

JOOST VAN ROSMALEN

Segmentation and Dimension Reduction

Exploratory and Model-Based Approaches



**Segmentation and Dimension Reduction:
Exploratory and Model-Based Approaches**

Segmentation and Dimension Reduction: Exploratory and Model-Based Approaches

(Segmentatie en dimensiereductie: exploratieve en modelmatige benaderingen)

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Erasmus Universiteit Rotterdam
op gezag van de rector magnificus

Prof.dr. S.W.J. Lamberts

en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op

donderdag 9 april 2009 om 13:30 uur

door

JOOST VAN ROSMALEN
geboren te Rotterdam.



Promotiecommissie

Promotor: Prof.dr. P.J.F. Groenen
Overige leden: Prof.dr. R. Paap
Prof.dr.ir. A. Smidts
Prof.dr. J.K. Vermunt

Co-promotor: Dr. A.J. Koning

Erasmus Research Institute of Management - ERIM
Rotterdam School of Management (RSM)
Erasmus School of Economics (ESE)
Erasmus University Rotterdam
Internet: <http://www.irim.eur.nl>

ERIM Electronic Series Portal: <http://hdl.handle.net/1765/1>

ERIM PhD Series in Research in Management, 165
Reference number ERIM: EPS-2009-165-MKT
ISBN 978-90-5892-201-4

©2009, Joost van Rosmalen

Design: B&T Ontwerp en advies www.b-en-t.nl
Print: Haveka www.haveka.nl

All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means electronic or mechanical, including photocopying, recording, or by any information storage and retrieval system, without permission in writing from the author.

Acknowledgements

Although a PhD thesis has only one author, a thesis is seldom written alone, and my thesis is no exception. I certainly could not have written this thesis without the help I received from many colleagues and friends, and here I wish to thank them for their efforts. Below, I give an overview of the people who, either directly or indirectly, contributed to my PhD thesis.

First and foremost, I am indebted to my promotor Patrick Groenen and my co-promotor Alex Koning. Patrick taught me how to do research, gave me new ideas and introduced me to new scientific fields, and somehow always found the time to read all my work. Alex was the first person to stimulate me to join a PhD program, and I enjoyed working together with him on Chapter 2. I also wish to thank the other members of the dissertation committee for taking the time to evaluate my thesis.

Over the years, I have collaborated with several other researchers. In the project on modeling response styles, I worked together with Hester van Herk, whom I thank for her cooperation and enthusiasm. This project resulted in what I consider the greatest success of my PhD: a publication in the *Journal of Marketing Research*. Not only did this project motivate me and help me learn the virtues of patience and perseverance, but it also made me realize that the methods I was developing could actually be useful for practitioners. I also wish to thank Bram van Dijk and Richard Paap for their cooperation. By working together with them on Chapter 6, we managed to create an interesting new paper in a very short time; hopefully, this paper will be published some day. Finally, my thesis has also benefitted from the useful comments I received from Ale Smidts, Bas Donkers, Dennis Fok, Ludo Waltman, Richard Paap, Jeroen Vermunt, and various editors and reviewers of journals.

Writing a PhD thesis can sometimes be a solitary experience. Fortunately, there were always my fellow PhD candidates to cheer me up. I was blessed to have a variety of smart, funny, interesting, and friendly colleagues. I certainly enjoyed the combined lunches, parties, and the lengthy coffee breaks in which every possible topic was discussed. Therefore, I would like to thank Amy, Bram, Carlos, Chen, Daniel, Georgi, Ludo, Jan Frederik, Jos, Lenny, Mariëlle, Martijn, Merel, Nees Jan, Nuno, Rene, Rianne, Robin, Sandra, Tim, Ward, Wilco, and all the other PhD candidates at the Erasmus School of Economics. Special thanks go to my paranymphs, Bram and Merel, who helped me organize the day of the defense. I am also grateful to the staffs of the Econometric Institute, the

Tinbergen Institute, and ERIM, who supported my PhD in many practical ways. Finally, I thank my parents for supporting me during these years.

Joost van Rosmalen
Rotterdam, February 2009

Contents

Acknowledgements	v
1 Introduction	1
1.1 Exploratory and Model-Based Data Analysis	2
1.2 Dimension Reduction	3
1.3 Segmentation	5
1.4 Summary	5
I Bilinear Decomposition Models	9
2 Optimal Scaling of Interaction Effects in Generalized Linear Models	11
2.1 Introduction	11
2.2 Optimal Scaling of Interactions Model	13
2.3 Extensions	17
2.3.1 Multidimensional extensions	17
2.3.2 Multilinear models for higher-way interactions	20
2.4 Empirical Applications	21
2.4.1 STAR Data Set	21
2.4.2 General Social Survey Data	26
2.5 Discussion	29
2.A Model Reparameterization	30
3 Identifying Response Styles: A Latent-Class Bilinear Multinomial Logit Model	33
3.1 Introduction	33
3.2 A Latent-Class Bilinear Multinomial Logit Model	36
3.3 Empirical Application: The List of Values	41
3.4 Results	42
3.5 Conclusions and Discussion	57
3.A Parameter Restrictions	59
3.B Optimization Algorithm	60

3.C	Model Identifiability	61
3.D	Simulation Study	61
3.E	Confidence Ellipses	64
4	Modeling Attribute Interactions in Choice-Based Conjoint Analysis	67
4.1	Introduction	67
4.2	Overview of Choice-Based Conjoint Analysis	68
4.3	The Conjoint-OSI model	69
4.4	Empirical Application	71
4.5	Optimal Designs	74
4.6	Conclusions	76
II	Two-Mode Partitioning	77
5	Optimization Strategies for Two-Mode Partitioning	79
5.1	Introduction	79
5.2	Overview of Optimization Problem	81
5.2.1	Alternating Exchanges Algorithm	82
5.2.2	Two-Mode k -Means Algorithm	83
5.3	Meta-Heuristics	84
5.3.1	Simulated Annealing	84
5.3.2	Tabu Search	85
5.4	Fuzzy Two-Mode Partitioning	85
5.4.1	The Two-Mode Fuzzy c -Means Algorithm	86
5.4.2	Fuzzy Steps	87
5.5	Simulation Study	88
5.5.1	Setup of the Simulation Study	88
5.5.2	Simulation Results	93
5.6	Empirical Application	99
5.7	Conclusions	102
6	A Bayesian Approach to Two-Mode Clustering	105
6.1	Introduction	105
6.2	The Latent-Class Two-Mode Clustering Model	107
6.2.1	The Likelihood Function	107
6.2.2	Parameter Estimation	109
6.3	Posterior Simulator	110
6.3.1	The Gibbs Sampler	110
6.3.2	Selecting the Numbers of Segments	113
6.4	Application 1: Supreme Court Voting Data	115
6.5	Application 2: Roll Call Voting Data	119
6.5.1	Data	119

6.5.2	Parameter Estimates	120
6.5.3	Interpretation of Segments	123
6.6	Conclusions	126
7	Discussion	127
7.1	Summary of the Main Findings	127
7.2	Limitations and Avenues for Future Research	129
	Nederlandse samenvatting (Summary in Dutch)	131
	Bibliography	135
	Curriculum Vitae	145

Chapter 1

Introduction

Representing the information in a data set in a concise way is an important part of data analysis. The data sets that are used in modern-day empirical research can often be hard to represent or interpret without specialized statistical techniques. This is because these data sets often comprise many respondents or objects and many variables, leading to large matrices of observed data. In addition, we often have limited information on how the variables are related and what kinds of effects can be expected, which exacerbates the problem. To understand the relations among variables in such large data matrices, we consider exploratory statistical techniques for analyzing multivariate data. That is, we do not have an explicit theory that we want to test, we only wish to better understand the data. These exploratory techniques make relatively few assumptions and emphasize flexible searching for clues and evidence (see, for example, Hoaglin, Mosteller, & Tukey, 1983). We specifically look for techniques that can summarize the information in a data set in ways that are easy to understand, preferably using graphical representations. At the same time, we wish to retain the advantages of imposing statistical assumptions on the process that generates the data; these advantages include the possibility to conduct statistical inference and hypothesis testing. To do so, we imbed techniques for summarizing multivariate data in statistical models.

In this thesis, we present a number of methodological innovations with respect to exploratory and model-based multivariate data analysis techniques. The goal of the thesis is to develop new models and methodologies for representing data in ways that are simple and easy to understand. We seek to develop models that can summarize and explain the information in a data set in a parsimonious way. The focus is on techniques for segmentation (which includes clustering and finite mixture modeling) and dimension reduction. To illustrate the usefulness of the techniques, the thesis includes a variety of empirical applications in marketing, psychometrics, and political science. In most of the chapters in this thesis, we propose new statistical models. The first part of this thesis focuses on bilinear decomposition models, and the second part considers techniques for two-mode partitioning. We treat a variety of subject areas, and each chapter presents a

unique methodological contribution. To show how the methods can be applied, we use at least one empirical data set in each chapter.

This thesis contributes to the literature in the following ways. First, we develop a general model for representing two-way interaction effects in generalized linear models using graphical representations. Second, we show how the methods discussed in this thesis can be used to model the response behavior of respondents in rating scale surveys with multiple items. Our method can provide better nuance on how response styles and opinions affect the response behavior than previous methods. Third, we show how bilinear decompositions can also be used to model attribute interactions in conjoint analysis experiments. Fourth, we compare several optimization methods for deterministic two-mode partitioning methods and show which methods yield the best results. Finally, we develop a new Bayesian approach to model-based two-mode clustering. This Bayesian approach gives more informative results and can be easier to implement than previous classical estimation methods for model-based two-mode clustering.

In this introduction, we first explain what kinds of models and data analysis techniques are considered. We then give a brief overview of the two types of techniques used in this thesis. Finally, we discuss an outline of the individual chapters, including their scientific contribution and publication status.

1.1 Exploratory and Model-Based Data Analysis

Many exploratory techniques for analyzing multivariate data, such as k -means clustering (see, for example, Hartigan, 1975) and principal components analysis (Jolliffe, 2002), are procedural or deterministic and do not rely on statistical assumptions. That is, these techniques only consist of an algorithm or a criterion function that needs to be optimized; no assumptions are made about the process that generated the data. Not making such assumptions has the advantage that these techniques are easy-to-apply, but can also limit their usefulness, as the statistical significance of the results cannot be known.

In this thesis, we strive to develop model-based techniques for analyzing multivariate data. By making explicit assumptions on the distribution of the data, we can determine whether the effects of variables are statistically significant and provide confidence intervals for the parameter estimates. Using information criteria, we can also estimate the complexity of the models (that is, the number of segments/clusters and components/dimensions). A disadvantage of the model-based techniques in this thesis is that they need to be designed for a specific type of data set. Simpler techniques, such as k -means clustering and principal components analysis, can be applied to almost any data set presented in this thesis, but would generally provide less informative results.

A large number of techniques can be considered as a form of exploratory multivariate data analysis, and giving a complete overview of these techniques is beyond the scope of this thesis. Instead, we focus on two techniques in this thesis: segmentation and dimension reduction. Specifically, we consider models based on a bilinear parameter structure (dimension reduction) and techniques for two-mode partitioning (segmentation). Below,

we give a general overview of the techniques of dimension reduction and segmentation and also a description of how these techniques are implemented in this thesis.

1.2 Dimension Reduction

Dimension reduction is the general technique of representing high-dimensional data in a low-dimensional space, preferably using graphical representations. Well-known examples of such techniques are principal components analysis, multiple correspondence analysis, and multidimensional scaling. To reduce the dimensionality of the data, one typically assigns a point in a Euclidean space to each respondent, object, or (category of a) variable. The data can then be visualized by plotting the respondents, objects, or variables in a figure. The resulting graphical displays are a major advantage of dimension reduction techniques, because they can help interpret the data. If different kinds of objects (for example, respondents and variables) are jointly represented in a single figure, the resulting figures are called *biplots* (Gower & Hand, 1996).

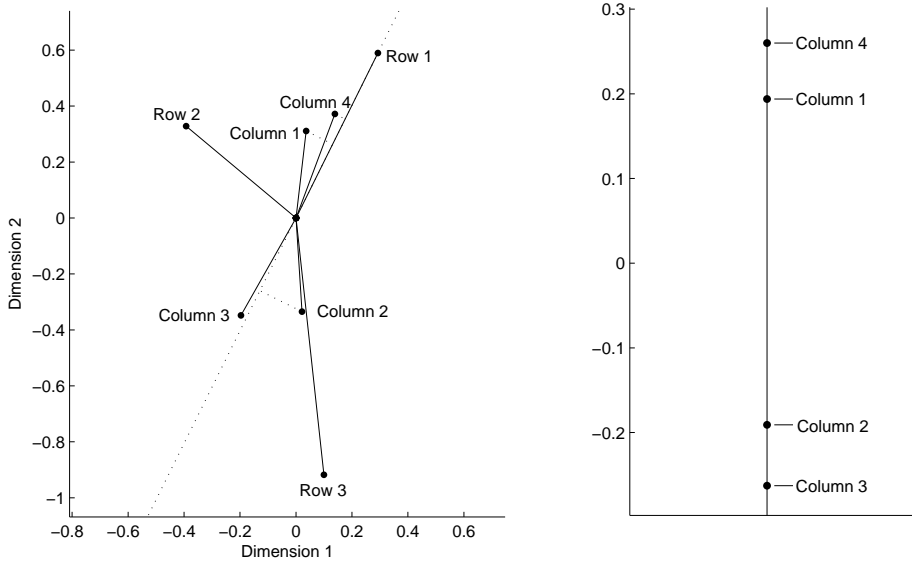
In this thesis, we incorporate dimension reduction methods in statistical models, so that the locations of the points must be estimated as parameters. We use bilinear parameter structures, in which a matrix of parameters is approximated by parameter matrices of a lower rank. This lower rank equals the dimensionality of the representation.

Mathematically, a bilinear parameter structure can be imposed as follows. Suppose that the effects of certain variables in a statistical model are estimated by the $k \times l$ parameter matrix \mathbf{H} . Such parameter matrices often occur in practice, especially in models in which several (predictor or response) variables have a categorical measurement level, such as analysis of variance. Instead of estimating \mathbf{H} directly, we can obtain a bilinear parameter structure by imposing that $\mathbf{H} = \mathbf{F}\mathbf{G}'$, in which \mathbf{F} and \mathbf{G} are $k \times P$ and $l \times P$ parameter matrices. The number of columns P is the rank of the estimated matrix \mathbf{H} and equals the dimensionality of the representation. For P large enough, the matrix \mathbf{H} can take any form and can thus approximate the effects in the data as well as possible. Note that, for large k and l , accurately estimating \mathbf{H} requires a large amount of data. Imposing a bilinear parameter structure can then reduce the number of parameters to be estimated, and the remaining parameters can be estimated more precisely.

To represent the results of such a bilinear decomposition in a single figure, we can plot the rows of the matrices \mathbf{F} and \mathbf{G} as points in a P -dimensional space. Figure 1.1 contains an example of such a figure, based on a 3×4 matrix \mathbf{H} . To easily determine the estimated effects, all points are connected to the origin using vectors. The estimated effects can be derived from such a figure in the same way as in principal components analysis biplots. Suppose we are interested in determining element (1,4) of \mathbf{H} . To do so, we may proceed as follows.

1. Find the points in the figure that are associated with row 1 and column 4.
2. Project the vector of column 4 onto the vector of row 1.

Figure 1.1: Hypothetical example of a graphical representation



3. Multiply the length of the projection with the length of the vector of row 1.
4. The resulting length is the estimated effect, which also equals the inner product of the two vectors.

The projections of the vectors of all columns onto the vector of Row 1 are shown in the left panel of Figure 1.1. In the right panel, the estimated effects for Row 1 are shown. The effect of element (1,4) is approximately 0.26.

Although any user of a bilinear decomposition model should be able to implement this stepwise approach, we can often simplify the interpretation of the effects by assigning interpretations to the dimensions of the low-dimensional representation. In that case, the users of a model often do not need to compute the exact estimated effects to understand the results. The bilinear specification using inner products contrasts the distance-based interpretation that is used in multidimensional scaling and other techniques. With a distance-based method, the distances between points in a graphical representation can be used to determine the estimated effects.

In this thesis, we use bilinear decompositions to model the effects of multiple categorical variables. The parameter matrix that is decomposed either represents the effects of a categorical explanatory variable on a categorical dependent variable or the two-way interaction effect of two categorical explanatory variables. We also link dimension

reduction to the methodology of *optimal scaling* (see, for example, Young, 1981; Gifi, 1990; Linting, Meulman, Groenen, & Van der Kooij, 2007). This methodology assigns numeric values to categorical variables in such a way that a criterion function is optimized. We incorporate optimal scaling in some of the models that we develop, and we show that, for some dimension reduction techniques, optimal scaling is equivalent to reducing the dimensionality to 1.

1.3 Segmentation

Segmentation techniques divide a set of respondents or objects into groups. Typically, the goal of such an approach is to make the respondents or objects in a group as similar as possible; the clusters should have internal cohesion and external isolation (Cormack, 1971). Segmentation techniques include cluster analysis and finite mixture modeling. Cluster analysis (see, for example, Sneath, 1957) is the automatic grouping of objects into groups on the basis of properties of the objects or the dissimilarities between them. Cluster analysis comprises both hierarchical methods, such as Ward's method, and non-hierarchical methods, such as k -means clustering. Finite mixture modeling (see, for example, McLachlan & Peel, 2000) can be seen as a model-based version of clustering. In finite mixture modeling, certain (sets of) parameters of a statistical model are allowed to vary among latent classes of respondents/objects. An advantage of finite mixture models is that it is possible to calculate the probability that a respondent or an object belongs to a certain segment. This allows us to determine the uncertainty in the classification, which is usually not possible in cluster analysis.

In this thesis, we apply both finite mixture modeling and more traditional clustering techniques. We use both methodologies for partitioning (that is, clustering in such a way that each object belongs to exactly one cluster) two-mode two-way data matrices; in two-mode partitioning, respondents and variables are clustered simultaneously. We first compare optimization methods for the k -means clustering problem in the case of two-mode data matrices. We then propose a new Bayesian segmentation method for two-mode data matrices, based on finite mixture modeling. Finally, in one of the models in which dimension reduction is applied, finite mixture modeling is also used to account for unobserved heterogeneity in the respondents. We will thus show how dimension reduction and segmentation can be meaningfully combined in a single model.

1.4 Summary

This thesis is divided into two parts. In Part I, the main focus is on representing interactions effects using bilinear decompositions (dimension reduction). Part II focuses on techniques for two-mode partitioning. Parts I and II comprise three and two chapters, respectively. Each chapter contains an abstract and can be read independently of the other chapters. Below, we give a brief overview of each chapter.

In Chapter 2, we consider a general model specification for representing two-way interaction effects in a parsimonious way. We discuss an optimal scaling model for analyzing the content of interaction effects in generalized linear models with any number of categorical predictor variables. This model, which we call the optimal scaling of interactions (OSI) model, is a parsimonious, one-dimensional multiplicative interaction model. We discuss how the model can be used to visually interpret the interaction effects. Several extensions of the one-dimensional model are also explored. Finally, two data sets are used to show how the results of the model can be applied and interpreted. Chapter 2 is based on Van Rosmalen, Koning, and Groenen (2009) and has been published in *Multivariate Behavioral Research*.

In Chapter 3, we consider the topic of response styles (that is, the way respondents use rating scales), which threaten the validity of rating scale responses. We investigate how response style and content of the items affect rating scale responses. We develop a novel model that accounts for different kinds of response styles, content of items, and background characteristics of respondents. By imposing a bilinear parameter structure on a multinomial logit model, the effects on the response behavior of the characteristics of a respondent and the content of an item are graphically distinguished. This approach is combined with finite mixture modeling, yielding two segmentations of the respondents: one for response style and one for item content. This latent-class bilinear multinomial logit (LC-BML) model is applied to the well-known List of Values in a cross-national context. Chapter 3 is based on Van Rosmalen, Van Herk, and Groenen (in press) and has been accepted for publication in the *Journal of Marketing Research*.

In Chapter 4, we apply the modeling principles of Chapter 2 to conjoint analysis data. In conjoint choice experiments, interaction effects of attributes are seldom fully modeled. Modeling all interaction effects tends to require a large number of parameters and a large conjoint choice design. We consider a novel approach for modeling two-way interactions effects, based on the OSI model of Chapter 2. The aim of this approach is to model the two-way interaction effects in choice-based conjoint analysis effectively, while keeping the design size and the number of parameters relatively small. We apply the resulting models to an empirical conjoint choice data set. We also sketch how good choice designs can be constructed for our models. Chapter 4 is joint work with Alex Koning and Patrick Groenen.

In Chapter 5, we compare various optimization strategies for two-mode partitioning. Two-mode partitioning is a relatively new form of clustering that clusters both rows and columns of a data matrix. We consider deterministic two-mode partitioning methods in which a criterion similar to k -means is optimized. A variety of optimization methods have been proposed for this type of problem. However, it is still unclear which method should be used, as various methods may lead to non-global optima. We review and compare several optimization methods for two-mode partitioning. Several known methods are discussed, and a new fuzzy steps method is introduced. The performances of all methods are compared in a large simulation study. Finally, an empirical data set is used to give a practical example

of two-mode partitioning. Chapter 5 is based on Van Rosmalen, Groenen, Trejos, and Castillo (in press) and has been accepted for publication in the *Journal of Classification*.

In Chapter 6, we develop a new Bayesian approach to estimate the parameters of a latent-class model for the joint clustering of both modes of two-mode data matrices. Posterior results are obtained using a Gibbs sampler with data augmentation. Our Bayesian approach has three advantages over existing methods. First, we are able to do statistical inference on the model parameters, which would not be possible using frequentist estimation procedures. In addition, the Bayesian approach allows us to provide statistical criteria for determining the optimal numbers of clusters. Finally, our Gibbs sampler has fewer problems with local optima in the likelihood function and empty classes than the EM algorithm used in a frequentist approach. We apply the Bayesian estimation method of the latent-class two-mode clustering model to two empirical data sets. The first data set is the Supreme Court voting data set of Doreian, Batagelj, and Ferligoj (2004). The second data set comprises the roll call votes of the United States House of Representatives in 2007. For both data sets, we show how the results can provide useful insight into the data. Chapter 6 is joint work with Bram van Dijk and Richard Paap.

In the Discussion, the thesis is ended with a short overview of the findings and a discussion of the limitations and avenues for future research.

Part I

Bilinear Decomposition Models

Chapter 2

Optimal Scaling of Interaction Effects in Generalized Linear Models*

Abstract

Multiplicative interaction models, such as Goodman's RC(M) association models, can be a useful tool for analyzing the content of interaction effects. However, most models for interaction effects are only suitable for data sets with two or three predictor variables. Here, we discuss an optimal scaling model for analyzing the content of interaction effects in generalized linear models with any number of categorical predictor variables. This model, which we call the optimal scaling of interactions (OSI) model, is a parsimonious, one-dimensional multiplicative interaction model. We discuss how the model can be used to visually interpret the interaction effects. Several extensions of the one-dimensional model are also explored. Finally, two data sets are used to show how the results of the model can be applied and interpreted. The first data set is based on the Student/Teacher Achievement Ratio (STAR) project and is used to investigate the effects of class size on the performance of primary school students. The second data set comprises four questions from the 1994 General Social Survey (Davis & Smith, 1996) on attitudes of the labor roles of women.

2.1 Introduction

The analysis of data sets with categorical variables often requires studying interaction effects between these variables. If the relationship between the response variable and the predictor variables is linear, the interaction effects can be studied using analysis of variance (ANOVA). If this relationship is nonlinear, models from the class of generalized linear models (GLMs) are often used. Generalized linear models, which are extensions of the general linear model, can be used to model various kinds of relationships between variables

*This chapter is based on Van Rosmalen et al. (2009).

and can account for nonnormality and nonlinearity. GLMs have been thoroughly described in Nelder and Wedderburn (1972) and McCullagh and Nelder (1989). Including all two-way interactions in a GLM may often require the estimation of a large number of parameters, especially if there are many categorical variables and if these variables have many levels. Because of the large number of parameters, the estimated individual interaction effects are often not interpreted, and only their combined effect is tested for significance.

Models for representing interaction effects parsimoniously have been proposed before, especially for the case of two categorical predictor variables. For example, Goodman (1981) proposed row-column (RC(M)) association models for the analysis of two-way contingency tables. RC(M) association models can be considered as a special case of generalized additive main effects and multiplicative interaction (GAMMI) models, which are mainly used in agricultural science (see, for example, Van Eeuwijk, 1995, 1996). Similar models were proposed by Gabriel (1998). Algorithmic approaches for these kinds of models were discussed by De Falguerolles and Francis (1992). These types of models often use *biplots* (see, for example, Gower & Hand, 1996) to represent interaction effects between two variables, by plotting the categories of both variables in a two-dimensional space. Specialized models for the case of three categorical predictor variables also exist (see, for example, Clogg, 1982; C. J. Anderson, 1996; Choulakian, 1996; Siciliano & Mooijaart, 1997; Wong, 2001).

For the case of more than three predictor variables, Groenen and Koning (2006) proposed the interaction decomposition model. They sketched an outline of an algorithm for parameter estimation in this model and gave graphical representations of their results. For log-linear analysis (a special case of generalized linear modeling) with more than three variables, a variety of models were proposed by C. J. Anderson and Vermunt (2000). In their article, the interaction effects are parsimoniously modeled by assuming the presence of latent variables.

In this chapter, we use the methodology of *optimal scaling* for modeling interaction effects parsimoniously. Optimal scaling (see, for example, Young, 1981; Gifi, 1990; Linting et al., 2007) is a methodology originating from psychometrics that assigns numeric values to categorical variables in an optimal way. The history of optimal scaling can also be traced back to correspondence analysis (Richardson & Kuder, 1933; Fisher, 1940; Guttman, 1941); see Benzécri (1982) for an overview and further details. Gifi (1990) discusses a host of multivariate analysis techniques (multiple correspondence analysis, nonlinear principal components analysis, generalized nonlinear canonical correlation analysis, etc.) all having in common that the variables are categorical and that some optimal recoding is being done. That is, the categories of the original categorical variables are replaced by their so-called category quantifications, and from then on the variables are considered to be quantitative variables. The word optimal refers to the fact that these category quantifications are chosen in such a way that they help optimize the criterion. Optimal scaling has also been applied in a regression context, with techniques such as MONANOVA (Kruskal, 1965), ADDALS (De Leeuw, Young, & Takane, 1976), MORALS (Young, De Leeuw, & Takane, 1976), ACE (Breiman & Friedman, 1985), and generalized additive models (Hastie & Tibshirani, 1990).

We describe a model for representing interaction effects in generalized linear models with any number of categorical predictor variables in a clear and parsimonious way. Typically, only main effects and two-way interaction effects are empirically relevant, and higher-way interaction effects are often not required in empirical applications. Therefore, we primarily focus on modeling two-way interaction effects. The main assumption of our model is that interactions between categorical predictor variables can be modeled using continuous predictor variables on which we have partial knowledge. This assumption leads to a model in which the estimated parameters may be interpreted in terms of an optimal scaling of the categorical predictor variables. Because of this assumption, we refer to our model as the *optimal scaling of interactions* (OSI) model. The OSI model requires a number of parameters that is only linear in the total number of categories of the categorical variables and quadratic in the number of variables. By contrast, a standard two-way interaction model requires a number of parameters that is quadratic in the total number of categories of the variables.

With our model, we construct one-dimensional graphical representations of the interaction effects, which can help interpret these effects. As a one-dimensional model may be restrictive in some cases, we also explore multidimensional extensions of our model. Furthermore, we briefly show how the idea of the OSI model could be extended to model three-way and higher-way interaction effects. Finally, two empirical data sets are used to show how the results of the OSI model and its multidimensional extensions can be applied and interpreted. The first data set is based on the Student/Teacher Achievement Ratio (STAR) project in Tennessee and is used to investigate the effects of class size on the academic performance of primary school students. The second data set comprises four questions from the 1994 General Social Survey (Davis & Smith, 1996) on attitudes of the labor roles of women. To fit the OSI model, we have written specialized software in the matrix programming language MATLAB. This software package includes the example data sets discussed in this paper and is available at <http://people.few.eur.nl/groenen>. The Optimization Toolbox in MATLAB is required to use this software.

The outline of this chapter is as follows. In the next section, we introduce some notation and our optimal scaling of interactions model. In Section 2.3, we discuss several extensions of this model. Section 2.4 describes the application of our model to two empirical data sets. The final section summarizes our findings.

2.2 Optimal Scaling of Interactions Model

The model we propose is based on generalized linear modeling (see, for example, Nelder & Wedderburn, 1972; McCullagh & Nelder, 1989). The observations $z_i, i = 1 \dots n$ are assumed to be independently distributed with $E(z_i) = \mu_i$. Each z_i has a distribution in the exponential family, with probability density function given by

$$f(z_i; \theta, \phi) = \exp \left\{ \frac{z_i \theta - b(\theta)}{a(\phi)} + c(z_i, \phi) \right\}, \quad (2.1)$$

where $a(\cdot)$, $b(\cdot)$, and $c(\cdot)$ are given functions, θ is the so-called natural parameter, and ϕ is the dispersion parameter. The exponential family includes the normal, Poisson, binomial, gamma, and inverse Gaussian distributions; see McCullagh and Nelder (1989) for a comprehensive overview of possible distributions. The systematic part of a generalized linear model consists of a predictor η_i , which typically is a linear function of the predictor variables and the parameters. A *link function* $h(\cdot)$ relates the linear predictor η_i to the response variable according to

$$\eta_i = h(\mu_i). \quad (2.2)$$

Common link functions are the identity, inverse, logarithm, and logit functions. In practice, one often uses *canonical links*, such as a logarithm link in combination with a Poisson error distribution. The canonical links can be derived from the theory of sufficient statistics.

In this paper, we aim to model interaction effects using a generalized linear modeling framework. Suppose continuous predictor variables \mathbf{x}_j , $j = 1, \dots, m$ are known. Then, the main effects and the two-way interaction effects of these variables can be modeled according to

$$\eta_i = c + \sum_{j=1}^m b_j x_{ij} + \sum_{j=1}^{m-1} \sum_{l=j+1}^m w_{jl} s_{jl} x_{ij} x_{il}, \quad (2.3)$$

where c is a constant term, b_j is the main effect of variable \mathbf{x}_j , and s_{jl} is the size of the interaction effect of variables \mathbf{x}_j and \mathbf{x}_l . The $m \times m$ upper-triangular matrix $\mathbf{W} = (w_{jl})$ specifies which interaction effects are to be estimated in the GLM, with $w_{jl} = 1$ if the interaction between predictor variables j and l is taken into account and $w_{jl} = 0$ otherwise. The purpose of the matrix \mathbf{W} is to increase the flexibility of the model; if a researcher believes that there is no interaction effect between variables j and l , he or she can set $w_{jl} = 0$ in advance. The diagonal elements of \mathbf{W} are not used, as these elements refer to main effects that are already modeled by the second term in (2.3).

Here, we restrict ourselves to cases in which all predictor variables are categorical instead of continuous, so that (2.3) cannot be used directly. The central assumption of our model is that interaction effects between the categorical predictor variables can be modeled in approximately the same way as interaction effects between continuous predictor variables. To do so, we apply the idea of optimal scaling (see, for example, Gifi, 1990) to the categorical predictor variables for modeling their interaction effects, hence the name optimal scaling of interactions (OSI) model.

To be able to introduce optimal scaling in model (2.3), we need some notation. Let there be m categorical predictor variables with each variable having k_j categories. To code the categorical predictor variables, we use indicator matrices \mathbf{G}_j with rows \mathbf{g}_{ij} of length k_j ; element l of \mathbf{g}_{ij} has value 1 if observation z_i belongs to category l of predictor variable j and 0 otherwise. In the OSI model, we use separate optimally scaled variables for the main effects and the interaction effects, so that

$$\eta_i = c + \sum_{j=1}^m b_j r_{ij} + \sum_{j=1}^{m-1} \sum_{l=j+1}^m w_{jl} s_{jl} q_{ij} q_{il}, \quad (2.4)$$

where \mathbf{r}_j is the optimally scaled variable for the main effect of variable j , and \mathbf{q}_j is the optimally scaled variable that is used for the interaction effects of variable j . In principle, one could also use the same optimally scaled variables for both the main effects and the interaction effects, so that $\mathbf{r}_j = \mathbf{q}_j$. However, we find this approach too restrictive, and we therefore do not explore it here.

The values of the continuous, optimally scaled predictor variables \mathbf{r}_j and \mathbf{q}_j are not known in our model and need to be estimated. The \mathbf{r}_j s are related to the categorical predictor variables according to $\mathbf{r}_j = \mathbf{G}_j \mathbf{a}_j$, where \mathbf{a}_j is a $k_j \times 1$ parameter vector that contains the category quantifications for the main effects of variable j . The \mathbf{q}_j s are constructed similarly as $\mathbf{q}_j = \mathbf{G}_j \mathbf{y}_j$, with \mathbf{y}_j a $k_j \times 1$ parameter vector that contains the category quantifications for the interaction effects of variable j . Instead of (2.4), the OSI model can also be described as

$$\eta_i = c + \sum_{j=1}^m b_j \mathbf{g}'_{ij} \mathbf{a}_j + \sum_{j=1}^{m-1} \sum_{l=j+1}^m w_{jl} s_{jl} \mathbf{g}'_{ij} \mathbf{y}_j \mathbf{y}'_{il} \mathbf{g}_{il}. \quad (2.5)$$

In this way, the main effects appear in the same manner as in an ordinary GLM with categorical predictor variables. For the interaction effects, the OSI model uses a multiplicative specification that is relatively parsimonious. The parameter vector \mathbf{y}_j reflects the content of the interaction effects of variable j . The goal of the parameter s_{jl} is to estimate the size of the interaction effect between variables j and l . Therefore, we refer to s_{jl} as a *scaling factor*.

Once the category quantifications \mathbf{a}_j and \mathbf{y}_j are estimated and hence are known, the optimally scaled variables can be treated as ordinary continuous variables. Then, the parameters b_j and s_{jl} can be computed using the ordinary GLM in (2.3). Because the \mathbf{q}_j s are not known, and the way the interactions appear in (2.4), the predictor η_i is a nonlinear function of the model parameters. Therefore, the OSI model is not an ordinary GLM. As the \mathbf{q}_j s are restricted by $\mathbf{q}_j = \mathbf{G}_j \mathbf{y}_j$, and \mathbf{y}_j can be estimated from the data, the OSI model can be seen as a GLM with optimal scaling of the categorical predictor variables.

Several parameter constraints, including location and scale constraints, are required for model identification. We use the following parameter constraints, which originate from the optimal scaling methodology and differ from the constraints typically used in multiplicative interaction models. We impose that the optimally scaled variables \mathbf{r}_j and \mathbf{q}_j have mean zero and variance one. For the interaction effects, this results in the location constraints $\mathbf{1}' \mathbf{q}_j = \sum_{i=1}^n \mathbf{g}'_{ij} \mathbf{y}_j = 0$ and the scale constraints $\mathbf{q}'_j \mathbf{q}_j = \sum_{i=1}^n \mathbf{g}'_{ij} \mathbf{y}_j \mathbf{y}'_{il} \mathbf{g}_{il} = n$, where $\mathbf{1}$ denotes a vector of ones of length n . For the main effects, we impose that $\mathbf{1}' \mathbf{r}_j = \sum_{i=1}^n \mathbf{g}'_{ij} \mathbf{a}_j = 0$ and $\mathbf{r}'_j \mathbf{r}_j = \sum_{i=1}^n \mathbf{g}'_{ij} \mathbf{a}_j \mathbf{a}'_{il} \mathbf{g}_{il} = n$. In addition, the value of the scaling factor s_{jl} cannot be estimated if $w_{jl} = 0$; therefore, we set $s_{jl} = 0$ whenever $w_{jl} = 0$. Finally, simultaneously changing the signs of the elements of \mathbf{y}_j and s_{jl} for all l does not affect the predictor η_i . To improve the interpretability of the model parameters, we simultaneously reflect the \mathbf{y}_j s and the scaling factors s_{jl} in such a way, that the sum of the estimated scaling factors is maximized. To do so, each of the 2^m possible combinations of reflections of the \mathbf{y}_j s is

considered, and the combination that maximizes $\sum_{j=1}^{m-1} \sum_{l=j+1}^m s_{jl}$ is used to interpret the results of the model.

Additional parameter constraints may be required if few observations are available, or if $w_{jl} = 0$ for many values of j and l . Whether such additional constraints are necessary can be determined empirically, for example, by checking whether the estimated parameters are unique maximizers of the log-likelihood function. This can be done by estimating the model parameters multiple times using randomly chosen starting values; if no additional parameter constraints are necessary, the estimated parameters must be the same in every instance.

The categorical predictor variables can have either a nominal or an ordinal measurement level. For ordinal predictor variables, it is possible to impose their ordering on \mathbf{y}_j . However, imposing such ordinality constraints may not be appropriate, as the interaction effects can reflect nonmonotonic relations between the predictor variables and the response variable. Therefore, we do not impose the ordering of ordinal predictor variables on the model parameters. The OSI model can also be extended to include continuous predictor variables, for example, by modeling the \mathbf{y}_j s as (spline) transformations of these continuous variables. In that case, one again needs to consider whether such transformations need to be monotonic. More information on splines and other nonlinear transformations is given in Gifi (1990).

As the OSI model is not an ordinary GLM, a special algorithm for parameter estimation is needed. In our implementation, the parameters are estimated by maximizing the log-likelihood function using the BFGS quasi-Newton optimization routine in the MATLAB Optimization Toolbox (version 3.0.4). To ensure that the global maximum of the log-likelihood function is found, this optimization routine should be run several multiple times with randomly chosen starting values. The parameters of the OSI model are not identified without the parameter constraints that were discussed above. To be able to compute standard errors, we used the reparameterization of the model parameters that is discussed in Appendix 2.A; in this reparameterization, the parameters are identified without imposing any constraints. Standard errors of the estimated parameters are computed using the negative inverse of the Hessian (the matrix of second-order partial derivatives of the log-likelihood function), evaluated at the final parameter estimates in the reparameterized model.

The OSI model has several relationships with existing models for interaction effects. A standard GLM with two-way interaction effects can be described as

$$\eta_i = c + \sum_{j=1}^m b_j \mathbf{g}'_{ij} \mathbf{a}_j + \sum_{j=1}^{m-1} \sum_{l=j+1}^m w_{jl} \mathbf{g}'_{ij} \mathbf{B}_{jl} \mathbf{g}_{il}, \quad (2.6)$$

where \mathbf{B}_{jl} is a $k_j \times k_l$ parameter matrix of interaction effects between variables j and l . The OSI model can be obtained from the full two-way interaction GLM by imposing that each interaction matrix \mathbf{B}_{jl} equals a matrix $\bar{\mathbf{B}}_{jl}$ with

$$\bar{\mathbf{B}}_{jl} = s_{jl} \mathbf{y}_j \mathbf{y}'_l. \quad (2.7)$$

Thus, the OSI model implicitly approximates each matrix of interaction effects \mathbf{B}_{jl} by a matrix of rank one.

The OSI model also resembles a few multiplicative interaction models that have been proposed previously. If there are only two categorical predictor variables, the OSI model is equivalent with the generalized additive main effects and multiplicative interaction models discussed by Van Eeuwijk (1995, 1996), which are a generalization of the RC association models discussed by Goodman (1981). For log-linear analysis (that is, generalized linear modeling with link function $\eta_i = \log(\mu_i)$ and a Poisson probability distribution), the OSI model is equivalent with equation (20) of C. J. Anderson and Vermunt (2000); they interpreted the \mathbf{y}_{js} as latent variables. For the special case of log-linear analysis with three predictor variables, the OSI model can be obtained by imposing consistent score restrictions in equation (4.9) in Clogg (1982); see also Wong (2001).

2.3 Extensions

The one-dimensional OSI model (2.5) is relatively straightforward to interpret. However, this model may yield an inadequate fit for some data sets. In that case, a less restrictive model may be considered. Here, we discuss several ways to generalize the one-dimensional OSI model to a multidimensional model, which should provide a better fit. We also discuss what identification constraints are required for such models and how they are related to previously proposed models. Finally, we briefly explore how the idea of the OSI model can be used to model higher-way interaction effects.

2.3.1 Multidimensional extensions

The most natural generalization of the one-dimensional OSI model consists of allowing for multiple optimally scaled variables per categorical predictor variable. We use this approach for our general multidimensional model, so that it is given by

$$\eta_i = c + \sum_{j=1}^m b_j r_{ij} + \sum_{j=1}^{m-1} \sum_{l=j+1}^m w_{jl} \sum_{p=1}^P s_{jlp} q_{ijp} q_{ilp}, \quad (2.8)$$

where q_{ijp} is the score of person i on the p -th optimally scaled variable for categorical variable j , and s_{jlp} is the coefficient of the p -th optimally scaled variable for the interaction between categorical variables j and l . By writing this model in terms of the categorical predictor variables, it can be also described as

$$\eta_i = c + \sum_{j=1}^m b_j \mathbf{g}'_{ij} \mathbf{a}_j + \sum_{j=1}^{m-1} \sum_{l=j+1}^m w_{jl} \mathbf{g}'_{ij} \mathbf{Y}_j \mathbf{S}_{jl} \mathbf{Y}'_l \mathbf{g}_{il}, \quad (2.9)$$

where \mathbf{Y}_j and \mathbf{S}_{jl} are matrices of sizes $k_j \times P$ and $P \times P$ respectively, with P the dimensionality of the model. The matrix \mathbf{S}_{jl} is constrained to be diagonal. For $P = 1$, (2.9) simplifies to the one-dimensional OSI model.

For the general multidimensional model (2.9), we impose similar location constraints as for the one-dimensional model, so that $\mathbf{1}'\mathbf{r}_j = \sum_{i=1}^n \mathbf{g}'_{ij}\mathbf{a}_j = 0$ and $\mathbf{1}'\mathbf{Q}_j = \sum_{i=1}^n \mathbf{g}'_{ij}\mathbf{Y}_j = \mathbf{0}'$, in which $\mathbf{Q}_j = \mathbf{G}_j\mathbf{Y}_j$ contains the P optimally scaled variables for variable j , and $\mathbf{1}$ and $\mathbf{0}$ are vectors of appropriate length. For the main effects, the scale constraints are $\mathbf{r}'_j\mathbf{r}_j = \sum_{i=1}^n \mathbf{g}'_{ij}\mathbf{a}_j\mathbf{a}'_j\mathbf{g}_{ij} = n$; for the interaction effects, we require that $\mathbf{Q}'_j\mathbf{Q}_j = \mathbf{Y}'_j\mathbf{G}'_j\mathbf{G}_j\mathbf{Y}_j$ has diagonal elements equal to n . In addition, we must set $s_{jlp} = 0$ for every $w_{jl} = 0$. Furthermore, just as in the one-dimensional OSI model, we change the signs of the columns of \mathbf{Y}_j and, correspondingly, s_{jlp} in such a way that the elements of $\sum_{j=1}^{m-1} \sum_{l=j+1}^m \mathbf{S}_{jl}$ are maximized. Finally, in a multidimensional model, it is often convenient to ensure that the amount of explained variation decreases with the dimension, so that the first dimension is the most important one. For the general multidimensional model (2.9), we accomplish this by requiring that the diagonal elements of $\sum_{j=1}^m \sum_{l=1}^m |\mathbf{S}_{jl}|$ are decreasing.

Determining the number of degrees of freedom in model (2.9) may be somewhat difficult, as the degrees of freedom are influenced by characteristics of both the model and the research design (that is, the values of the predictor variables in the data set). However, if all interaction terms are present in the model (that is $w_{jl} = 1$ for all $j < l$), and the number of observations is sufficiently large, the number of parameters in the general multidimensional model equals $1 + Pm(m-1)/2 + (1+P)\sum_{j=1}^m k_j$, and the total number of parameter restrictions is $(P+2)m$. In that case, the number of degrees of freedom required by this model is given by

$$df = 1 + (1+P)\sum_{j=1}^m k_j - (1+2P)m + \frac{1}{2}Pm(m-1). \quad (2.10)$$

To represent the results of (2.9) graphically in a way that is easy to interpret, we can construct a biplot for each interaction term separately. To do so, one may calculate a compact singular value decomposition of $\tilde{\mathbf{B}}_{jl} = w_{jl}\mathbf{Y}_j\mathbf{S}_{jl}\mathbf{Y}'_l$, so that $\mathbf{U}\Sigma\mathbf{V}' = w_{jl}\mathbf{Y}_j\mathbf{S}_{jl}\mathbf{Y}'_l$, where Σ is $P \times P$ diagonal matrix, and \mathbf{U} and \mathbf{V} are orthogonal (so that $\mathbf{U}'\mathbf{U} = \mathbf{V}'\mathbf{V} = \mathbf{I}$). Matrices \mathbf{U} , Σ , and \mathbf{V} that meet these requirements must exist, as the rank of $\tilde{\mathbf{B}}_{jl}$ cannot be greater than P . A biplot can then be constructed by plotting $\mathbf{U}\Sigma^{1/2}$ and $\mathbf{V}\Sigma^{1/2}$ simultaneously in one figure.

Restricted Multidimensional Models

For some data sets, the general multidimensional model (2.9) may require prohibitively many parameters, leading to instability of the estimated parameters. In addition, interpreting the estimated interaction effects using graphical representations may be difficult if there are many predictor variables. In such cases, alternative generalizations of the one-dimensional OSI model with fewer parameters can be considered. Here, we discuss three such generalizations, which consist of restricting the parameters in \mathbf{S}_{jl} to be equal for each interaction term or for each dimension.

First, we can restrict the elements of \mathbf{S}_{jl} to be equal across dimensions, which leads to

the model

$$\eta_i = c + \sum_{j=1}^m b_j \mathbf{g}'_{ij} \mathbf{a}_j + \sum_{j=1}^{m-1} \sum_{l=j+1}^m w_{jl} s_{jl} \mathbf{g}'_{ij} \mathbf{Y}_j \mathbf{Y}'_l \mathbf{g}_{il}. \quad (2.11)$$

Here, the scale constraint that $\mathbf{Q}'_j \mathbf{Q}_j$ (with $\mathbf{Q}_j = \mathbf{G}_j \mathbf{Y}_j$) must have diagonal elements equal to n cannot be imposed without loss of generality. Instead, we impose that $\text{tr}(\mathbf{Q}'_j \mathbf{Q}_j) = n$, in which $\text{tr}(\cdot)$ refers to the trace operator (the sum of the diagonal elements). For this model, the scale constraints for the main effects and the location constraints are identical to those of the general multidimensional model. As for any orthogonal rotation matrix \mathbf{T} , $\mathbf{Y}_j \mathbf{Y}'_l = \mathbf{Y}_j \mathbf{T} \mathbf{T}' \mathbf{Y}'_l$, simultaneously rotating the matrices \mathbf{Y}_j does not alter the values of η_i . Therefore, rotation restrictions are also required for this model. We require that $\sum_{j=1}^m \mathbf{Q}'_j \mathbf{Q}_j = \sum_{j=1}^m \mathbf{Y}'_j \mathbf{G}'_j \mathbf{G}_j \mathbf{Y}_j$ is diagonal, which yields $P(P-1)/2$ parameter restrictions on the \mathbf{Y}_j s.

A second type of restricted model can be obtained by imposing that $\mathbf{S}_{jl} = \mathbf{S}$ for every interaction term, so that

$$\eta_i = c + \sum_{j=1}^m b_j \mathbf{g}'_{ij} \mathbf{a}_j + \sum_{j=1}^{m-1} \sum_{l=j+1}^m w_{jl} \mathbf{g}'_{ij} \mathbf{Y}_j \mathbf{S} \mathbf{Y}'_l \mathbf{g}_{il}, \quad (2.12)$$

where \mathbf{S} is diagonal. Again, a special scale constraint is required for the interaction effects; here, we impose that $\sum_{j=1}^m \mathbf{Q}'_j \mathbf{Q}_j$ has diagonal elements equal to mn . The other scale and location constraints are identical to those of the general multidimensional model. In this model, no rotation constraints can be imposed without loss of generality. For log-linear modeling, this model coincides with equation (24) of C. J. Anderson and Vermunt (2000), though the parameter restrictions used in their article are different.

Finally, one may restrict the parameter matrices \mathbf{S}_{jl} to be equal for every interaction term and for every dimension (so that $\mathbf{S}_{jl} = \mathbf{I}$), essentially removing these parameters from the model. In that case, the model is

$$\eta_i = c + \sum_{j=1}^m b_j \mathbf{g}'_{ij} \mathbf{a}_j + \sum_{j=1}^{m-1} \sum_{l=j+1}^m w_{jl} \mathbf{g}'_{ij} \mathbf{Y}_j \mathbf{Y}'_l \mathbf{g}_{il}, \quad (2.13)$$

and we obtain the interaction decomposition model proposed by Groenen and Koning (2006). Here, the sizes of the interaction effects are determined by the parameter matrices \mathbf{Y}_j . The same location and rotation constraints as in (2.11) can be imposed; however, no scale constraints can be imposed on the \mathbf{Y}_j s without loss of generality. If the estimated interactions effects have similar sizes, the results of (2.13) can be conveniently visualized using a biplot in which the \mathbf{Y}_j s are simultaneously plotted in a P -dimensional space. However, if various interaction effects differ in size significantly, the visualization may break down, and the model may fit poorly.

Model (2.13) is a special case of (2.12), which can be obtained by setting $s_p = 1$ for all p ; the results of these two models may appear to be almost identical, though they are not equivalent. Model (2.13) can be obtained from (2.12) by multiplying the elements of

Table 2.1: Maximum degrees of freedom associated with various models

Model	Degrees of freedom
Main effects only GLM	$1 + \sum_{j=1}^m k_j - m$
One-dimensional OSI model (2.5)	$1 + 2 \sum_{j=1}^m k_j - 3m + m(m-1)/2$
General multidim. model (2.9)	$1 + (1+P) \sum_{j=1}^m k_j - (1+2P)m + Pm(m-1)/2$
Restricted multidim. model (2.11)	$1 + (1+P) \sum_{j=1}^m k_j - (2+P)m + m(m-1)/2 - P(P-1)/2$
Restricted multidim. model (2.12)	$1 + (1+P) \sum_{j=1}^m k_j - (1+P)m$
Restricted multidim. model (2.13)	$1 + (1+P) \sum_{j=1}^m k_j - (1+P)m - P(P-1)/2$
Full two-way interaction GLM	$1 + \sum_{j=1}^m k_j - m + \sum_{j=1}^{m-1} \sum_{l=j+1}^m (k_j - 1)(k_l - 1)$

the p -th column of \mathbf{Y}_j with $\sqrt{s_p}$, but this is only possible if all s_p are nonnegative, so that (2.12) is more flexible than (2.13).

Experimentation with these multidimensional models suggests that unique parameter estimates that maximize the log-likelihood function may not always exist. In that case, the parameter estimates that optimization algorithms produce may fail to converge to finite values and could approach infinity instead. If the parameters of a multidimensional model are not uniquely identified, another type of model should be considered. This effect does not appear to occur in the one-dimensional OSI model and also does not occur for all data sets; for some data sets, the parameters of multidimensional models can be uniquely estimated. Therefore, we still believe that the multidimensional extensions described above can be useful.

Table 2.1 gives an overview of the maximum numbers of degrees of freedom for a number of models, based on the location, scale, and rotation constraints that were described previously. The values in this table are upper bounds on the actual degrees of freedom; they can only be attained if all interaction terms are taken into account, and both the number of observations and the number of variables are large enough.

2.3.2 Multilinear models for higher-way interactions

In some cases, modeling two-way interactions is not enough, and higher-way interactions should also be taken into account. The idea of the OSI model can also be extended to model three-way and higher-way interactions. Here, we sketch how the OSI model can be extended to handle three-way interactions.

Modeling three-way interactions only makes sense if all main effects and two-way interaction effects are already taken into account. Applying the idea of the OSI model for modeling three-way interactions results in optimally scaled variables for the three-way interactions. If we add the three-way interaction effects of such optimally scaled variables to the full two-way interactions model (2.6), we obtain

$$\eta_i = c + \sum_{j=1}^m b_j \mathbf{g}'_{ij} \mathbf{a}_j + \sum_{j=1}^{m-1} \sum_{l=j+1}^m \mathbf{g}'_{ij} \mathbf{B}_{jl} \mathbf{g}_{il} + \sum_{j=1}^{m-2} \sum_{l=j+1}^{m-1} \sum_{r=l+1}^m w_{jlr} s_{jlr} q_{ij} q_{il} q_{ir}, \quad (2.14)$$

in which \mathbf{q}_j is the optimally scaled variable for modeling the three-way interaction effects of variable j , and s_{jlr} is the scaling factor for the three-way interaction between variables j , l , and r . The constants w_{jlr} determine which three-way terms are modeled; $w_{jlr} = 1$ if the three-way interaction effect between variables j , l , and r is taken into account and $w_{jlr} = 0$ otherwise. Writing this model in terms of the category quantifications \mathbf{y}_j yields

$$\eta_i = c + \sum_{j=1}^m b_j \mathbf{g}'_{ij} \mathbf{a}_j + \sum_{j=1}^{m-1} \sum_{l=j+1}^m \mathbf{g}'_{ij} \mathbf{B}_{jl} \mathbf{g}_{il} + \sum_{j=1}^{m-2} \sum_{l=j+1}^{m-1} \sum_{r=l+1}^m w_{jlr} s_{jlr} \mathbf{g}'_{ij} \mathbf{y}_j \mathbf{g}'_{il} \mathbf{y}_l \mathbf{g}'_{ir} \mathbf{y}_r. \quad (2.15)$$

For each three-way interaction term, this model performs a decomposition that is similar to the one-dimensional version of the PARAFAC/CANDECOMP model that was proposed by Harshman (1970) and Carroll and Chang (1970). For log-linear analysis with three variables, this model is equivalent to equation (3) in Siciliano and Mooijaart (1997) and equation (17) in Wong (2001). It is possible to construct multidimensional versions of this model. In addition, the idea of optimal scaling could be applied to four-way and higher-way interaction effects. However, we believe such interactions effects rarely need to be modeled in practice.

2.4 Empirical Applications

To determine how the OSI model performs in practice, we apply it to two empirical data sets. We also compare its usefulness with other models and show how the interaction effects can be visually represented and interpreted.

2.4.1 STAR Data Set

The first data set we use is based on the STAR data set, which can be found in the “Ecdat” package in the R programming language. This data set contains the results of 5,748 Tennessee primary school students on tests of math and reading skills. The data were collected as a part of the Student/Teacher Achievement Ratio (STAR) project (see <http://www.heros-inc.org/star.htm> for additional information). This project investigates the effects of class size on the performance of primary school students. Each student was assigned to either a small class (13 to 17 students per teacher), a regular size class (22 to 25 students per teacher), or a regular-with-aide class (22 to 25 students with a full-time teacher’s aide). The data set also contains personal background characteristics of the students, the level of experience of the teacher, and the school at which the test was taken.

The aim is to explain the results of the math skills test using six categorical predictor variables that are also in the STAR data set. We will focus on the two-way interaction effects of these predictor variables.

- Class size: A categorical predictor with levels “small”, “regular”, and “regular with aide”

Table 2.2: Residual degrees of freedom, log-likelihood values, and values of the Akaike information criterion of various models for the STAR data set

Model	Residual df	Log-likelihood	AIC
GLM with main effects only	5659	-29,483.34	59,144.68
OSI model	5647	-29,450.89	59,103.78
GLM with all two-way categorical interactions	5624	-29,433.69	59,115.38

- Teaching experience: A categorical predictor with levels “< 5 years”, “5-9 years”, “10-14 years”, “15-19 years”, and “> 19 years”
- Gender: A categorical predictor with levels “boy” and “girl”
- Race: A categorical predictor with levels “white” and “black”
- Free lunch: A categorical predictor with levels “Free lunch” and “No free lunch”
- School id: A categorical predictor with 79 levels, which identifies the school at which the test was taken.

The effects of the variables gender and race are combined, so that a new predictor variable (denoted by “Gender, race”) with four levels is obtained. We do not take the interaction effects of “School id” into account in our analysis, as the levels of this factor have no meaning to the reader; only the main effects of “School id” are modeled. In addition, “School id” is modeled as a random factor (the schools used in the study are a sample of the population of schools in Tennessee), whereas all other predictor variables are fixed factors. The math score is a continuous variable and ranges from 320 (worst performance) to 626 (best performance), with an average score of 486. As the response variable is continuous and approximately normally distributed, generalized linear modeling with an identity link and a normal error distribution (which is analysis of variance) seems most appropriate.

Table 2.2 contains the results for the one-dimensional OSI model, a standard GLM with only main effects, and a standard GLM with full two-way interaction effects. From this table, we can observe that the OSI model accounts for most of the interaction effects, as the difference in log-likelihood between the OSI model and a full two-way interaction model is relatively small. Therefore, a full two-way interaction model does not seem necessary. To formally choose among these models, we cannot use a likelihood ratio test, as the associated test statistic is generally not asymptotically chi-square distributed. This is because the parameters of the category quantifications in the OSI model are not identified under the null hypothesis that there are no interaction effects (see Davies, 1977). Instead of a likelihood ratio test, we use the often-used Akaike information criterion (AIC) to determine which model should be preferred. We note that the AIC should be regarded as a ‘heuristic figure of merit’ in this case, as the regularity conditions used to derive the AIC are similar to the conditions on which the chi-square distribution of the likelihood ratio test statistic is based. The OSI model has a lower value for the AIC than either a

Table 2.3: ANOVA table with all two-way interactions for STAR data set

Source	Type III sum sq.	d.f.	Mean sq.	F	p -value	Partial η^2
School id	2,230,156	78	28,591.7	17.04	0.000	0.191
Class size	50,189	2	25,094.3	14.96	0.000	0.005
Teaching exp	16,425	4	4,106.3	2.45	0.044	0.002
Gender, race	80,821	3	26,940.3	16.06	0.000	0.008
Free lunch	158,946	1	158,945.8	11.53	0.001	0.017
Class size \times Teaching exp	71,125	8	8,890.6	5.30	0.000	0.007
Class size \times Gender, race	16,631	6	2,771.7	1.65	0.129	0.002
Class size \times Free lunch	209	2	104.6	0.06	0.940	0.000
Teaching exp \times Gender, race	47,845	12	3,987.1	2.38	0.005	0.005
Teaching exp \times Free lunch	1,737	4	434.3	0.26	0.904	0.000
Gender, race \times Free lunch	23,268	3	7,755.9	4.62	0.003	0.002
Error	9,436,513	5,624	1,677.9			
Total	13,115,339	5,747				

Table 2.4: Main effects of OSI model for STAR data set

Category	Main effect	Category	Main effect
<i>Class size</i>		<i>Gender, race</i>	
Small	6.423	White boy	3.451
Regular with aide	-2.442	White girl	8.325
Regular	-3.106	Black boy	-16.456
		Black girl	-7.912
<i>Teaching experience</i>		<i>Free lunch</i>	
< 5 years	-0.479	No free lunch	9.240
5 – 9 years	-1.483	Free lunch	-9.899
10 – 14 years	-0.167		
15 – 19 years	3.270		
> 19 years	3.997		

model with only main effects or a full two-way interaction model, so that the OSI model is preferred by this criterion. Table 2.3 gives the ANOVA table for the model with full two-way interactions. This table shows that the interactions of “Teaching experience” with “Class size” and “Gender, race” are relatively large and statistically significant at the 5% level.

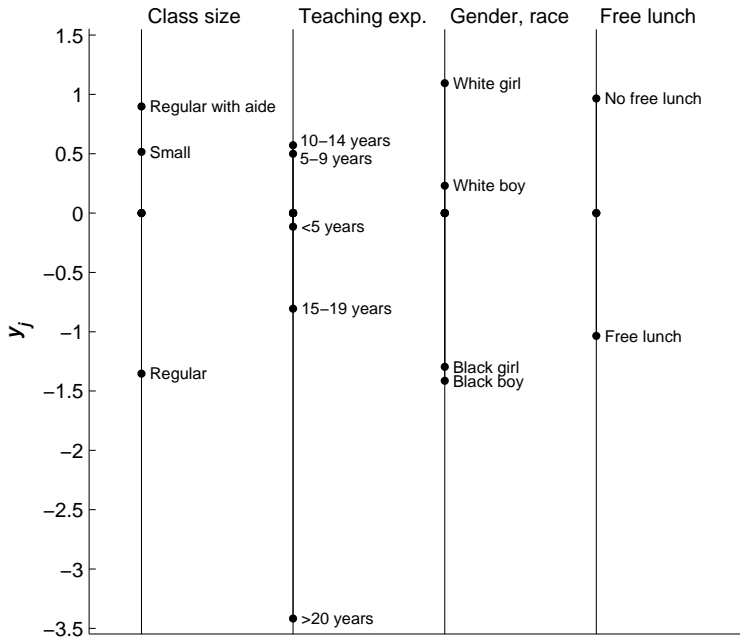
We discuss the main effects and the interaction effects of the OSI model for the STAR data set separately. Table 2.4 shows the values of the main effects (that is, the values of $b_j \mathbf{a}_j$) of the four predictor variables “Class size”, “Teaching experience”, “Gender, race”, and “Free lunch”. The main effects of “School id” have little meaning to the reader and are not reported here. The main effects show that math performance tends to be higher for girls, white children, and students who are not eligible to receive free lunches (that is, students whose household income is not low). Being in a small class and having an

experienced teacher also seem to positively affect a student's math performance, though these effects are somewhat smaller.

To interpret the interactions, we can construct visualizations of the estimated model parameters, which should provide an understanding of the interaction effects. The \mathbf{y}_j s may be graphically represented using an *interaction plot* that shows the elements of these parameter vectors. In such a figure, each level of each categorical variable is represented by a single parameter. Such an interaction plot is similar to a one-dimensional version of a biplot (see Gower & Hand, 1996) used in, for example, principal components analysis and correspondence analysis. The scaling factors s_{ji} constitute a matrix, which can be shown in a simple table.

Figure 2.1 shows the interaction plot and the estimated scaling factors of the OSI model for the STAR data set.¹ To improve the interpretability of this figure, the estimated category quantifications are shown in a separate axis for each predictor variable. The estimated standard errors of the scaling factors are shown in parentheses in Figure 2.1. Using these results, we can interpret the content of the interaction effects as follows. The values of the scaling factors s_{ji} determine the relative sizes of the interaction effects; large absolute values of these scaling factors correspond to large interaction effects. The scaling factors in Figure 2.1 show that the interaction effects between "Teaching experience" and "Class size" ($s_{12} = 3.861$) and between "Teaching experience" and "Gender, race" ($s_{23} = 2.265$) are relatively large. The content of the interaction terms can be determined using the \mathbf{y}_j s. If the corresponding scaling factor is positive, pairs of categories of different variables with quantifications \mathbf{y}_j of the same sign have positive estimated interaction effects. For the interaction between "Class size" and "Teaching experience", \mathbf{y}_1 and \mathbf{y}_2 show that there are fairly high positive estimated interaction effects between a high level of teaching experience and regular size classes; this is also true for a high level of teaching experience in combination with black students. Therefore, we can conclude that teachers with more than 15 years of experience appear more capable of handling regular size classes and classes with black students than other teachers. It seems best to assign small classes and classes with few black students to less experienced teachers. There is also a strong interaction between "Gender, race" and "Free lunch". The variable "Free lunch" is mainly determined by the household income of the student. There appear to be more severe negative effects of having a low household income on math performance for white students (especially for white girls) than for black students. The values of the quantifications for the variable "Gender, race" suggest the presence of three-way interaction effects between gender, race, and the other predictor variables. Therefore, combining the variables gender and race into one variable seems to be appropriate.

¹The figures in this chapter have been created using the programming language MATLAB. Programming code for creating such figures is included in the package that can be downloaded from <http://people.few.eur.nl/groenen>.



	Class size	Teaching exp.	Gender, race	Free lunch
Class size	-	3.861 (0.675)	-0.702 (0.707)	0.124 (0.606)
Teaching exp.	-	-	2.265 (0.693)	-0.262 (0.622)
Gender, race	-	-	-	2.593 (0.699)
Free lunch	-	-	-	-

Figure 2.1: Interaction plot and corresponding scaling factors s_{jl} of the OSI model for the STAR data set. The estimated standard errors of the scaling factors are shown in parentheses.

Table 2.5: Residual degrees of freedom, deviance values, and values of the AIC of various models for the General Social Survey data set

Model	Residual df	Deviance	AIC
GLM with main effects only	187	1,063.25	1,089.25
OSI model	173	277.85	331.85
Model (2.11), $P = 2$	162	168.46	244.46
GLM with all two-way categorical interactions	136	117.93	245.93

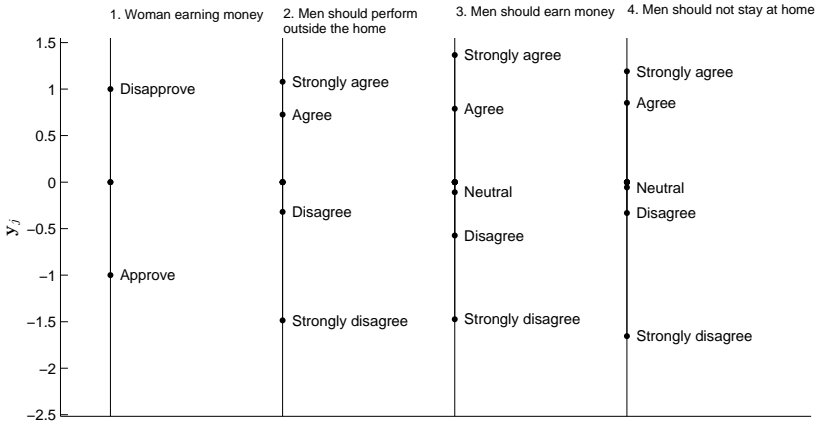
2.4.2 General Social Survey Data

The second data set used in this chapter is based on the 1994 General Social Survey (Davis & Smith, 1996). This data set contains the responses of 899 respondents on four questions on attitudes of the labor roles of women and was also used in C. J. Anderson and Vermunt (2000). The four questions were as follows:

1. Woman earning money: “Do you approve or disapprove of a married woman earning money in business or industry if she has a husband capable of supporting her?” (approve, disapprove).
2. Men should perform outside the home: “It is much better for everyone involved if the man is the achiever outside the home and the woman takes care of the home and family.” (strongly agree, agree, disagree, strongly disagree).
3. Men should earn money: “A man’s job is to earn money; a woman’s job is to look after the home and family.” (strongly agree, agree, neither agree nor disagree, disagree, strongly disagree).
4. Men should not stay at home: “It is not good if the man stays at home and cares for the children and the woman goes out to work.” (strongly agree, agree, neither agree nor disagree, disagree, strongly disagree).

From this data set, we construct a contingency table, so that we can apply log-linear analysis (generalized linear modeling with a log link and a Poisson error distribution) with the four questions as predictor variables; for each of the $2 \times 4 \times 5 \times 5 = 200$ combinations of the response categories of the questions, the response variable is the number of respondents ticking that combination of response categories.

Table 2.5 gives results of a GLM with only main effects, the OSI model, the multidimensional model (2.11) with $P = 2$, and a full two-way interaction model for the GSS data set. The deviance values in Table 2.5 show that a model with only main effects and the OSI model fit the data poorly. Therefore, we also consider the multidimensional model (2.11) with $P = 2$; we chose this model because it gives good results and unique parameter estimates for the GSS data set. Based on the deviance values in Table 2.5, this model and a GLM with full two-way interaction effects appear to fit the data. As the two-dimensional version of model (2.11) provides an adequate fit, there is no need to consider



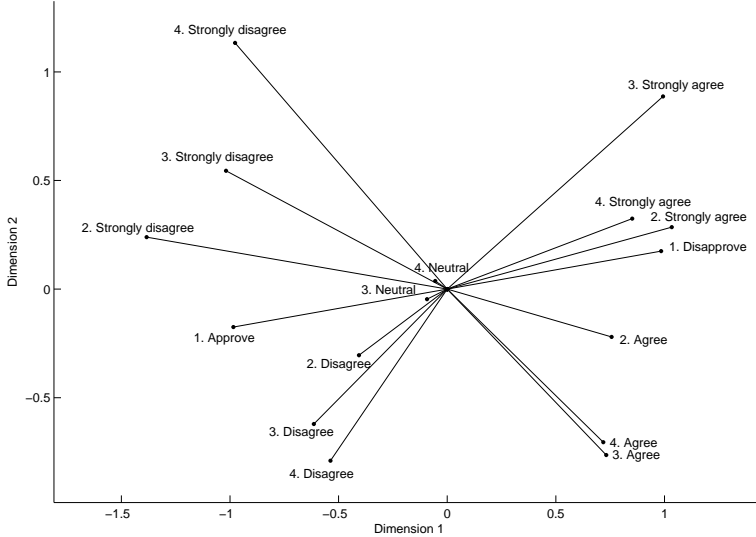
Item	1	2	3	4
1. Woman earning money	-	0.161 (0.079)	0.224 (0.073)	0.171 (0.069)
2. Men should perform outside the home	-	-	1.288 (0.129)	0.383 (0.078)
3. Men should earn money	-	-	-	0.817 (0.090)
4. Men should not stay at home	-	-	-	-

Figure 2.2: Interaction plot and scaling factors s_{jl} of the OSI model for the General Social Survey data set. The estimated standard errors of the scaling factors are shown in parentheses.

higher-dimensional versions of this model or models that include three-way interaction effects. The values of the Akaike information criterion (AIC) indicate that model (2.11) with $P = 2$ is preferable to the other three models in Table 2.5.

For the GSS data set, we focus on the interaction effects and we do not report any main effects, as the main effects only reflect the relative frequencies with which each of the response categories have been chosen. Although the two-dimensional model (2.11) should be preferred to the one-dimensional OSI model, the results of the OSI model can still help us interpret the interaction effects. Therefore, we report and interpret the results of both models here.

Figure 2.2 shows the interaction plot and the estimated scaling factors of the OSI model. As all estimated scaling factors are positive, the interaction plot shows that the respondents tend to have similar opinions for items 2, 3, and 4. The similarities are largest between items 2 and 3 ($s_{23} = 1.288$) and between items 3 and 4 ($s_{34} = 0.817$). The categories of item 1 appear inverted compared with the other three items. This was to be expected, as a negative response to item 1 implies a conservative attitude towards gender roles; for the other three items, a positive response indicates such a conservative attitude. The reported scaling factors correspond to fairly large effects. For example, for a respondent who responds “strongly disagree” to items 2 and 3, the additive interaction effect between these two items on η_i is $1.288 \times -1.485 \times -1.473 = 2.82$. For a log-linear model, the



Item	1	2	3	4
1. Woman earning money	-	0.159 (0.083)	0.282 (0.099)	0.246 (0.096)
2. Men should perform outside the home	-	-	1.452 (0.241)	0.475 (0.110)
3. Men should earn money	-	-	-	0.951 (0.118)
4. Men should not stay at home	-	-	-	-

Figure 2.3: Interaction plot and scaling factors s_{jl} of model (2.11) with $P = 2$ for the General Social Survey data set. The estimated standard errors of the scaling factors are shown in parentheses.

expected frequency μ_i is calculated as $\mu_i = \exp(\eta_i)$. Therefore, the estimated interaction effect between items 2 and 3 increases the probability of responding “strongly disagree” to both items by a factor of $\exp(2.82) = 16.8$.

Figure 2.3 shows the interaction plot and the estimated scaling factors of the two-dimensional model (2.11). The interpretation of such an interaction plot is similar to that of a principal components analysis biplot. Here, category k of item j is represented by a vector from the origin with coordinates \mathbf{y}_{kj} , that is, row k of \mathbf{Y}_j . Due to the scale constraints that have been imposed for this model, the weighted average of the squared lengths of the vectors equals one for the categories of each variable. The interaction effect of category k of variable j and category ℓ of variable l on the linear predictor η_i is modeled as $s_{jl}\mathbf{y}'_{kj}\mathbf{y}_\ell$ in (2.11), which is the product of the inner product $\mathbf{y}'_{kj}\mathbf{y}_\ell$ and the corresponding scaling s_{jl} . The inner product $\mathbf{y}'_{kj}\mathbf{y}_\ell$ is equivalent to the projection of \mathbf{y}_{kj} onto \mathbf{y}_ℓ multiplied by the length of \mathbf{y}_ℓ . Therefore, if two vectors are perpendicular, the

predicted interaction effect is zero. If the two vectors coincide in direction, then this inner product is positive and equal to the product of their two lengths. As all estimated scaling factors in Figure 2.3 are positive, pairs of vectors with angles smaller than 90° describe positive interaction effects and those with a larger angle negative effects.

In the two-dimensional interaction plot, similar response categories of items 2, 3, and 4 tend to be close together. This shows that the respondents have a tendency to give the same response for these three items. To help interpret such an interaction plot, it can sometimes be useful to assign interpretations to the dimensions. Here, we can interpret the first dimension as measuring conservatism with respect to the labor roles of women. The second dimension seems to represent the extremity of the respondent's answers (that is, his or her response tendency). The estimated scaling factors of this two-dimensional model resemble those of the one-dimensional OSI model. Therefore, we believe that these scaling factors adequately represent the sizes of the interaction effects.

2.5 Discussion

Optimal scaling is a useful methodology for modeling the effects of categorical predictor variables. In this chapter, we have applied this methodology to modeling two-way interactions effects in generalized linear models. The resulting optimal scaling of interactions (OSI) model is a multiplicative interaction model that can help interpret the content of interaction effects. This model has the additional advantages that it requires fewer parameters than a full two-way interaction model and that it can be used to construct (graphical) representations of the interaction effects. The OSI model can be seen as an extension of several models for parsimoniously representing interaction effects, including Goodman's RC(M) association models and models that were proposed by C. J. Anderson and Vermunt (2000) and Groenen and Koning (2006).

Using two empirical data sets, we have shown how the OSI model can be applied in practice and we have compared its usefulness with other models, including a multidimensional model. Based on the results, the one-dimensional OSI model appears to be useful, as it is easy to apply and gives good results. The results of multidimensional models tend to be harder to interpret. Based on our experience with these models, we recommend using the one-dimensional OSI model if the fit of this model is good enough. We believe that this model can be useful for interpreting interaction effects in an applied setting.

An advantage of the OSI model is that it uses different sets of parameters to model the strength of an interaction term (using the scaling factors s_{ji}) and the content of an interaction term (using \mathbf{y}_j). Separating the strength and the content of an interaction effect in this way may help to understand the results. A limitation of the one-dimensional OSI model is that it may have an inadequate fit for some data sets. This problem may be solved by applying one of the multidimensional extensions in Section 2.3. In these extensions, degenerate solutions can sometimes occur that are avoided in the one-dimensional OSI model. To account for this problem, researchers applying a multidimensional OSI model should always check whether the parameters are identified (for example, by determining

whether the optimization routine returns the same parameter estimates with different starting values). If the parameters are not identified, we recommend using either a one-dimensional OSI model or a full two-way interaction model. If even the fit of a full two-way interaction model is insufficient, the idea of optimal scaling can potentially be used to obtain a parsimonious specification for the three-way interaction effects. However, we have focused on two-way interaction effects in this chapter, as we believe that two-way interactions are the most important ones to explore in practice.

2.A Model Reparameterization

In this appendix, we discuss a parameterization for the quantifications of the optimally scaled variables in the one-dimensional OSI model. In this reparameterization, no parameter constraints are necessary, which facilitates the easy calculation of standard errors of the estimated parameters. This reparameterization consists of replacing the quantifications for the main effects and the quantifications for the interaction effects by functions of new parameters, so that the resulting optimally scaled variables automatically have mean 0 and variance 1.

The constraints imposing mean 0 and variance 1 can respectively be described by

$$\sum_{k=1}^{K_j} f_k x_k = 0 \quad (2.16)$$

and

$$\sum_{k=1}^{K_j} f_k x_k^2 = 1, \quad (2.17)$$

in which x_k is the quantification (for either the main effects or the interaction effects of variable j) of category k , and f_k denotes the relative frequency of category k . We assume that the f_k s are positive and sum to one, so that $\sum_{k=1}^{K_j} f_k = 1$.

We use an inductive approach to determine a reparameterization of the quantifications that satisfies the constraints (2.16) and (2.17). Suppose that quantifications x_1, \dots, x_{r-1} have already been determined and x_r , $r < K_j - 1$ must be determined. Consider the equations

$$\sum_{k=1}^r f_k x_k^2 + R_{\min}(x_1, \dots, x_r) \leq 1 \quad (2.18)$$

and

$$\sum_{k=1}^r f_k x_k^2 + R_{\max}(x_1, \dots, x_r) \geq 1, \quad (2.19)$$

where $R_{min}(x_1, \dots, x_r)$ and $R_{max}(x_1, \dots, x_r)$ are respectively defined as

$$R_{min}(x_1, \dots, x_r) = \min_{x_{r+1}, \dots, x_{K_j}} \sum_{k=r+1}^{K_j} f_k x_k^2 \quad \text{subject to} \quad \sum_{k=r+1}^{K_j} f_k x_k = - \sum_{k=1}^r f_k x_k,$$

$$R_{max}(x_1, \dots, x_r) = \max_{x_{r+1}, \dots, x_{K_j}} \sum_{k=r+1}^{K_j} f_k x_k^2 \quad \text{subject to} \quad \sum_{k=r+1}^{K_j} f_k x_k = - \sum_{k=1}^r f_k x_k.$$

It is straightforward to see that x_r , given x_1, \dots, x_{r-1} , can be an element of a sequence x_1, \dots, x_{K_j} that satisfies (2.16) and (2.17) if and only if it satisfies (2.18) and (2.19).

For $r < K_j - 1$, the solution of the maximization problem that occurs in the definition of $R_{max}(x_1, \dots, x_r)$ is always infinity, as x_{K_j-1} and x_{K_j} can be set arbitrarily large with opposite signs. Therefore, (2.19) is always satisfied. An application of the Lagrange relaxation method shows that the solution of the minimization problem in $R_{min}(x_1, \dots, x_r)$ is given by $x_s = -(\sum_{k=1}^r f_k x_k) / \sum_{k=r+1}^{K_j} f_k$, $r < s \leq K_j$, so that $R_{min}(x_1, \dots, x_r) = (\sum_{k=1}^r f_k x_k)^2 / \sum_{k=r+1}^{K_j} f_k$. The permissible values for x_r , $r < K_j - 1$ are therefore given by

$$\sum_{k=1}^r f_k x_k^2 + \left(\sum_{k=1}^r f_k x_k \right)^2 / \sum_{k=r+1}^{K_j} f_k \leq 1. \quad (2.20)$$

Rewriting this inequality as

$$\begin{aligned} & (f_r + f_r^2 / \sum_{k=r+1}^{K_j} f_k) x_r^2 + 2f_r \left(\sum_{k=1}^{r-1} f_k x_k \right) / \left(\sum_{k=r+1}^{K_j} f_k \right) x_r + \\ & \left(\sum_{k=1}^{r-1} f_k x_k \right)^2 / \sum_{k=r+1}^{K_j} f_k + \sum_{k=1}^{r-1} f_k x_k^2 - 1 \leq 0 \end{aligned}$$

and solving for x_r shows that this inequality holds if

$$-\frac{\sum_{k=1}^{r-1} f_k x_k}{\sum_{k=r}^{K_j} f_k} - \frac{\sqrt{Q_r}}{\sum_{k=r}^{K_j} f_k} \leq x_r \leq \frac{-\sum_{k=1}^{r-1} f_k x_k}{\sum_{k=r}^{K_j} f_k} + \frac{\sqrt{Q_r}}{\sum_{k=r}^{K_j} f_k}, \quad (2.21)$$

in which Q_r is defined as

$$Q_r = \left(\sum_{k=1}^{r-1} f_k x_k \right)^2 - \left(\left(\sum_{k=r+1}^{K_j} f_k \right)^2 / f_r + \sum_{k=r+1}^{K_j} f_k \right) \left(\sum_{k=1}^{r-1} f_k x_k \right)^2 / \sum_{k=r+1}^{K_j} f_k + \sum_{k=1}^{r-1} f_k x_k^2 - 1.$$

For determining x_{K_j-1} , $R_{min}(x_1, \dots, x_{K_j-1}) = R_{max}(x_1, \dots, x_{K_j-1})$ so that (2.18) should hold with equality. This equality holds if x_{K_j-1} equals one of two endpoints in the interval given in (2.21). Based on (2.21), we construct a reparameterization of

the parameters x_1, \dots, x_{K_j} using the parameters $\theta_1, \dots, \theta_{K_j-2}$ according to the following recursive formulas

$$\begin{aligned} x_r &= \frac{-\sum_{k=1}^{r-1} f_k x_k + (\exp(\theta_r) - 1)/(\exp(\theta_r) + 1)\sqrt{Q_r}}{\sum_{k=r}^{K_j} f_k}, r < K_j - 1 \\ x_{K_j-1} &= \frac{-\sum_{k=1}^{K_j-2} f_k x_k + \sqrt{Q_{K_j-1}}}{f_{K_j-1} + f_{K_j}}, \\ x_{K_j} &= \frac{-\sum_{k=1}^{K_j-1} f_k x_k}{f_{K_j}}. \end{aligned}$$

The sign of $\sqrt{Q_r}$ in the definition of x_{K_j-1} can be chosen to be either positive or negative without influencing the general results, so that x_{K_j-1} could alternatively be defined as $x_{K_j-1} = (-\sum_{k=1}^{K_j-2} f_k x_k - \sqrt{Q_r})/(f_{K_j-1} + f_{K_j})$. Now, the parameters $\theta_1, \dots, \theta_{K_j-2}$ can be allowed to vary freely, but these parameters can still represent all optimally scaled variables with mean 0 and variance 1.

Chapter 3

Identifying Response Styles: A Latent-Class Bilinear Multinomial Logit Model*

Abstract

Respondents can vary strongly in the way they use rating scales. Specifically, respondents can exhibit a variety of response styles, which threatens the validity of the responses. The purpose of this chapter is to investigate how response style and content of the items affect rating scale responses. The authors develop a novel model that accounts for different kinds of response styles, content of items, and background characteristics of respondents. By imposing a bilinear parameter structure on a multinomial logit model, the effects on the response behavior of the characteristics of a respondent and the content of an item are graphically distinguished. This approach is combined with finite mixture modeling, yielding two segmentations of the respondents: one for response style and one for item content. This latent-class bilinear multinomial logit (LC-BML) model is applied to the well-known List of Values in a cross-national context. The results show large differences in the opinions and the response styles of respondents and reveal previously unknown response styles. Some response styles appear valid communication styles, whereas other response styles often concur with inconsistent opinions of the items and seem to be response bias.

3.1 Introduction

In recent years, the interest in response styles has increased in marketing research (Baumgartner & Steenkamp, 2001; De Jong, Steenkamp, Fox, & Baumgartner, 2008) and

*This chapter is based on Van Rosmalen, Van Herk, and Groenen (in press).

related disciplines (Cheung & Rensvold, 2000; Van Herk, Poortinga, & Verhallen, 2004). Response styles are now considered a source of concern in domestic (Hui & Triandis, 1989; Greenleaf, 1992a) and international research (Johnson, Kulesa, Cho, & Shavitt, 2005).

A response style can be defined as a person's tendency to respond to questionnaire items regardless of item content (Paulhus, 1991). A major reason for concern about response styles is that they are considered content-irrelevant and that they may threaten the validity of responses given to substantive questions. Furthermore, marketing managers may perceive differences in response styles between countries as differences in content and may make unfavorable marketing decisions based on these assumed differences (Usunier & Lee, 2005). Therefore, researchers have suggested removing response styles from the data (for example, Hofstede, 2001; Baumgartner & Steenkamp, 2001). Purging response styles is possible if researchers have a large data set with heterogeneous items; however, in smaller data sets, the response style components are hard to separate from the real attitude information. In that case, purging the response styles may also remove attitude information. Moreover, several researchers (Greenleaf, 1992a; Smith, 2004) state that certain response styles contain attitude information only and no bias, and should be considered a communication style.

Many well-known statistical techniques such as k -means clustering (Hartigan, 1975) have problems distinguishing between response style and item content. For example, when applying k -means clustering to ratings containing extreme response style or acquiescence, some clusters are likely to emerge that reflect those response styles instead of attitude information. Other possible unwanted effects of response styles are the inflation of mean scores and the inflation of correlations between constructs that are both affected by the same response style (Baumgartner & Steenkamp, 2001). For example, Van Herk et al. (2004) found that Greek women systematically gave more positive answers on cooking attitudes than French women. However, this result was not reflected in their behavior. Targeting the Greek market for cooking products would not have led to the expected positive effect on sales, because the differences in responses between French and Greek respondents were not content-related.

When differences in responses are caused by response bias, correcting for response styles is necessary. Therefore, several methods to remove response style from data have been suggested: (a) Hofstede (2001) and Bond (1988) employ ipsatizing to correct for response styles in their respective values studies; (b) in his large values survey, Schwartz (1992) recommends using the respondent's mean across all items as a control variable in partial correlation or regression analysis; (c) Baumgartner and Steenkamp (2001) recommend calculating various response style indices across many heterogeneous items in a large data set and correcting for them using regression techniques; (d) other researchers include parameters in their model to correct for scale usage heterogeneity. For example, Ter Hofstede, Steenkamp, and Michel (1999) model heterogeneity in response behavior in pick any/ n data, Lenk, Wedel, and Böckenholt (2006) take scale usage heterogeneity into account when fitting circumplex structures, and Rossi, Gilula, and Allenby (2001) correct for extreme and acquiescent response tendencies in a small set of ordinal rating scales.

Although the idea of correcting for response styles seems generally accepted, there is no consensus on which approach should be used. Moreover, many cross-national studies do not even mention correcting for response styles (Schwarz, 2003). The fact that the validity of the information after correction is unclear (see, for example, Fischer, 2004) may be a major reason for the lack of consensus on whether or not to purge response styles.

In this chapter, we focus on the simultaneous effects of response style (which can be either bias or communication style) and attitude (item content) on response behavior in data sets with multiple items and constructs measured using a rating scale. As people are heterogeneous with respect to their opinions and response styles, we use a segmentation for each effect. In doing so, we aim to distinguish between the effects of response style and the effects of attitude on response behavior. We also show how these effects interact, so that we provide insight into what attitude-based segments are more prone to certain response styles than other attitude-based segments. In addition to content and response style, we also model the effects of background characteristics, as those are known to affect response behavior (Greenleaf, 1992a). Moreover, in contrast to other studies on response styles (Cheung & Rensvold, 2000; De Jong et al., 2008), we do not limit ourselves to the prominent response styles extreme response style and acquiescence, and we also do not predefine what kinds of response styles exist. Instead, the segmentation procedure allows us to infer what possibly a priori unknown response styles are present in the data.

The present study contributes to the literature in two ways. First, we simultaneously model the effects of item content and the effects of response style. By segmenting respondents with respect to their opinions and the way in which the respondents use the rating scale, we provide insight into what determines response behavior. Unlike other methods (De Jong et al., 2008; Rossi et al., 2001), we graphically show how item content affects response behavior, using biplots (see, for example, Gower & Hand, 1996; DeSarbo, Grewal, & Scott, 2008; Hoffman & Franke, 1986). The graphical displays facilitate insight into the extent to which responses to specific items are more likely to get, for example, extreme or positive responses. Second, the natural ordering of rating scale categories is not imposed on the model, as it could restrict how content and personal characteristics affect response behavior. As J. A. Anderson (1984, p. 3) remarks: “There is no merit in fitting an ordered relationship as a routine, simply because the response variable is ordered.” Our approach of assuming a nominal scale has the additional advantage that missing values and ‘don’t know’ response options can be modeled in the same way as ordinary rating scale categories.

The outline of this chapter is as follows. First, we introduce the LC-BML model that is proposed in this chapter. We then illustrate the merits of this model using a large international data set including the well-known List of Values (LOV) and compare our results with those of other methods for assessing response style. Finally, we discuss the implications of our findings and the study’s limitations, and we suggest possible topics for further research.

3.2 A Latent-Class Bilinear Multinomial Logit Model

We seek to model the entire response behavior, that is, the probability that a single respondent ticks certain rating scale categories for the items used in the study, given the content of the items and the respondent's background characteristics. To do so, we use the responses of all persons to all items in a single statistical model, which we call the Latent-Class Bilinear Multinomial Logit model (LC-BML model). In this model, we distinguish three sets of variables. The first set is a single variable called Rating that indicates the category of the rating scale ticked by the respondent. The second set is the single variable called Item that indicates the item being rated by the respondent. The third set contains variables indicating the background characteristics of the respondents. A nominal measurement level is assumed for all three sets of variables. To adequately distinguish between response style and item content, the items should represent a diverse set of constructs; throughout this chapter, we assume that the data satisfy this condition.

The basis of the LC-BML model is an unordered multinomial logit model with Item and the background variables as categorical predictor variables to explain the dependent categorical variable Rating. We adapt this multinomial logit model by incorporating ideas from the models proposed by J. A. Anderson (1984), C. J. Anderson and Vermunt (2000), and Groenen and Koning (2006). To limit the number of parameters, we apply the bilinear parameter structure used in these articles. In addition, we incorporate latent classes (see, for example, Wedel & Kamakura, 2000) to model (a) differences among respondents in reactions to content (that is, their opinions of the items) and (b) differences among respondents in response tendencies (that is, their use of rating scale response categories). Our model extends the stereotype model proposed by J. A. Anderson (1984) by using the bilinear parameter structure on categorical predictor variables and by incorporating latent classes. Our approach allows to determine how likely a given pattern of responses is for a single respondent, given the items used in the study and certain characteristics of the respondent.

Before introducing the LC-BML model, we define some notation.

i Index to specify the respondent, with $i = 1, \dots, n$.

t Index to indicate the item being rated, with $t = 1, \dots, T$.

j Index to indicate the rating, with $j = 1, \dots, J$.

k Index to indicate the background variable, with $k = 1, \dots, K$.

Y_{it} The random variable of the rating of person i on item t .

y_{it} The realized rating of person i on item t .

m_k The number of categories of background variable k .

The multinomial logit model predicts the probability of person i choosing rating j on item t as

$$\Pr(Y_{it} = j) = \frac{\exp(z_{ijt})}{\sum_{j'=1}^J \exp(z_{ij't})},$$

with z_{ijt} a linear combination of a constant a_j (the attractiveness of rating category j), the effects of the categories of the background variables, and the effect c_{jt} of rating j on item t , that is,

$$z_{ijt} = a_j + \sum_{k=1}^K \mathbf{b}'_{jk} \mathbf{x}_{ik} + c_{jt}. \quad (3.1)$$

Each background variable k can be represented by a matrix \mathbf{X}_k of dummy variables for each category, so that \mathbf{x}'_{ik} is row i of \mathbf{X}_k , and $\mathbf{b}'_{jk} \mathbf{x}_{ik}$ selects the element from vector \mathbf{b}_{jk} belonging to the category of respondent i on background variable k . As a result of Equation 3.1, separate model parameters need to be estimated for each rating scale category. To see this, consider a hypothetical example with $J = 3$ response categories of Rating, $T = 4$ items, and $K = 1$ background variable (that has $m_1 = 2$ categories). For this case, Table 3.1 contains the z_{ijt} s as the linear combination of the specific parameters. There are $J = 3$ attractiveness parameters a_j , $J \times m_1 = 6$ parameters for the effects of the two categories of the background variable, and $J \times T = 12$ parameters for the effect c_{jt} of item t on rating j .

To fully model the effects of the explanatory variables on the dependent variable, a parameter vector is required for each category of each explanatory variable. The number of parameters $J + J \sum_{k=1}^K m_k + JT$ gets large if the total number of categories is large. The resulting parameter vectors are generally not easy to represent or to interpret, which is one of the reasons for proposing the adaptations below.

Bilinear Decomposition of Parameters

The problem of the large number of parameters occurs in a wide variety of models for analyzing categorical data sets. For the case of log-linear analysis, C. J. Anderson and Vermunt (2000) proposed several models that construct a parsimonious representation of the relationships between categorical variables, based on a bilinear decomposition. Groenen and Koning (2006) proposed a similar model in the context of analysis of variance, with the aim of obtaining a simple low-dimensional visual display of the estimated interaction effects. Based on their interaction decomposition model, we adapt the multinomial logit model described above to parsimoniously represent the effects of the explanatory variables.

Table 3.1: The linear combination of the z_{ijt} s for a hypothetical example with $J = 3$ response categories of Rating, $T = 4$ items, and $K = 1$ background variable.

Rating j	Item 1	Item 2	Item 3	Item 4
1	$a_1 + \mathbf{b}'_1 \mathbf{x}_{i1} + c_{11}$	$a_1 + \mathbf{b}'_1 \mathbf{x}_{i1} + c_{12}$	$a_1 + \mathbf{b}'_1 \mathbf{x}_{i1} + c_{13}$	$a_1 + \mathbf{b}'_1 \mathbf{x}_{i1} + c_{14}$
2	$a_2 + \mathbf{b}'_2 \mathbf{x}_{i1} + c_{21}$	$a_2 + \mathbf{b}'_2 \mathbf{x}_{i1} + c_{22}$	$a_2 + \mathbf{b}'_2 \mathbf{x}_{i1} + c_{23}$	$a_2 + \mathbf{b}'_2 \mathbf{x}_{i1} + c_{24}$
3	$a_3 + \mathbf{b}'_3 \mathbf{x}_{i1} + c_{31}$	$a_3 + \mathbf{b}'_3 \mathbf{x}_{i1} + c_{32}$	$a_3 + \mathbf{b}'_3 \mathbf{x}_{i1} + c_{33}$	$a_3 + \mathbf{b}'_3 \mathbf{x}_{i1} + c_{34}$

An important objective of our approach is the graphical representation of the parameter estimates, to facilitate a better understanding of the effects. To be able to represent the results of the model graphically, the bilinear decomposition assigns a point in a low-dimensional space to each category of each variable.

To formally describe the bilinear decomposition, we gather the \mathbf{b}'_{jk} s (that is, the effect of background variable k on rating j) for all ratings j underneath each other as $\mathbf{B}_k = [\mathbf{b}_{1k} | \mathbf{b}_{2k} | \dots | \mathbf{b}_{Jk}]'$. To ensure that the categories of all variables can be represented by points in a low-dimensional space, we restrict each \mathbf{B}_k and \mathbf{C} (with elements c_{jt}) as

$$\mathbf{B}_k = \mathbf{F}\mathbf{G}'_k \text{ and } \mathbf{C} = \mathbf{F}\mathbf{H}'. \quad (3.2)$$

Here, \mathbf{F} is the $J \times P$ matrix of coordinates for the points of ratings in the visual representation, \mathbf{G}_k the $m_k \times P$ matrix of coordinates for background variable k , and the $T \times P$ matrix \mathbf{H} denotes the coordinates of the T items, where P is the dimensionality of the representation. The dimensionality P cannot be greater than $J - 1$; for $P = J - 1$, the effects of the explanatory variables are fully modeled, so that our bilinear model is equivalent to a standard multinomial logit model in that case. The model under the rank restrictions in Equation 3.2 can now be described as

$$z_{ijt} = a_j + \sum_{k=1}^K \sum_{p=1}^P f_{jp} \mathbf{g}'_{kp} \mathbf{x}_{ik} + \sum_{p=1}^P f_{jp} h_{tp},$$

where f_{jp} and h_{tp} are elements of \mathbf{F} and \mathbf{H} respectively, and \mathbf{g}_{kp} is the p -th column of \mathbf{G}_k .

Segmentation of Respondents

In marketing, and especially in international marketing, response behavior cannot be considered homogeneous across respondents. People differ with respect to their opinions and preferences, and heterogeneity in response behavior is expected. So far, the model has no possibility to take heterogeneity in the response behavior of respondents into account, and assumes that the T responses of every respondent are independent. To account for dependent observations and unobserved heterogeneity in the respondents, we use finite mixture modeling (see, for example, Wedel & Kamakura, 2000). We extend our model to take two types of heterogeneity into account. First, respondents may vary with respect to their response styles, so that they differ in the probabilities with which they generally tick different response categories of Rating. We account for these differences in response styles by allowing the constants a_j to vary between latent classes of respondents. Second, respondents can differ on their opinions of the relative importance of the items. These differences are incorporated by allowing the parameters h_{tp} to also vary between latent classes of respondents. We thus construct a joint segmentation of the respondents in the model, in which response style and content function as bases for the two segmentations. Here, we allow the prior probabilities of membership in the two types of segments to be dependent. It is also possible to model these prior probabilities as independent, which would be more restrictive.

Let there be R segments for the response tendencies and S segments for representing the opinions of the relative importance of the items. We refer to these types of segments as response style segments and item segments, respectively. The prior probability that a respondent simultaneously belongs to response style segment r and item segment s is denoted by u_{rs} and must satisfy $\sum_{r=1}^R \sum_{s=1}^S u_{rs} = 1$. The probability that respondent i evaluates item t with rating j can be calculated as

$$\Pr(Y_{it} = j) = \sum_{r=1}^R \sum_{s=1}^S u_{rs} \Pr(Y_{it} = j|r, s), \quad (3.3)$$

with

$$\Pr(Y_{it} = j|r, s) = \frac{\exp(z_{ijt|r,s})}{\sum_{j'=1}^J \exp(z_{ij't|r,s})}, \quad (3.4)$$

$$z_{ijt|r,s} = a_{j|r} + \sum_{k=1}^K \sum_{p=1}^P f_{jp} \mathbf{g}'_{kp} \mathbf{x}_{ik} + \sum_{p=1}^P f_{jp} h_{tp|s}.$$

The likelihood of observing responses y_{it} , $t = 1, \dots, T$, conditional on respondent i belonging to response style segment r and item segment s is

$$L_i(\mathbf{y}_i|r, s) = \prod_{t=1}^T \prod_{j=1}^J \Pr(Y_{it} = j|r, s)^{I(y_{it}=j)},$$

where $I()$ denotes the indicator function, and \mathbf{y}_i is the vector of length T with the observed responses of respondent i . As the unconditional likelihood for respondent i is

$$L(\mathbf{y}_i) = \sum_{r=1}^R \sum_{s=1}^S u_{rs} L_i(\mathbf{y}_i|r, s),$$

the likelihood of the entire model, as a function of the model parameters, is given by

$$\begin{aligned} & L(a_{\cdot|1}, \dots, a_{\cdot|R}, \mathbf{F}, \mathbf{G}_1, \dots, \mathbf{G}_K, \mathbf{H}_1, \dots, \mathbf{H}_S, \mathbf{U}) \\ &= \prod_{i=1}^n \sum_{r=1}^R \sum_{s=1}^S u_{rs} \prod_{t=1}^T \prod_{j=1}^J \Pr(Y_{it} = j|r, s)^{I(y_{it}=j)}. \end{aligned} \quad (3.5)$$

A number of parameter constraints are required to ensure parameter identification of the LC-BML model; these constraints are described in Appendix 3.A. For given values of the numbers of segments R and S and the dimensionality P , parameter estimates are obtained by maximizing the likelihood function in Equation 3.5. A description of the optimization algorithm can be found in Appendix 3.B. After imposing the necessary parameter constraints, the LC-BML model is identifiable (that is, unique maximum likelihood parameter estimates can exist; see also Teicher, 1963), if there are at least several items in the study and not too many segments; the identifiability of the LC-BML model is discussed further

in Appendix 3.C. Given the optimal parameter estimates, the posterior probability that respondent i belongs to response style segment r and item segment s can be calculated in a Bayesian way as

$$\Pr(i \in \Xi_{r,s}) = \frac{u_{rs} \prod_{t=1}^T \prod_{j=1}^J \Pr(Y_{it} = j|r, s)^{I(y_{it}=j)}}{\sum_{r'=1}^R \sum_{s'=1}^S u_{r's'} \prod_{t=1}^T \prod_{j=1}^J \Pr(Y_{it} = j|r', s')^{I(y_{it}=j)}}, \quad (3.6)$$

where $\Xi_{r,s}$ denotes the set of respondents belonging to response style segment r and item segment s .

The dimensionality and the numbers of segments of a model are often determined using an information criterion, such as AIC or BIC (see Andrews & Currim, 2003). To evaluate the performance of these information criteria for the LC-BML model, we have conducted a simulation study, which is discussed in Appendix 3.D. The results show that, for a variety of simulated data sets, AIC-3, BIC, and CAIC are usually able to recover the correct numbers of segments and the dimensionality. The simulation study also shows that the parameter recovery of the optimization algorithm is generally acceptable. In this chapter, we use the Bayesian information criterion (BIC) to determine the dimensionality and the numbers of segments. This criterion performs well with very large sample sizes (see Andrews & Currim, 2003) and is supported by the results of our simulation study.

Correcting for Response Styles

Because the LC-BML model accounts for differences in both response styles and opinions, it can also be used to correct for the presence of response styles (that is, remove the response styles from the data); such an approach will be used at the end of the Results section. For each observation, we can determine estimated posterior response probabilities as

$$\Pr(Y_{it} = j|y_i) = \sum_{r=1}^R \sum_{s=1}^S \Pr(i \in \Xi_{r,s}) \Pr(Y_{it} = j|r, s), \quad (3.7)$$

where $\Pr(i \in \Xi_{r,s})$ and $\Pr(Y_{it} = j|r, s)$ are calculated using Equations 3.4 and 3.6, respectively. These estimated posterior response probabilities can be corrected for the presence of response styles as follows. First, correct the conditional response probabilities $\Pr(Y_{it} = j|r, s)$ by setting $a_{j|r} = 0$ for all j and r and by setting all elements of each estimated \mathbf{g}_{kp} equal to 0 in Equation 3.4. Then, recalculate the posterior response probabilities using Equation 3.7. Note that the original (uncorrected) parameter estimates should be used to compute the posterior segment membership probabilities $\Pr(i \in \Xi_{r,s})$. The corrected response probabilities can be used to simulate a data set from which differences in response styles have been removed as much as possible. Such a simulated data set may be used in subsequent analyses.

3.3 Empirical Application: The List of Values

We apply the LC-BML model to the List of Values (Kahle, 1983). The List of Values (LOV) is a well-known values system that is often used in marketing in single country studies (see, for example, Chrysohoidis & Krystallis, 2005; Kamakura & Novak, 1992) as well as in studies in multiple countries (Wedel, Ter Hofstede, & Steenkamp, 1998; Kropp, Lavack, & Silvera, 2005). Moreover, it is known that different segments based on LOV exist (Kamakura & Novak, 1992) and that the items on the LOV represent a diverse set of constructs (Kahle, 1983). The LOV comprises nine items, of which respondents indicate the importance ‘as a guiding principle in their lives’. The nine items are (a) sense of belonging, (b) excitement, (c) warm relationships with others, (d) self-fulfillment, (e) being well-respected, (f) fun & enjoyment in life, (g) security, (h) self-respect, and (i) a sense of accomplishment.

The results of previous studies have shown similarities across countries and across subgroups within countries. The value ‘self-respect’ is the most important value in studies in the USA (Kahle, 1983) and one of the most important values in studies in, for example, Greece (Chrysohoidis & Krystallis, 2005) and Germany and Norway (Grunert & Scherhorn, 1990). The value ‘excitement’ is usually considered the least important one (Kamakura & Novak, 1992; Kropp et al., 2005). The importance of ‘warm relationships with others’ depends on the meaning respondents attach to it. If respondents see this value as hedonic, it is assessed similar to ‘fun & enjoyment’; if respondents see ‘warm relationships with others’ as true friendship, it is considered similar to ‘sense of belonging’. On the value ‘sense of belonging’, people tend to have a strong view that can be either positive or negative; in some countries (Germany, Norway), this value is considered very important, whereas it is not important in, for example, Canada, Greece, and the USA (Grunert & Scherhorn, 1990). For the other five LOV items, no consistent results have been found. Based on the aforementioned findings, we expect that: (1) in all item segments, the value ‘self-respect’ is considered (very) important; (2) ‘excitement’ is not considered important; (3) ‘sense of belonging’ is considered very important in some, but not all, segments.

Data

The data used in this study are part of a commercial survey performed in 1996 (see also Van Herk, 2000). The original sample consisted of 4916 male respondents from five European countries (France, Germany, Italy, Spain, and United Kingdom). For each respondent, in addition to items on personal care, the List of Values (Kahle, 1983) and background characteristics were measured.

The importance of each LOV item as a guiding principle in life was measured on a nine-point rating scale ranging from ‘1’ (‘very important’) to ‘9’ (‘not important at all’). The procedure used to establish translation equivalence was a committee approach using parallel translation (Craig & Douglas, 2005). All questionnaire items were translated from English into French, German, Italian, and Spanish by bilinguals. Next, the translations were

examined, and a discussion was held between the translators, two independent reviewers, and the project coordinators in the participating countries to solve inconsistencies.

Respondents who had missing values on the background variables, respondents who did not respond to any LOV item, and respondents who ticked the same rating category for each LOV item were removed from the sample. This procedure resulted in the elimination of 8% of the respondents, leading to a final sample of 4514 respondents. The respondents who had missing values for some but not all LOV items were included in the analyses. To do so, missing values were treated as a separate rating scale category, so that there effectively were ten rating scale categories for each LOV item.

The numbers of respondents were 936, 965, 994, 771, and 848 in France, Germany, Italy, Spain, and United Kingdom, respectively. Country is explicitly included in the model, because response styles differ among countries (see, for example, Johnson et al., 2005). In addition, we include the demographic variables age and level of education. Differences in response styles have often been reported for people differing with respect to these two demographic characteristics (see, for example, Greenleaf, 1992a, 1992b; Hamilton, 1968; Krosnick & Alwin, 1988). In his review, Hamilton (1968) already reported accumulated evidence that adolescents and elderly subjects gave more extreme responses than people aged 20-59. Later studies (see, for example, Greenleaf, 1992a) support this finding. Acquiescence was found positively related to age; older people tend to use the positive side of the rating scale more often (Winkler, Kanouse, & Ware, 1982; Moum, 1988; Greenleaf, 1992a, 1992b). Several researchers found that education is negatively associated with acquiescence (Greenleaf, 1992a; Krosnick & Alwin, 1988; Winkler et al., 1982) and with extreme response style (Greenleaf, 1992a, 1992b; Stening & Everett, 1984). Based on these findings, we expect that acquiescence has a curvilinear relationship with age and that extreme response style is positively related to age. In addition, we expect that both response styles are negatively related to education. To ensure equivalence across countries, education was coded in two levels: higher level of education (40%) and lower level of education (60%). Age was coded as a categorical variable with categories 15 – 24 (13%), 25 – 39 (32%), 40 – 54 (25%), and 55+ (30%).

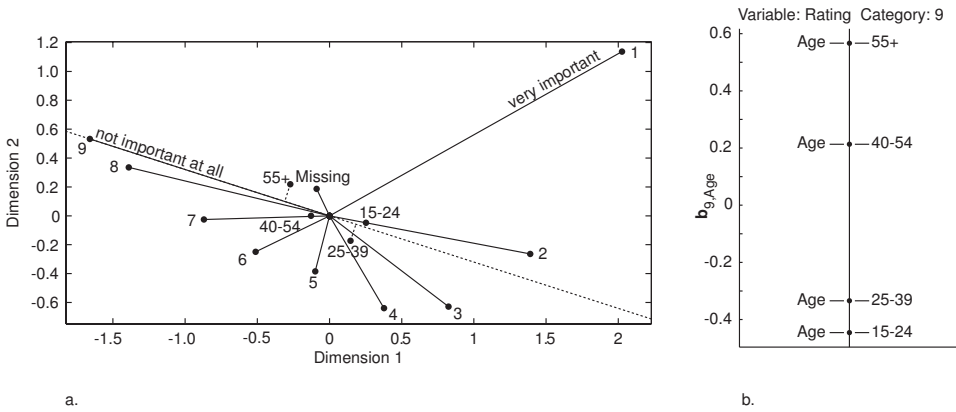
3.4 Results

In this section, we apply the LC-BML model to the List of Values data set and interpret the results. We seek to explain the observed rating scores using the LOV item and the background characteristics country, age, and education as predictor variables. In our analyses, we use each respondent's answers on how important each LOV item is in his daily life, so that we have $4514 \times 9 = 40626$ responses in total. To present the results, we first discuss how the dimensionality and the numbers of segments of the model are chosen. We then describe results for the two types of segmentations and how they are related. Finally, we study the statistical stability of the graphical representations and compare the predictive validity of our model with that of our methods.

To choose the dimensionality P , the number of response style segments R , and the

number of item segments S , we use the Bayesian information criterion (BIC). We estimate the model parameters for each combination of $P = 1, \dots, 3$ dimensions, $R = 1, \dots, 13$ response style segments, and $S = 1, \dots, 6$ item segments. As a safeguard against local optima, the optimization algorithm was run ten times with randomly chosen starting values for each combination of P , R , and S , and the best results out of these ten runs are reported here. The BIC values are shown in Table 3.2. The lowest overall BIC value (130,685) is attained with $P = 2$ dimensions, $R = 11$ response styles segments, and $S = 5$ item segments; the corresponding value of the log-likelihood is $-63,946.78$. We use these numbers of segments with two dimensions to interpret the results of the model. To facilitate distinguishing between the item segments and the response style segments, we number the response style segments and assign the letters A through E to the item segments. To easily interpret the segments, we order the response style segments based on the associated response styles, and we order the item segments according to their size.

Figure 3.1: Graphical Representation of Effects of Age on Response Behavior



Notes: Panel a shows the graphical representation with a vector for each category of Rating, a dot for each age category, and the projections of the age categories onto rating ‘9’. Panel b shows the projections multiplied by the length of rating ‘9’.

To interpret the results, we first show the effects of Item and the background characteristics on Rating using graphical representations. To help explain the graphical representations, only the points of the rating scale categories and the age categories are shown in Figure 3.1; this figure can be used to determine how rating scale usage varies among age categories. The categories of age are shown as dots, and the rating scale categories are shown as vectors. The effects of age in the multinomial logit model are related to the projections of the points of the age categories onto the vectors of Rating. For example, if one is interested in the estimated effect (that is, \mathbf{b}_{jk}) of age ‘55+’ on rating ‘9’, one may proceed as follows.

Table 3.2: BIC Values for $R = 1, \dots, 13$ Response Style Segments, $S = 1, \dots, 6$ Item Segments, and Dimensionalities $P = 1, \dots, 3$ in the LOV Data Set

Dim.	Response style segments	Item segments					
		1	2	3	4	5	6
$P = 1$	1	144,946	144,233	142,440	142,081	141,934	141,530
	2	139,964	138,791	138,379	137,668	137,705	137,547
	3	136,436	135,256	134,852	134,580	134,480	134,446
	4	135,428	134,158	133,666	133,397	133,175	133,154
	5	134,508	133,222	132,791	132,467	132,317	132,278
	6	134,063	132,694	132,156	131,858	131,622	131,595
	7	133,727	132,321	131,822	131,498	131,318	131,277
	8	133,471	132,072	131,570	131,255	131,105	131,081
	9	133,340	131,918	131,406	131,094	130,932	130,918
	10	133,270	131,846	131,333	130,979	130,839	130,850
	11	133,218	131,789	131,263	130,944	130,792	130,819
	12	133,189	131,753	131,246	130,909	130,789	130,820
	13	133,185	131,738	131,229	130,916	130,809	130,842
$P = 2$	1	144,324	142,145	140,528	140,023	139,895	139,476
	2	139,435	137,861	137,448	135,611	135,158	134,987
	3	136,022	134,962	134,606	133,910	133,895	133,733
	4	134,937	133,842	133,405	133,058	132,800	132,828
	5	134,055	132,957	132,539	132,144	131,940	131,960
	6	133,645	132,462	131,972	131,710	131,610	131,579
	7	133,329	132,149	131,631	131,361	131,186	131,313
	8	133,093	131,943	131,420	131,149	131,057	131,085
	9	132,980	131,786	131,247	130,963	130,870	130,967
	10	132,901	131,694	131,170	130,839	130,728	130,804
	11	132,825	131,613	131,098	130,791	130,685	130,793
	12	132,799	131,597	131,087	130,767	130,699	130,799
	13	132,808	131,590	131,061	130,765	130,705	130,813
$P = 3$	1	144,202	142,027	140,501	140,079	139,857	139,320
	2	139,309	137,761	137,429	135,661	135,200	135,149
	3	135,888	134,938	134,474	133,990	133,788	133,663
	4	134,820	133,787	133,361	132,954	132,855	132,374
	5	134,051	133,002	132,629	132,366	132,240	132,133
	6	133,642	132,544	132,121	131,914	131,855	131,898
	7	133,329	132,226	131,775	131,599	131,571	131,601
	8	133,091	132,018	131,561	131,371	131,340	131,472
	9	132,978	131,860	131,379	131,187	131,158	131,325
	10	132,909	131,768	131,294	131,060	131,017	131,173
	11	132,828	131,690	131,231	131,015	130,981	131,147
	12	132,800	131,663	131,206	130,990	130,998	131,151
	13	132,817	131,672	131,213	130,997	130,993	131,193

1. Project the point ‘55+’ onto rating vector ‘9’.
2. Multiply the length of the projection with the length of rating vector ‘9’.
3. The resulting length is the estimated effect, which also equals the inner product of the vector ‘9’ and the vector that connects the point ‘55+’ to the origin.

In Figure 3.1a, all points of age have been projected onto rating ‘9’. In the model, the projected effects are multiplied by the length of the vector ‘9’, and the result is shown in Figure 3.1b. As in vector models (see, for example, DeSarbo et al., 2008), the orientation of the rating vector points in the direction of higher utility for the corresponding rating category. The projection of the point of an age category onto the rating vector indicates the relative magnitude of the preference for that rating category among respondents of the corresponding age category. From Figure 3.1, it can be seen that ‘9’ is relatively more often ticked by older respondents than by younger respondents.

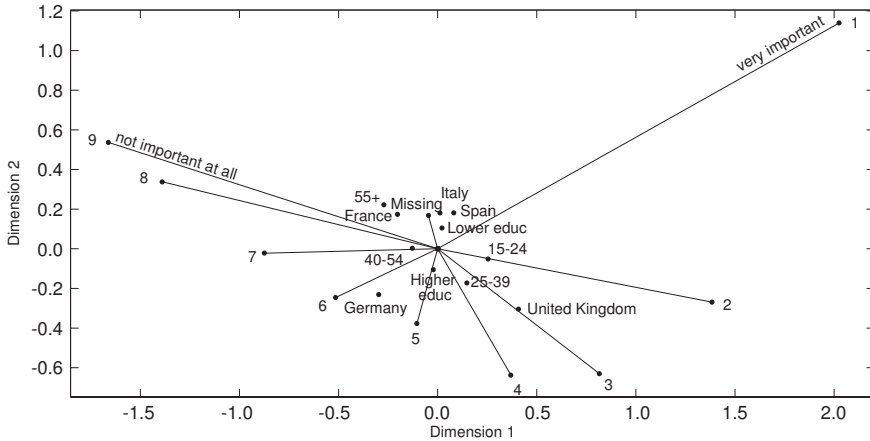
Thus, the representation has the following properties. Long vectors tend to correspond to large effects. To determine the sign of an effect, we consider the angle between a vector of Rating and the vector that connects a point of a predictor variable to the origin (note that the latter vectors are not shown in Figure 3.1b). If this angle is smaller than 90 degrees, the corresponding estimated effect is positive. If, on the other hand, the angle is greater than 90 degrees, the estimated effect is negative. The estimated effect becomes stronger as the angle gets closer to either 0 or 180 degrees. For example, the effect of ‘15-24’ on rating ‘9’ is negative, implying that younger respondents tick this category less often.

In Figure 3.2, we also include the points of country and education, so that this figure shows the effects of all three background variables. We can see that the structure of the categories of Rating, except for ‘missing’, resembles a U-shape. These rating scale categories have retained their natural ordering (from ‘1’ to ‘9’) with respect to the first dimension. As a result, the scores of the background characteristics on the first dimension are positively related to the probability of ticking positive rating categories such as ‘1’ or ‘2’. For example, respondents between 15 and 24 years of age tend to tick ‘1’ and ‘2’ more often than people in the other age categories. Therefore, it seems natural to interpret this first dimension as representing the effects of items and background characteristics on a scale of positivity-negativity.¹ On the second dimension, the most extreme rating categories ‘1’ and ‘9’ have the highest scores, whereas moderate responses such as ‘3’, ‘4’, and ‘5’ have low scores. Therefore, we can infer that the categories of the background characteristics on the second dimension are positively related to the probability of choosing extreme response categories; we can interpret this dimension as representing the effects of items and background characteristics on the extremity of a response. For example, elderly (55+) and lower educated respondents tend to use the extreme response categories more

¹However, one should be careful in assigning interpretations to the dimensions, as the graphical representations can, in principle, be rotated arbitrarily; here, they are rotated so that the greatest variation of the coordinates occurs along the first dimension, see Appendix 3.A.

often than other people. The point for the category of the missing values lies close to the origin, which means that the effects of the predictor variables on this response category are likely relatively small.

Figure 3.2: Graphical Representation of Effects of Background Characteristics on Response Behavior



Notes: The rating categories are represented by vectors, and the categories of the background variables are shown as dots.

In short, Figure 3.2 shows that respondents with a low education level are more prone to ticking extreme response categories than respondents with a high level of education. Age appears negatively related to the probability of ticking the positive side of the rating scale and seems to have a curvilinear relationship with the probability of ticking extreme response categories. Moreover, the respondents in Southern European countries show a stronger tendency to tick extreme response categories than respondents in the U.K. and Germany.

Item Segments

To show how content differs across the item segments, we provide a separate graphical representation of the effects of Item on Rating for each item segment. In Figures 3.3 and 3.4, the LOV items are shown as dots with labels, the categories of the background variables as dots without labels, and the rating categories are again shown using vectors. The effects of Item on the response behavior tend to be larger than the effects of the background variables, as the points belonging to Item are generally farther away from the origin. In addition, the differences in content among the item segments are quite large. For example, a large majority of the respondents in item segment C evaluated 'sense of

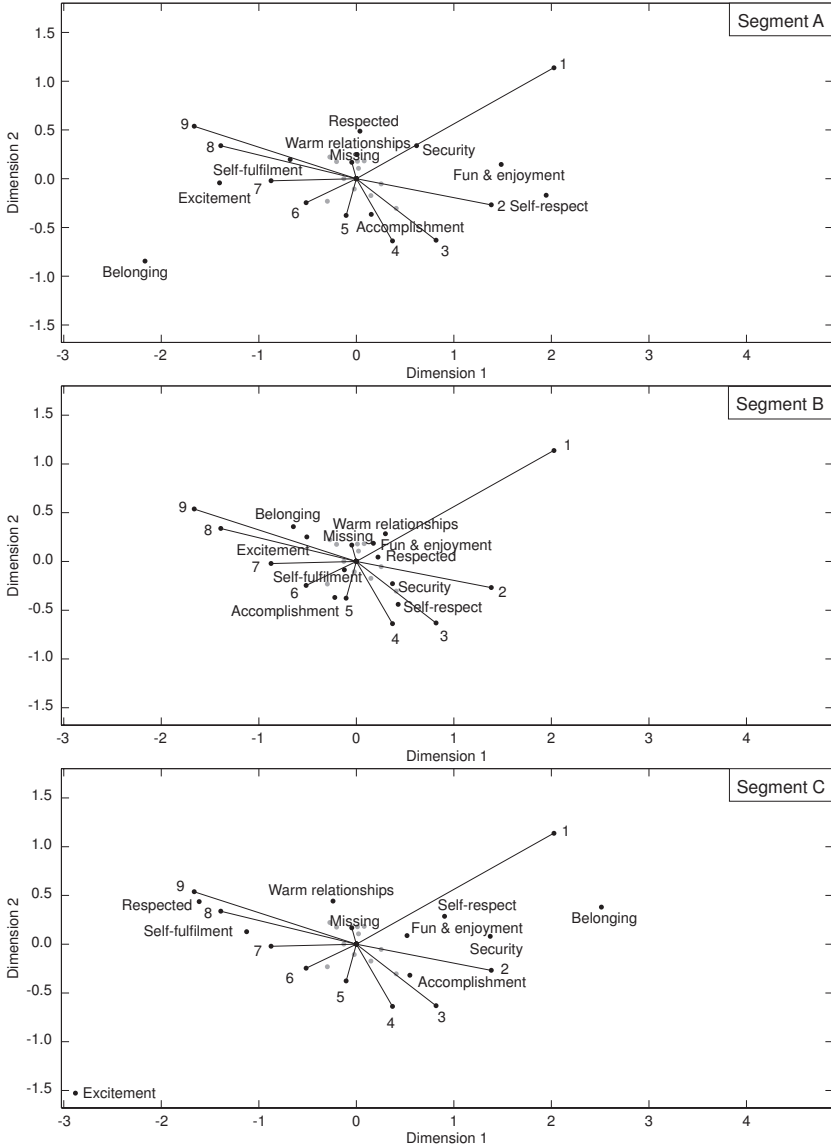
belonging' with a rating of '1', whereas almost no respondent in item segment A ticked '1' for this item.

The segments can be described as follows. In item segment A ('hedonists'), projections on the vectors '1' and '2' show that the LOV items 'fun & enjoyment' and 'self-respect' are considered important, whereas 'belonging' is not. The items 'being well-respected' and 'warm relationships' evoke the extreme responses '1' and '9' relatively more often than the other LOV items, implying that respondents who consider these LOV-items 'very important' and respondents who consider these items 'not important at all' belong to this segment. We label this segment 'hedonists', as the emphasis for those respondents seems on serving the individual interest and especially enjoyment. Segment B is labeled 'indifferent', because the effects of the LOV items are relatively small, as all item points lie close to the origin. The respondents in this segment seem to consider 'warm relationships', 'respected', 'security', and 'self-respect' somewhat more important than 'belonging' and 'excitement'. However, a clear opinion on the items (focusing on either group or individual interest) is not found; this segment also cannot be connected to earlier findings in the literature. In segment C ('group-oriented'), respondents clearly consider 'belonging' the most important item; 'excitement', which serves the individual interest more, is not considered important. 'Self-respect' and 'security' are also considered important. Such opinions on the LOV items have been found earlier by Grunert and Scherhorn (1990) and Kropp et al. (2005). Given the emphasis on values in the maturity and security domain, in which group interest is important (Kamakura & Novak, 1992), we label this segment 'group-oriented'. In segment D ('self-oriented'), 'self-respect' is considered the most important, though 'accomplishment' and 'security' also tend to get positive responses; 'belonging' is considered far less important, it relatively often gets the score '9'. People in this segment seem mainly concerned with achieving personal goals, and enjoyment is of secondary importance; therefore, we label this segment as 'self-oriented'. Previous studies (Chrysohoidis & Krystallis, 2005; Kamakura & Novak, 1992) found a similar preference for LOV items in certain groups. Finally, in segment E ('mixed opinions'), some items get relatively more extreme responses than other items; the items 'self-respect', 'accomplishment', and 'fun and enjoyment' all are often rated '1' or '9'. The items 'belonging' and 'excitement' tend to get scores from '2' to '4'. The answers given in this segment cannot be clearly linked to a values domain, and we therefore label segment E as 'mixed opinions'. The respondents in the latter segment, as in segment B ('indifferent'), do not appear to have strong opinions on what items are important to them. Moreover, there are respondents in segment E who appear to consider 'self-respect' not important at all and 'excitement' of moderate importance; both results have not been reported in the literature before.

Item Segments and Respondent Characteristics

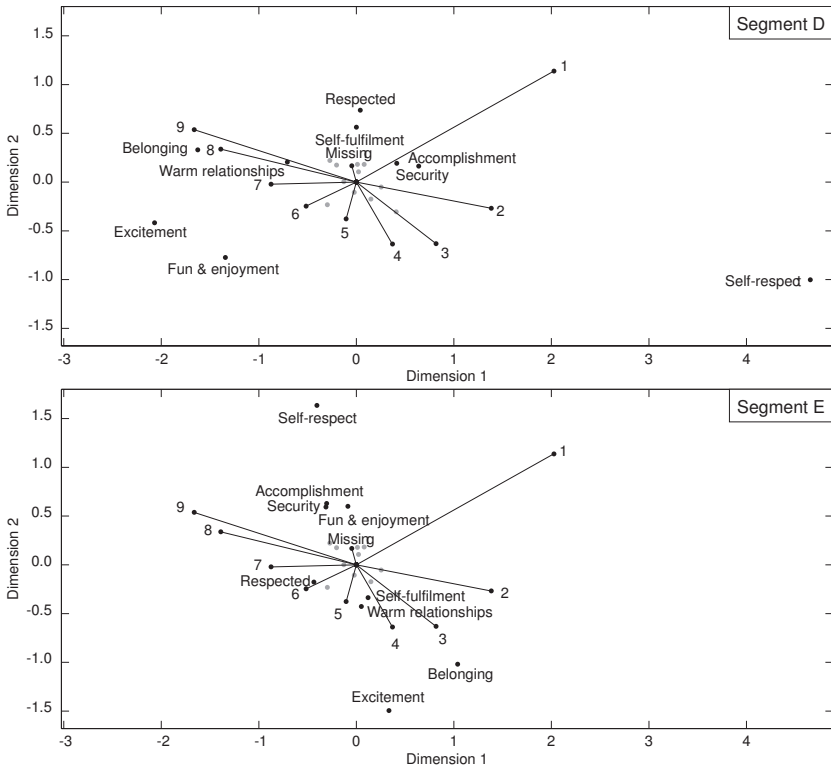
The item segments should satisfy the six criteria for effective lifestyle-based segments (see Wedel & Kamakura, 2000); these criteria are substantiality, identifiability, accessibility, stability, responsiveness, and actionability. All five segments have a substantial size

Figure 3.3: Graphical Representations of Effects of LOV Items on Response Behavior for Item Segments A through C.



Notes: The rating categories are represented by vectors, and the items and the categories of the background variables are shown as dots. The categories of the background variables are not labeled in these plots.

Figure 3.4: Graphical Representations of Effects of LOV Items on Response Behavior for Item Segments D and E.



Notes: The rating categories are represented by vectors, and the items and the categories of the background variables are shown as dots. The categories of the background variables are not labeled in these plots.

Table 3.3: Item Segment Membership Percentages for Each Category of Each Background Variable

Interpretation	Item segment				
	A Hedonists	B Indifferent	C Group-oriented	D Self-oriented	E Mixed opinions
<i>Country</i>					
United Kingdom	20.9	24.9	18.3	18.6	17.3
France	53.1	27.7	5.1	6.1	8.1
Germany	9.7	12.7	62.0	4.2	11.5
Italy	22.6	24.6	3.3	38.5	11.1
Spain	38.0	33.9	2.3	16.2	9.7
<i>Education</i>					
Lower educ	25.7	24.3	22.1	16.7	11.2
Higher educ	32.7	24.4	13.9	17.1	11.9
<i>Age</i>					
15-24	33.5	29.0	10.8	11.6	15.0
25-39	33.2	26.1	14.2	13.7	12.8
40-54	25.6	21.5	22.9	19.3	10.8
55+	23.5	22.8	24.1	20.7	9.0
Segment size	28.5	24.3	18.9	16.9	11.5

(see final row in Table 3.3), as even the smallest segment contains more than 10% of the sample. Identifiability corresponds to the extent to which distinct groups can be recognized using a segmentation. Each segment has a different profile on the LOV items and can be distinguished from the other segments. Furthermore, when the respondents' background characteristics are also different, the segments can be targeted by a company. Table 3.3 shows, in percentages, the distribution of respondents from each category of each background variable over the item segments. The percentages in this table are calculated using the posterior segment membership probabilities in Equation 3.6. The results in Table 3.3 show that the distribution of the respondents over the item segments depends on their age, education level, and country of residence, which suggests that the item segments are identifiable. We will come back to the other four criteria for effective segmentation in the discussion.

Using the information in Figures 3.3 and 3.4 and Table 3.3, we can further interpret the item segments. We distinguish three segments that seem to meet expectations regarding content (segments A, C, and D) and two segments that do not (segments B and E). In the largest item segment A (28.5%, 'hedonists'), respondents attach high importance to the LOV-items 'fun & enjoyment', and 'self-respect'; the values 'sense of belonging' and 'excitement' are considered much less important. In this segment, relatively many respondents are from France and Spain, and they are relatively young and highly educated. The relatively high percentage of young respondents confirms results of Kamakura and Novak (1992), who also found that 'fun & enjoyment' is considered important by younger people. Unlike the other segments, respondents in item segment C (18.9%, 'group-

oriented'), consider 'belonging' the most important value. The majority in this segment is from Germany, and a majority is lower educated and older than 40. As in most other segments, in item segment D (16.9%, 'self-oriented'), 'self-respect' is considered very important. Respondents in this segment are relatively older and often come from Italy. Segments C and D are also similar to results of previous studies; for example, segment C ('group-oriented') clearly resembles the results found for German respondents by Wedel et al. (1998). Item segment B ('indifferent') is relatively large (24.3%). People in this segment are relatively often from Spain, and a majority in this segment is young. Item segment E ('mixed opinions') is the smallest segment (11.5%) and contains relatively many young and British respondents. Unlike the three other segments, segments B and E are not similar to segments found in earlier studies and seem to include respondents with less outspoken and opaque opinions; we come back to this finding later.

Response Style Segments

To show what response styles are present in the data, we focus on the response style segments. Table 3.4 contains the observed proportions of the rating categories in percentages for the respondents in each response style segment. We do not show the values of the parameters $a_{j|r}$ (which model the differences in response styles among the response style segments), as the proportions in Table 3.4 more clearly show what rating categories have been used by the respondents in each segment. Two elements in this table are exactly zero, because the estimates of the corresponding parameters $a_{j|r}$ converged to minus infinity. We believe that these parameter estimates are unlikely to affect the results negatively, as it is conceivable that some subgroups of respondents do not use all rating scale categories.

The 11 response style segments identified by our model fit into the possible response styles distinguished by Broen and Wirt (1958). For example, in segments 1 through 3, respondents show different levels of acquiescence (a tendency to tick positive categories in rating scales, see Hofstede, 2001; often also defined as a tendency to agree with statements regardless of their content), and in segments 9 and 10, respondents show two levels of disacquiescence (a tendency to tick negative rating categories). In segment 7, respondents show an extreme response style, and in segments 4 and 5, respondents have a tendency to avoid extreme ratings. Interestingly, respondents in segment 8 show midpoint scoring and tick many extreme scores. In segment 11, respondents either tick the category '1' or do not tick a category (incomplete response). The respondents in segment 10 seem to have a high level of disacquiescence, as Table 3.4 shows that over 80% of their responses are either '7', '8', or '9'. However, a closer inspection of the raw data shows that values that are considered relatively important by most respondents are rated as relatively unimportant by the respondents in segment 10, and vice versa. Therefore, we conclude that most respondents in segment 10 must have misread the rating scale and incorrectly thought that '9' was the most positive rating value. This phenomenon was called "misresponse" by Swain, Weathers, and Niedrich (2008), who studied its prevalence for rating scales that contain both reversed and nonreversed items; our results indicate that misresponse can also occur in surveys without reversed items. Only a small number of respondents show

Table 3.4: Relative Frequencies of Rating Categories in Percentages and Response Style Indices for Each Response Style Segment

Segment	Size	1	2	3	4	5	6	7	8	9	Miss.	ARS	DARS	ERS	RR	MPR	NCR
1. Strong acquiescence	17.5	71.4	7.0	5.5	3.8	6.3	1.5	1.1	.8	2.6	.1	78.4	3.4	74.0	1.7	6.3	.8
2. Moderate acquiescence	16.7	46.7	32.5	12.0	4.2	2.2	.7	.7	.5	.6	.0	79.2	1.1	47.2	1.2	2.2	.8
3. Weak acquiescence	24.2	38.6	16.4	18.1	8.9	8.4	2.7	2.7	1.5	2.3	.5	55.2	3.8	41.1	1.9	8.4	1.4
4. Nuanced positive	14.1	12.0	42.0	20.3	10.0	7.9	3.3	2.0	1.2	1.2	.2	54.0	2.4	13.2	1.5	7.9	1.1
5. Moderate	6.5	4.4	11.6	33.8	23.6	14.8	5.1	3.4	1.7	1.0	.6	16.1	2.7	5.4	1.4	14.9	1.2
6. Wide response range	9.3	20.2	16.0	12.0	10.2	11.7	6.8	8.1	7.3	7.5	.1	36.2	14.8	27.7	2.5	11.7	2.3
7. Extreme scoring	4.1	58.8	2.7	2.3	5.9	3.1	1.8	1.2	1.8	21.4	1.0	62.1	23.5	81.0	3.1	3.1	2.4
8. Midpoint scoring	2.5	33.8	4.0	11.4	2.3	34.8	2.1	.6	.0	10.5	.5	38.0	10.6	44.5	2.4	35.0	2.0
9. Weak disacquiescence	2.3	1.4	2.2	5.2	6.7	21.4	21.0	19.3	15.3	6.6	1.0	3.6	22.1	8.0	1.6	21.6	1.4
10. Strong disacquiescence	1.8	1.4	.5	1.6	1.6	5.4	4.6	15.5	19.5	48.8	1.0	1.9	69.0	50.7	1.5	5.5	1.0
11. Incomplete response	.9	26.7	1.8	2.7	1.3	3.9	1.3	.3	.8	.7	60.6	72.3	3.8	69.5	1.0	9.8	.7
Entire sample		37.0	19.0	14.1	7.9	8.3	3.2	3.0	2.3	4.2	.8	56.6	6.6	41.6	1.7	8.4	1.3

Notes: The response style indices in the last six columns are ARS (acquiescence response style), DARS (disacquiescence), ERS (extreme response style), RR (response range), MPR (midpoint responding), and NCR (noncontingent responding).

misresponse (the size of segment 10 is 1.8%), but due to the severity of the mistake, it can have major consequences for the results of a study; the responses of some people in segment 9 may also suffer from misresponse.

The last columns of Table 3.4 show average values of six response style indices (see Baumgartner & Steenkamp, 2001, for an overview of these indices). These response styles indices are acquiescence response style (ARS), disacquiescence response style (DARS), extreme response style (ERS), response range (RR), midpoint responding (MPR), and noncontingent responding (NCR). Following Baumgartner and Steenkamp (2001), ARS is defined as the percentage of responses that is '1' or '2', DARS is the percentage of responses that is '8' or '9', ERS is the percentage of responses that is '1' or '9', RR is the estimated standard deviation of a respondent's ratings, MPR is the percentage of the responses that is '5', and NCR is defined here as the average difference between the ratings of pairs of items that have a correlation greater 40% in the entire sample. Missing observations have been excluded in the calculation of the response style indices.

These response style indices can help interpret the response style segments, as most segments can be characterized by a high level for one of the indices. An interesting result is that MPR seems to evoke the extreme categories as well; respondents showing midpoint scoring seem to use only a small part of the rating scale (the two extremes and the midpoint). Segment 11 shows that respondents with many missing values also tend to score high on ARS. To the best of our knowledge, the latter two response tendencies have not been described previously. In summary, using our model we find 11 response styles that closely resemble, but do not completely overlap with earlier descriptions of response styles.

Dependence between Segmentations

In the LC-BML model, the item segments and response style segments are not independent. Table 3.5 shows the segment sizes of all combinations of a response style segment and an item segment and, in parentheses, their expected values based on independence of the two segmentations. For item segments A, C, and D, the sizes of the segments are relatively close to their expected values based on independence of the segmentations. In item segments B and E, there is clear dependence between the two types of segments. In item segment B, there is an overrepresentation of response style segments 5 through 8, that is a moderate response style, wide response range, extreme scoring, and midpoint scoring. In item segment E, response style segments 1, 9, and 10 are overrepresented, implying that respondents in this item segment tend to show strong acquiescence (response style segment 1) or disacquiescence (segments 9 and 10) more often than respondents in the other item segments. These results indicate that less outspoken and opaque opinions, as in item segments B ('indifferent') and E ('mixed opinions'), coincide with specific response styles.

Table 3.5: Response Style Segments versus Item Segments, Segment Sizes in Percentages.

Response style segments	Item segments										Total
	A		B		C		D		E		
Interpretation	Hedonistic		Indifferent		Group-oriented		Self-oriented		Mixed opinions		
1. Strong acquiescence	4.8	(5.0)	.0	(4.3)	3.6	(3.3)	3.2	(3.0)	5.8	(2.0)	17.5
2. Moderate acquiescence	5.3	(4.7)	1.1	(4.1)	5.0	(3.1)	4.2	(2.8)	1.0	(1.9)	16.7
3. Weak acquiescence	9.2	(6.9)	6.1	(5.9)	5.3	(4.6)	3.6	(4.1)	.0	(2.8)	24.2
4. Nuanced positive	5.7	(4.0)	.9	(3.4)	2.8	(2.7)	4.6	(2.4)	.0	(1.6)	14.1
5. Moderate	1.8	(1.9)	3.8	(1.6)	.8	(1.2)	.2	(1.1)	.0	(.7)	6.5
6. Wide response range	.6	(2.7)	6.2	(2.3)	1.0	(1.8)	.9	(1.6)	.6	(1.1)	9.3
7. Extreme scoring	.4	(1.2)	2.8	(1.0)	.2	(.8)	.0	(.7)	.6	(.5)	4.1
8. Midpoint scoring	.3	(.7)	2.1	(.6)	.0	(.5)	.1	(.4)	.0	(.3)	2.5
9. Weak disacquiescence	.0	(.7)	.7	(.6)	.0	(.4)	.0	(.4)	1.6	(.3)	2.3
10. Strong disacquiescence	.1	(.5)	.0	(.4)	.0	(.3)	.0	(.3)	1.7	(.2)	1.8
11. Incomplete response	.1	(.3)	.6	(.2)	.1	(.2)	.0	(.2)	.1	(.1)	.9
Total	28.5		24.3		18.9		16.9		11.5		

Notes: Expected segment sizes based on independence of the segmentations have been calculated as the product of row and column totals and are shown in parentheses.

Reliability of Results

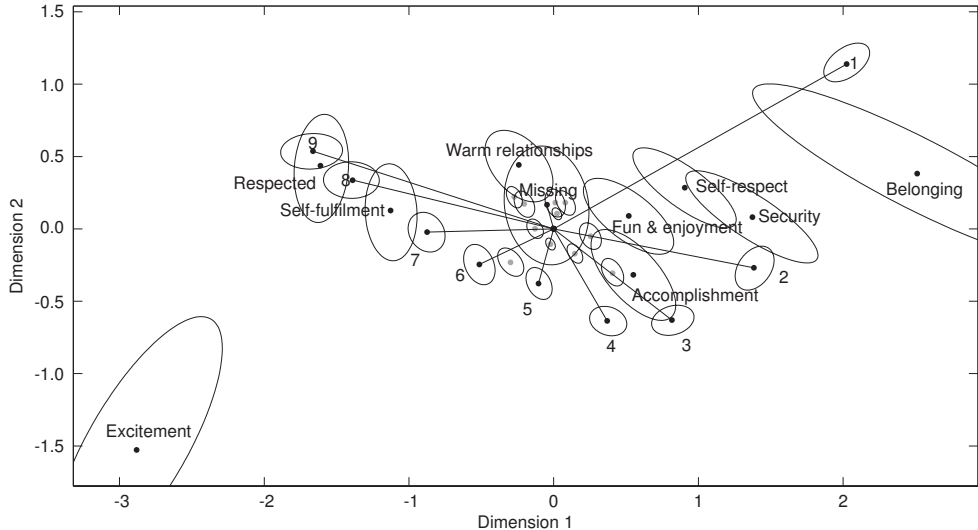
To show the statistical significance of our results, we have constructed confidence ellipses for the points in the graphical representations. Figure 3.5 shows 95%-confidence ellipses for all points shown in the graphical representation of item segment C; the other segments have comparable confidence ellipses. An explanation of how the confidence ellipses are constructed is given in Appendix 3.E.

The amount of uncertainty in the graphical representation appears to be low. The confidence ellipses of the items are larger than the ellipses of the ratings and the background variables, as only the respondents in item segment C determine the locations of the points of the items in this plot. The locations of all points (except for ‘missing’) are significantly different from the origin, so that the effects of all predictor variables are statistically significant. As the location of ‘missing’ is not significantly different from the origin, there is insufficient evidence to conclude that missing values are informative with respect to the opinions of respondents. The confidence ellipses of Rating do not overlap with each other, which means that all rating categories are distinguishable with respect to the predictor variables (see J. A. Anderson, 1984, for more information on the distinguishability of rating categories).

External Validation

We also use external criteria to give an indication of the effectiveness of the LC-BML model for assessing the opinions of respondents. The List of Values is a well-known general basis for segmentation (Wedel & Kamakura, 2000) and is thus expected to have some

Figure 3.5: Graphical Representation of Item Segment C with 95% Confidence Ellipses for All Points.



explanatory power. If our model works well, the results of the model should be able to help predict consumer behavior (Kamakura & Novak, 1992).

As was described previously, the results of the model can be used to construct response probability distributions that are corrected for differences in response styles. For each observation, it is also possible to construct a score that is corrected for differences in response styles, by taking the mean of the response distribution that is corrected for response styles (excluding the 'missing' category). These scores should be good predictors of a respondent's values and thus be related to other variables. If these scores accurately represent a respondent's values, there should be statistically significant correlations between these corrected scores and other variables. If such scores are still affected by differences in response styles, correlations with external variables that are unaffected by response styles should be expected to be smaller.

To properly assess the sizes of these correlations, external variables should be chosen that are unaffected by response styles, to avoid spurious correlations caused by response styles. We use behavioral variables as external variables, as such variables are typically considered to be much less sensitive to response styles than attitude or values scales (Greenleaf, 1992a). We compare the scores corrected for response styles using the LC-BML model with four other methods used in marketing research. These four methods are (a) using the original responses without correcting for response styles, (b) ipsatizing the LOV responses for each respondent, (c) correcting the observed scores by regressing them on various response style indices, and (d) the Bayesian hierarchical method for correcting

Table 3.6: Predictive Performance of Five Methods for Correcting for Response Styles Based on 20 Behavioral Variables

Method	Mean absolute correlation	Percentage of significant correlations
Original data	.0350	45.6
Ipsatized data	.0370	48.9
Scores based on response style indices	.0348	45.6
Scores based on Rossi et al. (2001)	.0373	48.9
Scores of LC-BML model	.0489	65.0

for scale usage heterogeneity proposed by Rossi et al. (2001). Method (c) uses the response style indices shown in Table 4 and was proposed by Baumgartner and Steenkamp (2001). Method (d) uses as scores the posterior means of the latent variables z_{ji} in Rossi et al. (2001). This method was performed using the bayesm package in the R programming language.

We use a set of 20 behavioral variables on the usage frequencies of personal care products, for the same respondents who evaluated the LOV items. For each of the five methods described above, we compute a matrix of correlation coefficients between the 20 items on usage frequencies of personal care products and the nine LOV items. Table 3.6 shows, for each method, the average of the absolute values of the correlations and the percentage of the correlation coefficients that are statistically significant at the 5% level. The mean absolute correlations are approximately .04, which seems low; such correlations are, however, common and acceptable in cases where values and behavior are related (see, for example, Kamakura & Novak, 1992). The scores of all five methods appear to be correlated with the behavior variables, as the percentages of significant correlations are clearly greater than 5%.

The elements of the correlation matrices cannot be assumed to be independent, as the methods for removing response styles do not ensure independence among the LOV items. However, it seems reasonable to suppose that the differences between the methods in average absolute correlations (with the averages taken over the nine LOV items) are independent. For 19 out of 20 behavioral variables (sign test, $p < .001$), the average absolute correlations for the LC-BML model are higher than the average absolute correlations for the original data; the same result is obtained when comparing scores of the LC-BML model with the ipsatized data and with the scores based on the model of Rossi et al. (2001). For all 20 variables (sign test, $p < .001$), the average absolute correlations for the LC-BML model are higher than those that were obtained using regressions on the response style indices. The scores of the LC-BML model thus appear to yield the best predictive performance, as the absolute correlations are highest for this method.

The scores of the LC-BML model are based on the posterior segment membership probabilities. To show how the predictive validity varies among the item segments, we also calculate the mean absolute correlations of the segment memberships and the behavioral variables for each item segment. This mean absolute correlation is relatively high for item segments A, C, and D (.049, .073, and .064, respectively) compared to segments B and E

(mean absolute correlations .027 and .039). These findings again indicate less useful results in the latter two segments.

3.5 Conclusions and Discussion

In this chapter, we have proposed a new model, called the LC-BML model, for explaining response behavior and distinguishing between content and response styles in data sets with rating scale responses. Specifically, we found a variety of response styles as suggested by Broen and Wirt (1958). The most prominent response style is acquiescence (58.4% of respondents), followed by avoiding the extreme scores (20.6%), using a wide range of responses (9.3%), extreme responding (4.1%), disacquiescence (4.1%), and midpoint scoring (2.5%). A small group of respondents (.9%) did not tick a rating category for most of the LOV items.

The LC-BML model distinguishes between the effects of the content of an item, background characteristics, and the response style of a respondent, which contributes to our understanding of response behavior in the following ways. First, the model uses graphical representations (biplots), enabling researchers to see how respondents use the rating scale and what kinds of opinions respondents have. Our results show that item content is highly effective in explaining differences in response behavior; there are large differences in the effects of the items on response behavior both within and among item segments. The effects of the background characteristics (that is, country, age, and education) on response behavior tend to be much smaller than the effects of the content of the items. Second, the LC-BML model reveals what response styles are used by the respondents without imposing any restrictions on the types of response styles. This freedom allows us, for example, to infer that some respondents reversed the scale and that others almost solely used the rating score '1' and the missing category. This latter group of respondents is not identified by traditional methods for assessing response styles. Third, unlike other models, the LC-BML model does not assume an ordering of the dependent variable Rating and can be used to investigate response styles in rating scales that include a 'don't know' category or have missing observations. By assigning a location to each rating category, our method can provide insight into the relative position of such categories. For the current data set, the location of the 'missing' category in the plot suggests that missing values can provide only limited information about a respondent's opinions. However, it is likely that missing values or 'don't know' responses can provide useful information in other surveys; for example, surveys including sensitive questions, in which item nonresponse is common (Tourangeau, Rips, & Rasinski, 2000), may benefit from our method.

The LC-BML model shows the possible detrimental effect of ignoring response styles in cross-national survey research. Our results show two item segments (B and E) together containing 36% of the sample in which respondents have biased response patterns. The interaction between content and response styles reveals that certain response styles, such as extreme response style, strong acquiescence, disacquiescence, and incomplete response, are overrepresented in these two item segments. Given the outspoken nature of these

response styles, we believe that these response styles are often not valid communication styles. Other response styles, such as moderate and weak acquiescence, appear to be valid communication styles. For marketing applications, it is better to drop item segments such as segments B and E from targeting activities. Focusing on segments that provide valid information can increase marketing effectiveness.

To provide a basis for the formulation of marketing strategies, the item segments should satisfy the six criteria for effective segmentation (Wedel & Kamakura, 2000). Our item segments are based on psychographics, which is a general unobservable segmentation basis. The criteria identifiability and substantiality were already established in our study. Stability is also likely, as values do not change dramatically over time (Schwartz, 1992). Further, accessibility refers to the extent to which the segments can be reached through promotion and distribution efforts. Especially the segments 'hedonists', 'group-oriented', and 'self-oriented' show differences with respect to the background characteristics country, age, and education; to reach those segments, the background information can be used to link the segments to existing commercial databases on, for example, media usage. Finally, responsiveness and actionability may not seem evident for the item segments. Responsiveness is satisfied when each segment responds in a unique way to the marketing mix, and actionability depends on whether the segments provide guidance for effective specification of the marketing instruments. Because of the psychographics used, we find only minor differences in behavior among segments. However, the substantial differences among segments in value structures and background characteristics suggest that differences in behavior exist. To increase responsiveness and actionability of the segments, product-specific attitudes measured using rating scales can be included in future studies. Unlike segments based on either demographics or behavior, segments based on product-specific (or domain-specific) attitudes are considered high on responsiveness and actionability (Wedel & Kamakura, 2000). Thus, applying the LC-BML model simultaneously to domain-specific variables and general psychographics or values may help find item segments that meet all criteria for effective segmentation.

As any study, the present study has some limitations. First, the number of items and the number of countries were limited. Second, the LC-BML model uses an optimization algorithm that may be quite time-consuming. This is especially true if the data set contains many observations and background variables, safeguards against local maxima are taken, and the numbers of segments are determined using an information criterion, so that the model parameters have to be optimized many times. Third, the LC-BML model cannot be estimated using conventional, commercially available statistics packages, so that specialized programs had to be written. These programs have been written in the matrix programming language MATLAB and are available from the authors upon request. Finally, the specific response styles found in our data set may differ from the response styles in other data sets. However, we believe that the response styles found here are representative of the response styles in many data sets in marketing research.

Our study offers several directions for future research. First, the LC-BML model has been designed for data sets with substantive scales measured on rating scales and

background characteristics measured on nominal scales. Many data sets in international marketing research fit this format. Future research can focus on using our model to assess the occurrence of various response styles worldwide, especially in emerging markets where different reactions to Western-based scales are common (Burgess & Steenkamp, 2006). In addition, the LC-BML model can be used to gain insight into the meaning of categories such as ‘don’t know’ or ‘missing’; this insight seems especially useful if those categories are not ticked completely at random. Future research may increase our understanding of how these response categories are used and whether such categories really reflect the absence of an opinion. Finally, our model gives estimated probability distributions for the response categories. Future studies can investigate whether such probability distributions can be meaningfully used for imputing missing observations in a data set.

In this research, we concentrated on marketing data from several countries. Marketing research providers are global enterprises. For example, ACNielsen currently has subsidiaries in 108 nations and conducts research all over the world. Such large research enterprises have proprietary ways to adjust their polling methods and surveys, as they operate in so many cultures and environments. Nonetheless, the method described here would be useful to any marketing research company (large or small) and is applicable to academic research, where there is increasing interest in testing traditional marketing models and practices in a variety of nations and cultures. In a marketing context, the LC-BML model is a useful tool to provide more insight into how rating scale responses are affected by response styles, the content of the items, and background variables. In brief, we believe that the LC-BML model adds a promising technique to the international marketing researcher’s toolkit.

3.A Parameter Restrictions

Several parameter restrictions are necessary for identification of the model parameters. Here, we describe the restrictions that we impose. First, we consider location restrictions, that is, restrictions that are necessary because adding a constant to a set of parameters would not change the estimated probabilities in the model. Location restrictions are needed for a_{jr} , \mathbf{F} , \mathbf{G}_k , and \mathbf{H}_s . We impose sum-to-zero constraints for all these parameters per set, that is, we require that $\sum_{j=1}^J a_{jr} = 0$, $\sum_{j=1}^J f_{jp} = 0$, $\sum_{t=1}^T h_{tp|s} = 0$, and that the elements of \mathbf{g}_{kp} sum to zero. As a result, the centroid of the points referring to a single variable (Rating, Item, or a background variable) is the origin in the plot.

Scale and rotation constraints are required for \mathbf{F} , \mathbf{G}_k , and \mathbf{H}_s . For notational convenience, let $\mathbf{B} = [\mathbf{B}_1 | \dots | \mathbf{B}_K]$, $\mathbf{C} = [\mathbf{C}_1 | \dots | \mathbf{C}_S]$, $\mathbf{G} = [\mathbf{G}'_1 | \dots | \mathbf{G}'_K]'$, and $\mathbf{H} = [\mathbf{H}'_1 | \dots | \mathbf{H}'_S]'$, where \mathbf{C}_s contains the effects of the ratings on the items in item segment s . Then, the rank restrictions in Equation 3.2 can be written as

$$[\mathbf{B}|\mathbf{C}] = \mathbf{F}[\mathbf{G}'|\mathbf{H}'] \tag{3.8}$$

Without loss of generality, we can transform the parameter matrices as $\mathbf{F} = \mathbf{F}\mathbf{T}$, $\mathbf{G} = (\mathbf{T}^{-1}\mathbf{G}')'$, and $\mathbf{H} = (\mathbf{T}^{-1}\mathbf{H}')'$ for any nonsingular $P \times P$ matrix \mathbf{T} , because

$$\mathbf{F}[\mathbf{G}'|\mathbf{H}'] = \mathbf{F}\mathbf{T}\mathbf{T}^{-1}[\mathbf{G}'|\mathbf{H}']$$

This freedom of scaling also occurs in principal components analysis and correspondence analysis (Gifi, 1990). We compute a constrained solution as follows. Let $[\mathbf{B}|\mathbf{C}]$ be obtained by some unconstrained \mathbf{F} , \mathbf{G} , and \mathbf{H} . In addition, let $\mathbf{P}\Phi\mathbf{Q}' = [\mathbf{B}|\mathbf{C}]$ be a compact singular value decomposition with \mathbf{P} and \mathbf{Q} orthogonal rotation matrices with P columns (with $\mathbf{P}'\mathbf{P} = \mathbf{Q}'\mathbf{Q} = \mathbf{I}$) and Φ a $P \times P$ diagonal matrix with positive monotonically decreasing values. As the rank of $[\mathbf{B}|\mathbf{C}]$ is not greater than P , matrices \mathbf{P} , Φ , and \mathbf{Q} that meet the requirements above must exist and typically are unique up to a reflection per dimension. We set $\mathbf{F} = \omega\mathbf{P}\Phi^{1/2}$ and $[\mathbf{G}'|\mathbf{H}']' = 1/\omega\mathbf{Q}\Phi^{1/2}$, where ω is a constant that determines the relative scaling of the points of the ratings compared to the points of the items and the background characteristics. Choosing \mathbf{F} and $[\mathbf{G}'|\mathbf{H}']'$ in this way ensures that the equality in Equation 3.8 is preserved, as

$$\mathbf{F}[\mathbf{G}'|\mathbf{H}'] = \omega\mathbf{P}\Phi^{1/2} \left(\frac{1}{\omega}\mathbf{Q}\Phi^{1/2} \right)' = \mathbf{P}\Phi\mathbf{Q}' = [\mathbf{B}|\mathbf{C}].$$

The value of ω can be adapted without altering the general results. We set

$$\omega = \left(\frac{J}{ST + \sum_{k=1}^K m_k} \right)^{1/4},$$

so that the average squared Euclidean distance of the points to the origin is the same for both \mathbf{F} and $[\mathbf{G}'|\mathbf{H}']'$. For the LOV data set, this choice amounts to setting $\omega \approx .65$. We also simultaneously reflect the columns of \mathbf{F} , \mathbf{G} , and \mathbf{H} in such a way that the first row of \mathbf{F} only has positive values.

Due to these parameter restrictions, the spread of the points in the graphical representations decreases with the dimension, and the dimensions are orthogonal with respect to \mathbf{F} and $[\mathbf{G}'|\mathbf{H}']'$. The point of Rating 1 must have positive values on all dimensions. A final parameter restriction is that the segment sizes must be nonnegative and should sum to one, so that $u_{rs} \geq 0$ and $\sum_{r=1}^R \sum_{s=1}^S u_{rs} = 1$.

3.B Optimization Algorithm

An Expectation-Maximization (EM) algorithm (Dempster, Laird, & Rubin, 1977) is used to estimate the model parameters by maximizing the likelihood function. The EM algorithm starts with initial parameter estimates and then iteratively performs an E-step and a M-step, until convergence has been achieved. In the E-step, the posterior segment membership probabilities $\Pr(i \in \Xi_{r,s})$ are computed using Equation 3.6, given the current parameter estimates. In the M-step, the expected complete log-likelihood is maximized with respect to the parameter estimates, given the segment membership probabilities computed in the E-step. In our implementation, every M-step consists of 10 iterations of the BFGS quasi-Newton optimization routine in the MATLAB Optimization Toolbox (version 3.0.4), with analytically computed gradients. Convergence is considered to have been achieved if the change in log-likelihood between two consecutive EM iterations is smaller than 2×10^{-5} .

It is possible that the EM algorithm converges to parameter estimates that are only locally optimal. To solve this problem, the EM algorithm was run 10 times for every value of R , S , and P with randomly chosen starting values, and the solution with the best likelihood value was retained.

3.C Model Identifiability

An important issue is whether the model parameters are identifiable (see, for example, Teicher, 1963). If there are few items in the study and many segments, the model may not be identifiable, so that unique parameter estimates that maximize the likelihood function may not exist. For example, with only one item in the study, it is not possible to distinguish more than one response style or item segment.

To ensure model identifiability, the parameters must be locally identifiable. In addition, we must account for trivial nonidentifiability, which occurs because the order of the segments can be interchanged without changing the likelihood value; this problem can be solved by imposing appropriate restrictions of the segment sizes. Local identifiability is guaranteed if the Hessian (that is, the matrix of second derivatives of the log-likelihood function at the final parameter estimates) has full rank, which can easily be checked after the estimating the model parameters. In principle, it may be possible to derive conditions in which the model is always identifiable, which can be checked before estimating the parameters. However, because we have two types of segments, deriving necessary and sufficient conditions for identifiability is very hard and beyond the scope of this chapter.

Instead, we solve this problem empirically. After estimating the parameters of a model, we check whether the Hessian is of full rank (using the alternative parametrization described in Appendix 3.E), and we impose a specific ordering of the segments. Based on numerical experimentation, we find that the model is almost always identifiable if there are more than three items, and the LC-BML model is also identified for the solution presented in this chapter for the LOV data set. Most marketing segmentation studies use at least ten items and distinguish at most five segments; such studies will thus almost always result in identifiable models. Therefore, we do not believe that the identifiability of the LC-BML model poses a problem for marketing practitioners.

3.D Simulation Study

To apply the LC-BML model, the numbers of segments R and S and the dimensionality P need to be chosen. To do so, we use an information criterion. However, many information criteria exist, and their performances tend to depend on the characteristics of the model and the data set (Yang & Yang, 2007). To determine how various information criteria perform for our model and to evaluate the performance of the optimization algorithm, we conduct a simulation study. In this simulation study, we generate data according to simulated model parameters and then apply the optimization algorithm.

Table 3.7: Factors in Simulation Study

Factor	Level 1	Level 2
Number of respondents (n)	$n = 1000$	$n = 4000$
Number of rating categories (J)	$J = 5$	$J = 9$
Background variables	no background variables	2 variables with 2 categories each
Numbers of segments (R_{sim}, S_{sim})	$R_{sim} = 2, S_{sim} = 2$	$R_{sim} = 3, S_{sim} = 3$
Dimensionality (P_{sim})	$P_{sim} = 1$	$P_{sim} = 2$
Number of items (T)	$T = 8$	$T = 16$
Variance of parameters	1.2	.6

To generate simulated data sets, we simulate model parameters as follows. The elements of $a_{\cdot|1}, \dots, a_{\cdot|R}$, \mathbf{F} , $\mathbf{G}_1, \dots, \mathbf{G}_k$, and $\mathbf{H}_1, \dots, \mathbf{H}_S$ are chosen as independently normally distributed variables, in such a way that the parameter restrictions that are described in Appendix 3.A are met (so that the columns of these parameter matrices must sum to zero). The variance of these parameters is one of the factors in the simulation study and equals either .6 or 1.2. The elements of U are chosen to be $1/(R \times S)$, so that all segments have the same size.

In our simulations, we vary seven factors with two levels each, using a full-factorial design. These factors and the levels of these factors are shown in Table 3.7. The levels of these factors have been chosen in such a way, that the simulated data are roughly comparable to empirical data sets such as the LOV data set; to limit the computation time, the numbers of segments have been set lower than in the empirical study. We generate one data set for each of the $2^7 = 128$ combinations of the levels of these factors. For each data set, we apply the optimization method for $R = 1, \dots, R_{sim}, R_{sim} + 1$, $S = 1, \dots, S_{sim}, S_{sim} + 1$, and $P = 1, \dots, P_{sim}, P_{sim} + 1$, in which R_{sim} , S_{sim} , and P_{sim} are the values of R , S , and P that have been used to simulate the data. The optimization method uses five random starts for each combination of R , S , and P in each data set, and the EM-algorithm is considered to have converged if the difference in log-likelihood between two iterations is smaller than 10^{-4} . We then determine the preferred model in each data set for each of four information criteria. These criteria are: AIC, AIC-3, BIC, and CAIC (see Andrews & Currim, 2003, for an overview).

The first four rows of Table 3.8 show, for each information criterion, the proportion of the data sets in which the information criterion identified the correct model; these proportions are computed for each level of each factor, averaged over the six remaining factors. In each situation, either the AIC-3 or the BIC performs best. CAIC tends to be slightly worse than BIC, and AIC seems to perform rather poorly (it often overestimates the numbers of segments). The differences in the performances between AIC-3, BIC, and CAIC appear to be rather small compared to the clearly inferior performance of AIC. For large data sets, BIC and CAIC are known to give good performances (Andrews & Currim, 2003), which is supported by the results in Table 3.8. As AIC-3 was only proposed as a criterion for determining numbers of segments (Andrews & Currim, 2003), and not for choosing a dimensionality, we decide to use BIC for the List of Values data set.

Table 3.8: Results of Simulation Study

	Factor 1 respondents 1000 4000	Factor 2 rating cats. 5 9	Factor 3 backgr. vars. 0 2	Factor 4 segments 2 2 3	Factor 5 dimensions 1 2	Factor 6 items 8 16	Factor 7 param. var. 1.2 .6						
<i>Information criterion.</i>													
AIC	.23 ^a	.09	.23 ^a	.09	.19	.14	.17	.16	.17	.16	.13	.20	.13
AIC-3	.94	.91	.92	.92	.92	.92	.94	.91	.94	.91	.88	.97	.94
BIC	.89	.92	.88	.94	.92	.89	.98 ^b	.83	.94	.88	.86	.95	.86
CAIC	.88	.92	.88	.92	.91	.89	.98 ^b	.81	.92	.88	.84 ^a	.95	.84
<i>Parameter set</i>													
<i>Average absolute error</i>													
$a_{.11}, \dots, a_{.1R}$.078 ^b	.038	.056	.060	.057	.059	.044 ^b	.073	.048 ^b	.068	.078 ^b	.039	.065 ^a
F	.054 ^b	.022	.033	.043	.046	.031	.030	.046	.025 ^b	.052	.049 ^a	.028	.034
$\mathbf{G}_1, \dots, \mathbf{G}_k$.033	.013	.014	.032	— ^c	.023	.013	.034	.011	.035	.033	.014	.020
$\mathbf{H}_1, \dots, \mathbf{H}_S$.051 ^b	.024	.035	.041	.026 ^a	.050	.031	.045	.027 ^a	.049	.045	.030	.033
P	.016 ^b	.007	.011	.012	.011	.012	.012	.012	.011	.012	.015 ^b	.009	.010 ^a
<i>Type of segment</i>													
<i>Average proportion of respondents placed in correct segment</i>													
Response style segment	.891	.907	.885	.912	.913	.885	.940 ^b	.858	.922 ^b	.876	.857 ^b	.941	.918 ^b
Item segment	.913	.912	.912	.913	.946 ^b	.879	.944 ^b	.881	.874 ^b	.951	.873 ^b	.952	.947 ^b

^a The factor has a significant effect at the 5% level.

^b The factor has a significant effect at the 1% level.

^c The parameter set $\mathbf{G}_1, \dots, \mathbf{G}_k$ is not used for data sets without background variables.

Notes: For the information criteria, the statistical significance of the effects of factors in the simulation study is determined using a binary logit model. For the measures of parameter recovery and segment membership recovery, the statistical significance of factors has been tested using analysis of variance with only main effects.

The results of the simulation study show that the optimization algorithm, in combination with a properly chosen information criterion, is capable of identifying the correct values for R , S , and P . We also wish to evaluate the parameter recovery of the optimization algorithm given that R , S , and P have been identified correctly. For each simulated data set, we compare the simulated parameters to the estimated parameters for the optimizations in which R , S , and P equal their correct values. The second part of Table 3.8 shows the mean absolute error for each parameter set, for each level of each factor. The mean absolute error is always smaller than .08 for each parameter set. As these estimated parameters (except for the elements of \mathbf{P}) have a variance of .6 or 1.2, we believe that the parameter recovery is acceptable.

The last two rows of Table 3.8 show the average proportions with which the respondents are assigned to their simulated clusters, for both the response styles segments and the item segments. To assign each respondent to one segment, each respondent is placed in the segment in which his posterior membership probability is greatest. These proportions tend to be around 90% and are relatively mildly affected by the characteristics of the data set.

Finally, we have tested the effects of each factor in the simulation study on the recovery performance of the LC-BML model, and the results of these tests are also shown in Table 3.8. For the information criteria, these tests have been conducted using a binomial logit model, and for the other results in Table 3.8, an ANOVA model has been used; both types of models contained a constant term, main effects of all factors, and no interaction effects. The results show that the parameter recovery is better if the numbers of respondents and items are higher. Increasing the numbers of segments and dimensions may negatively affect the parameter recovery and the probability that the correct model is identified, though the effects of these factors are often not statistically significant. For the other factors in the simulation study, no consistent results are found.

3.E Confidence Ellipses

The confidence ellipses in the Results section are based on maximum likelihood theory. Under regularity conditions, a maximum likelihood parameter estimator is normally distributed with covariance matrix equal to the inverse of the information matrix. The information matrix can be estimated as the negative of the matrix of second-order partial derivatives of the likelihood function with respect to all model parameters, evaluated at the maximum likelihood parameter estimates. However, as the model parameters are not identified without the parameter constraints described in Appendix 3.A, this matrix of second-order partial derivatives is not invertible. To circumvent this problem, we created an alternative model parametrization that does not require any parameter constraints for parameter identification and is equivalent to our original model. Several estimated parameters in the original parametrization represent probabilities of almost exactly 0. These parameters were fixed at these values for the alternative parametrization, so that the uncertainty in the associated parameter estimates is not taken into account in the confidence ellipses.

The confidence ellipses are constructed using the following procedure. First, the final parameter estimates of our original parametrization are transformed to the alternative one. Using these transformed parameter estimates, the matrix of second-order partial derivatives is calculated numerically. Then, 10,000 simulated parameter vectors are drawn from a multivariate normal distribution with covariance matrix equal to the negative inverse of the matrix of second-order derivatives. For each simulated parameter vector, the associated parameter vector in the original parametrization is computed. Analysis of the simulated parameters in the original parametrization shows that the locations of the points in the graphical representation have approximately a joint bivariate normal distribution. Finally, the simulated locations of the points are used to construct 95% normal theory confidence ellipses.

Chapter 4

Modeling Attribute Interactions in Choice-Based Conjoint Analysis*

Abstract

In conjoint choice experiments, interaction effects of attributes are seldom fully modeled. Modeling all interaction effects tends to require a large number of parameters and a large conjoint choice design. In this chapter, we consider a novel approach for modeling two-way interactions effects, based on the OSI model of Chapter 2. The aim of this approach is to model the two-way interaction effects in choice-based conjoint analysis effectively, while keeping the design size and the number of parameters relatively small. We apply the resulting models to an empirical conjoint choice data set. We also sketch how good choice designs can be constructed for our models.

4.1 Introduction

Conjoint choice experiments are a useful way to determine how product characteristics affect the preferences of consumers (see, for example, Green & Wind, 1975). Researchers have often focussed on the main effects of the product characteristics, and usually only limited attention is paid to interaction effects. However, it is well-known that combining various product characteristics can yield synergies or suppressions (Gofman, 2006), so that modeling interaction effects in conjoint analysis can be important.

There is often a need to also include interaction effects in such a model specification (see, for example, Carmone & Green, 1981). In conjoint analysis, modeling interaction effects is almost always confined to two-way interaction effects (Vriens, 1995). Compared to other types of models, modeling interactions in conjoint analysis can be hard, because the design (that is, the characteristics of the choice sets offered to the respondents) limits

*This chapter is joint work with Alex J. Koning and Patrick J. F. Groenen.

what kinds of effects can be measured. If a full-factorial design (that is, a design that contains all combinations of the levels of the attributes) is used, all main effects and higher-way interactions can be measured. However, the size of such a design increases exponentially with the number of product attributes and their number of levels. Alternatively, it is possible to model only selected two-way interaction terms, using fractional factorial designs (see Holland & Craves, 1973). For example, Carmone and Green (1981) suggest using *compromise designs*, which are fractional factorial designs that facilitate the estimation of all main effects and a subset of the two-way interactions. Fractional factorial designs typically are much smaller than the associated full-factorial designs. However, this approach has the disadvantage that one must determine in advance which two-way interaction terms are relevant; all other interaction effects are subsequently neglected.

In this chapter, we consider several ways to model interaction effects of product characteristics in conjoint choice experiments. We limit ourselves to choice-based conjoint analysis experiments, which is one of the most popular forms of conjoint analysis. We propose an alternative specification of the interaction effects that is based on the OSI model of Chapter 2. The aim is to construct a parsimonious model specification for two-way interaction effects in conjoint analysis. This model specification allows to account for all two-way interaction terms of both categorical and continuous product characteristics. The fact that the models discussed here require relatively few parameters has the advantage that the required choice design (that is, the sets of products that are presented to the respondents) can also be smaller. To estimate the parameters of a conjoint choice model efficiently, a good conjoint choice design must be available. The construction of optimized designs has received much attention in the marketing literature (see, for example, Kuhfeld, Tobias, & Garratt, 1994; Huber & Zwerina, 1996; Sándor & Wedel, 2001; Kessels, Goos, & Vandebroek, 2006). Therefore, we also explore the construction of good choice designs for our model in this chapter. Finally, we use empirical data to illustrate the models.

The outline of this chapter is as follows. In the next section, we give an overview of choice-based conjoint analysis, including the necessary notation. In Section 4.3, we introduce our model for analyzing interaction effects in conjoint analysis and show how it is related to other methods. The model is applied to an empirical data set in Section 4.4. In Section 4.5, we discuss the generation of optimal conjoint choice designs. In the final section, we summarize our findings.

4.2 Overview of Choice-Based Conjoint Analysis

The aim of conjoint choice experiments is to measure the preferences of consumers. A typical choice experiment presents each respondent with several sets of products; these sets of product are also called choice sets. Each product in each choice set comprises a number of attributes. For each choice set, each respondent is asked to select the product that he or she likes most. The choices made by the respondents can be used to infer what kind of product they would prefer most. Although many attributes of products, such as

the price, are continuous in nature, the levels of the attributes are typically chosen from a finite set.

To describe choice-based conjoint analysis mathematically, we consider the situation in which S choice sets are presented to each of n respondents. We assume that these choice sets are the same for each respondent. For each choice set, each respondent is asked to select exactly one of J profiles. Each profile represents a product that has m characteristics, which are denoted by $\mathbf{x}_{1js}, \dots, \mathbf{x}_{mjs}$. The vector \mathbf{x}_{ijs} is defined as an indicator vector; element k of \mathbf{x}_{ijs} equals 1 if attribute i of concept j in choice set s has level k , and 0 otherwise.

We use a multinomial logit model to model the choice behavior of the respondents. However, our approach for modeling interaction effects may also be used in other types of conjoint choice models. The utility that a respondent derives from profile j is modeled as

$$u_{js} = f(\mathbf{x}_{js}, \theta) + \epsilon_{js}, \quad (4.1)$$

in which \mathbf{x}_{js} represents the attributes of profile j in choice set s , θ is a parameter vector containing all model parameters, and ϵ_{js} is an error term with an i.i.d. extreme value distribution. The respondents are assumed to choose the profile with the highest utility in each choice set s . As a result, the probability that a respondent chooses profile j in choice set s is given by

$$p_{js} = \frac{\exp(f(\mathbf{x}_{js}, \theta))}{\sum_{j'=1}^J \exp(f(\mathbf{x}_{j's}, \theta))}. \quad (4.2)$$

As the observations are assumed to be independent, the log-likelihood function of the conjoint choice experiments is given by

$$\log L(f_{js}, p_{js}) = n \sum_{s=1}^S \sum_{j=1}^J f_{js} \log p_{js}, \quad (4.3)$$

in which f_{js} denotes the proportion of times that the respondents choose product j in choice set s . For a given choice design \mathbf{x}_{ijs} and response frequencies f_{js} , the parameters θ are estimated by maximizing the log-likelihood function.

4.3 The Conjoint-OSI model

If the product characteristics do not interact, the effects of the categorical attributes on the estimated utility can be fully specified as

$$f(\mathbf{x}_{js}, \mathbf{a}_1, \dots, \mathbf{a}_m) = \sum_{i=1}^m \mathbf{x}'_{ijs} \mathbf{a}_i, \quad (4.4)$$

in which \mathbf{a}_i is a parameter vector of main effects of attribute i and must sum to zero. The conjoint choice model with only main effects is well-known (see, for example, Sándor &

Wedel, 2001) and often applied. It is straightforward to augment this model with two-way interaction effects. Modeling all two-way interaction effects fully results in the model specification

$$f(\mathbf{x}_{js}, \mathbf{a}_1, \dots, \mathbf{a}_m, \mathbf{C}_{kl}) = \sum_{i=1}^m \mathbf{x}'_{ijs} \mathbf{a}_i + \sum_{k=1}^{m-1} \sum_{l=k+1}^m \mathbf{x}'_{kjs} \mathbf{C}_{kl} \mathbf{x}_{ljs}, \quad (4.5)$$

in which \mathbf{C}_{kl} is a parameter matrix that contains the interaction effect between attributes k and l ; all rows and columns of each \mathbf{C}_{kl} must sum to zero. For many applications, such a full two-way interaction specification requires estimating too many parameters, so that (4.5) is not often applied in practice.

In this chapter, we consider a novel way for modeling interaction effects in conjoint analysis. To do so, we use the modeling approach for interaction effects that was proposed in Chapter 2. This modeling approach is based on the methodology of optimal scaling (Young, 1981; Gifi, 1990). Optimal scaling assigns numeric values to categorical variables by optimizing a criterion (in this case, the likelihood function). In Chapter 2, optimal scaling was used to model two-way interaction effects in generalized linear models. In the OSI model, a continuous, optimally scaled variable is constructed for each predictor variable, to model its two-way interaction effects. The optimally scaled variables are constructed by assigning a parameter to each level of each categorical predictor variable. For each interaction term, the product of the corresponding two optimally scaled variables is then treated as an ordinary predictor variable. As a result, only one parameter is required for modeling the strength of the interaction effect of each pair of predictor variables.

Here, we adapt the methodology of the OSI model, so that it can be applied to (choice-based) conjoint analysis data. We call the resulting model the Conjoint-OSI model. To describe this model formally, we first write a model with only main effects in terms of optimally scaled variables as

$$f(\mathbf{x}_{js}, \mathbf{a}_1, \dots, \mathbf{a}_m, \gamma_i) = \sum_{i=1}^m \gamma_i \mathbf{x}'_{ijs} \mathbf{a}_i, \quad (4.6)$$

in which the vector \mathbf{a}_i contains the category quantifications of attribute i , and the γ_i are coefficients of the optimally scaled variables. The category quantifications are restricted to have mean 0 and variance 1, which makes it necessary to include the parameters γ_i . We distinguish two versions of the Conjoint-OSI model. The first version is obtained by adding the two-way interaction effects of a new set of optimally scaled variables to (4.6), which yields

$$f(\mathbf{x}_{js}, \mathbf{a}_1, \dots, \mathbf{a}_m, \mathbf{b}_1, \dots, \mathbf{b}_m, \gamma_i, \sigma_{kl}) = \sum_{i=1}^m \gamma_i \mathbf{x}'_{ijs} \mathbf{a}_i + \sum_{k=1}^{m-1} \sum_{l=k+1}^m \sigma_{kl} \mathbf{x}'_{kjs} \mathbf{b}_k \mathbf{b}'_l \mathbf{x}_{ljs}, \quad (4.7)$$

in which σ_{kl} is a parameter that measures the strength of the interaction between attributes k and l , and \mathbf{b}_k is the parameter vector of quantifications for the interaction effects of

Table 4.1: Degrees of freedom, log-likelihood values, and AIC values of various conjoint choice models.

Model	Degrees of freedom	Log-likelihood	AIC
Main effects only	11	-20,310.1	40,642.2
Conjoint-OSI model (unequal quantifications)	21	-20,298.6	40,639.2
Conjoint-OSI model (equal quantifications)	14	-20,301.0	40,630.0
All two-way interactions	44	-20,288.0	40,664.0

attribute k . The term $\mathbf{x}'_{kjs} \mathbf{b}_k$ is the value of optimally scaled variable k for product j in choice set s ; see Chapter 2 for more information on this model specification.

As an alternative model, we can also use the category quantifications of the main effects for modeling the two-way interaction effects (that is, impose that $\mathbf{b}_k = \mathbf{a}_k$), which leads to a more parsimonious model. This approach yields

$$f(\mathbf{x}_{.js}, \mathbf{a}_1, \dots, \mathbf{a}_m, \gamma, \sigma..) = \sum_{i=1}^m \gamma_i \mathbf{x}'_{ijs} \mathbf{a}_i + \sum_{k=1}^{m-1} \sum_{l=k+1}^m \sigma_{kl} \mathbf{x}'_{kjs} \mathbf{a}_k \mathbf{a}'_{ljs}. \quad (4.8)$$

In the remainder of this chapter, we consider both versions of the Conjoint-OSI model.

The Conjoint-OSI model partially models the two-way interaction effects. To model the interaction effects, this model constructs a continuous, optimally scaled variable for each attribute, by assigning a parameter to each level of each attribute. As in Chapter 2, the estimated quantifications (that is, \mathbf{b}_k and \mathbf{a}_k) are constrained in such a way that the optimally scaled variables have mean 0 and variance 1. For each pair of attributes, the Conjoint-OSI model uses a single parameter to model the size of the interaction effect. The interaction effect between two levels of different attributes is estimated as the product of the quantifications for the interaction effects of these two levels and the corresponding size parameter.

4.4 Empirical Application

We use a data set based on a commercial study¹ to test the Conjoint-OSI model and compare it with other approaches for modeling interaction effects. For this data set, 805 respondents were each shown 24 pairs of products. For each pair of products, each respondent was asked to choose his or her preferred product or to make no choice at all. The products had three attributes: ‘Brand’ (5 levels), ‘Benefit’ (4 levels), and ‘Format’ (4 levels). As the data are confidential, we do not report the meanings of the levels of these attributes here.

¹We would like to thank Graham Cleaver for providing this data set.

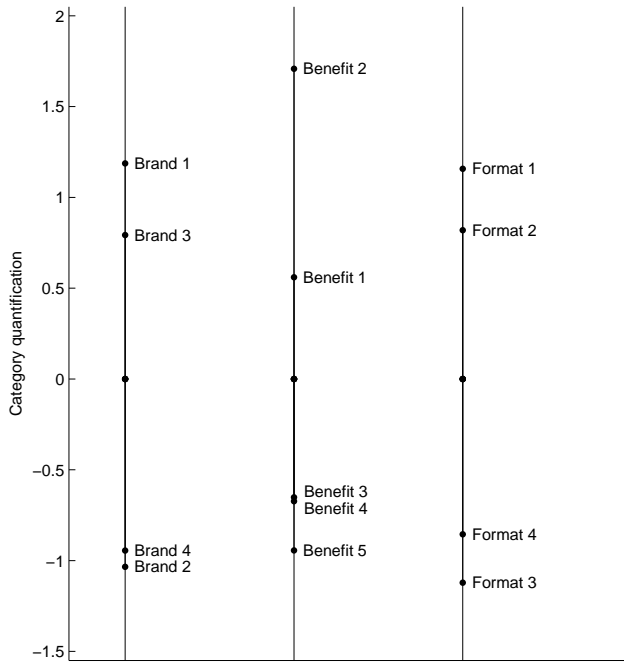
We estimate four conjoint choice models for this data set. These models are all based on a multinomial logit model and include a model with only main effects (Equation 4.4), a model that contains all two-way interaction terms (Equation 4.5), and the two variants of the Conjoint-OSI model (Equations 4.7 and 4.8). The parameters of the models were estimated by maximizing the log-likelihood function using a general-purpose quasi-Newton optimization routine, with appropriate constraints on the parameters; see also Chapter 2 of this thesis. Table 4.1 shows general results for these four models.

The Akaike information criterion (AIC) prefers both versions of the Conjoint-OSI model over the other two models, so that we are justified in considering the Conjoint-OSI model. Below, we interpret the results of the Conjoint-OSI model with equal quantifications, which resulted in the lowest value of the AIC.

The upper side of Figure 4.1 shows the estimated quantifications (that is, the parameters \mathbf{a}_i) of the attributes, which are used for both the main effects and the interaction effects. The first row of the lower part of Figure 4.1 contains the estimated sizes of the main effects (that is, the parameters γ_i). The main effects of Brand and Format appear to be stronger than the main effects of Benefit. To obtain the estimated main effects for each level of each attribute, we must multiply the sizes of the main effects with the corresponding quantifications. For example, Brand 2 has a positive main effect of $-0.144 \times -1.03 = 0.149$. Brands 2 and 4 are therefore preferred to Brands 1 and 3. A similar preference exists for Formats 1 and 2. However, the effects of the attributes on the respondents' choices seem relatively small, as the estimated parameters of the main effects are not large.

We also interpret the estimated interaction effects. The lower part of Figure 4.1 also displays the estimated sizes of the interaction effects; the interaction effects in this figure can be determined in the same way as in Figure 2.1. To obtain an estimated interaction effect between two levels of different attributes from Figure 4.1, the coefficient of the interaction in the lower part of Figure 4.1 must be multiplied with the corresponding interaction quantifications in the upper part of this figure. For example, the estimated interaction effect of Brand 1 and Benefit 1 is the coefficient 0.0433 times the product of the interaction quantifications 1.19 and 0.57, which equals 0.029. Of the three interaction effects, the interaction between Brand and Benefit appears to be the strongest; the attributes Benefit and Format do not appear to interact strongly.

Based on the estimated parameters of the Conjoint-OSI model, the most preferred combination of the levels is Brand 2, Benefit 2, and Format 1; the least preferred combination is Brand 1, Benefit 5, and Format 3. If these two products, which have the largest difference in estimated utility among all possible products, are compared using a no-choice option, the estimated probability that the respondent chooses the product with Brand 2, Benefit 2, and Format 1 equals 52.3%, and the estimated probability of choosing Brand 1, Benefit 5, and Format 3 equals 26.7%. With a probability of 21.0% percent, the respondent chooses neither of the two products. These results show that differences in utility among products exist, but they are relatively small.



	Brand	Benefit	Format
<i>Main effects (γ_i)</i>	-0.144	0.039	0.117
<i>Interaction effects (σ_{kl})</i>			
Brand	-	0.0433	0.0148
Benefit	-	-	0.0079
Format	-	-	-

Figure 4.1: Parameter estimates for the Conjoint-OSI model with a single quantification for each level of each attribute.

4.5 Optimal Designs

Before conducting a conjoint choice experiment, a design must first be chosen. Several criteria exist to evaluate the effectiveness of a design; see Kessels et al. (2006) for an overview of the performance of these criteria. The most commonly used criterion is the D-criterion (see, for example, Kuhfeld et al., 1994; Huber & Zwerina, 1996), and we discuss this criterion here. The D-criterion minimizes the D_p -error, which is defined as

$$\det[\mathbf{I}(\theta|\mathbf{X})]^{-1/k}, \quad (4.9)$$

in which $\mathbf{I}(\theta|\mathbf{X})$ is the information matrix and k denotes the number of free model parameters. Minimizing (4.9) corresponds to minimizing the volume of the confidence ellipsoid for θ . For model (4.2), the information matrix is given by

$$n \sum_{s=1}^S \frac{\partial f(\mathbf{x}_{js}, \theta)}{\partial \theta} (\mathbf{P}_s - \mathbf{p}_s \mathbf{p}_s') \frac{\partial f(\mathbf{x}_{js}, \theta)}{\partial \theta}, \quad (4.10)$$

in which the matrix \mathbf{X} contains the vectors \mathbf{x}_{js} , \mathbf{p}_s has elements p_{js} , and \mathbf{P}_s is defined as $\text{diag}(\mathbf{p}_s)$.

The D-criterion depends on the unknown values of the model parameters θ . In some cases, the D-criterion is optimized by assuming that all parameters are zero. However, for the Conjoint-OSI model, this assumption does not make sense, as the model parameters are not identified under this assumption.

An optimal design for the Conjoint-OSI model can be constructed as follows

1. Conduct a pilot study that does not use an optimized design to obtain initial parameter estimates.
2. Generate a random initial design and calculate the value of the D-criterion for this design.
3. Do the following, until no further improvements can be made to the design.
 - (a) Randomly choose a product in the design and replace it with another, randomly chosen product.
 - (b) Determine whether replacing the product improves the value of the D-criterion.
 - (c) If no improvement in the design has been obtained, revert to the previous design.

To obtain the globally optimal design, Steps 2 and 3 of the algorithm described above should be performed many times, and the best design found should be chosen. However, finding the globally optimal design is usually not necessary, and a good design will suffice. The design obtained using this algorithm should allow researchers to investigate the presence of interaction effects much more efficiently than an unoptimized or a random design.

Table 4.2: An example of an optimized design for the Conjoint-OSI model with equal quantifications.

Choice set	Profile 1			Profile 2		
	Brand	Benefit	Format	Brand	Benefit	Format
1	4	2	3	2	5	1
2	1	3	4	3	5	1
3	4	5	1	1	2	2
4	2	2	1	1	5	2
5	4	5	2	2	5	3
6	4	5	3	1	5	1
7	2	4	3	4	2	1
8	1	2	3	3	4	4
9	3	4	3	1	3	2
10	1	1	1	2	2	2
11	1	2	3	4	4	4
12	1	4	1	3	1	2
13	1	4	2	3	3	4
14	1	5	4	3	3	1
15	2	2	1	1	2	4
16	1	1	4	4	4	3
17	1	2	1	2	2	3
18	3	3	2	1	1	1
19	2	2	4	1	4	3
20	1	1	2	1	2	4
21	2	5	2	1	2	1
22	1	2	2	1	3	1
23	3	5	3	1	4	1
24	3	4	2	1	5	3

We have constructed an optimized design for the Conjoint-OSI model with equal quantifications, using the data set in this chapter as a pilot study. This design is shown in Table 4.2. It is also possible to construct optimized designs for the full two-way interaction model; however, such a design must be larger than the design shown in Table 4.2. We believe that the fact the Conjoint-OSI model requires a smaller design than a full two-way interaction model is an important advantage.

4.6 Conclusions

In this chapter, we have applied the optimal scaling methodology of Chapter 2 to modeling interaction effects in conjoint analysis. We find that imposing a bilinear parameter structure can also be useful in sophisticated techniques such as choice-based conjoint analysis. This approach has resulted in the new Conjoint-OSI model. We distinguish two versions of our model: a version with separate quantifications for the interaction effects and the main effects and a version with one quantification for both main effects and interaction effects. We believe that both versions of the model can be useful; which version is the best one likely depends on the data set used.

As interaction effects in conjoint analysis are usually relatively small, there is typically no need to consider multidimensional models or full two-way interaction models. Three-way interactions appear to be almost always negligible in conjoint analysis.

We have applied our models to a commercial data set. In this data set, the Conjoint-OSI model with equal quantifications for main and interaction effects is preferred over all other models by Akaike information criterion. Using only three parameters, this model can fit a large portion of the two-way attribute interactions; a full two-way interaction model would use 33 effective parameters to account for these interaction effects.

Finally, we have sketched how optimal designs can be generated for our model. Such optimal designs are easiest to construct if prior parameter estimates are available. We find that optimized designs for the Conjoint-OSI model can be smaller than designs for the full two-way interaction model. It remains a subject for further research how good conjoint choice designs can be generated if no prior information is available. Future research may also investigate how attribute interactions in other types of conjoint analysis experiments can be modeled effectively.

Part II

Two-Mode Partitioning

Chapter 5

Optimization Strategies for Two-Mode Partitioning*

Abstract

Two-mode partitioning is a relatively new form of clustering that clusters both rows and columns of a data matrix. In this chapter, we consider deterministic two-mode partitioning methods in which a criterion similar to k -means is optimized. A variety of optimization methods have been proposed for this type of problem. However, it is still unclear which method should be used, as various methods may lead to non-global optima. This chapter reviews and compares several optimization methods for two-mode partitioning. Several known methods are discussed, and a new fuzzy steps method is introduced. The fuzzy steps method is based on the fuzzy c -means algorithm of Bezdek (1981) and the fuzzy steps approach of Heiser and Groenen (1997) and Groenen and Jajuga (2001). The performances of all methods are compared in a large simulation study. In our simulations, a two-mode k -means optimization method most often gives the best results. Finally, an empirical data set is used to give a practical example of two-mode partitioning.

5.1 Introduction

Clustering can be seen as one of the cornerstones of classification. Consider a typical two-way two-mode data set of respondents by variables. Often, clustering algorithms are applied to just one mode of the data matrix, which can be done in a hierarchical or non-hierarchical way. Among the non-hierarchical methods, k -means clustering (Hartigan, 1975) is one of the most popular methods and has the advantage of a loss function being optimized. For an overview of clustering methodology for one-mode data, see Mirkin (2005).

*This chapter is based on Van Rosmalen, Groenen, et al. (in press).

A relatively new form of clustering is two-mode clustering. In two-mode clustering, both rows and columns of a two-mode data matrix are assigned to clusters. Each row of a two-mode data matrix is assigned to one or more row clusters, and each column to one or more column clusters. Elements of the data matrix that are in the same row cluster and in the same column cluster should be close. DeSarbo (1982) described such a two-mode clustering method, called the GENNCLUS model. A technique that is related to two-mode clustering is blockmodeling, which is often used in social network analysis (see, for example, Noma & Smith, 1985; Doreian, Batagelj, & Ferligoj, 2005) and can be used for both one-mode and two-mode data matrices (Doreian et al., 2004). Blockmodeling attempts to partition network actors into clusters, so that a block structure can be identified in a data matrix in which the rows and columns have been reordered in a specific way; the elements within each block should typically have similar values. An extensive overview of two-mode clustering methods can be found in Van Mechelen, Bock, and De Boeck (2004).

In this chapter, we focus on two-mode partitioning, that is, partitioning the sets of rows and columns so that each row and each column is assigned to exactly one cluster. To do so, we use a least-squares criterion that models the elements of the data matrix belonging to the same row and column cluster by their average. Other optimization criteria and methods that do not optimize a criterion can also be used to partition the rows and columns of a data matrix. However, in the remainder of this chapter, we focus on deterministic two-mode partitioning using a least-squares criterion and refer to this method simply as two-mode partitioning. Several optimization methods for finding good two-mode partitions based on this criterion are known from the literature. However, these methods are not guaranteed to find the global optimum and often get stuck in local minima. Therefore, we study the local minimum problem of two-mode partitioning for several optimization methods and determine which method tends to give the best local minimum among these methods. In addition, a new optimization method for two-mode partitioning is introduced, based on the fuzzy c -means algorithm of Bezdek (1981). Using a simulation study, we identify the methods that perform well under most circumstances, within a reasonable computational effort.

The remainder of this chapter is organized as follows. In the next section, we introduce the notation and give an overview of the optimization problem, including two hill climbing algorithms. Section 5.3 describes the implementation of two *meta-heuristics* for two-mode partitioning. Section 5.4 introduces the fuzzy steps two-mode partitioning method. In Section 5.5, we compare the performances of the methods using a simulation study. Section 5.6 uses an empirical data set to compare the methods and to give a practical example of two-mode partitioning. Finally, we draw conclusions and give recommendations for further research.

5.2 Overview of Optimization Problem

To define the two-mode partitioning problem, consider the following notation:

$\mathbf{X}_{n \times m} = (x_{ij})_{n \times m}$	Two-mode data matrix of n rows and m columns.
$\mathbf{P}_{n \times K} = (p_{ik})_{n \times K}$	Cluster membership matrix of the rows with K the number of row clusters, $p_{ik} = 1$ if row i belongs to row cluster k , and $p_{ik} = 0$ otherwise.
$\mathbf{Q}_{m \times L} = (q_{jl})_{m \times L}$	Cluster membership matrix of the columns with L the number of column clusters, $q_{jl} = 1$ if column j belongs to column cluster l , and $q_{jl} = 0$ otherwise.
$\mathbf{V}_{K \times L} = (v_{kl})_{K \times L}$	Matrix with cluster centers for row cluster k and column cluster l .
$\mathbf{E}_{n \times m} = (e_{ij})_{n \times m}$	Matrix with errors from cluster centers.

Usually, the rows of \mathbf{X} correspond to objects, and the columns of \mathbf{X} refer to variables. The elements of \mathbf{X} can be associations, confusions, fluctuations, etc., between row and column objects. Applying two-mode partitioning only makes sense if the data matrix is *matrix-conditional*, that is, its values can be compared among each other. Therefore, if one of the modes refers to variables, these variables must be comparable, standardized, or measured on the same scale. In the remainder of this chapter, we assume that the data satisfy this condition.

Two-mode partitioning assigns each element of \mathbf{X} to a row cluster and a column cluster. If L equals m , each column can be placed in a cluster by itself, so that two-mode partitioning reduces to one-mode k -means clustering, and the same is true if K equals n . The matrix \mathbf{V} can be interpreted as the combined cluster means. The cluster memberships are given by the matrices \mathbf{P} and \mathbf{Q} . Together, the three matrices \mathbf{P} , \mathbf{Q} , and \mathbf{V} approximate the information in \mathbf{X} by \mathbf{PVQ}' . To make this approximation as close to \mathbf{X} as possible, we use the additive model

$$\mathbf{X} = \mathbf{PVQ}' + \mathbf{E}, \quad (5.1)$$

where \mathbf{E} is the error of the model. Equation (5.1) can be seen as a special case of the additive box model proposed by Mirkin, Arabie, and Hubert (1995); the additive box model does not partition the rows and columns, but allows for overlapping clusters.

Two-mode partitioning searches for the optimal partition \mathbf{P} , \mathbf{Q} and cluster centers \mathbf{V} that minimize the sums of squares of \mathbf{E} . This objective amounts to minimizing the squared Euclidean distance of the data points to their respective clusters centers in \mathbf{V} . Therefore, the criterion to be minimized can be expressed as

$$f(\mathbf{P}, \mathbf{Q}, \mathbf{V}) = \|\mathbf{X} - \mathbf{PVQ}'\|^2 = \sum_{k=1}^K \sum_{l=1}^L \sum_{i=1}^n \sum_{j=1}^m p_{ik} q_{jl} (x_{ij} - v_{kl})^2. \quad (5.2)$$

Using the Euclidean metric is not mandatory; other metrics have been used as well, especially in one-mode clustering (see, for example, Bock, 1974). However, in this study, we restrict ourselves to the Euclidean metric.

The optimal cluster membership matrices must satisfy the following constraints.

1. The cluster memberships of each row and column object must sum to one, so that $\sum_{k=1}^K p_{ik} = 1$ and $\sum_{l=1}^L q_{jl} = 1$.
2. All cluster membership values must be either zero or one, so that $p_{ik} \in \{0, 1\}$ and $q_{jl} \in \{0, 1\}$.
3. None of the row or column clusters is empty, that is $\sum_{i=1}^n p_{ik} > 0$ and $\sum_{j=1}^m q_{jl} > 0$.

The first two constraints together require that each row of \mathbf{P} and \mathbf{Q} contains exactly one element with the value 1. Hence, each row and column object is assigned to exactly one cluster. These two constraints are necessary and sufficient for the second equality in (5.2) to hold. A partition that is optimal according to (5.2) typically does not have empty clusters, and, in principle, the third constraint is not required during the estimation. However, some algorithms may lead to a partition with empty clusters. Therefore, we adapt these algorithms to correct for potential empty clusters or prevent them from happening.

No known polynomial time algorithm is guaranteed to find the global minimum of $f(\mathbf{P}, \mathbf{Q}, \mathbf{V})$ in every instance. Even for relatively small n and m (say, $n = 20$ and $m = 20$), the number of possible partitions can become extremely large, and a complete enumeration of the possible solutions is almost always computationally infeasible. However, if two of the three matrices \mathbf{P} , \mathbf{Q} , and \mathbf{V} are known, the optimal value of the third matrix can be computed easily. If both \mathbf{P} and \mathbf{Q} are known, the optimal cluster centers \mathbf{V} can be computed as

$$v_{kl} = \frac{\sum_{i=1}^n \sum_{j=1}^m p_{ik} q_{jl} x_{ij}}{\sum_{i=1}^n \sum_{j=1}^m p_{ik} q_{jl}}, \quad (5.3)$$

which is the average of the elements of \mathbf{X} belonging to row cluster k and column cluster l . If \mathbf{V} and either \mathbf{P} or \mathbf{Q} are known, the problem of minimizing $f(\mathbf{P}, \mathbf{Q}, \mathbf{V})$ becomes a linear program that has a closed-form solution. When \mathbf{V} and \mathbf{Q} are known, the optimal value of \mathbf{P} can be computed as follows. Let $c_{ik} = \sum_{j=1}^m \sum_{l=1}^L q_{jl} (x_{ij} - v_{kl})^2$. Then,

$$p_{ik} = \begin{cases} 1 & \text{if } c_{ik} = \min_{1 \leq r \leq K} c_{ir}, \\ 0 & \text{otherwise.} \end{cases} \quad (5.4)$$

When \mathbf{P} and \mathbf{V} are known, the optimal matrix \mathbf{Q} can be computed in a similar fashion.

Several methods for finding an optimal two-mode partition, based on the minimization of (5.2), have been proposed in the literature. Two of these methods are discussed in the remainder of this section. Both methods are hill climbing algorithms and are guaranteed to find a local optimum (which may or may not be the global optimum). Three additional optimization methods for two-mode partitioning are discussed in the next two sections.

5.2.1 Alternating Exchanges Algorithm

The alternating exchanges algorithm was proposed by Gaul and Schader (1996). This algorithm tries to improve an initial partition by making a transfer of either a row or a column object and immediately recalculating \mathbf{V} . Our implementation of the alternating exchanges algorithm performs the following steps:

1. Choose initial \mathbf{P} and \mathbf{Q} , and calculate \mathbf{V} according to (5.3).
2. Repeat the following until there is no improvement of $f(\mathbf{P}, \mathbf{Q}, \mathbf{V})$ in either step.
 - (a) For each i and k , transfer row object i to row class k and re-calculate \mathbf{V} according to (5.3). Accept the transfer if it has improved $f(\mathbf{P}, \mathbf{Q}, \mathbf{V})$, otherwise return to the old \mathbf{P} and \mathbf{V} .
 - (b) For each j and l , transfer column object j to column class l and re-calculate \mathbf{V} according to (5.3). Accept the transfer if it has improved $f(\mathbf{P}, \mathbf{Q}, \mathbf{V})$, otherwise return to the old \mathbf{Q} and \mathbf{V} .

The alternating exchanges algorithm always converges to a local minimum, as the value of $f(\mathbf{P}, \mathbf{Q}, \mathbf{V})$ decreases in every iteration, the algorithm is defined on a finite set, and $f(\mathbf{P}, \mathbf{Q}, \mathbf{V})$ is bounded from below by 0.

5.2.2 Two-Mode k -Means Algorithm

The k -means algorithm (Hartigan, 1975) is one of the simplest and fastest ways to obtain a good partition, which accounts for its popularity in one-mode clustering. This algorithm can easily be extended to handle two-mode partitioning (see Baier, Gaul, & Schader, 1997; Vichi, 2001). The so-called two-mode k -means algorithm aims to improve an initial partition using (5.3) and (5.4) and comprises the following steps.

1. Choose initial \mathbf{P} and \mathbf{Q} .
2. Repeat the following, until there is no improvement of $f(\mathbf{P}, \mathbf{Q}, \mathbf{V})$ in any step.
 - (a) Update \mathbf{V} according to (5.3).
 - (b) Let $c_{ik} = \sum_{j=1}^m \sum_{l=1}^L q_{jl}(x_{ij} - v_{kl})^2$. Then update \mathbf{P} according to

$$p_{ik} = \begin{cases} 1 & \text{if } c_{ik} = \min_{1 \leq r \leq K} c_{ir}, \\ 0 & \text{otherwise.} \end{cases} \quad (5.5)$$

- (c) Update \mathbf{V} according to (5.3).
- (d) Let $d_{jl} = \sum_{i=1}^n \sum_{k=1}^K p_{ik}(x_{ij} - v_{kl})^2$. Then update \mathbf{Q} according to

$$q_{jl} = \begin{cases} 1 & \text{if } d_{jl} = \min_{1 \leq r \leq L} d_{jr}, \\ 0 & \text{otherwise.} \end{cases} \quad (5.6)$$

The two-mode k -means algorithm always converges to a local minimum, as the value of the criterion $f(\mathbf{P}, \mathbf{Q}, \mathbf{V})$ cannot increase in any step. However, one or more clusters may become empty after Step 2b or 2d. This situation is immediately corrected by transferring the row or column object with the highest value of $\sum_{k=1}^K p_{ik}c_{ik}$ or $\sum_{l=1}^L q_{jl}d_{jl}$ to the empty cluster.

5.3 Meta-Heuristics

The algorithms discussed in the previous section contain no provisions to avoid finding non-global optima. As no known polynomial time algorithm is capable of finding the global optimum in every instance, we also use optimization methods that are based on *meta-heuristics*. These meta-heuristics aim to increase the likelihood of finding the global optimum, though this cannot be guaranteed. Here, we discuss two optimization methods for two-mode partitioning that are applications of the meta-heuristics simulated annealing and tabu search. For the other three optimization methods in this chapter (the two algorithms discussed in the previous section and the fuzzy steps method that is proposed in the next section), the likelihood of finding a non-global optimum is reduced by performing multiple random starts. That is, these algorithms are performed several times with starting values that are chosen randomly every time an algorithm is run; the best result of all runs is retained as the final solution.

Other optimization methods have also been used for two-mode partitioning. Hansohm (2001) used a genetic algorithm, but found that its performance in two-mode partitioning is not as good as it is in one-mode k -means clustering. Gaul and Schader (1996) implemented a penalty algorithm, but found that it does not compare favorably with the alternating exchanges algorithm.

5.3.1 Simulated Annealing

Simulated annealing (see, for example, Van Laarhoven & Aarts, 1987) is a meta-heuristic that simulates the slow cooling of a physical system. Trejos and Castillo (2000) first used simulated annealing for two-mode partitioning. Their implementation of simulated annealing performs a local search that is based on the central idea of the alternating exchanges algorithm. To avoid getting stuck in local minima, transitions that increase $f(\mathbf{P}, \mathbf{Q}, \mathbf{V})$ are also accepted with a positive probability.

Here, we use the implementation of Trejos and Castillo (2000), except for the values of certain parameters. These parameters are the cooling rate $\gamma < 1$, the number of iterations R in which the temperature remains constant, the initial value of the temperature T , and the maximum number of iterations without accepted transitions, which is denoted by t_{max} . This implementation comprises the following steps.

1. Choose initial \mathbf{P} and \mathbf{Q} and calculate \mathbf{V} according to (5.3).
2. Choose the parameters R , γ , and an initial value of T .
3. Repeat the following until there is no change in \mathbf{P} and \mathbf{Q} for the last t_{max} values of T .
 - (a) Do the following R times:
 - i. Choose one of the two modes with equal probability.

- ii. Choose one of the objects of this mode with uniform probability and transfer it to another randomly chosen cluster.
 - iii. Update \mathbf{V} according to (5.3) and calculate Δf as the change in $f(\mathbf{P}, \mathbf{Q}, \mathbf{V})$ achieved by the transfer and the subsequent updating of \mathbf{V} .
 - iv. Always accept the transfer if $\Delta f < 0$, otherwise accept it with probability $\exp(-\Delta f/T)$.
- (b) Set $T = \gamma T$.

The partition with the lowest value of $f(\mathbf{P}, \mathbf{Q}, \mathbf{V})$ found during estimation is retained as the final solution.

5.3.2 Tabu Search

The tabu search meta-heuristic (see, for example, Glover, 1986) also performs a local search, but tries to avoid local optima by maintaining a tabu list. The tabu list is a list of solutions that are temporarily not accepted. We use the following implementation of tabu search for two-mode partitioning, which is based on the alternating exchanges algorithm. See Castillo and Trejos (2002) for a more detailed description of this implementation.

Define $Z(\mathbf{P}, \mathbf{Q}) = \min_{\mathbf{V}} f(\mathbf{P}, \mathbf{Q}, \mathbf{V})$. The algorithm performs the following steps.

1. Start with an initial partition (\mathbf{P}, \mathbf{Q}) and an empty tabu list. Set $(\mathbf{P}, \mathbf{Q})_{opt} = (\mathbf{P}, \mathbf{Q})$.
2. Choose the number of iterations S and a maximum length of the tabu list.
3. Perform the following steps S times.
 - (a) Generate a neighborhood of partitions N , consisting of the partitions that can be constructed by transferring one row or column object from (\mathbf{P}, \mathbf{Q}) to another cluster.
 - (b) Choose the partition $(\mathbf{P}, \mathbf{Q})_{cand}$ as the partition in N with the lowest value of $Z(\mathbf{P}, \mathbf{Q})$ that is not on the tabu list.
 - (c) Set $(\mathbf{P}, \mathbf{Q}) = (\mathbf{P}, \mathbf{Q})_{cand}$. If $Z((\mathbf{P}, \mathbf{Q})_{cand}) < Z((\mathbf{P}, \mathbf{Q})_{opt})$, then $(\mathbf{P}, \mathbf{Q})_{opt} = (\mathbf{P}, \mathbf{Q})_{cand}$.
 - (d) Add (\mathbf{P}, \mathbf{Q}) to the tabu list. Remove the oldest item from the tabu list, if the list exceeds its maximum length.

The final solution of the algorithm is given by $(\mathbf{P}, \mathbf{Q})_{opt}$.

5.4 Fuzzy Two-Mode Partitioning

Fuzzy methods relax the requirement that an object belongs to a single cluster, so that the cluster membership can be distributed over the clusters. For one-mode clustering, the best

known method is fuzzy c -means (Bezdek, 1981); for adaptations of this method see Tsao, Bezdek, and Pal (1994) and Groenen and Jajuga (2001). These methods try to make the optimization task easier by allowing for cluster membership values between 0 and 1. In this section, we extend one-mode fuzzy optimization methods to the two-mode case. First, we introduce a fuzzy two-mode partitioning criterion. We then discuss an algorithm for finding an optimal fuzzy partition, based on Bezdek (1981). Finally, we describe the fuzzy steps method, which reduces the fuzziness of the solution in steps until a crisp partition (that is, a partition in which all cluster membership values are either 0 or 1) is found.

Simply relaxing the constraint that the cluster membership values in (5.2) must be 0 or 1 by allowing for values between 0 and 1 does not guarantee an optimal partition that is fuzzy. A crisp partition will still minimize $f(\mathbf{P}, \mathbf{Q}, \mathbf{V})$ in that case, though a fuzzy partition might be equally good. Fuzzy optimal partitions can be obtained if the criterion $f(\mathbf{P}, \mathbf{Q}, \mathbf{V})$ is altered by raising the cluster membership values to a power s , with $s \geq 1$. Thus, the fuzzy two-mode partitioning criterion is defined as

$$f_s(\mathbf{P}, \mathbf{Q}, \mathbf{V}) = \sum_{k=1}^K \sum_{l=1}^L \sum_{i=1}^n \sum_{j=1}^m p_{ik}^s q_{jl}^s (x_{ij} - v_{kl})^2, \quad (5.7)$$

subject to the constraints

$$\sum_{k=1}^K p_{ik} = 1, \sum_{l=1}^L q_{jl} = 1, p_{ik} \geq 0, q_{jl} \geq 0, \sum_{i=1}^n p_{ik} > 0, \text{ and } \sum_{j=1}^m q_{jl} > 0. \quad (5.8)$$

This criterion yields fuzzy optimal partitions and the fuzziness parameter s determines how fuzzy the optimal partition is. For $s = 1$, the fuzzy criterion coincides with the crisp criterion.

5.4.1 The Two-Mode Fuzzy c -Means Algorithm

The algorithm for the optimization of $f_s(\mathbf{P}, \mathbf{Q}, \mathbf{V})$ is based on iteratively updating each set of parameters while keeping the other two sets fixed. Given optimal \mathbf{Q} and \mathbf{V} , the optimal \mathbf{P} can be found using the Lagrange method for each row i of \mathbf{P} . The Lagrangian is given by

$$L_i(\mathbf{P}, \mathbf{Q}, \mathbf{V}, \lambda) = \sum_{k=1}^K p_{ik}^s \sum_{l=1}^L \sum_{j=1}^m q_{jl}^s (x_{ij} - v_{kl})^2 - \lambda \left(\sum_{k=1}^K p_{ik} - 1 \right). \quad (5.9)$$

Defining $c_{ik} = \sum_{l=1}^L \sum_{j=1}^m q_{jl}^s (x_{ij} - v_{kl})^2$ and taking partial derivatives of L_i gives

$$\frac{\partial L_i}{\partial p_{ik}} = s p_{ik}^{s-1} c_{ik} - \lambda \quad \text{and} \quad \frac{\partial L_i}{\partial \lambda} = \sum_{k=1}^K p_{ik} - 1. \quad (5.10)$$

Setting these derivatives to zero and solving for p_{ik} yields

$$p_{ik} = \frac{c_{ik}^{1/(1-s)}}{\sum_{k=1}^K c_{ik}^{1/(1-s)}}. \quad (5.11)$$

However, (5.11) does not apply if any c_{ik} in row i is zero. In that case, any partition with $p_{ik} = 0$ whenever $c_{ik} > 0$ and $\sum_{k=1}^K p_{ik} = 1$ is optimal. Finding the optimal \mathbf{Q} given \mathbf{P} and \mathbf{V} can be done in a similar fashion. When s is large enough, the optimal values of the cluster memberships become $p_{ik} \approx 1/K$ and $q_{jl} \approx 1/L$, which can easily be derived from (5.11). In practice, the cluster membership values approach these values quite rapidly for reasonably large s . For $s = 3$, the cluster membership values often differ only slightly and for $s > 10$, they are usually equal to each other within the numerical accuracy of current computers. As s approaches 1 from above, $1/(1-s)$ approaches minus infinity, and the fuzzy optimization formula (5.11) becomes its crisp counterpart (5.4). In that case, the optimal partition based on (5.7) also approaches the optimal crisp partition. Therefore, higher values of s correspond to fuzzier optimal partitions, and s close to 1 to crisp partitions. The optimal \mathbf{V} can be obtained by setting the partial derivatives of (5.7) with respect to v_{kl} to zero and solving for v_{kl} , which yields

$$v_{kl} = \frac{\sum_{i=1}^n \sum_{j=1}^m p_{ik}^s q_{jl}^s x_{ij}}{\sum_{i=1}^n \sum_{j=1}^m p_{ik}^s q_{jl}^s}. \quad (5.12)$$

For a given value of s , the *two-mode fuzzy c-means algorithm* comprises the following steps.

1. Choose initial \mathbf{P} and \mathbf{Q} , which can be either crisp or fuzzy, and calculate \mathbf{V} according to (5.12).
2. Repeat the following, until the decrease in $f_s(\mathbf{P}, \mathbf{Q}, \mathbf{V})$ is small.
 - (a) Calculate $c_{ik} = \sum_{l=1}^L \sum_{j=1}^m q_{jl}^s (x_{ij} - v_{kl})^2$ and update \mathbf{P} according to $p_{ik} = \frac{c_{ik}^{1/(1-s)}}{\sum_{k=1}^K c_{ik}^{1/(1-s)}}$.
 - (b) Calculate $d_{jl} = \sum_{k=1}^K \sum_{i=1}^n p_{ik}^s (x_{ij} - v_{kl})^2$ and update \mathbf{Q} according to $q_{jl} = \frac{d_{jl}^{1/(1-s)}}{\sum_{k=1}^K d_{jl}^{1/(1-s)}}$.
 - (c) Update \mathbf{V} according to (5.12).

This algorithm lowers the value of $f_s(\mathbf{P}, \mathbf{Q}, \mathbf{V})$ in each iteration, until convergence has been achieved. Therefore, the algorithm always converges to a saddle point or a local minimum, which may or may not be a global minimum.

5.4.2 Fuzzy Steps

The two-mode fuzzy c -means algorithm generally converges to a fuzzy partition. To ensure that our fuzzy optimization method converges to a crisp partition, we use the idea of fuzzy steps, which was proposed by Heiser and Groenen (1997). Our fuzzy steps method for two-mode partitioning starts with an initial value of s that is greater than 1. This method uses the two-mode fuzzy c -means algorithm to minimize $f_s(\mathbf{P}, \mathbf{Q}, \mathbf{V})$ for a given value of s and gradually lowers s to avoid local minima and obtain a good crisp partition. Our fuzzy steps method performs the following steps.

1. Choose an initial value of s , a fuzzy step size $\gamma < 1$, and a threshold value s_{min} .
2. Choose initial \mathbf{P}_0 and \mathbf{Q}_0 and calculate \mathbf{V}_0 according to (5.12). The initial \mathbf{P}_0 and \mathbf{Q}_0 can be either crisp or fuzzy.
3. Repeat the following while $s > s_{min}$.
 - (a) Perform the two-mode fuzzy c -means algorithm starting with \mathbf{P}_0 , \mathbf{Q}_0 , and \mathbf{V}_0 . The results are in \mathbf{P}_1 , \mathbf{Q}_1 , and \mathbf{V}_1 .
 - (b) Set $s = 1 + \gamma(s - 1)$ and set $\mathbf{P}_0 = \mathbf{P}_1$, $\mathbf{Q}_0 = \mathbf{Q}_1$, and $\mathbf{V}_0 = \mathbf{V}_1$.
4. Apply the two-mode k -means algorithm starting from \mathbf{P}_0 , \mathbf{Q}_0 , and \mathbf{V}_0 .

The formula in Step 3b for decreasing s gives an exponential decay of $(s - 1)$. The value of s_{min} should generally be set to a value slightly higher than 1, for example, 1.001. The two-mode k -means algorithm is performed at the end of the fuzzy steps method to ensure that a crisp solution is found. Although the two-mode k -means algorithm was defined for crisp partitions in Section 5.2.2, it can also be used in combination with fuzzy initial partitions.

The fuzzy steps optimization method may sometimes get stuck in a saddle point, if two or more row or column clusters become equal. In that case, these clusters and their corresponding cluster membership values will remain equal for any value of s . Preliminary tests with this method suggest that it often reaches a saddle point, if the starting value of s is too high. Therefore, one should not set the starting value of s too high, for example, $s \leq 1.2$.

5.5 Simulation Study

To compare the performances of the optimization methods described in the previous sections, we conduct a simulation study. With this simulation study, we also aim to determine which methods perform well under most circumstances and how well the optimization methods can retrieve a clustering structure. First, we describe the setup of the simulation study and how the results are represented. We then give the results of the simulation study and interpret them.

5.5.1 Setup of the Simulation Study

We generate the data matrix \mathbf{X} in each problem instance by simulating \mathbf{P} , \mathbf{Q} , \mathbf{V} , and \mathbf{E} , and then using (5.1) to construct \mathbf{X} . Generating simulated data this way comes natural and has the advantage that some clustering structure exists in the data. Also, the \mathbf{P} , \mathbf{Q} , and \mathbf{V} used in generating \mathbf{X} can give a useful upper bound on the optimal value of $f(\mathbf{P}, \mathbf{Q}, \mathbf{V})$.

A large number of factors can be varied in a simulation study for two-mode partitioning, such as the values of n , m , K , and L , the size and the distribution of the errors, the numbers of elements in the clusters, and the locations of the cluster centers. As a full-factorial design with several levels for each of these factors would require a prohibitively large number of simulations, we limit the number of factors and the number of levels for each factor. The simulation study is set up using a four-factor design and is loosely based on the approach of Milligan (1980). The first factor is the size of the data matrix \mathbf{X} . The numbers of rows and columns are given by $n = m = 60$, $n = 150$ and $m = 30$, and $n = m = 120$ for the three levels of this factor. The second factor is the numbers of clusters; the three levels of this factor are $K = L = 3$, $K = L = 5$, and $K = L = 7$. The third factor is the size of the error perturbations. All elements of \mathbf{E} are independently normally distributed with mean 0 and standard deviation equal to 0.5, 1, or 2 for the three levels of this factor. The final factor is the distribution of the objects over the clusters. For the first level of this factor, all objects are divided over the clusters with equal probability. For the second and third levels, one cluster contains exactly 10% and 60% of the objects, respectively. The remaining objects are then divided over the remaining clusters with uniform probability. Constructing one cluster with 10% of the objects represents a small deviation from a uniform distribution, whereas a cluster with 60% constitutes a large deviation. The size of this deviation also depends on the numbers of clusters.

Empty clusters are not allowed in the generated data sets. If a simulated data set contains an empty cluster, this data set is discarded and another data set is simulated. The locations of the cluster centers v_{kl} are chosen by randomly assigning the numbers $\Phi(\frac{i}{K \times L + 1})$, $i = 1, \dots, K \times L$ to the elements of \mathbf{V} , where $\Phi(\cdot)$ is the inverse standard normal cumulative distribution function. As a result, the elements of \mathbf{V} appear standard normally distributed, and a fixed minimum distance between the cluster centers is ensured.

We show the effects of some of these choices in Figures 5.1 through 5.3. These figures give a visual representation of a simulated data matrix for various levels of the factors. In these figures, the rows and columns are ordered according to their cluster. The elements of the data matrix are represented by different shades of gray. The bars at the right-hand side show what values the shades of gray represent. Figure 5.1 represents a data set for which the levels of the factors ensure that the original clusters can easily be recognized. In Figure 5.2, the original clusters are more difficult to recognize, and in Figure 5.3, this is almost impossible. The optimal partition should generally be easier to find if the number of clusters is low, the error standard deviation is small, and the objects are evenly divided over the clusters.

The four factors in the simulation design give a total of $3^4 = 81$ possible combinations. To avoid drawing spurious conclusions based on a single data set, we simulate 50 data sets for each combination, and all five methods are performed for each data set.

All optimization methods require an initial choice for \mathbf{P} and \mathbf{Q} . For the alternating exchanges and two-mode k -means methods, \mathbf{P} and \mathbf{Q} are chosen randomly by assigning each row and column to a cluster with uniform probability. However, some methods may perform better if the initial partitions are relatively good. Therefore, the initial partitions

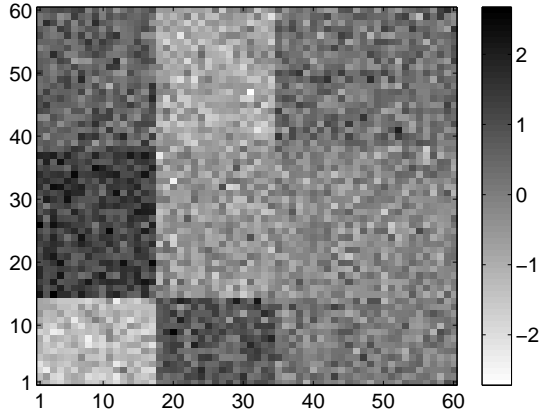


Figure 5.1: Graphical representation of simulated data set with sizes $n = m = 60$, $K = L = 3$, error standard deviation 0.5, and a uniform distribution of the objects over the clusters.

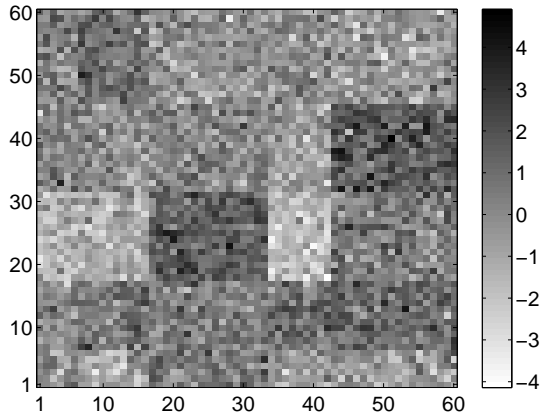


Figure 5.2: Graphical representation of simulated data set with sizes $n = m = 60$, $K = L = 5$, error standard deviation 1, and 10% of the objects in one cluster.

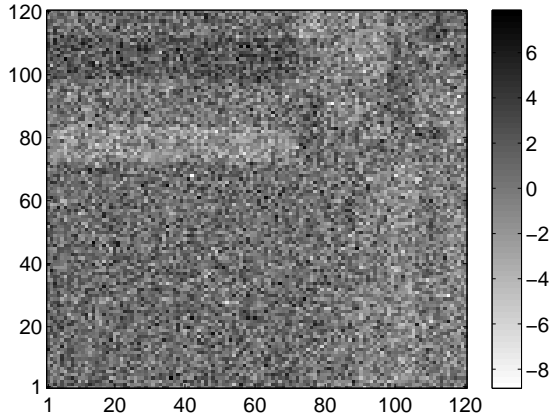


Figure 5.3: Graphical representation of simulated data set with sizes $n = m = 120$, $K = L = 7$, error standard deviation 2, and 60% of the objects in one cluster.

of the simulated annealing, tabu search, and fuzzy steps methods are chosen by applying the two-mode k -means algorithm to a random partition.

The optimization methods also require choosing additional parameters. For the optimization methods that use multiple random starts (that is, the alternating exchanges, two-mode k -means, and fuzzy steps methods), the numbers of random starts have been set so that the total computation time of these methods is approximately equal. This is also true for the numbers of iterations R and S that are respectively used in the simulated annealing and tabu search methods. We also wish to ensure a good balance between the performance and the computation time of the simulated annealing and tabu search methods for varying problem sizes. Therefore, the parameters R and S are chosen as a function of the size of the neighborhood used in these two methods, which equals $n(K - 1) + m(L - 1)$; as a result, the computation time should increase modestly with the size of the problem. The initial temperature in the simulated annealing method is chosen in such a way that the initial acceptance rate for the transfer of a row or a column to another cluster is at least 85%. As a result, the transfers of rows and columns should occur almost randomly in the initial iterations of this method. The cooling rate γ in this method should be chosen slightly lower than 1. We stop the simulated annealing method if there are no changes for $t_{max} = 10$ values of the temperature parameter, as further improvements are unlikely to occur in that case. Preliminary experimentation with the tabu search method shows that increasing the tabu list length typically has a positive effect on the performance of this method. We set the length of the tabu list to one third of the number of iterations S , as increasing the tabu list length any further would increase computation time without meaningfully improving the performance. In the fuzzy steps method, the fuzzy step size γ

is set to a value somewhat lower than 1. The threshold value of s (that is, the parameter s_{min}) is set to 1.001, as experimentation shows that the solution is almost always close to a crisp partition at that value of s . The initial value of s was optimized on a separate set of data sets, which were simulated in the same way as the data sets in the simulation study. Initial values of $s = 1.025, 1.05, \dots, 1.2$ were tested, and an optimal recovery was achieved with $s = 1.05$.

The parameters of the optimization methods are thus chosen as follows:

- Alternating exchanges: The alternating exchanges algorithm is run 30 times for every simulated data set.
- Two-mode k -means: The two-mode k -means algorithm is run 500 times for every simulated data set.
- Simulated annealing: Initial temperature $T = 1$, $\gamma = 0.95$, $t_{max} = 10$, and $R = 1/4(n(K-1) + m(L-1))$.
- Tabu search: The number of iterations is $S = 3(n(K-1) + m(L-1))^{1/2}$, and the length of the tabu list is $1/3S$.
- Fuzzy steps: The initial value of s is 1.05, the fuzzy step size γ is 0.85, and the threshold value s_{min} is 1.001. The fuzzy steps method is run 20 times for every simulated data set.

The values of the parameters that are a function of the neighborhood size $n(K-1) + m(L-1)$ are rounded to integer values.

We use three criteria to evaluate the results of the simulation study. The first criterion is the variance accounted for (VAF) criterion, which is comparable to the R^2 -measure used in regression analysis. The VAF criterion is defined as

$$\text{VAF} = 1 - \frac{\sum_{k=1}^K \sum_{l=1}^L \sum_{i=1}^n \sum_{j=1}^m p_{ik}q_{jl}(x_{ij} - v_{kl})^2}{\sum_{i=1}^n \sum_{j=1}^m (x_{ij} - \bar{x})^2}, \quad (5.13)$$

where $\bar{x} = 1/(nm) \sum_{i=1}^n \sum_{j=1}^m x_{ij}$. The optimal value of VAF ranges from 0 to 1, and maximizing VAF corresponds to minimizing $f(\mathbf{P}, \mathbf{Q}, \mathbf{V})$.

Second, we report the average adjusted Rand index (ARI), which was introduced by Hubert and Arabie (1985). This metric shows how well the partitions found by the optimization methods approximate the original partition. The ARI is invariant with respect to the ordering of the clusters, which is arbitrary in two-mode partitioning. The ARI is based on the original Rand index, which is defined as the fraction of the pairs of elements on which the two partitions agree. For two random partitions, the expected value of the Rand index depends on the parameters of the clustering problem. To ensure a constant expectation, the ARI is a linear function of the Rand index, so that its expectation for randomly chosen partitions is 0, and its maximum value is 1. The ARI is defined as

$$\text{ARI} = \frac{\sum_{i=1}^R \sum_{j=1}^R \binom{a_{ij}}{2} - \sum_{i=1}^R \binom{a_i}{2} \sum_{j=1}^R \binom{a_j}{2} / \binom{a}{2}}{[\sum_{i=1}^R \binom{a_i}{2} + \sum_{j=1}^R \binom{a_j}{2}] / 2 - \sum_{i=1}^R \binom{a_i}{2} \sum_{j=1}^R \binom{a_j}{2} / \binom{a}{2}}, \quad (5.14)$$

where a_{ij} is the number of elements of \mathbf{X} that simultaneously belong to cluster i of the original partition and cluster j of the retrieved partition, R is the total number of clusters, $a_{i.} = \sum_{j=1}^R a_{ij}$, $a_{.j} = \sum_{i=1}^R a_{ij}$, and $a = \sum_{i=1}^R \sum_{j=1}^R a_{ij} = nm$. Here we consider a pair of elements to be in the same cluster only if they belong to same row cluster and to the same column cluster, so that $R = K \times L$.

The final criterion is the average CPU time, which is of practical interest. In this study, the CPU time is also used as a control variable, as the parameters of the optimization methods have been chosen in such a way, that the average computation time is approximately equal for all methods. To provide a fair comparison, we made sure that the optimization methods were programmed in an efficient way. Efficient updating formulas for computing the change in $f(\mathbf{P}, \mathbf{Q}, \mathbf{V})$ when transferring an object from one cluster to another are given in Castillo and Trejos (2002). These updating formulas were used for the alternating exchanges, simulated annealing, and tabu search methods. For the two-mode k -means algorithm, an efficient implementation is given by Rocci and Vichi (2008). All computer programs are written in the matrix programming language MATLAB 7.2 and are executed on a Pentium IV 2.8 GHz computer. These programs are available from the authors upon request.

Dolan and Moré (2002) discuss a convenient tool, called *performance profiles*, for graphically representing the distribution of the results of a simulation study. Performance profiles are especially useful for determining which optimization methods perform reasonably well in almost every instance. They can be constructed as follows. First, one has to identify a performance measure. We use the VAF criterion for this and define $\text{VAF}^{(p,s)}$ as the VAF achieved by method s in problem instance p . We include the partition used to generate the data as one of the methods in the performance profiles. Then, the *performance ratio* $\rho^{(p,s)}$ is defined as

$$\rho^{(p,s)} = \frac{\max_s \text{VAF}^{(p,s)}}{\text{VAF}^{(p,s)}}. \quad (5.15)$$

Finally, the cumulative distribution function of the performance ratio can be computed as

$$\Psi^{(s)}(\tau) = \frac{1}{50} \sum_{p=1}^{50} I(\rho^{(p,s)} \leq \tau), \quad (5.16)$$

where $I()$ denotes the indicator function. By drawing $\Psi^{(s)}(\tau)$ in one figure for all optimization methods, their performances can be compared quite easily.

5.5.2 Simulation Results

We now present the results of the simulation study. Tables 5.1 through 5.3 show the average VAF values of all optimization methods and the original partitions for various combinations of the factors. Tables 5.1, 5.2, and 5.3 respectively show the effects of the size of the data set, the numbers of clusters, and the distribution of the objects over the clusters. As the error standard deviation strongly affects the magnitude of the VAF values,

Table 5.1: Average VAF for data sets with different error variances and numbers of rows and columns.

Size of data set Error st. dev.	$n = 60, m = 60$			$n = 150, m = 30$			$n = 120, m = 120$		
	.5	1	2	.5	1	2	.5	1	2
Alternating exchanges	.7260	.4128	.1685	.7288	.4140	.1750	.7251	.4058	.1532
Two-mode k -means	<i>.7283</i>	<i>.4134</i>	.1667	<i>.7322</i>	<i>.4146</i>	.1733	<i>.7293</i>	<i>.4082</i>	<i>.1535</i>
Simulated annealing	.7092	.4068	.1668	.7166	.4078	.1739	.7108	.3987	.1518
Tabu search	.6823	.4003	.1642	.6904	.4015	.1660	.6703	.3853	.1496
Fuzzy steps	.7248	.4129	<i>.1688</i>	.7299	.4143	<i>.1755</i>	.7199	.4064	.1531
Original partition	<i>.7292</i>	.4126	.1540	<i>.7326</i>	.4122	.1531	.7308	.4088	.1515

Table 5.2: Average VAF for data sets with different error variances and numbers of clusters.

Numbers of clusters Error st. dev.	$K = 3, L = 3$			$K = 5, L = 5$			$K = 7, L = 7$		
	.5	1	2	.5	1	2	.5	1	2
Alternating exchanges	.6843	.3558	.1294	.7399	.4260	<i>.1725</i>	.7557	.4508	.1948
Two-mode k -means	<i>.6843</i>	<i>.3558</i>	<i>.1296</i>	<i>.7440</i>	<i>.4276</i>	.1724	<i>.7614</i>	<i>.4529</i>	.1915
Simulated annealing	.6804	.3525	.1284	.7221	.4186	.1706	.7340	.4422	.1935
Tabu search	.6584	.3476	.1259	.6868	.4066	.1663	.6980	.4329	.1876
Fuzzy steps	.6838	.3558	.1288	.7394	.4264	.1725	.7515	.4515	<i>.1960</i>
Original partition	.6843	.3549	.1214	.7442	.4265	.1601	.7640	.4522	.1771

we do not report average VAF values over data sets with different error standard deviations; the effects of the error standard deviation are shown within each of these three tables. The results of the best performing optimization method are shown in italics, for each column in these tables.

Based on the VAF values, the two-mode k -means method appears to perform relatively well. This method always has the best average performance if the error standard deviation is .5 or 1. In addition, the partition found by this method is almost always as least as good as the original partition. For data sets with an error standard deviation of 2, the best performance is most often achieved by the fuzzy steps method, though the alternating

Table 5.3: Average VAF for data sets with different error variances and distributions of the objects over the clusters.

Object distribution Error st. dev.	uniform distribution			10% in one cluster			60% in one cluster		
	.5	1	2	.5	1	2	.5	1	2
Alternating exchanges	.7460	.4293	.1695	.7330	.4171	.1666	.7009	.3862	.1606
Two-mode k -means	<i>.7467</i>	<i>.4294</i>	.1683	<i>.7336</i>	<i>.4174</i>	.1655	<i>.7094</i>	<i>.3894</i>	.1596
Simulated annealing	.7352	.4255	.1684	.7199	.4105	.1650	.6814	.3774	.1593
Tabu search	.6858	.4127	.1641	.6818	.3993	.1606	.6755	.3751	.1552
Fuzzy steps	.7419	.4291	<i>.1698</i>	.7299	.4170	<i>.1668</i>	.7029	.3875	<i>.1608</i>
Original partition	<i>.7467</i>	.4288	.1602	<i>.7337</i>	.4167	.1566	.7121	.3881	.1418

exchanges and the two-mode k -means methods also perform well. The simulated annealing method and especially the tabu search method do not perform as well as the other optimization methods in the simulation study.

Table 5.4 gives the average values of the adjusted Rand Index, for each level of each factor, averaged over the levels of the three other factors; the best results are shown in italics. The two-mode k -means method again has the best average performance. This method almost always exactly retrieves the original partition for data sets with a small error standard deviation and a uniform distribution of the rows and columns over the clusters. The best performing optimization method usually has an average ARI above 90% or even 99% if the characteristics of the simulated data sets are favorable. However, this value rapidly decreases, when the problem instances become harder. The original partitions are especially hard to retrieve if the error standard deviation is 2 or if 60% of the objects are located in one cluster. The ARI never becomes negative or close to 0 for any optimization method, indicating that at least some structure can always be found in the data set. Another important conclusion is that the differences in the average adjusted Rand indices between the methods can be large. The difference in ARI can sometimes be as large as 20%, when the corresponding difference in VAF is just a few percentage points. Therefore, the choice of the optimization method can be quite important, if one wants to find the ‘true’ clustering.

Finally, we give the average CPU times in seconds used by the methods in Table 5.5, for each size of the data set and each number of clusters. Note that the average CPU times strongly depend on the type of computer used and how the methods have been implemented; they should only serve as a general indication of the amount of computation that optimization methods require. The effects of the distribution of the objects over the clusters and the error standard deviation on the computation time are not reported here. The computation time does not increase with m , n , K , and L in the same manner for all optimization methods. The computation time of the two-mode k -means method increases relatively slowly with the size of the data set and the numbers of clusters; the increase of the computation time is more rapid for the tabu search and fuzzy steps methods.

The average VAF values in Tables 5.1 through 5.3 do not show the distribution of the VAF values. For example, a method may usually perform well, but occasionally give very poor results. Here, we show performance profiles of the VAF values for all methods and the original partition, for various combinations of the levels of the four factors. As drawing these profiles for each of the 81 combinations of the factors requires too much space, we give three examples in Figures 5.4 through 5.6.

Figure 5.4 shows a performance profile of 50 data sets with a low error standard deviation. The value of the graph at the y -axis gives the fraction of times that a method achieved the best VAF value of all methods (including the original partition); this value is an upper bound on the fraction of times that a method managed to find the global optimum. The alternating exchanges, two-mode k -means method, and fuzzy steps methods found the best partition, which usually was the original partition, in every problem instance. The simulated annealing and tabu search methods usually also found the best partition, but

Table 5.4: Average adjusted Rand indices for each level of each factor.

Size of data set	$n = 60, m = 60$	$n = 150, m = 30$	$n = 120, m = 120$
Alternating exchanges	.739	.721	.846
Two-mode k -means	.759	.748	.893
Simulated annealing	.678	.673	.793
Tabu search	.636	.613	.729
Fuzzy steps	.745	.737	.851
Numbers of clusters	$K = 3, L = 3$	$K = 5, L = 5$	$K = 7, L = 7$
Alternating exchanges	.865	.772	.669
Two-mode k -means	.868	.815	.718
Simulated annealing	.830	.702	.612
Tabu search	.775	.637	.567
Fuzzy steps	.858	.783	.692
Error standard deviation	.5	1	2
Distribution of the objects	uniform distribution	10% in one cluster	60% in one cluster
Alternating exchanges	.915	.853	.538
Two-mode k -means	.974	.889	.537
Simulated annealing	.852	.788	.503
Tabu search	.782	.742	.454
Fuzzy steps	.934	.863	.537
Distribution of the objects	uniform distribution	10% in one cluster	60% in one cluster
Alternating exchanges	.874	.861	.571
Two-mode k -means	.873	.860	.668
Simulated annealing	.837	.805	.501
Tabu search	.741	.727	.510
Fuzzy steps	.868	.856	.610

Table 5.5: Average CPU times of all optimization methods in seconds.

Size of data sets	$n = 60, m = 60$			$n = 150, m = 30$			$n = 120, m = 120$		
Numbers of clusters (K, L)	3	5	7	3	5	7	3	5	7
Alternating exchanges	1.21	2.54	3.96	1.92	4.22	6.43	2.39	5.59	8.72
Two-mode k -means	1.54	2.36	3.09	1.96	3.58	4.57	2.50	5.31	7.66
Simulated annealing	1.08	2.30	3.54	1.75	3.82	5.78	2.42	5.47	8.49
Tabu search	.82	2.16	3.74	1.47	3.82	6.77	2.34	6.04	10.75
Fuzzy steps	.99	1.91	2.66	1.05	2.48	3.47	2.22	7.92	11.28

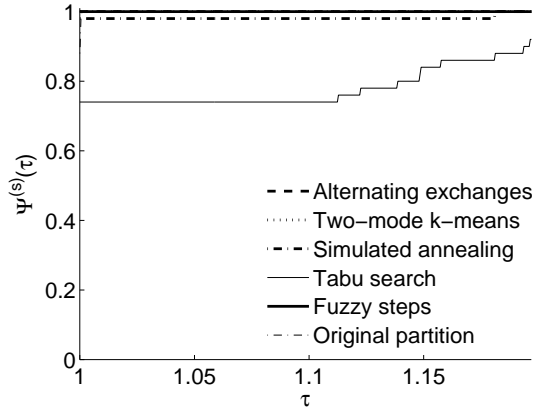


Figure 5.4: Performance profile of data sets with sizes $n = 150$, $m = 30$, $K = 3$, $L = 3$, error standard deviation $.5$, and a uniform distribution of the objects.

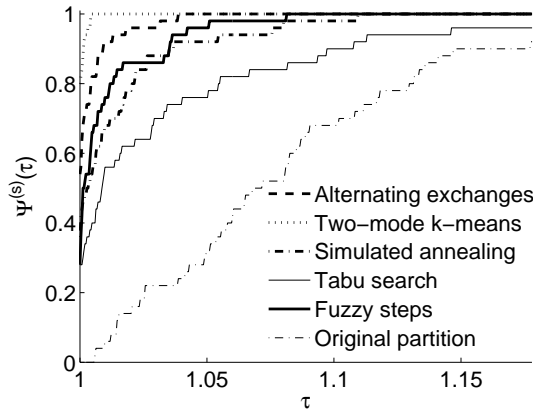


Figure 5.5: Performance profile of data sets with sizes $n = 60$, $m = 60$, $K = 5$, $L = 5$, error standard deviation 2 , and one cluster containing 10% of the objects.

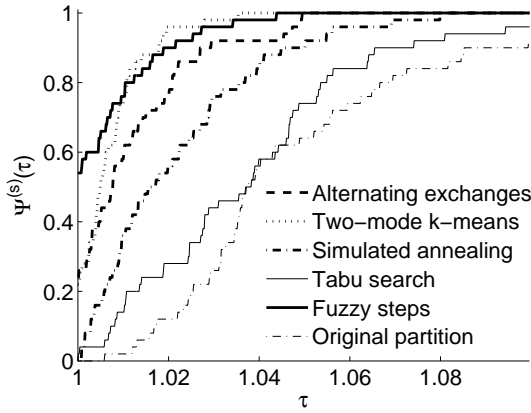


Figure 5.6: Performance profile of data sets with sizes $n = 120$, $m = 120$, $K = 7$, $L = 7$, error standard deviation 2, and one cluster containing 60% of the objects.

sometimes got stuck in an inferior solution. Figures 5.5 and 5.6 show performance profiles of data sets with a large error standard deviation, in which the optimal partition was hard to find. In these cases, no method performed at least as good as all others in every problem instance. In Figure 5.5, the two-mode k -means method performed best, followed by the alternating exchanges and fuzzy steps methods. This performance profile shows that each of these three methods almost always found a solution within 5 to 10 percent of the best solution found, in terms of the optimization criterion $f(\mathbf{P}, \mathbf{Q}, \mathbf{V})$. In Figure 5.6, each optimization method ended up in a non-global optimum for at least 46% of the problem instances. In this figure, the fuzzy steps and two-mode k -means methods had the best relative performances, followed by the alternating exchanges method. These three methods seldom find a solution that is more than 5 percent worse than the best solution found. The two implementations of meta-heuristics (simulated annealing and tabu search) do not perform as well as the other methods, but can usually find partitions that are better than the original partition.

As all optimization methods do not find a global optimum for a large number of problem instances represented in Figure 5.6, their performances may be considered unsatisfactory. If finding the global optimum is required, one may consider increasing the numbers of random starts in the optimization methods. To test the effectiveness of increasing the number of random starts, we applied the two-mode k -means and fuzzy steps methods to data sets similar to the ones used for Figure 5.6 with 50,000 and 2,000 random starts, respectively. The results show that increasing the numbers of random starts in this way (by a factor of 100) improves the performances of the methods only slightly; the average VAF values for the methods with increased numbers of random starts were approximately 0.2 percentage points higher than for the methods with the original numbers of random

starts. However, even with these large numbers of random starts, no optimization method performed as well as all others for a majority of the data sets. These results show that all optimization methods have great difficulty in finding the global optimum in conditions such as represented in Figure 5.6. This problem is much less severe for smaller data sets and data sets with lower numbers of clusters and lower error standard deviations. As strongly increasing the number of random starts tends to improve the VAF value only slightly, we find it unlikely that the solution found by the optimization methods are much worse than the global optimum, in terms of the VAF criterion.

5.6 Empirical Application

We use an empirical data set to illustrate two-mode partitioning and determine whether the conclusions of the simulation study are valid for empirical data. The data set is based on a questionnaire about the internet and comprises evaluations of 22 statements by 193 respondents. The statements were evaluated using a seven-point Likert scale, ranging from 1 (completely disagree) to 7 (completely agree). The average scores in the data set can differ significantly per individual and per statement. A sample run of the optimization methods shows that, if the raw data set is used, the individuals and the statements are mostly clustered based on their average scores. We correct for this problem by double centering the data matrix \mathbf{X} , that is, by replacing each x_{ij} with \tilde{x}_{ij} , where

$$\tilde{x}_{ij} = x_{ij} - \frac{1}{n} \sum_{i'=1}^n x_{i'j} - \frac{1}{m} \sum_{j'=1}^m x_{ij'} + \frac{1}{nm} \sum_{i'=1}^n \sum_{j'=1}^m x_{i'j'}, \quad (5.17)$$

so that the row and column averages of $\tilde{\mathbf{X}}$ are zero; double centering accounts for 34% of the variance in \mathbf{X} .

As good values for the numbers of clusters are not known in advance, we use the following procedure for choosing K and L . First, we perform the optimization methods on $\tilde{\mathbf{X}}$, for all K and L such that $K + L = i$, for $i = 4, \dots, 20$. For each value i and for each optimization method, we determine the values of K and L that yield the best VAF. The resulting VAF values are shown in Figure 5.7. The optimization methods were given the same parameters as in the simulation study. For this data set, the fuzzy steps method performed best on average, followed by the alternating exchanges, simulated annealing, and two-mode k -means methods. The tabu search method performed poorly. These results appear to support the conclusions of the simulation study for data sets in which the optimal partition is hard to find.

The maximum of the VAF values shown in Figure 5.7 increases smoothly with the numbers of clusters, so that it is not clear which values of K and L should be chosen. To determine the numbers of clusters for two-mode partitioning, several selection criteria are discussed by Schepers, Ceulemans, and Van Mechelen (2008); they found the best performance using the numerical convex hull method. This criterion prefers $K = 2$ and $L = 2$, with a VAF value of 8.8%. In this solution, both the statements and the respondents

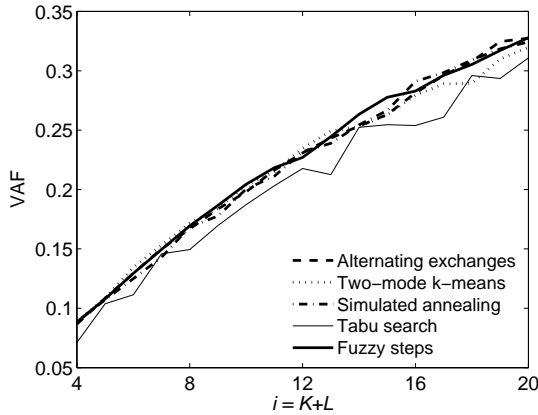


Figure 5.7: Best VAF values found in internet data set, for $K + L = 4 \dots 20$.

are divided into a cluster with a positive attitude and a cluster with a negative attitude. We find that this solution lacks a useful interpretation. As we seek to obtain a more fine-grained clustering, we select the solution with $K = 5$ and $L = 5$, based on the interpretability of the findings. The results of this solution are shown in Table 5.6. The associated VAF value is 20.4%, so that two-mode partitioning in combination with double centering accounts for 48% of the variance in the original data.

The statements in each statement cluster are shown in Table 5.7. Although the proportion of the variance of $\tilde{\mathbf{X}}$ explained by the clustering is relatively small (20.4%), it seems possible to interpret the clusters in a meaningful way. The statement clusters can be interpreted as follows. Cluster 1 comprises only one statement, which argues that the content of websites should be regulated. Cluster 2 consists of statements that show an enthusiastic attitude towards the internet. The statements in cluster 3 state that using the internet is expensive. The statements in cluster 4 are associated with experienced users of the internet. Finally, the statements in cluster 5 consider the internet unreliable. To

Table 5.6: Average evaluations per cluster, cluster sizes, and interpretations.

Respondent clusters	Statement clusters					Cluster	
	1	2	3	4	5	size	Interpretation
1	-2.03	0.19	0.42	0.40	-0.89	36	enthusiasts
2	1.86	0.35	0.15	-0.34	-0.41	54	safety-conscious
3	-1.42	-0.02	-1.42	0.81	-0.16	32	experts
4	1.61	-0.37	-0.46	-0.28	1.06	37	skeptics
5	-1.21	-0.33	1.15	-0.33	0.59	34	price-conscious
Cluster size	1	6	3	8	4		
Interpretation	regulation	enthusiastic	expensive	experience	unreliable		

Table 5.7: Statement clusters in internet data set.

Cluster 1 (regulation):
The content of websites should be regulated
Cluster 2 (enthusiastic):
Internet is easy to use
Internet is fast
Internet is the future's means of communication
Internet is user-friendly
Internet offers unbounded opportunities
Surfing the internet is easy
Cluster 3 (expensive):
Internet phone costs are high
The costs of surfing are high
The prices of internet subscriptions are high
Cluster 4 (experience):
I always attempt new things on the internet first
I know much about the internet
I like surfing
I like to be informed of important new things
I often speak with friends about the internet
I regularly visit websites recommended by others
Internet is addictive
Paying using the internet is safe
Cluster 5 (unreliable):
Internet is slow
Internet is unreliable
Internet offers many possibilities for abuse
Transmitting personal data using the internet is unsafe

interpret the respondent clusters, we use the interpretations of the statement clusters and the values of the cluster centers in Table 5.6. The respondents in cluster 1 are ordinary internet users with a positive and enthusiastic attitude. Cluster 2 consists of people who mainly want regulation of the content of websites. Cluster 3 consists of experienced internet users, who find using the internet quite cheap, but apparently are not enthusiastic about it. The respondents in cluster 4 seem to dislike using the internet and find it unreliable. The respondents in the final cluster seem to dislike using the internet because of the high costs.

5.7 Conclusions

Two-mode partitioning seems a powerful statistical technique, and a variety of optimization methods exist for finding good partitions. However, it remains unclear which optimization method has the best relative performance and is thus preferable in practice. We have tried to alleviate this problem by giving an overview of five optimization methods, one of which has been introduced in this chapter.

Using a simulation study, the effects of various characteristics of the clustering problem on the performances of the methods were evaluated. We found that both the error variance and the relative sizes of the clusters strongly affect how well a clustering structure can be retrieved and can also affect the relative performances of the optimization methods. The two-mode k -means method most often had the best performance in the simulated data sets, especially if the characteristics of the problem were favorable. When the optimal partition was hard to find, the alternating exchanges and fuzzy steps methods often performed best. The simulated annealing method also performed fairly well, but not as good as the three methods mentioned above. The performance of the tabu search method generally is inferior.

In sum, the best average performance is obtained using the two-mode k -means method, followed by the alternating exchanges and fuzzy steps methods. Which of these three methods gives the best performance depends on characteristics of the clustering problem. The performance profiles show that no optimization method can find the global optimum in a majority of the cases for every type of simulated data set. However, the alternating exchanges, two-mode k -means, and fuzzy steps methods almost always give a solution within 10% of the best solution found, in terms of the optimization criterion. As the two-mode k -means method requires relatively little CPU time and is easy to implement, we believe it is the best choice for data sets similar to the ones in this study. Note that performing a fairly large number (preferably more than 100) of random starts is necessary to ensure a good performance for this method. If finding a globally optimal partition is important, and computation time is not really an issue, the number of random starts should be set much higher. However, for data sets in which the optimal partition is hard to find, even a very large number of random starts may not ensure that the global optimum is found.

The results of the methods for the empirical data set are similar to the results of the

simulation study. This data set also gives a useful example of the potential applications of two-mode partitioning. We believe that the clustering found in this data set is meaningful and provides relevant insight.

There are a few limitations associated with this study. First, we can only evaluate the relative performances of the methods for situations similar to our simulation study. It is possible that the relative performances of the methods are different for data sets with different characteristics. However, we have tried to make the simulated data sets similar to data sets that are typically found in practice. Therefore, the results should be reasonably similar for many empirical data sets. Second, the results of such a comparison of optimization methods depend on the values of certain parameters of the methods and the way these methods have been implemented. Here, we have tried to choose the parameters of the optimization methods in a sensible way that gives good performance. In addition, we have tried to make our computer programs as fast as possible. To do so, the efficient updating formulas that were given by Castillo and Trejos (2002) and Rocci and Vichi (2008) turn out to be quite important.

Some avenues for further research exist. For example, we cannot exclude the possibility that some optimization methods can be improved by choosing their parameters differently or implementing them in a different way. Besides further theoretical research, we believe that further practical experience with two-mode partitioning is also required. Practice can show the real merits and drawbacks of using two-mode partitioning.

Chapter 6

A Bayesian Approach to Two-Mode Clustering*

Abstract

We develop a new Bayesian approach to estimate the parameters of a latent-class model for the joint clustering of both modes of two-mode data matrices. Posterior results are obtained using a Gibbs sampler with data augmentation. Our Bayesian approach has three advantages over existing methods. First, we are able to do statistical inference on the model parameters, which would not be possible using frequentist estimation procedures. In addition, the Bayesian approach allows us to provide statistical criteria for determining the optimal numbers of clusters. Finally, our Gibbs sampler has fewer problems with local optima in the likelihood function and empty classes than the EM algorithm used in a frequentist approach. We apply the Bayesian estimation method of the latent-class two-mode clustering model to two empirical data sets. The first data set is the Supreme Court voting data set of Doreian et al. (2004). The second data set comprises the roll call votes of the United States House of Representatives in 2007. For both data sets, we show how the results can provide useful insight into the data.

6.1 Introduction

Clustering algorithms divide a single set of objects into segments based on their similarities and properties or the dissimilarities between them (see, for example, Hartigan, 1975). Such methods typically operate on one mode (dimension) of a data matrix; we refer to these methods as one-mode clustering. Two-mode clustering techniques (Van Mechelen et al., 2004) cluster two sets of objects into segments based on their interactions. In two-mode clustering, both rows and columns of data matrix are clustered simultaneously.

*This chapter is joint work with Bram van Dijk and Richard Paap.

Many clustering methods, such as k -means clustering and Ward's method, lack a method to ascertain the significance of the results and rely on arbitrary methods to determine the number of clusters. To solve these problems, one may consider using model-based techniques for clustering data. For one-mode data, model-based clustering methods have been developed (see, for example, Fraley & Raftery, 1998; Frühwirth-Schnatter, 2006). These model-based clustering methods use statistical tools for inference.

In this chapter, we extend the model-based one-mode clustering approach to two-mode clustering. In two-mode clustering, we cluster both the rows and the columns of a data matrix into groups in such a way that the resulting block structure is homogenous within blocks but differs between blocks. This requires matrix-conditional data, which means that all elements must be comparable in size, standardized, or measured on the same scale. Methods for two-mode clustering are in general not model-based (see, for example, Candel & Maris, 1997; Doreian et al., 2004; Brusco & Steinley, 2006, and Chapter 5 of this thesis). Model-based methods usually rely on latent-class techniques. It is not straightforward to extend these techniques to two-mode data, because, unlike one-mode data, two-mode data cannot be assumed to be independent. Despite this problem, Govaert and Nadif (2003, 2008) have been able to use a latent-class approach to cluster two-mode data. They use a frequentist approach to estimate the parameters, but they are only able to optimize an approximation of their likelihood function using the EM algorithm (Dempster et al., 1977). In this chapter, we use the same likelihood function as Govaert and Nadif (2003, 2008), but we propose a Bayesian estimation procedure. This enables us to estimate the model parameters properly and to do statistical inference on the results.

The contribution of our Bayesian approach is threefold. First, our approach allows for statistical inference on the parameter estimates. Govaert and Nadif (2003, 2008) estimate the model parameters in a frequentist setting, but they are unable to compute standard errors of the estimated parameters. Our method provides posterior standard deviations and posterior distributions of the parameters. Therefore, our approach enables hypothesis testing, which is not feasible in the frequentist setting.

Second, our Bayesian method has fewer computational problems than the maximum likelihood approach. Using proper priors, we avoid some computational issues with empty classes, which is a known problem with the EM algorithm. Because we use natural conjugate priors, sampling from the posterior distributions is straightforward for a large variety of distributions, including all distributions from the exponential family. Furthermore, because of the more flexible way Markov Chain Monte Carlo methods search the parameter space, our Bayesian algorithm is less likely to get stuck in a local optimum of the likelihood function. This flexibility may cause *label switching* (see Celeux, Hurn, & Robert, 2000). However, solutions to this problem exist (see, for example, Geweke, 2007).

Finally, our method can help indicate the optimal number of segments. The Bayesian approach can be used to derive selection criteria such as *Bayes factors*. Methods previously proposed in the literature for selecting the optimal number of clusters (see, for example, Milligan & Cooper, 1985; Schepers et al., 2008) seem somewhat arbitrary and lack theoretical underpinnings.

We illustrate our Bayesian approach using two data sets. The first data set comprises votes of the Supreme Court of the United States and was also used by Doreian et al. (2004). Our approach results in a similar solution; however, the optimal numbers of segments are lower than in their solution. Our second application is a large data set concerning roll call voting in the United States House of Representatives. We use our model to cluster both the representatives and the bills simultaneously.

The remainder of this chapter is organized as follows. In Section 6.2, we introduce our new Bayesian approach for clustering two-mode data. We compare this Bayesian approach with the existing frequentist approaches of (Govaert & Nadif, 2003, 2008). In Section 6.3, we discuss the posterior simulator for our Bayesian approach and the selection of the numbers of segments. In Section 6.4, the Bayesian approach is illustrated on the Supreme Court voting data. Section 6.5 deals with our second application, which concerns roll call votes of the United States House of Representatives in 2007. Finally, Section 6.6 concludes.

6.2 The Latent-Class Two-Mode Clustering Model

In this section, we present our Bayesian approach to clustering both modes of two-mode data simultaneously. We first give a derivation of the likelihood function and then discuss Bayesian parameter estimation for the latent-class two-mode clustering model.

6.2.1 The Likelihood Function

For illustrative purposes, we start this discussion with one-mode data, that is, we have N observations denoted by $\mathbf{y} = (y_1, \dots, y_N)'$. These observations can be discrete or continuous, and one-dimensional or multidimensional. We assume that each observation comes from one of K segments, and that the elements within each segment are independently and identically distributed. As a result, all observations must be independent. Furthermore, we assume that the observations come from a known distribution that is the same for all segments; only the parameters of the distribution vary among the segments. This can be described by a mixture model. Let $k_i \in \{1, \dots, K\}$ be an indicator for the segment to which observation y_i belongs, and let $\mathbf{k} = (k_1, \dots, k_N)'$. The conditional density of y_i belonging to segment q only depends on the parameter vector θ_q and is denoted by $g(y_i|\theta_q)$. The segment membership is unknown. We assume that the probability that observation y_i belongs to segment q is given by κ_q for $q = 1, \dots, K$, with $\kappa_q > 0$ and $\sum_{q=1}^K \kappa_q = 1$. We collect the so-called mixing proportions κ_q in the vector $\boldsymbol{\kappa} = (\kappa_1, \dots, \kappa_K)'$. The likelihood function of this model is given by

$$l(\boldsymbol{\theta}, \boldsymbol{\kappa}) = \prod_{i=1}^N \left\{ \sum_{q=1}^K \kappa_q g(y_i|\theta_q) \right\}, \quad (6.1)$$

where $\boldsymbol{\theta} = (\theta_1, \dots, \theta_K)'$.

To cluster two-mode data, we would like to extend (6.1) to two-mode data matrices, with a simultaneous clustering of both rows and columns. We aim to construct a model in which the observations that belong to the same row cluster and the same column cluster are independently and identically distributed. In two-mode clustering, unlike in one-mode clustering, this assumption does not ensure that all observations are independent. As a result, a naive extension of the one-mode likelihood function to two modes will not adequately model the dependence structure in the data.

Assume that \mathbf{Y} is an $(N \times M)$ matrix, and that we want to cluster the rows into K latent classes and the columns into L latent classes. The naive extension of (6.1) to two-mode data yields

$$l(\theta, \kappa, \lambda) = \prod_{i=1}^N \prod_{j=1}^M \sum_{q=1}^K \sum_{r=1}^L \kappa_q \lambda_r g(Y_{i,j} | \theta_{q,r}), \quad (6.2)$$

where $\kappa = (\kappa_1, \dots, \kappa_K)'$ gives the size of each row segment, $\lambda = (\lambda_1, \dots, \lambda_L)'$ gives the size of each column segment, and $\theta_{q,r}$ contains the parameters of observations in row segment q and column segment r . Model (6.2) fails to impose that all elements in a row belong to the same row cluster and also does not impose that all elements in a column belong to the same column cluster; using this model, the data matrix \mathbf{Y} would effectively be modeled as a vector of one-mode data.

To solve this problem, we first rewrite the one-mode likelihood function (6.1) as

$$\begin{aligned} l(\theta, \kappa) &= \prod_i^N \left\{ \sum_{q=1}^K \kappa_q g(y_i | \theta_q) \right\} \\ &= \left\{ \sum_{q=1}^K \kappa_q g(y_1 | \theta_q) \right\} \left\{ \sum_{q=1}^K \kappa_q g(y_2 | \theta_q) \right\} \cdots \left\{ \sum_{q=1}^K \kappa_q g(y_N | \theta_q) \right\} \\ &= \sum_{k_1=1}^K \sum_{k_2=1}^K \cdots \sum_{k_N=1}^K \prod_{i=1}^N \kappa_{k_i} g(y_i | \theta_{k_i}) \\ &= \sum_{\mathbf{k} \in \mathcal{K}} \prod_{q=1}^K \kappa_q^{N_{\mathbf{k}}^q} \prod_{i=1}^N g(y_i | \theta_{k_i}), \end{aligned} \quad (6.3)$$

where we introduce some new notation in the last line. First, the set \mathcal{K} contains all possible divisions of items into the segments and thus has K^N elements if there are N items and K possible segments. Second, $N_{\mathbf{k}}^q$ equals the number of items belonging to segment q according to segmentation \mathbf{k} . Thus, $\sum_{q=1}^K N_{\mathbf{k}}^q = N$ for a fixed classification \mathbf{k} . The fact that these two representations of the likelihood function of a mixture model are equivalent was already noticed by Symons (1981).

Using this representation, we can extend the mixture model to clustering two modes

simultaneously. The resulting likelihood function for two modes is

$$l(\theta, \kappa, \lambda) = \sum_{\mathbf{k} \in \mathcal{K}} \sum_{\mathbf{l} \in \mathcal{L}} \prod_{q=1}^K \kappa_q^{N_q^{\mathbf{k}}} \prod_{r=1}^L \lambda_r^{M_r^{\mathbf{l}}} \prod_{i=1}^N \prod_{j=1}^M g(Y_{i,j} | \theta_{k_i, l_j}), \quad (6.4)$$

where \mathcal{L} denotes all possible divisions of the columns into L segments, and $M_r^{\mathbf{l}}$ equals the number of items belonging to segment r according to column segmentation $\mathbf{l} = (l_1, \dots, l_M)'$. Note that it is impossible to rewrite (6.4) as a product of likelihood contributions, which is possible in the one-mode case (6.1).

6.2.2 Parameter Estimation

The likelihood function (6.4) was already proposed by Govaert and Nadif (2003), who estimate the parameters of this model in a frequentist setting. However, this approach has several limitations. First, in contrast to the likelihood function in the one-mode case, the likelihood function (6.4) cannot be written as a product over marginal/conditional likelihood contributions; we only have a sample of size 1 from the joint distribution of \mathbf{Y} , \mathbf{k} , and \mathbf{l} . Therefore, the classical results for the asymptotic properties of the maximum likelihood estimator are not applicable.

Second, standard approaches to maximize the likelihood function (6.4) and estimate the model parameters are almost always computationally infeasible. Enumerating the $K^N L^M$ possible ways to assign the rows and columns to clusters in every iteration of an optimization routine is only possible for extremely small data sets. To solve this problem, Govaert and Nadif (2003) instead consider the so-called classification likelihood approach, in which \mathbf{k} and \mathbf{l} are parameters that need to be optimized. Hence, one maximizes

$$l(\theta, \kappa, \lambda, \mathbf{k}, \mathbf{l}) = \prod_{q=1}^K \kappa_q^{N_q^{\mathbf{k}}} \prod_{r=1}^L \lambda_r^{M_r^{\mathbf{l}}} \prod_{i=1}^N \prod_{j=1}^M g(Y_{i,j} | \theta_{k_i, l_j}) \quad (6.5)$$

with respect to θ , κ , λ , $\mathbf{k} \in \mathcal{K}$, and $\mathbf{l} \in \mathcal{L}$. As the parameter space contains discrete parameters \mathbf{k} and \mathbf{l} , standard asymptotic theory for maximum likelihood parameter estimation does not apply. Govaert and Nadif (2008) also consider the optimization of an approximation to the likelihood function (6.4). This approximation is based on the assumption that the two classifications (that is, the classification of the rows and the classification of the columns) are independent.

We solve the aforementioned problems by considering a Bayesian approach. This approach has several advantages. First, we do not have to rely on asymptotic theory for inference. We can use the posterior distribution to do inference on the model parameters. In addition, it turns out that we do not need to evaluate the the likelihood specification (6.4) to obtain posterior results. Posterior results can easily be obtained using a Markov Chain Monte Carlo (MCMC) sampler with data augmentation (Tanner & Wong, 1987). Data augmentation means that the latent variables (that is, \mathbf{k} and \mathbf{l}) are simulated alongside the model parameters θ , κ , and λ . This amounts to applying the Gibbs sampler to the

complete data likelihood in (6.5). As Tanner and Wong (1987) show, the posterior results for the complete data likelihood function are equal to the posterior results for the likelihood function. As we can rely on the complete data likelihood (6.5) and do not have to consider (6.4), obtaining posterior results is computationally feasible. Furthermore, unlike previous authors (see, for example, Govaert & Nadif, 2003, 2008), we can provide statistical rules for choosing the numbers of segments, which we will do in Section 6.3.2. Finally, our method does not suffer much from computational difficulties when searching the global optimum of the likelihood function. The EM algorithm is known to get stuck in local optima of the likelihood function, which often occurs in local optima with one or more empty segments. Because we rely on MCMC methods, our approach has fewer problems with local optima. Using proper priors, we can avoid solutions with empty segments; see also Dias and Wedel (2004) for similar arguments.

6.3 Posterior Simulator

As discussed previously, we rely on MCMC methods to estimate the posterior distributions of the parameters of the two-mode mixture model. We propose a Gibbs sampler (Geman & Geman, 1984) with data augmentation (Tanner & Wong, 1987), in which we sample the vectors \mathbf{k} and \mathbf{l} alongside the model parameters. This approach allows us to sample from the posterior distributions of the parameters without evaluating the full likelihood function and therefore requires limited computation time. We assume independent priors for the model parameters with density functions $f(\kappa)$, $f(\lambda)$, and $f(\theta)$. In Section 6.3.1, we derive the Gibbs sampler. Methods for choosing the numbers of segments are discussed in Section 6.3.2.

6.3.1 The Gibbs Sampler

In each iteration of the Gibbs sampler, we need to draw the parameters θ , κ , and λ together with the latent variables \mathbf{k} and \mathbf{l} from their full conditional distributions. The MCMC simulation scheme is as follows.

- Draw $\kappa, \lambda | \theta, \mathbf{k}, \mathbf{l}, \mathbf{Y}$
- Draw $\mathbf{k} | \kappa, \lambda, \theta, \mathbf{l}, \mathbf{Y}$
- Draw $\mathbf{l} | \kappa, \lambda, \theta, \mathbf{k}, \mathbf{Y}$
- Draw $\theta | \kappa, \lambda, \mathbf{k}, \mathbf{l}, \mathbf{Y}$

Below, we derive the full conditional posteriors, which are needed for the Gibbs sampler. The Gibbs sampler results in a series of draws from the posterior distributions of the parameters θ , κ , and λ . These draws can be used to compute posterior means, posterior standard deviations, and highest posterior density regions. Because we use data augmentation, we also obtain draws from the posterior distributions of \mathbf{k} and \mathbf{l} . This enables us to

compute the posterior distributions of each row of data and each column of data over the segments. We can store the posterior distributions in matrices \mathbf{Q} and \mathbf{R} , where \mathbf{Q} is of size $(N \times K)$, and \mathbf{R} is of size $(M \times L)$. Each row of \mathbf{Q} contains the posterior distribution of a row of data over the K possible row segments, and each row of \mathbf{R} contains the posterior distribution of a column of data over the L possible column segments.

Sampling of κ and λ

The full conditional density of κ is given by

$$\begin{aligned} f(\kappa|\theta, \lambda, \mathbf{k}, \mathbf{l}, \mathbf{Y}) &\propto l(\theta, \kappa, \lambda) f(\kappa) \\ &\propto \prod_{q=1}^K \kappa_q^{\sum_{i=1}^N I(k_i=q)} f(\kappa), \end{aligned} \quad (6.6)$$

where $f(\kappa)$ is the prior density of κ , and $I(\cdot)$ is an indicator function that equals 1 if the argument is true and 0 otherwise. The first part of (6.6) is the kernel of a Dirichlet distribution (see, for example, Frühwirth-Schnatter, 2006). If we specify a Dirichlet (d_1, d_2, \dots, d_K) prior distribution for κ , the full conditional posterior is also a Dirichlet distribution with parameters $\sum_{i=1}^N I(k_i = 1) + d_1, \sum_{i=1}^N I(k_i = 2) + d_2, \dots, \sum_{i=1}^N I(k_i = K) + d_K$.

If we take a Dirichlet (d_1, d_2, \dots, d_L) prior for λ , the λ parameters can be sampled in exactly the same way. The full conditional posterior density is now given by

$$f(\lambda|\theta, \kappa, \mathbf{k}, \mathbf{l}, \mathbf{Y}) \propto \prod_{r=1}^L \lambda_r^{\sum_{j=1}^M I(l_j=r)} f(\lambda), \quad (6.7)$$

where $f(\lambda)$ denotes the prior density. Hence, we can sample λ from a Dirichlet distribution with parameters $\sum_{j=1}^M I(l_j = 1) + d_1, \sum_{j=1}^M I(l_j = 2) + d_2, \dots, \sum_{j=1}^M I(l_j = L) + d_L$.

Sampling of \mathbf{k} and \mathbf{l}

We sample each element of \mathbf{k} and \mathbf{l} separately. The full conditional density of k_i is given by

$$\begin{aligned} p(k_i|\theta, \kappa, \lambda, \mathbf{k}_{-i}, \mathbf{l}, \mathbf{Y}) &\propto p(\mathbf{Y}_i|k_i, \theta, \kappa, \lambda, \mathbf{l}) \\ &\propto \kappa_{k_i} \prod_{j=1}^M g(Y_{i,j}|\theta_{k_i, l_j}), \end{aligned} \quad (6.8)$$

for $k_i = 1, \dots, K$, where \mathbf{k}_{-i} denotes \mathbf{k} without k_i , and \mathbf{Y}_i denotes the i th row of \mathbf{Y} . Hence, k_i can be sampled from a multinomial distribution. In a similar way, we can derive the full conditional density of l_j , which equals

$$p(l_j|\theta, \kappa, \lambda, \mathbf{k}, \mathbf{l}_{-j}, \mathbf{Y}) \propto \lambda_{l_j} \prod_{i=1}^N g(Y_{i,j}|\theta_{k_i, l_j}), \quad (6.9)$$

where \mathbf{l}_{-j} denotes \mathbf{l} without l_j . We can thus sample l_j from a multinomial distribution.

Sampling of θ

The sampling of the parameters θ depends on the specification of $g(Y_{i,j}|\theta_{q,r})$. In principle, the distribution $g(Y_{i,j}|\theta_{q,r})$ can be chosen at will. However, sampling from the posterior is much easier if $g(Y_{i,j}|\theta_{q,r})$ has a natural conjugate prior. If a natural conjugate prior is used, the posterior distribution is in the same family as the prior distribution (see Gelman, Carlin, Stern, & Rubin, 2003, for an overview of the natural conjugate priors). A natural conjugate prior exists if the distribution $g(Y_{i,j}|\theta_{q,r})$ is chosen from an exponential family of distributions. A large number of distributions belong to an exponential family, including the binomial, gamma, normal, and Poisson distributions. Therefore, our Bayesian approach to two-mode clustering can easily be used with many distributions used in practice. Below, we discuss the prior and posterior distributions of the parameters if $Y_{i,j}$ has a Bernoulli or a normal distribution.

Example 1: Sampling of θ if \mathbf{Y} Has a Bernoulli Distribution

If $Y_{i,j}$ is a binary random variable with a Bernoulli distribution, with probability $p_{q,r}$ for an element in row segment q and column segment r , then

$$g(Y_{i,j}|\theta_{q,r}) = Y_{i,j}^{p_{q,r}}(1 - Y_{i,j})^{1-p_{q,r}}. \quad (6.10)$$

Let \mathbf{P} denote the $(K \times L)$ matrix containing these probabilities for each combination of a row segment and a column segment, so that $\theta = \mathbf{P}$.

To sample $p_{q,r}$, we need to derive its full conditional density, which is given by

$$\begin{aligned} f(p_{q,r}|\mathbf{P}_{-q,r}, \kappa, \lambda, \mathbf{k}, \mathbf{l}, \mathbf{Y}) & \\ \propto f(\mathbf{Y}|\mathbf{P}_{q,r}, \mathbf{P}_{-q,r}, \kappa, \lambda, \mathbf{k}, \mathbf{l})f(p_{q,r}) & \\ \propto \prod_{i \in \mathcal{Q}} \prod_{j \in \mathcal{R}} p_{q,r}^{Y_{i,j}}(1 - p_{q,r})^{1-Y_{i,j}} f(p_{q,r}) & \\ \propto p_{q,r}^{\sum_{i=1}^N \sum_{j=1}^M I(k_i=q)I(l_j=r)Y_{i,j}} (1 - p_{q,r})^{\sum_{i=1}^N \sum_{j=1}^M I(k_i=q)I(l_j=r)(1-Y_{i,j})} f(p_{q,r}), & \quad (6.11) \end{aligned}$$

where \mathcal{Q} is the set containing all rows that belong to segment q , \mathcal{R} contains all columns that belong to segment r , $\mathbf{P}_{-q,r}$ denotes \mathbf{P} without $p_{q,r}$, and $f(p_{q,r})$ denotes the prior density of $p_{q,r}$. The first part of (6.11) is the kernel of a beta distribution. If we specify a Beta(b_1, b_2) prior distribution, the full conditional posterior distribution is also a beta distribution; the parameters of this posterior distribution are $\sum_{i=1}^N \sum_{j=1}^M I(k_i = q)I(l_j = r)Y_{i,j} + b_1$ and $\sum_{i=1}^N \sum_{j=1}^M I(k_i = q)I(l_j = r)(1 - Y_{i,j}) + b_2$.

Example 2: Sampling of θ if \mathbf{Y} Has a Normal Distribution

If $Y_{i,j}$ is a normally distributed variable, with mean $\mu_{q,r}$ and variance $\sigma_{q,r}^2$ in row segment q and column segment r , then

$$g(Y_{i,j}|\theta_{q,r}) = \frac{1}{\sqrt{2\pi\sigma_{q,r}^2}} \exp \left\{ -\frac{1}{2} \frac{(Y_{i,j} - \mu_{q,r})^2}{\sigma_{q,r}^2} \right\}. \quad (6.12)$$

Let μ and Σ denote the $(K \times L)$ matrices containing the means and variances, respectively, for each combination of a row segment and a column segment; hence, $\theta = \{\mu, \Sigma\}$.

To sample $\mu_{q,r}$, we need to derive its full conditional distribution, which is given by

$$\begin{aligned}
& f(\mu_{q,r} | \mu_{-q,r}, \Sigma, \kappa, \lambda, \mathbf{k}, \mathbf{l}, \mathbf{Y}) \\
& \propto f(\mathbf{Y} | \mu_{q,r}, \mu_{-q,r}, \Sigma, \kappa, \lambda, \mathbf{k}, \mathbf{l}) f(\mu_{q,r}) \\
& \propto \exp \left[-\frac{\sum_{i \in \mathcal{Q}} \sum_{j \in \mathcal{R}} (Y_{i,j} - \mu_{q,r})^2}{2\sigma_{q,r}^2} \right] f(\mu_{q,r}) \\
& \propto \exp \left[-\frac{(\mu_{q,r} - 1/N_{\mathbf{k},\mathbf{l}}^{q,r} \sum_{i \in \mathcal{Q}} \sum_{j \in \mathcal{R}} Y_{i,j})^2}{2\sigma_{q,r}^2 / N_{\mathbf{k},\mathbf{l}}^{q,r}} \right] f(\mu_{q,r}), \tag{6.13}
\end{aligned}$$

where $\mu_{-q,r}$ denotes μ without $\mu_{q,r}$, $f(\mu_{q,r})$ denotes the prior of $\mu_{q,r}$, and the number of elements that are both in row segment q and column segment r according to segmentations \mathbf{k} and \mathbf{l} is denoted by $N_{\mathbf{k},\mathbf{l}}^{q,r} = \sum_{i=1}^N \sum_{j=1}^M I(k_i = q) I(l_j = r)$. As some segments may become empty in one of the iterations, we use a proper prior. In this case, the first part of (6.13) is the kernel of a normal distribution, so that using a normal prior is an obvious choice. Let the prior distribution have mean μ_0 and variance σ_0^2 . Then, the full conditional posterior distribution is also normal with

$$f(\mu_{q,r} | \mu_{-q,r}, \Sigma, \kappa, \lambda, \mathbf{k}, \mathbf{l}, \mathbf{Y}) \sim \mathcal{N} \left(\frac{\sigma_0^{-2}}{\sigma_0^{-2} + s^{-2}} \mu_0 + \frac{s^{-2}}{\sigma_0^{-2} + s^{-2}} \bar{\mu}, (\sigma_0^{-2} + s^{-2})^{-1} \right), \tag{6.14}$$

where $\bar{\mu} = \sum_{i \in \mathcal{Q}} \sum_{j \in \mathcal{R}} Y_{i,j} / N_{\mathbf{k},\mathbf{l}}^{q,r}$, the sample average within the cluster, and $s^2 = \sigma_{q,r}^2 / N_{\mathbf{k},\mathbf{l}}^{q,r}$.

The full conditional distribution of $\sigma_{q,r}^2$ is given by

$$\begin{aligned}
& f(\sigma_{q,r}^2 | \mu, \Sigma_{-q,r}, \kappa, \lambda, \mathbf{k}, \mathbf{l}, \mathbf{Y}) \\
& \propto f(\mathbf{Y} | \mu, \sigma_{q,r}^2, \Sigma_{-q,r}, \kappa, \lambda, \mathbf{k}, \mathbf{l}) f(\sigma_{q,r}^2) \\
& \propto (\sigma_{q,r}^2)^{N_{\mathbf{k},\mathbf{l}}^{q,r}/2} \exp \left[-\frac{\sum_{i \in \mathcal{Q}} \sum_{j \in \mathcal{R}} (Y_{i,j} - \mu_{q,r})^2}{2\sigma_{q,r}^2} \right] f(\sigma_{q,r}^2), \tag{6.15}
\end{aligned}$$

where $\Sigma_{-q,r}$ denotes Σ without $\sigma_{q,r}^2$, and $f(\sigma_{q,r}^2)$ denotes the prior of $\sigma_{q,r}^2$. The first part of (6.15) is the kernel of an inverted gamma-2 distribution. Again, using a proper prior seems advisable. If we specify an inverted gamma-2 prior with parameters g_1 and g_2 , the full conditional posterior has an inverted gamma-2 distribution with parameters $N_{\mathbf{k},\mathbf{l}}^{q,r} + g_1$ and $\sum_{i \in \mathcal{Q}} \sum_{j \in \mathcal{R}} (Y_{i,j} - \mu_{q,r})^2 + g_2$. The prior essentially adds g_1 observations with an average variance of g_2/g_1 to each combination of a row segment and a column segment.

6.3.2 Selecting the Numbers of Segments

The usual way to determine the numbers of clusters in a finite mixture model in a frequentist framework is to use information criteria such as AIC, AIC-3, BIC, and CAIC

(see, for example, Fraley & Raftery, 1998; Andrews & Currim, 2003). The reason for this is that standard tests for determining the optimal number of classes in latent-class models are not valid due to the Davies (1977) problem. Within a Bayesian framework, we can avoid this problem by computing Bayes factors (see, for example, Berger, 1985; Kass & Raftery, 1995; Han & Carlin, 2001). Unlike hypothesis tests, Bayes factors can be used to compare several possibly nonnested models simultaneously; Bayes factors naturally penalize complex models. The Bayes factor for comparing Model 1 with Model 2 is defined as

$$B_{21} = \frac{f(\mathbf{Y}|M_2)}{f(\mathbf{Y}|M_1)}, \quad (6.16)$$

where $f(\mathbf{Y}|M_i)$ denotes the marginal likelihood of model M_i . The marginal likelihood is defined as the expected value of the likelihood function with respect to the prior (see, for example, Gelman et al., 2003).

Computing the value of the marginal likelihood is not an easy task. Theoretically, its value can be estimated by averaging the likelihood function over draws from the prior distribution. If the support of the prior distribution does not completely match with the support of the likelihood function, the resulting estimate will be very poor. Another strategy is to use the harmonic mean estimator of Newton and Raftery (1994). However, this estimator can be quite unstable. In this chapter, we estimate the marginal likelihood using the fourth estimator proposed by Newton and Raftery (1994, p. 22), which is also used by DeSarbo, Fong, Liechty, and Saxton (2004) in a similar model. This estimator uses importance sampling to compute the marginal likelihood value. The importance sampling function is a mixture of the prior and the posterior distribution with mixing proportion δ . Using the fact that the marginal likelihood is the expected value of the likelihood function with respect to the prior, it can be shown that the marginal likelihood $f(\mathbf{Y})$ can be estimated using the iterative formula

$$\widehat{f(\mathbf{Y})} = \frac{\delta m / (1 - \delta) + \sum_{i=1}^m (f(\mathbf{Y}|\vartheta^{(i)}) / (\delta \widehat{f(\mathbf{Y})} + (1 - \delta) f(\mathbf{Y}|\vartheta^{(i)}))}{\delta m / (1 - \delta) \widehat{f(\mathbf{Y})} + \sum_{i=1}^m (\delta \widehat{f(\mathbf{Y})} + (1 - \delta) f(\mathbf{Y}|\vartheta^{(i)}))^{-1}}, \quad (6.17)$$

where m denotes the number of draws $\vartheta^{(i)}$ from the posterior distribution; for notational convenience, we drop the model indicator M_i . To apply this formula, we need to choose the value δ ; Newton and Raftery (1994) recommend using a low value of δ , which we set to 0.001 in our application below.

Obtaining an accurate value of the marginal likelihood for any moderately sophisticated model tends to be hard, as was noted by Han and Carlin (2001). Therefore, we also propose a simpler alternative method to choose the numbers of segments, based on information criteria. Simulations in Andrews and Currim (2003) suggest that the AIC-3 of Bozdogan (1994) performs well as a criterion for selecting numbers of segments. To evaluate the AIC-3, we need the maximum likelihood value and the number of parameters. To compute the maximum likelihood value, we take the highest value of the likelihood function (6.5) across the sampled parameters.

Determining the number of parameters is not as straightforward as usual. The parameters θ , κ , and λ contain wKL , $K - 1$, and $L - 1$ parameters, respectively, where w denotes the number of parameters in θ per combination of a row segment and a column segment. Although \mathbf{k} and \mathbf{l} contain the same numbers of parameters for all numbers of latent classes, the number of possible values for each parameter increases. We can think of \mathbf{k} as representing an $(N \times K)$ indicator matrix, where each row indicates to which segment an object belongs. This means that \mathbf{k} and \mathbf{l} represent $N(K - 1)$ and $M(L - 1)$ free parameters, respectively. Hence, the effective total number of parameters is $wKL + NK + ML + K + L - M - N - 2$.

6.4 Application 1: Supreme Court Voting Data

We apply the latent-class two-mode clustering model to two empirical data sets. The first data set, which is discussed in this section, is the Supreme Court voting data of Doreian et al. (2004). We use this data set to compare the results of our approach with the results of previous authors, and we discuss this data set relatively briefly. The second data set will be analyzed in greater detail in the next section. The Supreme Court voting data set comprises the decisions of the nine Justices of the United States Supreme Court on 26 important issues. The data are displayed in Table 6.1. In this table, a 1 reflects that the Justice voted with the majority, and a 0 means that the Justice voted with the minority.

To describe the votes, we use a Bernoulli distribution with a Beta(1,1) prior for the probability, which is equivalent to a uniform prior on (0,1). Furthermore, we use an uninformative Dirichlet(1,1,...,1) prior for both κ and λ . To determine the optimal numbers of segments, we compute the marginal likelihoods for several values of K and L , based on an MCMC chain of 100,000 draws for each combination of K and L . Table 6.2 displays the values of $\ln f(\mathbf{Y})$ for each combination of $K = 1, \dots, 6$ rows segments and $L = 1, \dots, 6$ column segments. The highest marginal likelihood is achieved with $K = 2$ segments for the issues and $L = 3$ segments for the Justices.

To analyze the results, it is possible to weight the results with different numbers of segments according to the posterior model probabilities that follow from the marginal likelihoods. However, we find it more convenient to consider the results for only one value of K and L . Therefore, we focus on the solution with the highest marginal likelihood value, that is, $K = 2$ segments of issues and $L = 3$ segments of Justices. Note that we find fewer segments than Doreian et al. (2004), who applied blockmodeling to this data set and found 7 clusters for the issues and 4 clusters for the Justices, and Brusco and Steinley (2006), who found 5 clusters for the issues and 3 clusters for the Justices.

We experienced label switching in our MCMC sampler. Two of the segments of Justices switched places twice in the MCMC chain of 100,000 draws. However, we could easily identify where these switchings occurred. As suggested by Geweke (2007), we solved the label switching problem by sorting the draws in an appropriate way.

The posterior means and standard deviations of \mathbf{P} , κ , and λ are shown in Table 6.3. Tables 6.4 and 6.5 show the marginal posterior distributions of the issues and the Justices

Table 6.1: The Supreme Court Voting Data

Issue	Supreme Court Justice									Issue	Supreme Court Justice								
	Br	Gi	So	St	OC	Ke	Re	Sc	Th		Br	Gi	So	St	OC	Ke	Re	Sc	Th
2000 Presidential Election	0	0	0	0	1	1	1	1	1	Title VI Disabilities	0	0	0	0	1	1	1	1	1
Illegal Search 1	1	1	1	1	1	1	0	0	0	PGA vs. Handicapped	1	1	1	1	1	1	1	0	0
Illegal Search 2	1	1	1	1	1	1	0	0	0	Immigration Jurisdiction	1	1	1	1	0	1	0	0	0
Illegal Search 3	1	1	1	0	0	0	0	1	1	Deporting Criminal Aliens	1	1	1	1	1	0	0	0	0
Seat Belts	0	0	1	0	0	1	1	1	1	Detaining Criminal Aliens	1	1	1	1	0	1	0	0	0
Stay of Execution	1	1	1	1	1	1	0	0	0	Citizenship	0	0	0	1	0	1	1	1	1
Federalism	0	0	0	0	1	1	1	1	1	Legal Aid for the Poor	1	1	1	1	0	1	0	0	0
Clean Air Act	1	1	1	1	1	1	1	1	1	Privacy	1	1	1	1	1	0	0	0	0
Clean Water	0	0	0	0	1	1	1	1	1	Free Speech	1	0	0	0	1	1	1	1	1
Cannabis for Health	0	1	1	1	1	1	1	1	1	Campaign Finance	1	1	1	1	1	0	0	0	0
United Foods	0	0	1	1	0	1	1	1	1	Tobacco Ads	0	0	0	0	1	1	1	1	1
New York Times Copyright	0	1	1	0	1	1	1	1	1	Labor Rights	0	0	0	0	1	1	1	1	1
Voting Rights	1	1	1	1	1	1	0	0	0	Property Rights	0	0	0	0	1	1	1	1	1

Table 6.2: Log marginal likelihoods for the Supreme Court Voting Data

Segments of issues	Segments of Justices					
	1	2	3	4	5	6
1	-155.56	-160.96	-164.17	-166.28	-168.13	-169.96
2	-172.98	-117.69	-109.98	-111.00	-112.27	-114.02
3	-183.29	-124.17	-112.17	-113.83	-114.63	-116.56
4	-189.60	-129.92	-115.57	-117.72	-119.27	-118.61
5	-194.88	-133.45	-118.78	-120.00	-121.50	-121.51
6	-199.67	-137.27	-122.29	-122.66	-125.84	-126.28

over the segments. We find that Justices *Ginsburg*, *Stevens*, *Breyer*, and *Souter* constitute the liberal wing (that is, the left wing) of the Supreme Court. The Court's moderate wing comprises Justices *O'Connor* and *Kennedy*, and the conservative wing (that is, the right wing) consists of Justices *Rehnquist*, *Scalia*, and *Thomas*. The segments of the issues consist of issues that resulted in liberal decisions (segment 1) and issues that resulted in conservative decisions (segment 2). We find strong partisan tendencies in the Supreme Court: liberal Justices support liberal decisions in 97% of the cases, and conservative Justices also support conservative decisions with a 97% probability. The liberal Justices sometimes (in 26% of the cases) vote for a conservative decision, whereas conservative Justices seldom support a liberal decision. Because of their centrist position in the court, the moderate Justices usually are in the majority. However, the moderate Justices are slightly more likely to support conservative decisions than liberal decisions. In general, the uncertainty in these classifications is low, especially given the relatively small size of the data set. The Justices and almost all issues can be assigned to one segment with a posterior probability close to 1.

The segmentation of the Justices, as displayed in Table 6.4, resembles the one found by Doreian et al. (2004), who divide the Justices into four segments. The segmentation of the issues deviates more from the solution of Doreian et al. (2004), who find 7 segments for the issues. Brusco and Steinley (2006) also find more segments of Justices and issues than the numbers of segments found here. We believe that the methods used by these authors can overestimate the numbers of segments in the data.

Table 6.3: Mean posterior results, with posterior standard deviations in parentheses, for $K = 2$ and $L = 3$ in the Supreme Court Data.

Segment of issues Interpretation	Segment of Justices			Posterior segment size
	1 liberal	2 moderate	3 conservative	
1 (liberal majority)	0.97 (0.03)	0.68 (0.10)	0.10 (0.07)	0.46 (0.10)
2 (conservative majority)	0.26 (0.07)	0.84 (0.07)	0.97 (0.03)	0.54 (0.10)
Posterior segment size	0.42 (0.14)	0.25 (0.12)	0.33 (0.13)	

Table 6.4: Marginal posterior distribution of the Justices over the segments.

Justice	1	2	3
Breyer	1.00	0.00	0.00
Ginsburg	1.00	0.00	0.00
Stevens	1.00	0.00	0.00
Souter	1.00	0.00	0.00
O'Connor	0.00	1.00	0.00
Kennedy	0.00	0.98	0.02
Rehnquist	0.00	0.00	1.00
Thomas	0.00	0.00	1.00
Scalia	0.00	0.00	1.00
Interpretation	liberal	moderate	conservative

Table 6.5: Marginal posterior distribution of the issues over the segments.

Issue \ Segment	1	2	Issue \ Segment	1	2
2000 Presidential Election	0.00	1.00	Clean Air Act	0.20	0.80
Federalism	0.00	1.00	Illegal Search 3	0.57	0.43
Clean Water	0.00	1.00	PGA vs. Handicapped	1.00	0.00
Title VI Disabilities	0.00	1.00	Illegal Search 1	1.00	0.00
Tobacco Ads	0.00	1.00	Illegal Search 2	1.00	0.00
Labor Rights	0.00	1.00	Stay of Execution	1.00	0.00
Property Rights	0.00	1.00	Privacy	1.00	0.00
Citizenship	0.00	1.00	Immigration Jurisdiction	1.00	0.00
Free Speech	0.00	1.00	Detaining Criminal Aliens	1.00	0.00
Seat Belts	0.00	1.00	Legal Aid for the Poor	1.00	0.00
United Foods	0.00	1.00	Voting Rights	1.00	0.00
New York Times Copyright	0.00	1.00	Deporting Criminal Aliens	1.00	0.00
Cannabis for Health	0.01	0.99	Campaign Finance	1.00	0.00
Majority	liberal	conservative	Majority	liberal	conservative

6.5 Application 2: Roll Call Voting Data

6.5.1 Data

To apply our method to a larger data set, we consider the voting behavior of the entire United States House of Representatives. The details of each roll call vote of the United States congress are published on the website <http://www.GovTrack.us>. We gathered data on all roll call votes from the House of Representatives in 2007. We only use data on votes that are related to a bill. We thus obtain data on 766 roll call votes from 427 members of the House of Representatives in 2007. There are four possible types of votes: *yea*, *nay*, *no vote*, and *present*. A *no vote* means that the representative was absent at the moment of voting; this is the case for 3.5% of the observations. A *present* vote means that the representative is present, but votes neither *yea* nor *nay*, which happens only 143 times (0.04%).

We did not recode our data in such a way that voting with the majority or voting with the Democrats is always recorded as a 1 in the data matrix. Such recodings have been done in the past, for example with the Supreme Court voting data from Section 6.4. The main argument for recoding the data is that each bill is equivalent to a bill for which each representative votes the opposite, in terms of the information that the votes of the representatives provide. However, this argument does not take into account the fact that there is a status quo, from which a bill deviates. We find it is interesting to see in which direction a bill deviates from the status quo; this information would be lost if we recoded the votes.

We collected some additional information on the representatives from GovTrack.us. We have data on their party membership, gender, age on January 1st 2007, and state

Table 6.6: Sample means of the individual characteristics for the whole House, Democrats, and Republicans.

	House	Democrats	Republicans
Size	1.00	0.54	0.46
Democrat	0.54	1.00	0.00
Female	0.16	0.20	0.10
Age	55.73	56.54	54.78
<i>Region</i>			
Pacific	0.16	0.19	0.12
Mountain	0.07	0.05	0.09
West North Central	0.07	0.07	0.08
East North Central	0.15	0.14	0.17
Middle Atlantic	0.14	0.18	0.10
New England	0.05	0.09	0.01
West South Central	0.11	0.08	0.15
East south Central	0.06	0.05	0.08
South Atlantic	0.18	0.15	0.22

from which they were elected. Table 6.6 shows the means for these variables for the entire House of Representatives and for the Democrats and Republicans separately. In 2007, the Democrats had a majority in the House of 53.9%, and there were no third-party or independent representatives. There is a fairly large difference in the share of female representatives between the Democrats (20.4%) and the Republicans (10.1%). The age is about the same for representatives from both parties. We aggregated the representatives' home states into nine regions.

Table 6.7: The numbers of bills prepared by each House Committee

Administration	14	Intelligence (Permanent Select)	15
Agriculture	18	Judiciary	51
Appropriations	291	Natural Resources	50
Armed Services	43	Oversight and Government Reform	57
Budget	14	Rules	9
Education and Labor	51	Science and Technology	43
Energy and Commerce	44	Small Business	26
Financial Services	98	Transportation and Infrastructure	69
Foreign Affairs	44	Veterans' Affairs	15
Homeland Security	55	Ways and Means	42

We also collected more information on the bills. Before a bill comes to a vote in the House of Representatives, it is prepared by at least one of the 20 House Committees. Table 6.7 shows the committees and how many bills they prepared. The committee that handles the largest number of bills is *Appropriations*, which controls the disbursement of funds. The *Rules* committee influences what is discussed and voted upon; this committee is not primarily concerned with bills and only prepared nine of them. Most other committees deal with specific topics. The committee(s) that prepared a bill is an indication of the subject of the bill. Having this information should allow us to interpret the segments of bills. Identifying the segments of bills may help us understand the segments of representatives in a better way, as we know what types of bills they support and oppose.

Roll call votes have analyzed before. Poole and Rosenthal (1991), Heckman and Snyder Jr. (1997), and Nelson (2002) try to estimate latent preferences of representatives, based on their voting behavior. De Leeuw (2006) plots the relative positions of representatives into a two-dimensional space. The paper that most closely resembles the analysis in this chapter is Hartigan (2000), who clusters the members of the United States Senate, as well as the bills on which they vote. However, Hartigan (2000) does not cluster the two dimensions simultaneously, but alternates between clustering one dimension conditional on the segmentation of the other dimension, until convergence.

6.5.2 Parameter Estimates

We apply the latent-class two-mode clustering model to the roll call voting data. We assign a 1 to *yea* votes and a 0 to *nay* votes; we treat the response options *no vote*

and *present* as missing observations. Again, we describe the individual votes using a Bernoulli distribution with a Beta(1, 1) prior for the probability. Furthermore, we use an uninformative Dirichlet(1, 1, ..., 1) prior for both κ and λ .

To determine the numbers of segments, we now opt for the AIC-3 criterion as described in Section 6.3.2. To prevent the Gibbs sampler from getting stuck in a local optimum of the likelihood function, we sample 10 sets of 10 MCMC chains, and each of the 100 MCMC chains has length 200. For each set, the MCMC chain that attains the highest likelihood value is chosen, and this MCMC chain is allowed to run for an additional 3,000 iterations. The highest likelihood value that is attained during these 3,000 iterations over all sets of MCMC chains is then used as the final maximum likelihood value. This likelihood value serves as input for the AIC-3 information criterion.

Table 6.8: AIC-3 values for $K = 2, \dots, 12$ segments of bills and $L = 2, \dots, 10$ segments of representatives.

Segments of bills	Segments of representatives								
	2	3	4	5	6	7	8	9	10
2	251,889	245,490	242,752	242,630	243,333	244,062	245,130	246,267	247,433
3	206,373	197,752	193,995	193,132	193,360	193,966	194,899	195,951	197,020
4	191,907	181,359	177,459	176,081	176,193	176,556	177,376	178,328	179,349
5	186,493	175,547	170,848	169,409	169,370	169,698	170,476	171,396	172,353
6	181,998	170,774	165,959	164,315	164,261	164,480	165,229	166,115	167,061
7	179,889	168,300	163,112	161,363	161,263	161,449	162,178	163,044	163,920
8	179,620	167,790	162,484	160,722	160,516	160,698	161,421	162,244	163,136
9	179,565	167,459	162,080	160,182	160,015	160,056	160,745	161,567	162,449
10	179,531	167,209	161,744	159,715	<i>159,530</i>	159,553	160,194	161,011	161,837
11	180,904	168,414	162,822	160,696	160,443	160,346	160,986	161,810	162,621
12	182,500	169,944	164,292	162,167	161,802	161,671	162,248	163,065	163,854

Table 6.8 displays the AIC-3 values. The lowest AIC-3 value is attained with $K = 10$ segments of bills and $L = 6$ segments of representatives; the corresponding log-likelihood value is $-66,108.70$. For these numbers of segments, we sample an additional 100,000 iterations from the chain that had the highest likelihood value. Due to the large size of the data set, we have no problems with label switching. In the remainder of this section, we present and interpret the results for this model specification.

Table 6.9 shows the posterior means and standard deviations for \mathbf{P} , κ , and λ . The first thing to note is that, except for segments (of bills) 9 and 10, the posterior means of the *yea* voting probabilities are monotonously increasing or decreasing in each row. For segments 9 and 10, there are only deviations from monotonicity in segment of representatives 6. These results imply that the political preferences in the House are one-dimensional. Bills from segment 2 are approved more or less unanimously, and bills from segments 7 and 9 are also widely supported. Bills from other segments seem to be backed only by representatives from either segments 1 through 3 or segments 4 through 6. In the next subsection, we show that these segments mainly contain Democrats and Republicans, respectively.

Figure 6.1: Graphical Representation of Voting Data Set Before (Upper Panel) and After (Lower Panel) Reordering of Rows and Columns.

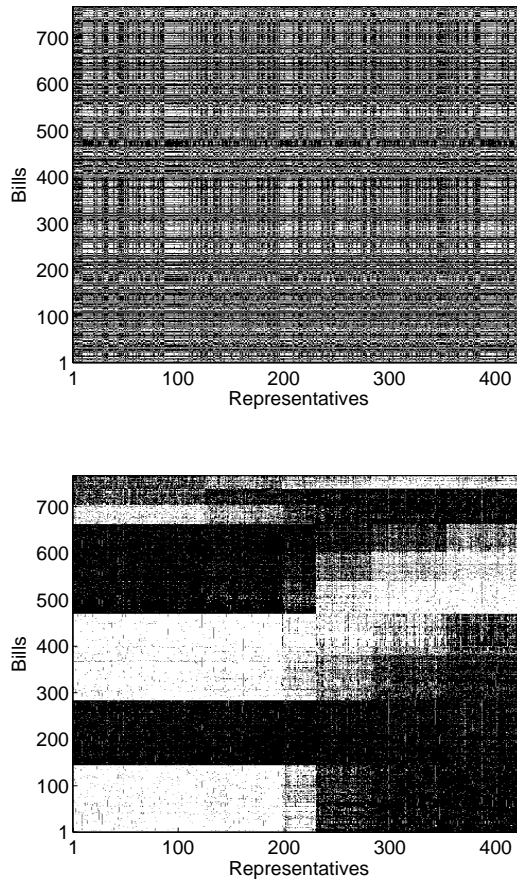


Table 6.9: Posterior means and posterior standard deviations in parentheses of \mathbf{P} , κ , and λ .

Segment of bills	Segment of representatives						κ
	1	2	3	4	5	6	
1	0.00 (0.00)	0.01 (0.00)	0.20 (0.01)	0.87 (0.01)	0.98 (0.00)	0.99 (0.00)	0.19 (0.01)
2	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	0.98 (0.00)	0.18 (0.01)
3	0.01 (0.00)	0.03 (0.00)	0.17 (0.01)	0.45 (0.01)	0.83 (0.01)	0.93 (0.00)	0.13 (0.01)
4	0.01 (0.00)	0.01 (0.00)	0.07 (0.01)	0.13 (0.01)	0.30 (0.01)	0.79 (0.01)	0.12 (0.01)
5	0.99 (0.00)	0.97 (0.00)	0.84 (0.01)	0.14 (0.01)	0.02 (0.00)	0.02 (0.00)	0.09 (0.01)
6	1.00 (0.00)	0.99 (0.00)	0.95 (0.01)	0.65 (0.01)	0.25 (0.01)	0.10 (0.01)	0.08 (0.01)
7	1.00 (0.00)	1.00 (0.00)	0.99 (0.00)	0.95 (0.00)	0.83 (0.01)	0.48 (0.01)	0.08 (0.01)
8	0.11 (0.01)	0.41 (0.01)	0.78 (0.01)	0.93 (0.01)	0.97 (0.00)	0.98 (0.00)	0.05 (0.01)
9	0.71 (0.01)	0.91 (0.01)	0.95 (0.01)	0.97 (0.00)	0.98 (0.00)	0.97 (0.00)	0.05 (0.01)
10	0.74 (0.01)	0.49 (0.02)	0.34 (0.02)	0.25 (0.01)	0.09 (0.01)	0.15 (0.01)	0.04 (0.01)
λ	0.29 (0.02)	0.17 (0.02)	0.08 (0.01)	0.12 (0.02)	0.17 (0.02)	0.17 (0.02)	

To show the effectiveness of our two-mode clustering method, we show graphical representations of the roll call voting data set before and after reordering the rows and columns according to their segments in Figure 6.1. For this reordering, we used the segmentation \mathbf{k} and \mathbf{l} that yielded the highest likelihood value. Before reordering the rows and columns, it is already apparent that some structure exists in the data; after reordering, the nature of the block structure becomes clear.

6.5.3 Interpretation of Segments

For each row (bill) and for each column (representative), we compute the marginal posterior distribution over the segments. This allow us to compute the means of the explanatory variables within each segment. Table 6.10 shows the posterior means of the individual characteristics of the representatives for each segment of representatives. The main result is that the first three segments consist of Democrats, and the last three contain Republicans. We know from Table 6.9 that voting behavior is monotonous; therefore, we can interpret segments 1 and 6 as very partisan Democrats and Republicans, respectively. Segments 2 and 5 seem to be typical Democrats and Republicans, respectively, and the representatives

in segments 3 and 4 are relatively moderate. Note that segments 3 and 4 are not completely homogenous, which means that there is a little overlap between these moderate Democrats and Republicans.

Table 6.10: Posterior means of personal variables for each segment of representatives.

Segment of representatives	1	2	3	4	5	6	Total
Democrat	1.00	1.00	0.97	0.02	0.00	0.00	0.54
Female	0.29	0.08	0.16	0.16	0.08	0.08	0.16
Age	59.30	57.44	50.79	57.68	55.19	55.06	55.73
<i>Region 9</i>							
Pacific	0.27	0.12	0.03	0.10	0.14	0.12	0.16
Mountain	0.03	0.06	0.09	0.08	0.03	0.15	0.07
West North Central	0.03	0.12	0.09	0.08	0.07	0.08	0.07
East North Central	0.13	0.14	0.16	0.26	0.14	0.15	0.15
Middle Atlantic	0.22	0.12	0.13	0.27	0.06	0.03	0.14
New England	0.12	0.07	0.00	0.02	0.00	0.00	0.05
West South Central	0.04	0.12	0.15	0.00	0.21	0.19	0.11
East South Central	0.01	0.06	0.16	0.02	0.15	0.04	0.06
South Atlantic	0.14	0.18	0.19	0.18	0.20	0.25	0.18
<i>Region 4</i>							
West	0.30	0.18	0.13	0.18	0.17	0.27	0.22
Midwest	0.17	0.26	0.25	0.34	0.21	0.23	0.23
Northeast	0.34	0.20	0.13	0.29	0.06	0.03	0.19
South	0.19	0.36	0.50	0.20	0.57	0.48	0.36
Segment size	0.29	0.17	0.08	0.12	0.17	0.17	1.00

Further, we can see that there are relatively more women in the left wing. Not only are women more often Democrats than Republicans, but they also seem to be on the left side within their parties. There appear to be no effects of age within the Republican party, but within the Democratic party, the younger representatives seem more moderate than the older ones.

There are also some clear regional patterns. Representatives from states in the West are more extreme in their voting behavior, as there are few representatives from these states that are in the moderate clusters 3 and 4. The representatives from the Pacific are responsible for the left wing, while the right-wing representatives seem to come mainly from the Mountain states. Representatives from the Midwest seem to be more moderate than the national average, though this effect is not very strong. In the Northeast, we find that the Democrats are relatively liberal and that the Republicans are relatively moderate. Finally, in the South, we find that the Democrats are moderate, whereas the Republicans often belong to the most conservative segments.

Table 6.11 contains the posterior means of the committees for the segments of bills. These results are less pronounced than the posterior means for the representatives. For example, the posterior means for segment 1 closely resemble the entire sample (that is, the

Table 6.11: Posterior means of committees for each segment of bills.

Segment of bills	1	2	3	4	5	6	7	8	9	10	Total
Administration	0.01	0.03	0.00	0.00	0.04	0.02	0.00	0.00	0.09	0.03	0.02
Agriculture	0.01	0.01	0.03	0.00	0.03	0.00	0.02	0.02	0.00	0.28	0.02
Appropriations	0.36	0.05	0.66	0.92	0.27	0.24	0.07	0.58	0.37	0.38	0.38
Armed Services	0.06	0.05	0.02	0.00	0.12	0.00	0.05	0.02	0.14	0.27	0.06
Budget	0.03	0.00	0.00	0.00	0.07	0.05	0.02	0.02	0.00	0.00	0.02
Education and Labor	0.07	0.05	0.08	0.00	0.11	0.15	0.09	0.00	0.06	0.10	0.07
Energy and Commerce	0.07	0.11	0.01	0.00	0.07	0.08	0.05	0.00	0.03	0.10	0.06
Financial Services	0.24	0.09	0.15	0.03	0.03	0.24	0.14	0.12	0.03	0.07	0.13
Foreign Affairs	0.05	0.08	0.03	0.00	0.06	0.02	0.09	0.00	0.09	0.35	0.06
Homeland Security	0.08	0.05	0.02	0.02	0.19	0.05	0.05	0.12	0.14	0.13	0.07
Intelligence (Permanent Select)	0.04	0.00	0.00	0.00	0.07	0.02	0.02	0.05	0.03	0.00	0.02
Judiciary	0.04	0.14	0.00	0.00	0.07	0.03	0.16	0.02	0.23	0.07	0.07
Natural Resources	0.08	0.09	0.03	0.00	0.10	0.14	0.05	0.02	0.03	0.10	0.07
Oversight and Government Reform	0.03	0.17	0.02	0.00	0.10	0.07	0.14	0.05	0.09	0.10	0.07
Rules	0.01	0.00	0.00	0.00	0.04	0.03	0.00	0.00	0.06	0.03	0.01
Science and Technology	0.04	0.10	0.04	0.01	0.05	0.02	0.09	0.07	0.09	0.10	0.06
Small Business	0.05	0.05	0.01	0.00	0.01	0.02	0.11	0.00	0.00	0.10	0.03
Transportation and Infrastructure	0.06	0.09	0.05	0.03	0.17	0.18	0.15	0.07	0.06	0.10	0.09
Veterans' Affairs	0.00	0.11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02
Ways and Means	0.06	0.04	0.00	0.01	0.18	0.11	0.05	0.02	0.11	0.00	0.05
Segment size	0.19	0.18	0.13	0.12	0.09	0.08	0.08	0.05	0.05	0.04	1.00

final column in the table), except that segment 1 contains relatively many bills from the Financial Services committee.

Nevertheless, there are some striking results. Bills from the Veterans' Affairs committee all belong to segment 2, which contains bills that receive nearly unanimous support. Transportation and Infrastructure is relatively common in segments 5, 6, 7, and 10, which are all primarily favored by Democrats. Bills from the Judiciary committee can primarily be found in segments 2, 7, and 9. For these segments, the voting is almost unanimously *yea*. Segment 4 almost solely contains bills from the Appropriations committee. Only the hard-line Republicans from segment 6 vote in majority (79%) *yea* for these bills. To a lesser extent, this is also true for bills from segment 3, though there is a little more support for these bills, even from some of the moderate Democrats in segment 3.

6.6 Conclusions

We have developed a Bayesian approach to do inference in a latent-class two-mode clustering model, which has several advantages over frequentist parameter estimation methods. First, our method allows for statistical inference on the model parameters, which is not possible using a maximum likelihood approach. The Bayesian approach also allows us to do statistical inference on the number of segments using marginal likelihoods. An alternative way to select the numbers of segments is to consider information criteria. The third advantage of using Bayesian estimation techniques is that fewer computational problems occur during estimation.

We have applied our model to the Supreme Court voting data set of Doreian et al. (2004). The marginal likelihoods used to determine the optimal number of segments indicate fewer segments than were found in these previous studies. In the second example, we consider roll call votes from the United States House of Representatives in 2007. We detect six segments of representatives and ten segments of bills. Three of the individual segments contain Democrats and the other three segments contain Republicans, though there is a little overlap. There are also clear regional effects on voting behavior.

Another advantage of our approach is that it can easily be extended in several directions. First, it can be adapted for use with data matrices with arbitrary distributions. Although we have only derived posterior samplers for Bernoulli and normally distributed data, it is straightforward to derive posterior samplers for all distributions that belong to an exponential family, as these distributions have natural conjugate priors. Second, our method can be adapted for three-mode data (see Schepers, Van Mechelen, & Ceulemans, 2006). Third, explanatory variables can be added, either with segment-dependent effects or as concomitant variables, that is, variables explaining why a row (or column) belongs to a certain segment (see Dayton & MacReady, 1988).

Chapter 7

Discussion

In this chapter, we first give a short summary of the main findings of each chapter. We then proceed to discuss limitations and some technical issues of the models that have been proposed in this thesis. We also sketch some directions for future research.

7.1 Summary of the Main Findings

The main contribution of this thesis is the introduction of new statistical models that incorporate techniques for segmentation and dimension reduction. Because each chapter of this thesis discusses a different topic, it is hard to give overall conclusions. Instead, we give a brief overview of the main findings of each chapter and how each chapter contributes to the literature.

In Chapter 2, we have developed the optimal scaling of interactions (OSI) model for two-interaction effects of categorical predictor variables in generalized linear models. We find that a one-dimensional version of this model usually yields the best results. Two-dimensional models and models for analyzing three-way interaction effects can also be constructed. However, these versions of the OSI model often do not yield appropriate results, as unique parameter estimates that maximize the likelihood function do not always exist in that case. We have also devised ways to graphically represent the results of the OSI model, based on suitably chosen parameter constraints. We believe that these graphical representations can provide insight into the interaction effects. Furthermore, we have shown the usefulness of such methods using two empirical applications. A final contribution of Chapter 2 is the iterative reparameterization for estimating the standard errors of a set of parameter estimates that are subject to location and scale constraints. We believe that this reparameterization may be of use in future research.

In Chapter 3, we proposed a new model for analyzing response behavior in rating scale surveys. The main contribution is that rating scale responses can be treated as nominal response options, without a significant loss of information. To limit the number of parameters, we incorporated a bilinear parameter structure, which can be used to construct graphical representations of the estimated effects. Another important novelty

is the separation of content and response style using two types of segmentations. We find that our LC-BML model performs well for a data set that contains the List of Values. By applying our model to this data set, we also managed to discover what appear to be novel response styles. For example, we find misresponse (that is, some respondents accidentally interpret the rating scale in the opposite order) even when the survey contains no reverse-keyed items.

In Chapter 4, we have extended the OSI model of Chapter 2 to the area of conjoint analysis. The resulting Conjoint-OSI model is a multiplicative interaction model for analyzing attribute interactions in choice-based conjoint analysis experiments. We believe that this method can be useful, because until now researchers have often estimated only a subset of the two-way interaction effects in conjoint analysis experiments. Fully estimating all two-way interactions typically requires a large and costly design. We show that our model can lead to smaller designs. For an empirical data set, we find that the Conjoint-OSI model outperforms traditional approaches for modeling interactions.

In Chapter 5, we compared various optimization strategies for two-mode partitioning, which is a form of clustering that partitions the rows and the columns of a data matrix simultaneously. We have given an overview of customary optimization strategies for this optimization problem. In addition, we have introduced a new fuzzy steps optimization method. Using a large simulation study, we have compared the performances of these optimization methods. We find that, in the simulation study, the two-mode k -means method performs best on average, followed by the alternating exchanges and fuzzy steps methods. Which of these three methods gives the best performance depends on characteristics of the clustering problem. An empirical application shows the practical applicability of two-mode partitioning.

In Chapter 6, we have developed a new Bayesian approach to estimate the parameters of a latent-class model for the joint clustering of both modes of two-mode data matrices. A Gibbs sampler with data augmentation can be used to obtain posterior results. We have shown how such a Gibbs sampler can be designed for a variety of prior distributions. Our Bayesian approach has several advantages over existing methods. First, we are able to do statistical inference on the model parameters, which would not be possible using frequentist estimation procedures. In addition, the Bayesian approach allows us to provide statistical criteria for determining the optimal numbers of clusters. Finally, our Gibbs sampler has fewer problems with local optima in the likelihood function and empty classes than the EM algorithm used in a frequentist approach. We have applied the Bayesian estimation method of the latent-class two-mode clustering model to two empirical data sets that comprise voting behavior in the United States. Using these data sets, we have shown how the results of the Bayesian approach to two-mode clustering can provide useful insight into the data.

7.2 Limitations and Avenues for Future Research

Throughout this thesis, we have encountered some limitations and technical issues that can occur in the types of models that we have discussed. Below, we give an overview of these issues. For some limitations, solutions may be found in future research; this will be discussed below. We also discuss other ways in which future research may build upon the work that has been done in this thesis.

After proposing a new model, a way to estimate the parameters must be determined; in most chapters of this thesis, the parameters were estimated using the principle of maximum likelihood. Finding the global maximum of the likelihood function is not always trivial for finite mixture models and models that include a bilinear parameter structure. In this thesis, we have mostly relied on general-purpose optimization routines (such as BFGS) for estimating parameters. Such optimization routines start from randomly chosen initial parameter estimates and then iteratively try to improve these estimates until convergence. To increase the probability that such an optimization routine can locate the global maximum of the likelihood function, we have used multiple random starts of the optimization routines. For finite mixture models, the probability of finding the global optimum can also be increased by applying the EM algorithm. We find that these two solution strategies together are usually sufficient to find the global optimum of the likelihood function, as long as the numbers of segments and dimensions are not very large.

To be able to conduct statistical inference on the parameters of a model, the global maximum of the likelihood function must also be unique, that is, the model parameters must be identified. The parameters of the models in this thesis are not identified without imposing suitably chosen parameter restrictions. Finding parameter constraints that ensure that the model parameters are identified is an important part of designing a model. In principle, constrained optimization can be used to optimize the likelihood function under the necessary parameter constraints. However, we have not used this approach in this thesis. Instead, we first estimate the parameters without any constraints and then transform the parameters estimates so that they do satisfy the constraints. An alternative solution is constructing a reparameterization in which no parameter constraints are necessary (see Appendix 2.A). For some multidimensional versions of the OSI model of Chapter 2, we found that the parameter estimates produced by an optimization routine do not always converge to finite values. This may mean that unique parameter estimates that maximize the likelihood function do not always exist. An easy solution to this problem is to use the one-dimensional version of the OSI model, which does not suffer from this problem and works well in practice. However, future research may investigate in greater detail what causes this problem for multidimensional versions of the OSI model and how it can be solved.

Another important issue is determining the appropriate complexity of the model, that is, the number(s) of segments and dimensions. This can only be done after estimating the parameters for varying numbers of segments and/or dimensions. To choose the numbers of segments and dimensions, we have relied on information criteria such as BIC and AIC-3, which penalize the maximum value of the likelihood function based on the number of

parameters in the model. We find that these information criteria have better theoretical underpinnings than the selection criteria that are often used in techniques that are not model-based (such as the scree plots commonly used in principal components analysis). In future research, we could study which information criterion performs best in practice and if there are any good alternatives to using information criteria.

Some of the strategies outlined above for estimating the model parameters and determining the appropriate complexity of a model can lead to a large computation time. This was especially so for the LC-BML model of Chapter 3, where we used 10 random starts for each combination of the number of response style segments, the number of item segments, and the number of dimensions. This strategy was likely successful in determining the optimal parameter estimates, but may lead to a prohibitively large computation time for larger data sets. Future research may investigate how to limit the computation time associated with estimating the model parameters.

A final and important issue is the interpretability of graphical representations based on bilinear decompositions. The possibility to construct graphical representations of the parameter estimates is an important advantage of the bilinear parameter structures used in this thesis. However, some explanation is often necessary to allow the users of a model to determine the exact estimated effects. Future research could focus on ways to make such graphical representations easier to understand and explain.

Above, we have given an overview of some technical issues and pitfalls that can occur when developing a model similar to the ones proposed in this thesis. Although this may seem like a long list of issues, we believe that our models, now that they have been developed, can be easy to apply. By making appropriate software publicly available, we aim to make it easy for practitioners to use the models proposed in this thesis. We sincerely hope that our techniques will be applied in practice more often.

Nederlandse samenvatting

(Summary in Dutch)

Voor de statistische analyse van data is het belangrijk om de informatie op een overzichtelijke manier te representeren. Diverse multivariate statistische technieken zijn hiervoor ontwikkeld, zoals k -means-clusteranalyse en principale-componentenanalyse. Deze technieken zijn vaak gebaseerd op de beginselen van segmentatie (het in groepen onderverdelen van de observaties) en dimensiereductie (het maken van een visuele, laag-dimensionale weergave van een dataset). Echter, deze technieken maken meestal geen statistische aannames over het proces dat de data genereert, wat tot gevolg heeft dat de statistische significantie van de resultaten vaak onbekend is.

In dit proefschrift hebben we de modelleer-principes van segmentatie en dimensiereductie verwerkt in statistische modellen. We ontwikkelen nieuwe modellen om de informatie in een dataset samen te vatten, opdat we de data makkelijker kunnen begrijpen en verklaren. De nadruk ligt op dimensiereductie met behulp van bilineaire parameterstructuren en technieken voor het clusteren van de rijen en kolommen van een datamatrix. Aan de hand van diverse toepassingen uit de marketing, de psychometrie en de politicologie, laat dit proefschrift zien hoe de voorgestelde technieken extra inzicht kunnen geven in de data.

Dit proefschrift bestaat uit twee delen. Deel I richt zich op het representeren van interactie-effecten door middel van bilineaire decomposities (dimensiereductie). Deel II bespreekt technieken voor het simultaan clusteren van de rijen en kolommen van datamatrices. Deel I bestaat uit drie hoofdstukken en deel II bestaat uit twee hoofdstukken. Hieronder geven we een korte omschrijving van elk hoofdstuk.

In hoofdstuk 2 bespreken we een algemeen model voor het representeren van tweeweg-interactie-effecten met een beperkt aantal parameters. We bespreken een nieuw model, gebaseerd op de methodologie van optimaal schalen (*optimal scaling*), voor het analyseren van de inhoud van de interactie-effecten in gegeneraliseerde lineaire modellen met een willekeurig aantal onafhankelijke variabelen. Dit model, dat we het optimaal schalen van interacties (OSI) model noemen, is een eendimensionaal multiplicatief interactiemodel dat weinig parameters gebruikt. We bespreken hoe het model kan worden gebruikt om de interactie-effecten op grafische wijze weer te geven en te interpreteren. Tevens bespreken we diverse uitbreidingen van het eendimensionale model. Ten slotte pas we het model toe op twee datasets, om te laten zien hoe het model in de praktijk kan worden gebruikt.

Hoofdstuk 2 is gebaseerd op Van Rosmalen et al. (2009) en is gepubliceerd in *Multivariate Behavioral Research*.

Hoofdstuk 3 behandelt het modelleren van antwoordstijlen (de manier waarop respondenten ratingschalen gebruiken), die een risico vormen voor de validiteit van ratingschalen. We onderzoeken hoe de verschillende antwoordstijlen en de meningen over de items de antwoorden op ratingschalen beïnvloeden. Hiervoor ontwikkelen we een nieuw model dat rekening houdt met verschillen in antwoordstijlen, verschillen in mening over de items en achtergrondgegevens van de respondenten. Door een bilineaire parameterstructuur op te leggen op de parameters van een multinomiaal logitmodel, kunnen we op grafische wijze onderscheiden hoe persoonlijke kenmerken en inhoudelijke meningen het antwoordgedrag beïnvloeden. We combineren deze methode met segmentaties op basis van latente klassen, zodat we twee segmentaties van de respondenten krijgen. Een segmentatie groepeerde de respondenten op basis van hun meningen; de andere segmentatie deelt de respondenten op in groepen op basis van hun antwoordstijlen. Dit bilineaire multinomiaal logitmodel met latente klassen wordt toegepast op een internationale dataset waarin de *List of Values* gemeten is. Hoofdstuk 3 is gebaseerd op Van Rosmalen, Van Herk, en Groenen (in druk) en is geaccepteerd voor publicatie in het *Journal of Marketing Research*.

In hoofdstuk 4 passen we de modelleer-technieken van hoofdstuk 2 toe op conjunctanalyse. De interactie-effecten van de attributen worden zelden volledig gemodelleerd bij het analyseren van de resultaten van conjunctanalyse. Een reden hiervoor is dat het modelleren van alle interactie-effecten in het algemeen veel parameters vereist en veel keuzesets per respondent. Daarom introduceren we een nieuwe aanpak voor het modelleren van de tweeweginteracties, gebaseerd op het OSI model van hoofdstuk 2. De bedoeling is om de tweeweginteractie-effecten zo goed mogelijk te schatten met een beperkt design en een beperkt aantal te schatten parameters. Het op deze wijze ontworpen model passen we toe op een conjunctanalysedataset. Tevens laten we zien hoe conjunctanalysedesigns geconstrueerd kunnen worden op basis van ons model. Alex Koning and Patrick Groenen zijn co-auteurs van hoofdstuk 4.

Hoofdstuk 5 vergelijkt diverse optimalisatiemethoden voor het simultaan partitioneren van de rijen en kolommen van datamatrices. Het op deze wijze partitioneren van datamatrices is een relatief nieuwe clustertechniek. We richten ons op partitioneringsmethodes die een criterium optimaliseren dat vergelijkbaar is met het criterium van *k*-means-clusteranalyse. Voor dit type probleem bestaan diverse optimalisatiemethoden. Welke methode de beste resultaten oplevert, is tot nog toe niet duidelijk, aangezien de meeste methodes niet-globale optima van de criteriumfunctie kunnen opleveren. We bespreken en vergelijken diverse optimalisatiemethoden voor het partitioneren van de rijen en kolommen van datamatrices. Naast het behandelen van bestaande methodes, introduceren we ook een nieuwe methode, gebaseerd op het principe van *fuzzy steps*. De effectiviteit van de methodes vergelijken we met behulp van een grote simulatiestudie. Ten slotte gebruiken we een empirische (niet gesimuleerde) dataset om de praktische toepasbaarheid van de techniek te illustreren. Hoofdstuk 5 is gebaseerd op Van Rosmalen, Groenen, et al. (in druk) en is geaccepteerd voor publicatie in het *Journal of Classification*

In hoofdstuk 6 behandelen we een nieuwe Bayesiaanse aanpak voor het schatten van de parameters van een latenteklassenmodel voor het simultaan clusteren van de rijen en kolommen van datamatrices. Een Gibbs sampler met data-augmentatie wordt gebruikt om de posteriorresultaten te verkrijgen. Deze Bayesiaanse aanpak heeft drie voordelen ten opzichte van bestaande methodes. Ten eerste kunnen we statistische inferentie toepassen op de modelparameters, hetgeen niet mogelijk zou zijn met frequentistische schattingsmethodes. Tevens stelt deze Bayesiaanse methode ons in staat, statistische criteria voor het bepalen van de optimale aantallen clusters te definiëren. Ten slotte leidt het gebruik van de Gibbs sampler minder vaak tot problemen met lege klassen of lokale optima in de aannemelijkheidsfunctie dan het EM-algoritme, dat vaak gebruik wordt in frequentistische methodes. Om de praktische toepasbaarheid te illustreren, passen we de Bayesiaanse schattingsmethode toe op twee datasets. De eerste dataset bestaat uit het stemgedrag van het Hoogerechtshof van de Verenigde Staten (zie ook Doreian et al., 2004). De tweede dataset beschrijft de *roll call votes* van het Amerikaanse Huis van Afgevaardigden in 2007. Voor beide datasets laten we zien hoe de resultaten kunnen leiden tot praktisch inzicht in de data. Bram van Dijk en Richard Paap zijn co-auteurs van hoofdstuk 6.

Bibliography

- Anderson, C. J. (1996). The analysis of three-way contingency tables by three-mode association models. *Psychometrika*, *61*, 465-483.
- Anderson, C. J., & Vermunt, J. K. (2000). Log-multiplicative association models as latent variable models for nominal and/or ordinal data. *Sociological Methodology*, *30*, 81-122.
- Anderson, J. A. (1984). Regression and ordered categorical variables. *Journal of the Royal Statistical Society. Series B (Methodological)*, *46*(1), 1-30.
- Andrews, R. L., & Currim, I. S. (2003). A comparison of segment retention criteria for finite mixture logit models. *Journal of Marketing Research*, *40*(2), 235-243.
- Baier, D., Gaul, W., & Schader, M. (1997). Two-mode overlapping clustering with applications in simultaneous benefit segmentation and market structuring. In R. Klar & O. Opitz (Eds.), *Classification and knowledge organization* (p. 557-566). Heidelberg: Springer.
- Baumgartner, H., & Steenkamp, J.-B. E. M. (2001). Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research*, *38*(2), 143-156.
- Benzécri, J.-P. (1982). *Histoire et préhistoire de l'analyse des données*. Paris: Dunod.
- Berger, J. (1985). *Statistical decision theory and Bayesian analysis*. New York: Springer-Verlag.
- Bezdek, J. C. (1981). *Pattern recognition with fuzzy objective function algorithms*. New York: Plenum Press.
- Bock, H. H. (1974). *Automatische klassifikation*. Göttingen: Vandenhoeck & Ruprecht.
- Bond, M. H. (1988). Finding universal dimensions of individual variation in multicultural studies of values: The Rokeach and Chinese value surveys. *Journal of Personality and Social Psychology*, *55*(6), 1009-1015.
- Bozdogan, H. (1994). Mixture-model cluster analysis using model selection criteria and a new information measure of complexity. In H. Bozdogan (Ed.), *Proceedings of the first US/Japan conference on the frontiers of statistical modeling: An informational approach* (Vol. 2, p. 69-113). Boston: Kluwer.
- Breiman, L., & Friedman, J. H. (1985). Estimating optimal transformation for multiple regression and correlation (with discussion). *Journal of the American Statistical*

- Association*, 80, 580-619.
- Broen, W. E. J., & Wirt, R. D. (1958). Varieties of response sets. *Journal of Consulting Psychology*, 22(3), 237-240.
- Brusco, M., & Steinley, D. (2006). Inducing a blockmodel structure of two-mode binary data using seriation procedures. *Journal of Mathematical Psychology*, 50, 468-477.
- Burgess, S. M., & Steenkamp, J.-B. E. M. (2006). Marketing renaissance: How research in emerging markets advances marketing science and practice. *International Journal of Research in Marketing*, 23(4), 337-356.
- Candel, M. J. J. M., & Maris, E. (1997). Perceptual analysis of two-way two-mode frequency data: Probability matrix decomposition and two alternatives. *International Journal of Research in Marketing*, 14, 321-339.
- Carmone, F. J., & Green, P. E. (1981). Model misspecification in multiattribute parameter estimation. *Journal of Marketing Research*, 18(1), 87-93.
- Carroll, J., & Chang, J. (1970). Analysis of individual differences in multidimensional scaling via an N -way generalization of 'Eckart-Young' decomposition. *Psychometrika*, 35, 283-319.
- Castillo, W., & Trejos, J. (2002). Two-mode partitioning: Review of methods and application of tabu search. In K. Jajuga, A. Sololowksi, & H. Bock (Eds.), *Classification, clustering and data analysis* (p. 43-51). Berlin: Springer.
- Celeux, G., Hurn, M., & Robert, C. P. (2000). Computational and inferential difficulties with mixture posterior distributions. *Journal of the American Statistical Association*, 95, 957-970.
- Cheung, G. W., & Rensvold, R. B. (2000). Assessing extreme and acquiescence response sets in cross-cultural research using structural equations modeling. *Journal of Cross-Cultural Psychology*, 31(2), 187-212.
- Choulakian, V. (1996). Generalized bilinear models. *Psychometrika*, 61, 271-283.
- Chrysosoidis, G. M., & Krystallis, A. (2005). Organic consumers' personal values research: Testing and validating the List of Values (LOV) scale and implementing a value-based segmentation task. *Food Quality and Preference*, 17(6), 585-599.
- Clogg, C. C. (1982). Some models for the analysis of association in multiway cross-classifications having ordered categories. *Journal of the American Statistical Association*, 77, 803-815.
- Cormack, R. M. (1971). A review of classification. *Journal of the Royal Statistical Society, Series A*, 134, 321-367.
- Craig, C., & Douglas, S. P. (2005). *International marketing research*. Chichester: Wiley.
- Davies, R. B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, 64(2), 247-254.
- Davis, J. A., & Smith, T. W. (1996). *General Social Surveys 1972-1996: Cumulative codebook*. Chicago, IL: National Opinion Research Center.

- Dayton, C. M., & MacReady, G. B. (1988). Concomitant variable latent class models. *Journal of the American Statistical Association*, *83*, 173-178.
- De Leeuw, J., Young, F. W., & Takane, Y. (1976). Additive structure in qualitative data: An alternating least squares method with optimal scaling features. *Psychometrika*, *41*(4), 471-503.
- De Falguerolles, A., & Francis, B. (1992). Algorithmic approaches for fitting bilinear models. In Y. Dodge & J. Whittaker (Eds.), *Compstat 1992* (Vol. 1, p. 77-82). Heidelberg, Germany: Physica-Verlag.
- De Jong, M. G., Steenkamp, J.-B. E. M., Fox, J.-P., & Baumgartner, H. (2008). Using item response theory to measure extreme response style in marketing research: A global investigation. *Journal of Marketing Research*, *45*(1), 104-115.
- De Leeuw, J. (2006). Principal component analysis of binary data by iterated singular value decomposition. *Computational Statistics & Data Analysis*, *50*, 21-39.
- Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM-algorithm. *Journal of the Royal Statistical Society, Series B*, *39*(1), 1-38.
- DeSarbo, W. S. (1982). GENCLUS: New models for general nonhierarchical clustering analysis. *Psychometrika*, *47*(4), 449-475.
- DeSarbo, W. S., Fong, D. K. H., Liechty, J. C., & Saxton, M. K. (2004). A hierarchical Bayesian procedure for two-mode cluster analysis. *Psychometrika*, *69*, 547-572.
- DeSarbo, W. S., Grewal, R., & Scott, C. J. (2008). A clusterwise bilinear multidimensional scaling methodology for simultaneous segmentation and positioning analysis. *Journal of Marketing Research*, *45*(3), 280-292.
- Dias, J. G., & Wedel, M. (2004). An empirical comparison of EM, SEM and MCMC performance for problematic Gaussian mixture likelihoods. *Statistics and Computing*, *14*, 323-332.
- Dolan, E. D., & Moré, J. J. (2002). Benchmarking optimization software with performance profiles. *Mathematical Programming*, *91*, 201-213.
- Doreian, P., Batagelj, V., & Ferligoj, A. (2004). Generalized blockmodeling of two-mode network data. *Social Networks*, *26*, 29-53.
- Doreian, P., Batagelj, V., & Ferligoj, A. (2005). *Generalized blockmodeling*. Cambridge University Press.
- Fischer, R. (2004). Standardization to account for cross-cultural response bias: A classification of score adjustment procedures and review of research in JCCP. *Journal of Cross-Cultural Psychology*, *35*(3), 263-282.
- Fisher, R. A. (1940). The precision of discriminant functions. *Annals of Eugenics*, *10*, 422-429.
- Fraley, C., & Raftery, A. E. (1998). How many clusters? Which clustering method? Answers via model-based cluster analysis. *The Computer Journal*, *41*, 578-588.

- Frühwirth-Schnatter, S. (2006). *Finite mixture and markov switching models*. New York: Springer.
- Gabriel, K. R. (1998). Generalised bilinear regression. *Biometrika*, *85*(3), 689-700.
- Gaul, W., & Schader, M. (1996). A new algorithm for two-mode clustering. In H. Bock & W. Polasek (Eds.), *Data analysis and information systems* (p. 15-23). Heidelberg: Springer.
- Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2003). *Bayesian data analysis*. Chapman & Hall.
- Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *6*, 721-741.
- Geweke, J. (2007). Interpretation and inference in mixture models: Simple MCMC works. *Computational Statistics & Data Analysis*, *51*, 3529-3550.
- Gifi, A. (1990). *Nonlinear multivariate analysis*. Wiley.
- Glover, F. (1986). Future paths for integer programming and links to artificial intelligence. *Computers and Operations Research*, *13*, 533-549.
- Gofman, A. (2006). Emergent scenarios, synergies and suppressions uncovered within conjoint analysis. *Journal of Sensory Studies*, *21*, 373-414.
- Goodman, L. A. (1981). Association models and canonical correlation in the analysis of cross-classifications having ordered categories. *Journal of the American Statistical Association*, *76*, 320-334.
- Govaert, G., & Nadif, M. (2003). Clustering with block mixture models. *Pattern Recognition*, *36*, 463-473.
- Govaert, G., & Nadif, M. (2008). Block clustering with Bernoulli mixture models: Comparison of different approaches. *Computational Statistics & Data Analysis*, *52*, 3233-3245.
- Gower, J. C., & Hand, D. J. (1996). *Biplots* (No. 54). London: Chapman & Hall.
- Green, P. E., & Wind, Y. (1975). New way to measure consumers' judgments. *Harvard Business Review*, 107-117.
- Greenleaf, E. A. (1992a). Improving rating scale measures by detecting and correcting bias components in some response styles. *Journal of Marketing Research*, *29*(2), 176-188.
- Greenleaf, E. A. (1992b). Measuring extreme response style. *Public Opinion Quarterly*, *56*(3), 328-351.
- Groenen, P. J. F., & Jajuga, K. (2001). Fuzzy clustering with squared Minkowski distances. *Fuzzy Sets and Systems*, *120*, 227-237.
- Groenen, P. J. F., & Koning, A. J. (2006). A new model for visualizing interactions in analysis of variance. In M. Greenacre & J. Blasius (Eds.), *Multiple correspondence analysis and related methods* (p. 487-502). Chapman & Hall.
- Grunert, S. C., & Scherhorn, G. (1990). Consumer values in West Germany underlying

- dimensions and cross-cultural comparison with North America. *Journal of Business Research*, 20(2), 97-107.
- Guttman, L. (1941). The quantification of a class of attributes: A theory and method of scale construction. In The Committee on Social Adjustment (Ed.), *The prediction of personal adjustment* (p. 319-348). New York: Social Science Research Council.
- Hamilton, D. L. (1968). Personality attributes associated with extreme response style. *Psychological Bulletin*, 69(3), 192-203.
- Han, C., & Carlin, B. P. (2001). Markov chain Monte Carlo methods for computing Bayes factors: A comparative review. *Journal of the American Statistical Association*, 96, 1122-1133.
- Hansohm, J. (2001). Two-mode clustering with genetic algorithms. In *Classification, automation, and new media* (p. 87-93). Berlin: Springer.
- Harshman, R. (1970). Foundations of the PARAFAC procedure: models and conditions for an "exploratory" multimodal factor analysis. *UCLA Working Papers in Phonetics*, 16, 184.
- Hartigan, J. A. (1975). *Clustering algorithms*. New York: John Wiley and Sons.
- Hartigan, J. A. (2000). Bloc voting in the United States senate. *Journal of Classification*, 17, 29-49.
- Hastie, T. J., & Tibshirani, R. J. (1990). *Generalized additive models*. New York: Chapman and Hall.
- Heckman, J. J., & Snyder Jr., J. M. (1997). Linear probability models of the demand for attributes with an empirical application to estimating the preferences of legislators. *RAND Journal of Economics*, 28, S142-S189.
- Heiser, W. J., & Groenen, J. (1997). Cluster differences scaling with a within-clusters loss component and a fuzzy successive approximation strategy to avoid local minima. *Psychometrika*, 62(1), 63-83.
- Hoaglin, D. C., Mosteller, F., & Tukey, J. W. (1983). *Understanding robust and exploratory data analysis*. Wiley.
- Hoffman, D. L., & Franke, G. R. (1986). Correspondence analysis: Graphical representation of categorical data in marketing research. *Journal of Marketing Research*, 23(3), 213-227.
- Hofstede, G. (2001). *Culture's consequences: Comparing values, behaviors, institutions, and organizations across nations*. Thousand Oaks, CA: Sage.
- Holland, C. W., & Craves, D. W. (1973). Fractional factorial experimental designs in marketing research. *Journal of Marketing Research*, 10(3), 270-276.
- Huber, J., & Zwerina, K. (1996). The importance of utility balance in efficient choice designs. *Journal of Marketing Research*, 33(3), 307-317.
- Hubert, L., & Arabie, P. (1985). Comparing partitions. *Journal of Classification*, 2, 193-218.

- Hui, C. H., & Triandis, H. C. (1989). Effects of culture and response format on extreme response style. *Journal of Cross-Cultural Psychology*, *20*(3), 296-309.
- Johnson, T., Kulesa, P., Cho, Y. I., & Shavitt, S. (2005). The relation between culture and response styles: Evidence from 19 countries. *Journal of Cross-Cultural Psychology*, *36*(2), 264-277.
- Jolliffe, I. (2002). *Principal component analysis* (2nd ed.). Springer.
- Kahle, L. R. (1983). *Social values and social change: Adaptation to life in America*. New York: Praeger.
- Kamakura, W. A., & Novak, T. P. (1992). Value-system segmentation: Exploring the meaning of LOV. *Journal of Consumer Research*, *19*(1), 119-132.
- Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*, 773-795.
- Kessels, R., Goos, P., & Vandebroek, M. (2006). A comparison of criteria to design efficient choice experiments. *Journal of Marketing Research*, *43*(3), 409-419.
- Kropp, F., Lavack, A. M., & Silvera, D. H. (2005). Values and collective self-esteem as predictors of consumer susceptibility to interpersonal influence among university students. *International Marketing Review*, *22*(1), 7-33.
- Krosnick, J. A., & Alwin, D. F. (1988). A test of the form-resistant correlation hypothesis: Ratings, rankings, and the measurement of values. *Public Opinion Quarterly*, *52*(4), 526-538.
- Kruskal, J. B. (1965). Analysis of factorial experiments by estimating monotone transformations of the data. *Journal of the Royal Statistical Society. Series B (Methodological)*, *27*, 251-263.
- Kuhfeld, W. F., Tobias, R. D., & Garratt, M. (1994). Efficient experimental designs with marketing research applications. *Journal of Marketing Research*, *21*, 545-557.
- Lenk, P., Wedel, M., & Böckenholt, U. (2006). Bayesian estimation of circumplex models subject to prior theory constraints and scale-usage bias. *Psychometrika*, *71*(1), 33-55.
- Linting, M., Meulman, J. J., Groenen, P. J. F., & Van der Kooij, A. J. (2007). Nonlinear principal components analysis: Introduction and application. *Psychological Methods*, *12*, 336-358.
- McCullagh, P., & Nelder, J. (1989). *Generalized linear models* (Second ed.). Chapman & Hall.
- McLachlan, G., & Peel, D. (2000). *Finite mixture models*. Wiley.
- Milligan, G. W. (1980). An examination of the effect of six types of error perturbation on fifteen clustering algorithms. *Psychometrika*, *45*(3), 325-342.
- Milligan, G. W., & Cooper, M. C. (1985). An examination of procedures for determining the number of clusters in a data set. *Psychometrika*, *50*(2), 159-179.
- Mirkin, B. (2005). *Clustering for data mining: A data recovery approach*. Chapman &

- Hall.
- Mirkin, B., Arabie, P., & Hubert, L. J. (1995). Additive two-mode clustering: The error-variance approach revisited. *Journal of Classification*, *12*, 243-263.
- Moum, T. (1988). Yea-saying and mood-of-the-day effects in self-reported quality of life. *Social Indicators Research*, *20*(2), 117-139.
- Nelder, J., & Wedderburn, R. (1972). Generalized linear models. *Journal of The Royal Statistical Society. Series A (General)*, *135*(3), 370-384.
- Nelson, J. P. (2002). "Green" voting and ideology: LCV scores and roll-call voting in the U.S. Senate, 1988-1998. *The Review of Economics and Statistics*, *84*, 518-529.
- Newton, M. A., & Raftery, A. E. (1994). Approximate Bayesian inference by the weighted likelihood bootstrap. *Journal of the Royal Statistical Society, Series B*, *56*, 3-48.
- Noma, E., & Smith, D. R. (1985). Benchmark for the blocking of sociometric data. *Psychological Bulletin*, *97*(3), 583-591.
- Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver, & L. S. Wrightman (Eds.), *Measures of personality and social psychological attitudes* (p. 17-59). San Diego, CA, US: Academic Press, Inc.
- Poole, K. T., & Rosenthal, H. (1991). Patterns of congressional voting. *American Journal of Political Science*, *35*, 228-278.
- Richardson, M., & Kuder, G. F. (1933). Making a rating scale that measures. *Personnel Journal*, *12*, 36-40.
- Rocci, R., & Vichi, M. (2008). Two-mode multi-partitioning. *Computational Statistics & Data Analysis*, *52*, 1984-2003.
- Rossi, P. E., Gilula, Z., & Allenby, G. M. (2001). Overcoming scale usage heterogeneity: A bayesian hierarchical approach. *Journal of the American Statistical Association*, *96*(453), 20-31.
- Sándor, Z., & Wedel, M. (2001). Designing conjoint choice experiments using managers' prior beliefs. *Journal of Marketing Research*, *38*(4), 430-444.
- Schepers, J., Ceulemans, E., & Van Mechelen, I. (2008). Selecting among multi-mode partitioning models of different complexities: A comparison of four model selection criteria. *Journal of Classification*, *25*, 67-85.
- Schepers, J., Van Mechelen, I., & Ceulemans, E. (2006). Three-mode partitioning. *Computational Statistics & Data Analysis*, *51*, 1623-1642.
- Schwartz, S. H. (1992). Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 25, p. 1-65). New York: Academic Press.
- Schwarz, N. (2003). Sex differences in value priorities: Cross-cultural and multimethod studies. *Journal of Consumer Research*, *29*(4), 588-594.
- Siciliano, R., & Mooijaart, A. (1997). Three-factor association models for three-way contingency tables. *Computational Statistics & Data Analysis*, *24*, 337-356.

- Smith, P. B. (2004). Acquiescent response bias as an aspect of cultural communication style. *Journal of Cross-Cultural Psychology, 35*(1), 50-61.
- Sneath, P. (1957). The application of computers to taxonomy. *Journal of General Microbiology, 17*, 201-226.
- Stening, B. W., & Everett, J. E. (1984). Response styles in a cross-cultural managerial study. *Journal of Social Psychology, 122*(2), 151-156.
- Swain, S. D., Weathers, D., & Niedrich, R. W. (2008). Assessing three sources of misresponse to reversed likert items. *Journal of Marketing Research, 45*(1), 116-131.
- Symons, M. J. (1981). Clustering criteria and multivariate normal mixtures. *Biometrics, 37*, 35-43.
- Tanner, M. A., & Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association, 82*, 528-540.
- Teicher, H. (1963). Identifiability of finite mixtures. *The Annals of Mathematical Statistics, 34*(4), 1265-1269.
- Ter Hofstede, F., Steenkamp, J.-B. E. M., & Michel, W. (1999). International market segmentation based on consumer-product relations. *Journal of Marketing Research, 36*(1), 1-17.
- Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge, United Kingdom: Cambridge University Press.
- Trejos, J., & Castillo, W. (2000). Simulated annealing optimization for two-mode partitioning. In W. Gaul & R. Decker (Eds.), *Classification and information at the turn of the millenium* (p. 135-142). Heidelberg: Springer.
- Tsao, E. C. K., Bezdek, J. C., & Pal, N. R. (1994). Fuzzy Kohonen clustering networks. *Pattern Recognition, 27*(5), 757-764.
- Usunier, J.-C., & Lee, J. A. (2005). *Marketing across cultures*. Essex, England: Pearson Education Limited.
- Van Eeuwijk, F. A. (1995). Multiplicative interaction in generalized linear models. *Biometrics, 51*, 1017-1032.
- Van Eeuwijk, F. A. (1996). *Between and beyond additivity and non-additivity: The statistical modelling of genotype by environment interaction in plant breeding*. Unpublished doctoral dissertation, Wageningen Agricultural University, Wageningen, the Netherlands.
- Van Herk, H. (2000). *Equivalence in a cross-national context: Methodological & empirical issues in marketing research*. Unpublished doctoral dissertation, Tilburg University, Tilburg, the Netherlands.
- Van Herk, H., Poortinga, Y. H., & Verhallen, T. M. M. (2004). Response styles in rating scales: Evidence of method bias in data from six EU countries. *Journal of Cross-Cultural Psychology, 35*(3), 346-360.

- Van Laarhoven, P. J. M., & Aarts, E. H. L. (1987). *Simulated annealing: Theory and applications*. Eindhoven: Kluwer Academic Publishers.
- Van Mechelen, I., Bock, H. H., & De Boeck, P. (2004). Two-mode clustering methods: A structured overview. *Statistical Methods in Medical Research*, *13*, 363-394.
- Van Rosmalen, J., Groenen, P. J. F., Trejos, J., & Castillo, W. (in press). Optimization strategies for two-mode partitioning. *Journal of Classification*.
- Van Rosmalen, J., Koning, A. J., & Groenen, P. J. F. (2009). Optimal scaling of interaction effects in generalized linear models. *Multivariate Behavioral Research*, *44*(1), 59-81.
- Van Rosmalen, J., Van Herk, H., & Groenen, P. J. F. (in press). Identifying response styles: A latent-class bilinear multinomial logit model. *Journal of Marketing Research*.
- Vichi, M. (2001). Double k -means clustering for simultaneous classification of objects and variables. In S. Borra, R. Rocchi, & M. Schader (Eds.), *Advances in classification and data analysis. studies in classification, data analysis, and knowledge organization* (p. 43-52). Heidelberg: Springer.
- Vriens, M. (1995). *Conjoint analysis in marketing*. Unpublished doctoral dissertation, Rijksuniversiteit Groningen, Groningen, the Netherlands.
- Wedel, M., & Kamakura, W. A. (2000). *Market segmentation: Conceptual and methodological foundations* (2nd ed.). Dordrecht: Kluwer.
- Wedel, M., Ter Hofstede, F., & Steenkamp, J.-B. E. M. (1998). Mixture model analysis of complex samples. *Journal of Classification*, *15*(2), 225-244.
- Winkler, J. D., Kanouse, D. E., & Ware, J. E. (1982). Controlling for acquiescence response set in scale development. *Journal of Applied Psychology*, *67*, 555-561.
- Wong, R.-K. (2001). Multidimensional association models: A multilinear approach. *Sociological Methods and Research*, *30*(2), 197-24.
- Yang, C.-C., & Yang, C.-C. (2007). Separating latent classes by information criteria. *Journal of Classification*, *24*(2), 183-203.
- Young, F. W. (1981). Quantitative analysis of qualitative data. *Psychometrika*, *46*, 357-388.
- Young, F. W., De Leeuw, J., & Takane, Y. (1976). Regression with qualitative and quantitative variables: An alternating least squares method with optimal scaling features. *Psychometrika*, *41*(4), 505-529.

Curriculum Vitae



Joost van Rosmalen (1982) obtained his master's degree in Econometrics and Management Science with honors from Erasmus University Rotterdam in 2004. In the same year, he started his PhD research. His main research interests are in quantitative marketing research, psychometrics, and econometrics. His research has been presented at several international conferences and has resulted in several papers, of which three have been accepted for publication in *Journal of Classification*, *Journal of Marketing Research*, and *Multivariate Behavioral Research*.

ERASMUS RESEARCH INSTITUTE OF MANAGEMENT (ERIM)

ERIM PH.D. SERIES
RESEARCH IN MANAGEMENT

ERIM Electronic Series Portal: <http://hdl.handle.net/1765/1>

Althuizen, N.A.P., *Analogical Reasoning as a Decision Support Principle for Weakly Structured Marketing Problems*, Promotor: Prof.dr.ir. B. Wierenga, EPS-2006-095-MKT, ISBN: 90-5892-129-8, <http://hdl.handle.net/1765/8190>

Alvarez, H.L., *Distributed Collaborative Learning Communities Enabled by Information Communication Technology*, Promotor: Prof.dr. K. Kumar, EPS-2006-080-LIS, ISBN: 90-5892-112-3, <http://hdl.handle.net/1765/7830>

Appelman, J.H., *Governance of Global Interorganizational Tourism Networks: Changing Forms of Coordination between the Travel Agency and Aviation Sector*, Promotors: Prof.dr. F.M. Go & Prof.dr. B. Nooteboom, EPS-2004-036-MKT, ISBN: 90-5892-060-7, <http://hdl.handle.net/1765/1199>

Assem, M.J. van den, *Deal or No Deal? Decision Making under Risk in a Large-Stake TV Game Show and Related Experiments*, Promotor: Prof.dr. J. Spronk, EPS-2008-138-F&A, ISBN: 978-90-5892-173-4, <http://hdl.handle.net/1765/13566>

Baquero, G., *On Hedge Fund Performance, Capital Flows and Investor Psychology*, Promotor: Prof.dr. M.J.C.M. Verbeek, EPS-2006-094-F&A, ISBN: 90-5892-131-X, <http://hdl.handle.net/1765/8192>

Berens, G., *Corporate Branding: The Development of Corporate Associations and their Influence on Stakeholder Reactions*, Promotor: Prof.dr. C.B.M. van Riel, EPS-2004-039-ORG, ISBN: 90-5892-065-8, <http://hdl.handle.net/1765/1273>

Berghe, D.A.F. van den, *Working Across Borders: Multinational Enterprises and the Internationalization of Employment*, Promotors: Prof.dr. R.J.M. van Tulder & Prof.dr. E.J.J. Schenk, EPS-2003-029-ORG, ISBN: 90-5892-05-34, <http://hdl.handle.net/1765/1041>

Berghman, L.A., *Strategic Innovation Capacity: A Mixed Method Study on Deliberate Strategic Learning Mechanisms*, Promotor: Prof.dr. P. Mattysens, EPS-2006-087-MKT, ISBN: 90-5892-120-4, <http://hdl.handle.net/1765/7991>

Bijman, W.J.J., *Essays on Agricultural Co-operatives: Governance Structure in Fruit and Vegetable Chains*, Promotor: Prof.dr. G.W.J. Hendrikse, EPS-2002-015-ORG, ISBN: 90-5892-024-0, <http://hdl.handle.net/1765/867>

Bispo, A., *Labour Market Segmentation: An investigation into the Dutch hospitality industry*, Promotors: Prof.dr. G.H.M. Evers & Prof.dr. A.R. Thurik, EPS-2007-108-ORG, ISBN: 90-5892-136-9, <http://hdl.handle.net/1765/10283>

Blindenbach-Driessen, F., *Innovation Management in Project-Based Firms*, Promotor: Prof.dr. S.L. van de Velde, EPS-2006-082-LIS, ISBN: 90-5892-110-7, <http://hdl.handle.net/1765/7828>

- Boer, C.A., *Distributed Simulation in Industry*, Promotors: Prof.dr. A. de Bruin & Prof.dr.ir. A. Verbraeck, EPS-2005-065-LIS, ISBN: 90-5892-093-3, <http://hdl.handle.net/1765/6925>
- Boer, N.I., *Knowledge Sharing within Organizations: A situated and Relational Perspective*, Promotor: Prof.dr. K. Kumar, EPS-2005-060-LIS, ISBN: 90-5892-086-0, <http://hdl.handle.net/1765/6770>
- Boer-Sorbán, K., *Agent-Based Simulation of Financial Markets: A modular, Continuous-Time Approach*, Promotor: Prof.dr. A. de Bruin, EPS-2008-119-LIS, ISBN: 90-5892-155-0, <http://hdl.handle.net/1765/10870>
- Boon, C.T., *HRM and Fit: Survival of the Fittest!?*, Promotors: Prof.dr. J. Paauwe & Prof.dr. D.N. den Hartog, EPS-2008-129-ORG, ISBN: 978-90-5892-162-8, <http://hdl.handle.net/1765/12606>
- Braun, E., *City Marketing: Towards an Integrated Approach*, Promotor: Prof.dr. L. van den Berg, EPS-2008-142-MKT, ISBN: 978-90-5892-180-2, <http://hdl.handle.net/1765/13694>
- Brito, M.P. de, *Managing Reverse Logistics or Reversing Logistics Management?* Promotors: Prof.dr.ir. R. Dekker & Prof.dr. M. B. M. de Koster, EPS-2004-035-LIS, ISBN: 90-5892-058-5, <http://hdl.handle.net/1765/1132>
- Brohm, R., *Polycentric Order in Organizations: A Dialogue between Michael Polanyi and IT-Consultants on Knowledge, Morality, and Organization*, Promotors: Prof.dr. G. W. J. Hendrikse & Prof.dr. H. K. Letiche, EPS-2005-063-ORG, ISBN: 90-5892-095-X, <http://hdl.handle.net/1765/6911>
- Brumme, W.-H., *Manufacturing Capability Switching in the High-Tech Electronics Technology Life Cycle*, Promotors: Prof.dr.ir. J.A.E.E. van Nunen & Prof.dr.ir. L.N. Van Wassenhove, EPS-2008-126-LIS, ISBN: 978-90-5892-150-5, <http://hdl.handle.net/1765/12103>
- Burgers, J.H., *Managing Corporate Venturing: Multilevel Studies on Project Autonomy, Integration, Knowledge Relatedness, and Phases in the New Business Development Process*, Promotors: Prof.dr.ir. F.A.J. Van den Bosch & Prof.dr. H.W. Volberda, EPS-2008-136-STR, ISBN: 978-90-5892-174-1, <http://hdl.handle.net/1765/13484>
- Campbell, R.A.J., *Rethinking Risk in International Financial Markets*, Promotor: Prof.dr. C.G. Koedijk, EPS-2001-005-F&A, ISBN: 90-5892-008-9, <http://hdl.handle.net/1765/306>
- Chen, H., *Individual Mobile Communication Services and Tariffs*, Promotor: Prof.dr. L.F.J.M. Pau, EPS-2008-123-LIS, ISBN: 90-5892-158-1, <http://hdl.handle.net/1765/11141>
- Chen, Y., *Labour Flexibility in China's Companies: An Empirical Study*, Promotors: Prof.dr. A. Buitendam & Prof.dr. B. Krug, EPS-2001-006-ORG, ISBN: 90-5892-012-7, <http://hdl.handle.net/1765/307>
- Damen, F.J.A., *Taking the Lead: The Role of Affect in Leadership Effectiveness*, Promotor: Prof.dr. D.L. van Knippenberg, EPS-2007-107-ORG, <http://hdl.handle.net/1765/10282>
- Daniševská, P., *Empirical Studies on Financial Intermediation and Corporate Policies*, Promotor: Prof.dr. C.G. Koedijk, EPS-2004-044-F&A, ISBN: 90-5892-070-4, <http://hdl.handle.net/1765/1518>
- Delporte-Vermeiren, D.J.E., *Improving the Flexibility and Profitability of ICT-enabled Business Networks: An Assessment Method and Tool*, Promotors: Prof. mr. dr. P.H.M. Vervest & Prof.dr.ir. H.W.G.M. van Heck, EPS-2003-020-LIS, ISBN: 90-5892-040-2, <http://hdl.handle.net/1765/359>
- Derwall, J.M.M., *The Economic Virtues of SRI and CSR*, Promotor: Prof.dr. C.G. Koedijk, EPS-2007-101-F&A, ISBN: 90-5892-132-8, <http://hdl.handle.net/1765/8986>

Diepen, M. van, *Dynamics and Competition in Charitable Giving*, Promotor: Prof.dr. Ph.H.B.F. Franses, EPS-2009-159-MKT, ISBN: 978-90-5892-188-8, <http://hdl.handle.net/1765/1>

Dijksterhuis, M., *Organizational Dynamics of Cognition and Action in the Changing Dutch and US Banking Industries*, Promotors: Prof.dr.ir. F.A.J. van den Bosch & Prof.dr. H.W. Volberda, EPS-2003-026-STR, ISBN: 90-5892-048-8, <http://hdl.handle.net/1765/1037>

Eijk, A.R. van der, *Behind Networks: Knowledge Transfer, Favor Exchange and Performance*, Promotors: Prof.dr. S.L. van de Velde & Prof.dr.dr. W.A. Dolfsma, EPS-2009-161-LIS, ISBN: 978-90-5892-190-1, <http://hdl.handle.net/1765/1>

Elstak, M.N., *Flipping the Identity Coin: The Comparative Effect of Perceived, Projected and Desired Organizational Identity on Organizational Identification and Desired Behavior*, Promotor: Prof.dr. C.B.M. van Riel, EPS-2008-117-ORG, ISBN: 90-5892-148-2, <http://hdl.handle.net/1765/10723>

Erken, H.P.G., *Productivity, R&D and Entrepreneurship*, Promotor: Prof.dr. A.R. Thurik, EPS-2008-147-ORG, ISBN: 978-90-5892-179-6, <http://hdl.handle.net/1765/14004>

Fenema, P.C. van, *Coordination and Control of Globally Distributed Software Projects*, Promotor: Prof.dr. K. Kumar, EPS-2002-019-LIS, ISBN: 90-5892-030-5, <http://hdl.handle.net/1765/360>

Fleischmann, M., *Quantitative Models for Reverse Logistics*, Promotors: Prof.dr.ir. J.A.E.E. van Nunen & Prof.dr.ir. R. Dekker, EPS-2000-002-LIS, ISBN: 35-4041-711-7, <http://hdl.handle.net/1765/1044>

Flier, B., *Strategic Renewal of European Financial Incumbents: Coevolution of Environmental Selection, Institutional Effects, and Managerial Intentionality*, Promotors: Prof.dr.ir. F.A.J. van den Bosch & Prof.dr. H.W. Volberda, EPS-2003-033-STR, ISBN: 90-5892-055-0, <http://hdl.handle.net/1765/1071>

Fok, D., *Advanced Econometric Marketing Models*, Promotor: Prof.dr. Ph.H.B.F. Franses, EPS-2003-027-MKT, ISBN: 90-5892-049-6, <http://hdl.handle.net/1765/1035>

Ganzaroli, A., *Creating Trust between Local and Global Systems*, Promotors: Prof.dr. K. Kumar & Prof.dr. R.M. Lee, EPS-2002-018-LIS, ISBN: 90-5892-031-3, <http://hdl.handle.net/1765/361>

Gilsing, V.A., *Exploration, Exploitation and Co-evolution in Innovation Networks*, Promotors: Prof.dr. B. Nooteboom & Prof.dr. J.P.M. Groenewegen, EPS-2003-032-ORG, ISBN: 90-5892-054-2, <http://hdl.handle.net/1765/1040>

Gijsbers, G.W., *Agricultural Innovation in Asia: Drivers, Paradigms and Performance*, Promotor: Prof.dr. R.J.M. van Tulder, EPS-2009-156-ORG, ISBN: 978-90-5892-191-8, <http://hdl.handle.net/1765/1>

Govers, R., *Virtual Tourism Destination Image: Global Identities Constructed, Perceived and Experienced*, Promotors: Prof.dr. F.M. Go & Prof.dr. K. Kumar, EPS-2005-069-MKT, ISBN: 90-5892-107-7, <http://hdl.handle.net/1765/6981>

Graaf, G. de, *Tractable Morality: Customer Discourses of Bankers, Veterinarians and Charity Workers*, Promotors: Prof.dr. F. Leijnse & Prof.dr. T. van Willigenburg, EPS-2003-031-ORG, ISBN: 90-5892-051-8, <http://hdl.handle.net/1765/1038>

Groot, E.A. de, *Essays on Economic Cycles*, Promotors: Prof.dr. Ph.H.B.F. Franses & Prof.dr. H.R. Commandeur, EPS-2006-091-MKT, ISBN: 90-5892-123-9, <http://hdl.handle.net/1765/8216>

Gutkowska, A.B., *Essays on the Dynamic Portfolio Choice*, Promotor: Prof.dr. A.C.F. Vorst, EPS-2006-085-F&A, ISBN: 90-5892-118-2, <http://hdl.handle.net/1765/7994>

Hagemijer, R.E., *The Unmasking of the Other*, Promotors: Prof.dr. S.J. Magala & Prof.dr. H.K. Letiche, EPS-2005-068-ORG, ISBN: 90-5892-097-6, <http://hdl.handle.net/1765/6963>

Halderen, M.D. van, *Organizational Identity Expressiveness and Perception Management: Principles for Expressing the Organizational Identity in Order to Manage the Perceptions and Behavioral Reactions of External Stakeholders*, Promotor: Prof.dr. S.B.M. van Riel, EPS-2008-122-ORG, ISBN: 90-5892-153-6, <http://hdl.handle.net/1765/10872>

Hartigh, E. den, *Increasing Returns and Firm Performance: An Empirical Study*, Promotor: Prof.dr. H.R. Commandeur, EPS-2005-067-STR, ISBN: 90-5892-098-4, <http://hdl.handle.net/1765/6939>

Hermans, J.M., *ICT in Information Services: Use and Deployment of the Dutch Securities Trade, 1860-1970*, Promotor: Prof.dr. drs. F.H.A. Janszen, EPS-2004-046-ORG, ISBN 90-5892-072-0, <http://hdl.handle.net/1765/1793>

Hessels, S.J.A., *International Entrepreneurship: Value Creation Across National Borders*, Promotor: Prof.dr. A.R. Thurik, EPS-2008-144-ORG, ISBN: 978-90-5892-181-9, <http://hdl.handle.net/1765/13942>

Heugens, P.P.M.A.R., *Strategic Issues Management: Implications for Corporate Performance*, Promotors: Prof.dr.ir. F.A.J. van den Bosch & Prof.dr. C.B.M. van Riel, EPS-2001-007-STR, ISBN: 90-5892-009-9, <http://hdl.handle.net/1765/358>

Heuvel, W. van den, *The Economic Lot-Sizing Problem: New Results and Extensions*, Promotor: Prof.dr. A.P.L. Wagelmans, EPS-2006-093-LIS, ISBN: 90-5892-124-7, <http://hdl.handle.net/1765/1805>

Hoedemaekers, C.M.W., *Performance, Pinned down: A Lacanian Analysis of Subjectivity at Work*, Promotors: Prof.dr. S. Magala & Prof.dr. D.H. den Hartog, EPS-2008-121-ORG, ISBN: 90-5892-156-7, <http://hdl.handle.net/1765/10871>

Hooghiemstra, R., *The Construction of Reality: Cultural Differences in Self-serving Behaviour in Accounting Narratives*, Promotors: Prof.dr. L.G. van der Tas RA & Prof.dr. A.Th.H. Pruyn, EPS-2003-025-F&A, ISBN: 90-5892-047-X, <http://hdl.handle.net/1765/871>

Hu, Y., *Essays on the Governance of Agricultural Products: Cooperatives and Contract Farming*, Promotors: Prof.dr. G.W.J. Hendrkse & Prof.dr. B. Krug, EPS-2007-113-ORG, ISBN: 90-5892-145-1, <http://hdl.handle.net/1765/10535>

Huij, J.J., *New Insights into Mutual Funds: Performance and Family Strategies*, Promotor: Prof.dr. M.C.J.M. Verbeek, EPS-2007-099-F&A, ISBN: 90-5892-134-4, <http://hdl.handle.net/1765/9398>

Huurman, C.I., *Dealing with Electricity Prices*, Promotor: Prof.dr. C.D. Koedijk, EPS-2007-098-F&A, ISBN: 90-5892-130-1, <http://hdl.handle.net/1765/9399>

Iastrebova, K., *Manager's Information Overload: The Impact of Coping Strategies on Decision-Making Performance*, Promotor: Prof.dr. H.G. van Dissel, EPS-2006-077-LIS, ISBN: 90-5892-111-5, <http://hdl.handle.net/1765/7329>

Iwaarden, J.D. van, *Changing Quality Controls: The Effects of Increasing Product Variety and Shortening Product Life Cycles*, Promotors: Prof.dr. B.G. Dale & Prof.dr. A.R.T. Williams, EPS-2006-084-ORG, ISBN: 90-5892-117-4, <http://hdl.handle.net/1765/7992>

- Jansen, J.J.P., *Ambidextrous Organizations*, Promotors: Prof.dr.ir. F.A.J. Van den Bosch & Prof.dr. H.W. Volberda, EPS-2005-055-STR, ISBN: 90-5892-081-X, <http://hdl.handle.net/1765/6774>
- Jaspers, F.P.H., *Organizing Systemic Innovation*, Promotor: Prof.dr.ir. J.C.M. van den Ende, EPS-2009-160-ORG, ISBN: 978-90-5892-197-), <http://hdl.handle.net/1765/1>
- Jennen, M.G.J., *Empirical Essays on Office Market Dynamics*, Promotors: Prof.dr. C.G. Koedijk & Prof.dr. D. Brounen, EPS-2008-140-F&A, ISBN: 978-90-5892-176-5, <http://hdl.handle.net/1765/13692>
- Jong, C. de, *Dealing with Derivatives: Studies on the Role, Informational Content and Pricing of Financial Derivatives*, Promotor: Prof.dr. C.G. Koedijk, EPS-2003-023-F&A, ISBN: 90-5892-043-7, <http://hdl.handle.net/1765/1043>
- Keizer, A.B., *The Changing Logic of Japanese Employment Practices: A Firm-Level Analysis of Four Industries*, Promotors: Prof.dr. J.A. Stam & Prof.dr. J.P.M. Groenewegen, EPS-2005-057-ORG, ISBN: 90-5892-087-9, <http://hdl.handle.net/1765/6667>
- Kijkuit, R.C., *Social Networks in the Front End: The Organizational Life of an Idea*, Promotor: Prof.dr. B. Nooteboom, EPS-2007-104-ORG, ISBN: 90-5892-137-6, <http://hdl.handle.net/1765/10074>
- Kippers, J., *Empirical Studies on Cash Payments*, Promotor: Prof.dr. Ph.H.B.F. Franses, EPS-2004-043-F&A, ISBN: 90-5892-069-0, <http://hdl.handle.net/1765/1520>
- Klein, M.H., *Poverty Alleviation through Sustainable Strategic Business Models: Essays on Poverty Alleviation as a Business Strategy*, Promotor: Prof.dr. H.R. Commandeur, EPS-2008-135-STR, ISBN: 978-90-5892-168-0, <http://hdl.handle.net/1765/13482>
- Knapp, S., *The Econometrics of Maritime Safety: Recommendations to Enhance Safety at Sea*, Promotor: Prof.dr. Ph.H.B.F. Franses, EPS-2007-096-ORG, ISBN: 90-5892-127-1, <http://hdl.handle.net/1765/7913>
- Kole, E., *On Crises, Crashes and Comovements*, Promotors: Prof.dr. C.G. Koedijk & Prof.dr. M.J.C.M. Verbeek, EPS-2006-083-F&A, ISBN: 90-5892-114-X, <http://hdl.handle.net/1765/7829>
- Kooij-de Bode, J.M., *Distributed Information and Group Decision-Making: Effects of Diversity and Affect*, Promotor: Prof.dr. D.L. van Knippenberg, EPS-2007-115-ORG, <http://hdl.handle.net/1765/10722>
- Koppius, O.R., *Information Architecture and Electronic Market Performance*, Promotors: Prof.dr. P.H.M. Vervest & Prof.dr.ir. H.W.G.M. van Heck, EPS-2002-013-LIS, ISBN: 90-5892-023-2, <http://hdl.handle.net/1765/921>
- Kotlarsky, J., *Management of Globally Distributed Component-Based Software Development Projects*, Promotor: Prof.dr. K. Kumar, EPS-2005-059-LIS, ISBN: 90-5892-088-7, <http://hdl.handle.net/1765/6772>
- Krauth, E.I., *Real-Time Planning Support: A Task-Technology Fit Perspective*, Promotors: Prof.dr. S.L. van de Velde & Prof.dr. J. van Hillegersberg, EPS-2008-155-LIS, ISBN: 978-90-5892-193-2, <http://hdl.handle.net/1765/1>
- Kuilman, J., *The Re-Emergence of Foreign Banks in Shanghai: An Ecological Analysis*, Promotor: Prof.dr. B. Krug, EPS-2005-066-ORG, ISBN: 90-5892-096-8, <http://hdl.handle.net/1765/6926>
- Langen, P.W. de, *The Performance of Seaport Clusters: A Framework to Analyze Cluster Performance and an Application to the Seaport Clusters of Durban, Rotterdam and the Lower Mississippi*, Promotors: Prof.dr. B. Nooteboom & Prof. drs. H.W.H. Welters, EPS-2004-034-LIS, ISBN: 90-5892-056-9, <http://hdl.handle.net/1765/1133>

Le Anh, T., *Intelligent Control of Vehicle-Based Internal Transport Systems*, Promotors: Prof.dr. M.B.M. de Koster & Prof.dr.ir. R. Dekker, EPS-2005-051-LIS, ISBN: 90-5892-079-8, <http://hdl.handle.net/1765/6554>

Le-Duc, T., *Design and Control of Efficient Order Picking Processes*, Promotor: Prof.dr. M.B.M. de Koster, EPS-2005-064-LIS, ISBN: 90-5892-094-1, <http://hdl.handle.net/1765/6910>

Leeuwen, E.P. van, *Recovered-Resource Dependent Industries and the Strategic Renewal of Incumbent Firm: A Multi-Level Study of Recovered Resource Dependence Management and Strategic Renewal in the European Paper and Board Industry*, Promotors: Prof.dr.ir. F.A.J. Van den Bosch & Prof.dr. H.W. Volberda, EPS-2007-109-STR, ISBN: 90-5892-140-6, <http://hdl.handle.net/1765/10183>

Lentink, R.M., *Algorithmic Decision Support for Shunt Planning*, Promotors: Prof.dr. L.G. Kroon & Prof.dr.ir. J.A.E.E. van Nunen, EPS-2006-073-LIS, ISBN: 90-5892-104-2, <http://hdl.handle.net/1765/7328>

Li, T., *Informedness and Customer-Centric Revenue Management*, Promotors: Prof.dr. P.H.M. Vervest & Prof.dr.ir. H.W.G.M. van Heck, EPS-2009-146-LIS, ISBN: 978-90-5892-195-6, <http://hdl.handle.net/1765/1>

Liang, G., *New Competition: Foreign Direct Investment and Industrial Development in China*, Promotor: Prof.dr. R.J.M. van Tulder, EPS-2004-047-ORG, ISBN: 90-5892-073-9, <http://hdl.handle.net/1765/1795>

Liere, D.W. van, *Network Horizon and the Dynamics of Network Positions: A Multi-Method Multi-Level Longitudinal Study of Interfirm Networks*, Promotor: Prof.dr. P.H.M. Vervest, EPS-2007-105-LIS, ISBN: 90-5892-139-0, <http://hdl.handle.net/1765/10181>

Loef, J., *Incongruity between Ads and Consumer Expectations of Advertising*, Promotors: Prof.dr. W.F. van Raaij & Prof.dr. G. Antonides, EPS-2002-017-MKT, ISBN: 90-5892-028-3, <http://hdl.handle.net/1765/869>

Maeseire, W., de, *Essays on Firm Valuation and Value Appropriation*, Promotor: Prof.dr. J.T.J. Smit, EPS-2005-053-F&A, ISBN: 90-5892-082-8, <http://hdl.handle.net/1765/6768>

Londoño, M. del Pilar, *Institutional Arrangements that Affect Free Trade Agreements: Economic Rationality Versus Interest Groups*, Promotors: Prof.dr. H.E. Haralambides & Prof.dr. J.F. Francois, EPS-2006-078-LIS, ISBN: 90-5892-108-5, <http://hdl.handle.net/1765/7578>

Maas, A.A., van der, *Strategy Implementation in a Small Island Context: An Integrative Framework*, Promotor: Prof.dr. H.G. van Dissel, EPS-2008-127-LIS, ISBN: 978-90-5892-160-4, <http://hdl.handle.net/1765/12278>

Maeseire, W., de, *Essays on Firm Valuation and Value Appropriation*, Promotor: Prof.dr. J.T.J. Smit, EPS-2005-053-F&A, ISBN: 90-5892-082-8, <http://hdl.handle.net/1765/6768>

Mandele, L.M., van der, *Leadership and the Inflection Point: A Longitudinal Perspective*, Promotors: Prof.dr. H.W. Volberda & Prof.dr. H.R. Commandeur, EPS-2004-042-STR, ISBN: 90-5892-067-4, <http://hdl.handle.net/1765/1302>

Meer, J.R. van der, *Operational Control of Internal Transport*, Promotors: Prof.dr. M.B.M. de Koster & Prof.dr.ir. R. Dekker, EPS-2000-001-LIS, ISBN: 90-5892-004-6, <http://hdl.handle.net/1765/859>

Mentink, A., *Essays on Corporate Bonds*, Promotor: Prof.dr. A.C.F. Vorst, EPS-2005-070-F&A, ISBN: 90-5892-100-X, <http://hdl.handle.net/1765/7121>

Meyer, R.J.H., *Mapping the Mind of the Strategist: A Quantitative Methodology for Measuring the Strategic Beliefs of Executives*, Promotor: Prof.dr. R.J.M. van Tulder, EPS-2007-106-ORG, ISBN: 978-90-5892-141-3, <http://hdl.handle.net/1765/10182>

Miltenburg, P.R., *Effects of Modular Sourcing on Manufacturing Flexibility in the Automotive Industry: A Study among German OEMs*, Promotors: Prof.dr. J. Paauwe & Prof.dr. H.R. Commandeur, EPS-2003-030-ORG, ISBN: 90-5892-052-6, <http://hdl.handle.net/1765/1039>

Moerman, G.A., *Empirical Studies on Asset Pricing and Banking in the Euro Area*, Promotor: Prof.dr. C.G. Koedijk, EPS-2005-058-F&A, ISBN: 90-5892-090-9, <http://hdl.handle.net/1765/6666>

Moitra, D., *Globalization of R&D: Leveraging Offshoring for Innovative Capability and Organizational Flexibility*, Promotor: Prof.dr. K. Kumar, EPS-2008-150-LIS, ISBN: 978-90-5892-184-0, <http://hdl.handle.net/1765/14081>

Mol, M.M., *Outsourcing, Supplier-relations and Internationalisation: Global Source Strategy as a Chinese Puzzle*, Promotor: Prof.dr. R.J.M. van Tulder, EPS-2001-010-ORG, ISBN: 90-5892-014-3, <http://hdl.handle.net/1765/355>

Mom, T.J.M., *Managers' Exploration and Exploitation Activities: The Influence of Organizational Factors and Knowledge Inflows*, Promotors: Prof.dr.ir. F.A.J. Van den Bosch & Prof.dr. H.W. Volberda, EPS-2006-079-STR, ISBN: 90-5892-116-6, <http://hdl.handle.net/1765>

Mulder, A., *Government Dilemmas in the Private Provision of Public Goods*, Promotor: Prof.dr. R.J.M. van Tulder, EPS-2004-045-ORG, ISBN: 90-5892-071-2, <http://hdl.handle.net/1765/1790>

Muller, A.R., *The Rise of Regionalism: Core Company Strategies Under The Second Wave of Integration*, Promotor: Prof.dr. R.J.M. van Tulder, EPS-2004-038-ORG, ISBN: 90-5892-062-3, <http://hdl.handle.net/1765/1272>

Nalbantov G.I., *Essays on Some Recent Penalization Methods with Applications in Finance and Marketing*, Promotor: Prof. dr P.J.F. Groenen, EPS-2008-132-F&A, ISBN: 978-90-5892-166-6, <http://hdl.handle.net/1765/13319>

Nguyen, T.T., *Capital Structure, Strategic Competition, and Governance*, Promotor: Prof.dr. A. de Jong, EPS-2008-148-F&A, ISBN: 90-5892-178-9, <http://hdl.handle.net/1765/14005>

Nieuwenboer, N.A. den, *Seeing the Shadow of the Self*, Promotor: Prof.dr. S.P. Kaptein, EPS-2008-151-ORG, ISBN: 978-90-5892-182-6, <http://hdl.handle.net/1765/14223>

Ning, H., *Hierarchical Portfolio Management: Theory and Applications*, Promotor: Prof.dr. J. Spronk, EPS-2007-118-F&A, ISBN: 90-5892-152-9, <http://hdl.handle.net/1765/10868>

Noeverman, J., *Management Control Systems, Evaluative Style, and Behaviour: Exploring the Concept and Behavioural Consequences of Evaluative Style*, Promotors: Prof.dr. E.G.J. Vosselman & Prof.dr. A.R.T. Williams, EPS-2007-120-F&A, ISBN: 90-5892-151-2, <http://hdl.handle.net/1765/10869>

Oosterhout, J., van, *The Quest for Legitimacy: On Authority and Responsibility in Governance*, Promotors: Prof.dr. T. van Willigenburg & Prof.mr. H.R. van Gunsteren, EPS-2002-012-ORG, ISBN: 90-5892-022-4, <http://hdl.handle.net/1765/362>

Paape, L., *Corporate Governance: The Impact on the Role, Position, and Scope of Services of the Internal Audit Function*, Promotors: Prof.dr. G.J. van der Pijl & Prof.dr. H. Commandeur, EPS-2007-111-MKT, ISBN: 90-5892-143-7, <http://hdl.handle.net/1765/10417>

Pak, K., *Revenue Management: New Features and Models*, Promotor: Prof.dr.ir. R. Dekker, EPS-2005-061-LIS, ISBN: 90-5892-092-5, <http://hdl.handle.net/1765/362/6771>

Pattikawa, L.H., *Innovation in the Pharmaceutical Industry: Evidence from Drug Introduction in the U.S.*, Promotors: Prof.dr. H.R. Commandeur, EPS-2007-102-MKT, ISBN: 90-5892-135-2, <http://hdl.handle.net/1765/9626>

Peeters, L.W.P., *Cyclic Railway Timetable Optimization*, Promotors: Prof.dr. L.G. Kroon & Prof.dr.ir. J.A.E.E. van Nunen, EPS-2003-022-LIS, ISBN: 90-5892-042-9, <http://hdl.handle.net/1765/429>

Pietersz, R., *Pricing Models for Bermudan-style Interest Rate Derivatives*, Promotors: Prof.dr. A.A.J. Pelsser & Prof.dr. A.C.F. Vorst, EPS-2005-071-F&A, ISBN: 90-5892-099-2, <http://hdl.handle.net/1765/7122>

Poel, A.M. van der, *Empirical Essays in Corporate Finance and Financial Reporting*, Promotors: Prof.dr. A. de Jong & Prof.dr. G.M.H. Mertens, EPS-2007-133-F&A, ISBN: 978-90-5892-165-9, <http://hdl.handle.net/1765/13320>

Popova, V., *Knowledge Discovery and Monotonicity*, Promotor: Prof.dr. A. de Bruin, EPS-2004-037-LIS, ISBN: 90-5892-061-5, <http://hdl.handle.net/1765/1201>

Pouchkarev, I., *Performance Evaluation of Constrained Portfolios*, Promotors: Prof.dr. J. Spronk & Dr. W.G.P.M. Hallerbach, EPS-2005-052-F&A, ISBN: 90-5892-083-6, <http://hdl.handle.net/1765/6731>

Prins, R., *Modeling Consumer Adoption and Usage of Value-Added Mobile Services*, Promotors: Prof.dr. Ph.H.B.F. Franses & Prof.dr. P.C. Verhoef, EPS-2008-128-MKT, ISBN: 978/90-5892-161-1, <http://hdl.handle.net/1765/12461>

Puvasanvari Ratnasingam, P., *Interorganizational Trust in Business to Business E-Commerce*, Promotors: Prof.dr. K. Kumar & Prof.dr. H.G. van Dissel, EPS-2001-009-LIS, ISBN: 90-5892-017-8, <http://hdl.handle.net/1765/356>

Quak, H.J., *Sustainability of Urban Freight Transport: Retail Distribution and Local Regulation in Cities*, Promotor: Prof.dr. M.B.M. de Koster, EPS-2008-124-LIS, ISBN: 978-90-5892-154-3, <http://hdl.handle.net/1765/11990>

Quariguasi Frota Nero, J., *Eco-efficient Supply Chains for Electrical and Electronic Products*, Promotors: Prof.dr.ir. J.A.E.E. van Nunen & Prof.dr.ir. H.W.G.M. van Heck, EPS-2008-152-LIS, ISBN: 978-90-5892-192-5, <http://hdl.handle.net/1765/1>

Radkevitch, U.L., *Online Reverse Auction for Procurement of Services*, Promotor: Prof.dr.ir. H.W.G.M. van Heck, EPS-2008-137-LIS, ISBN: 978-90-5892-171-0, <http://hdl.handle.net/1765/13497>

Rinsum, M. van, *Performance Measurement and Managerial Time Orientation*, Promotor: Prof.dr. F.G.H. Hartmann, EPS-2006-088-F&A, ISBN: 90-5892-121-2, <http://hdl.handle.net/1765/7993>

Romero Morales, D., *Optimization Problems in Supply Chain Management*, Promotors: Prof.dr.ir. J.A.E.E. van Nunen & Dr. H.E. Romeijn, EPS-2000-003-LIS, ISBN: 90-9014078-6, <http://hdl.handle.net/1765/865>

Roodbergen, K.J., *Layout and Routing Methods for Warehouses*, Promotors: Prof.dr. M.B.M. de Koster & Prof.dr.ir. J.A.E.E. van Nunen, EPS-2001-004-LIS, ISBN: 90-5892-005-4, <http://hdl.handle.net/1765/861>

Rook, L., *Imitation in Creative Task Performance*, Promotor: Prof.dr. D.L. van Knippenberg, EPS-2008-125-ORG, <http://hdl.handle.net/1765/11555>

Samii, R., *Leveraging Logistics Partnerships: Lessons from Humanitarian Organizations*, Promotors: Prof.dr.ir. J.A.E.E. van Nunen & Prof.dr.ir. L.N. Van Wassenhove, EPS-2008-153-LIS, ISBN: 978-90-5892-186-4, <http://hdl.handle.net/1765/1>

Schaik, D. van, *M&A in Japan: An Analysis of Merger Waves and Hostile Takeovers*, Promotors: Prof.dr. J. Spronk & Prof.dr. J.P.M. Groenewegen, EPS-2008-141-F&A, ISBN: 978-90-5892-169-7, <http://hdl.handle.net/1765/13693>

Schauten, M.B.J., *Valuation, Capital Structure Decisions and the Cost of Capital*, Promotors: Prof.dr. J. Spronk & Prof.dr. D. van Dijk, EPS-2008-134-F&A, ISBN: 978-90-5892-172-7, <http://hdl.handle.net/1765/13480>

Schramade, W.L.J., *Corporate Bonds Issuers*, Promotor: Prof.dr. A. De Jong, EPS-2006-092-F&A, ISBN: 90-5892-125-5, <http://hdl.handle.net/1765/8191>

Schweizer, T.S., *An Individual Psychology of Novelty-Seeking, Creativity and Innovation*, Promotor: Prof.dr. R.J.M. van Tulder, EPS-2004-048-ORG, ISBN: 90-5892-077-1, <http://hdl.handle.net/1765/1818>

Six, F.E., *Trust and Trouble: Building Interpersonal Trust Within Organizations*, Promotors: Prof.dr. B. Nooteboom & Prof.dr. A.M. Sorge, EPS-2004-040-ORG, ISBN: 90-5892-064-X, <http://hdl.handle.net/1765/1271>

Slager, A.M.H., *Banking across Borders*, Promotors: Prof.dr. R.J.M. van Tulder & Prof.dr. D.M.N. van Wensveen, EPS-2004-041-ORG, ISBN: 90-5892-066-6, <http://hdl.handle.net/1765/1301>

Sloot, L., *Understanding Consumer Reactions to Assortment Unavailability*, Promotors: Prof.dr. H.R. Commandeur, Prof.dr. E. Peelen & Prof.dr. P.C. Verhoef, EPS-2006-074-MKT, ISBN: 90-5892-102-6, <http://hdl.handle.net/1765/7438>

Smit, W., *Market Information Sharing in Channel Relationships: Its Nature, Antecedents and Consequences*, Promotors: Prof.dr.ir. G.H. van Bruggen & Prof.dr.ir. B. Wierenga, EPS-2006-076-MKT, ISBN: 90-5892-106-9, <http://hdl.handle.net/1765/7327>

Sonnenberg, M., *The Signalling Effect of HRM on Psychological Contracts of Employees: A Multi-level Perspective*, Promotor: Prof.dr. J. Paauwe, EPS-2006-086-ORG, ISBN: 90-5892-119-0, <http://hdl.handle.net/1765/7995>

Speklé, R.F., *Beyond Generics: A closer Look at Hybrid and Hierarchical Governance*, Promotor: Prof.dr. M.A. van Hoepen RA, EPS-2001-008-F&A, ISBN: 90-5892-011-9, <http://hdl.handle.net/1765/357>

Stam, D.A., *Managing Dreams and Ambitions: A Psychological Analysis of Vision Communication*, Promotor: Prof.dr. D.L. van Knippenberg, EPS-2008-149-ORG, <http://hdl.handle.net/1765/14080>

Stienstra, M., *Strategic Renewal in Regulatory Environments: How Inter- and Intra-organisational Institutional Forces Influence European Energy Incumbent Firms*, Promotors: Prof.dr.ir. F.A.J. Van den Bosch & Prof.dr. H.W. Volberda, EPS-2008-145-STR, ISBN: 978-90-5892-184-0, <http://hdl.handle.net/1765/13943>

Szkudlarek, B.A., *Spinning the Web of Reentry: [Re]connecting reentry training theory and practice*, Promotor: Prof.dr. S.J. Magala, EPS-2008-143-ORG, ISBN: 978-90-5892-177-2, <http://hdl.handle.net/1765/13695>

Teunter, L.H., *Analysis of Sales Promotion Effects on Household Purchase Behavior*, Promoters: Prof.dr.ir. B. Wierenga & Prof.dr. T. Kloek, EPS-2002-016-MKT, ISBN: 90-5892-029-1, <http://hdl.handle.net/1765/868>

Tims, B., *Empirical Studies on Exchange Rate Puzzles and Volatility*, Promotor: Prof.dr. C.G. Koedijk, EPS-2006-089-F&A, ISBN: 90-5892-113-1, <http://hdl.handle.net/1765/8066>

Tuk, M.A., *Is Friendship Silent When Money Talks? How Consumers Respond to Word-of-Mouth Marketing*, Promoters: Prof.dr.ir. A. Smidts & Prof.dr. D.H.J. Wigboldus, EPS-2008-130-MKT, ISBN: 978-90-5892-164-2, <http://hdl.handle.net/1765/12702>

Valck, K. de, *Virtual Communities of Consumption: Networks of Consumer Knowledge and Companionship*, Promoters: Prof.dr.ir. G.H. van Bruggen & Prof.dr.ir. B. Wierenga, EPS-2005-050-MKT, ISBN: 90-5892-078-X, <http://hdl.handle.net/1765/6663>

Valk, W. van der, *Buyer-Seller Interaction Patterns During Ongoing Service Exchange*, Promoters: Prof.dr. J.Y.F. Wynstra & Prof.dr.ir. B. Axelsson, EPS-2007-116-MKT, ISBN: 90-5892-146-8, <http://hdl.handle.net/1765/10856>

Valk, W. van der, *Buyer-Seller Interaction Patterns During Ongoing Service Exchange*, Promoters: Prof.dr. J.Y.F. Wynstra & Prof.dr.ir. B. Axelsson, EPS-2007-116-MKT, ISBN: 90-5892-146-8, <http://hdl.handle.net/1765/10856>

Verheul, I., *Is There a (Fe)male Approach? Understanding Gender Differences in Entrepreneurship*, Prof.dr. A.R. Thurik, EPS-2005-054-ORG, ISBN: 90-5892-080-1, <http://hdl.handle.net/1765/2005>

Verwijmeren, P., *Empirical Essays on Debt, Equity, and Convertible Securities*, Promoters: Prof.dr. A. de Jong & Prof.dr. M.J.C.M. Verbeek, EPS-2009-154-F&A, ISBN: 978-90-5892-187-1, <http://hdl.handle.net/1765/14312>

Vis, I.F.A., *Planning and Control Concepts for Material Handling Systems*, Promoters: Prof.dr. M.B.M. de Koster & Prof.dr.ir. R. Dekker, EPS-2002-014-LIS, ISBN: 90-5892-021-6, <http://hdl.handle.net/1765/866>

Vlaar, P.W.L., *Making Sense of Formalization in Interorganizational Relationships: Beyond Coordination and Control*, Promoters: Prof.dr.ir. F.A.J. Van den Bosch & Prof.dr. H.W. Volberda, EPS-2006-075-STR, ISBN 90-5892-103-4, <http://hdl.handle.net/1765/7326>

Vliet, P. van, *Downside Risk and Empirical Asset Pricing*, Promotor: Prof.dr. G.T. Post, EPS-2004-049-F&A, ISBN: 90-5892-07-55, <http://hdl.handle.net/1765/1819>

Vlist, P. van der, *Synchronizing the Retail Supply Chain*, Promoters: Prof.dr.ir. J.A.E.E. van Nunen & Prof.dr. A.G. de Kok, EPS-2007-110-LIS, ISBN: 90-5892-142-0, <http://hdl.handle.net/1765/10418>

Vries-van Ketel E. de, *How Assortment Variety Affects Assortment Attractiveness: A Consumer Perspective*, Promoters: Prof.dr. G.H. van Bruggen & Prof.dr.ir. A. Smidts, EPS-2006-072-MKT, ISBN: 90-5892-101-8, <http://hdl.handle.net/1765/7193>

Vromans, M.J.C.M., *Reliability of Railway Systems*, Promoters: Prof.dr. L.G. Kroon, Prof.dr.ir. R. Dekker & Prof.dr.ir. J.A.E.E. van Nunen, EPS-2005-062-LIS, ISBN: 90-5892-089-5, <http://hdl.handle.net/1765/6773>

Vroomen, B.L.K., *The Effects of the Internet, Recommendation Quality and Decision Strategies on Consumer Choice*, Promotor: Prof.dr. Ph.H.B.F. Franses, EPS-2006-090-MKT, ISBN: 90-5892-122-0, <http://hdl.handle.net/1765/8067>

Waal, T. de, *Processing of Erroneous and Unsafe Data*, Promotor: Prof.dr.ir. R. Dekker, EPS-2003-024-LIS, ISBN: 90-5892-045-3, <http://hdl.handle.net/1765/870>

Watkins Fassler, K., *Macroeconomic Crisis and Firm Performance*, Promotors: Prof.dr. J. Spronk & Prof.dr. D.J. van Dijk, EPS-2007-103-F&A, ISBN: 90-5892-138-3, <http://hdl.handle.net/1765/10065>

Wennekers, A.R.M., *Entrepreneurship at Country Level: Economic and Non-Economic Determinants*, Promotor: Prof.dr. A.R. Thurik, EPS-2006-81-ORG, ISBN: 90-5892-115-8, <http://hdl.handle.net/1765/7982>

Wielemaker, M.W., *Managing Initiatives: A Synthesis of the Conditioning and Knowledge-Creating View*, Promotors: Prof.dr. H.W. Volberda & Prof.dr. C.W.F. Baden-Fuller, EPS-2003-28-STR, ISBN: 90-5892-050-X, <http://hdl.handle.net/1765/1042>

Wijk, R.A.J.L. van, *Organizing Knowledge in Internal Networks: A Multilevel Study*, Promotor: Prof.dr.ir. F.A.J. van den Bosch, EPS-2003-021-STR, ISBN: 90-5892-039-9, <http://hdl.handle.net/1765/347>

Wolters, M.J.J., *The Business of Modularity and the Modularity of Business*, Promotors: Prof. mr. dr. P.H.M. Vervest & Prof. dr. ir. H.W.G.M. van Heck, EPS-2002-011-LIS, ISBN: 90-5892-020-8, <http://hdl.handle.net/1765/920>

Yang, J., *Towards the Restructuring and Co-ordination Mechanisms for the Architecture of Chinese Transport Logistics*, Promotor: Prof.dr. H.E. Harlambides, EPS-2009-157-LIS, ISBN: 978-90-5892-198-7, <http://hdl.handle.net/1765/1>

Yu, M., *Enhancing Warehouse Performance by Efficient Order Picking*, Promotor: Prof.dr. M.B.M. de Koster, EPS-2008-139-LIS, ISBN: 978-90-5892-167-3, <http://hdl.handle.net/1765/13691>

Zhang, X., *Strategizing of Foreign Firms in China: An Institution-based Perspective*, Promotor: Prof.dr. B. Krug, EPS-2007-114-ORG, ISBN: 90-5892-147-5, <http://hdl.handle.net/1765/10721>

Zhu, Z., *Essays on China's Tax System*, Promotors: Prof.dr. B. Krug & Prof.dr. G.W.J. Hendrikse, EPS-2007-112-ORG, ISBN: 90-5892-144-4, <http://hdl.handle.net/1765/10502>

Zwart, G.J. de, *Empirical Studies on Financial Markets: Private Equity, Corporate Bonds and Emerging Markets*, Promotors: Prof.dr. M.J.C.M. Verbeek & Prof.dr. D.J.C. van Dijk, EPS-2008-131-F&A, ISBN: 978-90-5892-163-5, <http://hdl.handle.net/1765/12703>

SEGMENTATION AND DIMENSION REDUCTION EXPLORATORY AND MODEL-BASED APPROACHES

Representing the information in a data set in a concise way is an important part of data analysis. A variety of multivariate statistical techniques have been developed for this purpose, such as *k*-means clustering and principal components analysis. These techniques are often based on the principles of segmentation (partitioning the observations into distinct groups) and dimension reduction (constructing a low-dimensional representation of a data set). However, such techniques typically make no statistical assumptions on the process that generates the data; as a result, the statistical significance of the results is often unknown.

In this thesis, we incorporate the modeling principles of segmentation and dimension reduction into statistical models. We thus develop new models that can summarize and explain the information in a data set in a simple way. The focus is on dimension reduction using bilinear parameter structures and techniques for clustering both modes of a two-mode data matrix. To illustrate the usefulness of the techniques, the thesis includes a variety of empirical applications in marketing, psychometrics, and political science. An important application is modeling the response behavior in surveys with rating scales, which provides novel insight into what kinds of response styles exist, and how substantive opinions vary among respondents. We find that our modeling approaches yield new techniques for data analysis that can be useful in a variety of applied fields.

ERIM

The Erasmus Research Institute of Management (ERIM) is the Research School (Onderzoeksschool) in the field of management of the Erasmus University Rotterdam. The founding participants of ERIM are Rotterdam School of Management (RSM), and the Erasmus School of Economics (ESE). ERIM was founded in 1999 and is officially accredited by the Royal Netherlands Academy of Arts and Sciences (KNAW). The research undertaken by ERIM is focussed on the management of the firm in its environment, its intra- and interfirm relations, and its business processes in their interdependent connections.

The objective of ERIM is to carry out first rate research in management, and to offer an advanced doctoral programme in Research in Management. Within ERIM, over three hundred senior researchers and PhD candidates are active in the different research programmes. From a variety of academic backgrounds and expertises, the ERIM community is united in striving for excellence and working at the forefront of creating new business knowledge.



ERIM PhD Series Research in Management

Erasmus Research Institute of Management - ERIM
Rotterdam School of Management (RSM)
Erasmus School of Economics (ESE)
P.O. Box 1738, 3000 DR Rotterdam
The Netherlands

Tel. +31 10 408 11 82
Fax +31 10 408 96 40
E-mail info@erim.eur.nl
Internet www.erim.eur.nl