardized interval-valued variables,
  ## it receives the matrix X containing the interval variables
```

```

## such that the lower and upper limits of each variable are side by side
## and returns a matrix of standardized interval variables.
## center=T each variable will be centered to have mean 0
## scale=T each variable will be scaled to have variance 1

## If there is an odd number of columns, it will stop
if (ncol(X)%2 != 0 || is.null(ncol(X)) )
  stop("There should be an even number of columns, since an interval variable
       has a column of lower and another of upper limits")

p <- ncol(X)/2 ## number of interval variables
n <- nrow(X)   ## number of observations
## vector containing the indexes of the lower bound of the p variables
index_Xlower <- seq(from=1,to=(2*p),by=2)
## vector containing the indexes of the upper bound of the p variables
index_Xupper <- seq(from=2,to=(2*p),by=2)

## initializes the matrix that will receive the standardized data
standar <- matrix(NA,ncol=2*p,nrow=n)
for (j in 1:p){
  ## calculates the symbolic mean for the jth interval variable
  if (center==T) {
    meanS <- mean_symb(X[,c(index_Xlower[j],index_Xupper[j])])
  } else {
    ## if center==F, sets the mean to 0, so it won't affect the calculations
    meanS <- 0 }
  ## calculates the symbolic standard deviation for the jth interval variable
  if (scale==T) {
    sdS <- sqrt(var_symb(X[,c(index_Xlower[j],index_Xupper[j])]))
  } else {
    ## if scale==F, sets the sd to 1, so it won't affect the calculations
    sdS <- 1 }
  ## standardize the lower limit of the jth variable and all observations
  standar[,index_Xlower[j]] <- (X[,index_Xlower[j]]-meanS)/sdS
  ## standardize the upper limit of the jth variable and all observations
  standar[,index_Xupper[j]] <- (X[,index_Xupper[j]]-meanS)/sdS
}
return(standar)
}

```

Centering variables

```
center_symb <- function(X){
  ## This function obtains centered data,
  ## bringing it back to classical observations.
  ## It receives the matrix X containing the interval variables
  ## such that the lower and upper limits of each variable are side by side
  ## and returns a matrix of centered (classical) data.

  ## If there is an odd number of columns, it will stop
  if (ncol(X)%2 != 0 || is.null(ncol(X)) )
    stop("There should be an even number of columns, since an interval
         variable has a column of lower and another of upper limits")

  p <- ncol(X)/2 ## number of interval variables
  n <- nrow(X)    ## number of observations
  ## vector containing the indexes of the lower bound of the p variables
  index_Xlower <- seq(from=1,to=(2*p),by=2)
  ## vector containing the indexes of the upper bound of the p variables
  index_Xupper <- seq(from=2,to=(2*p),by=2)

  ## initializes the matrix that will receive the centered data
  centered <- matrix(NA,ncol=p,nrow=n)
  for (j in 1:p){
    ## calculates the center of the interval of each obs. for the jth variable
    centered[,j] <- apply(X[,c(index_Xlower[j],index_Xupper[j])],1,mean)
  }
  return(centered)
}
```

Linear regression

```
lm_symb <-function(X,Y,intercept=T,type=3,method='old'){
  ## This function conducts linear regression on interval valued variables
  ## It receives the interval-valued explanatory variables (X), and
  ## the interval response variable (Y). This function accepts more than
  ## one explanatory variable, but only one response variable.
  ## The lower and upper limits of each variable are side by side.
  ## 'intercept' indicates whether the estimation should include intercept,
  ## the default is to include the intercept.
  ## 'type' indicates what linear regression method should be used
```

```

## type=1 is for the center method
## type=2 is for the center and range method
## type=3 is for the symbolic covariance method
## Can be done using any of the two covariance formulas for interval data
## method=old is the method presented in Billard and Diday (2006a)
## method=new if the method presented in Billard (2008)
## The function returns the estimated coefficients, interval-valued predicted
## values and interval-valued residuals

## If there is an odd number of columns in X, it will stop
if (ncol(X)%2 != 0 || is.null(ncol(X)))
  stop("There should be an even number of columns, since an interval
       variable has a column of lower and another of upper limits")
## If Y has more or less than 2 columns, it will stop
if (ncol(Y) != 2)
  stop("The matrix with the response variables must have 2 columns,
       one with the lower and other with the upper bound")

## Let's check if the type was correctly identified
if (type=="cent" || type=="center" || type=="centered" || type==1) type=1
if (type=="centerrange" || type=="rangecenter" ||
    type=="centerandrange" || type==2) type=2
if (type=="covariance" || type=="covar" || type=="cov" || type==3) type=3

## Let's check if the method was correctly identified, or it will stop
if (method=='n' || method=='ne' || method=='new' || method=='N' ||
    method=='Ne' || method=='NE' || method=='New' || method=='NEW') {method='new'
} else {
  if (method=='o' || method=='ol' || method=='old' || method=='O' ||
      method=='Ol' || method=='OL' || method=='Old' || method=='OLD')
    {method='old'
  } else {stop(" 'method' should be defined either as 'old' or 'new'.")}}

p <- ncol(X)/2 ## number of interval variables
n <- nrow(X)   ## number of observations

if (type == 1 || type==2){
  ## for the center and center and range methods, if the estimation includes
  ## an intercept, add two columns of 1 in the design matrix, and make p=p+1
  if (intercept==TRUE) {
    X <- cbind(rep(1,times=n),rep(1,times=n),X)
  }
}

```

```

    p <- p+1
  }
}
## vector containing the indexes of the lower bound of the p variables
index_Xlower <- seq(from=1,to=(2*p),by=2)
## vector containing the indexes of the upper bound of the p variables
index_Xupper <- seq(from=2,to=(2*p),by=2)

if (type==1){ ## for center method
  Y_c <- center_symb(Y) ## obtains the centered response variable
  X_c <- center_symb(X) ## obtains the centered design matrix

  ## calculates the coefficients
  coefficients <- solve(t(X_c)%*%X_c)%*%t(X_c)%*%Y_c

  ## obtains the lower bound predictors
  predictionsLower <- X[,index_Xlower]%*%coefficients
  ## obtains the upper bound predictors
  predictionsUpper <- X[,index_Xupper]%*%coefficients

  residuals <- Y-cbind(predictionsLower,predictionsUpper)
  return(list(coef=coefficients, pred=
             cbind(predictionsLower,predictionsUpper),resid=residuals))
}
if (type==2){ ## for center and range method
  ## obtains the centered vector of response variable
  Y_c <- matrix(center_symb(Y), ncol=1)
  ## obtains the vector of half ranges of the response variable
  Y_r <- matrix((Y[,2]-Y[,1])/2, ncol=1 ,nrow=n)
  ## obtains the centered design matrix
  X_c <- matrix(center_symb(X), ncol=p)
  ## obtains the matrix of half ranges of the design matrix
  X_r <- (X[,index_Xupper]-X[,index_Xlower])/2

  ## makes sure the first column of the range matrix of design matrix is 1
  if (intercept==TRUE) X_r[,1] <- rep(1,times=n)

  ## estimates the center coefficients
  coefficients_c <- solve(t(X_c)%*%X_c)%*%t(X_c)%*%Y_c
  ## estimates the range coefficients
  coefficients_r <- solve(t(X_r)%*%X_r)%*%t(X_r)%*%Y_r
}

```

```

## calculates the lower predictors
predictionsLower <- X_c%%coefficients_c - X_r%%coefficients_r
## calculates the upper predictors
predictionsUpper <- X_c%%coefficients_c + X_r%%coefficients_r

residuals <- Y-cbind(predictionsLower,predictionsUpper)
return(list(coef_center=coefficients_c,coef_range=coefficients_r, pred=
          cbind(predictionsLower,predictionsUpper),resid=residuals))
}
if (type==3){ ## for symbolic covariance method
  if (intercept==F) {
    ## installs and load the crayon package into R, to make note red
    if(!require("crayon")) { install.packages("crayon"); require("crayon") }
    ## returns the following message
    cat(red("The estimation of the coefficients (other than the intercept) does
           not differ whether it is with or without intercept. We can force the
           intercept to be 0, by centering the data, but the estimation of the
           other parameters will be the same, since the covariance between
           variables is the same when they are centered or not."))
    X <- stand_symb(X,scale=F) ## centers the explanatory variables
    Y <- stand_symb(Y,scale=F) ## centers the response variable
  }

Cov_X_matrix <- cov_symb(X,method) ## calculates the covariance matrix of X
## calculates the covariance between the response and the design matrix, and
## obtain just the vector of covariance between Y and X
Cov_YX_matrix <- matrix(cov_symb(cbind(Y,X),method)[-1,1],ncol=1)

## estimates the coefficients
coefficients <- solve(n*Cov_X_matrix)%%(n*Cov_YX_matrix)

predictionsLower <- X[,index_Xlower]%%coefficients ## predicts lower bound
predictionsUpper <- X[,index_Xupper]%%coefficients ## predicts upper bound

## then we need to estimate the intercept,
## if intercept=F, it will be 0 (or almost0), since the var. were centered
beta0 <- mean_symb(Y)- t(coefficients)%%mean_symb(X)
coefficients <- rbind(beta0,coefficients)
predictionsLower <- predictionsLower + as.numeric(beta0)
predictionsUpper <- predictionsUpper + as.numeric(beta0)

```

```

## makes sure the predicted lower bound is the minimum of the predictions,
## and the predicted upper bound is the maximum
predictions <- t(apply(cbind(predictionsLower,predictionsUpper),1,sort))
residuals <- Y-predictions
return(list(coef=coefficients, pred=predictions,resid=residuals))
}
}

```

Interval-valued PLS-PM

```

pls_symb <- function(data, blocks, path, Mode=NULL, scheme="centroid",
                    maxiter=100, tol=10^(-6),scaled=F,type="cov",method="old",
                    gof=FALSE,ambiguity=FALSE){
## This function conducts the PLS-PM estimation for interval-valued data.
## data is a matrix with the interval valued manifest variables (MV)
## such that the upper and lower bounds for each MV should be side by side.
## blocks is a list, each element of the list contains a vector of indices
## of the variables that correspond to that latent variable (LV).
## path is a square matrix indicating the relationship between LV
## the rows indicate the LV that have an arrow from the columns variable
## pointing at it, and columns are the LV that have the other end of the arrow.
## Mode is a vector with the mode of each LV
## Mode "A" is used when the MV relate to the LV in a reflective way
## Mode "B" is used when the MV relate to the LV in a formative way
## scheme can be "centroid", "factor" or "path"
## maxiter is the maximum iterations to try if it does not converge
## before it reaches it. Default is 100.
## tol is the tolerance acceptable for convergence, default is 10^(-6)
## scaled=F means that the MV have not been standardized yet, and
## the function will standardize it. If scaled=T, will assume that it has been
## standardized, and will center the interval variables, just to guarantee that
## the variables are centered.
## 'type' indicates what linear regression method should be used,
## does not accept the center and range method.
## type=1 or type="center" is for the center method
## type=3 or type="cov" is for the symbolic covariance method
## method indicates the covariance formula to be used
## method=old is the method presented in Billard and Diday (2006a)
## method=new if the method presented in Billard (2008)
## if gof=T the model validation measures will be calculated

```

```

## ambiguity=T checks and corrects for sign ambiguity

## installs and load the packages into R
if(!require("turner")) { install.packages("turner"); require("turner") }
if(!require("Matrix")) { install.packages("Matrix"); require("Matrix") }

## Let's check if the type was correctly identified
if (type=="cov" || type=="covar" || type=="covariance" || type==3) {
  type=3
} else if (type=="cen" || type=="center" || type=="centered" || type==1) {
  type=1
} else stop("This function will use either the centered ('center') or
           the symbolic covariance ('covariance') regression methods")

## Let's check if the method was correctly identified, or it will stop
if (method=='n' || method=='ne' || method=='new' || method=='N' ||
    method=='Ne' || method=='NE' || method=='New' || method=='NEW') {method='new'
} else {
  if (method=='o' || method=='ol' || method=='old' || method=='O' ||
      method=='Ol' || method=='OL' || method=='Old' || method=='OLD')
    {method='old'
  } else {stop(" 'method' should be defined either as 'old' or 'new'.")}}

if (ncol(data)%2 != 0)
  stop("There should be an even number of columns, since an interval
       variable has a column of lower and another of upper limits")

lvs <- nrow(path) ## number of latent variables
mv <- ncol(data)/2 ## number of manifest variables

Row_names <- LV_names <- NULL ## receive the var. names, if they are provided
## if the manifest variables had names when loaded, this will store the names
if (!is.null(colnames(data))) Row_names <- colnames(data)

## sees if either the row or columns of the paths had names
if (!is.null(colnames(path)) || !is.null(rownames(path))) {
  if (!is.null(colnames(path))==TRUE) {
    LV_names <- colnames(path) ## if the columns had names, stores it
  } else {LV_names <- rownames(path)} ## otherwise stores from the rows
}

```

```

## if Mode is not provided, assume that it is A
if (is.null(Mode)) Mode <- rep("A", times=lvs)
if (length(Mode)!=lvs) stop("If Mode is defined, there should be one
                             Mode specified for each latent variable")

## Checks if the blocks were defined correctly
if (sum(unlist(blocks)!=seq(1:mv))>1) stop("The data and blocks should be
      in the same order, i.e. block 1 should go from manifest variables
      1,...,k1, block 2 from k1+1,...,k2, and so on.")

## vector containing the indexes of the lower bound of the p variables
index_lower <- seq(from=1,to=(2*mv),by=2)
## vector containing the indexes of the upper bound of the p variables
index_upper <- seq(from=2,to=(2*mv),by=2)

## standardizes the data if it has not been done previously
if (scaled==F) data <- stand_symb(data)
## makes sure that the data is centered if it has already been scaled
if (scaled==T) data <- stand_symb(data,center=T,scale=F)

## Let's check the dimensionality of the blocks
h <- 1
Alpha <- Rho <- c()
for (j in 1:lvs){
  kj <- length(blocks[[j]]) ## obtains number of MV of jth LV
  S <- cov_symb(data[,index_lower[h]:index_upper[(h+kj-1)]],method)
  Alpha_N <- sum(S)-sum(diag(S))
  Alpha_D <- sum(S)
  Alpha[j] <- (Alpha_N/Alpha_D)*(kj/(kj-1)) ## calculates alpha
  t1 <- matrix(NA,ncol=2,nrow=length(data[,1]))
  t1[,1] <- data[,index_lower[h:(h+kj-1)]]%*%matrix(eigen(S)$vectors[,1],
                                                    ncol=1)
  t1[,2] <- data[,index_upper[h:(h+kj-1)]]%*%matrix(eigen(S)$vectors[,1],
                                                    ncol=1)
  Rho_N <- sum(cov_symb(cbind(t1,data[,index_lower[h]:index_upper[(h+kj-1)]]),
                        method=method)[-1,1])^2
  Rho_D <- Rho_N + kj - sum(cov_symb(cbind(t1,data[,index_lower[h]:
                                                    index_upper[(h+kj-1)]]),
                        method=method)[-1,1]^2)
  Rho[j] <- Rho_N/Rho_D ## calculates rho
  h <- h +kj
}

```

```

}
unidim <- cbind(Alpha,Rho)
## Then, we assign the names to the unidim, if there were names
if (!is.null(LV_names)) rownames(unidim) <- LV_names

## using the weights function, we calculate the weights, and store it
w_list <- weights_symb(data=data,blocks=blocks,path=path,maxiter=maxiter,
                      tol=tol,Mode=Mode,scheme=scheme,mv=mv,lvs=lvs,
                      index_lower=index_lower,index_upper=index_upper,
                      type=type,method=method,ambiguity=ambiguity)
w <- w_list$W ## obtains the weights
iter <- w_list$iter ## stores the number of iterations until convergence

y <- stand_symb(data%*%w) ## estimates the LV using the weights calculated

## This matrix will receive the coef. for the structural model
mss <- matrix(0,ncol=lvs,nrow=lvs)
Load <- c() ## will receive the coefficients for the measurement model
for (j in 1:lvs){
  ## Start by estimating the coefficients for structural model

  ## if the path indicates the link between two LV, estimate the coefficient,
  ## where the Y is the LV referred to in the jth row, X is the one (or more)
  ## variables referred in the columns for the row in Y.
  if (sum(path[j,])>0) { ## checks if the variables acts are response var.
    ## index_col_X obtains the indices of the columns of y that act as indep.
    ## variable in this case
    index_col_X <- sort(c(index_lower[1:lvs][path[j,]==1],
                        index_upper[1:lvs][path[j,]==1]))
    ## index_col_Y obtains the indices of the columns of the LV that act as dep.
    ## variable in this case
    index_col_Y <- c(index_lower[j],index_upper[j])
    if (type==1) {
      mss[j,path[j,]==1] <- lm_symb(X=y[,index_col_X],Y=y[,index_col_Y],
                                   type=type,intercept=F)$coef
    }
    if (type==3) {
      mss[j,path[j,]==1] <- lm_symb(X=y[,index_col_X], Y=y[,index_col_Y],
                                   type=type,intercept=T,method=method)$coef[-1]
      ## remove the first coef. because cov method always outputs intercept
    }
  }
}

```

```

}
## Now we estimate the loadings
## For each LV we calculate the loadings for each MV
if (Mode[j]=="A"){
  for (i in unlist(blocks[j])){
    if (type==1) {
      Load[i] <- lm_symb(X=y[,c(index_lower[j],index_upper[j])],
                        Y=data[,c(index_lower[i],index_upper[i])],
                        type=type,intercept=F)$coef
    }
    if (type==3) {
      Load[i] <- lm_symb(X=y[,c(index_lower[j],index_upper[j])],
                        Y=data[,c(index_lower[i],index_upper[i])],
                        type=type,intercept=T,method=method)$coef[-1]
    }
  }
}
if (Mode[j]=="B"){
  for (i in unlist(blocks[j])){
    if (type==1) {
      Load[i] <- lm_symb(Y=y[,c(index_lower[j],index_upper[j])],
                        X=data[,c(index_lower[i],index_upper[i])],
                        type=type,intercept=F)$coef
    }
    if (type==3) {
      Load[i] <- lm_symb(Y=y[,c(index_lower[j],index_upper[j])],
                        X=data[,c(index_lower[i],index_upper[i])],
                        type=type,intercept=T,method=method)$coef[-1]
    }
  }
}
}

## The next part of the code consists of organizing the output and
## naming rows and columns of objects that will be in the output.

loadings <- data.frame(Load) ## makes the loadings into a data frame
## organizes the weights to be output as a matrix
w <- list_to_dummy(blocks)*(apply(w,1,sum)[seq(from=1, to=(2*mv),by=2)])
y <- data.frame(y) ## makes LV into data frame, so we can name the columns

## if the MV had names when loaded, it assigns them to the rows.

```

```

if (!is.null(Row_names)) {
  new_row_names <- c()
  for(i in 1:mv){
    ## since we have two columns for each variable but weights and loading
    ## are the same for lower and upper bounds, need to concatenate the name
    ## for the lower and upper bounds.
    new_row_names[i] <- paste0(Row_names[index_lower[i]],"/",
                              Row_names[index_upper[i]])
  }
  rownames(loadings) <- new_row_names
  rownames(w) <- new_row_names
}
## if the LV had names, assign it to the variables that will be in the output
if (!is.null(LV_names)){
  colnames(mss) <- rownames(mss) <- LV_names
  colnames(w) <- LV_names
  new_col_names <- c()
  for (j in 1:lvs){
    ## the scores have lower and upper bounds, need to adjust the columns names
    new_col_names[index_lower[j]] <- paste0(LV_names[j], "_l")
    new_col_names[index_upper[j]] <- paste0(LV_names[j], "_u")
  }
  colnames(y) <- new_col_names
}

## If gof=TRUE, need to calculate the model validation estimates
if(gof==TRUE){
  communality <- kj<- redundancy <- R2 <- c() ## initializes vectors
  h <- 1 ## auxiliary index
  for (j in 1:lvs){
    kj[j] <- length(blocks[[j]])
    ## dat_j separates the MV associated with jth LV
    dat_j <- data[,index_lower[h]:index_upper[(h+kj[j]-1)]]
    ## then calculates the communality
    communality[j] <- sum(cor_symb(as.matrix(cbind(
      y[,index_lower[j]:index_upper[j]],dat_j)),
      method=method)[-1,1]^2)/kj[j]
    ## calculates the r2 and redundancy for LV that act as dependent var.
    if (sum(path[j,])>0){
      ## dat_r2 obtains the columns that will be correlated for r2 for jth LV
      dat_r2 <- cbind(y[,index_lower[j]:index_upper[j]],y[,-c(index_lower[j],

```

```

                                                                    index_upper[j]))
    r2_prov <- mss[j,-j]%*%cor_symb(as.matrix(dat_r2),method=method)[-1,1]
    R2 <- c(R2,r2_prov)
    redundancy_prov <- communality[j]*r2_prov
    redundancy <- c(redundancy,redundancy_prov)
  }
  h <- h+kj[j] ## updates h
}
average_communality <- sum(communality*kj)/sum(kj)
average_R2 <- mean(R2)
GoF <- sqrt(average_communality*average_R2)

## Next we organize the model validation results to be output
validation <- list(Communality=communality,Communality_bar =
                  average_communality, R2=R2,Redundancy=redundancy,GoF=GoF)
return(list(weights=w,scores=y, path_coef= mss,loadings=loadings,iter=iter,
           manifest=data,unidimensionality=unidim,gof=validation))
}

if(gof==FALSE) return(list(weights=w,scores=y, path_coef= mss,loadings=loadings,
                          iter=iter,manifest=data,unidimensionality=unidim))
}

```

```

weights_symb <- function(data,blocks, path, maxiter, tol, Mode, scheme, mv, lvs,
                        index_lower, index_upper, type, method,ambiguity=FALSE){
  ## Function to be used in pls_symb().
  ## The arguments are the same as used in pls_symb().
  ## This function calculates the weights for the PLS-PM for interval-valued data
  ## After an iterative procedure, this function will return the weights
  ## associated with each manifest variable.

  ## installs and load the package into R
  if(!require("turner")) { install.packages("turner"); require("turner") }

  ## kronecker is a block multiplication that will be used to guarantee
  ## that the lower and upper bounds of each variable is accessed correctly.
  ## For this first multiplication, it multiplies and replaces each element
  ## of the matrix created by the structure in the blocks by an identity
  ## matrix of 2 rows and columns.
  ## W_tilda refers to the weights before they have been standardized

```

```

W_tilda <- kronecker(list_to_dummy(blocks),diag(2))
## Then we store the weight to verify convergence later
W_old <- apply(list_to_dummy(blocks),1,sum)

## L refers to the external estimations
L <- stand_symb(data%%W_tilda) ## obtain and stand. 1st external estimation
if (ambiguity==T) {
  ## corrects if there is sign ambiguity, leave as is if there isn't
  L <- sign_ambiguity_intervaldata(L,data,blocks=blocks,method=method)}

iter=0 ;err <- 1 ## initializes the error and iteration

while(iter<=maxiter && err>tol){
  ## This loop will keep running until either the maximum number of iteration
  ## is reached, or the error is smaller than the tolerance
  iter = iter+1 #updates the iteration number

  ## based on the type of scheme, we need to calculate the matrix E
  if (scheme=="centroid") E <- (path+t(path))*sign(cor_symb(L,method=method))
  if (scheme=="factor") E <- (path+t(path))*cor_symb(L,method=method)
  if (scheme=="path"){
    ## follower matrix is the path matrix and predictor matrix is its transpose
    follower_mat <- path==1
    predictor_mat <- t(path)==1
    E <- matrix(0, ncol=lvs, nrow=lvs)
    for(j in 1:lvs){
      if (sum(follower_mat[,j])>0) {
        E[follower_mat[,j],j] <- cor_symb(L,method=method)[j,follower_mat[,j]]
      }
      if (sum(predictor_mat[,j])>0) {
        ## index_lower_col selects the columns of L to be used as the
        ## explanatory variables.
        index_lower_col <- sort(c(index_lower[1:lvs][predictor_mat[,j]],
                                index_upper[1:lvs][predictor_mat[,j]]))

        if (type==1) {
          E[predictor_mat[,j],j] <- lm_symb(X=L[,index_lower_col],
                                             Y= L[,c(index_lower[j],index_upper[j])],
                                             type=type,intercept=F)$coef
        }
        if (type==3){
          E[predictor_mat[,j],j] <- lm_symb(X=L[,index_lower_col],

```

```

                                Y=L[,c(index_lower[j],index_upper[j])],
                                type=type,intercept=T,
                                method=method)$coef[-1]
                                }
                                }
                                }
}
## adjusts the E matrix for the interval data, by duplicating it
## since the weights don't differ for lower and upper bounds
E <- kronecker(E,diag(2))

## Z refers to the inner estimations
Z <- stand_symb(L%%E) ## updates inner estimates

W_tilda <- c()
for (j in 1:lvs){
  ## this loop will estimate the w_tilda for each latent variable
  if (Mode[j]=="A"){
    for (i in unlist(blocks[j])){
      if (type==1) {
        W_tilda[i] <- lm_symb(X=Z[,c(index_lower[j],index_upper[j])],
                              Y=data[,c(index_lower[i],
                                          index_upper[i])],
                              type=type,intercept=F)$coef
      }
      if (type==3) {
        W_tilda[i] <- lm_symb(X=Z[,c(index_lower[j],index_upper[j])],
                              Y=data[,c(index_lower[i],index_upper[i])],
                              type=type,intercept=T,
                              method=method)$coef[-1]
      }
    }
  }
}
if (Mode[j]=="B"){
  if (type==1) {
    W_tilda[unlist(blocks[j])] <- lm_symb(Y=Z[,c(index_lower[j],
                                                  index_upper[j])],
                                          X=data[,sort(c(index_lower[
                                                                unlist(blocks[j])),
                                                                index_upper[unlist(
                                                                blocks[j])])])],

```

```

                                type=type,intercept=F)$coef
}
if (type==3) {
  W_tilda[unlist(blocks[j])] <- lm_symb(Y=Z[,c(index_lower[j],
                                             index_upper[j])],
                                       X=data[,sort(c(index_lower[
                                             unlist(blocks[j])),
                                             index_upper[unlist(
                                             blocks[j])])])],
                                       type=type,intercept=T,
                                       method=method)$coef[-1]
}
}
}
## adjusts matrix for the interval data
W <- kronecker(list_to_dummy(blocks)*W_tilda,diag(2))

L <- stand_symb(data%%W) ## updates the external estimates
if (ambiguity==T) {
  ## corrects if there is sign ambiguity, leave as is if there isn't
  L <- sign_ambiguity_intervaldata(L,data,blocks=blocks,method=method)}

W_new <- W_tilda ## updates the weight matrix
err <- max(abs(W_old-W_new))
W_old <- W_new ## updates the value of previous weight
}
## organizes and standardizes the estimated weight to be output
W_st <- W%%diag(rep(1/sqrt(var_symb(data%%W)),each=2))
return(list(W=W_st,iter=iter))
}

```

```

sign_ambiguity_intervaldata <- function(y,x,blocks,method){
  ## Function to be used in pls_symb().
  ## This function checks if there is sign ambiguity between the response y
  ## and this set of explanatory interval-valued variables x.
  ## The response variable for the pls_symb will be the matrix
  ## with the predicted latent variables
  ## The explanatory variable for the pls_symb is the set of MV associated
  ## with that LV.
  ## It also receives the list of blocks from the pls_symb().
  ## If there is, it returns the vector y adjusted so that the majority of

```

```

## explanatory variables is positively correlated with it. If there is no
## sign ambiguity, it returns the initial vector y.

p <- ncol(y)/2
mv <- ncol(x)/2
## vector containing the indexes of the lower bound of the p variables
index_lower <- seq(from=1,to=(2*mv),by=2)
## vector containing the indexes of the upper bound of the p variables
index_upper <- seq(from=2,to=(2*mv),by=2)

for (j in 1:p){
  ## checks which manifest variables are associated with jth lv
  mv_var_j <- blocks[[j]]
  ## obtains the data associated only with j
  x_j <- x[,index_lower[min(mv_var_j)]:index_upper[max(mv_var_j)]]
  Sign <- sum(sign(cor_symb(
    cbind(y[,c(index_lower[j],index_upper[j])],x_j),method=method))[-1,1])
  ## for each lv we check how many negative correlations between
  ## the lv and interval mv.
  ## if the sum is 0, there is the same number of positive and negative
  ## correlations, nothing needs to be done
  ## if the sum is positive, there is more positive than negative cor.,
  ## that's what we want
  ## if the sum is negative, there is more negative than positive cor.,
  ## there is sign ambiguity, need to adjust the sign of y
  if (Sign<0) {
    y[,c(index_lower[j],index_upper[j])] <- t(apply(cbind(
      y[,index_lower[j]]*(-1),y[,index_upper[j]]*(-1)),1,sort))
  }
}
return(y)
}

```

C.2 NFL example

```
nfl <- read.table("nfl_interval.txt",header=T)
rownames(nfl) <- nfl[,1]
nfl <- nfl[,-1]
mv <- ncol(nfl)/2

index_lower <- seq(from=1,to=ncol(nfl),by=2)
index_upper <- seq(from=2,to=ncol(nfl),by=2)

(nfl_blocks <- list(1:3,4:9,10:12))
```

```
## [[1]]
## [1] 1 2 3
##
## [[2]]
## [1] 4 5 6 7 8 9
##
## [[3]]
## [1] 10 11 12
```

```
pass <- c(0,0,0)
offense <- c(1,0,0)
scoring <- c(0,1,0)
nfl_path <- rbind(pass,offense,scoring)
colnames(nfl_path) <- rownames(nfl_path)
nfl_path
```

```
##      pass offense scoring
## pass      0      0      0
## offense   1      0      0
## scoring   0      1      0
```

```
fit_new_center <- pls_symb(data=nfl, blocks=nfl_blocks, path=nfl_path,
                           type="center",ambiguity=T, gof=T)
```

```
## Loading required package: turner
```

```
##
```

```
## Attaching package: 'turner'
```

```
## The following object is masked from 'package:base':  
##  
## lengths
```

```
## Loading required package: Matrix
```

```
fit_new_center$unidimensionality
```

```
##           Alpha           Rho  
## pass      0.7992530 0.8491417  
## offense   0.7981325 0.8128779  
## scoring   0.9731991 0.9656233
```

```
fit_new_center$weights
```

```
##           pass      offense      scoring  
## pass1_l/pass1_u      0.1330795 0.00000000 0.00000000  
## pass2_l/pass2_u      0.4423277 0.00000000 0.00000000  
## pass3_l/pass3_u      0.4604160 0.00000000 0.00000000  
## offense1_l/offense1_u 0.0000000 0.24357170 0.00000000  
## offense2_l/offense2_u 0.0000000 0.07862755 0.00000000  
## offense3_l/offense3_u 0.0000000 0.17613760 0.00000000  
## offense4_l/offense4_u 0.0000000 0.20620736 0.00000000  
## offense5_l/offense5_u 0.0000000 0.26058878 0.00000000  
## offense6_l/offense6_u 0.0000000 0.13272919 0.00000000  
## scoring1_l/scoring1_u 0.0000000 0.00000000 0.3450681  
## scoring2_l/scoring2_u 0.0000000 0.00000000 0.3320104  
## scoring3_l/scoring3_u 0.0000000 0.00000000 0.3297052
```

```
fit_new_center$path_coef
```

```
##           pass      offense      scoring  
## pass      0.0000000 0.000000 0  
## offense   0.7656701 0.000000 0  
## scoring   0.0000000 1.091452 0
```

```
fit_new_center$loadings
```

```
##           Load  
## pass1_l/pass1_u      0.3402622  
## pass2_l/pass2_u      1.0444392
```

```

## pass3_l/pass3_u      1.0701922
## offense1_l/offense1_u 1.0202651
## offense2_l/offense2_u 0.4181014
## offense3_l/offense3_u 0.8211034
## offense4_l/offense4_u 0.8928388
## offense5_l/offense5_u 1.1449231
## offense6_l/offense6_u 0.6895754
## scoring1_l/scoring1_u 0.9942013
## scoring2_l/scoring2_u 0.9795900
## scoring3_l/scoring3_u 1.0060467

```

```
fit_new_center$gof
```

```

## $Communality
## [1] 0.6724179 0.4441603 0.9037524
##
## $Communality_bar
## [1] 0.6161227
##
## $R2
## [1] 0.6255205 0.8904310
##
## $Redundancy
## [1] 0.2778314 0.8047292
##
## $GoF
## [1] 0.6833784

```

```
fit_new_cov <- pls_symb(data=nfl, blocks=nfl_blocks,path=nfl_path,
                        type="cov",ambiguity=T,gof=T)
```

```
fit_new_cov$unidimensionality
```

```

##           Alpha      Rho
## pass      0.7992530 0.8491417
## offense  0.7981325 0.8128779
## scoring  0.9731991 0.9656233

```

```
fit_new_cov$weights
```

```

##           pass      offense      scoring
## pass1_l/pass1_u 0.1667636 0.0000000 0.0000000

```

```

## pass2_l/pass2_u      0.4642162 0.0000000 0.0000000
## pass3_l/pass3_u      0.4105325 0.0000000 0.0000000
## offense1_l/offense1_u 0.0000000 0.29602602 0.0000000
## offense2_l/offense2_u 0.0000000 0.05257489 0.0000000
## offense3_l/offense3_u 0.0000000 0.18084473 0.0000000
## offense4_l/offense4_u 0.0000000 0.22683173 0.0000000
## offense5_l/offense5_u 0.0000000 0.21294269 0.0000000
## offense6_l/offense6_u 0.0000000 0.12597230 0.0000000
## scoring1_l/scoring1_u 0.0000000 0.0000000 0.3494996
## scoring2_l/scoring2_u 0.0000000 0.0000000 0.3316097
## scoring3_l/scoring3_u 0.0000000 0.0000000 0.3256889

```

```
fit_new_cov$path_coef
```

```

##          pass  offense scoring
## pass      0.0000000 0.0000000      0
## offense  0.8187483 0.0000000      0
## scoring  0.0000000 0.8182818      0

```

```
fit_new_cov$loadings
```

```

##          Load
## pass1_l/pass1_u      0.4530313
## pass2_l/pass2_u      0.9959992
## pass3_l/pass3_u      0.9050091
## offense1_l/offense1_u 0.7780324
## offense2_l/offense2_u 0.2894339
## offense3_l/offense3_u 0.6819428
## offense4_l/offense4_u 0.8017282
## offense5_l/offense5_u 0.8201955
## offense6_l/offense6_u 0.4417566
## scoring1_l/scoring1_u 0.9284756
## scoring2_l/scoring2_u 0.9235803
## scoring3_l/scoring3_u 0.9981755

```

```
fit_new_cov$gof
```

```

## $Communality
## [1] 0.6720978 0.4441317 0.9038073
##
## $Communality_bar

```

```
## [1] 0.6160421
##
## $R2
## [1] 0.6703488 0.6695850
##
## $Redundancy
## [1] 0.2977231 0.6051758
##
## $GoF
## [1] 0.642439
```

C.3 Simulations

```
simulation_plspm_interval <- function(n,s,aggr,path_coef_value,structural,
                                     Load_value,Load_size,mean_center_xi,
                                     lambda,sigma_error=1, mu_error=0,
                                     type="cov",method="old"){
  ## This function conducts 's' repetitions of the PLS-PM for interval-valued data
  ## with 'n' observations in each repetition and returns summary stats from them.
  ## aggr is the number of obs. to be aggregated to generate each interval data
  ## path_coef_value can be a number or vector, and is the value of the parameters
  ## in the structural model. If all should have a fixed set to a single number or
  ## if it is be randomly chosen out of some numbers give the vector as attribute
  ## Load_value has the same structure as path_coef_value
  ## Load_size should indicate how many MV are associated with each LV, for this
  ## simulation all LV should have the same number of MV associated with them.
  ## mean_center_xi is a vector containing the mean of the center of the indep. LV
  ## lambda is a vector with parameters to generate the range of each indep. LV
  ## sigma_error is the standard deviation of the errors added to the variables
  ## mu_error is the mean of the errors added to the variables, 0 for all studies
  ## 'type' indicates what linear regression method should be used,
  ## does not accept the center and range method.
  ## type="center" is for the center method
  ## type="cov" is for the symbolic covariance method
  ## Can be done using any of the two covariance formulas for interval data
  ## method=old is the method presented in Billard and Diday (2006a)
  ## method=new if the method presented in Billard (2008)
  ## structural is the square matrix indicating the structures in the structural
  ## model or in the path diagram.
  ## for this function to work, independent LV should come first in the structural
  ## model, the other should be in the correct order that is necessary for them
  ## to be estimated.
  ## For eg, if xi4 depends on xi3, xi3 needs to come first than xi4.
  ## The functions for interval data in the dissertation need to be loaded

  ## installs and load the packages into R
  if(!require("MASS")) { install.packages("MASS"); require("MASS") }
  if(!require("turner")) { install.packages("turner"); require("turner") }

  #####
  ## This next part of the code is going to initialize variables and list,
```

```

## as well as make sure they have the needed structure
#####

path <- (structural!=0)*1 ## makes sure the path matrix is 0/1
## ind_lv will store indices of independent LV in the struct. model
## dep_lv will store indices of dependent LV in the struct. model
ind_lv <- dep_lv <- c()
for (j in 1:nrow(path)){
  if (sum(path[j,])==0) {ind_lv <- c(ind_lv,j)}
  if (sum(path[j,])!=0) {dep_lv <- c(dep_lv,j)}
}
## measurement creates the structure for the blocks in pls_symb
measurement <- vector("list",nrow(path))
for (j in 1:nrow(path)){
  measurement[[j]] <- rep(0,times=Load_size)
}

p_ind_lv <- length(ind_lv) ## counts how many ind. LV
p_dep_lv <- length(dep_lv) ## counts how many dep. LV
p_lv <- nrow(path) ## counts the number of LV
## Then we obtain a vector with the # of MV associated with each LV
mv_each <- lengths(measurement)
mv_total <- sum(mv_each) ## counts total number of MV

## This condition makes sure that the dep. LV are in the correct
## order, that will allow the data generation process to work
if (sum(path[dep_lv,dep_lv][upper.tri(path[dep_lv,dep_lv])])+
    sum(diag(path[dep_lv,dep_lv]))!=0)
  stop("Check the structure of 'structural'")

blocks <- measurement ## copies the list structure
mv_index <- 1:mv_total ## sequence from 1 to mv_total
## Then we need to fill the list blocks with the indices
## of the MV that are associated with each LV.
## mv_index will be used to update the indices that were used
for (j in 1:p_lv){
  blocks[[j]] <- mv_index[1:length(blocks[[j]])]
  mv_index <- mv_index[-(1:length(blocks[[j]]))]
}

## Finds the maximum number of associations in the struc. model

```

```

max_association <- max(apply(path,1,sum))
## Finds the maximum number of loadings associated with LV
max_loadings <- max(lengths(blocks))

## Then we generate empty matrices and arrays to store the results
Coef <- rb_coef <- array(NA,dim=c(s,max_association,p_dep_lv))
loadings <- rb_load <- array(NA,dim=c(s,max_loadings,p_lv))
## the number of coluns in each third dimension of the array will
## always be the same, and will be the maximum number needed.
## Extra columns will remain wil NA, this is why I chose NA instead of 0.

rmse <- matrix(NA,ncol=2*p_lv,nrow=s) ## RMSE is interval-valued
r <- matrix(NA,ncol=p_lv,nrow=s)

redundancy <- matrix(NA,ncol=p_dep_lv,nrow=s)
communality <- matrix(NA, ncol=p_lv,nrow=s)
goodness <- c()

## vector containing the indexes of the lower bound of the p variables
index_lower <- seq(from=1,to=(2*max(mv_total,p_lv)),by=2)
## vector containing the indexes of the upper bound of the p variables
index_upper <- seq(from=2,to=(2*max(mv_total,p_lv)),by=2)

for (sim in 1:s){
  ## In this loop, at each repetition data is generated,
  ## PLS-PM is conducted and results stored.

  ## Randomly generate the parameters from the values provided and
  ## assign it to the structural and measurement matrices
  structural[structural!=0] <- sample(path_coef_value,size=sum(path),
                                   replace=T)

  for (j in 1:p_lv){
    measurement[[j]] <- sample(Load_value,size=lengths(measurement)[j],
                              replace=T)
  }

  #####
  ## Then we can generate the data
  #####

  xi <- matrix(0,ncol=(2*p_lv),nrow=n) ## will receive predicted LV

```

```

x <- matrix(0,ncol=(2*mv_total),nrow=n) ## will receive temp x's
SD <- rep(NA,times=p_lv) ## will receive the sd of data

## Then we can generate the interval independent LV
## Step 1 of Algorithm 1
xi_center <- mvrnorm(n,mu=mean_center_xi,Sigma=diag(p_ind_lv))
for (j in 1:p_ind_lv){
  xi_range <- matrix(rexp(n,rate=lambda[j]))
  xi[,index_lower[ind_lv[j]]] <- xi_center[,j]-0.5*xi_range
  xi[,index_upper[ind_lv[j]]] <- xi_center[,j]+0.5*xi_range
  SD[ind_lv[j]] <- sqrt(var_symb(
    xi[,index_lower[ind_lv[j]]:index_upper[ind_lv[j]]]))
}

xi <- stand_symb(xi)

## Then, for each observation we can generate the dependent LV
## Step 2 of Algorithm 1
for (u in 1:n){
  xi_temp <- matrix(0,ncol=p_lv,nrow=aggr)
  for (j_ind in ind_lv){
    xi_temp[,j_ind] <- runif(aggr,xi[u,index_lower[j_ind]],
      xi[u,index_upper[j_ind]])
  }
  #for each dependent lv we add the regression effect
  for (j_dep in dep_lv){
    for (j in ((1:p_lv)[(1:p_lv)!=j_dep])){
      xi_temp[,j_dep] <- xi_temp[,j_dep]+structural[j_dep,j]*
        xi_temp[,j]
    }
    ## then we add the error term
    xi_temp[,j_dep] <- xi_temp[,j_dep]+rnorm(aggr,mean=mu_error,
      sd=sigma_error)

    ## obtain Q1 and Q3
    xi[u,index_lower[j_dep]] <- quantile(xi_temp[,j_dep],probs=0.25)
    xi[u,index_upper[j_dep]] <- quantile(xi_temp[,j_dep],probs=0.75)

    ## restarts for next step, in case it uses the variable again
    xi_temp[,j_dep] <- runif(aggr,xi[u,index_lower[j_dep]],
      xi[u,index_upper[j_dep]])
  }
}

```

```

}
## Stores the sd and standardize the new LV
for (j_dep in dep_lv){
  SD[j_dep] <- sqrt(var_symb(xi[,index_lower[j_dep]:
                             index_upper[j_dep]]))
  xi[,index_lower[j_dep]:index_upper[j_dep]] <-
    stand_symb(xi[,index_lower[j_dep]:index_upper[j_dep]])
}

## Step 3 of Algorithm 1
### Generating the manifest variables

for (u in 1:n){
  h <- 0 ## auxiliari index to reach the correct MV
  for (j in 1:p_lv){
    data_unif_x <- runif(aggr,xi[u,index_lower[j]],
                        xi[u,index_upper[j]])
    for (k in 1:length(measurement[[j]])){
      h <- h+1
      x_temp2 <- measurement[[j]][k]*data_unif_x+
                rnorm(aggr,mean=mu_error,sd=sigma_error)
      x[u,index_lower[h]] <- quantile(x_temp2,probs=0.25)
      x[u,index_upper[h]] <- quantile(x_temp2,probs=0.75)
    }
  }
}

## Stores the sd and standardize the MV
SD_load <- sqrt(var_symb(x))
mv <- stand_symb(x)

#####
## Finally we can run the model and obtain the estimates
#####

model_plspm <- pls_symb(data=mv,blocks,path,gof=TRUE,type=type,method=method)

### Then need to organize the outputs, and we need to weight the estimates
### by the size of the reduction of the effect, which is the sd of the data

for (j_dep in 1:p_dep_lv){

```

```

index_path_j_dep <- which(path[dep_lv[j_dep],] != 0)
for (j in 1:length(index_path_j_dep)){
  ## Obtains and corrects the estimates for the struc. model
  Coef[sim,j,j_dep] <- model_plspm$path_coef[dep_lv[j_dep],
                                             index_path_j_dep[j]]*SD[dep_lv[j_dep]]
  ## then we calculate the mean absolute relative bias
  rb_coef[sim,j,j_dep] <- abs(Coef[sim,j,j_dep]-
                              structural[dep_lv[j_dep],index_path_j_dep[j]])/
                              structural[dep_lv[j_dep],index_path_j_dep[j]]
}
redundancy[sim,j_dep] <- model_plspm$gof$Redundancy[[j_dep]]
}

## Storing the results from measurement model, model validation and fit
h <- 0
for(j in 1:p_lv){
  for (k in 1:lengths(measurement)[j]){
    h <- h+1
    loadings[sim,k,j] <- model_plspm$loadings[h,1]*SD_load[h]
    rb_load[sim,k,j] <- abs(loadings[sim,k,j]-measurement[[j]][k])/
                          measurement[[j]][k]
  }
  communality[sim,j] <- model_plspm$gof$Communality[[j]]
}

rmse[sim,] <- sqrt(apply(((xi-model_plspm$scores)^2),2,mean))
for (j in 1:p_lv){
  ## dat1 has the LV that the MV came from
  dat1 <- xi[,index_lower[j]:index_upper[j]]
  ## dat2 has the estimated LV, from model
  dat2 <- as.matrix(model_plspm$scores[,index_lower[j]:index_upper[j]])
  r[sim,j] <- cov_symb(cbind(dat1,dat2),method=method)[2,1]/
              sqrt(var_symb(dat1)*var_symb(dat2))
}

goodness[sim] <- model_plspm$gof$GoF
## end of replications
}
## Then we obtain the average for each estimate, and organize it
mean_coef <- apply(Coef,c(2,3),mean)
sd_coef <- apply(Coef,c(2,3),sd)

```

```

mean_rb_coef <- apply(rb_coef,c(2,3),mean)
sd_rb_coef <- apply(rb_coef,c(2,3),sd)
## Stores the values for 95% coming from simulations
lb_rb_coef <- apply(rb_coef,c(2,3),quantile,probs=0.025,na.rm=T)
ub_rb_coef <- apply(rb_coef,c(2,3),quantile,probs=0.975,na.rm=T)

coef_output <- matrix(NA,nrow=sum(path),ncol=7)
colnames(coef_output) <- c("beta","mean_beta","sd_beta","mean_rb",
                          "sd_rb","ll_rb","ul_rb")

h <- 0
for (l in 1:p_dep_lv){
  index_path_j_dep <- which(path[dep_lv[l],]!=0)
  for (g in 1:length(index_path_j_dep)){
    h <- h+1
    coef_output[h,1] <- structural[dep_lv[l],index_path_j_dep[g]]
    coef_output[h,2] <- mean_coef[g,1]
    coef_output[h,3] <- sd_coef[g,1]
    coef_output[h,4] <- mean_rb_coef[g,1]
    coef_output[h,5] <- sd_rb_coef[g,1]
    coef_output[h,6] <- lb_rb_coef[g,1]
    coef_output[h,7] <- ub_rb_coef[g,1]
  }
}

mean_load <- apply(loadings,c(2,3),mean)
sd_load <- apply(loadings,c(2,3),sd)
mean_rb_load <- apply(rb_load,c(2,3),mean)
sd_rb_load <- apply(rb_load,c(2,3),sd)
lb_rb_load <- apply(rb_load,c(2,3),quantile,probs=0.025,na.rm=T)
ub_rb_load <- apply(rb_load,c(2,3),quantile,probs=0.975,na.rm=T)

load_output <- matrix(NA,nrow=mv_total,ncol=7)
colnames(load_output) <- c("loading","mean_loading","sd_loading",
                          "mean_rb","sd_rb","ll_rb","ul_rb")

h <- 0
for (j in 1:p_lv){
  for (i in 1:lengths(measurement)[j]){
    h <- h+1
    load_output[h,1] <- measurement[[j]][i]
  }
}

```

```

    load_output[h,2] <- mean_load[i,j]
    load_output[h,3] <- sd_load[i,j]
    load_output[h,4] <- mean_rb_load[i,j]
    load_output[h,5] <- sd_rb_load[i,j]
    load_output[h,6] <- lb_rb_load[i,j]
    load_output[h,7] <- ub_rb_load[i,j]
  }
}

mean_commun <- apply(communality,2,mean)
sd_commun <- apply(communality,2,sd)
mean_redund <- apply(redundancy,2,mean)
sd_redund <- apply(redundancy,2,sd)

com_red_output <- matrix(NA,nrow=p_lv,ncol=4)
colnames(com_red_output) <- c("mean_communality","sd_communality",
                             "mean_redundancy","sd_redundancy")
if (!is.null(rownames(path))) rownames(com_red_output) <- rownames(path)

h <- 0
for (j in 1:p_lv){
  com_red_output[j,1] <- mean_commun[j]
  com_red_output[j,2] <- sd_commun[j]
  if (sum(path[j,])>=1) {
    h <- h+1
    com_red_output[j,3] <- mean_redund[h]
    com_red_output[j,4] <- sd_redund[h]
  }
}

model_fit <- matrix(apply(rmse,2,mean),ncol=2,byrow=T)
colnames(model_fit) <- c("lower","upper")
model_fit <- cbind(model_fit,matrix(apply(r,2,mean)))
colnames(model_fit)[3] <- "r"

return(list(Coefficients=coef_output,Loadings=load_output,
           Fit= model_fit,gof=mean(goodness),
           commu_redund=com_red_output,All_rb=rb_coef))
}

```

C.4 R code - Chapter 5

```
sim_misspecification <- function(s,n,aggr,mean_center,sigma_center,lambda_range,
                                sigma_error,beta,interc=TRUE,method_cov="old"){
  ## This function does the simulation on Algorithm 2
  ## s is the number of replications
  ## n is the sample size - number of interval-valued observations
  ## aggr is the number of points generated to be aggregated for
  ## the response variable (RV)
  ## mean_center is a vector of means for the center of the explanatory variables
  ## sigma_center is a square matrix of covariance for the center of the
  ## explanatory variables (EV)
  ## lambda_range is a vector of lambdas for the exponential dist. for the ranges
  ## sigma_error is the sd of the error term added to the response variable
  ## beta is a vector containing the original parameters, including intercept
  ## interc=TRUE indicates that the estimation is with intercept,
  ## if interc=FALSE, estimation without intercept
  ## type can be 1, 2 or 3, for the 'center', 'center and range' and
  ## 'symbolic covariance'regression methods, respectively
  ## Can be done using any of the two covariance formulas for interval data
  ## method=old is the method presented in Billard and Diday (2006a)
  ## method=new if the method presented in Billard (2008)
  #####

  ## installs and load the packages into R
  if(!require("MASS")) { install.packages("MASS"); require("MASS") }
  if(!require("triangle")) { install.packages("triangle"); require("triangle") }
  if(!require("mc2d")) { install.packages("mc2d"); require("mc2d") }
  if(!require("ggplot2")) { install.packages("ggplot2"); require("ggplot2") }
  if(!require("ggpubr")) { install.packages("ggpubr"); require("ggpubr") }
  if(!require("cowplot")) { install.packages("cowplot"); require("cowplot") }

  p <- length(beta) ## finds the number of EV
  if (interc==T) p <- p-1

  ## Checks if the attributes were defined correctly
  if (length(mean_center)!=p) stop("Incorrect dimension")
  if (length(sigma_center[1,])!=p) stop("Incorrect dimension")
  if (length(lambda_range)!=p) stop("Incorrect dimension")
  if (!(method_cov=="old")||(method_cov=="new"))
    stop("method_cov should be 'old' or 'new'")
}
```

```

## vector containing the indexes of the lower bound of the variables
index_lower <- seq(from=1,to=(2*(p+1)),by=2)
## vector containing the indexes of the upper bound of the variables
index_upper <- seq(from=2,to=(2*(p+1)),by=2)

## Then we initialize the matrices that will receive the results
## we do 4* because the center range needs space of 2
sim_coef_spec <- sim_coef_miss <- matrix(NA,ncol=4*length(beta),nrow=s)
sim_rb_spec <- sim_rb_miss <- matrix(NA,ncol=3*length(beta),nrow=s)
# column of min and max times 3 methods
sim_Sum_resid_spec <- sim_Sum_resid_miss <- sim_rmse_spec <-
  sim_rmse_miss <- matrix(NA,ncol=3*2,nrow=s)

Range_spec <- Range_miss <- matrix(NA, ncol=(p+1),nrow=s)

## We can start doing the simulations
for (sim in 1:s){
  ## We start by initializing the matrices needed in each sim
  data_spec <- data_miss <- matrix(0,ncol=2*(p+1),nrow=n)
  x <- matrix(0,ncol=2*p,nrow=n)
  y_spec <- y_miss <- matrix(0,ncol=2,nrow=n)

  ## Then we can generate the explanatory variables
  ## See Algorithm 2, first three lines
  x_center <- mvrnorm(n=n, mu=mean_center,Sigma=sigma_center)
  x_range <- matrix(0,ncol=p,nrow=n)
  for (j in 1:p)
    x_range[,j] <- matrix(rexp(n,rate=lambda_range[j]))
  x[,index_lower[1:p]] <- x_center-0.5*x_range
  x[,index_upper[1:p]] <- x_center+0.5*x_range

  ## Then we need to generate the response variables for
  ## each observation.
  for (u in 1:n){
    ## Start by generating the model correctly identified
    specification <- F
    while (specification==F){
      ## This will be repeated until the response variable generated
      ## pass the uniformity test, in which case specification=T
      y_temp_spec <- rep(beta[1],times=aggr)
    }
  }
}

```

```

x_temp_spec <- matrix(NA, ncol=p,nrow=aggr)

for (j in 1:p){
  x_temp_spec[,j] <- runif(aggr,x[u,index_lower[j]],x[u,index_upper[j]])
  y_temp_spec <- y_temp_spec+beta[j+1]*x_temp_spec[,j]
}
error_spec <- rnorm(aggr,mean=0, sd=sigma_error)
y_temp_spec <- y_temp_spec+error_spec
data_temp_spec <- cbind.data.frame(y_temp_spec,
                                  x_temp_spec)[order(y_temp_spec),]
data_temp_spec <- data_temp_spec[floor(aggr*0.25):floor(aggr*0.75),]
## after we generate the aggr observations of y for observation i, check
## if the assumption of uniformity is met. If not do it again.
## Here I chose to use a non-conservative p-value of 0.1
if (ks.test(data_temp_spec[,1],"punif",min(data_temp_spec[,1]),
            max(data_temp_spec[,1]))$p.value>=0.1) {
  specification <- T
}
}
## Then we generate data with misspecification
misspecification <- F
while (misspecification==F){
  ## This will be repeated until the response variable generated
  ## fails the uniformity test, in which case misspecification=T
  y_temp_miss <- rep(beta[1],times=aggr)
  x_temp_miss <- matrix(NA, ncol=p,nrow=aggr)
  for (j in 1:p){
    x_temp_miss[,j] <- rpert(aggr,min=x[u,index_lower[j]],
                            max=x[u,index_upper[j]],
                            mode=(x[u,index_lower[j]]+0.1))
    y_temp_miss <- y_temp_miss+beta[j+1]*x_temp_miss[,j]
  }
  error_miss <- rnorm(aggr,mean=0, sd=sigma_error)
  y_temp_miss <- y_temp_miss+error_miss
  data_temp_miss <- cbind.data.frame(y_temp_miss,
                                    x_temp_miss)[order(y_temp_miss),]
  data_temp_miss <- data_temp_miss[floor(aggr*0.25):floor(aggr*0.75),]
  ## after we generate the aggr observations of y for observation i, check
  ## if the assumption of uniformity is met. Do this until it is not met.
  ## Here I chose to use a non-conservative p-value of 0.1
  if (ks.test(data_temp_miss[,1],"punif",min(data_temp_miss[,1]),
              max(data_temp_miss[,1]))$p.value>=0.1) {
    misspecification <- T
  }
}

```

```

        max(data_temp_miss[,1]))$p.value<0.1){
misspecification <- T
    }
}
## Then we get the lower and upper bounds for the response
## variables in both studies. Since we already reduced the data
## to Q1 and Q3 earlier, now we can just get their min and max.
data_spec[u,1] <- min(data_temp_spec[,1])
data_spec[u,2] <- max(data_temp_spec[,1])
data_miss[u,1] <- min(data_temp_miss[,1])
data_miss[u,2] <- max(data_temp_miss[,1])
}
## Then we join the response and explanatory variables
data_spec[,-c(1,2)] <- x
data_miss[,-c(1,2)] <- x

## Calculate the range of the mean range for the results of this sim
Range_miss[sim,] <- apply(data_miss[,index_upper[1:(p+1)]]-
                        data_miss[,index_lower[1:(p+1)]] ,2,mean)
Range_spec[sim,] <- apply(data_spec[,index_upper[1:(p+1)]]-
                        data_spec[,index_lower[1:(p+1)]] ,2,mean)

## Then, calculate the regression using the three methods, for both studies
fit_miss_1 <- lm_symb(X=data_miss[,-c(1,2)],Y=data_miss[,1:2],
                    intercept=interc,type=1,method=method_cov)
fit_spec_1 <- lm_symb(X=data_spec[,-c(1,2)],Y=data_spec[,1:2],
                    intercept=interc,type=1,method=method_cov)

fit_miss_2 <- lm_symb(X=data_miss[,-c(1,2)],Y=data_miss[,1:2],
                    intercept=interc,type=2,method=method_cov)
fit_spec_2 <- lm_symb(X=data_spec[,-c(1,2)],Y=data_spec[,1:2],
                    intercept=interc,type=2,method=method_cov)

fit_miss_3 <- lm_symb(X=data_miss[,-c(1,2)],Y=data_miss[,1:2],
                    intercept=interc,type=3,method=method_cov)
fit_spec_3 <- lm_symb(X=data_spec[,-c(1,2)],Y=data_spec[,1:2],
                    intercept=interc,type=3,method=method_cov)

## Then we store the residuals, rmse, coefficients and relative biases for
## this simulation.
sim_Sum_resid_spec[sim,] <- apply(cbind(fit_spec_1$resid,

```

```

                                fit_spec_2$resid,fit_spec_3$resid),2,sum)
sim_rmse_spec[sim,] <- sqrt(apply(cbind(fit_spec_1$resid,
                                fit_spec_2$resid,
                                fit_spec_3$resid)^2,2,mean))
sim_Sum_resid_miss[sim,] <- apply(cbind(fit_miss_1$resid,
                                fit_miss_2$resid,fit_miss_3$resid),2,sum)
sim_rmse_miss[sim,] <- sqrt(apply(cbind(fit_miss_1$resid,
                                fit_miss_2$resid,fit_miss_3$resid)^2,2,mean))

sim_coef_spec[sim,] <- c(fit_spec_1$coef,fit_spec_2$coef_center,
                        fit_spec_2$coef_range,fit_spec_3$coef)
sim_rb_spec[sim,] <- (c(fit_spec_1$coef,fit_spec_2$coef_center,
                        fit_spec_3$coef)-rep(beta,times=3))/rep(beta,times=3)
sim_coef_miss[sim,] <- c(fit_miss_1$coef,fit_miss_2$coef_center,
                        fit_miss_2$coef_range,fit_miss_3$coef)
sim_rb_miss[sim,] <- (c(fit_miss_1$coef,fit_miss_2$coef_center,
                        fit_miss_3$coef)-rep(beta,times=3))/rep(beta,times=3)

## end of replications
}
## Need to organize the data to be output

coef_output_spec <- matrix(NA,ncol=5,nrow=4*length(beta))
coef_output_miss <- matrix(NA,ncol=5,nrow=4*length(beta))

## True value of the coefficients for specification
coef_output_spec[,1] <- c(rep(beta,times=2),rep(NA,times=length(beta)),beta)
## Calculates the average estimates for specification
coef_output_spec[,2] <- round(apply(sim_coef_spec,2,mean),5)
## Calculates the sd of the estimates for specification
coef_output_spec[,3] <- round(apply(sim_coef_spec,2,sd),5)
## Calculates the mean relative biases for specification
coef_output_spec[,4] <- round(c(apply(sim_rb_spec[,1:(2*length(beta))],2,mean),
                                rep(NA,times=length(beta)),
                                apply(sim_rb_spec[, (2*length(beta)+1):
                                (3*length(beta))],2,mean)),5)
## Calculates the sd of the relative biases for specification
coef_output_spec[,5] <- round(c(apply(sim_rb_spec[,1:(2*length(beta))],2,sd),
                                rep(NA,times=length(beta)),apply(
                                sim_rb_spec[, (2*length(beta)+1):
                                (3*length(beta))],2,sd)),5)
colnames(coef_output_spec) <- c("beta","mean_coef","sd_coef","mean_rb","sd_rb")

```

```

## True value of the coefficients for misspecification
coef_output_miss[,1] <- c(rep(beta,times=2),rep(NA,times=length(beta)),beta)
## Calculates the average estimates for misspecification
coef_output_miss[,2] <- round(apply(sim_coef_miss,2,mean),5)
## Calculates the sd of the estimates for misspecification
coef_output_miss[,3] <- round(apply(sim_coef_miss,2,sd),5)
## Calculates the mean relative biases for misspecification
coef_output_miss[,4] <- round(c(apply(sim_rb_miss[,1:(2*length(beta))],2,mean),
                               rep(NA,times=length(beta)),apply(
                                   sim_rb_miss[, (2*length(beta)+1):
                                                (3*length(beta))],2,mean)),5)
## Calculates the sd of the relative biases for misspecification
coef_output_miss[,5] <- round(c(apply(sim_rb_miss[,1:(2*length(beta))],2,sd),
                               rep(NA,times=length(beta)),apply(
                                   sim_rb_miss[, (2*length(beta)+1):
                                                (3*length(beta))],2,sd)),5)
colnames(coef_output_miss) <- c("beta","mean_coef","sd_coef","mean_rb","sd_rb")

## Then need to organize the output with residuals and rmse for both studies

model_fit_spec <- model_fit_miss <- matrix(0,ncol=2*4,nrow=3)

model_fit_spec[,1:2] <- matrix(apply(sim_Sum_resid_spec,2,mean),ncol=2,byrow=T)
model_fit_spec[,3:4] <- matrix(apply(sim_Sum_resid_spec,2,sd),ncol=2,byrow=T)
model_fit_spec[,5:6] <- matrix(apply(sim_rmse_spec,2,mean),ncol=2,byrow=T)
model_fit_spec[,7:8] <- matrix(apply(sim_rmse_spec,2,sd),ncol=2,byrow=T)
colnames(model_fit_spec) <- c("sum_resid_lower","sum_residupper",
                             "SD_sum_resid_lower","SD_sum_residupper",
                             "RMSE_lower","RMSE_upper","SD_RMSE_lower",
                             "SD_RMSE_upper")
rownames(model_fit_spec) <- c("Center","Center and Range","Symb. Cov.")

model_fit_miss[,1:2] <- matrix(apply(sim_Sum_resid_miss,2,mean),ncol=2,byrow=T)
model_fit_miss[,3:4] <- matrix(apply(sim_Sum_resid_miss,2,sd),ncol=2,byrow=T)
model_fit_miss[,5:6] <- matrix(apply(sim_rmse_miss,2,mean),ncol=2,byrow=T)
model_fit_miss[,7:8] <- matrix(apply(sim_rmse_miss,2,sd),ncol=2,byrow=T)
colnames(model_fit_miss) <- c("sum_resid_lower","sum_residupper",
                             "SD_sum_resid_lower","SD_sum_residupper",
                             "RMSE_lower","RMSE_upper","SD_RMSE_lower",
                             "SD_RMSE_upper")

```

```
rownames(model_fit_miss) <- c("Center", "Center and Range", "Symb. Cov.")

return(list(coefficients_Specification=coef_output_spec,
           fit_Specification=model_fit_spec,
           average_Range_Specification=apply(Range_spec, 2, mean),
           coefficients_Misspecification=coef_output_miss,
           fit_Misspecification=model_fit_miss,
           average_Range_Misspecification=apply(Range_miss, 2, mean)))
}
```