Complex Linear Regression Based on Structural Latent Variable Modeling

Graduate School of Culture and Information Science,
Doshisha University

48161004

Duong Thi Binh An

Supervisor Prof. Hiroshi Yadohisa

Abstract

Models are created in order to simulate and solve real-world problems. Linear models have a similar purpose. The more a model describes the actual condition, the more effective it is as a reference while making decisions. Linear models are widely applied in practice and quantitative research fields because they are easy to understand and interpret. However, problems in practice or research are complicated, requiring these models to be modified; for example, the number of variables is increased compared with the original model, assumptions are added to the models, or constraints are replaced. While limitations on statistical techniques, such as a lack of follow-up models to solve problems, or theories exist, there is no fixed method that can be applied to modified cases, or the estimation method exists but has many restrictions that make it difficult to apply.

Within the scope of this dissertation, we aim to make two key contributions to the multivariate linear model literature with these complex models expressed through structural latent variables:

- In Chapter 3, we propose an estimation method for a linear model containing numerous independent and dependent variables that all have errors. It can also be seen as a modification of a multivariate linear regression, wherein the number of dependent variables is increased and assumptions pertaining to errors are added to the model.
- In Chapter 4, we focus on a more complex model, namely, structural equation modeling (SEM). Essentially, SEM is a set of multivariate linear regressions wherein the dependent variable in one equation can be an independent variable in another equation and vice versa. The estimation method proposed in this chapter is a constraint improvement for the generalized maximum entropy (GME) for SEM.

Before going to key chapters, a general context is also provided in Chapter 1. Finally, we present the conclusion to summarize the major contributions of the proposed methods in Chapter 5.

In Chapter 3, we propose a multivariate multiple orthogonal linear regression (MMOLR). The MMOLR expresses the relationship between two sets of dependent and independent

variables. The MMOLR especially considers the advantages of the errors-in-variables (EIV) model, that is, the errors are included in all independent and dependent variables. Consequently, the assumptions of the model are easy to satisfy in practice. Our contribution is deriving an estimation method. It is in this context of total least squares that we reveal the relationship between the MMOLR and the canonical regression model.

Next, in Chapter 4, we derive a novel estimation method for SEM. SEM is widely used in many fields such as psychology, behavioral science, and marketing, to measure unobservable concepts and explore complicated relationships. Over the past four decades, especially, the maximum likelihood (ML) has become the predominant estimation method for SEM. However, this method relies upon sample size to ensure a normal distribution for data and non-negative degree of freedom: The more complicated the model is, the larger the sample size that is required. This is particularly evident in the present world, wherein researchers must consider increasingly multi-parameter problems. This increase in concepts gives rise to an upsurge in relationship paths. At the same time, research subject matters and/or populations with sensitive issues (e.g., managers of enterprises or patients with social diseases) present challenging barriers to data collection. Therefore, an estimation method that does not require assumptions on data distribution but works with minimum sample sizes is necessary. We contribute to this by proposing a new estimation method, the K-means GME for SEM. The K-means GME-SEM can reduce the volume of calculation and redundant constraints through the K-means centroid as a representative for data in consistency constraints. The simulation results also confirm that the new method improves the appropriate goodness of fit. Although there are many limitations to the proposed method, the results of this study are a major step toward a new approach for an information-theoretic-based SEM.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This chapter presents a general picture of multivariate data analysis as an overview of the research area, its implications, and its contributions.

Multivariate analysis refers to a set of models, or statistical techniques, for use in numerous fields (McArdle and Anderson, 2001). Some of the more common models used to analyze two sets of variables include canonical regression (Bartlett, 1951; Tatsuoka and Lohnes, 1988; Lutz and Eckert, 1994), two-block partial least squares (Wold, 1975), and multivariate linear regression. The choice of method depends on the type of data and the information needs of the researchers (Elo and Kyngäs, 2008). These complex linear models are expressed through structural latent variables, which, in turn, refer to the linear relationship between the measurement variables and their latent construct. These structural latent variables are also known as unobservable variables because they cannot be measured or inspected directly; they can only be indirectly measured through observable variables; for example, the physical condition is an abstract concept that can be evaluated through height, weight, and blood, among others (see Figure 1.1).

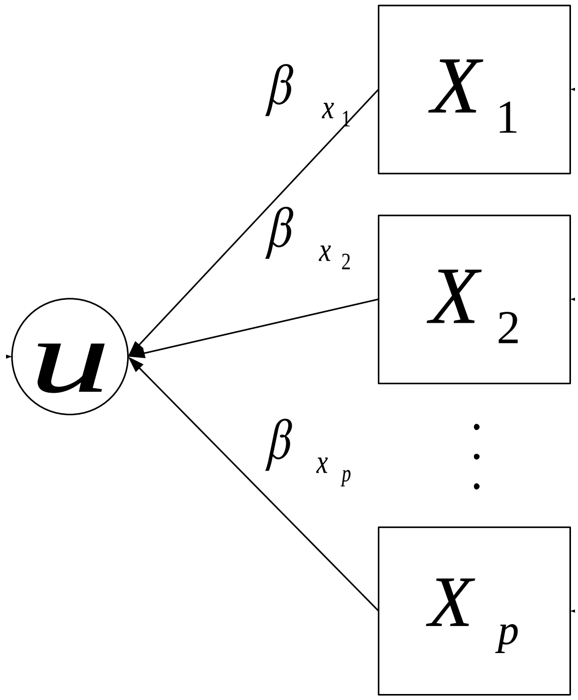


Figure 1.1: Example of a structural latent variable

Because of their wide applicability and appeal to researchers from diverse fields, multivariate processing methods have been studied in theory as well as from the application perspective. These methods are classified according to the criteria used; the criteria include the following:

1. Characteristics of the variable: For example, are there any variables that be considered dependent variables? If yes, how many dependent variables are there, and what types are they? Accordingly, the common methods in multivariate processing are divided into two main groups.

    1.a. "Dependent" methods, wherein the relationship of two sets of variables is considered. Common methods that can be included in this group are multiple regression, discriminant analysis, canonical correlation, and multivariate analysis of variance.

    1.b. "Interdependent" methods, wherein all variables have the same role and there is no distinction between "independent" or "dependent" variables. These methods often investigate the latent structure between variables or elements. Common methods in this group are principal component analysis, factor analysis, and cluster analysis.

2. Purpose of the processing: There are four main groups of methods based on this criterion.

2.a. Multivariate processing methods that investigate relationships between two groups of variables using correlation and regression methods.

2.b. Those that evaluate the difference among groups of elements [1] using analysis of variance methods.

2.c. Those that group elements using clustering methods, logistic regression, and discriminant analysis.

2.d. Those that find the underlying structure of data using principal component analysis and factor analysis.

Following these two criteria, the scope of this thesis contributes to multivariate processing methods that belong to group 1.a. or 2.a. in terms of the estimation method.

Generally, each analysis method requires certain conditions for the data. Some common conditions may include

- Absence of outliers: As points for which the values are very different from the rest of the data, outliers affect the normalization of the data and estimation results. However, outliers also contain certain information; hence, the decision to eliminate these data points or retain them is a matter of debate.

- Normality: The value of each variable must have a normal distribution.

- Standardization: Data normalization is the process of removing the influence of the unit of measure, that is, placing the variables on the same scale. After normalization, the variables have a variance of 1 (standard deviation) and a mean of 0. Data normalization methods include:

  + Subtracting the mean and dividing by the standard deviation;

  + Subtracting mean;

  + Dividing by the standard deviation;

  + Subtracting the first value, and then dividing by the second; and

  + Making range from start to end.

- Homoskedasticity: Homoskedasticity is a statistical phenomenon wherein the errors, that is, residuals, do not follow a special rule. This means the results are unbiased estimates.

[1]Elements can be understood to be data points or observations

- Absence of multicollinearity : Multicollinearity refers to a phenomenon wherein there is a significant correlation between the independent variables in the regression model. It can cause problems when fitting the model and interpreting the results.

Different methods of analyses require different conditions to be met. However, in reality, some conditions are difficult to satisfy. Therefore, it is necessary to develop estimation methods that can be applied under unusual or diverse conditions.

This thesis consists of two estimation methods to deal with complex structures. The first is a novel estimation method combining orthogonal and canonical regressions called multivariate multiple orthogonal linear regression (MMOLR)(Duong et al., 2018).By combining the aforementioned estimation methods, the latent structure is estimated when the dependent variables include error terms. MMOLR is also applicable when the data feature multicollinearity (see Chapter 3 for more details). The second model is a novel estimation method combining the generalized maximum entropy (GME) estimation of a structural equation modeling and the K-means method(Duong et al., 2021). GME has advantageous properties, such as robustness. However, GME is an unstable solution when the sample size is small because the number of constraints on parameters is large. To resolve this issue, the number of constraints on the parameter are reduced by K-means. Moreover, the K-means GME can be applied without considering normality (see Chapter 3). These estimation methods are derived for latent structure modeling because complex structured data can be decomposed into a simple structure between the observed and latent variables and among the latent variables. Figure 1.2 shows the overview of the proposed method.

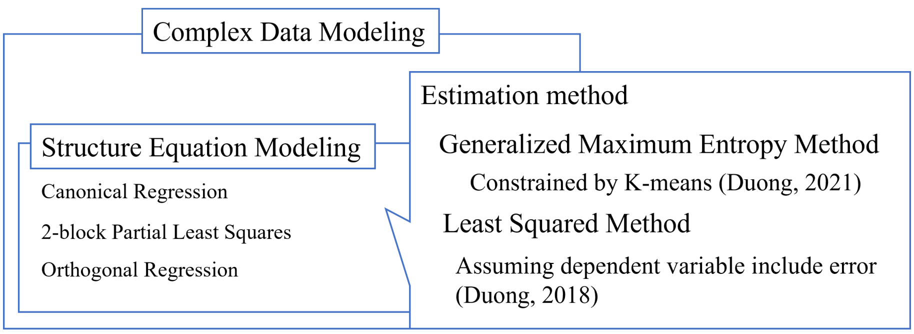


Figure 1.2: Focus of the proposed method

This complex relationship is thus unraveled from less to more complexity along with the content of this thesis. First, the simple model with two latent variables is considered in Chapter 3. Then, the more complicated model with multi-latent constructs is discussed Chapter 4; this model employs structural equation modeling. Subsequent chapters will describe the other advantages of these two new estimation models.

Chapter 2

Notation

In this section, we introduce and define some notations to aid in readability. We denote a matrix, a vector, and scalar by a capital bold italic letter, a bold italic letter, and an italic letter, respectively. For example, $\boldsymbol{X}_{(n \times p)}$ is $n \times p$ matrix. $\boldsymbol{X} \in \mathbb{R}^{n \times p}$ denotes that $\boldsymbol{X}$ is an $n \times p$ matrix whose elements are real values. $\boldsymbol{x} \in \mathbb{R}^{p}$ denotes that $\boldsymbol{x}$ is a p-dimensional vector whose elements are real values. In this thesis, no distinction is made between random variables and data by notation, unless to avoid confusion.

Now, we will denote parameters and unobserved variables by Greek letters. The capitalization rule is applied to the Greek letter, that is, $\boldsymbol{\beta}$ denotes a parameter vector.

The table 2.1 lists the specific notation used in this thesis.

Table 2.1: Notations used in the study

Notation	Explanation
N	Sample size
p,p_x,p_y	Number of variables. Suffix shows data or random variables corresponding to the number of variables.
r,r_x,r_y	Number of dimensions of latent variables. Suffix shows data or random variables corresponding to the number of dimensions of latent variables.
K, L	Number of support points in a generalized maximum entropy estimation.
q	Number of clusters in k-means clustering
ε,ε_x,ε_y	Error term. Suffix shows data or random variables corresponding to the error term.
$\boldsymbol{\beta}, \boldsymbol{\beta}_x, \boldsymbol{\beta}_y$	Coefficient vector. Suffix shows data or random variables corresponding to the coefficient vector.
$\boldsymbol{\Sigma} = (\sigma_{ij})$, $\boldsymbol{\Sigma}_x$, $\boldsymbol{\Sigma}_y$	Covariance matrix. Suffix shows data or random variables corresponding to the coefficient vector.
$\|\boldsymbol{x}\|$	Euclidean norm.
$E(\boldsymbol{x})$	Expectation value of $\boldsymbol{x}$.
$\mathbf{Cov}(\boldsymbol{x}, \boldsymbol{x})$	Covariance matrix of $\boldsymbol{x}$.
$\mathbf{Cov}(\boldsymbol{x}, \boldsymbol{y})$	Covariance matrix between $\boldsymbol{x}$ and $\boldsymbol{y}$.
$\mathrm{corr}(\boldsymbol{x}, \boldsymbol{y})$	Correlation matrix between $\boldsymbol{x}$ and $\boldsymbol{y}$.
$\mathrm{diag}(\boldsymbol{x})$	Diagonal matrix whose ith diagonal elements is the ith element of $\boldsymbol{x}$.
$\boldsymbol{X}'$	Transpose matrix of $\boldsymbol{X}$.

Chapter 3

Multivariate Multiple Orthogonal Linear Regression

3.1 Overview

We now present an overview of the background and context of our research. We discuss the ordinary least squares (OLS) and total least squares (TLS) methods as a pair of contrasting concepts.

3.1.1 Model formula

Linear Regression

Considering the model formula to be a linear regression, the corresponding linear regression model is defined as follows:

$$y = \boldsymbol{x}'\boldsymbol{\beta} + \varepsilon_y, \tag{3.1}$$

where

$y \in \mathbb{R}$ is the dependent variable (DV),

$\boldsymbol{x} \in \mathbb{R}^p_x$ are the independent variables (IVs),

$\boldsymbol{\beta} \in \mathbb{R}^p_x$ is the coefficients' vector, and

$\varepsilon_y \in \mathbb{R}$ is the error term.

This model formula, a popular one in research, is rewritten in the context of the total least squares method:

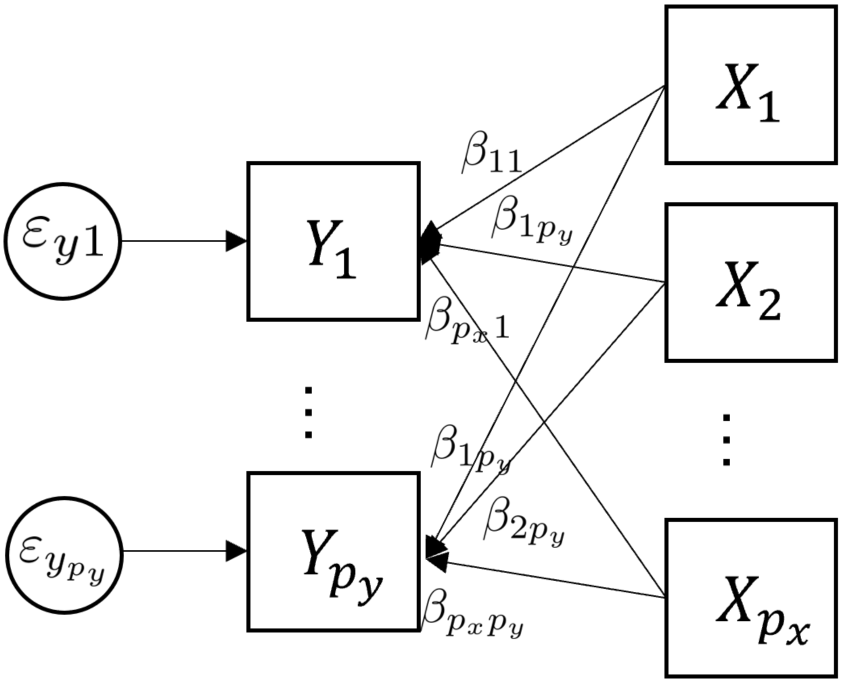


Figure 3.1: Path diagram of multivariate linear regression

$$y = (\boldsymbol{x} - \boldsymbol{\varepsilon}_x)'\boldsymbol{\beta} + \varepsilon_y, \tag{3.2}$$

where $\boldsymbol{\varepsilon}_x$ is the error term for the dependent variables.

Multivariate Linear Regression

The multivariate linear regression (MvLR) is an extended form of the linear regression; then, the general model under the OLS approach is described as follows:

$$\boldsymbol{y} = \boldsymbol{x}\boldsymbol{B} + \boldsymbol{\varepsilon}_y, \tag{3.3}$$

where $\boldsymbol{B} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \dots, \boldsymbol{\beta}_{p_y})$. Given the observation $\boldsymbol{y}_i$ ($i = 1, 2, \dots, N$) and $\boldsymbol{x}_i$, the least squares estimation of the coefficient is $\boldsymbol{B} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}\boldsymbol{Y}$, where $\boldsymbol{Y} = (\boldsymbol{y}_1, \boldsymbol{y}_2, \dots, \boldsymbol{y}_N)'$ and $\boldsymbol{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_N)'$.

However, equation (3.3) is only equivalent to a set of multiple regressions. Hence, the correlation among the DVs is not considered (Velu and Reinsel, 2013), and thus, the relationship among the response variables is also not considered.

We then pose the following question: With the same set of predictors, which dependent variable will be affected the most or least? In other words, how do we compare DVs given that these variables are influenced by another set of variables? To answer these question, we discuss the extended MvLR model herein.

Canonical Regression

Canonical regression, also known as canonical correlation analysis (CCA), is a model that expresses the relationship between two sets of p_x IVs and p_y DVs (Bartlett, 1951; Tatsuoka and Lohnes, 1988; Lutz and Eckert, 1994). The CCA is an extension of the OLS regression (Dattalo, 2013) and has numerous applications, such as testing the omnibus null hypothesis [1], assessing the overall model fit, testing a composite hypothesis [2], and model validation (Dattalo, 2013). The CCA focuses more on correlation than on regression, and thus, the term "canonical correlation" is more commonly used.

Canonical analysis entails finding combining vectors $\boldsymbol{u} \in \mathbb{R}^p$ and $\boldsymbol{v} \in \mathbb{R}^p$ that maximize the correlation between the linear combination of IVs and DVs. Specifically, its purpose is to find $k\ (\leq \min\{p_y, p_x\})$ pairs of linear combinations such that the correlation of each pair is maximized and the successive pair is uncorrelated with the previous pairs (Breiman and Friedman, 1997).

$$Z = \sum_{j=1}^{p_y} u_j Y_j = \boldsymbol{u}'\boldsymbol{y}. \tag{3.4}$$

$$W = \sum_{j=1}^{q_x} v_j X_j = \boldsymbol{v}'\boldsymbol{x}. \tag{3.5}$$

$$r_{ZW} = \max \operatorname{corr}\left(\boldsymbol{u}'\boldsymbol{y}, \boldsymbol{v}'\boldsymbol{x}\right). \tag{3.6}$$

In the aforementioned equations, Z and W are the canonical variates (see Anderson (1962); Lutz and Eckert (1994); Tatsuoka and Lohnes (1988)). The solution for (3.6) is obtained from the eigen-equation, as follows:

$$(E(\boldsymbol{y}\boldsymbol{y}')^{-1} E\left(\boldsymbol{y}\boldsymbol{x}'\right) E(\boldsymbol{x}\boldsymbol{x}')^{-1} E\left(\boldsymbol{x}\boldsymbol{y}'\right) - \mu^2 I)\boldsymbol{u} = 0. \tag{3.7}$$

Two-block partial least squares

Contrasting the CCA, the two-block partial least squares method (2B-PLS) proposed by Wold (1975) is used to analyze the covariance between two sets of variables (Rohlf and Corti, 2000). In this method, the weighted vectors $\boldsymbol{u}\ (\|\boldsymbol{u}\| = 1)$, and $\boldsymbol{v}\ (\|\boldsymbol{v}\| = 1)$ are defined as follows:

[1]The omnibus test checks the significance of multiple parameters in a model simultaneously, that is, when the null hypothesis contains two or more parameters.

[2]A composite hypothesis has two disjointed parameter spaces for null hypothesis H_0 and alternative hypothesis H_1

$$\{\boldsymbol{u}, \boldsymbol{v}\} = \arg\max_{\{\boldsymbol{u},\boldsymbol{v}\}} \operatorname{Cov}\left(\boldsymbol{u}'\boldsymbol{y}, \boldsymbol{v}'\boldsymbol{x}\right). \tag{3.8}$$

The 2B-PLS describes the relationship between the latent variables and latent constructs of the two variables. It can handle collinearity by extracting the latent variables when we apply this method to linear regression (**?**). It is also consistent with a large number of predictor variables.

3.1.2 Ordinary Least Squares and Total Least Squares

Ordinary Least Squares

The OLS is a widely used parametric[3] estimation method for regression equations. An OLS estimator is obtained by minimizing the sum of squares of the distances from the data points to the projection in the direction parallel to the axis representing the dependent variable on the regression line/plane.

In the case of MvLR, the OLS estimator of $\boldsymbol{B}$ is obtained as $\boldsymbol{B} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$, where $\boldsymbol{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n)' \in \mathbb{R}^{n\times p_x}, \boldsymbol{Y} = (\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_n)' \in \mathbb{R}^{n\times p_y}$. The data are represented as a matrix. $\boldsymbol{x}_i (i = 1, 2, \ldots, n)$ is the value of the independent variables of object i and $\boldsymbol{y}_i$ is the value of the dependent variables of object i.

Total Least Squares

The error-in-variables (EIV) model by Durbin (1954) has a well-established relationship with TLS. Explaining TLS through the EIV concept provides a comprehensive understanding of TLS and its implications.

In their study, Schaffrin and Wieser (2008) described the EIV model as follows:

$$y = (\boldsymbol{x} - \boldsymbol{\varepsilon}_x)\,\boldsymbol{\beta} + \varepsilon_y, \tag{3.9}$$

where $\boldsymbol{\varepsilon}_x$ is error of the IVs and the elements are assumed to be independent zero-means.

Further, the characteristics of the errors are described as follows:

$$\begin{bmatrix} \varepsilon_y \\ \boldsymbol{\varepsilon}_x \end{bmatrix} \sim \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix} \begin{bmatrix} \sigma_y^2 & 0 \\ 0 & \boldsymbol{\Sigma}_x \end{bmatrix} \right), \tag{3.10}$$

where σ_y^2 is the variance of ε_y and $\boldsymbol{\Sigma}_y$ is the covariance matrix of ε_x.

Under the assumption of homoskedasticity [4], $\boldsymbol{\Sigma}_x = \sigma_y I$. This model represents the

[3]Parametric statistical estimation is based on assumptions about certain distributions of data.

[4]Homoskedasticity, that is, homogeneity of variance, is a statistical phenomenon wherein the errors, that is, residuals, do not follow a special rule. This makes the results unbiased estimates. The opposite of

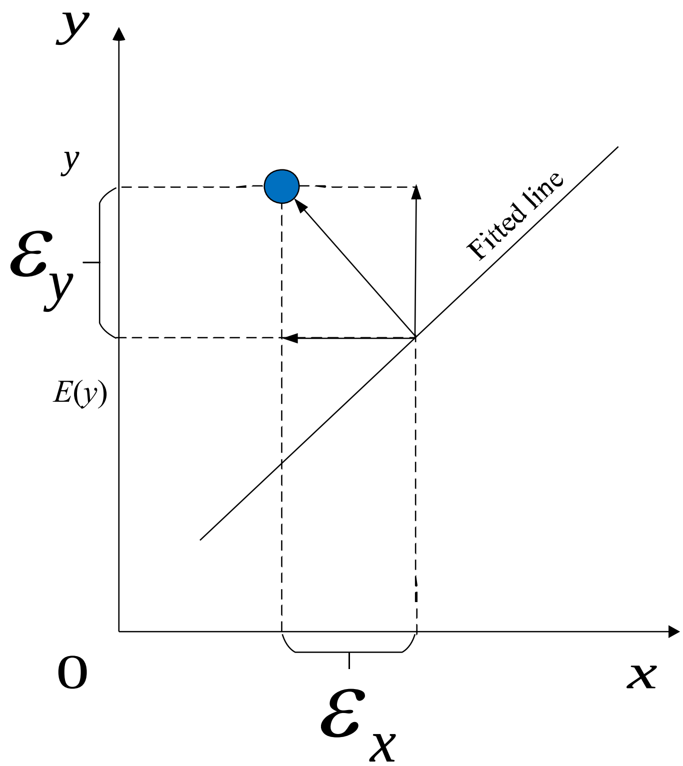


Figure 3.2: Linear regression by EIV model under the TLS approach in a two-dimensional space

TLS problem. Thus, TLS and EIV are aligned and inseparable. TLS, first proposed by Golub and Van Loan (1979), is among the most important applications to estimate the parameters of the EIV model. Using the TLS instead of OLS in typical applications has been shown to increase accuracy by 10—15% (Van Huffel and Vandewalle, 1991).

In the TLS approach, the error is the distance from the data points to the orthogonal projection on a fitted line/plane (Markovsky and Van Huffel, 2007). ε is synthesized from the errors of all IVs and DVs. For easy visualization, $\vec{\varepsilon} = \vec{\varepsilon_x} + \vec{\varepsilon_y}$ in the two dimensions (see Figure 3.2).

In the following sections, we present an overview of related models in multivariate analysis.

3.1.3 Research Problem

The more popular and less limited interpretation in applications involves MvLR, which is a prediction model wherein the values of responses can be predicted from a set of predictors (?Velu and Reinsel, 1998). The model is also used to estimate the linear association

homoskedasticity is heteroskedasticity.

between predictors and responses (Harrell Jr, 2015). However, each individual DV in this model is regressed separately on the IVs. Therefore, the inter-correlation among the DVs is left unexamined (Everitt et al., 2001).

From the perspective of applications, the variables generally measure related aspects, and thus, the relational view of the DVs in a discrete way would lack merit and reduce the significance of analyzing the variables in sets. To consider the association between IDs and DVs, the interpretation of a large number of coefficients simultaneously is unwieldy (Velu and Reinsel, 2013).

All three models take the OLS approach, which considers only the errors in IVs. According to Berkson (1950) and Durbin (1954), it is difficult to retain this assumption strictly in practice. To extend the interpretation using the TLS approach, we propose the MMOLR, which can describe linear relationships between two sets of variables, such as MvLR, and the relationship among the DVs that are integrated into the model. Because of its canonical coefficients, the proposed model is very similar to canonical regression. We thus reveal the relationship between MMOLR and canonical regression based on the TLS.

3.2 Multivariate Multiple Orthogonal Linear Regression (MMOLR)

3.2.1 Model Formula of MMOLR

In the MMOLR model, the word "multivariate" refers to the existence of multiple DVs as well as a large number of IVs; further, the model applies an orthogonal linear regression.

The following equation describes the MMOLR model (intercept excluded):

$$(\boldsymbol{Y} - \boldsymbol{E}_y)\,\boldsymbol{\beta}_y = (\boldsymbol{X} - \boldsymbol{E}_x)\,\boldsymbol{\beta}_x, \tag{3.11}$$

where

$$\begin{aligned}
&\boldsymbol{Y} = (\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_{p_y}), \quad \boldsymbol{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_{p_x}),\\
&\boldsymbol{E}_y = (\boldsymbol{\varepsilon}_{y_1}, \boldsymbol{\varepsilon}_{y_2}, \ldots \boldsymbol{\varepsilon}_{y_{p_y}}), \quad \boldsymbol{E}_x = (\boldsymbol{\varepsilon}_{x_1}, \boldsymbol{\varepsilon}_{x_2}, \ldots \boldsymbol{\varepsilon}_{x_{p_x}}),\\
&\boldsymbol{\beta}_y = (\beta_{y_1}, \beta_{y_2}, \ldots, \beta_{y_{p_y}})', \quad \boldsymbol{\beta}_x = (\beta_{x_1}, \beta_{x_2}, \ldots, \beta_{x_{p_x}})',\\
&\boldsymbol{x}_i, \boldsymbol{y}_j, \boldsymbol{\varepsilon}_{x_i}, \boldsymbol{\varepsilon}_{y_j} \in \mathbb{R}^n, \quad (i = 1, 2, \ldots, p_x;\ j = 1, 2, \ldots, p_y).
\end{aligned}$$

The components and relationships of equation (3.11) are schematically depicted in Figure 3.3.

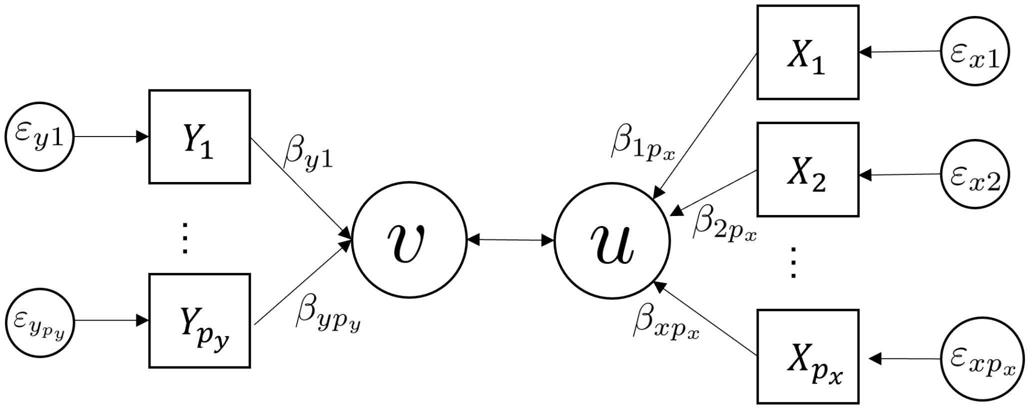


Figure 3.3: Path diagram of multivariate multiple orthogonal linear regression

To estimate the coefficients, (3.11) is transformed by setting data matrix $\boldsymbol{A} = \begin{pmatrix} \boldsymbol{Y}, \boldsymbol{X} \end{pmatrix} = (\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_{p_y+p_x}) = (\boldsymbol{f}_1, \boldsymbol{f}_2, \ldots, \boldsymbol{f}_n)'$ with k column vectors $\boldsymbol{a}_j$ and m row vectors $\boldsymbol{f}_i$; then, coefficient matrix $\boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}_y, -\boldsymbol{\beta}_x \end{pmatrix}^T = (\beta_1, \beta_2, \ldots, \beta_{p_y+p_x})'$ and error matrix $\boldsymbol{E} = \begin{pmatrix} \boldsymbol{E}_y, \boldsymbol{E}_x \end{pmatrix} = (\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \ldots, \boldsymbol{\varepsilon}_{p_y+p_x})$.

We then rewrite equation (3.11) as follows:

$$\boldsymbol{A\beta} = \boldsymbol{E\beta}. \tag{3.12}$$

We also rewrite equation (3.12) as

$$\beta_1 \boldsymbol{a}_1 + \ \beta_2 \boldsymbol{a}_2 \ + \cdots + \ \beta_{p_y+p_x} \boldsymbol{a}_{p_y+p_x} \ = \ \boldsymbol{\varepsilon}, \tag{3.13}$$

where $\boldsymbol{\varepsilon} = \beta_1 \boldsymbol{\varepsilon}_1 + \ \beta_2 \boldsymbol{\varepsilon}_2 \ + \cdots + \ \beta_{p_y+p_x} \boldsymbol{\varepsilon}_{p_y+p_x}$.

Estimation Method of MMOLR

Equation (3.13) describes the general expression of the hyperplane H with normal vector $(\beta_1, \beta_2, \ \ldots, \beta_k)$. Thus, we can assume that the estimation of coefficients is intended to find H, such that the errors from data points to the perpendicular projections corresponding to H are the smallest. H can be written as $\boldsymbol{P f}_i = \boldsymbol{\alpha}_i$, where $\boldsymbol{P}$ is a projection matrix with orthogonal rows, and hence, $\boldsymbol{PP}^T = \boldsymbol{I}$. The square distance of f to H is $\| \boldsymbol{P f}_i \|^2$. Then, the objective function $g(\boldsymbol{P})$ is

$$\begin{array}{c} g(\boldsymbol{P}) = \sum\limits_{i=1}^{m} \| \boldsymbol{P f}_i \|^2 \to \text{minimize} \\ \text{Subject to } \boldsymbol{PP}^T = \boldsymbol{I} \end{array} . \tag{3.14}$$

The solution of (3.14) is a matrix whose column vectors are eigenvectors corresponding to the smallest eigenvalues of the covariance matrix of $\boldsymbol{A}$ (Zamar, 1989; Maronna, 2005).

citeGolubandReinsch1970 argue that singular value decomposition (SVD) can be used to solve the least squares problem. Golub and Van Loan (1980) go further by claiming that the eigenvector corresponding to the smallest eigenvalue (denoted by $\boldsymbol{n}$) of SVD is the solution to the TLS problem. Therefore, orthogonal regression deals with the directions of $\boldsymbol{n}$ (Maronna, 2005). In other words, errors are minimized if we project the data to H with the direction of $\boldsymbol{n}$, that is, $\boldsymbol{n}$ is a normal vector of H. Thus, $\boldsymbol{n} = (n_1, n_2, \ldots, n_k) = (\beta_1, \beta_2 \ldots, \beta_k)$. However, $\boldsymbol{n}$ is notably very sensitive to outliers (Brown, 1982), and hence, it is better to use robust covariance matrixes when finding $\boldsymbol{n}$.

3.2.2 Relationship between MMOLR and the Canonical Regression Model

The difference between the objective function of the proposed model and canonical regression is the optimizing design. Canonical regression maximizes the correlation or covariance between IVs and DVs. However, the proposed model minimizes it. This difference is caused by the following equation:

$$\mathrm{tr}(\boldsymbol{\beta}'(\boldsymbol{A}+\boldsymbol{E})'(\boldsymbol{A}-\boldsymbol{E})\boldsymbol{\beta}) = \|\boldsymbol{A}\boldsymbol{\beta}\|^2 - \|\boldsymbol{E}\boldsymbol{\beta}\|^2. \tag{3.15}$$

From equation (3.15), we can derive the objective function for the maximizing equation (3.15) by two types of design. The first is by maximizing $\|\boldsymbol{A}\boldsymbol{\beta}\|^2$; the second is by minimizing $\|\boldsymbol{E}\boldsymbol{\beta}\|^2$. We extend the objective function of the proposed method to include canonical regression in the context of TLS. The objective function g, extended from the proposed method, is defined as follows:

$$\begin{gathered} g(\boldsymbol{\beta}) = \|\boldsymbol{E}\boldsymbol{\beta}\|^2 \rightarrow \text{minimize} \\ \text{Subject to } \boldsymbol{\beta}'\boldsymbol{\beta} = 1,\ \boldsymbol{\beta}'\boldsymbol{A}'\boldsymbol{A}\boldsymbol{\beta} \geq r. \end{gathered}$$

Using the Lagrange multiplier and assuming that the linear space spanned by $\boldsymbol{\beta}$ and $\boldsymbol{E}$ is orthogonal and the direct sum is $\boldsymbol{\beta}\dot{\oplus}\boldsymbol{E}$, the objective function g is rewritten as follows:

$$\begin{gathered} g(\boldsymbol{\beta}) = \|\boldsymbol{E}\boldsymbol{\beta}\|^2 - \lambda\|\boldsymbol{A}\boldsymbol{\beta}\|^2 \rightarrow \text{minimize} \\ \text{Subject to } \boldsymbol{\beta}'\boldsymbol{\beta} = 1, \end{gathered}$$

where $\lambda \geq 0$ represents the tuning parameters. The role of tuning parameters λ is to adjust the maximizing $\|\boldsymbol{A}\boldsymbol{\beta}\|^2$ or minimizing $\|\boldsymbol{E}\boldsymbol{\beta}\|^2$. When $\lambda = 0$, the objective function g is the same as that of the proposed model. However, g is equivalent to the canonical regression when we set a sufficiently large λ.

From this fact, we can affirm that the objective function g is an extension from the objective function of canonical regression and the MMOLR.

We then adopt the algorithm for alternative least squares to optimize g. Given $\boldsymbol{E}$, the estimator of $\boldsymbol{\beta}$ satisfies the following equation:

$$(\boldsymbol{E}'\boldsymbol{E} - \lambda \boldsymbol{A}'\boldsymbol{A})\boldsymbol{\beta} = \eta\boldsymbol{\beta}, \tag{3.16}$$

where η is the Lagrange multiplier. Equation (3.16) is the same as the eigenvalue problem. Therefore, we obtain the estimator by the eigenvector whose eigenvalue is the minimum. However, given $\boldsymbol{\beta}$, the estimator of $\boldsymbol{E}$ satisfies the following equation:

$$\boldsymbol{E}'\boldsymbol{\beta}\boldsymbol{\beta}' = \boldsymbol{O}. \tag{3.17}$$

Equation (3.17) indicates that the error matrix is the orthogonal from $\boldsymbol{\beta}\boldsymbol{\beta}'$, that is, the space of error is the orthogonal space of $\boldsymbol{\beta}$. We set $\boldsymbol{E} = \boldsymbol{A} - \boldsymbol{A}\boldsymbol{\beta}\boldsymbol{\beta}'$, although it is arbitrary.

As mentioned earlier, we define the algorithm of the extended proposed model as follows:

Step 1 Set $\boldsymbol{E} = \boldsymbol{A}$.

Step 2 Update $\boldsymbol{\beta}$.

Step 3 Update $\boldsymbol{E}$.

Step 4 Repeat steps 2 and 3 until $\boldsymbol{\beta}$ converge.

3.3 Evaluating MMOLR

3.3.1 Numerical Example

In this section, we use the prediction error as a criterion to compare the proposed model with CCA and canonical covariance analysis(CCoVA) which is similar method of 2B-PLS. These two models can be considered special cases of MMOLR. We evaluate the variance and bias of the estimators through the prediction error.

The simulation data are generated in following steps: First, coefficient $\boldsymbol{\beta} = (\boldsymbol{\beta}_y', -\boldsymbol{\beta}_x')$ is generated under the condition $\beta_j \overset{i.i.d.}{\sim} U(-1,\ 1)$ $(j = 1,\ 2,\ \cdots,\ p_y + p_x)$, where $U(-1,\ 1)$ shows uniform distribution from -1 to 1. Then, IVs $\boldsymbol{x}_i \overset{i.i.d.}{\sim} N(\boldsymbol{0}_{p_x},\ \sigma^2\boldsymbol{I}_{p_x})(i = 1,\ 2,\ \cdots,\ n)$, where $\boldsymbol{0}_{p_x}$ is the p_x dimensional vector whose elements is 0, $\boldsymbol{I}_p$ is the p dimensional identical matrix and n is the number of objects. In this example, we set the number of observations at 300. By generating the coefficients randomly, we could evaluate the average prediction error in various scenarios. Dependent variable $\boldsymbol{y}_i$ is generated as $\boldsymbol{y}_i = \boldsymbol{x}_i\boldsymbol{\beta}_x\boldsymbol{\beta}_y'$. Therefore, the relationship between $\boldsymbol{x}_i$ and $\boldsymbol{y}_i$ is the ideal relationship for regression. We set σ as $1,\ 2$, the number of IVs $p_x = 7$, and the number of DVs $p_y = 5$.

The error matrix is highly correlated and the error variables are generated as $\boldsymbol{\varepsilon}_{xi} \overset{i.i.d.}{\sim} N(\mathbf{0},\ \boldsymbol{\Sigma}_x)$ and $\boldsymbol{\varepsilon}_{yi} \overset{i.i.d.}{\sim} N(\mathbf{0},\ \boldsymbol{\Sigma}_y)$. The $(i,\ j)$th element of $\boldsymbol{\Sigma}_x$ and $\boldsymbol{\Sigma}_y$, which is represented as $\sigma^x_{ij},\ \sigma^y_{ij}$, respectively, is as follows:

$$\sigma^x_{ij} = r^{|i-j|},\ \sigma^y_{o\ell} = r^{|o-\ell|}$$
$$(i,\ j = 1,\ 2,\ \cdots,\ p_x;\ o,\ \ell = 1,\ 2,\ \cdots,\ p_y),$$

where r is the parameter for correlations and is set as $0.5,\ 0.7,\ 0.9$. We obtain the data $\boldsymbol{X}^*, \boldsymbol{Y}^*$ by $\boldsymbol{X}^* = \boldsymbol{X} + \boldsymbol{E}_x$ and $\boldsymbol{Y}^* = \boldsymbol{Y} + \boldsymbol{E}_x$.

The criterion of evaluation is the prediction error, which is defined as follows:

$$\|(\boldsymbol{Y}_{pr}, \boldsymbol{X}_{pr})\boldsymbol{\beta} - (\boldsymbol{Y}^*_{pr}, \boldsymbol{X}^*_{pr})\hat{\boldsymbol{\beta}}\|^2,$$

where $\boldsymbol{Y}_{pr}$, $\boldsymbol{X}_{pr}$, $\boldsymbol{Y}^*_{pr}$, and $\boldsymbol{X}^*_{pr}$ are the test data obtained in the same way as $\boldsymbol{Y}$, $\boldsymbol{X}$, $\boldsymbol{Y}^*$, and $\boldsymbol{X}^*$, respectively. $\hat{\boldsymbol{\beta}}$ is the value of the estimator. We set the iteration to 100 times.

Figure 3.4 shows the boxplots of the prediction error. The boxplots depict, from left, the CCA, canonical covariance analysis, and the proposed model. Our model has the best result in all cases. Therefore, it is more effective than the CCA for highly correlated errors.

Table 3.1 shows the mean of prediction errors in all cases. The blanked values show the standard deviation of the prediction error. The highlighted cells indicate the best result under each condition. Table 3.1 shows that our model demonstrates the best results under all cases.

The standard deviation of CCoVA tends to increase when the correlation among the error variables is higher. However, the standard deviation of the proposed method does not increase.

3.3.2 Practical Example

From the perspective of application, we analyze a typical example of biochemical data extracted from the book *Multivariate Reduced-Rank Regression: Theory and Applications* (see Velu and Reinsel (2013) [Appendix 1]). The data include 33 observations with five DVs, comprising pigment creatinine (y_1), concentration of phosphate (y_2), phosphorus

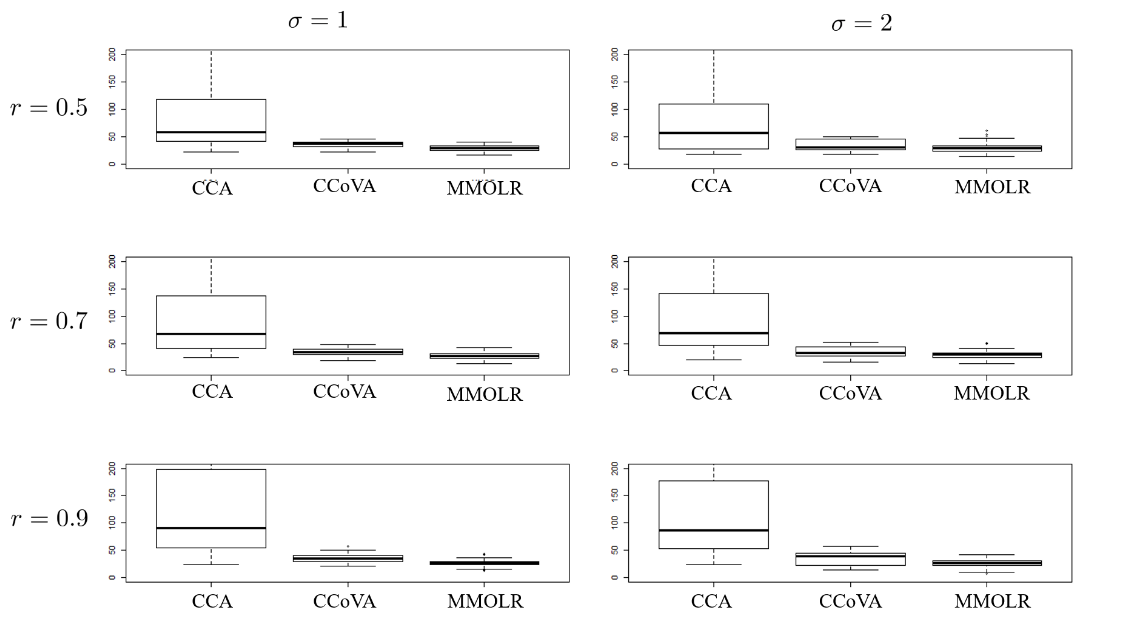


Figure 3.4: Boxplots of square prediction errors

Table 3.1: Mean and standard deviation of prediction error

		CCA	CCoVA	MMOLR
	r=0.5	176.355 (508.636)	35.440 (5.221)	29.053 (5.722)
$\sigma = 1$	r=0.7	215.599 (593.429)	34.566 (6.439)	26.818 (6.054)
	r=0.9	283.832 (743.962)	35.440 (6.927)	26.042 (5.050)
	r=0.5	565.195 (4629.857)	34.969 (9.731)	28.764 (8.861)
$\sigma = 2$	r=0.7	474.120 (2040.812)	35.046 (10.421)	28.728 (7.202)
	r=0.9	881.278 (5244.808)	34.578 (12.038)	25.897 (6.859)

(y_3), creatinine (y_4), and choline (y_5); and three IVs, namely, weight (x_1), volume (x_2), and x_3=100 (*specific gravity*-1). In Chapter 1 of Velu and Reinsel (2013), the MvLR model was applied for these data to measure the involvement of IVs x_1, x_2, and x_3 in DVs y_1, y_2, y_3, y_4, and y_5.

The result of the MvLR model is summarized as follows:

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{pmatrix} = \begin{pmatrix} 15.2809 \\ 1.4159 \\ 2.0187 \\ 1.8717 \\ -0.8902 \end{pmatrix} + \begin{pmatrix} -2.9090 & 1.9631 & 0.2043 \\ 0.6044 & -0.4816 & 0.2667 \\ 0.5768 & -0.4245 & -0.0401 \\ 0.6160 & -0.5781 & 0.3518 \\ 1.3798 & -0.6289 & 2.8908 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \tag{3.18}$$

According to the MvLR model, how the variables of y interact with each other (all coefficients of the variables y are set to 1) cannot be known, although an analysis of the correlation matrix of the DVs shows them to be correlated, and the correlation between variables y_1, y_3, and y_4 is especially high (see Figure 4.7).

In addition, the following two questions arise:

- i. For an individual or the same individuals, that is, when the values of x_1, x_2, and x_3 are determined, what is the relationship among the variables of y_1, y_2, y_3, y_4, and y_5?
- ii. Which are the most affected or weakest DVs?

The MvLR model itself cannot compare DVs, although these variables are affected by the same set of IVs. Thus, it does not make sense to examine the relationships of variables in a set. We apply the MMOLR model to answer these questions.

To compare variables, the intercept is not considered, and the data are centered. In this example, we use a package that is robust to the "covRob" command in R to compute the covariance matrix of the data. The default of covRob allows auto-selection among the Donoho–Stahel projection-based estimators, the fast minimum covariance determinant algorithm of Rousseeuw and Van Driessen, and the orthogonalized quadrant correlation pairwise estimator for good estimates in a reasonable amount of time (Wang et al., 2017). From the covariance matrix, eight eigenvalues and eigenvectors are computed. In addition, the smallest eigenvalue is 0.009 given eigenvector $(-0.004, -0.682, 0.634, -0.062, 0.005, 0.116, -0.008, 0.329)$. Thus, the equation expressing the relationship of the variables is as follows:

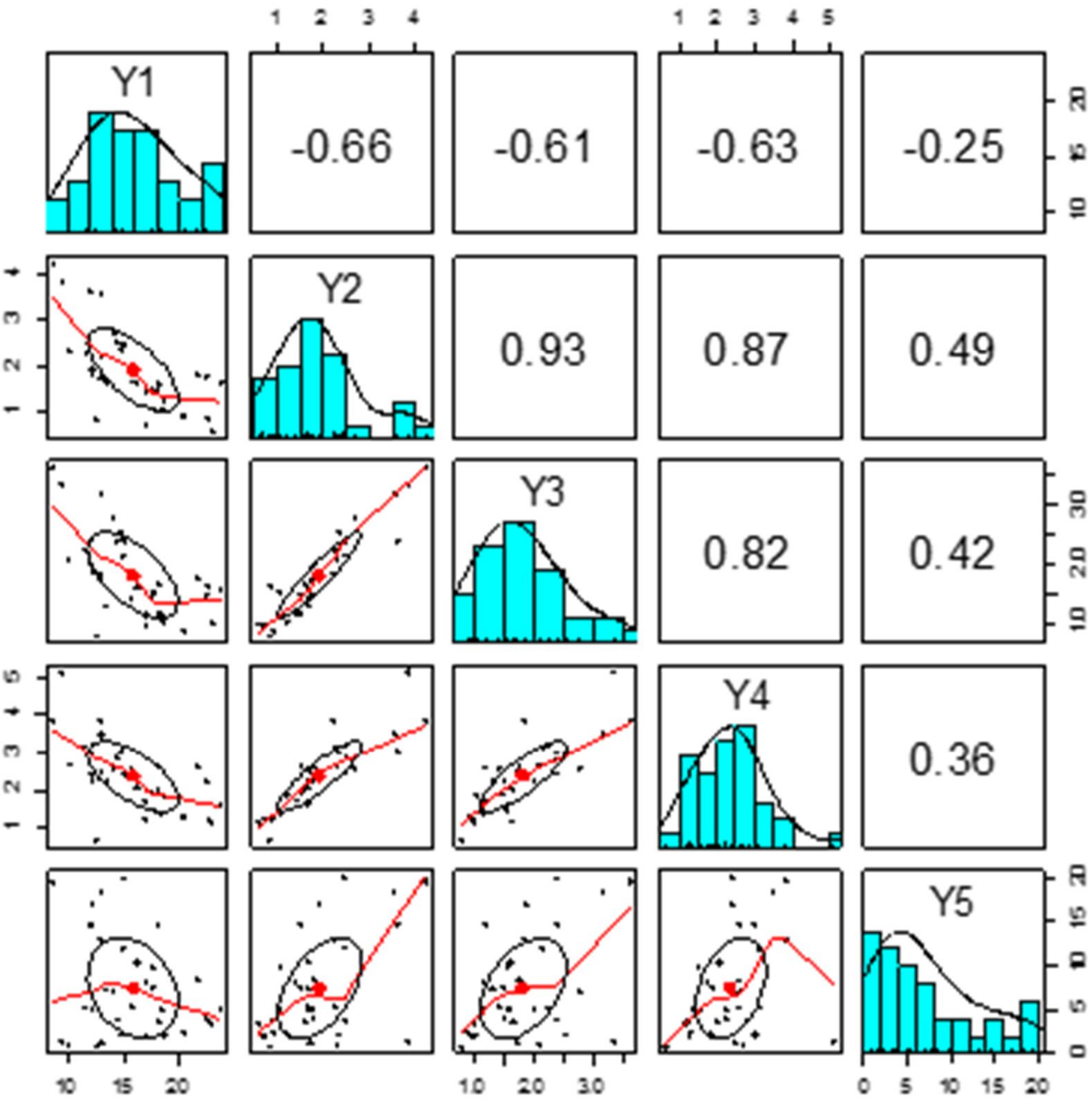


Figure 3.5: Correlation of responses in biochemical data

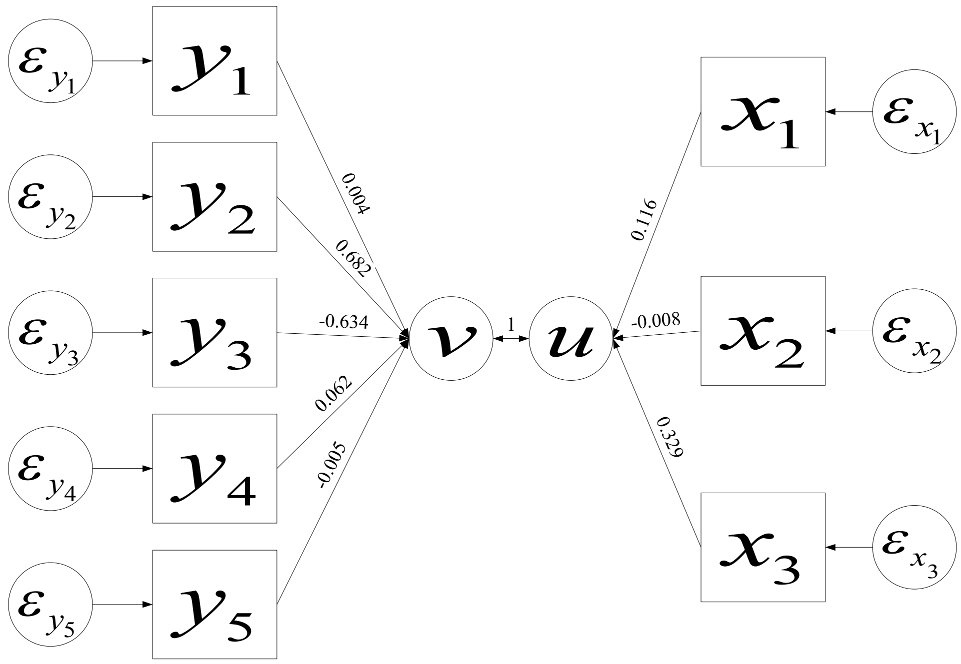
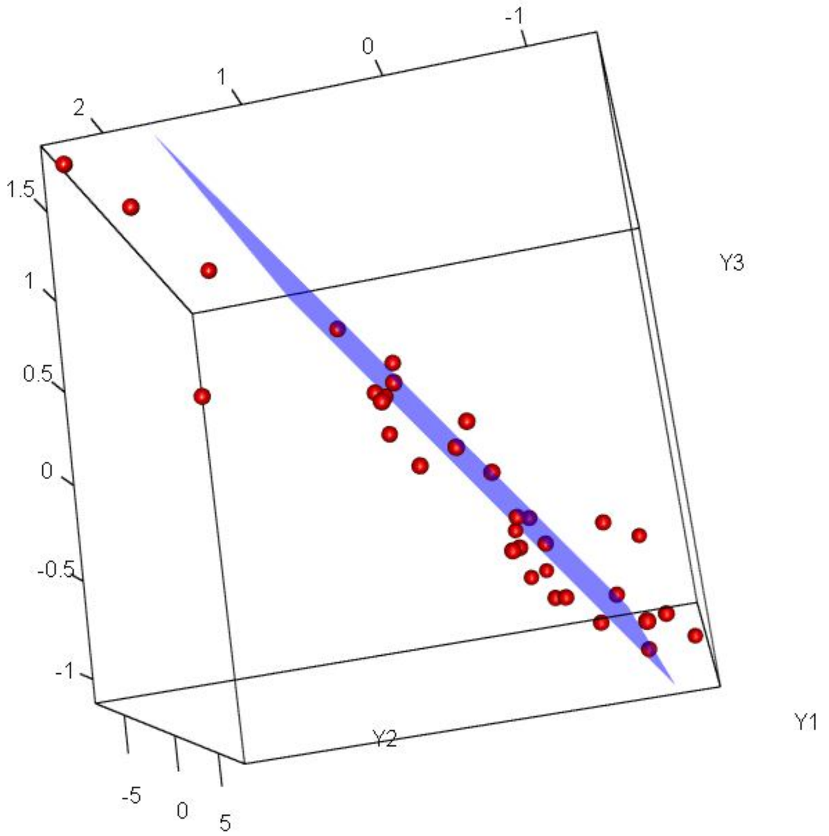


Figure 3.6: Path diagram of the MMOLR model

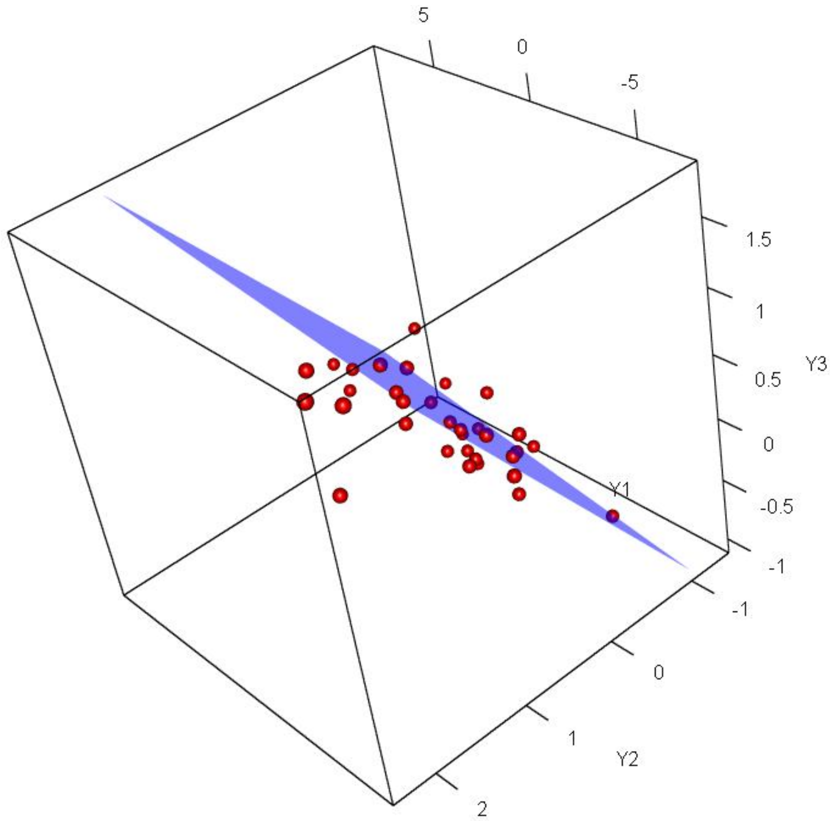
$$0.004y_1 + 0.682y_2 - 0.634y_3 + 0.062y_4 - 0.005y_5 = 0.116x_1 - 0.008x_2 + 0.329x_3. \quad (3.19)$$

The general result is depicted in Figure (3.6) and the fitted plane in view of the three-dimensional space in Figure (3.7).

With the same effect of the variables x_1, x_2, and x_3, the DV y_2 is most sensitive to the changes to the set of IVs, whereas y_1 is the least sensitive variable.

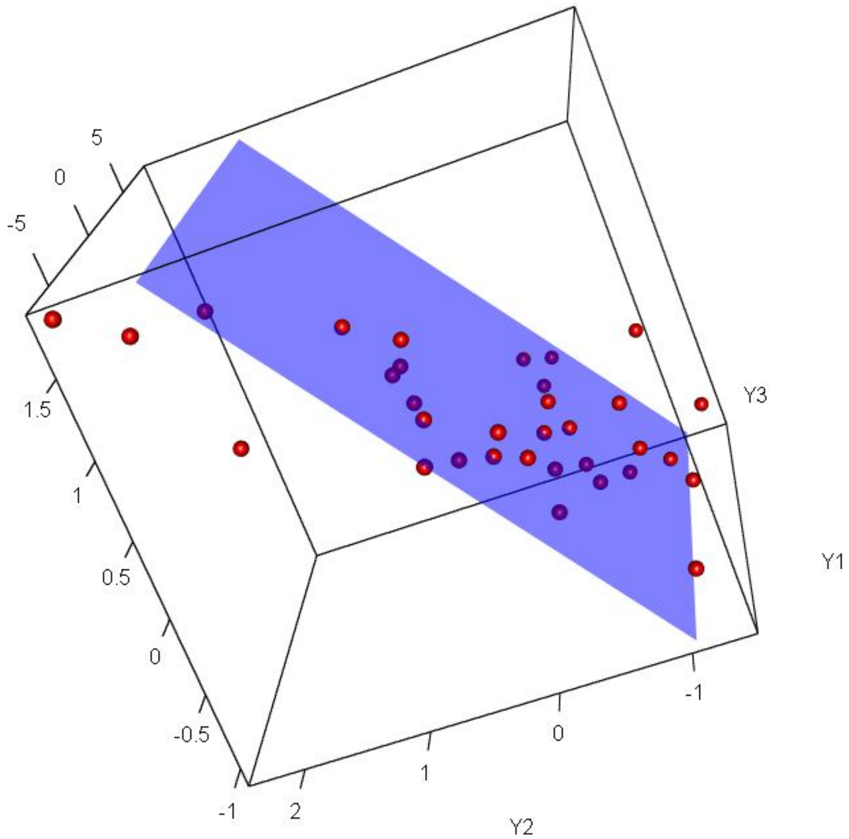


(a) View 1

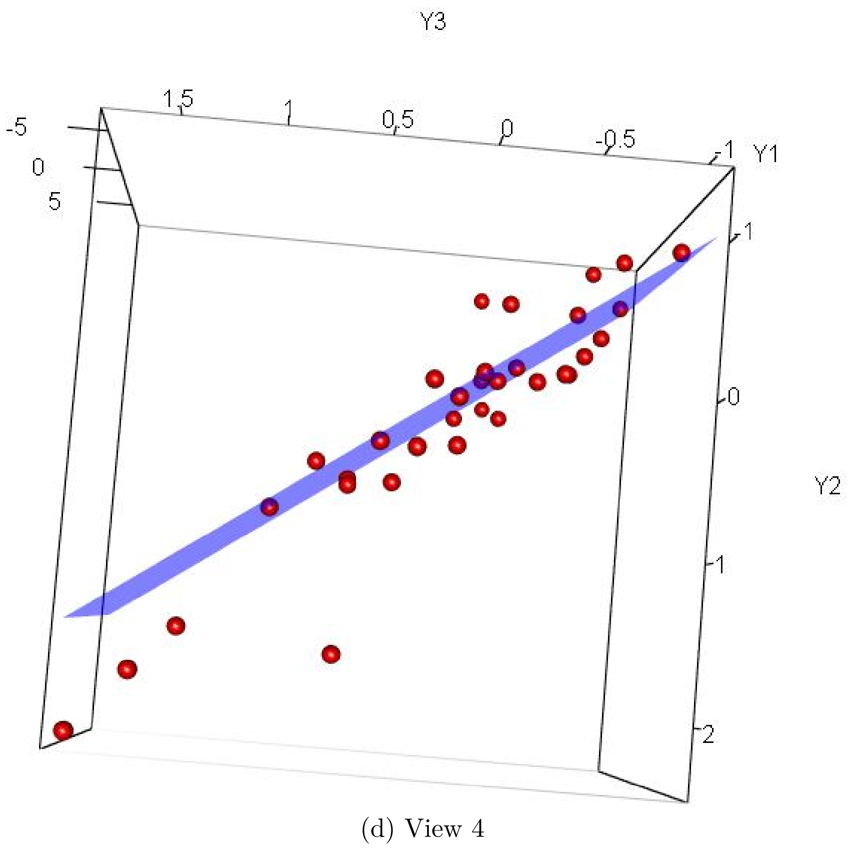


(b) View 2

Figure 3.7: Fitted plane of the practical example

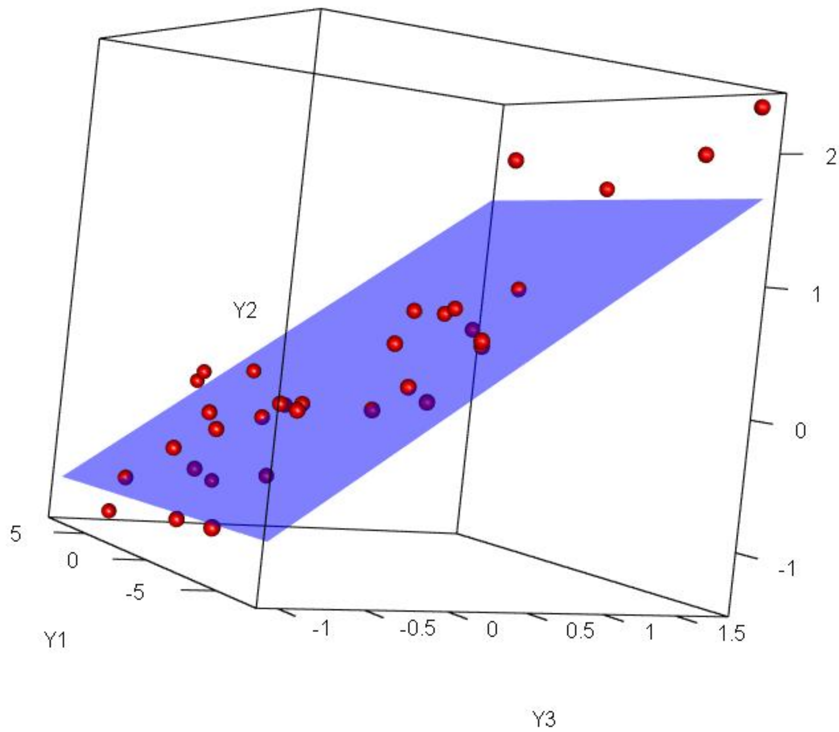


(c) View 3

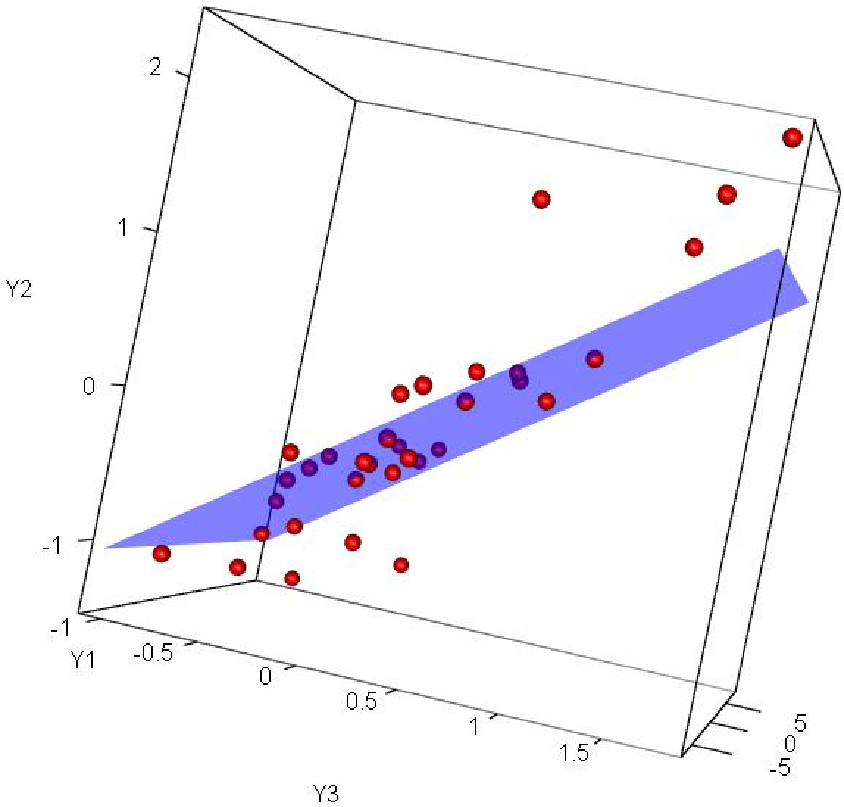


(d) View 4

Figure 3.7: Fitted plane of the practical example (continue)

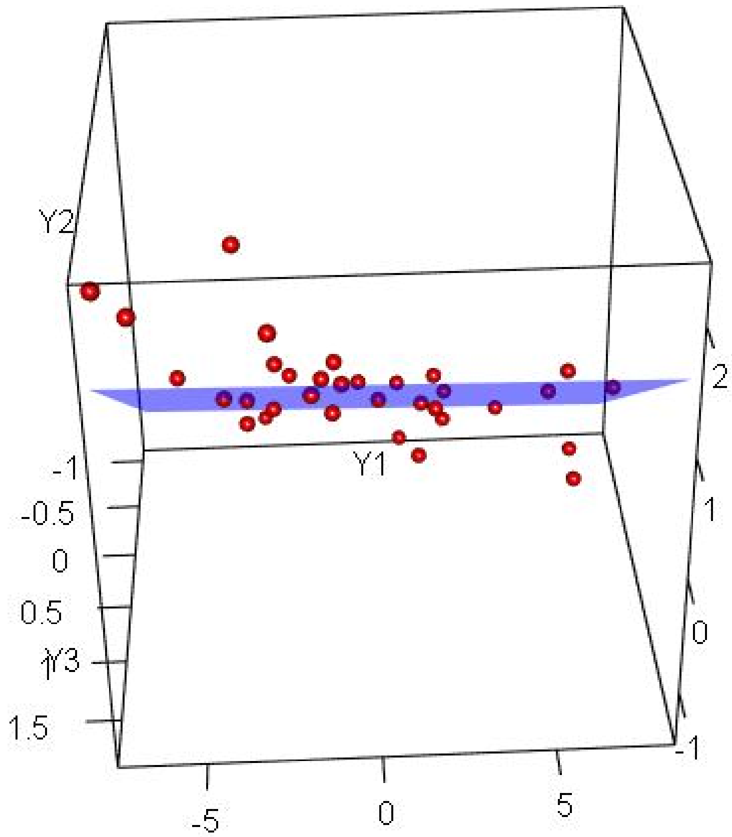


(e) View 5



(f) View 6

Figure 3.7: Fitted plane of the practical example (continue)



(g) View 7

Figure 3.7: Fitted plane of the practical example (continue)

3.3.3 Concluding remarks

Our analysis model is a follow-up to the MvLR for estimating the linear association between predictors and responses. In terms of the statistical calculus, there is no difference among all the variables, including the DVs and IVs in the model; therefore, it is possible to use the model to compare variables. When the number of DVs and IVs is one, the analytical model transforms into a Deming regression, which is a method widely used for the comparative evaluation of equipment in chemistry.

Notably, the smallest eigenvector is the solution of the MMOLR model; however, this is very sensitive to outliers, as stated at the end of subsection 3.2.1. Therefore, the effect of outliers on the estimation results is a suggestion for future research.

In the next chapter, a more complex linear model is mentioned. In Chapter 3, we enhance the estimation methods for the linear model, which is widely used in behavioral science, psychology, and supply-chain management, among others.

Chapter 4

K-means Structural Equation Modeling via Generalized Maximum Entropy

4.1 Overview

Structure equation model (SEM) is a multivariate analytical technique widely used in research areas such as psychology and marketing. It is especially used in studies on supply chain management. A common feature of SEM-applicable domains is that there are concepts that cannot be measured directly. Sometimes, SEM is used to describe the causal relationship between these concepts. For example, in the field of marketing, researchers evaluate brand awareness (BA) as well as customer loyalty (CL), and the relationship between these two concepts. However, these two concepts cannot be measured or directly observed. SEM allows us to develop and evaluate the concept through clearly measurable manifestations.

Specifically, we assume that a customer

i. who repeatedly returns to buy a product (CL1),

ii. willing/has introduced products to acquaintances (CL2), and

iii. has no intention of switching to another similar product (CL3)

is rated as a loyal customer. We can evaluate CL by observing CL1, CL2, and CL3. In this relationship, CL is the latent variable/construct and variables CL1, CL2, and CL3 are called observed variables; the same holds for BA.

Currently, there are three techniques to estimate the parameters in SEM: total least squares (TLS), maximum likelihood (ML), and generalized maximum entropy (GME). TLS and ML techniques are applied when the statistical assumption is "as observed variables with normal distribution." However, ML is more widely applied in application research communities and commercial software using SEM. GME has a long history of being employed in other industries. It does not require statistical assumptions of the normal distribution, which is an advantage compared with TLS and ML. However, in the context of SEM, the GME is not widely accepted because GME–SEM has many constrained formulation for parameters. Many constrained formulation for parameters make the estimation of parameters difficult. Moreover, the GME–SEM still has many shortcomings that will be discussed more in the upcoming sections.

This study aims to

i. synthesize theories to provide full knowledge and elevate GME to a clear branch of SEM and

ii. to improve current technologies.

In this section, we present the summary of the ML method and the shortcomings of applying this estimation method, thereby showing why it is necessary to include a parameter estimation method in SEM.

4.1.1 Structural Equation Modeling

SEM, as introduced by Jöreskog (1978), is a multivariate statistical analysis technique for discovering or describing complex systems. Jöreskog specifically separates the model into measurement and structural models:

- *The measurement model*, depicted by B.12 and 4.2, is as follows:

$$\boldsymbol{x} = \boldsymbol{\Lambda}^{x}{}_{(p_x \times r_x)}\boldsymbol{\xi} + \boldsymbol{\delta}, \tag{4.1}$$

  where $\boldsymbol{x}$ is the vector of p_x exogenous observed variables, $\boldsymbol{\Lambda}^x$ is a matrix of coefficients, and $\boldsymbol{\xi}$ is a vector of n exogenous latent constructs and disturbance vector $\boldsymbol{\delta}$.

  We interpret the latent variables (i.e., concepts) using $\boldsymbol{\Lambda}^x$.

$$\boldsymbol{y} = \boldsymbol{\Lambda}^{y}{}_{(p_y \times r_y)}\boldsymbol{\eta} + \boldsymbol{\varepsilon}, \tag{4.2}$$

where $\boldsymbol{y}$ is the vector of p_y endogenous observed variables, $\boldsymbol{\Lambda}^y$ is a matrix of the coefficients, and $\boldsymbol{\eta}$ is a vector of m endogenous latent variables and disturbance vector $\boldsymbol{\varepsilon}$.

We interpret the relationship between $\boldsymbol{y}$ and latent variables using $\boldsymbol{\Lambda}^x$.

- *The structural model* can be defined as follows:

$$\boldsymbol{\eta} = \boldsymbol{B}_{(r_y \times r_y)}\boldsymbol{\eta} + \boldsymbol{\Gamma}_{(r_y \times r_x)}\boldsymbol{\xi} + \boldsymbol{\zeta}_{(m \times 1)}, \tag{4.3}$$

where $\boldsymbol{B}$ is also a matrix of the coefficients, $\boldsymbol{\Gamma}$ is a matrix of the relationships between the exogenous and endogenous latent variables, and $\boldsymbol{\zeta}$ is the vector of disturbances/errors. From $\boldsymbol{B}$ and $\boldsymbol{\Gamma}$, we interpret the relationship among the latent variables.

The most popular estimation method for this model is the maximum likelihood (Anderson and Gerbing, 1988). This method compares the observed covariance matrix ($\boldsymbol{S}$), that is, the covariance matrix of data, with the implied covariance matrix ($\boldsymbol{\Sigma}$), which can be calculated as 4.4 to gain a certain model-fit.

$$\boldsymbol{\Sigma} = \begin{bmatrix} \mathbf{Cov}(\boldsymbol{x}, \boldsymbol{x}) & \mathbf{Cov}(\boldsymbol{x}, \boldsymbol{y}) \\ \mathbf{Cov}(\boldsymbol{y}, \boldsymbol{x}) & \mathbf{Cov}(\boldsymbol{y}, \boldsymbol{y}) \end{bmatrix}, \tag{4.4}$$

where

$$\begin{aligned} \mathbf{Cov}(\boldsymbol{x}, \boldsymbol{x}) &= E(\boldsymbol{x}\boldsymbol{x}') \\ &= E[(\boldsymbol{\Lambda}^x\boldsymbol{\xi} + \boldsymbol{\delta})(\boldsymbol{\Lambda}^x\boldsymbol{\xi} + \boldsymbol{\delta})'] \\ &= E[\boldsymbol{\Lambda}^x\boldsymbol{\xi}\boldsymbol{\xi}'\boldsymbol{\Lambda}^{x\prime} + \boldsymbol{\delta}\boldsymbol{\xi}'\boldsymbol{\Lambda}^{x\prime} + \boldsymbol{\Lambda}^x\boldsymbol{\xi}\boldsymbol{\delta}' + \boldsymbol{\delta}\boldsymbol{\delta}'] \\ &= \boldsymbol{\Lambda}^x E(\boldsymbol{\xi}\boldsymbol{\xi}')\boldsymbol{\Lambda}^{x\prime} + E(\boldsymbol{\delta}\boldsymbol{\xi}')\boldsymbol{\Lambda}^{x\prime} + \boldsymbol{\Lambda}^x E(\boldsymbol{\xi}\boldsymbol{\delta}') + E(\boldsymbol{\delta}\boldsymbol{\delta}'). \end{aligned} \tag{4.5}$$

Assuming that the error terms of indicators are independent of the exogenous concepts, this then means that $E(\boldsymbol{\delta}\boldsymbol{\xi}') = 0$, $E(\boldsymbol{\xi}\boldsymbol{\delta}') = 0$ and set $E(\boldsymbol{\xi}\boldsymbol{\xi}') = \boldsymbol{\Phi}$, $E(\boldsymbol{\delta}\boldsymbol{\delta}') = \boldsymbol{\Theta}_\delta$.

Then,

$$\mathbf{Cov}(\boldsymbol{x}, \boldsymbol{x}) = \boldsymbol{\Lambda}^x\boldsymbol{\Phi}\boldsymbol{\Lambda}^{x\prime} + \boldsymbol{\Theta}_\delta. \tag{4.6}$$

Similarly, we have

$$
\begin{aligned}
\mathbf{Cov}(\boldsymbol{y},\boldsymbol{y}) &= E(\boldsymbol{y}\boldsymbol{y}') \\
&= E[(\boldsymbol{\Lambda}^y\boldsymbol{\eta}+\boldsymbol{\varepsilon})(\boldsymbol{\Lambda}^y\boldsymbol{\eta}+\boldsymbol{\varepsilon})'] \\
&= E[\boldsymbol{\Lambda}^y\boldsymbol{\eta}\boldsymbol{\eta}'\boldsymbol{\Lambda}^{y\prime}+\boldsymbol{\varepsilon}\boldsymbol{\eta}'\Lambda^{y\prime}+\boldsymbol{\Lambda}^y\boldsymbol{\eta}\boldsymbol{\varepsilon}'+\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'] \\
&= \boldsymbol{\Lambda}^y E(\boldsymbol{\eta}\boldsymbol{\eta}')\boldsymbol{\Lambda}^{y\prime}+E(\boldsymbol{\varepsilon}\boldsymbol{\eta}')\boldsymbol{\Lambda}^{y\prime}+\boldsymbol{\Lambda}^y E(\boldsymbol{\eta}\boldsymbol{\varepsilon}')+E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'). \qquad (4.7)
\end{aligned}
$$

With the same assumption as for $\boldsymbol{x}$, the errors are independent of the latent constructs, that is, $E(\boldsymbol{\varepsilon}\boldsymbol{\eta}')=0$, $E(\boldsymbol{\eta}\boldsymbol{\varepsilon}')=0$, and set $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}')=\boldsymbol{\Theta}_\varepsilon$. The following equations will help us find $E(\boldsymbol{\eta}\boldsymbol{\eta}')$ and determine this matrix; we can then find $\mathbf{Cov}(\boldsymbol{y},\boldsymbol{y})$.

First, with some of the transformation steps from (B.16), we obtain $\boldsymbol{\eta}=(\boldsymbol{I}-\boldsymbol{B})^{-1}(\boldsymbol{\Gamma}\boldsymbol{\xi}+\boldsymbol{\zeta})$ (4.8), and then

$$
\begin{aligned}
E(\boldsymbol{\eta}\boldsymbol{\eta}') &= E([(\boldsymbol{I}-\boldsymbol{B})^{-1}(\boldsymbol{\Gamma}\boldsymbol{\xi}+\boldsymbol{\zeta})][(\boldsymbol{I}-\boldsymbol{B})^{-1}(\boldsymbol{\Gamma}\boldsymbol{\xi}+\boldsymbol{\zeta})]') \\
&= E[(\boldsymbol{I}-\boldsymbol{B})^{-1}(\boldsymbol{\Gamma}\boldsymbol{\xi}+\boldsymbol{\zeta})(\boldsymbol{\Gamma}\boldsymbol{\xi}+\boldsymbol{\zeta})'(\boldsymbol{I}-\boldsymbol{B})^{-1\prime}] \\
&= (\boldsymbol{I}-\boldsymbol{B})^{-1}E[\boldsymbol{\Gamma}\boldsymbol{\xi}\boldsymbol{\xi}'\boldsymbol{\Gamma}'+\boldsymbol{\zeta}\boldsymbol{\xi}'\boldsymbol{\Gamma}'+\boldsymbol{\Gamma}\boldsymbol{\xi}\boldsymbol{\zeta}'+\boldsymbol{\zeta}\boldsymbol{\zeta}'](\boldsymbol{I}-\boldsymbol{B})^{-1\prime} \qquad (4.9) \\
&= (\boldsymbol{I}-\boldsymbol{B})^{-1}[\boldsymbol{\Gamma}E(\boldsymbol{\xi}\boldsymbol{\xi}')\boldsymbol{\Gamma}'+E(\boldsymbol{\zeta}\boldsymbol{\xi}')\boldsymbol{\Gamma}'+\boldsymbol{\Gamma}E(\boldsymbol{\xi}\boldsymbol{\zeta}')+E(\boldsymbol{\zeta}\boldsymbol{\zeta}')](\boldsymbol{I}-\boldsymbol{B})^{-1\prime}
\end{aligned}
$$

and set $E(\boldsymbol{\zeta}\boldsymbol{\zeta}')=\boldsymbol{\psi}$.

$$
\Rightarrow E(\boldsymbol{\eta}\boldsymbol{\eta}')=(\boldsymbol{I}-\boldsymbol{B})^{-1}(\boldsymbol{\Gamma}\boldsymbol{\Phi}\boldsymbol{\Gamma}'+\boldsymbol{\psi})(\boldsymbol{I}-\boldsymbol{B})^{-1\prime}. \qquad (4.10)
$$

Therefore,

$$
\mathbf{Cov}(\boldsymbol{y},\boldsymbol{y})=\boldsymbol{\Lambda}^y(\boldsymbol{I}-\boldsymbol{B})^{-1}(\boldsymbol{\Gamma}\boldsymbol{\Phi}\boldsymbol{\Gamma}'+\boldsymbol{\psi})(\boldsymbol{I}-\boldsymbol{B})^{-1\prime}\boldsymbol{\Lambda}^{y\prime}+\boldsymbol{\Theta}_\varepsilon. \qquad (4.11)
$$

Finally, $\mathbf{Cov}(\boldsymbol{x},\boldsymbol{y})=\mathbf{Cov}(\boldsymbol{y},\boldsymbol{x})$, where

$$
\begin{aligned}
\mathbf{Cov}(\boldsymbol{x},\boldsymbol{y}) &= E(\boldsymbol{x}\boldsymbol{y}') \\
&= E[(\boldsymbol{\Lambda}^x\boldsymbol{\xi}+\boldsymbol{\delta})(\boldsymbol{\Lambda}^y\boldsymbol{\eta}+\boldsymbol{\varepsilon})'] \\
&= E[\boldsymbol{\Lambda}^x\boldsymbol{\xi}\boldsymbol{\eta}'\boldsymbol{\Lambda}^{y\prime}+\boldsymbol{\delta}\boldsymbol{\eta}'\boldsymbol{\Lambda}^{y\prime}+\boldsymbol{\Lambda}^x\boldsymbol{\xi}\boldsymbol{\eta}'\boldsymbol{\Lambda}^{y\prime}+\boldsymbol{\delta}\boldsymbol{\varepsilon}'] \\
&= \boldsymbol{\Lambda}^x E(\boldsymbol{\xi}\boldsymbol{\eta}')\boldsymbol{\Lambda}^{y\prime}+E(\boldsymbol{\delta}\boldsymbol{\eta}')\boldsymbol{\Lambda}^{y\prime}+\boldsymbol{\Lambda}^x E(\boldsymbol{\xi}\eta')\boldsymbol{\Lambda}^{y\prime}+E(\boldsymbol{\delta}\boldsymbol{\varepsilon}') \\
&= \boldsymbol{\Lambda}^x E(\boldsymbol{\xi}\boldsymbol{\eta}')\boldsymbol{\Lambda}^{y\prime}; \qquad (4.12)
\end{aligned}
$$

and

$$
\begin{aligned}
E(\boldsymbol{\xi}\boldsymbol{\eta}') &= E[\boldsymbol{\xi}(\boldsymbol{I}-\boldsymbol{B})^{-1}(\boldsymbol{\Gamma}\boldsymbol{\xi}+\boldsymbol{\zeta})'] \\
&= E[\boldsymbol{\xi}(\boldsymbol{\xi}'\boldsymbol{\Gamma}'+\boldsymbol{\zeta}')(\boldsymbol{I}-\boldsymbol{B})^{-1'}] \\
&= E[\boldsymbol{\xi}\boldsymbol{\xi}'\boldsymbol{\Gamma}'(\boldsymbol{I}-\boldsymbol{B})^{-1'}+\boldsymbol{\xi}\boldsymbol{\zeta}'(\boldsymbol{I}-\boldsymbol{B})^{-1'}] \\
&= E(\boldsymbol{\xi}\boldsymbol{\xi}')\boldsymbol{\Gamma}'(\boldsymbol{I}-\boldsymbol{B})^{-1'}+E(\boldsymbol{\xi}\boldsymbol{\zeta}')(\boldsymbol{I}-\boldsymbol{B})^{-1'} \\
&= \boldsymbol{\Phi}\boldsymbol{\Gamma}'(\boldsymbol{I}-\boldsymbol{B})^{-1'}.
\end{aligned} \tag{4.13}
$$

From (4.12) and (4.13), we gain (4.14)

$$
\mathbf{Cov}(\boldsymbol{x},\boldsymbol{y}) = \boldsymbol{\Lambda}^{x}\boldsymbol{\Phi}\boldsymbol{\Gamma}'(\boldsymbol{I}-\boldsymbol{B})^{-1'}\boldsymbol{\Lambda}^{y'}. \tag{4.14}
$$

The maximum likelihood estimation function is

$$
\mathbf{F}_{ML} = \ln|\boldsymbol{\Sigma}| - \ln|\boldsymbol{S}| + \mathrm{tr}[\boldsymbol{S}\boldsymbol{\Sigma}^{-1}] - (p_x + p_y). \tag{4.15}
$$

The parameters/unknowns are only estimated if and only if the implied model satisfies the following identification rules:

- **Degree of freedom**:
  Model degree of freedom (d_f) is defined as the subtraction of the number of observed variance/covariance (p^*) and the number of parameters to estimate (t).

  $$
  d_f = p^* - t \tag{4.16}
  $$

  Typically, $p^* = p(p+1)/2$, where $p = (p_x + p_y)$ and the value of t depends on the specific model implied.

  There are three cases of degrees of freedom:

  - *Over-identified* ($d_f > 0$): there is no exact solution; only an approximate solution exists.
  - *Just-identified* ($d_f = 0$): there is only one solution for the model.
  - *Under-identified* ($d_f < 0$): there are no solutions for the model; a model with more free parameters than data points (i.e., too many paths in the model) is generally under-identified.

  A model is identified if it is possible to find one or more solutions from the data, that is, $d_f \geqslant 0$. However, an over-identified model may not fit as well as a just-identified model; therefore, we need statistical hypotheses tests, including a global model fit (Loehlin, 1998).

An additional complication that can arise is *empirical under-identification.* This occurs when we perform the inversion of a parameter matrix that establishes a model identification that has a very small (close to zero) estimate.

For structural under-identification, the only solution is to re-specify the model. Empirical under-identification may be correctable by collecting more data or re-specifying the model.

- **The non-recursive rule**: There are no loops in the relationships among endogenous variables and uncorrelated latent disturbances.

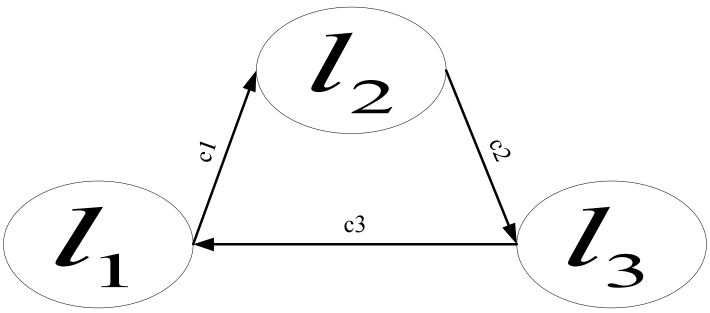


Figure 4.1: Example of a recursive model.

4.1.2 Research Problem

First, we explain the motivative example. In a joint research project, we applied SEM to depict numerous concepts and their relationships in supply chain management. The model of this research includes 10 latent variables as in Figure 4.3 and 50 manifests as in Table 4.1. Hair et al. (1998) and Hoelter (1983) recommend that the minimum sample size should be from between 100 to 150 or even from 200 when using the ML estimation method. Therefore, with this model, we need data with 200 objects.

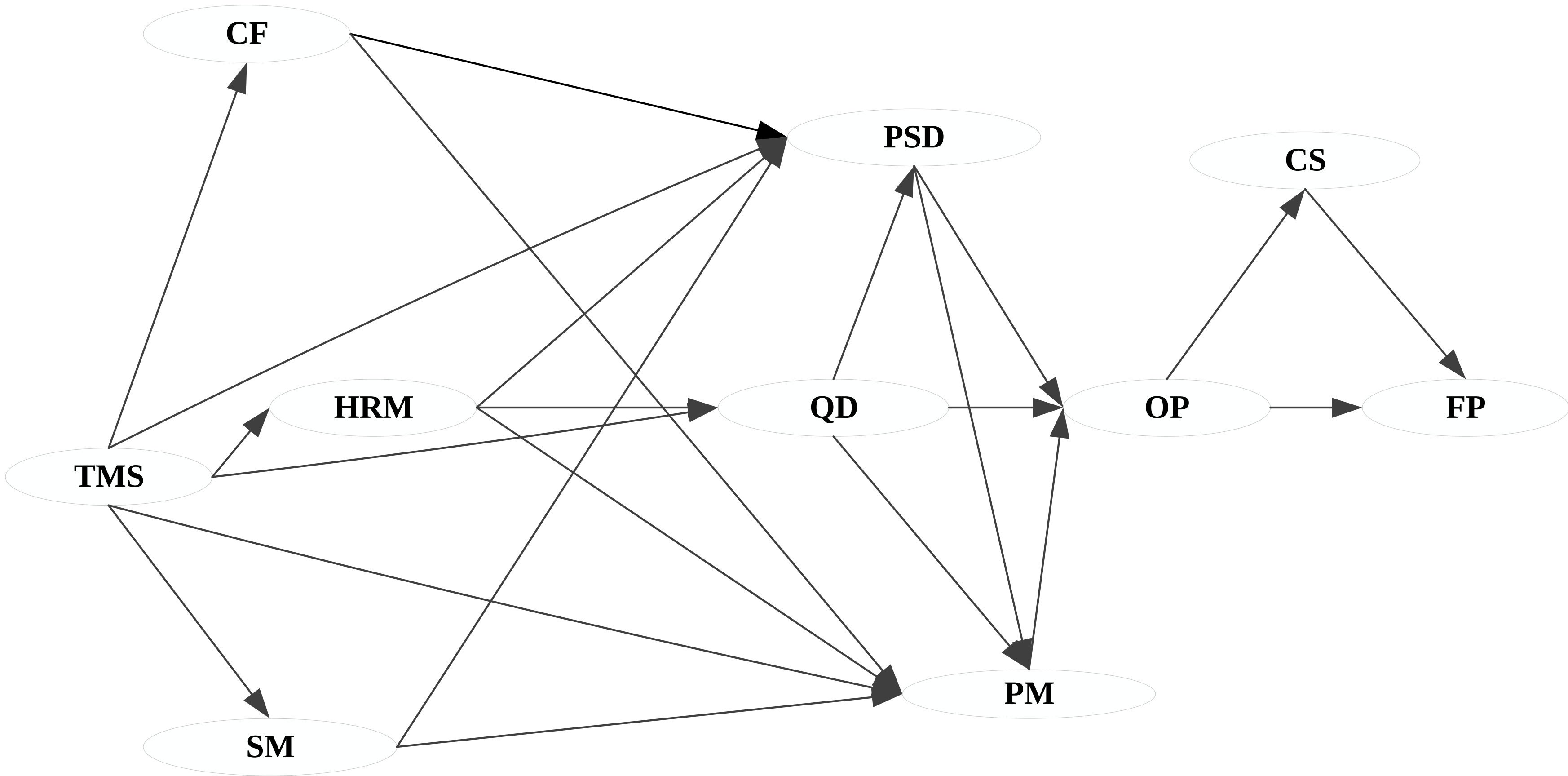


Figure 4.2: Research model: empirical example

Table 4.1: Designed measurement model: empirical example

No.	Constructs	Observed items
1	QD	QD1 QD2 QD3 QD4 QD5
2	PSD	PSD6 PSD7 PSD8 PSD9 PSD10
3	PM	PM11 PM12 PM13 PM14 PM15
4	CF	CF16 CF17 CF18 CF19 CF20
5	SM	SM21 SM22 SM23 SM24 SM25
6	TMS	TMS26 TMS27 TMS28 TMS29 TMS30 TMS 31

Continued on the next page

No.	Constructs	Observed items
7	HRM	HRM32 HRM33 HRM34 HRM35 HRM36 HRM37 HRM38 HRM39
8	OP	OP40 OP41 OP42 OP43 OP44
9	CS	CS45 CS46
10	FP	FP47 FP48 FP49 FP50

Collecting data is a difficult task, especially when the objects are managers. We first emailed around 50 structural questionnaires to managers. The response rate, however, was very low. Only three people responded, and they were those who already knew about our research work. Finally, substantial resources were spent on contacting the research population over a three-month period, which resulted in 179 valid questionnaires.

This empirical research shows that the prerequisite on sample size causes certain barriers to research, such as the high cost of collecting data or the inaccessibility of population.

The entropy principle was first introduced by Shannon (1948) and was based on the principle of indifference. Following this, Jaynes (1957, 1984) developed Shannon's approach to produce the maximum entropy principle (MEP) with which to estimate the probability distribution by maximizing the entropy function. This new approach was considered a powerful approach for both ill-posed (where there are more variables than observations) and well-posed data. The MEP combines constraints into the model but does not directly solve the objective function; therefore, the problem of under-identifying in complicated models such as SEM is also eliminated.

As computer power has increased, the MEP has gradually been applied to statistical estimations as well (Berger et al., 1996). Golan et al. (1996) went further by re-parameterizing and reformulating the MEP and accompanying it with the necessary inferences. This process is now known as the generalized maximum entropy (GME) estimation (Al-Nasser, 2003). GME-SEM is thus an estimation method without a strict distribution assumption of data or sample size. However, in this method, each data point becomes a constraint that might lead to redundant constraints when the number of observations increases, leading to over-fitting and heavy calculation cost.

Therefore, in this study, the method of K-means-GME-SEM is proposed for reducing the number of constraints and heavy calculation cost. In the new method, instead of using whole data as a consistency constraint, some means of groups in data are used, thus reducing the number of constraints.

4.2 K-means Generalized Maximum Entropy Estimation for Structural Equation Modeling

This section presents the fundamental and most important theoretical foundation, along with the algorithm for K-means GME-SEM method. The highlight of the study here is not only the synthesizing and arranging of the theoretical parts of GME for SEM, but the improvements through a combination of K-means clustering and GME-SEM.

4.2.1 Generalized Maximum Entropy

Information Theory

The basis of information theory is to quantify the volume of information of a message/data transmitted through a noisy channel. In this theory, information can be considered the resolution of uncertainty, that is, as information is treated as a set of possible messages of a communication over a noisy channel, it can be reconstructed with low probability of error (Shannon, 1948). The argument is that nearly error-free communication/information can be achieved through the channel and has a greater capacity than the entropy of the source. By this approach, a statistical sample could be considered a noisy channel, and the message is about parameters with prior distribution (Golan, 2002). The purpose, then, is to retrieve that distribution. The idea is to recover the distribution using entropy as a measure of disorder in a dataset and information to gain a measure of the decrease in disorder achieved by partitioning the original dataset.

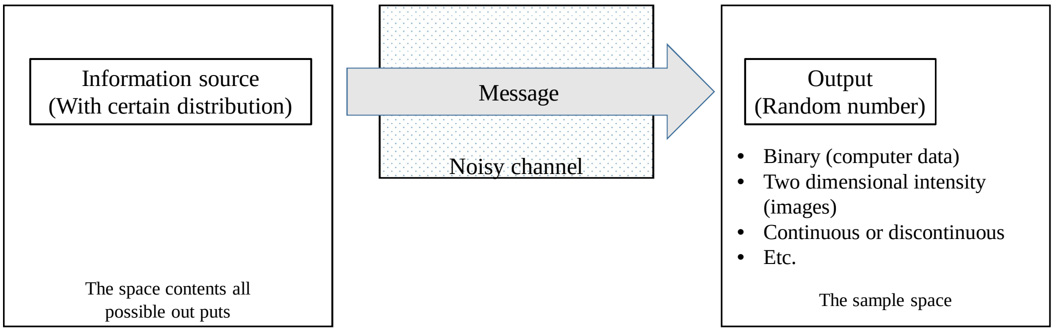


Figure 4.3: Information theory model

In particular, maximum entropy information theory based on probability theory and statistics gives a certain criterion for finding probability distributions from the basis of partial knowledge. This leads to a statistical inference that is known as the maximum entropy, which is considered as a least biased estimation method (Jaynes, 1957). Furthermore, maximum entropy is taken to be a general approach of logit models (Soofi, 1994).

Entropy

Given a finite set $A = \{a_1, a_2, \ldots a_m\}$ with a probability mass function π on A, the entropy function is defined by Shannon (1948) as follows:

$$H(\pi) \equiv -\sum_{i=1}^{m} \pi_i \ln \pi_i, \tag{4.17}$$

where $\pi_i = P(a_i)$. The probability distribution must satisfy the normalization condition (4.18), and $0 \ln 0 = 0$.

$$\sum_{i=1}^{m} \pi_i = 1. \tag{4.18}$$

Entropy is a measure of the volume of information transmitted in a message through a channel.

Maximum Entropy Principle

The entropy measures uncertainty or the average amount of information provided by p (Jaynes, 1957, 1984; Golan et al., 1996; Golan, 2002). Uncertainty refers to situations with imperfect or unknown information and can be applied to predictions of future events, that is, uncertainty means the range of possible values within which the true value of the measurement lies (Golan, 2002).

The MEP gives the probability density function with certain constraints and has entropy as large as possible, leading to a decision with the least surprising or without unwarranted information (Jaynes, 1957, 1984; Golan et al., 1996; Conrad, 2004; Bickel, 2015). Moreover, MEP provides a guide with which to solve the constraints of a model rather than directly concerning the criterion function (Berger et al., 1996; Golan et al., 1996). According to Conrad (2004), all Gaussians have the same value of entropy on infinitely maximum entropy distribution with a fixed variance; therefore, the difference in solution is from constraints. It allows the largest amount of information to be extracted from any given small data with a unique solution (Paris and Howitt, 1998; Conrad, 2004; Berger et al., 1996).

MEP is a robust statistical inference that is applied in recovering techniques or to inverse problems. Many studies have shown surprising results in recovering images (Studholme et al., 1999), estimating coefficients for dynamic systems (Golan et al., 1996), classifying (Nigam et al., 1999), and sampling (Wynn, 1993).

In the case of ill-posed data [1], the maximum likelihood estimation is non-unique (Csiszar, 1991) and the least squares objective function vanishes (Paris and Howitt, 1998). However, the MEP approach provides a unique solution (Golan et al., 1996; Paris and Howitt, 1998).

With well-posed data, an MEP estimation will be more reliable (Paris and Howitt, 1998). It seems that we can solve almost any complex problem with very limited data. This powerful characteristic comes from the fact that no distribution assumptions are required (Gupta et al., 2007). The MEP' s constraint rule is unsatisfactory (Uffink, 1996) and we should not ignore these debates. However, any approach has certain controversies; yet, MEP still shows remarkable efficiency in many fields.

Generalized Maximum Entropy Estimation

The fundamentals of MEP come from probability, but the coefficients of many models do not have the properties of probabilities. Therefore, a re-parameterization and reformulation procedure was proposed by Golan et al. (1996) to overcome this.

- Re-parameterization:

  The idea here is that a coefficient can be written as a linear combination of discrete points with corresponding weights, that is, a parameter β is a convex combination

[1]The number of unknown parameters is larger than the number of data points/observations.

of finite K support points z_k and probabilities π_k so that $\pi_k \geqslant 0$ and $\sum_{k=1}^{K} \pi_k = 1$.

$$\beta = \sum_{k=1}^{K} z_k \pi_k = \boldsymbol{z}'\boldsymbol{\pi}, \tag{4.19}$$

where $\boldsymbol{z}' = (z_1, z_2, \ldots, z_K)$ and $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots, \pi_K)'$.

The figure (4.4) gives a simple explanation of a convex combination, where P is a convex combination of z_1 z_2 z_3, but Q is not.

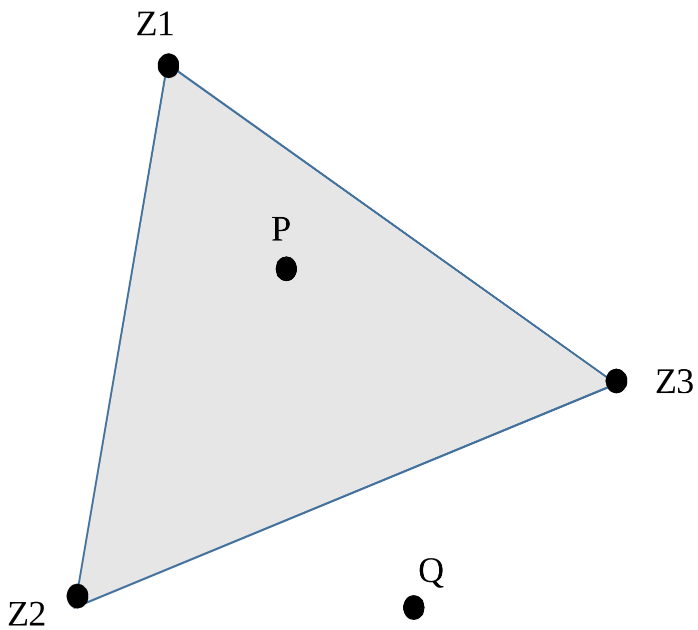


Figure 4.4: Example of a convex combination

Now, instead of estimating β, we face the problem of recovering $\boldsymbol{\pi}$ with the chosen $\boldsymbol{z}$. Note that choosing $\boldsymbol{z}$ will span a space for each parameter and this could affect the estimation results. Furthermore, the recovered parameters are not affected by the number of support points, but are variant to different end-points (Paris and Howitt, 1998). To assure the recovered β is not aberrant, the support interval should be around the usual value of the parameters. For example, if we believe or empirical research shows that $0 \leqslant \beta \leqslant 1$ and choose $K = 5$, we must then specify $\boldsymbol{z}' = (0, 0.25, 0.5, 0.75, 1)$.

- Re-formulation:

The main purpose of this step is to define the objective function and constraints. To make use of MEP, the objective function is the maximum nonlinear Shannon's entropy function. Furthermore, there are two kinds of constraints: consistency and normalization. The consistency constraint comes from data and the statistic model (e.g., linear regression), whereas the normalization constraint comes from the definition of a convex combination and/or when the total probability is one. In the next section, the procedure of generating an objective function and constraints for SEM will be described in detail.

To measure the goodness of the recovered model, the normalized entropy measure $S(\tilde{\pi})$ (Soofi, 1992; Golan, 2008) is given. Where $0 \leqslant S(\tilde{\pi}) \leqslant 1$, the value 0 implies perfect certainty, whereas 1 implies total uncertainty, and the calculation of $S(\tilde{\pi})$ is expressed as follows:

$$S(\tilde{\pi}) = \frac{-\tilde{\pi}' \ln(\tilde{\pi})}{M \ln(D)}, \tag{4.20}$$

where $\tilde{\pi}$ is the estimated value of π and the number of parameters to estimate M.

4.2.2 K-means Clustering

The K-means algorithm was first introduced by Lloyd (1957). It is a method of vector quantization and a popular clustering algorithm. This algorithm starts with a first group of randomly selected centroids, and then repeatedly calculates to optimize the positions of the centroids. The iterative halts when the centroids have stabilized, or the defined number of iterations has been achieved. In the result, N observations are partitioned into q groups/clusters with the nearest mean[2]. Below is the mathematical depiction of K-means clustering.

Given a set of data points $(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N)$, $\boldsymbol{x}_i \in \mathbb{R}_x^p$, these N observations are assigned into $q \leqslant N$ sets $S = (S_1, S_2, \ldots, S_q)$ by minimizing the objective function (4.21):

$$F = \sum_{i=1}^{n} \sum_{j=1}^{q} w_{ij} \|\boldsymbol{x}_i - \boldsymbol{\mu}_j\|^2, \tag{4.21}$$

where $w_{ij} = 1$ if data point x_i belongs to cluster S_j; otherwise, $w_{ij} = 0$; and $\boldsymbol{\mu}_j$ is the mean/centroid of cluster S_j.

The iteration to minimize the objective function is conducted in two main step: assigning the data points into the closest cluster (E-step) and computing the centroid of each cluster (M-step).

- E-step: assign the x_i into the closest cluster judged by sum of squared (squared Euclidean) distance from cluster' s centroid:

[2]minimum the sum of the squared distance between the data points and the cluster' s centroid

$$\frac{\partial F}{\partial w_{ij}} = \sum_{i=1}^{n}\sum_{j=1}^{q} \left\|\boldsymbol{x}_i - \boldsymbol{\mu}_j\right\|^2 \tag{4.22}$$

$$\Rightarrow w_{ij} = \begin{cases} 1 \text{ if } j = \arg\min_j \left\|\boldsymbol{x}_i - \boldsymbol{\mu}_j\right\|^2 \\ 0 \text{ otherwise} \end{cases}$$

- M-step: recomputing the centroid of each cluster:

$$\frac{\partial F}{\partial \boldsymbol{\mu}_j} = 2\sum_{i=1}^{n} w_{ij}\left(\boldsymbol{x}_i - \boldsymbol{\mu}_j\right) = 0 \tag{4.23}$$

$$\Rightarrow \boldsymbol{\mu}_j = \frac{\sum\limits_{i=1}^{n} w_{ij}\boldsymbol{x}_i}{\sum\limits_{i=1}^{m} w_{ij}}$$

Note: at the start of the algorithm, different initializations may lead to different results because the algorithm can be stuck in local optimum, that is, not converge to global optimum.

When changing the distance calculation, we use a different algorithm. Beside Euclidean distance, there are two other popular distances used in K-means algorithm: (weighted) Manhattan distance (see Figure 4.6) and (weighted) Minkowski distance (Singh et al., 2013).

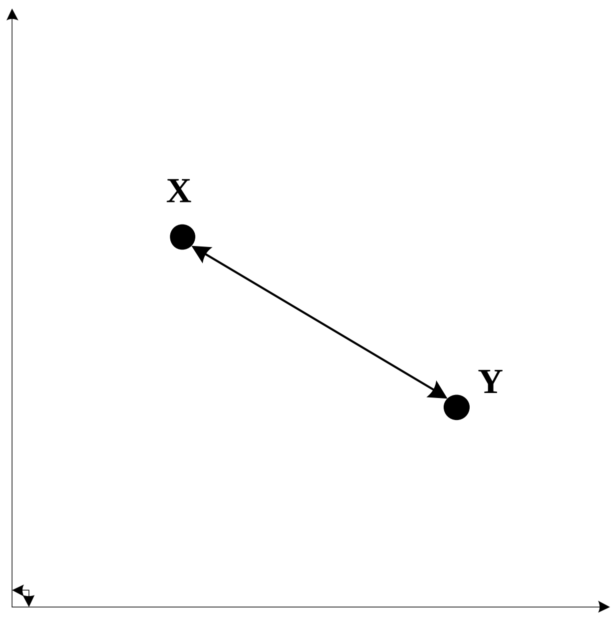


Figure 4.5: Euclidean distance

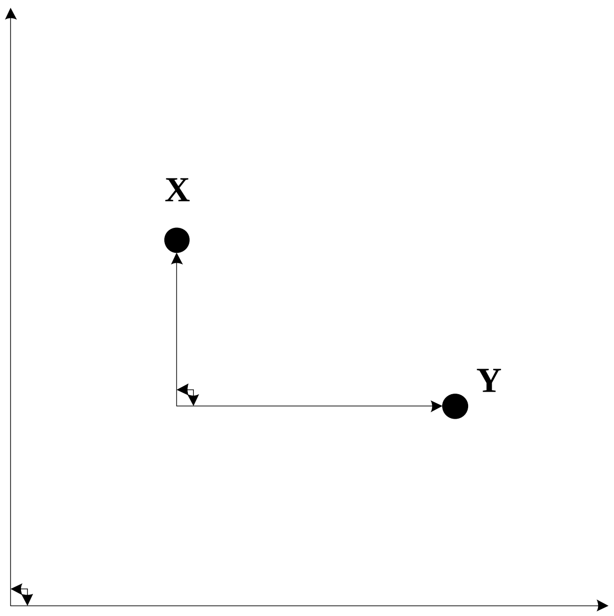


Figure 4.6: Manhattan distance

- (weighted) Manhattan distance [3] is calculated by summing the horizontal and vertical distances on a grid, that is, summing the absolute differences of their Cartesian coordinates. For example, given $X = (x_1, x_2)$ and $Y = (y_1, y_2)$, the Manhattan distance is $|x_1 - y_1| + |x_2 - y_2|$. In general, the Manhattan distance is calculated by equation (4.24).

$$Manhattan(X, Y) = \sum_{j=1}^{p} w_i \left(|x_i - y_i|\right), \tag{4.24}$$

  where p is number of dimensions and $\boldsymbol{w} = (w_i)$ is the weight vector.

- (weighted) Minkowski distance is a generalized metric form of Manhattan distance and Euclidean distance.

$$Minkowski(X, Y, \lambda) = \sqrt[\lambda]{\sum_{j=1}^{p} w_p^{\lambda} (|x_p - y_p|)^{\lambda}}, \tag{4.25}$$

  where λ is the degree of Minkowski; when $\lambda = 1$, the Minkowski distance equals the Manhattan distance, and when $\lambda = 2$, the Minkowski distances equals the Euclidean distance.

[3] also known as city block distance

4.2.3 Generalized Maximum Entropy Estimation for Structural Equation Modeling (GME-SEM)

SEM is depicted in a general form as follows:

- Measurement model:

$$\boldsymbol{x} = \boldsymbol{A}_{(p_x \times r)}\boldsymbol{\gamma} + \boldsymbol{\varepsilon} \tag{4.26}$$

- Structural model:

$$\boldsymbol{\gamma} = \boldsymbol{B}_{(r \times r)}\boldsymbol{\gamma} + \boldsymbol{\varphi}, \tag{4.27}$$

where $\boldsymbol{x}$ is the vector of the observed variables (manifests), $\boldsymbol{\gamma}$ is the latent construct vector, $\boldsymbol{A} = (\alpha_{ij})$ is the matrix of coefficients that reflect the relationship between the manifests and latent constructs, $\boldsymbol{B} = (\beta_{ef})$ is the matrix of the coefficients reflecting the relationship between the latent constructs, and $\boldsymbol{\epsilon}$ and $\boldsymbol{\varphi}$ are the vectors of disturbances with elements ε and φ, respectively.

The objective is to estimate the coefficients $\boldsymbol{A}$, $\boldsymbol{B}$ and disturbances $\boldsymbol{\epsilon}$ and $\boldsymbol{\varphi}$. To use GME based on the procedure from Golan et al. (1996), these unknowns are reparameterized. Choosing K as the number of support points, every coefficient is rewritten in the form of $\alpha_{ij} = \boldsymbol{z}_{ij}^{\alpha\prime}\boldsymbol{\pi}_{ij}^{\alpha}$, $\beta_{ef} = \boldsymbol{z}_{ef}^{\beta\prime}\boldsymbol{\pi}_{ef}^{\beta}$, where $\boldsymbol{z}_{ij}^{\alpha\prime} = (z_1^{(\alpha_{ij})}, z_2^{(\alpha_{ij})}, \dots, z_K^{(\alpha_{ij})})$; $\boldsymbol{z}_{ef}^{\beta\prime} = (z_1^{(\beta_{ef})}, z_2^{(\beta_{ef})}, \dots, z_K^{(\beta_{ef})})$; $\boldsymbol{\pi}_{ij}^{\alpha} = (\pi_1^{(\alpha_{ij})}, \pi_2^{(\alpha_{ij})}, \dots, \pi_K^{(\alpha_{ij})})'$ and $\boldsymbol{\pi}_{ef}^{\beta} = (\pi_1^{(\beta_{ef})}, \pi_2^{(\beta_{ef})}, \dots, \pi_K^{(\beta_{ef})})'$ with normalization constraints $\sum\limits_{k=1}^{K} \pi_k^{(\alpha_{ij})} = \sum\limits_{k=1}^{K} \pi_k^{(\beta_{ef})} = 1$; $\pi_k^{(\alpha_{ij})}, \pi_k^{(\beta_{ef})} \geqslant 0$; for all $i = 1, 2, \dots, N$, $j = 1, 2, \dots, p_x$; $e = 1, 2, \dots, r_x$; $f = 1, 2, \dots, r_x$. Similarly, by the number of support points L, disturbances are also in the form of $\varepsilon_i = \boldsymbol{v}_i^{\varepsilon\prime}\boldsymbol{\pi}_i^{\varepsilon}$; $\varphi_e = \boldsymbol{v}_e^{\varphi\prime}\boldsymbol{\pi}_e^{\varphi}$; where $\boldsymbol{v}_i^{\varepsilon\prime} = (v_1^{(\varepsilon_i)}, v_2^{(\varepsilon_i)}, \dots, v_L^{(\varepsilon_i)})$, $\boldsymbol{v}_e^{\varphi\prime} = (v_1^{(\varphi_e)}, v_2^{(\varphi_e)}, \dots, v_L^{(\varphi_e)})$, $\boldsymbol{\pi}_i^{\varepsilon} = (\pi_1^{(\varepsilon_i)}, \pi_2^{(\varepsilon_i)}, \dots, \pi_L^{(\varepsilon_i)})'$ and $\boldsymbol{\pi}_e^{\varphi} = (\pi_1^{(\varphi_e)}, \pi_2^{(\varphi_e)}, \dots, \pi_L^{(\varphi_e)})'$ with normalization constraints $\sum\limits_{\ell=1}^{L} \pi_\ell^{(\varepsilon_i)} = \sum\limits_{\ell=1}^{L} \pi_\ell^{(\varphi_e)} = 1$; $\pi_\ell^{(\varepsilon_i)}, \pi_k^{(\varphi_e)} \geqslant 0$; for all $i = 1, 2, \dots, N$, $e = 1, 2, \dots, r_x$.

From empirical research and simulation, Golan et al. (1996) revealed that the mean squared error (MSE) would be better if $K = L = 5$. If the value of coefficients belong to $[-\alpha; \alpha]$, then the support vector $\boldsymbol{z}' = \alpha(-1, -0.5, 0, 0.5, 1)$. The value of α should come from theoretical or empirical research; if not, we should perform a sensitivity analysis by considering the results with different α.

Because the end-points of the support vectors for the disturbances proposed follow the three sigma rule (Pukelsheim, 1994) by Ciavolino and Al-Nasser (2009), then $\boldsymbol{v}' = 3\sigma(-1, -0.5, 0, 0.5, 1)$, where σ is the empirical standard deviation of the standardized errors. However, because choosing support spaces relies heavily on empirical research, this is one of the disadvantages to GME-SEM, as it could cause confusion in innovative research.

Once the support vectors have been established, unknowns are simply estimated through the recovering probabilities by maximizing the entropy function (4.28) subjected to constraints.

$$\begin{aligned} H(\boldsymbol{\pi}^{\mathrm{A}}, \boldsymbol{\pi}^{\mathrm{B}}, \boldsymbol{\pi}^{\varepsilon}, \boldsymbol{\pi}^{\varphi}) &= H(\boldsymbol{\pi}^{\mathrm{A}}) + H(\boldsymbol{\pi}^{\mathrm{B}}) + H(\boldsymbol{\pi}^{\varepsilon}) + H(\boldsymbol{\pi}^{\varphi}) \\ &= -\boldsymbol{\pi}^{\mathrm{A}\prime} \ln \boldsymbol{\pi}^{\mathrm{A}} - \boldsymbol{\pi}^{\mathrm{B}\prime} \ln \boldsymbol{\pi}^{\mathrm{B}} - \boldsymbol{\pi}^{\varepsilon\prime} \ln \boldsymbol{\pi}^{\varepsilon} - \boldsymbol{\pi}^{\varphi\prime} \ln \boldsymbol{\pi}^{\varphi}, \end{aligned} \tag{4.28}$$

where $\boldsymbol{\pi}^{\mathrm{A}} = (\boldsymbol{\pi}_{11}^{\alpha}{}', \boldsymbol{\pi}_{12}^{\alpha}{}', \ldots, \boldsymbol{\pi}_{1r_x}^{\alpha}{}', \boldsymbol{\pi}_{21}^{\alpha}{}', \ldots, \boldsymbol{\pi}_{Nr_x}^{\alpha}{}')'$;
$\boldsymbol{p}^{\mathrm{B}} = (\boldsymbol{\pi}_{11}^{\beta}{}', \boldsymbol{\pi}_{12}^{\beta}{}', \ldots, \boldsymbol{\pi}_{1m}^{\beta}{}', \boldsymbol{\pi}_{21}^{\beta}{}', \ldots, \boldsymbol{\pi}_{r_x r_x}^{\beta}{}')'$; $\boldsymbol{\pi}^{\varepsilon} = (\boldsymbol{\pi}_1^{\varepsilon\prime}, \boldsymbol{\pi}_2^{\varepsilon\prime}, \ldots, \boldsymbol{\pi}_N^{\varepsilon}{}')'$;
$\boldsymbol{\pi}^{\varphi} = (\boldsymbol{\pi}_1^{\varphi\prime}, \boldsymbol{\pi}_2^{\varphi}, \ldots, \boldsymbol{\pi}_{r_x}^{\varphi}{}')'$, and $\ln \boldsymbol{\pi}$ is element wise logarithmic function means $\ln \boldsymbol{\pi} = (\ln \pi_i)$.

Following GME-SEM, we eliminate the latent constructs and re-parameterize; (4.26) and (4.27) could be combined into a consistency constraint for each data point as (4.29).

$$\boldsymbol{x}_i = \boldsymbol{A}(\boldsymbol{I} - \boldsymbol{B})^{-1} \boldsymbol{\varphi}_i + \boldsymbol{\varepsilon}_i, \tag{4.29}$$

where $\boldsymbol{I}$ is an r_x by r_x identity matrix, $i = 1, 2, \ldots, n$.

However, as mentioned in the introduction, when using data size n of N variables as constraints, and the model wherein the number of coefficients is u and the number of disturbances is v, consequently a large number of consistency constraints is $N \times n$ and the number of normalization constraints is $(u+v) \times N$. This can cause problems of redundant constraints, over-fitting, heavy calculating volume, and is even difficult to converge when using numerical optimization to solve objective function. Furthermore, as GME-SEM is a full-estimation method that is supposed to depend on large-sample properties, this makes this situation more serious.

4.2.4 K-means Generalized Maximum Entropy Estimation for Structural Equation Modeling

To overcome issues of GME-SEM, data points are allocated into q groups, and then the means of these groups are used as consistency constraints. The purpose of this grouping is to gain clusters that simply compress data points and preserve the nature of data to assure the representativeness of constraints. This is the procedure of thinning the constraint grid. The consistent constrained of K-means GME-SEM is described as follows:

$$\bar{\boldsymbol{x}}^{(k)} = \boldsymbol{A}(\boldsymbol{I} - \boldsymbol{B})^{-1} \boldsymbol{\varphi}_k + \boldsymbol{\varepsilon}_k, \tag{4.30}$$

where $\bar{\boldsymbol{x}}^{(k)}$ is the centroid of cluster k, $(k = 1, 2, \ldots, q)$. We choose the cluster algorithm as K-means. Among cluster algorithms, K-means is still the most popular method because data points are partitioned into groups because of their inherent characteristics[4]. This algorithm is also supposed as a simplest clustering algorithm (Jain, 2010). Moreover, K-means creates tighter groups, and the sparse of the grid can be adjusted by choosing q. Therefore, in the context of this study, K-means clustering is suitable for thinning the constraint grid.

It is noteworthy that the number of consistency constraints now reduce to some $q \times N \ll n \times N$, leading to a decrease in the number of normalization constraints. Thus, $u + v \times q$. Each mean-value represents the characteristic of each group and, as such, the constraints of the new method still assure that all the core information is covered.

Here, we maximize a nonlinear objective function subject to linear constraint equations. The solution can be obtained by the theory of constrained optimization with the first order of Lagrange multipliers (Berger et al., 1996; Golan et al., 1994).

4.3 Evaluating the Performance of K-means GME-SEM

In this section, we evaluate the performance of K-means GME-SEM through two main tasks:

a. First, we verify whether the algorithm works effectively with a simulation. This simulation will also guide readers on how to apply K-means GME-SEM in model parameter estimation. It is also a comparison of the efficiency of the two algorithms.

b. Second, we show the practical significance of this new method when applied to real data. Key to our process is that K-means GME-SEM can be applied in some situations wherein the ML, the current most popular model estimation method for SEM, cannot be solved. We consider this the strongest contribution of “the new branch,” and it also enhances the importance of GME in SEM. However, to the best of our knowledge, prior studies have not recognized this importance in practice.

4.3.1 Simulation

We conduct a simulation to compare the performance of the proposed method with the original one. The data reflected in the model (Figure 4.7) were generated with $N = 150$ by `Simsem` package in R and are summarized in Table (4.2). The input for generating data includes factor loading matrix FL, factor covariance matrix FC, and $\sigma_{error} = 0.7$

[4]Nature of data.

(see Appendix C).

$$FL = \begin{pmatrix} 0.5 & 0 & 0 \\ 0.6 & 0 & 0 \\ 0.7 & 0 & 0 \\ 0 & 0.7 & 0 \\ 0 & 0.8 & 0 \\ 0 & 0.9 & 0 \\ 0 & 0 & 0.6 \\ 0 & 0 & 0.7 \\ 0 & 0 & 0.8 \end{pmatrix} \qquad FC = \begin{pmatrix} 0.8 & 0.5 & 0.6 \\ 0.5 & 0.9 & 0.7 \\ 0.6 & 0.7 & 0.7 \end{pmatrix}$$

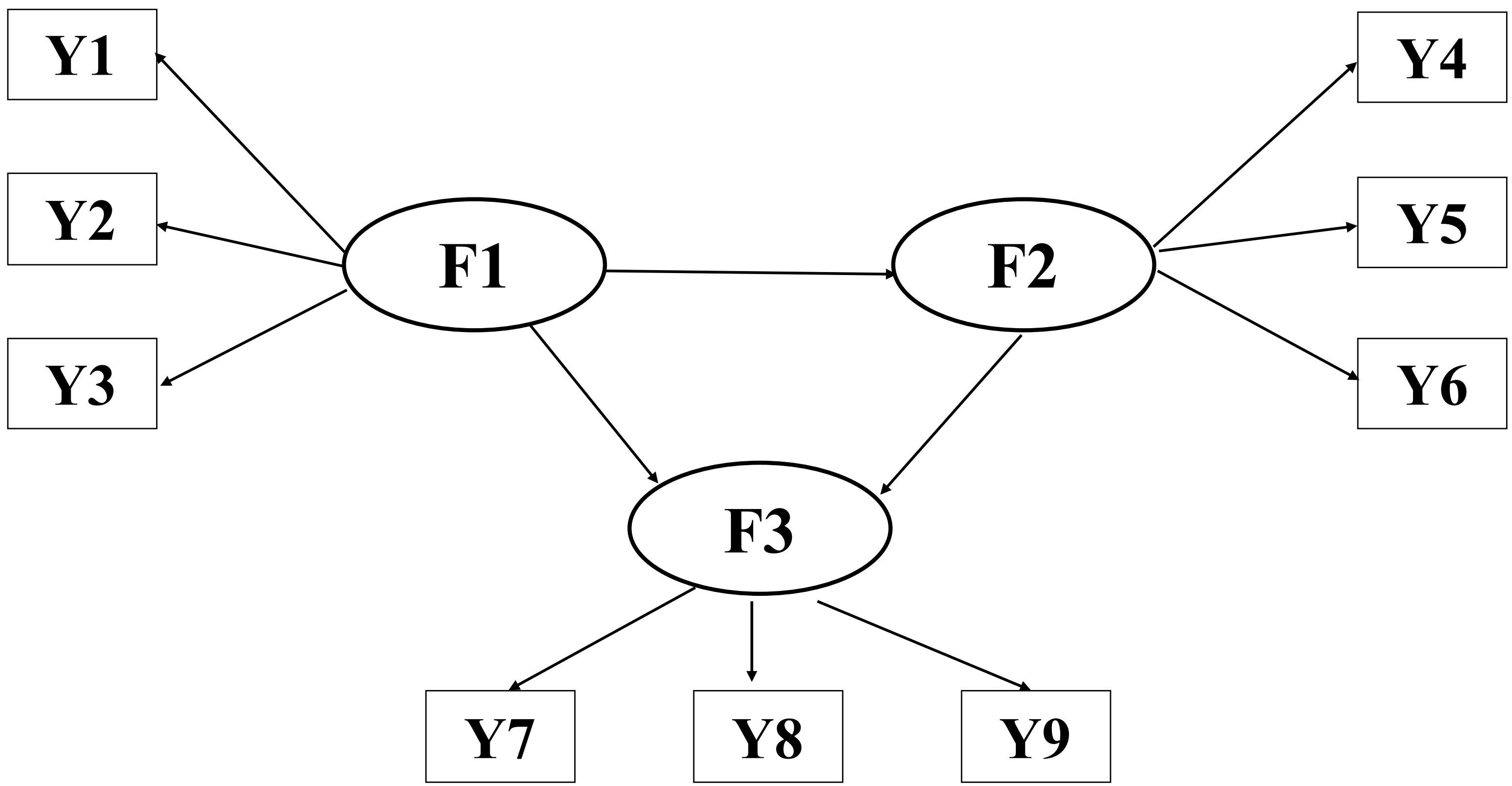


Figure 4.7: Simulated model

Table 4.2: Summary simulated data

	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7	Y_8	Y_9
Min.	−3.14473	−2.51634	−2.29591	−2.88968	−3.31046	−2.979841	−2.3686	−2.91442	−2.7230
1st Qu.	−1.01143	−0.64152	−0.64014	−0.88832	−0.94101	−0.871963	−0.9064	−0.60007	−0.8718
Median	0.01491	0.11814	0.03062	−0.17732	0.01892	−0.006795	−0.1100	−0.00741	−0.1189
Mean	−0.07445	0.05934	−0.07563	−0.09364	−0.08321	−0.056249	−0.0653	−0.04113	−0.1416
3rd Qu.	0.72455	0.74786	0.53325	0.84780	0.78896	0.635936	0.7478	0.57295	0.5637
Max.	2.77612	2.39873	2.03529	2.26036	2.26461	2.856819	2.9235	2.30902	2.3876

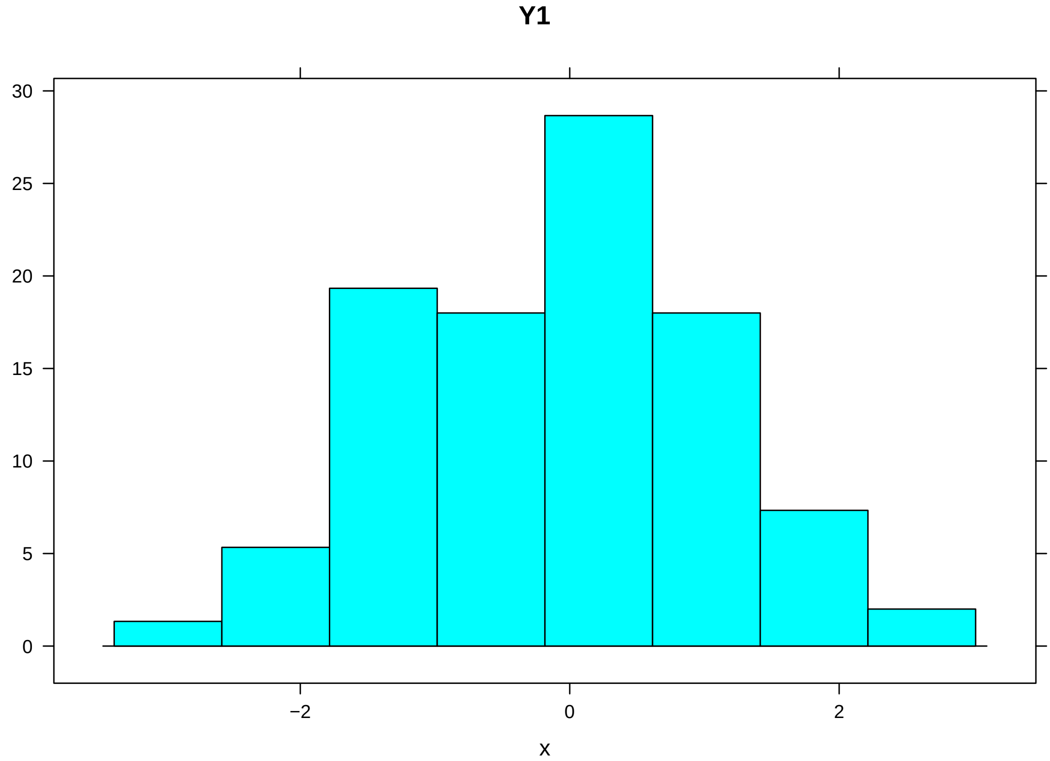


Figure 4.8: Histogram of simulated data: Y_1

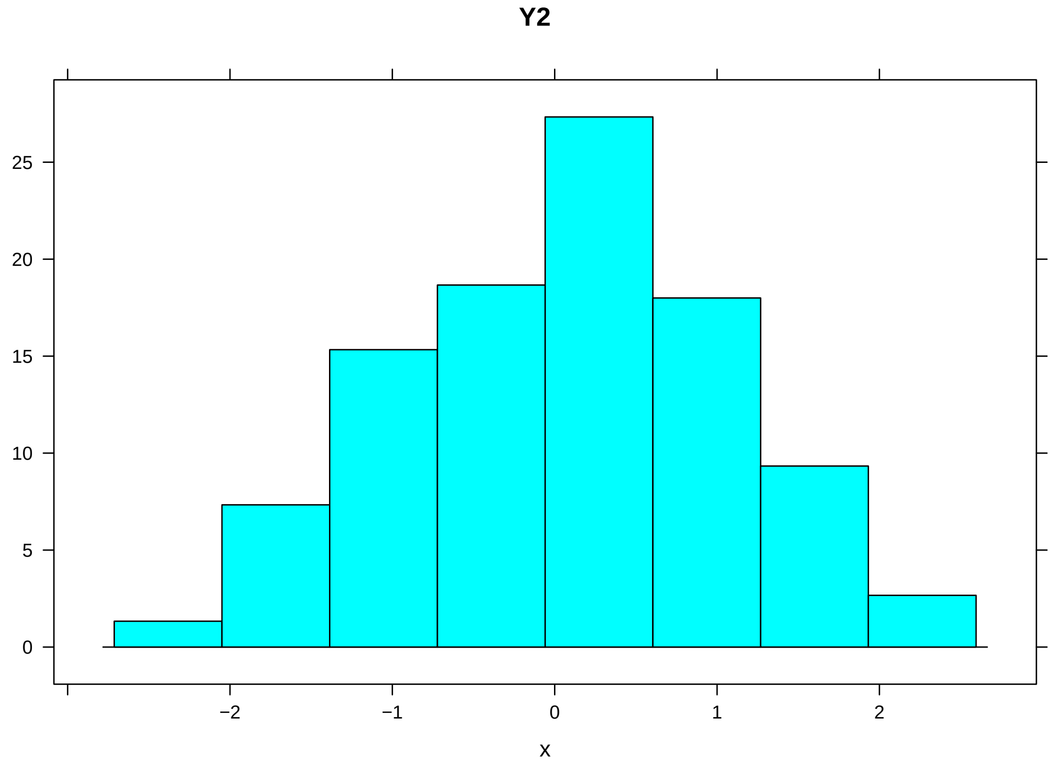


Figure 4.9: Histogram of simulated data: Y_2

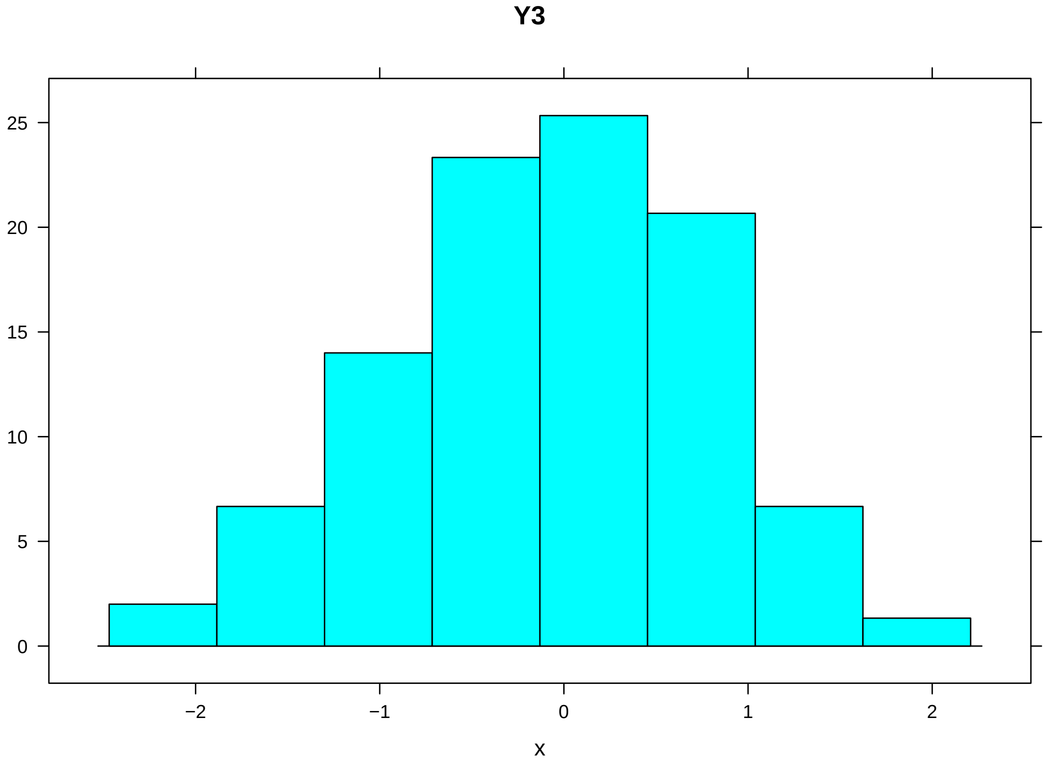


Figure 4.10: Histogram of simulated data: Y_3

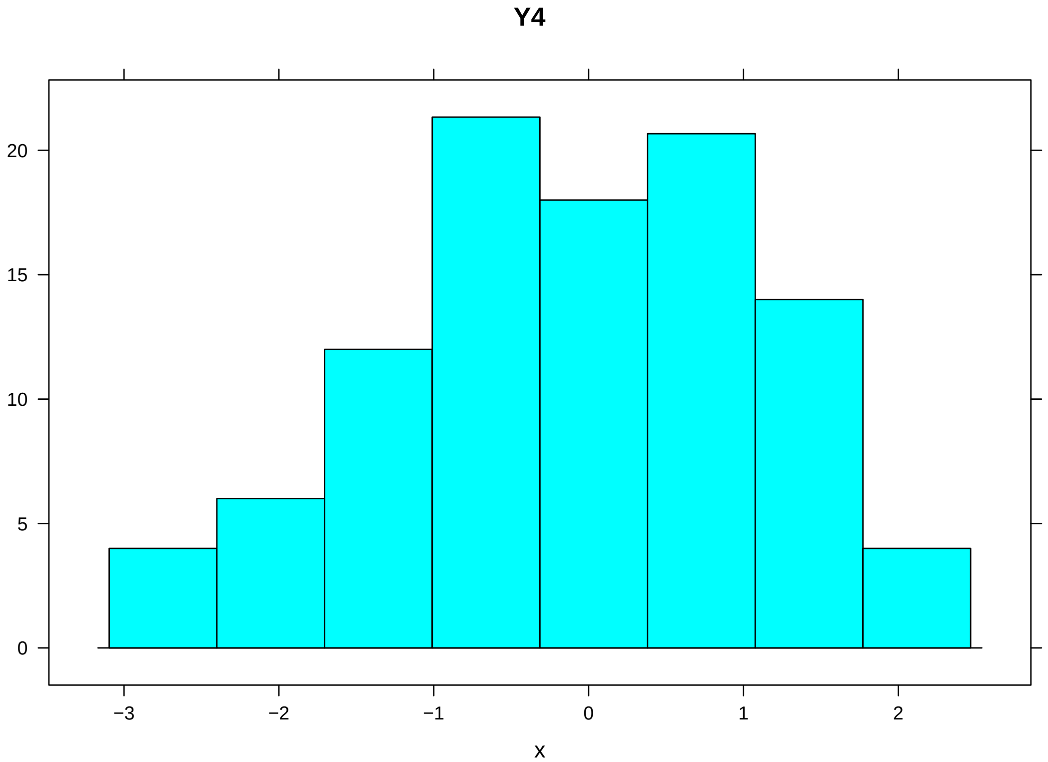


Figure 4.11: Histogram of simulated data: Y_4

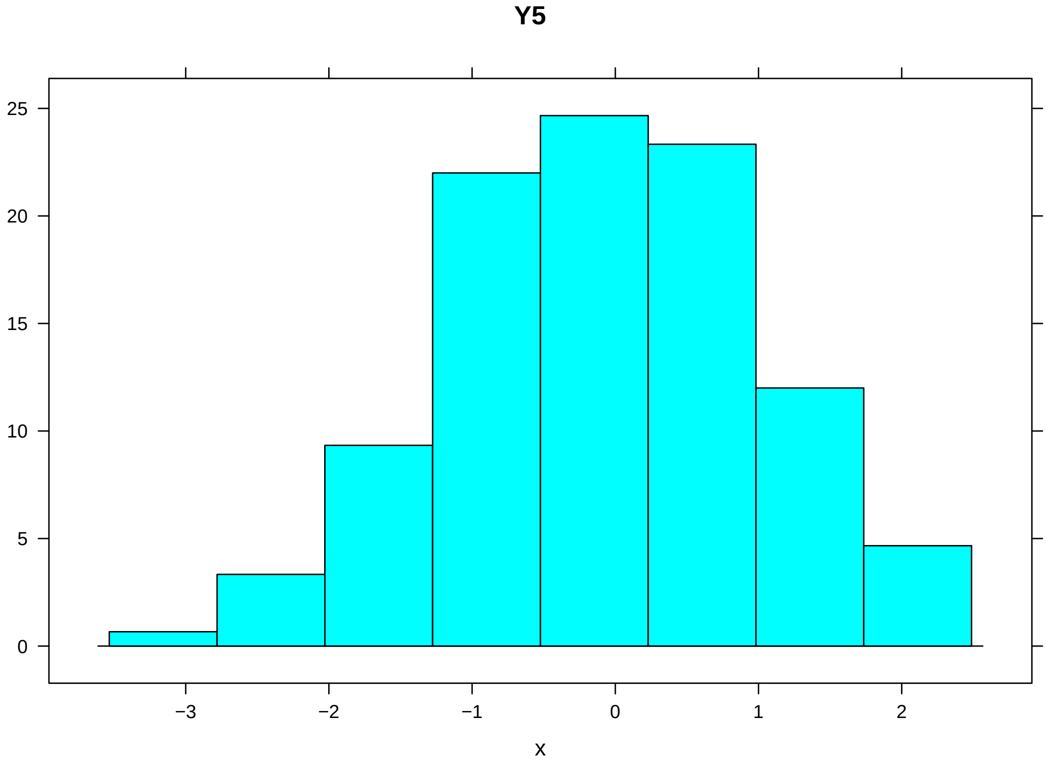


Figure 4.12: Histogram of simulated data: Y_5

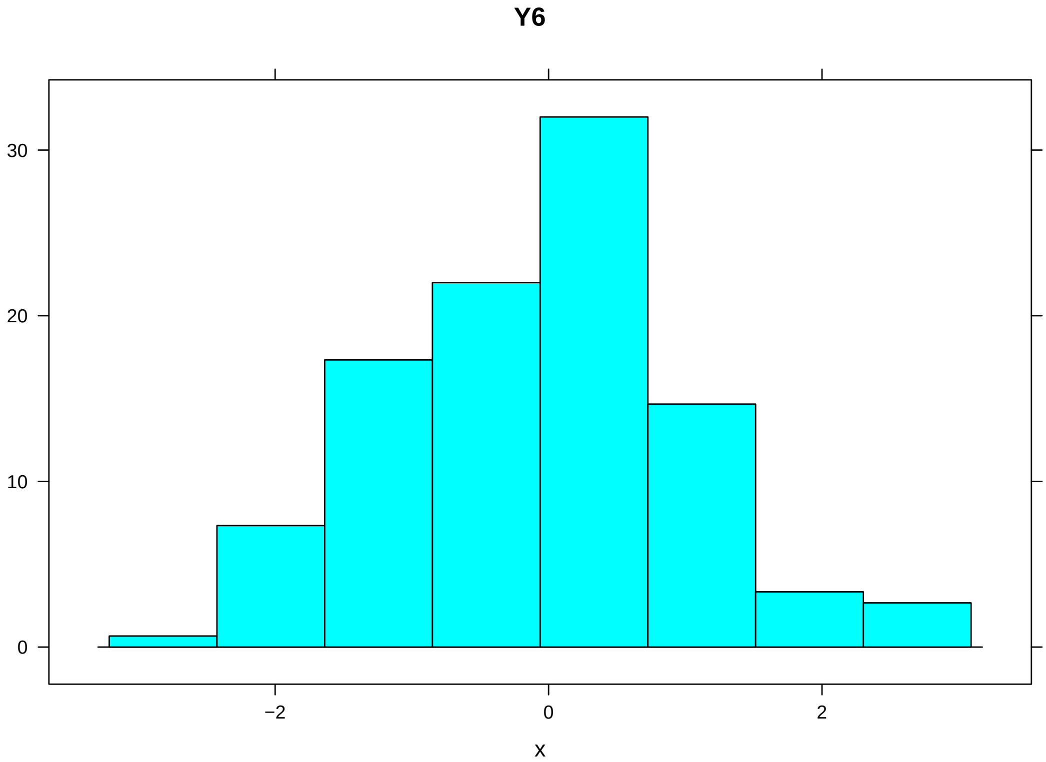


Figure 4.13: Histogram of simulated data: Y_6

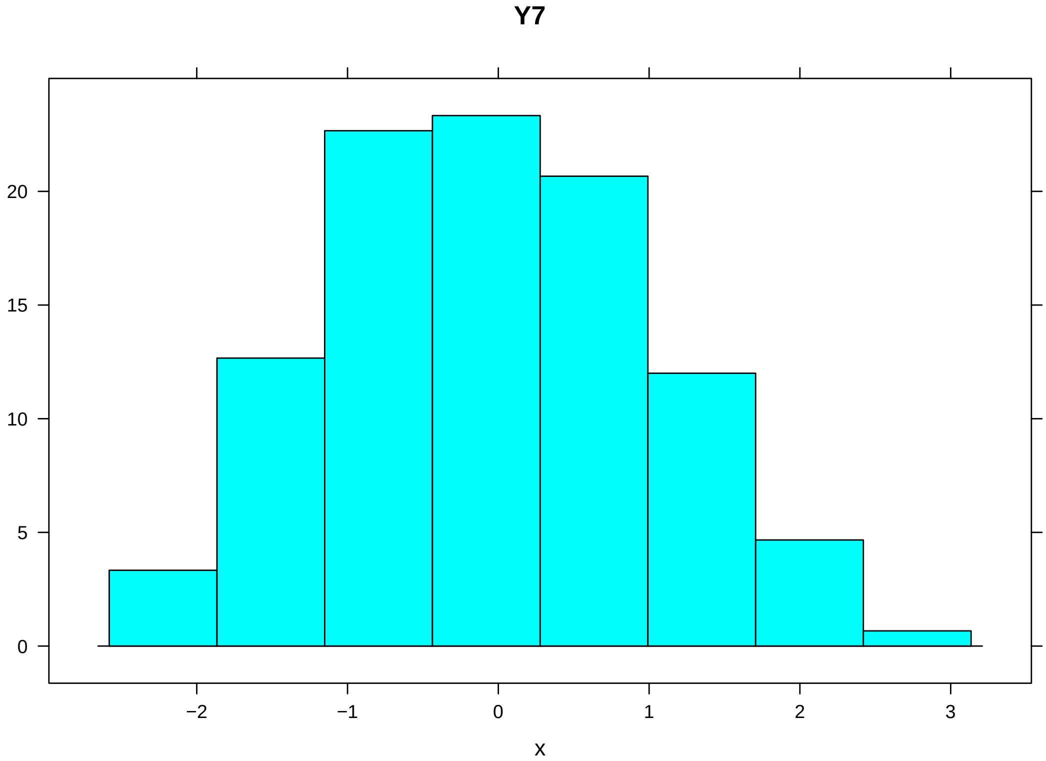


Figure 4.14: Histogram of simulated data: Y_7

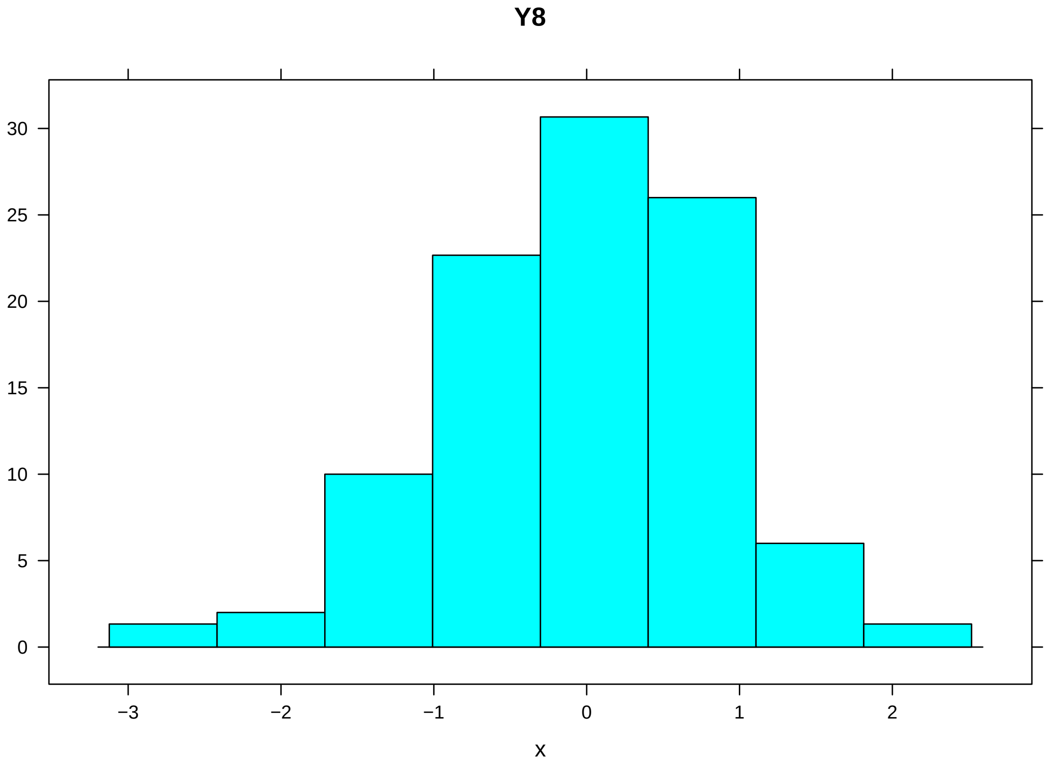


Figure 4.15: Histogram of simulated data: Y_8

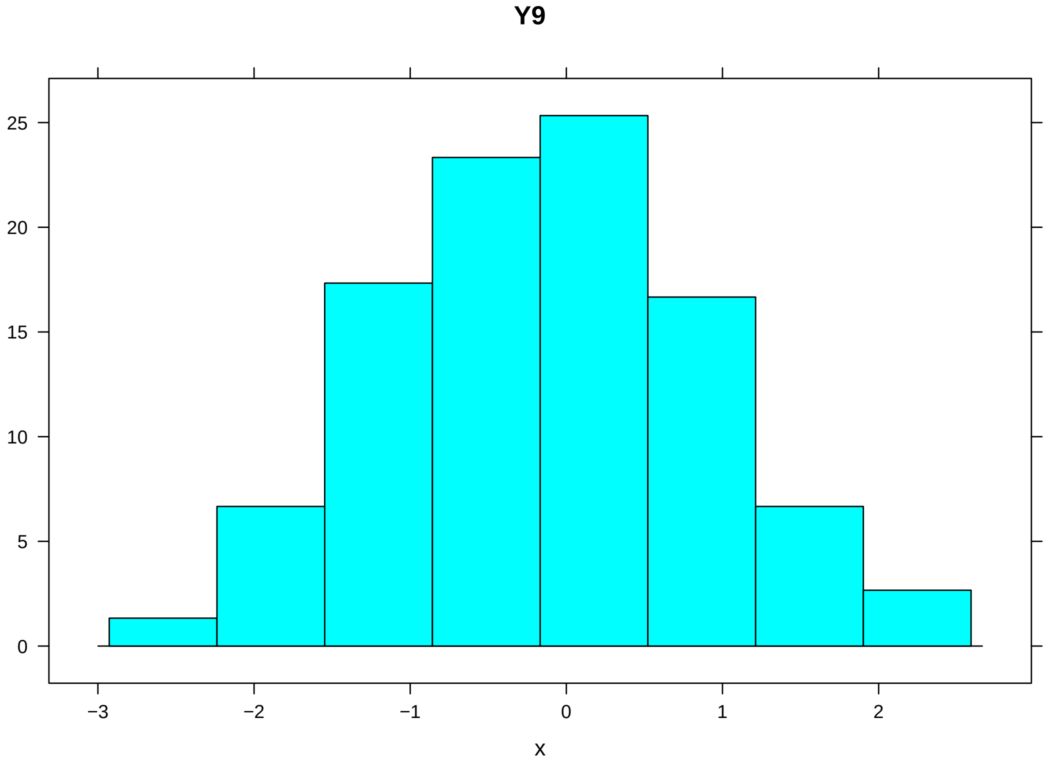


Figure 4.16: Histogram of simulated data: Y_9

The Solnp function in the Rsolnp package (Ghalanos and Theussl, 2012) is used to solve the optimization problem. Solnp is a non-linear optimization function using an augmented Lagrange method to solve the minimization problem in form of (4.31).

$$\min f(x). \tag{4.31}$$

$$\textit{subject to}$$

$$g(x) = 0.$$

$$LB_h \leqslant h(x) \leqslant UB_h.$$

$$LB_x \leqslant x \leqslant UB_x.$$

In the above, LB is the lower bound and UB is upper bound, and are optional.

Because our objective function needs to be maximized, a small adjustment is applied here: maximizing $f(x)$ is equivalent to minimizing $-f(x)$. Proof for this statement is below:

Given $x_0 \in X$ is an absolute maximum point of $f : X \to \mathbb{R}$ if $f(x_0) \geqslant f(x)$ $(\forall x \in X)$. Similarly, $x_0 \in X$ is an absolute minimum point of $-f : X \to \mathbb{R}$ if $-f(x_0) \leqslant -f(x)$ $(\forall x \in X)$.

Proof.

$$
\begin{aligned}
-f(x_0) &\leqslant -f(x) \\
f(x) - f(x) - f(x_0) &\leqslant f(x_0) - f(x_0) - f(x) \\
f(x) + (-f(x) - f(x_0)) &\leqslant f(x_0) + (-f(x_0) - f(x)).
\end{aligned} \tag{4.32}
$$

remove $-f(x_0) - f(x)$ from both sizes, then:

$$f(x) \leqslant f(x_0)$$

Done.

After this adjustment, our problem changes from maximizing function (4.28) to minimizing function (4.33), subject to the same constraints.

$$
\begin{aligned}
f(x) = -H(\boldsymbol{\pi}^{\mathrm{A}}, \boldsymbol{\pi}^{\mathrm{B}}, \boldsymbol{\pi}^{\varepsilon}, \boldsymbol{\pi}^{\varphi}) &= -H(\boldsymbol{\pi}^{\mathrm{A}}) - H(\boldsymbol{\pi}^{\mathrm{B}}) - H(\boldsymbol{\pi}^{\varepsilon}) - H(\boldsymbol{\pi}^{\varphi}) \\
&= \boldsymbol{\pi}^{\mathrm{A}\prime} \ln \boldsymbol{\pi}^{\mathrm{A}} + \boldsymbol{\pi}^{\mathrm{B}\prime} \ln \boldsymbol{\pi}^{\mathrm{B}} + \boldsymbol{\pi}^{\varepsilon\prime} \ln \boldsymbol{\pi}^{\varepsilon} + \boldsymbol{\pi}^{\varphi\prime} \ln \boldsymbol{\pi}^{\varphi}.
\end{aligned} \tag{4.33}
$$

$g(x)$ is a set of constraints and $LB_x = 0$ because $p_i \geqslant 0$.

Within the scope of this simulation, the control parameters are set at outer.iter $= 10$, inner.iter $= 800$, delta $= 1e - 13$, and tol $= 1e - 14$, while considering $q = 2$, $q = 6$, and $q = 10$.

The results are summarized in Table (4.3). Note that the optimization procedure of GME-SEM stopped after two iterations with a caution about the "Problem Inverting Problem." According to Gill and King (2004), non-invertible Hessians can be a signal of a nonsensical model. However, with the same design but different constraints, the models of $q = 2, 6, 10$ gave normal results. It can thus be understood that the problem was caused by constraints. Temporarily, as a limitation of this study, the results are compared without taking the warning into consideration.

The value pseudo-R^2 measures the goodness of fit, where $0 \leqslant R^2 \leqslant 1$, 1 means it is perfectly fitted and 0 means no informational value (Ciavolino and Al-Nasser, 2009; Soofi, 1992). The R^2 value of GME-SEM near 1 is a sign of over-fitting; in the model with $q = 2$, this value quite low. The model $q = 6$ and $q = 10$ give acceptable goodness of fit values, with a balanced number of constraints and calculation volume.

In practice, good calculation speed for SEM is required because researchers often have many candidate models for SEM. For this reason, within the scope of this simulation, the K-means GME-SEM with 6 and 10 clusters shows dominant results compared with just

Table 4.3: Summary of the simulation results

	K-means GME-SEM			GME-SEM
Size	$q = 2$	$q = 6$	$q = 10$	$N = 150^{*}$
No. CC	18	54	90	$1,350$
No. NC	36	84	132	$1,812$
$S(\tilde{\pi})$	0.666	0.222	0.133	0.007
Pseudo-R^2	0.333	0.778	0.866	0.992
Time	7.716 (s)	51.810 (s)	5.615 (m)	1.554 (h)
Convergence	0	0	0	2

* Problem inverting Hessian

No. CC: Number of consistency constraints

No. NC: Number of normalization constraints

$S(\tilde{p})$: normalized entropy

Pseudo-$R^2 = 1 - S(\tilde{\pi})$

s: seconds; m: minutes; h: hours

0: converged ; 1: not converged; 2: not converged with warnings

the GME-SEM.

Next, we compare the proposed method with SEM based on ML by parameter recoveries. The generating model is the same as in Fig 4.7. The setting of Rsolnp is also the same as in the first simulation. We use `sem` package in R for the parameter estimation of SEM based on ML. The evaluation criterion is the mean square error (MSE) of matrices of coefficients $\boldsymbol{A}$ and $\boldsymbol{B}$. The number of repetitions is 20. The number of objects N is 150 and 300.

Table 4.4: MSE of $\boldsymbol{A}$

N	K-means GNE-SEM			ML
	q=2	q= 6	q=10	
150	3.239	3.789	2.103	2.516
300	3.235	2.097	3.325	2.549

Table 4.5: MSE of $\boldsymbol{B}$

N	K-means GNE-SEM			ML
	q=2	q= 6	q=10	
150	3.240	3.919	4.220	1.073
300	3.240	3.479	3.782	1.159

Tables 4.4 and 4.5 show the MSE of $\boldsymbol{A}$ and $\boldsymbol{B}$ by each method, respectively. When N is 150, in the proposed method, q= 10 is the best result among the MSEs of $\boldsymbol{A}$. When N is 300, in the proposed method, q= 6 is the best result among the MSEs of $\boldsymbol{A}$. Therefore, the proposed method recovers measurement model $\boldsymbol{A}$ when we select a suitable number of q. However, the proposed method is the estimation of the parameter of the latent structure $\boldsymbol{B}$. This property may be caused by a constraint of the proposed method. The proposed model is constrained by only means. Hence, the constrained disturbance vector $\boldsymbol{\varphi}$ is not sufficient as the disturbance vector $\boldsymbol{\varepsilon}$ is. The results of the simulation show that, with large sample size, in most cases, the MSEs of the ML estimation method are still better than that of K-means GME-SEM. For a comprehensive understanding of the proposed method, a large number of simulations are required for evaluating the MSE of the covariance matrix and for checking the convergence and the effect of misspecification. However, the proposed method still has a role to play (as will be discussed in the next section) in the application field.

4.3.2 A Real Data Example

To describe some characteristics of K-means GME-SEM in practice, an example using real data [5] from empirical research in the construction sector in the field of supply chain management is given. The model included 13 observations abbreviated as TMS1, TMS2, TMS3, and TMS4 to measure for the latent variable TMS; HRM5, HRM6, HRM7, HRM8, HRM9, and HRM10 to measure for latent variable HRM; and QD11, QD12, and QD13 to measure for latent variable QD on a seven-point Likert scale. Furthermore, the structure of the model includes three coefficients that reflect the effect TMS has on QD and HRM, as seen in Figure 4.17.

In the original study, based on the theory, the authors proposed two hypotheses that need to be tested:

* H1: TMS positively affects HRM.

* H2: TMS positively affects QD.

However, in this study, we must discover one more relationship, that is, the relationship of TMS on QD.

[5] The data are extracted from part of the study "Supply chain management and organizational performance: the resonant influence," published by Emerald

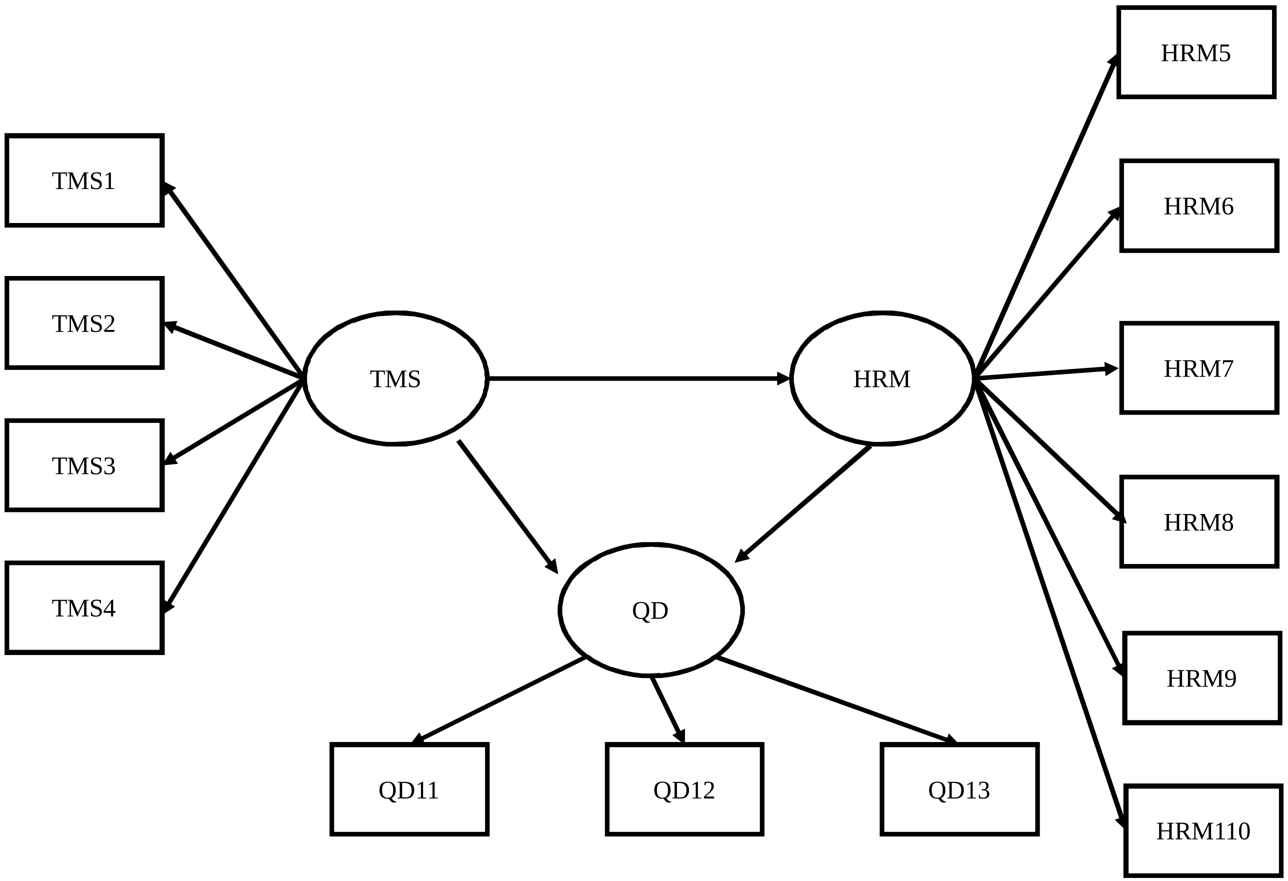


Figure 4.17: Model for real data

For the CB-SEM with the ML estimation method, the recommended minimum sample size should be from 100 to 150 (Hair et al., 1998), or if more rigid, it should be 200 (Hoelter, 1983). In a complicated model with many observed variables, the number of data points could be higher in order to assure the degree of freedom is non-negative. However, in this study, the subjects in the design who are tasked with answering the questionnaire are managers from enterprises. This is a difficult population to reach, and only 50 valid answers were collected. The data gained also does not have normal distributions of each variable (see Figures from 4.18 to 4.30, and Table 4.6). With this sample size and distribution, applying the available CB-SEM estimation methods is not possible. In this situation, K-means GME-SEM is an alternative solution.

Table 4.6: Summary real data

	TMS1	TMS2	TMS3	TMS4	HRM5	HRM6	HRM7	HRM8	HRM9	HRM10	QD11	QD12	QD13
Min.	1.00	1.00	1.00	2.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1st Qu.	3.00	4.00	3.25	4.00	4.00	4.00	4.00	3.00	4.00	4.00	5.00	4.25	5.00
Median	5.00	4.50	5.00	4.00	5.00	5.00	5.00	5.00	5.00	5.00	6.00	6.00	6.00
Mean	4.74	4.72	4.70	4.62	4.90	4.76	4.82	4.76	4.80	5.08	5.52	5.26	5.16
3rd Qu.	6.00	6.00	5.75	5.00	6.00	5.75	6.00	6.00	6.00	6.00	7.00	7.00	7.00
Max.	7.00	7.00	7.00	7.00	7.00	7.00	7.00	7.00	7.00	7.00	7.00	7.00	7.00

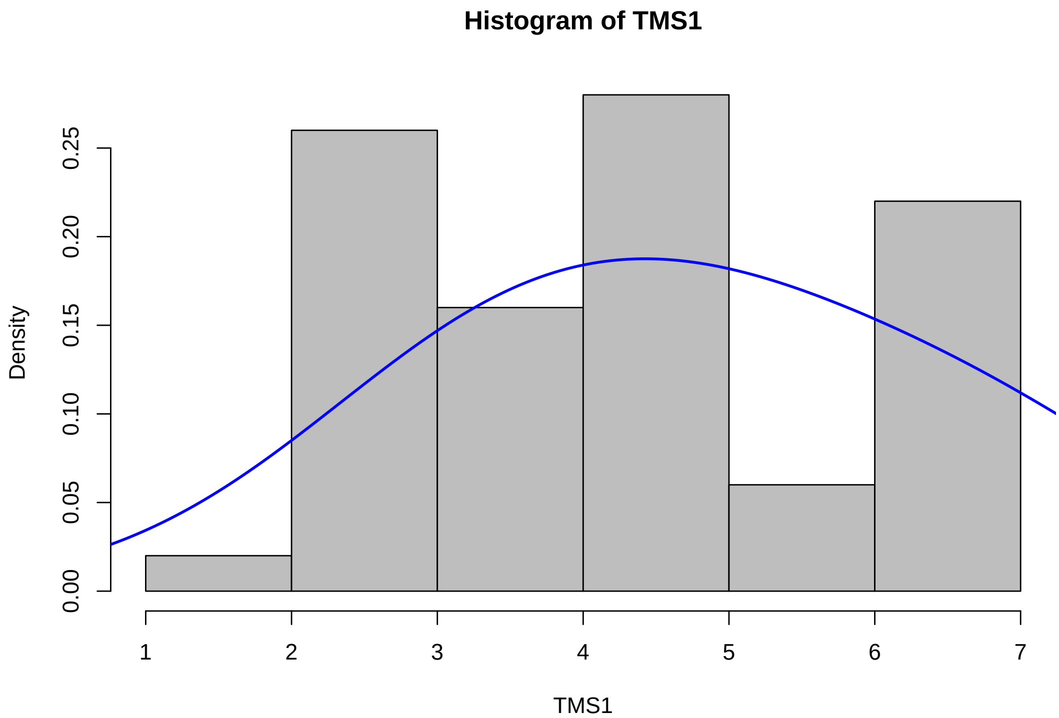


Figure 4.18: Distribution of real data: variable TMS1

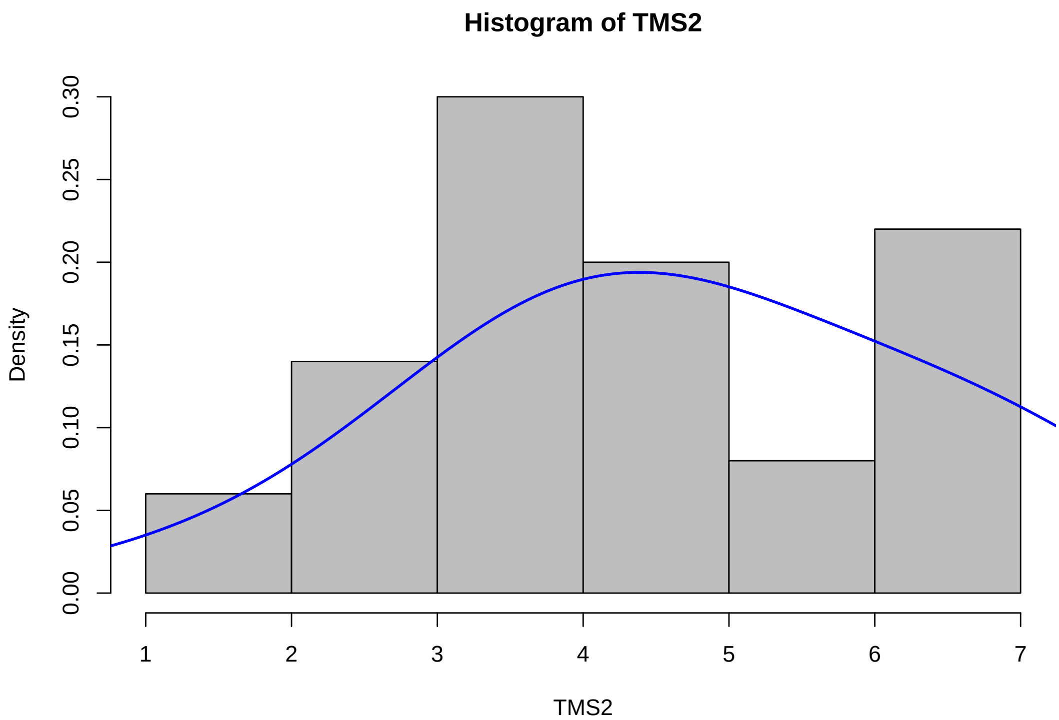


Figure 4.19: Distribution of real data: variable TMS2

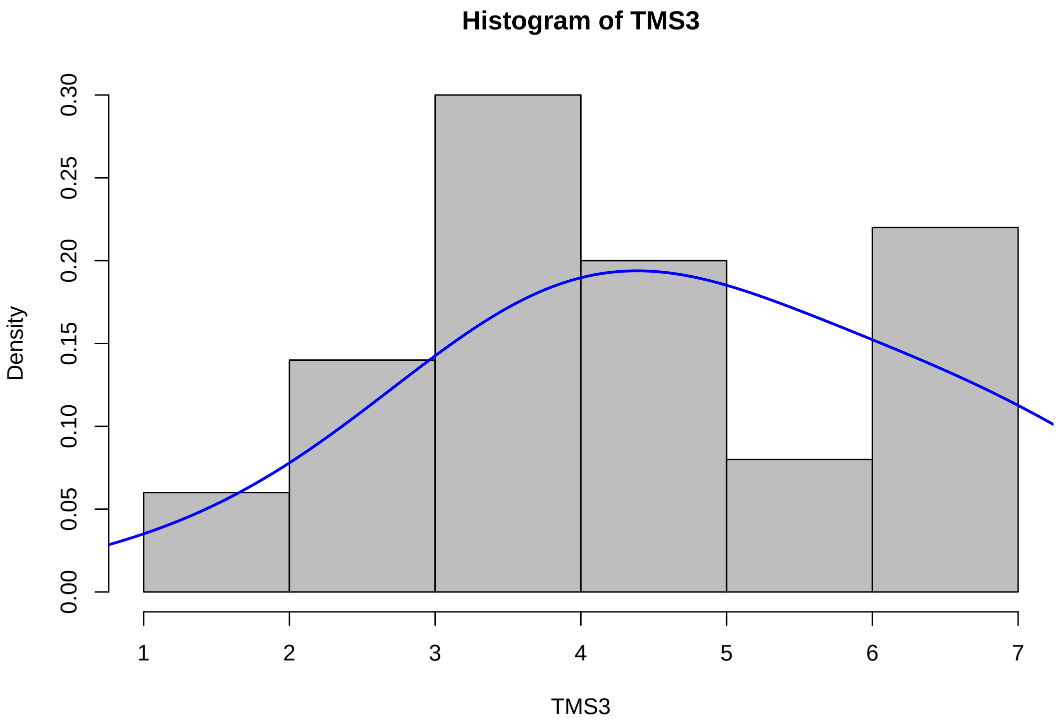


Figure 4.20: Distribution of real data: variable TMS3

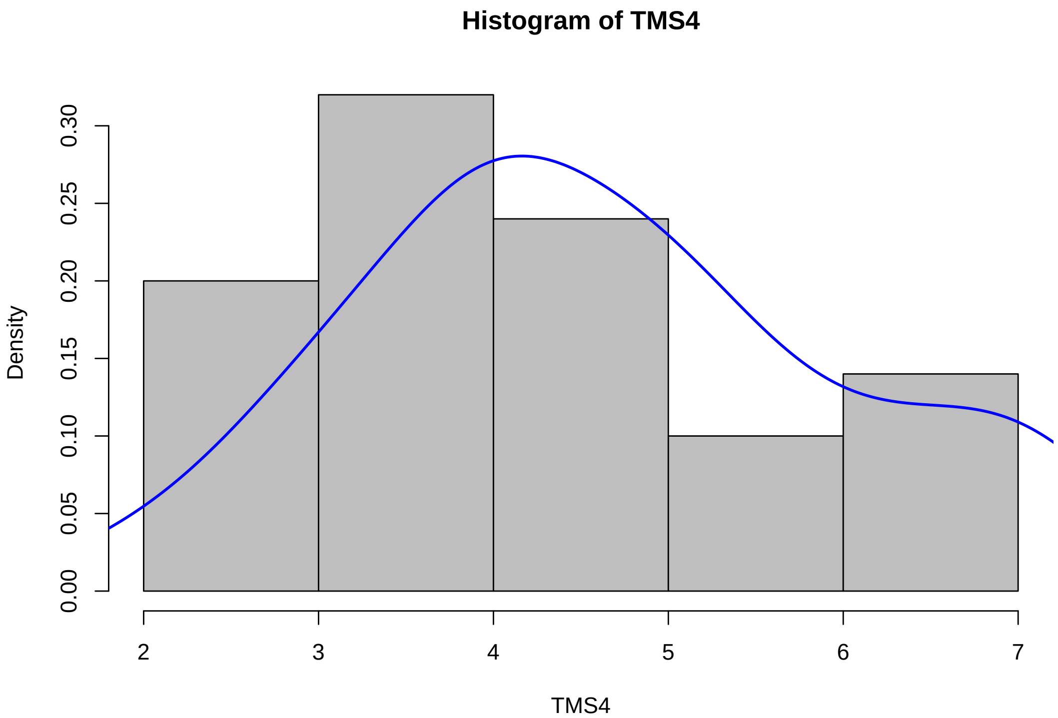


Figure 4.21: Distribution of real data: variable TMS4

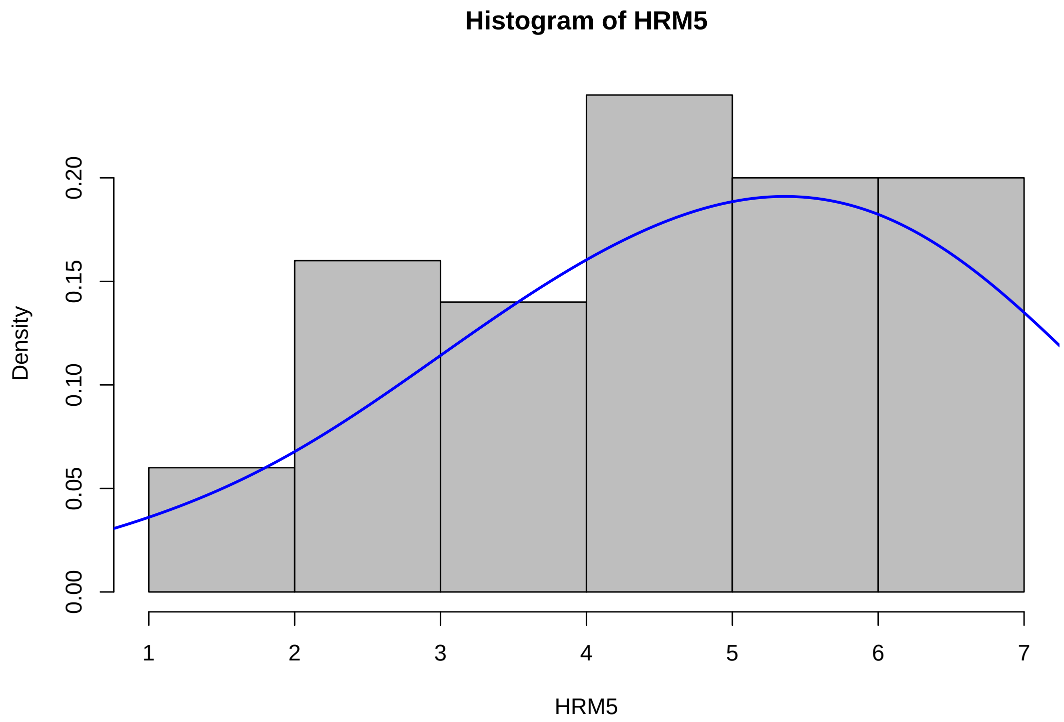


Figure 4.22: Distribution of real data: variable HRM5

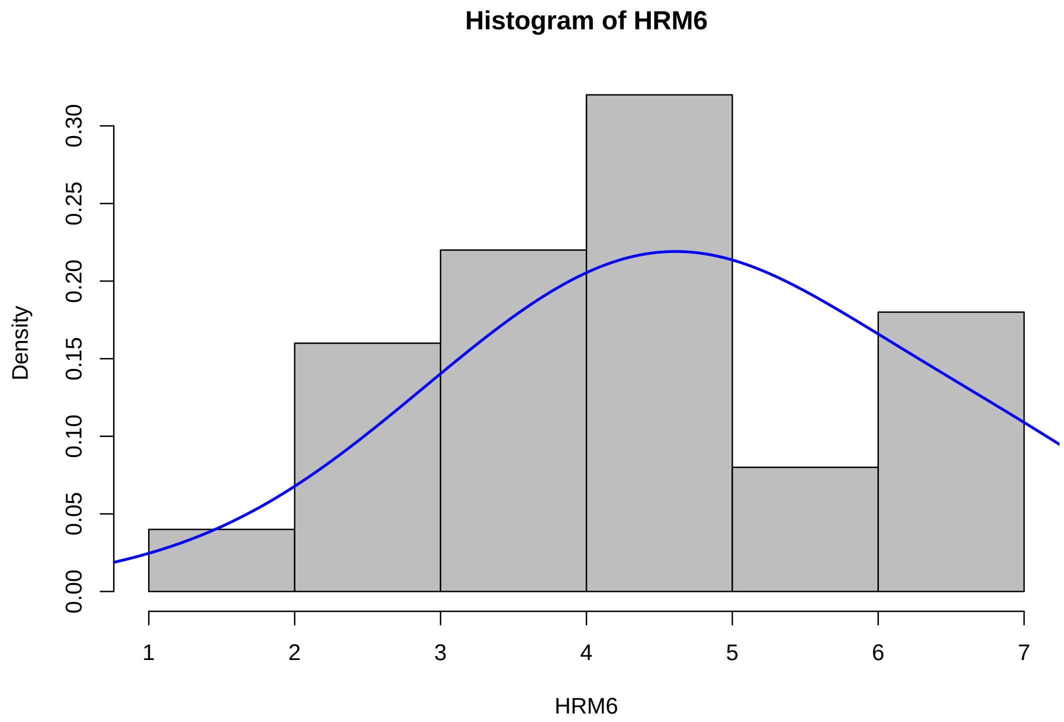


Figure 4.23: Distribution of real data: variable HRM6

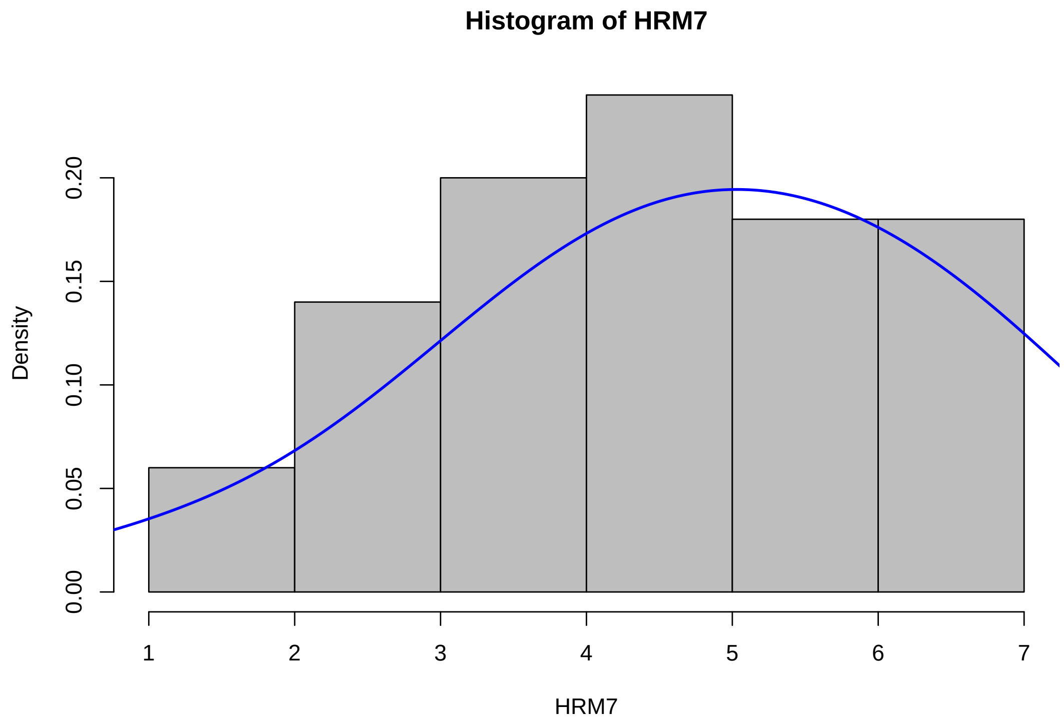


Figure 4.24: Distribution of real data: variable HRM7

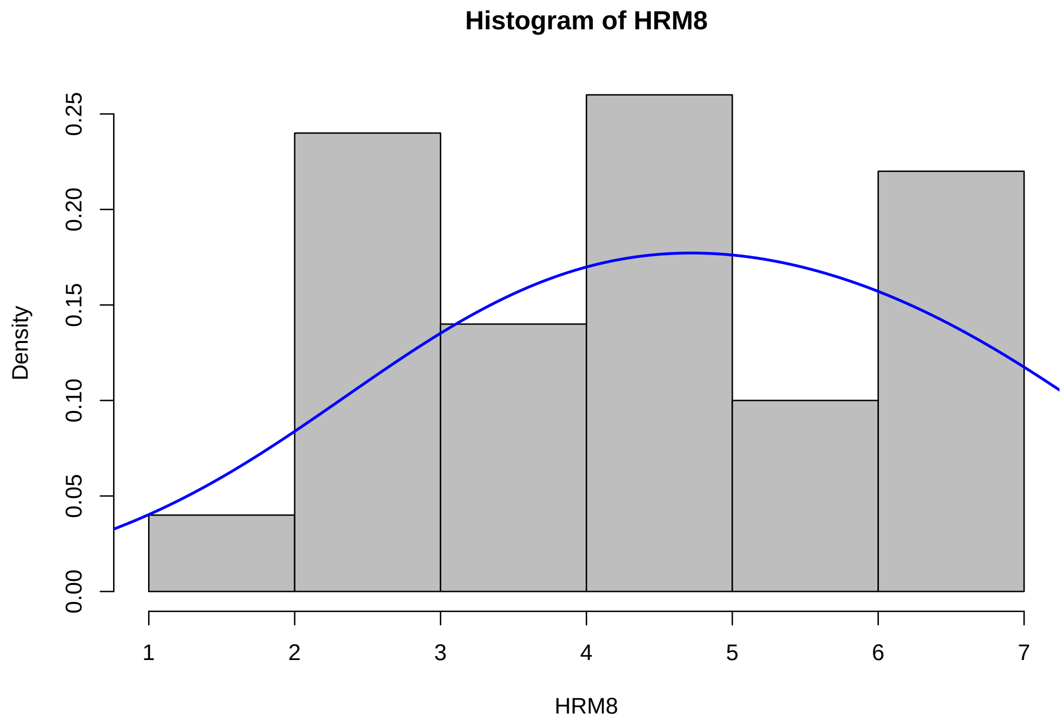


Figure 4.25: Distribution of real data: variable HRM8

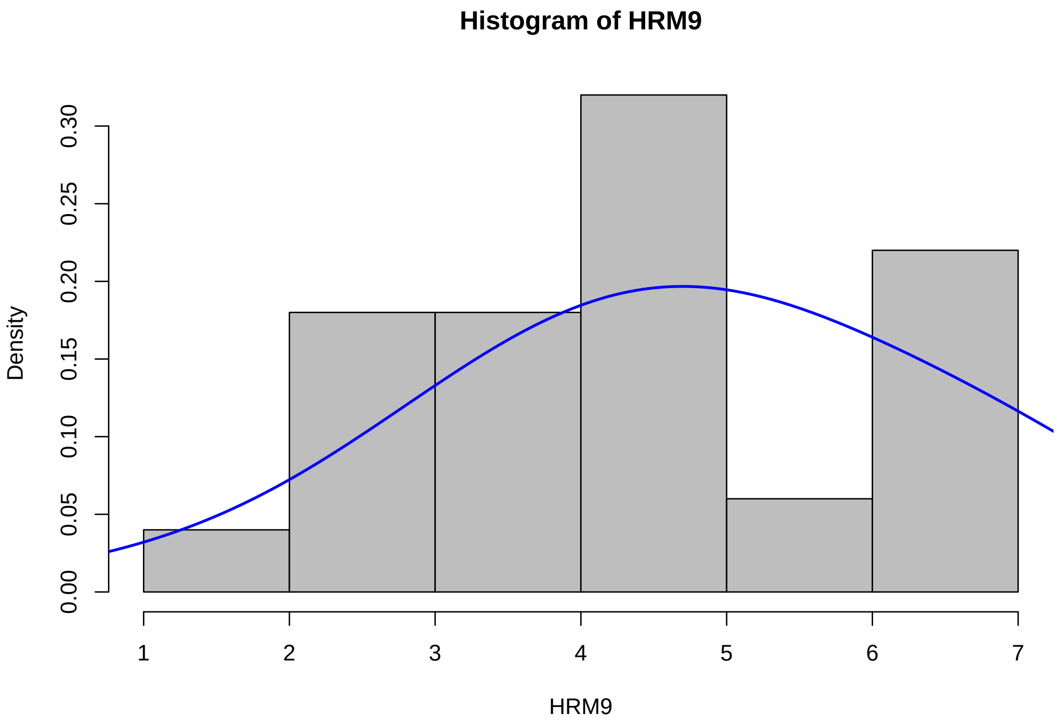


Figure 4.26: Distribution of real data: variable HRM9

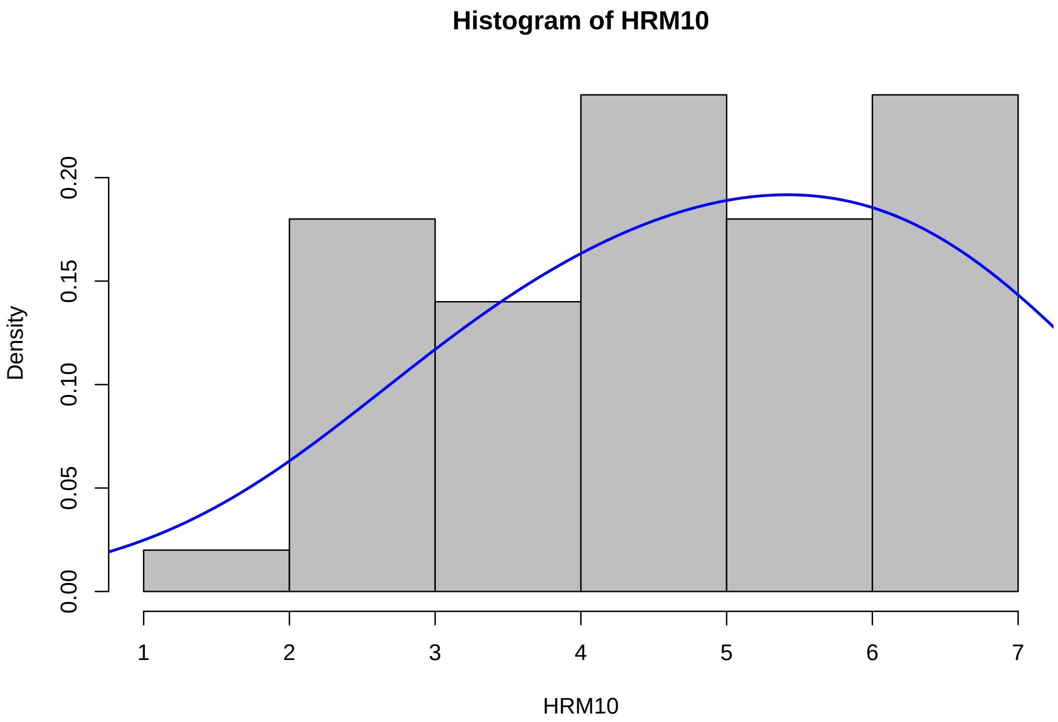


Figure 4.27: Distribution of real data: variable HRM10

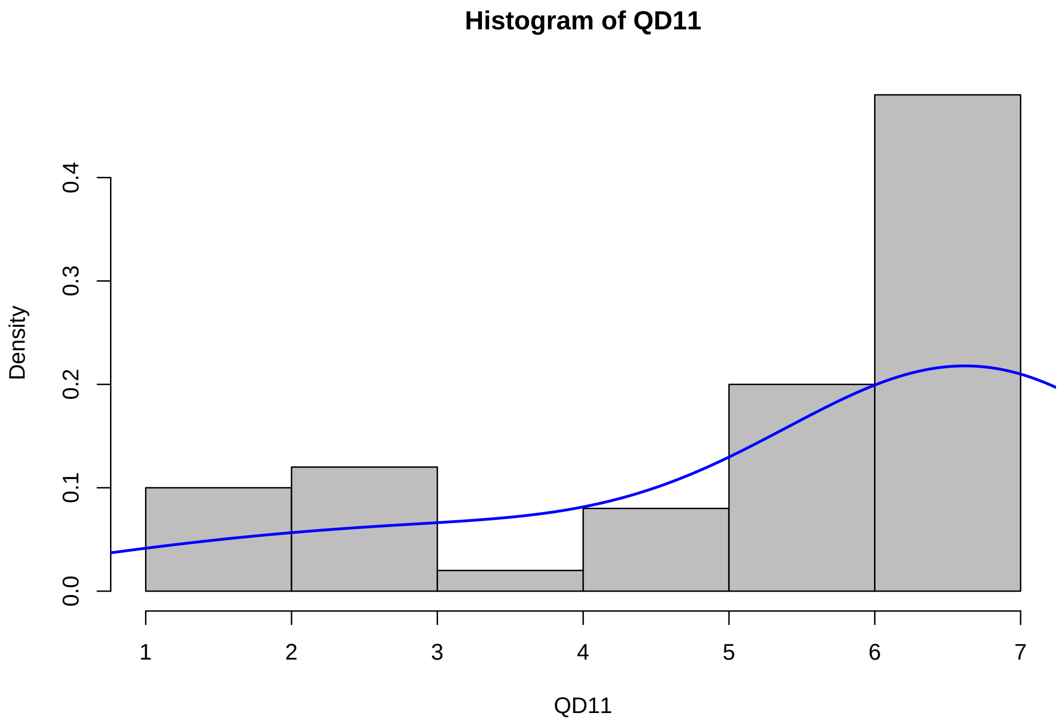


Figure 4.28: Distribution of real data: variable QD11

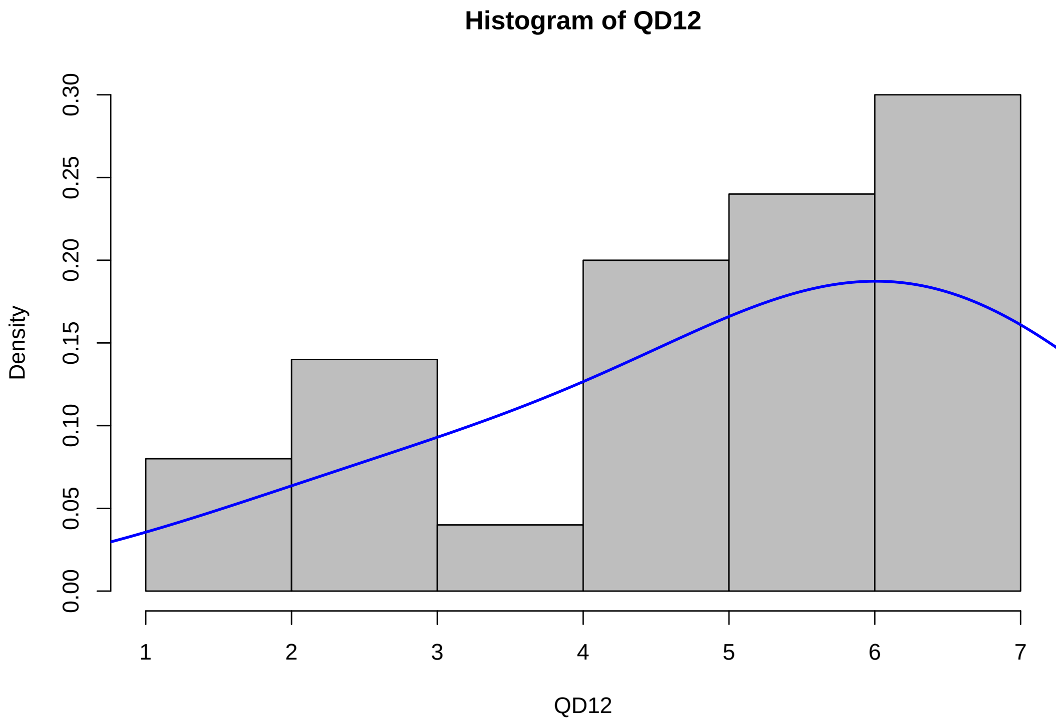


Figure 4.29: Distribution of real data: variable QD12

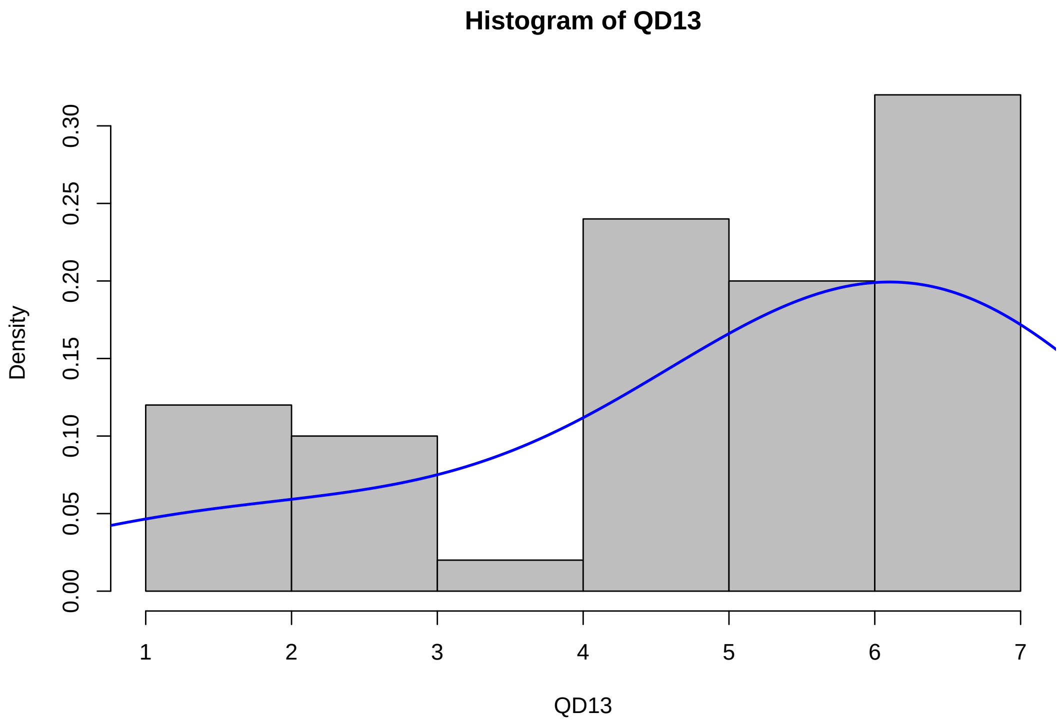


Figure 4.30: Distribution of real data: variable QD13

Applying K-means GME-SEM starts with deciding the constraint grid. The value of q should be considered by combining the optimal number of clusters from the Elbow, Silhouette, and Gap statistic methods. Figures 4.31, 4.32, and 4.33 depicting $q = 2, 3, 4$ are set as the three cases of the constraint grid.

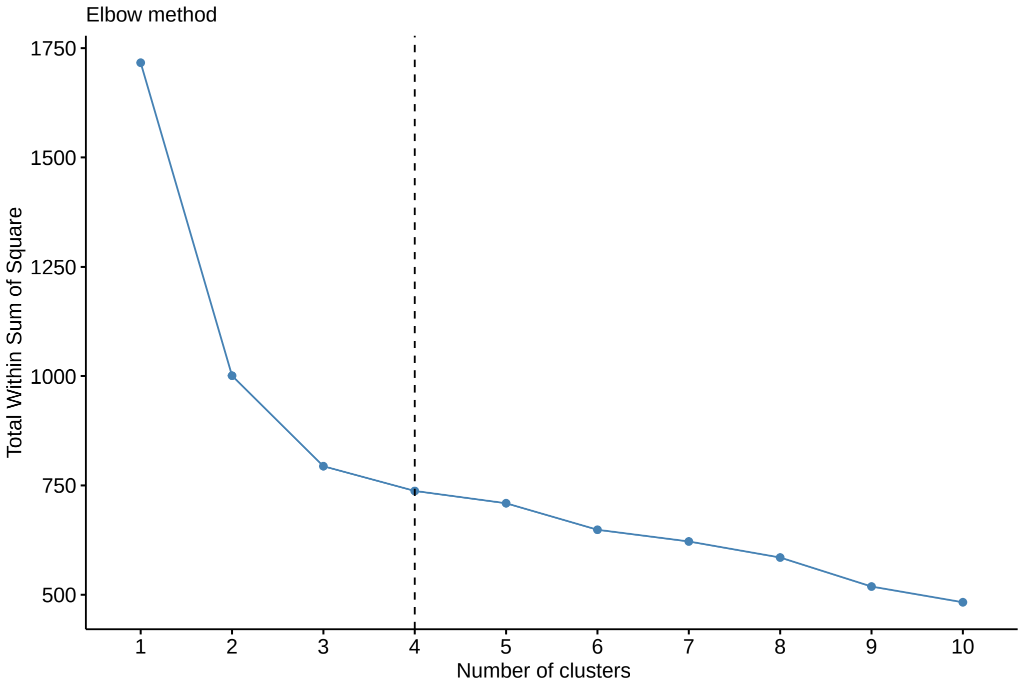


Figure 4.31: Optimal number of constraints: Elbow

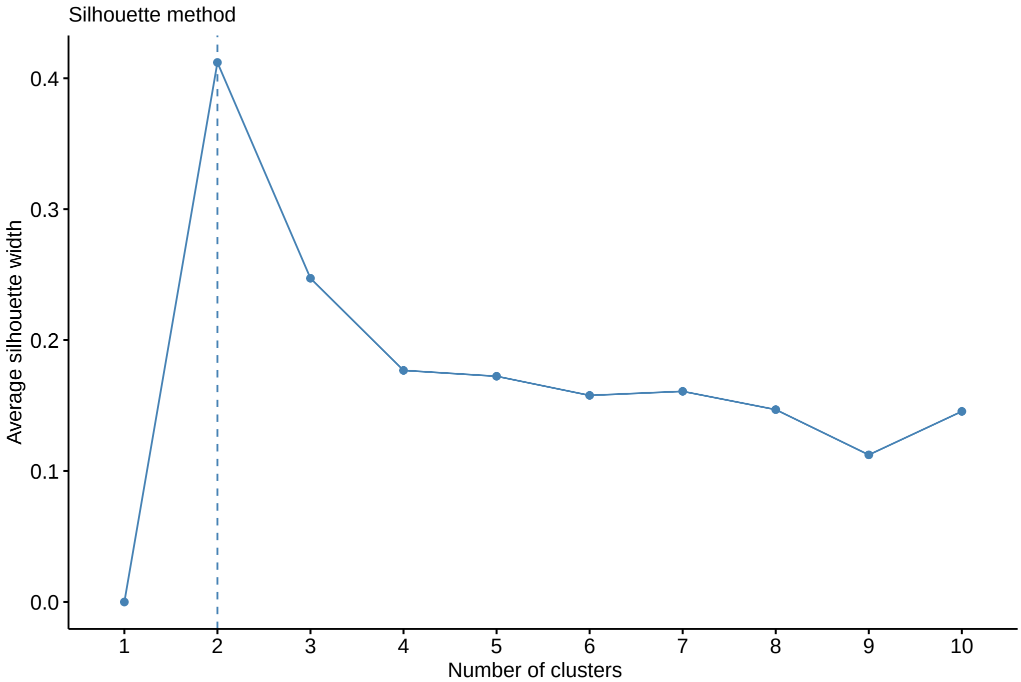


Figure 4.32: Optimal number of constraints: Silhouette

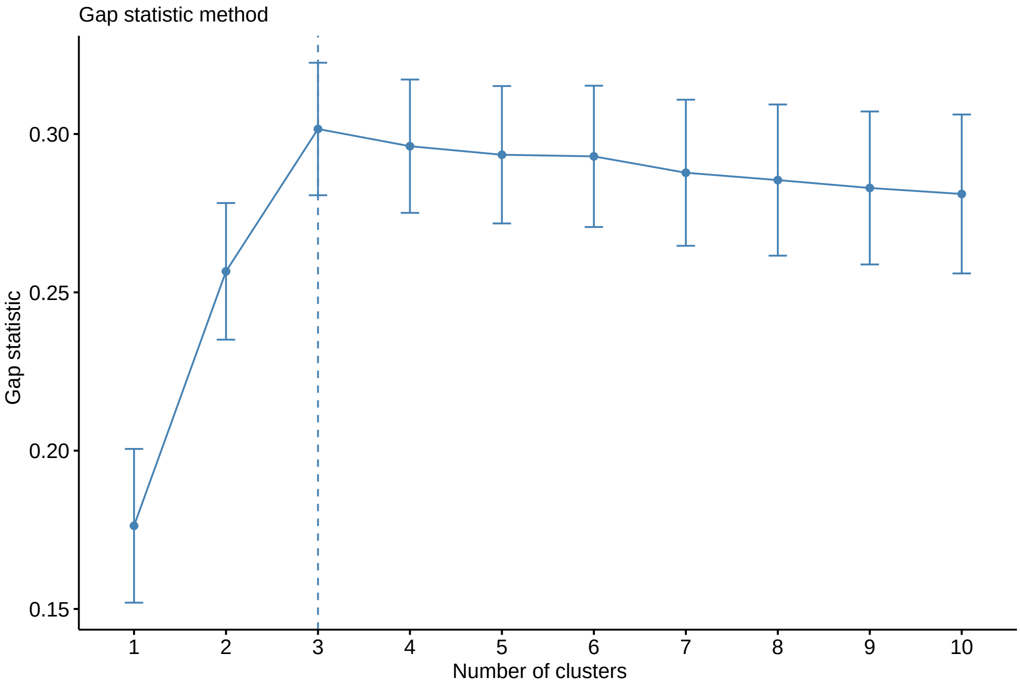


Figure 4.33: Optimal number of constraints: Gap

According to the viewpoint of the author, the dimension of spanned space for parameters should be equal, that is, $K = L$, to maintain the uniformity in measurement system. In this example, $K = L = 5$, as the guideline of Golan et al. (1996), and $z = (-1, -0.5, 0, 0.5, 1)$, $v = (-3, -1.5, 0, 1.5, 3)$, and control parameters outer.iter=10, inner.iter = 800, delta = 1e-13, tol = 1e-13.

Table 4.7: Estimation results

			$q = 2$	$q = 3$	$q = 4$
TMS1	<−	TMS	0.667	−0.795	0.445
TMS2	<−	TMS	0.668	−0.798	0.446
TMS3	<−	TMS	0.661	−0.793	0.447
TMS4	<−	TMS	0.657	−0.790	0.459
HRM5	<−	HRM	0.543	−0.700	0.574
HRM6	<−	HRM	0.534	−0.690	0.565
HRM7	<−	HRM	0.530	−0.694	0.570
HRM8	<−	HRM	0.521	−0.679	0.566
HRM9	<−	HRM	0.531	−0.692	0.567
HRM10	<−	HRM	0.564	−0.729	0.576
QD11	<−	QD	0.575	0.718	0.580
QD12	<−	QD	0.497	0.686	0.554
QD13	<−	QD	0.476	0.688	0.560
HRM	<−	TMS	0.889	0.976	0.587
QD	<−	TMS	0.558	−0.456	0.506
QD	<−	HRM	0.508	−0.791	0.608
$S(\tilde{p})$			0.462	0.247	0.136
Pseudo-R^2			0.538	0.753	0.864

As the results in Table (4.7) show, $q = 4$ has the best goodness of fit value. Besides, there is a tendency that the goodness of fit increases when q increases (see Figure (4.34)). If this value is too high, the result could imply over fitting. As with the simulations, and the application on real data, q in the range of 4 to 6 usually implies acceptable goodness of fit for the model.

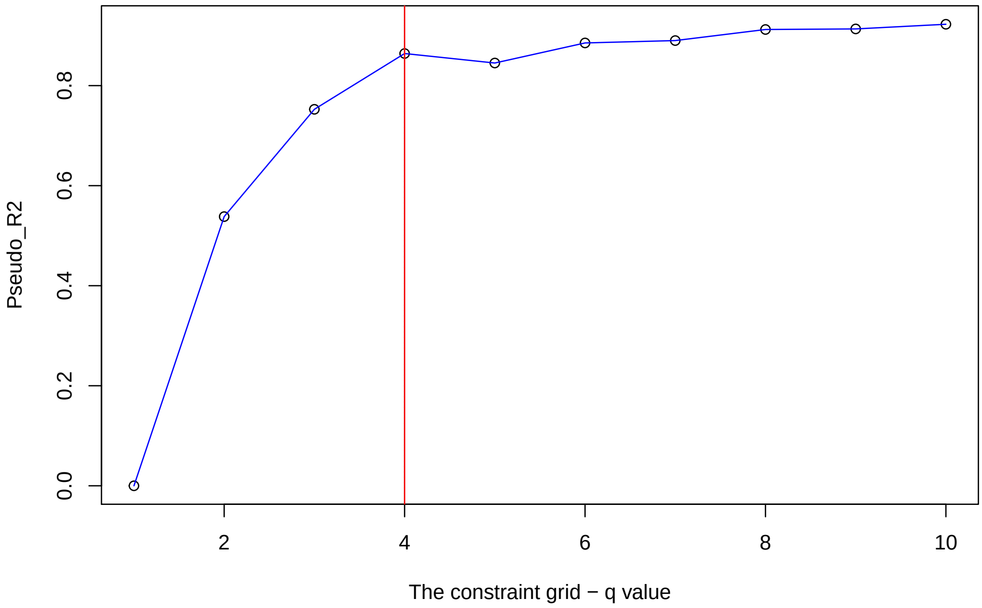


Figure 4.34: Goodness of fit when q increase

In addition, using the optimal case with $q = 4$ as benchmarking, the three-dimension spanned space for parameters, that is, $K = L = 3$, is also tested to consider the existence of any differences between the number of support points. The estimated coefficients are 0.5678220, 0.5576318, 0.5664214, 0.5631794, 0.5675743, 0.5535214, 0.5586630, 0.5466499, 0.5491629, 0.5569383, 0.5060573, 0.5678182, 0.5552965, 0.5676999, 0.5307815, 0.5676955, and $S(\tilde{p}) = 0.1075043$, with pseudo-$R^2 = 0.8924957$. With this result, there are almost no large differences between $K = L = 3$ and $K = L = 5$. These results consolidate the results of previous studies, that is, that the number of support points has no effect on the recovered parameters.

In the published study, ML was used to estimate the parameters. To be able to use the ML estimation method, this research team had to distribute 2,147 questionnaires; with a response rate of 11.5%, only 246 usable questionnaires were collected. Whereas with K-means GME-SEM, only 50 valid questionnaires are needed, that is, only 20–25% of the effort is needed.

The results of the original study are as follows:

*H1: TMS positively affects HRM at 0.638.

*H2: HRM positively affects QD at 0.612.

In this study, with K-means GME-SEM, the results are as follows:

*H1: TMS positively affects HRM at 0.587.

*H2: HRM positively affects QD at 0.608.

The two results are almost equivalent. Importantly, we discover that TMS has a positive effect on QD at 0.506.

4.3.3 Concluding Remarks

A problem in applying SEM is that, in complex models, where many relationships/paths are expressed, the degrees of freedom are quickly exhausted, leading to an underdetermined model. SEM also requires a set of assumptions such as multivariate normality and minimum sample size with various criteria (Hair et al., 2011; Boomsma, 1985)[6] to be fulfilled (Hair et al., 2011). This, in turn, restricts CB-SEM.

Given these limitations, GME-SEM is an information-theoretic-based approach that evaluates results by quantifying the reduction in information uncertainty, that is, it measures the level of information in the data under the recovered model. These criteria could be used to measure the goodness of the estimated model regardless of the sample size and distribution of data. Together, this makes GME-SEM more flexible. However, some challenges persist, and to overcome them, we proposed K-means GME-SEM. K-means GME-SEM inherits the advantages of GME-SEM but outperforms it in terms of reducing calculation volume with acceptable goodness of fit, thus becoming a hands-on estimation method.

This research contributes one more step in developing the new branch of SEM: the information-theoretic-based branch. However, SEM research not only refers to estimation methods, but also combines other techniques. After estimating coefficients, we must also calculate the modification indices for K-means GME-SEM and improve the model to achieve a final suitable model. We hope to explore this in the near future.

The simulation in this research shows that the estimation results are sensitive to consistency constraints. This effect has not been examined carefully. Our simulation used the Hartigan–Wong K-means algorithm; how other K-means algorithms, such as the Lloyd algorithm or Forgy algorithm, might affect the result of the new SEM estimation is yet unknown. Moreover, K-means algorithm has many local optima. Thus, the standard error of estimator of proposed method may be larger than ML and TLS. For investigating the performance of proposed method in some practical cases, we should conduct simulation including various scenarios. This is future works.

[6]Larger sample size when applying SEM is widely accepted as a “rule of thumb”

K-means is also not the only method that can be used to thin the constraint grid. A study that compares alternative choices, such as hierarchical clustering when incorporated with GME-SEM, should be conducted as well. One limitation of our model is that when its complexity increases, that is, the number of parameters increases, the proposed model becomes heavy. A possible solution is to integrate the approach of Huang et al. (2017) to balance the complexity and the goodness of fit for the model. The effects of misspecification should also be investigated for the proposed method to comprehensively compare it with the CB-SEM, as in Kolenikov and Bollen (2012). We consider these to be fruitful directions for future research on the newly proposed method in SEM.

Chapter 5

Conclusion

In this study , we develop two estimation methods for complex linear models to reduce the burden of assumptions for data. The reviewed linear models provide a baseline for the more intricate steps in data analysis.

In Chapter 2, we develop a total least squares (TLS) model, which has been a staple of fitting methods for decades . Although we propose a regression model to describe the relationship between two sets of variables, the true contribution of this study lies in its specifying an application of the TLS and orthogonal regression. Multivariate multiple orthogonal linear regression (MMOLR) is suitable for applications in medical and chemical fields, where numerous variables are known to interact in a complex process that is affected by control variables. In fact, 2B-PSL that is related MMOLR, is developed in chemical research fields. In these research field, linear model is widely used because interpretation of parameters is easy and prediction performance of linear model is often sufficient for practical use. From the result of numerical example, MMOLR has superior prediction performance. Thus, MMOLR will be widely used instead of 2B-PSL and canonical regression.

The MMOLR can be considered a solution for data having multicollinearity, that is, when the independent variables in a regression model are strongly correlated. This phenomenon is especially common in medical or chemical data owing to the specific characteristics of these sectors.

In Chapter 3, we employ the generalized maximum entropy (GME) estimator as an appropriate method for cases without distributional assumption, that is, normality can be omitted when running data analysis. We find that the K-means GME-SEM (structural equation modeling), originally developed for small data, is also suitable for big data. Within the scope of this thesis, the K-means GME was developed for SEM; nevertheless, without changing the theoretical basis, this method can still be applied to other simple linear regression models[1].

Although the estimation method in Chapter 4, K-means GME, can be applied to the model in Chapter 3, it is, at the same time, recommended to use MMOLR for the model with two latent variables because of the brevity of this estimation method.

There are two common points in the proposed estimation methods . The first is assuming a complex relationship among the observed variables caused by the latent variables described in the simple linear model. The two proposed estimation methods estimate the latent variables because the research interests compose the concept from the data. In MMOLR, the TLS is used for the estimation of the latent variable, whereas the GME-SEM method uses the GME. The second point is that the independent variables include error terms. In MMOLR, this is a key assumption. From this assumption, we extend MvMR to MMOLR. In the GME-SEM approach, the error term is modeled and estimated. Modeling the error term by linear combination is the key assumption of GME-SEM because this assumption raises the number of parameters and makes the estimator more robust. Although this robustness is inherent to the K-means GME-SEM, the K-means method reduces the number of constraints in the GME-SEM.

Now, we consider the usefulness of the proposed method in real data analysis. In the context of Industrial Revolution 4.0, our two estimation methods can be applied to big data analysis, machine learning, artificial intelligent systems, and so on. Industrial Revolution 4.0, also known as the 4.0 Revolution toward digital transformation, has been theorized as a concept since 2011. It refers to a paradigm shift wherein all fields of life and production are enhanced through the integration of the physical and digital in order to support real-time data and the Internet of Things. Within this paradigm, emergent concepts such as "smart"—for example, smart city, smart industrial park, smart device, and smart supply chain—have come to exist, and they imply increased productivity and responsiveness of production and service systems through digitalization. This "revolution" is now a global development. In this era, data are collected and processed continuously (real-time data); as a result, they are chaotic, and it is difficult to satisfy the assumptions of distribution. Our estimation method could thus be useful for building linear regression models that can be applied in machine learning and artificial intelligent systems regardless of the chaotic nature of the data.

[1] for a similar approach, see Corral et al. (2017) for GME of linear models.

Acknowledgments

This study has been done under the supervision, advice, encouragement, support, and motivation from multiple people and organizations.

I would like to express my sincere gratitude to my supervisor; without Professor Yadohisa Hiroshi, none of this would have been possible. I am deeply grateful for Sensei’s valuable support, encouragement, advice, patience, and guidance throughout this entire process.

I would also like to thank Mr. Jun Tsuchida for helping me whenever needed and for collaborating with me.

I am deeply grateful to the respected Dean of Faculty of Culture and Information Science, who gave me the valuable opportunity to conduct my study and educated me about research ethics.

Last but not the least, this study is supported by Doshisha University - Young Researcher Project and the Japan Society for the Promotion of Science KAKENHI Grant Number JP18J11566. My sincerest gratitude to these organizations

Appendix A

Singular Value Decomposition

Any matrix $\boldsymbol{X}_{(n\times m)}$ with rank r can be decomposed as

$$\boldsymbol{X}_{(n\times m)} = \boldsymbol{U}_{(n\times r)}\boldsymbol{D}_{(r\times r)}\boldsymbol{V}_{(r\times m)}, \tag{A.1}$$

where

- $\boldsymbol{U}$ is the column orthogonal matrix ($\boldsymbol{U}'\boldsymbol{U} = I$) contains the eigenvectors of $\boldsymbol{X}\boldsymbol{X}'$;
- $\boldsymbol{D} = \mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_r)$ and $\sigma_1 > \sigma_2 > \cdots > \sigma_r$;
- $d_i = \sqrt{\lambda_i}$, with λ_i as the eigenvalue of $\boldsymbol{X}'\boldsymbol{X}$; and
- $\boldsymbol{V}$ is the column orthogonal matrix ($V^T V = I$) that contains the eigenvectors of $\boldsymbol{X}'\boldsymbol{X}$.

Appendix B

Structural Equation Modeling

SEM has developed to be one of the main techniques of data analysis and attracted many scholars across different disciplines, especially those in the social sciences (Barrett, 2007; Kelloway, 1995). SEM is a technique to analyze multiple and interrelated relationships among the constructs for model building (Tabachnick and Fidell, 2007; Hair et al., 2014; Byrne, 2013).

SEM has its own language. To fully understand the complexity of SEM, we summarize its definitions/notations.

SEM includes a set of regression equations simultaneously, where the exogenous or upstream variables are the independent variables. These are assumed to be measured without errors. The endogenous or downstream variables are the dependent or mediator variables [1]. For example, in figure B.1, l_1 is the exogenous variable, l_2 is the mediator [2], and l_3 is the endogenous variable.

[1] A mediator variable is the variable that mediates between the dependent and the independent variables, that is, it explains the indirect relationship between the dependent and the independent variable.

[2] The mediator is also treated as an endogenous variable.

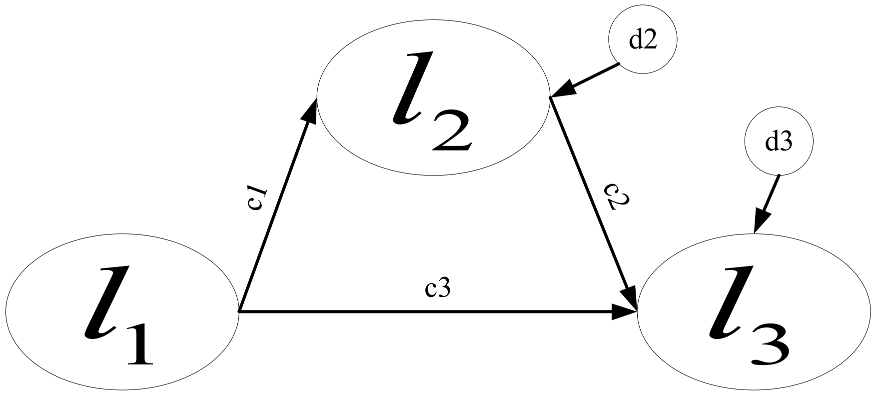


Figure B.1: Relationship in SEM

Manifest/observed variables/indicators are measured by appliers through sampling or from data, whereas latent/unobserved variables are inferred/built from measured variables in the analysis. The relationships among observed and unobserved variables are represented using path diagrams, where latent variables are denoted as ovals or circles, and rectangles or squares represent the measured variables [3]. The single-headed arrows represent causal effects/regression relations in the path diagram. Figure B.2 is an example of the relationships among the manifest constructs v_1, v_2, and v_3 and the latent construct l_1.

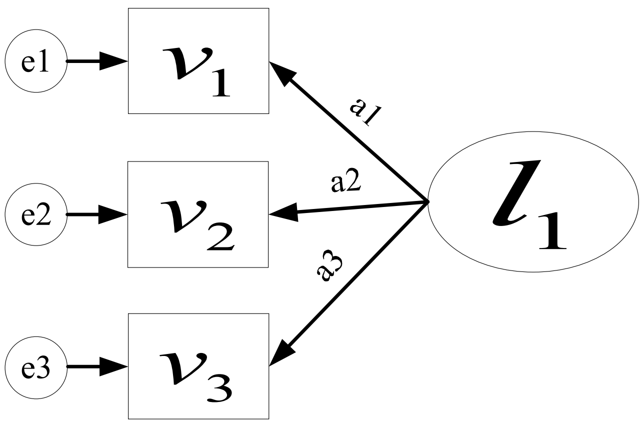


Figure B.2: Example of the relationships between the manifest and latent construct

The numbers on the arrow-shaft are the coefficients or loading of dependent variables on the independent variables. A two-headed arrow represents correlation.

[3]Residuals/errors/disturbances are always unobserved; hence, they are depicted by ovals or circles.

B.1 Expression of an SEM

We now explain how a matrix in an SEM model is formed. As mentioned earlier, single-head arrows represent causal effects. The direction of the arrow distinguishes the independent variable from the dependent variable in a relationship. The arrowhead points to the dependent variable; thus, we could write equations for SEM from the path diagram, or in contrast, draw a path diagram from equations. For instance, in the path diagram in Figure B.2, the arrowhead points to v_1, v_2, and v_3, which yields three equations, (B.1), (B.2), and (B.3).

$$v_1 = a_1 l_1 + e_1. \tag{B.1}$$

$$v_2 = a_2 l_1 + e_2. \tag{B.2}$$

$$v_3 = a_3 l_1 + e_3. \tag{B.3}$$

Similarly, in Figure B.1, the arrowhead points to $L2$ and $L3$, which can be expressed as (B.4) and (B.5).

$$l_2 = c_1 l_1 + d_2. \tag{B.4}$$

$$l_3 = c_3 l_1 + c_2 l_2 + d_3. \tag{B.5}$$

In more complicated models, such as in Figure B.3, we have 11 equations besides (B.1),(B.2), (B.3), (B.4), and (B.5), that is, from (B.6) to (B.11).

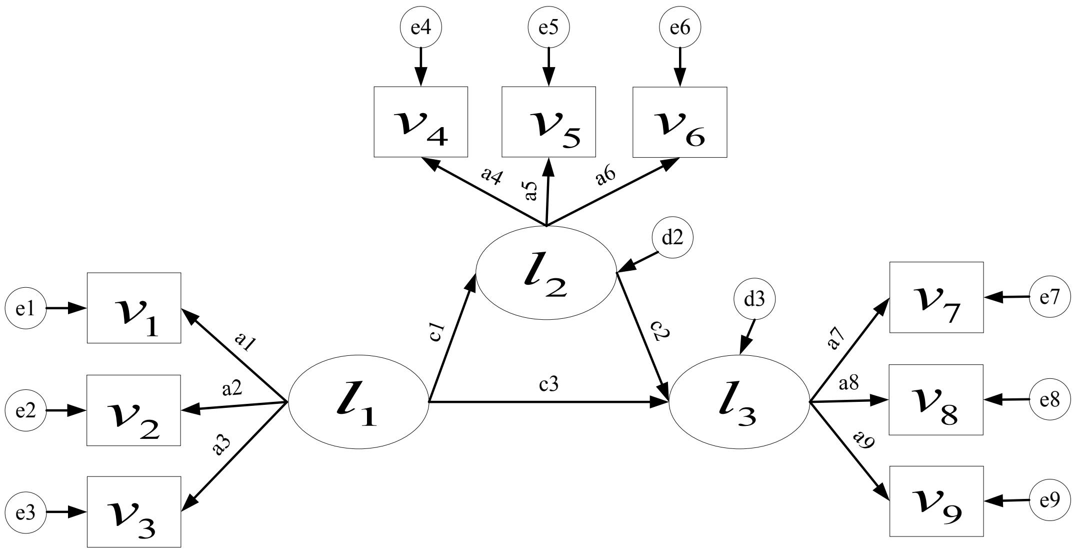


Figure B.3: Prototype path diagram of an SEM

$$v_4 = a_4 l_2 + e_4. \tag{B.6}$$

$$v_5 = a_5 l_2 + e_5. \tag{B.7}$$

$$v_6 = a_6 l_2 + e_6. \tag{B.8}$$

$$v_7 = a_7 l_3 + e_7. \tag{B.9}$$

$$v_8 = a_8 l_3 + e_8. \tag{B.10}$$

$$v_9 = a_9 l_3 + e_9. \tag{B.11}$$

In practice, the models are even more complicated; hence, the best way to express SEM is systematizing through matrices. Matrices allow us to easily write the algorithm for estimating parameters simultaneously. In the next section, we demonstrate the ways to depict SEM in a matrix term.

B.2 SEM in the Matrices Term

Jöreskog (1970) first introduced a general model for SEM; later research developed the basics, that is, theories, terms, and rules, for SEM on this model. One such later model is the Jöreskog' model, which is separated as follows:

- The **measurement model** specifies the relationships between the observed and unobserved variables, for example, Figure B.2. The Jöreskog's model is divided into exogenous and endogenous measurement models.

  - The *exogenous measurement model* includes the relationship between the observed variables and the upstream latent constructs, such as equations (B.1) to (B.3) in the aforementioned example. This model is depicted generally in matrices, as in (B.12).

$$\boldsymbol{x} = \boldsymbol{\Lambda}^{x}{}_{(p_x \times r_x)}\boldsymbol{\xi} + \boldsymbol{\delta}, \tag{B.12}$$

  where $\boldsymbol{x}$ is a vector of p_x exogenous observed variables, $\boldsymbol{\Lambda}^x$ is the matrix of coefficients, $\boldsymbol{\xi}$ is a vector of r_x exogenous latent construct, and $\boldsymbol{\delta}$ is the disturbance vector.

For example, equations (B.1)–(B.3) are arranged in matrices as follows:

$$\begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} (l_1) + \begin{pmatrix} e_1 \\ e_2 \\ e_3 \end{pmatrix}. \tag{B.13}$$

– The *endogenous measurement model* expresses the relationships between the observed variables and the downstream latent constructs, for example, see equations (B.6) to (B.11). (B.14) is the matrix term thereof.

$$\boldsymbol{y} = \boldsymbol{\Lambda}^{y}{}_{(p_y \times r_y)} \boldsymbol{\eta} + \boldsymbol{\varepsilon}, \tag{B.14}$$

where $\boldsymbol{y}$ is a vector of r endogenous observed variables, $\boldsymbol{\Lambda}^y$ is the matrix of coefficients, $\boldsymbol{\eta}$ is a vector of m endogenous latent variables, and $\boldsymbol{\varepsilon}$ is the disturbance vector.

For reference, equations (B.6)–(B.11) are compiled in matrices as (B.15):

$$\begin{pmatrix} v_4 \\ v_5 \\ v_6 \\ v_7 \\ v_8 \\ v_9 \end{pmatrix} = \begin{pmatrix} a_4 & 0 \\ a_5 & 0 \\ a_6 & 0 \\ 0 & a_7 \\ 0 & a_8 \\ 0 & a_9 \end{pmatrix} \begin{pmatrix} l_2 \\ l_3 \end{pmatrix} + \begin{pmatrix} e_4 \\ e_5 \\ e_6 \\ e_7 \\ e_8 \\ e_9 \end{pmatrix} \tag{B.15}$$

- The **structural model** expresses the relationship between the latent variables as (B.16); here, $\mathbf{B}$ is the matrix of coefficients between the endogenous unobserved variables, $\boldsymbol{\xi}$ is the matrix of relationships between the exogenous and endogenous latent variables, and $\boldsymbol{\zeta}$ is the vector of disturbances/errors.

$$\boldsymbol{\eta} = \mathbf{B}_{(r_y \times r_y)} \boldsymbol{\eta} + \boldsymbol{\Gamma}_{(r_y \times r_x)} \boldsymbol{\xi} + \boldsymbol{\zeta}. \tag{B.16}$$

In the prototype model, the structural model in the matrices term is as follows (B.17):

$$\begin{pmatrix} l_2 \\ l_3 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ c_2 & 0 \end{pmatrix} \begin{pmatrix} l_2 \\ l_3 \end{pmatrix} + \begin{pmatrix} 0 \\ c_3 \end{pmatrix} (l_1) + \begin{pmatrix} d_2 \\ d_3 \end{pmatrix}. \tag{B.17}$$

Appendix C

R Code and Simulated Data for Evaluating K-means GME-SEM

The R code for generating data:

```
# Generate data:
N<-150
library("simsem")


# Factor loading - FL
loading <- matrix(0, 9, 3)
loading[1:3, 1] <- c(1, NA, NA)
loading[4:6, 2] <- c(1, NA, NA)
loading[7:9, 3] <- c(1, NA, NA)
loadingVal <- matrix(0, 9, 3)
loadingVal[1:3, 1] <- c(0.5, 0.6, 0.7)
loadingVal[4:6, 2] <- c(0.7, 0.8, 0.9)
loadingVal[7:9, 3] <- c(0.6, 0.7, 0.8)
FL <- bind(loading, loadingVal)


# Factor covariance - FC
facCov <- matrix(NA, 3, 3)
facCovVal <- diag(c(0.8, 0.9, 0.7))
facCovVal[lower.tri(facCovVal)] <- c(0.5, 0.6, 0.7)
facCovVal[upper.tri(facCovVal)] <- c(0.5, 0.6, 0.7)
```

```
FC <- binds(facCov, facCovVal)

# Error covariance
error.var <- matrix(0, 9, 9)
diag(error.var) <- 0.49
TE <- binds(error.var)

# Sigma
err.sig<- 0.7

# Model
CFA.model <- model(LY=FL, PS = FC, TE=TE, modelType="CFA")

# Simulate data
simulated_data_Eigen_GME_SEM <- generate(CFA.model, N, seed = 2)
```

Table C.1: Generated data

Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7	Y_8	Y_9
0.257116767	-0.435759785	0.871574611	-1.090387449	-1.224475746	-1.981917693	-1.776901527	-0.343584556	-0.262242547
-1.436397578	-1.548944369	-1.864633867	-2.088875243	-1.813535573	-2.044601138	-1.836957663	-2.162874727	-2.300472636
1.636564934	0.467052200	-0.189202273	-1.440965458	-1.672976117	-1.746902792	-0.180619781	0.475905259	-0.445401711
0.042672959	-1.450156938	0.17911200	1.054896258	0.062982282	0.699625548	0.699224696	-0.06972893	0.151912064
0.102162028	-1.577846851	0.528058637	0.607253943	-1.317750701	2.308836214	0.010569853	-0.776912857	-0.58698816
-1.349885393	-0.339474585	-1.32167044	-2.889675423	-1.848192094	-2.272699702	-1.374106551	0.149206648	-2.722980067
0.035448832	-1.572102968	-0.932519481	-1.795810185	-2.364913533	-1.071622679	-0.73753275	-0.923740979	-1.293804177
-2.569912456	-0.880190016	-2.295906177	-2.612315385	-0.380481107	0.010640065	-1.433669799	-1.570368858	-1.943093109
-1.311654085	-2.353437033	-1.625895634	-1.172195515	-0.220404797	-0.905268829	-0.502530796	1.821679536	-1.51496348
0.462644373	0.035675869	0.135401057	-1.024260756	-2.056188849	-1.16955423	-1.150202916	-1.432206902	-0.141401318
-0.517266586	-1.487237511	0.7933923	-2.585271725	-1.107376216	-0.331635682	-0.984270191	0.7685597950	0.4502430400
0.690765615	1.071268946	1.115770453	0.475095343	0.900339713	0.236146848	0.124857651	-0.013857927	0.64879101
1.166862311	-0.841581748	0.290340635	1.119395185	1.8461684	1.110905736	0.239508274	-0.603335319	0.709901498
-0.040487811	0.005004798	-0.222850297	0.249234197	0.275775249	0.004735965	-1.483005478	0.06558115	0.411280963
0.601392634	1.357109704	0.879395258	0.92548046	1.768516608	0.520415516	0.922181479	0.639815378	0.565508213
-1.962211446	-0.949089257	-1.841674039	0.395096817	-0.053618695	0.54717211	-0.758173023	0.860613833	0.188159244
-0.481112302	1.114987147	-1.608900973	1.512186284	0.941187572	0.005772442	0.703763068	0.978677568	0.465654925
-2.826787242	-0.51115975	0.980302629	-0.519928455	-1.389122353	0.003624004	-0.733593802	0.507720926	-0.626900798
0.880212626	1.439793891	0.626681549	0.566003582	-0.1081835	1.762219239	-1.010962247	0.31745779	1.32317504

Table C.1: Generated data

Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7	Y_8	Y_9
1.110957771	-0.450663568	-1.941364734	1.028670564	-0.016041275	0.184605809	1.134752728	-0.323516337	60.583961915
-0.665875476	0.161496129	1.02847046	-1.287571415	0.131333887	-1.399811173	0.164499906	-0.786754685	1.357038001
-0.450248517	0.734179085	0.154262102	-0.63763519	-0.554565197	-0.055485332	-1.415512903	0.945152609	-0.886031954
-1.112039207	0.115917521	0.064091708	-0.068310912	-0.674757774	-0.514262919	-1.135170295	0.416054139	-0.276852351
-0.39970317	0.519805104	-0.644413551	0.205326051	-1.543845362	-0.882550185	-1.141952593	-0.074224273	-0.41646639
-2.157267993	-0.929568237	-0.920950143	-1.092570834	-0.672270451	-0.823065736	0.232072473	-1.371017738	-1.165823079
-1.110946402	1.446911273	0.194415762	0.086789228	1.394338573	0.420135206	0.609250027	-0.7899865	0.413041349
0.202775206	0.120359631	0.607320685	-0.661024319	-0.683810814	-1.477568938	-0.055350297	0.39250316	0.320209691
-0.104179009	-0.729708531	-0.002786221	-2.265135531	-1.531400913	-2.000321048	-1.175785507	-0.857731713	-0.001272258
0.072606457	1.649867228	0.76080263	-0.434029748	-1.042057018	-1.1429862	-1.125945809	-0.720219145	-0.369946571
-0.565584132	-0.670933213	0.043397758	0.375683112	-0.35790936	0.47983668	0.754458297	0.102778362	-0.170800809
-0.55737979	-0.477967021	-0.256677649	-0.340077242	0.512854759	0.177036616	-0.123332338	-1.36308503	-1.716144794
-1.200682009	-1.186524374	-1.125597689	-1.27265388	-1.652760173	-1.473482362	-1.097238509	-1.082929698	-0.819358937
-0.718611448	-0.279188935	-0.988119788	-1.484429715	-1.121518226	-1.152199162	-0.582608252	-0.32868167	-1.561236969
1.266654414	-0.361050362	-0.848031395	-0.375214016	0.011155984	-0.070056766	-0.922836864	0.871037808	-0.562816445
0.058239315	-0.196510255	-0.273448994	-0.963426993	-0.142254031	-1.418481474	-1.35330204	-0.241743104	-0.684104049
-0.481425127	-1.141526574	-0.36747267	-1.593029555	-1.139088956	-0.562434779	-0.667058032	-1.944107622	0.037479381
-0.89085085	-1.679936238	-0.589609113	-0.47249018	-0.659173259	-1.465393697	-2.368628817	0.392478159	-1.237601537
0.132715609	0.452703567	-1.656974641	-0.30860897	-0.62331044	-1.30941986	-1.083329644	0.013556691	-0.963274003
-0.399170046	1.937532936	1.84177636	0.805488037	-0.33714057	1.403191215	2.26614702	-0.260451927	1.477708313

Table C.1: Generated data

Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7	Y_8	Y_9
-0.676874879	0.925093544	0.357895342	-0.154806321	1.081952034	2.624346915	-0.528897021	0.880069596	1.077734426
-1.343338112	-0.418018659	-1.292731487	1.516713631	0.670565317	0.118261234	0.260271959	-0.12193574	0.353070295
0.478635049	-0.94004966	-0.15144405	-0.544716037	0.382870156	0.638338388	0.720431432	-0.130212217	-0.788883532
-1.14612688	0.110533652	-0.921190667	0.390146117	-0.687553005	1.789274174	60.003118219	-0.276802246	-0.844725844
-2.160950374	-1.381488134	-0.836339452	-0.241811727	0.262727209	-0.95794986	-0.554285874	0.656302362	-0.653654763
-0.923858658	-0.357313976	-1.043196116	-1.656813491	-0.938470649	-0.477351861	-1.446442198	-1.596916123	1.002117849
0.472012643	.018764389	-0.627306164	0.427403994	-1.830427425	0.628728445	0.822157888	0.26056058	0.22126662
-1.459908108	-0.530241183	-1.067419578	-0.855600879	-0.285624799	-0.234967484	-0.076490818	0.115895339	0.813433546
-0.408236102	0.941463142	-0.143494047	1.990631817	1.210822127	1.847712679	1.83368418	0.029028352	-0.674881845
1.274165389	0.24612278	0.9698283	1.305871486	1.625056951	-0.190886678	2.230769063	1.584317865	-0.290263747
-0.089633837	-0.792396996	-1.182975598	-1.21642068	0.005528958	0.612354439	-0.376939711	-0.017679969	-2.187474643
-0.840863631	0.208345992	-0.249496905	0.859361802	-0.929650275	-0.184876859	-0.269539077	-0.547917757	-1.246111785
0.742891078	0.896334942	0.246088073	-0.238842328	-0.336465244	0.348237735	-0.968632205	1.000727067	0.229538278
-0.459894707	-0.039937052	-0.231728452	1.012160276	0.958424541	0.832623036	0.62433881	-0.675140606	1.21718512
2.167489947	1.846319453	1.178830678	1.398816758	1.267349696	2.324387867	1.244981311	1.037796318	2.276495381
0.537660187	-0.064152791	-0.19307453	-0.442659832	1.085671596	-0.102988124	0.220269684	2.30901971	0.91072053
-0.013099032	-0.138967527	0.459041798	-0.373523587	-1.067995237	-0.937683226	1.140437607	0.285061222	-0.407977972
0.551786395	1.356244233	-0.89029594	0.103262387	-0.361499549	-0.965454574	0.304076817	-0.092743143	-1.388866024
-0.993264038	0.037634862	1.412208771	-0.280218854	0.19551611	-0.348696762	-0.823178675	-0.315225481	0.701094916
-1.018458764	-1.162289141	0.231973728	-0.199829487	-0.996710054	-1.443243703	-2.038299535	-0.955278762	-1.96601707

Table C.1: Generated data

Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7	Y_8	Y_9
-1.039305809	-1.169180329	-1.092510897	0.579930629	0.446594067	0.209937397	-0.830949608	0.674503825	-1.436448581
-1.812296929	-0.657339284	-0.409247966	-1.427172352	-1.030903189	0.170130137	0.02222482	-0.335569486	-0.598657631
-3.144731319	-1.005331189	-0.308912218	-2.030708049	-1.6662654	-2.07388907	-0.836831807	-1.464626355	-1.463624817
0.215165795	-0.01464725	-0.504470634	-0.650804865	1.285949095	0.04838027	1.291626383	0.155157279	0.383740314
2.131446369	0.361765884	2.035294294	1.378860673	0.137238399	0.696731468	1.592443173	-0.26887248	0.273371462
-1.108022446	-1.616022524	-1.155719552	0.733242646	0.671374294	0.160381844	0.29818253	-0.0022536	0.893202906
-0.745878596	0.726388596	0.413224978	1.186911909	1.508370969	1.437341225	1.539338546	0.528114498	1.22794182
-1.098774134	0.334723206	-0.240144645	0.085113463	0.833214628	0.71851861	1.430578129	1.438560569	-0.342718523
-1.017480448	0.135171007	-1.292234523	0.071551552	1.732835591	-1.216784727	-1.225756519	-0.318697186	0.173592481
2.758036741	1.078390188	0.686491661	-1.598529258	0.341719483	0.157254606	-0.062032509	-0.154021893	-1.222983325
-0.023428384	-0.020481854	-0.714525685	-0.375366539	0.90152868	-1.396205464	-1.671065905	-0.989197949	-0.364392708
-1.165933532	-1.456506576	-0.317891359	-0.902910896	-0.324402613	-1.122992865	-0.153254115	0.426659242	0.035330526
1.785764224	0.915610009	1.418347157	0.487930181	0.942572526	0.709002181	1.983224145	0.817275851	0.519545396
0.235128546	-0.594080591	0.017842228	0.325144607	-0.994092332	-0.184988587	-1.160417808	0.177876856	-0.281603689
0.449368882	0.161989947	0.611403071	1.714394446	0.260227953	0.400784715	0.699615873	0.898631843	-0.07145629
0.852873574	2.156152173	0.235126026	1.642938863	1.29395805	-0.15135773	1.436122904	0.015391633	2.179935156
1.35254713	-0.300041387	0.623315734	0.928261425	0.971845121	-0.354385694	-0.279147872	-0.686445573	0.570214895
1.493604433	0.47345855	0.044617477	0.789446286	-0.392761785	1.24531905	1.886538039	-0.054814582	0.802591136
-0.946856535	0.700049738	-1.406259391	-2.546773674	-2.617348712	-1.269311906	0.964701214	0.571040265	-0.096372602
1.267850754	0.427402815	1.334593155	1.157810364	1.029627275	2.856818817	1.16979801	0.09904029	-0.306794038

Table C.1: Generated data

Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7	Y_8	Y_9
0.108903327	-0.008954011	-0.103157234	-0.290711515	-0.34974446	-0.389386367	1.054679441	0.551458544	2.38760717
0.387562372	-0.077256086	0.098016264	-0.673020519	0.790754036	0.001476149	-0.94500172	-0.590271013	-0.427069015
0.558253464	0.586782865	-0.157280398	1.440121287	0.611439226	1.290641841	-0.492125319	0.670303338	1.727330974
-1.208781276	-1.101073913	-1.606342204	-0.853585484	-0.723254754	-0.16061766	-1.557597048	-1.193851558	-1.12908645
0.220560146	-0.480196929	0.551323976	0.00999396	0.421781068	0.150427017	0.28630786	-0.613800914	1.221709269
-0.901218793	0.705199471	-0.464770111	-2.147285023	-0.30184026	-0.075470929	-0.846068217	-1.509273926	-0.780787688
-1.664880596	-1.523607012	0.688582174	-0.887992542	-1.931718435	-1.679893077	-1.23131572	-1.207744491	0.01623792
-1.175575421	-0.087214442	0.982584821	0.621860505	0.387159957	0.322114744	-2.155919691	0.944686866	-1.830105719
0.590522181	-1.161369058	-0.614195351	-2.574493417	-3.310464257	-2.024565504	-2.235084542	-2.21643433	-1.924878185
0.730787536	0.480070199	-0.051011777	-1.434205987	-1.50561639	-0.301492793	-0.384458597	0.120786388	-0.633286947
1.740044852	2.398727485	0.401600712	0.782385575	0.440715608	0.351064762	1.174310201	-0.495815767	1.94023058
1.392371856	0.81890461	-0.438380706	1.788137422	1.781087715	1.049609922	-0.399165913	0.573593076	-0.998153129
-0.325550312	-0.319506169	-1.210766301	-2.78324271	0.336797704	-0.840199667	-1.046383942	-0.787652031	-1.380945825
-0.055326832	0.129491541	0.137620918	-2.316030413	0.086708938	-1.034779441	-0.565618936	-0.934413401	-1.639913877
-1.38184517	-2.516343511	-0.395237278	-0.702599147	-0.112359319	-1.796114686	-1.019133912	0.281178982	-1.49442129
-0.11273105	-0.479509432	0.181237863	-0.285510544	-0.722133479	0.233796113	0.777949557	-0.235097738	-0.184714742
0.773271535	1.341632086	-0.209543724	-0.281293398	0.798606278	0.988769332	-0.322342894	0.427171819	-0.924811561
0.970084719	1.009845269	-0.220642696	1.313632436	0.0758452	-0.547197563	1.897599927	-0.263108728	0.460707833
-0.019844669	-0.31022012	-0.570041272	-0.999002424	-2.266219844	-1.034365237	-0.602832334	-0.558634665	-1.214477123
0.39697368	1.49300407	0.318384581	-0.976272775	-0.662692691	-0.80249963	-0.857014637	0.108327495	0.108448743

Table C.1: Generated data

Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7	Y_8	Y_9
0.750284624	0.215366388	-0.645628078	1.172469878	-0.294676085	1.100544911	-0.013526375	-0.496607457	0.079233007
0.705845901	0.208236425	-1.10709235	0.473688247	0.783576533	-0.099275808	-0.613315651	-0.752535537	-0.363427415
-0.965787217	1.141336646	-0.512625831	0.846093374	0.881622671	-0.040988189	0.413668475	0.222062956	0.178529664
-0.230220488	0.047710324	0.618348961	0.828324471	-0.02297357	0.355523658	-0.030876718	0.751167263	0.356792198
-1.59141365	0.723035711	0.238134543	-0.864295365	-1.189387031	0.044817	-0.389414152	-0.053408542	-0.266967133
1.924750145	1.717881595	1.028406457	0.7469764	1.77562086	1.377399964	1.139928517	-0.071576216	1.117424228
0.164055046	0.318157581	0.545194377	0.985317743	1.729878785	0.316594236	0.287763616	-0.838826675	0.250242792
-2.100568377	-1.291721405	-2.134464996	-0.734214688	0.026676931	-1.222707478	-1.164871421	-0.20497257	-1.154780808
0.181221516	-0.833878556	0.398398775	1.927520489	2.264609144	1.438417548	0.94154976	1.047226622	0.055899584
-0.005636858	0.705510352	1.524397646	-0.68094351	0.164587569	-0.077717223	0.026712954	-0.024314468	-0.819274986
-1.028788689	-1.621667348	-1.178393984	-0.364940629	0.192588054	-0.271969365	0.849175005	0.630206952	-0.789164949
-1.62108207	-0.397023199	0.534980421	1.053062975	-0.329506139	0.003473586	0.136625099	-0.331543479	-0.754747355
-1.251133516	-0.295661802	0.643625095	-0.38543017	-1.121356001	-0.286186896	-1.213394179	0.836891495	-1.63011021
-1.267126292	-0.93286282	-0.381081649	-0.38590526	-1.154593319	-1.057188354	-1.700501299	-1.020692309	-0.785483722
0.421835101	-1.626705053	-0.689779726	2.2603618	1.813785846	1.351039417	1.087635873	0.91742847	0.444348928
0.48948923	0.577020031	0.11528716	1.674194494	0.440270835	-0.102622591	0.655947198	-0.330612062	0.869963311
1.075694616	0.831970157	-0.265663244	2.202758299	0.893919391	0.757748682	1.267019603	0.424127689	0.640714689
-1.094951656	0.203149805	-0.564455578	0.742596598	0.595534405	0.657924558	0.436641291	-0.401781346	0.745393184
1.314705557	0.960350035	0.138957654	-0.888434356	-0.94185473	-0.455181631	1.312090268	0.193588608	0.969503893
1.470481671	2.108613907	0.504183801	-0.300840307	0.581027805	1.140648037	0.890353477	1.465592486	-0.54102989

Table C.1: Generated data

Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7	Y_8	Y_9
-0.806798535	0.523406667	-0.038329507	1.043850665	0.124213161	1.880017692	0.782966385	0.516279833	1.56423092
0.966039544	0.827324111	0.386281373	-0.319400883	0.225449493	0.18053736	-0.12659895	1.001018331	0.253003263
1.698224003	1.200124587	0.438636527	-0.345620559	-0.840667651	-2.97984127	-0.197786076	0.618157714	1.189542865
0.588661339	0.363811003	0.717499405	-2.037117165	-0.599959935	1.059829163	-0.25802138	0.520167442	0.618440212
1.091340953	1.241339724	0.458358529	0.349782925	0.055131518	-0.044856808	-0.771096224	0.204774607	-1.167166219
-2.294761476	-1.271005051	-0.780050397	-1.831594972	-1.007948309	-2.297885305	-2.10703027	-2.914420463	-1.385904874
0.410494952	-1.086216986	-0.373407292	0.035336722	-2.147296593	0.480272286	-0.529658808	-1.330233464	-1.472039129
-1.428691569	-1.063753205	-1.548548508	0.57479568	0.236862767	1.026759305	-0.666843788	0.034538248	0.098476939
-1.428691569	-1.063753205	-1.548548508	0.57479568	0.236862767	1.026759305	-0.666843788	0.034538248	0.098476939
1.853577662	1.271524006	0.522348241	1.738075687	1.410863711	-0.08337952	1.218154157	0.021338557	-1.144738994
-0.6685709	1.092813687	0.423083106	0.566647363	0.494358215	1.486580644	-0.096607623	1.511717811	0.347511504
-1.532105518	-1.147532109	-0.60486152	-0.242084588	-0.962452156	-1.316374628	-0.966708876	-0.369021792	-0.269062627
0.867429567	0.523992035	0.167777029	-0.392817342	-1.045835019	0.353293066	-0.242147272	-0.397038219	0.430619646
0.762057288	1.167447587	-0.015378919	1.489130852	1.297830817	-0.134461113	1.383567262	1.329878644	0.742846888
-0.333695752	0.315181624	0.800536339	-1.568271411	-1.612865146	-2.203063282	0.310089251	-0.649705919	-0.955537818
-1.163077052	-0.586867195	-1.180074726	-1.996348818	-0.634545132	-0.359718204	-1.499477535	-1.386501039	-1.847723948
0.087163754	0.379057656	0.357785417	1.123931295	0.302918627	1.786068906	-0.189140313	1.536410217	-0.746140562
-0.335299058	0.749670817	0.151217655	-0.020967792	-0.165192781	0.757871389	-1.292083618	-0.271933188	-0.16546296
-2.14689875	-0.969812536	-1.510186335	-1.439431863	-1.079935395	0.030415792	-0.291156463	-2.435455721	-0.885001215
1.374532363	0.742437279	0.070723312	1.640767316	0.812128377	-0.088933166	0.826947025	1.475770005	0.161537971

Table C.1: Generated data

Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7	Y_8	Y_9
2.776119724	0.233885226	0.557931754	0.848373225	-1.230293229	0.934302035	0.498340704	0.714010092	0.600197965
0.25905568	1.836528788	1.052677716	1.08659357	1.297905745	0.42551534	1.267087777	0.764957562	0.107349378
0.930028936	0.460550996	0.945979031	-1.166236849	1.154757758	0.804761144	2.923527174	1.723948972	1.02052543
2.470455748	1.144210573	0.452246879	1.447792034	1.879039976	1.029957311	1.904990633	0.87852016	0.459744187
-0.935597239	-0.555864761	0.298449863	0.347656806	0.170259692	-0.627536714	0.379584434	-0.607321829	1.58434846
0.297245061	1.608581569	0.888557663	1.677555575	1.181531467	0.938615783	0.43995822	0.755943489	1.443944232
1.071306866	-0.21600825	1.529149269	-0.797773158	0.1773895	-0.015066655	0.469154715	-0.012565864	-0.880766736
1.499212057	0.242952049	1.470245771	1.823816933	0.849116317	0.43074002	0.72798756	1.135945848	0.922650466
0.151084762	0.872171747	-0.975856644	-1.215216373	-1.409422347	-0.08756513	-0.362210523	-1.245667923	-1.523937432
0.549461019	0.456428272	0.244565958	0.934796049	1.02071317	0.688596041	0.820729979	0.205072484	0.444355303
1.165536142	1.370733988	0.942094152	0.186388918	0.046106985	-1.113419703	0.034774155	-1.684064334	0.558449736

Appendix D

Real Data Description

The real data used in Chapter 4 is extracted from a research project on supply chain management, which published the results under the title "Supply chain management and organizational performance: the resonant influence" (DOI https://doi.org/10.1108 /IJQRM-11-2017-0245 by Emerald in Sight). This project was supported by the European Commission to advance the understanding of the ASEAN region —in this case, Vietnam. The scope of the data in this project is much broader; however, we only describe a part of it in this thesis.

The scales of constructs were developed as shown in Table D.1 based on an extensive literature review. The data were collected using a structural questionnaire whose answers were rated on a seven-point Likert scale, ranging from 1 indicating "strongly disagree" to 7 indicating "strongly agree." This way, we extracted the different attitudes of the respondents, that is, Vietnam-based garment enterprises.

* Note: To provide more information about the data, hereafter the more in-depth terms of supply chain management will be used.

Table D.1: Questionnaire for collecting data

No.	Constructs	Observed items
1	Top management support (TMS)	TMS1: Offer of innovation and continuous improvement policies. TMS2: Provision of necessary resources for processes. TMS3: Promotion of partners' involvement in firm's activities. TMS4: Participation of top management in supply chain improvement process.

Continued on the next page

No.	Constructs	Observed items
2	Human resource management (HRM)	HRM5: Relationship between human resource objectives and strategy. HRM6: Role of environment on the development of all employees. HRM7: Promoting the motivation of employees HRM8: Involvement in determining training needs. HRM9: Timely training program for employees. HRM10: Responsibility in employees’ tasks.
3	Reporting and analysis of quality data (QD)	QD11: The collection of quality data. QD12: Display of quality data, control charts... at workstations. QD13: Delivery feedback of quality data to employees.

The data are presented in Table D.2.

Table D.2: Extracted real data

TMS1	TMS2	TMS3	TMS4	HRM5	HRM6	HRM7	HRM8	HRM9	HRM10	QD11	QD12	QD13
4.00	5.00	4.00	7.00	5.00	5.00	7.00	5.00	6.00	6.00	7.00	5.00	7.00
7.00	7.00	4.00	4.00	7.00	4.00	5.00	5.00	7.00	7.00	6.00	6.00	7.00
5.00	6.00	7.00	5.00	7.00	7.00	7.00	5.00	5.00	7.00	7.00	5.00	5.00
5.00	6.00	5.00	6.00	6.00	7.00	7.00	6.00	5.00	5.00	6.00	7.00	6.00
3.00	4.00	5.00	5.00	4.00	3.00	3.00	5.00	5.00	3.00	6.00	4.00	7.00
5.00	3.00	5.00	4.00	5.00	5.00	4.00	3.00	5.00	3.00	5.00	5.00	6.00
1.00	1.00	1.00	3.00	3.00	4.00	5.00	3.00	3.00	5.00	1.00	3.00	1.00
6.00	4.00	5.00	5.00	7.00	6.00	4.00	4.00	6.00	5.00	7.00	7.00	6.00
3.00	4.00	3.00	3.00	7.00	7.00	6.00	4.00	4.00	6.00	7.00	7.00	7.00
4.00	4.00	4.00	6.00	4.00	4.00	4.00	5.00	4.00	6.00	7.00	7.00	5.00
3.00	3.00	5.00	3.00	3.00	5.00	3.00	3.00	5.00	5.00	7.00	6.00	7.00
6.00	5.00	6.00	7.00	5.00	7.00	6.00	7.00	7.00	4.00	7.00	7.00	5.00
7.00	5.00	4.00	4.00	6.00	4.00	4.00	5.00	5.00	7.00	6.00	5.00	6.00
3.00	5.00	5.00	4.00	5.00	3.00	5.00	3.00	5.00	3.00	1.00	3.00	3.00
5.00	3.00	5.00	5.00	1.00	3.00	3.00	1.00	3.00	3.00	3.00	3.00	3.00
3.00	5.00	3.00	5.00	4.00	3.00	4.00	3.00	4.00	4.00	7.00	6.00	5.00
7.00	7.00	4.00	7.00	3.00	3.00	4.00	3.00	5.00	3.00	7.00	6.00	7.00
7.00	7.00	7.00	7.00	6.00	4.00	4.00	7.00	5.00	7.00	6.00	7.00	6.00
7.00	4.00	6.00	4.00	7.00	7.00	7.00	6.00	5.00	4.00	5.00	7.00	7.00

Table D.2: Extracted real data

TMS1	TMS2	TMS3	TMS4	HRM5	HRM6	HRM7	HRM8	HRM9	HRM10	QD11	QD12	QD13
3.00	5.00	4.00	3.00	3.00	5.00	5.00	4.00	3.00	4.00	7.00	6.00	5.00
5.00	7.00	7.00	4.00	7.00	6.00	7.00	7.00	7.00	6.00	7.00	5.00	5.00
5.00	7.00	7.00	5.00	7.00	7.00	7.00	4.00	7.00	7.00	6.00	5.00	4.00
5.00	3.00	3.00	5.00	3.00	5.00	5.00	4.00	3.00	3.00	7.00	7.00	7.00
3.00	3.00	3.00	5.00	1.00	1.00	1.00	1.00	3.00	1.00	3.00	2.00	3.00
4.00	7.00	7.00	6.00	5.00	5.00	5.00	5.00	7.00	7.00	7.00	4.00	7.00
5.00	7.00	7.00	6.00	5.00	5.00	4.00	5.00	4.00	5.00	7.00	7.00	6.00
6.00	5.00	5.00	4.00	6.00	4.00	6.00	6.00	6.00	7.00	5.00	7.00	7.00
7.00	4.00	4.00	4.00	5.00	7.00	5.00	6.00	4.00	5.00	7.00	5.00	5.00
4.00	7.00	5.00	5.00	6.00	5.00	6.00	7.00	7.00	5.00	7.00	7.00	5.00
7.00	4.00	7.00	4.00	4.00	4.00	6.00	7.00	7.00	7.00	7.00	6.00	5.00
7.00	6.00	7.00	4.00	7.00	7.00	5.00	5.00	7.00	4.00	7.00	5.00	7.00
3.00	5.00	5.00	3.00	5.00	5.00	5.00	5.00	5.00	5.00	7.00	6.00	6.00
3.00	2.00	3.00	2.00	5.00	5.00	3.00	3.00	4.00	3.00	3.00	2.00	3.00
5.00	4.00	3.00	3.00	3.00	5.00	3.00	3.00	5.00	4.00	3.00	2.00	2.00
4.00	4.00	3.00	3.00	6.00	6.00	6.00	7.00	7.00	6.00	7.00	7.00	7.00
5.00	7.00	4.00	7.00	6.00	4.00	7.00	7.00	7.00	7.00	6.00	5.00	6.00
7.00	7.00	7.00	7.00	7.00	4.00	7.00	5.00	5.00	7.00	7.00	5.00	7.00
7.00	4.00	7.00	5.00	4.00	7.00	6.00	7.00	4.00	7.00	7.00	6.00	5.00
4.00	4.00	4.00	7.00	7.00	5.00	6.00	6.00	7.00	6.00	4.00	7.00	5.00

Table D.2: Extracted real data

TMS1	TMS2	TMS3	TMS4	HRM5	HRM6	HRM7	HRM8	HRM9	HRM10	QD11	QD12	QD13
3.00	5.00	3.00	4.00	1.00	2.00	1.00	3.00	1.00	3.00	1.00	3.00	1.00
5.00	4.00	5.00	3.00	3.00	3.00	1.00	3.00	1.00	3.00	3.00	3.00	1.00
3.00	3.00	3.00	5.00	5.00	5.00	3.00	3.00	3.00	5.00	3.00	3.00	1.00
3.00	3.00	3.00	4.00	4.00	4.00	3.00	3.00	3.00	5.00	6.00	6.00	7.00
7.00	4.00	7.00	6.00	6.00	6.00	6.00	7.00	4.00	6.00	6.00	7.00	7.00
5.00	7.00	4.00	4.00	4.00	4.00	7.00	5.00	5.00	6.00	7.00	6.00	6.00
5.00	4.00	5.00	4.00	5.00	3.00	4.00	4.00	3.00	5.00	5.00	6.00	7.00
4.00	6.00	5.00	4.00	6.00	5.00	4.00	7.00	5.00	6.00	7.00	7.00	5.00
5.00	5.00	5.00	5.00	6.00	5.00	5.00	7.00	4.00	7.00	6.00	6.00	6.00
4.00	4.00	3.00	4.00	3.00	3.00	5.00	5.00	3.00	5.00	2.00	1.00	1.00
3.00	1.00	2.00	2.00	5.00	5.00	5.00	4.00	5.00	4.00	1.00	3.00	3.00

Bibliography

Al-Nasser, A. D.(2003). Customer satisfaction measurement models: Generalised Maximum Entropy approach, *Pakistan journal of statistics - all series-*, **19** (2), 213–226.

Anderson, J. C. and Gerbing, D. W.(1988). Structural equation modeling in practice: A review and recommended two-step approach., *Psychological bulletin*, **103** (3), 411–423.

Anderson, T. W.(1962). An introduction to multivariate statistical analysis, Technical report, Wiley New York.

Barrett, P.(2007). Structural equation modelling: Adjudging model fit, *Personality and Individual differences*, **42** (5), 815–824.

Bartlett, M. S.(1951). An inverse matrix adjustment arising in discriminant analysis, *The Annals of Mathematical Statistics*, **22** (1), 107–111. https://www.jstor.org/stable/2236707.

Berger, A. L., Pietra, V. J. D., and Pietra, S. A. D.(1996). A maximum entropy approach to natural language processing, *Computational linguistics*, **22** (1), 39–71.

Berkson, J.(1950). Are there two regressions? *Journal of the american statistical association*, **45** (250), 164–180.

Bickel, D. R.(2015). Blending Bayesian and frequentist methods according to the precision of prior information with applications to hypothesis testing, *Statistical methods & applications*, **24** (4), 523–546.

Boomsma, A.(1985). Nonconvergence, improper solutions, and starting values in LISREL maximum likelihood estimation, *Psychometrika*, **50** (2), 229–242.

Breiman, L. and Friedman, J. H.(1997). Predicting multivariate responses in multiple linear regression, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **59** (1), 3–54.

Brown, M. L.(1982). Robust line estimation with errors in both variables, *Journal of the American Statistical Association*, **77** (377), 71–79.

Byrne, B. M.(2013). *Structural equation modeling with LISREL, PRELIS, and SIMPLIS: Basic concepts, applications, and programming*, Psychology Press.

Ciavolino, E. and Al-Nasser, A. D.(2009). Comparing generalised maximum entropy and partial least squares methods for structural equation models, *Journal of nonparametric statistics*, **21** (8), 1017–1036.

Conrad, K.(2004). Probability distributions and maximum entropy, *Entropy*, **6** (452), 10–37.

Corral, P., Kuehn, D., and Jabir, E.(2017). Generalized maximum entropy estimation of linear models, *The Stata Journal*, **17** (1), 240–249.

Csiszar, I.(1991). Why least squares and maximum entropy? An axiomatic approach to inference for linear inverse problems, *The annals of statistics*, **19** (4), 2032–2066.

Dattalo, P.(2013). *Analysis of multiple dependent variables*, Oxford University Press.

Duong, T. B. A., Tsuchida, J., and Yadohisa, H.(2018). Multivariate Multiple Orthogonal Linear Regression, in *International Conference on Intelligent Decision Technologies*, 44–53, Springer.

Duong, T. B. A., Tsuchida, J., and Yadohisa, H.(2021). K-means generalized maximum entropy estimation for structural equation modeling, *Behaviormetrika*, **48**, 103–115.

Durbin, J.(1954). Errors in variables, *Revue de l'institut International de Statistique*, 23–32.

Elo, S. and Kyngäs, H.(2008). The qualitative content analysis process, *Journal of advanced nursing*, **62** (1), 107–115.

Everitt, B. S., Dunn, G. et al.(2001). *Applied multivariate data analysis*, **2**, Wiley Online Library.

Ghalanos, A. and Theussl, S.(2012). Rsolnp: general non-linear optimization using augmented Lagrange multiplier method, *R package version*, **1**.

Gill, J. and King, G.(2004). What to do when your Hessian is not invertible: Alternatives to model respecification in nonlinear estimation, *Sociological methods & research*, **33** (1), 54–87.

Golan, A.(2002). 『Information and entropy econometrics 窶覇 ditor's view』 *Journal of econometrics*, **107** (1-2), 1–15.

Golan, A.(2008). 『Information and entropy econometrics 窶尿 review and synthesis』 *Foundations and trends® in econometrics*, **2** (1–2), 1–145.

Golan, A., Judge, G. G., and Miller, D.(1996). *Maximum entropy econometrics: Robust estimation with limited data*, Wiley New York.

Golan, A., Judge, G., and Karp, L.(1996). A maximum entropy approach to estimation and inference in dynamic models or counting fish in the sea using maximum entropy, *Journal of economic dynamics and control*, **20** (4), 559–582.

Golan, A., Judge, G., and Robinson, S.(1994). Recovering information from incomplete or partial multisectoral economic data, *The Review of economics and statistics*, 541–549.

Golub, G. H. and Van Loan, C.(1979). Total least squares, in *Smoothing Techniques for Curve Estimation*, Springer, 69–76.

Golub, G. H. and Van Loan, C. F.(1980). An analysis of the total least squares problem, *SIAM journal on numerical analysis*, **17** (6), 883–893.

Gupta, A. K., Harrar, S. W., and Pardo, L.(2007). On testing homogeneity of variances for nonnormal models using entropy, *Statistical methods and applications*, **16** (2), 245–261.

Hair, J. F., Ringle, C. M., and Sarstedt, M.(2011). PLS-SEM: Indeed a silver bullet, *Journal of marketing theory and practice*, **19** (2), 139–152.

Hair, J. F., Anderson, R. E., Tatham, R. L., and Black, W. C.(1998). Multivariate data analysis. Englewood Cliff, *New Jersey, USA*, **5** (3), 207–2019.

Hair, J. F., Anderson, R. E., Tatham, R. L., and William, C.(1998). Black (1998), *Multivariate data analysis*, **5**, 87–135.

Hair, J. F., Gabriel, M., and Patel, V.(2014). AMOS covariance-based structural equation modeling (CB-SEM): Guidelines on its application as a marketing research tool, .

Harrell Jr, F. E.(2015). *Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis*, **0172–7397**, Springer-Verlag New York.

Hoelter, J. W.(1983). The analysis of covariance structures: Goodness-of-fit indices, *Sociological Methods & Research*, **11** (3), 325–344.

Huang, P.-H., Chen, H., and Weng, L.-J.(2017). A penalized likelihood method for structural equation modeling, *psychometrika*, **82** (2), 329–354.

Jain, A. K.(2010). Data clustering: 50 years beyond K-means, *Pattern recognition letters*, **31** (8), 651–666.

Jaynes, E. T.(1957). Information theory and statistical mechanics, *Physical review*, **106** (4), 620–630.

Jaynes, E. T.(1984). Prior information and ambiguity in inverse problems, *Inverse problems*, **14**, 151–166.

Jöreskog, K. G.(1970). A general method for estimating a linear structural equation system, *ETS research bulletin series*, **1970** (2), i–41.

Jöreskog, K. G.(1978). Structural analysis of covariance and correlation matrices, *Psychometrika*, **43** (4), 443–477.

Kelloway, E. K.(1995). Structural equation modelling in perspective, *Journal of Organizational Behavior*, **16** (3), 215–224.

Kolenikov, S. and Bollen, K. A.(2012). Testing negative error variances: Is a Heywood case a symptom of misspecification? *Sociological methods & research*, **41** (1), 124–167.

Lloyd, S. P.(1957). Least Squares Quantitization in PCM, *IEEE Transactions on Information.*

Loehlin, J. C.(1998). *Latent variable models: An introduction to factor, path, and structural analysis*, Lawrence Erlbaum Associates Publishers.

Lutz, J. G. and Eckert, T. L.(1994). The relationship between canonical correlation analysis and multivariate multiple regression, *Educational and Psychological Measurement*, **54** (3), 666–675.

Markovsky, I. and Van Huffel, S.(2007). Overview of total least-squares methods, *Signal processing*, **87** (10), 2283–2302.

Maronna, R.(2005). Principal components and orthogonal regression based on robust scales, *Technometrics*, **47** (3), 264–273.

McArdle, B. H. and Anderson, M. J.(2001). Fitting multivariate models to community data: a comment on distance-based redundancy analysis, *Ecology*, **82** (1), 290–297.

Nigam, K., Lafferty, J., and McCallum, A.(1999). Using maximum entropy for text classification, in *IJCAI-99 workshop on machine learning for information filtering*, **1**, 61–67.

Paris, Q. and Howitt, R. E.(1998). An analysis of ill-posed production problems using maximum entropy, *American journal of agricultural economics*, **80** (1), 124–138.

Pukelsheim, F.(1994). The three sigma rule, *The american statistician*, **48** (2), 88–91.

Rohlf, F. J. and Corti, M.(2000). Use of two-block partial least-squares to study covariation in shape, *Systematic Biology*, **49** (4), 740–753.

Schaffrin, B. and Wieser, A.(2008). On weighted total least-squares adjustment for linear regression, *Journal of geodesy*, **82** (7), 415–421.

Shannon, C. E.(1948). A mathematical theory of communication, *Bell system technical journal*, **27** (3), 379–423.

Singh, A., Yadav, A., and Rana, A.(2013). K-means with Three different Distance Metrics, *International Journal of Computer Applications*, **67** (10).

Soofi, E. S.(1992). A generalizable formulation of conditional logit with diagnostics, *Journal of the american statistical association*, **87** (419), 812–816.

Soofi, E. S.(1994). Capturing the intangible concept of information, *Journal of the american statistical association*, **89** (428), 1243–1254.

Studholme, C., Hill, D. L., and Hawkes, D. J.(1999). An overlap invariant entropy measure of 3D medical image alignment, *Pattern recognition*, **32** (1), 71–86.

Tabachnick, B. and Fidell, L.(2007). Multivariate analysis of variance and covariance, *Using multivariate statistics*, **3**, 402–407.

Tatsuoka, M. M. and Lohnes, P. R.(1988). *Multivariate analysis: Techniques for educational and psychological research*, Macmillan Publishing Co, Inc.

Uffink, J.(1996). The constraint rule of the maximum entropy principle, *Studies in history and philosophy of science part B: studies in history and philosophy of modern physics*, **27** (1), 47–79.

Van Huffel, S. and Vandewalle, J.(1991). *The total least squares problem: computational aspects and analysis*, SIAM.

Velu, R. and Reinsel, G. C.(1998). *Multivariate reduced-rank regression: theory and applications*, Springer. https://doi.org/10.1007/978-1-4757-2853-8.

Velu, R. and Reinsel, G. C.(2013). *Multivariate reduced-rank regression: theory and applications*, **136**, Springer Science & Business Media.

Wang et al.(2017). *Package 'robust'*.

Wold, H.(1975). Path models with latent variables: The NIPALS approach, in *Quantitative sociology*, Elsevier, 307–357.

Wynn, H.(1993). The foundation of experimental design and observation, *Journal of the italian statistical society*, **2** (2), 137–158.

Zamar, R. H.(1989). Robust estimation in the errors-in-variables model, *Biometrika*, **76** (1), 149–160.