Methods for Modelling Response Styles

Methods for Modelling Response Styles

Methoden voor de Modellering van Responsstijlen

Thesis

to obtain the degree of Doctor from the
Erasmus University Rotterdam
by command of the
rector magnificus

Prof.dr. H.A.P. Pols

and in accordance with the decision of the Doctorate Board

The public defense shall be held on

Tuesday, 15 December 2015 at 15:30 hrs

by

Pieter Cornelis Schoonees
born in Stellenbosch, South Africa.

Erasmus University Rotterdam

Doctoral Committee:

Promotor:	Prof.dr. P.J.F. Groenen
Committee members:	Prof.dr. D. Fok
	Prof.dr. F.A. van Eeuwijk
	Prof.dr.ing. P.H.C. Eilers
Co-promotor:	Dr. M. van de Velden

Erasmus Research Institute of Management - ERIM
The joint research institute of the Rotterdam School of Management (RSM)
and the Erasmus School of Economics (ESE) at the Erasmus University Rotterdam
Internet: http://www.erim.eur.nl

ERIM Electronic Series Portal: `http://repub.eur.nl/pub`

ERIM PhD Series in Research in Management, 348
ERIM reference number: EPS–2015–348–MKT
ISBN 978–90–5892–431–5

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

This publication (cover and interior) is printed by haveka.nl on recycled paper, Revive®.
The ink used is produced from renewable resources and alcohol free fountain solution.
Certifications for the paper and the printing production process: Recycle, EU Flower, FSC, ISO14001.
More info: http://www.haveka.nl/greening

For Lindy

Acknowledgements

The journey that takes another turn with the defense of this thesis would not have been possible, or remotely as pleasant, if it were not for quite a few people. I want to thank Patrick Groenen and Michel van de Velden, my promotor and co-promotor respectively, for giving me the opportunity to pursue my PhD in Rotterdam. You have proven to be very pleasant collaborators and mentors, also willing to provide non-academic advice to an African lost in your cold and flat country. The academic knowledge and experience you have imparted to me over the past four plus years are extremely valuable, yet I also appreciate that adjusting to life in Rotterdam could have been much harder without your support in this regard. I appreciate the time and ideas you have both given to this thesis, and I am grateful for having learnt a lot from your reasoning, enthusiasm, networking, management and optimism. It is my hope that we can continue to collaborate in the future.

I am very much indebted to Niël le Roux from the University of Stellenbosch for his inspiration in my pursuit of mathematical statistics. It was by his doing that I was able to obtain a bursary for my masters studies through Sasol, and it was also Niël that encouraged me to apply for the PhD position that culminated in this thesis. I strongly suspect that his recommendation to Patrick is mainly what made all this possible. Thank you for your time, support, advice, collaboration, and the various dinners and excursions at many conferences. Long may it continue.

Thank you to Carolin Strobl for making me feel very welcome at the Department of Psychology (Psychological Methods, Evaluation and Statistics) at the Universität Zürich for three months during the summer and fall of 2014. I had a tremendous time working in your department (and on the Zürichberg), and even though our work is not included in this thesis, it inspired Chapter 4. I expect that building on this work will lead to very nice publications in the future. Having yourself, Nina, Michel, Matt, Christina and Marina as colleagues was very pleasant.

Thank you to Fred van Eeuwijk, Paul Eilers and Dennis Fok for agreeing to serve on the inner committee for my thesis, and to Eva Ceulemans, Hester van Herk and Ale Smidts for serving on the larger committee. A special thanks also to Hester and Joost van Rosmalen, whose collaboration made Chapter 5 possible. I also thank my new colleagues at the Department of Marketing Management at the Rotterdam School of Management, and in particular Ale, for welcoming me in their midst. Furthermore, I am indebted to Trevor and Jeff for initiating the research in Appendix C of Chapter 2, for providing the data used in that chapter, and for two pleasant visits to the United Kingdom.

Many people at the Econometric Institute, the Erasmus School of Economics, the Erasmus Research Institute of Management, and the Rotterdam School of Management contributed towards making my time as a PhD candidate an enjoyable experience. For this, I want to thank (in no particular order): Frank, Gertjan, Koen, Ilse, Kristiaan, Zahra, Judith, Mathijn, Harwin, Twan, Remy, Willem, Thomas, Diana, Riccardo, Panagiotis, Ronald, Andreas, Murat, Nufer, all the other PhD candidates I crossed paths with, as well as the support staff.

My parents, Koos and Anelene, have always supported me with time and sacrifice, and for this I will always be indebted to them. I value your example and advice, and am acutely aware that I would not be where I am today without the springboard you have provided me. Thank you to my parents-in-law, Jacques and Patricia, for unendingly and unconditionally supporting Lindy and I from all those kilometres away. I also want to thank our friends and other family members (sisters, grandmothers, and extended family) for their support and warm reception whenever we got a chance to visit South Africa. Thank you for supporting our endeavours even if it meant that we had to move to a different continent and that we would not be seeing each other very much. It is extremely heartening to have friendships which can withstand the test of time and space in this way. Last but certainly not least, a special thanks also to our (geographically) nearest-neighbour friends, especially the Kraaije and Prinsloo's.

I dedicate this thesis to my wife, Lindy. None of this would have been possible without you. It was tough being apart during 2011 and 2012 when you were still living in Stellenbosch, but we persevered and I am very happy that we did. Thank you for your sacrifice in leaving Africa, Blue Horizon Bay, Stellenbosch, friends and family behind to start anew with me in Rotterdam. Thank you for believing in us and for sharing your life with me. We have already accumulated so many happy memories together; here is to many more (without a thesis to write)!

Soli deo gloria.

Pieter C. Schoonees
Rotterdam, July 2015

Table of Contents

CHAPTER 1

Introduction

This dissertation is concerned with the study of categorical data, and, more specifically, those that arise from surveys of human respondents. It is primarily a study of new statistical methods and algorithms for the analysis of such data (Chapters 2 to 4), but also included is an empirical study utilizing an existing advanced statistical model (Chapter 5). While many monographs are devoted to the design of such surveys, such as Lohr (1999), the human element means that even the most well-designed study can suffer from unobserved heterogeneity arising from individual differences between respondents.

One such problem is the occurrence of response styles, a main theme of the studies included here. Response styles arise because individuals use rating scales, most often Likert (1932) scales, differently. An often cited definition, attributed to Paulhus (1991), is that a response style is a systematic tendency to respond to survey items on some basis other than what the questions were designed to measure. To illustrate what the problem is, consider this example taken from the European Social Survey:

> How much do you agree or disagree with each of the following statements?
>
> Q1: *In general I feel very positive about myself.*
>
> Q2: *I'm always optimistic about my future.*

In order to measure the responses on a unified scale, the researcher will typically specify a rating scale which the respondent is obliged to use in answering the question. This is a constraint imposed by our limited measurement ability: there is no way for us to objectively measure opinion. We simply have to ask for it, which introduces an element of subjectivity.

In our example, the answer sheet may look like this:

	1 – Strongly agree	2 – Agree	3 – Neutral	4 – Disagree	5 – Strongly disagree
Q1	☐	☐	☐	☐	☐
Q2	☐	☐	☐	☐	☐

The response style problem is that there is heterogeneity in how persons map their opinions to the given rating scale. Suppose that two persons strongly agree with the first statement. The first person maps this opinion to the expected rating category, namely “1 – Strongly agree.” However, the second person exhibits a tendency to use only the middle three rating categories – a response style known as ‘midpoint scoring’ (more on this and other styles in Chapter 2). Consequently, this person, being adverse to using the extremes of the rating scale, answers the question with “2 – Agree.” Hence, even though these persons had exactly the same opinion, they have arrived at different answers because of the difference in response style. Concluding that these two different answers are driven solely by a difference in opinion obfuscates the truth.

There is growing evidence that response styles are widespread in survey data, as the present study also shows. However, ordinary, established statistical analyses take the meaning of the rating categories in survey items at face value. It is by far the simplest approach, and is perfectly justifiable in situations where categorical data arise naturally, such as when the political party that a person voted for is recorded. However, in surveys, where we have to ask for a person’s unobservable opinion, the heterogeneity introduced by different interpretations of the rating scale can contaminate the conclusions drawn from standard statistical analyses.

In this thesis, statistical methods and algorithms, which account for differences in response styles, are formulated and evaluated. A number of principles underly these methods. First, we can learn about what different rating categories mean to a specific individual by looking at how he used the rating scale across multiple questions. If, for the second person in the above example, we observe that he never uses rating category ‘1 – Strongly agree’ across a set of diverse questions, we can conclude that this rating category probably corresponds to a higher level of agreeance than the same category for the first hypothetical individual. The task then becomes how to quantify the difference in meaning.

A second principle is that if we can identify groups of individuals who exhibit the same response styles, we can apply standard analyses within each of these groups. This is because response styles are only a problem when individual differences are present. If all persons in the above survey example exhibits midpoint scoring similar to the second person, we can safely assume that the rating categories mean the same thing for all persons and proceed with standard methods. The question is then how to identify these groups, and how we can determine how many of these groups there are.

This dissertation makes methodological and empirical contributions to the literature on modelling response styles, related areas such as psychometrics, and potentially to any area utilizing rating data. The main contributions of the four substantive chapters can be summarized as follows:

Chapter 2 proposes a new numerical method for detecting and purging response styles;

Chapter 3 formulates a fast least-squares method for estimating the latent-class bilinear multinomial logit (LC-BML) model for response style detection, and extends this to more general bilinear decompositions coupled with crisp clustering;

Chapter 4 implements a novel algorithm for estimating the rating scale model (RSM), a simple but powerful item response theory (IRT) model for analyzing rating data; and

Chapter 5 applies the LC-BML model to the Dutch political landscape, uncovering an array of response styles in data from the European Social Survey.

These chapters are accompanied by software packages for R (R Core Team, 2015), a free software environment for statistical computing and graphics. I strongly believe in open science and reproducibility of research, and hence this software aims to allow the work in this thesis to be disseminated and reproduced as far as possible. The **cds** package (Schoonees, 2015b) for Chapter 2 is available for download from the Comprehensive R Archive Network (CRAN) at `http://cran.r-project.org/web/packages/cds`. An unpublished accompanying package, **muk**, implements the incomplete constrained dual scaling routines outlined in Appendix C of Chapter 2. These routines will be merged into **cds** in due course. Chapter 3 is accompanied by the CRAN package **lsbclust** (Schoonees, 2015a), available at `http://CRAN.R-project.org/package=lsbclust`. A package implementing the routines of Chapter 4, currently named **polymix**, is available on request. Releasing this code too on CRAN is planned once other related algorithms have been added.

The next and final section of this chapter elaborates on the content of Chapters 2 to 5.

1.1 Chapter Synopsis

Chapter 2, based on Schoonees et al. (2015b) and co-authored with Michel van de Velden and Patrick Groenen, introduces an optimal scaling method for detecting and purging response styles. The method assigns optimal scores (numerical values) to the rating categories, which quantifies how similar the rating categories are. Multiple response styles are allowed for by identifying groups of individuals with similar rating scale use. Within each group, different optimal scores are assigned. The method relies on a special characteristic of dual scaling for successive categories data (Nishisato, 1980), which is closely related to correspondence analysis (e.g, Greenacre, 2007). We show by simulations under which circumstances the method works best, and that the optimal scores derived from the method can be used to obtain a version of the original data set which is purged of the response styles. The method is applied to sensory data to illustrate it's practical application.

The topic of Chapter 3 (Schoonees et al., 2015a), co-authored with Patrick Groenen and Michel van de Velden, is the analysis of three-way data. Three-way data arise when each observation in the data set is a matrix, and can be visualized as a stack of such matrices. An example of three-way data arise in consumer studies when consumers assess a set of products on a set of attributes. For each consumer, we observe the ratings of each of the products, e.g. chocolate bars, on a range of attributes, such as colour, texture, firmness, melt rate and taste. Such three-way data also arise in many other contexts, such as in longitudinal studies. Our method combines cluster analysis with two-way matrix decompositions, and can be interpreted with simple graphical displays. As such it is a useful alternative to the multi-way decompositions available in the literature. A special case of our method, where two-way data are transformed to three-way data by adding the rating scale, can be used to detect response styles. This version of our method can be seen as a nonparametric alternative to the latent-class bilinear multinomial logit (LC-BML) model (Van Rosmalen et al., 2010), an advanced and computationally intensive method for analyzing data while accounting for response styles.

In Chapter 4, a new algorithm is formulated in collaboration with Patrick Groenen for estimating the rating scale model (RSM; Andrich, 1978a), an important item response theory (IRT) model from the psychometric literature. This RSM models all responses to a set of items by three sets of parameters: person parameters describing the person's location on a latent continuum, such as political orientation, that we are trying to measure; item parameters indicating how difficult an item is; and rating category parameters for the rating scale used. We derive a majorization algorithm (see De Leeuw, 1994, for example) which maximizes the joint maximum likelihood formulation of the model by iterative least squares operations. The advantages of this approach is that missing observations can be handled very easily, and that regularization methods, which penalize the parameters in the model, can be incorporated very easily, compared to existing algorithms. Regularization methods can be used to improve the predictive performance of these models and to overcome identifiability issues which arise relatively often in these methods. The algorithm can also be applied to other IRT models, and it provides a building block for the development of models that account for different subgroups in the data by finite mixture modelling.

Chapter 5, co-authored with Hester van Herk, Patrick Groenen and Joost van Rosmalen, concerns an application of the LC-BML model to five waves of Dutch data from the European Social Survey. We use data from roughly 9500 Dutch persons collected in 2002, 2004, 2006, 2008 and 2010 respectively, to identify segments of individuals in the Dutch population who have similar human value preferences. We then relate these segments to which political parties persons voted for, which gives insight into what kinds of voters are attracted to which parties, and how this attraction evolved over time. For example, a subset of the Dutch population considers tradition to be a very important value, while other segments exist which do not think tradition is important at all. We evaluate 480 different instances of the LC-BML model in order to select the model which best describes the data, a computationally intensive task. Seven different value segments are identified, as well as 20 different response styles. This provides more evidence of response style contamination in survey data, and by relating the value segments to political parties we gain insight into which type of individuals are attracted to which parties.

CHAPTER 2

Constrained Dual Scaling for Detecting Response Styles in Categorical Data

2.1 Introduction

A major issue in questionnaire-based research is the presence of response styles. A response style, sometimes also known as response bias or scale usage heterogeneity, can be described as systematic bias due to a respondent's tendency to respond to survey items regardless of its content (Van Rosmalen et al., 2010). Paraphrasing, a response style is the manner in which a person uses a rating scale, an example being extreme response style where the respondent, *for no substantial reason,* prefers to use the endpoints of the Likert scale more often than the intermediate rating categories.

Response styles can invalidate statistical analyses since they are completely confounded with the substantial information contained in the data and hence biases results in non-trivial ways (Baumgartner and Steenkamp, 2001). The problem manifests itself when different respondents resort to different response styles within the same data set. Advanced methods, such as the latent-class multinomial logit model of Van Rosmalen et al. (2010), the multidimensional ordinal IRT model of De Jong and Steenkamp (2010), or the ordinal regression model with heterogeneous thresholds of Johnson (2003), have been developed to deal with the data analysis when response style contamination is relevant. None of these appear to have achieved much popularity in practice.

Existing models often require a substantial investment of resources for its implementation, estimation and/or interpretation. As an alternative, the method presented in this paper results in a data set cleaned of the effects of response styles so that any analyses appropriate for the continuous nature of this cleaned data can be conducted. Furthermore, this method has three

This chapter is based on Schoonees et al. (2015b).

additional purposes, namely to (i) determine whether different response styles are present in categorical data; (ii) identify the respondents associated with each response style; and to (iii) classify the identified response styles into four different types. Software which implements the method in the R software environment (R Core Team, 2015) are available from the first author.

The proposed method is a variant of dual scaling (DS) for rating data (Nishisato, 980a), also referred to as successive categories data in the DS literature. DS is an exploratory multivariate method, akin to correspondence analysis or CA (e.g. Greenacre, 2007). In the special case of rating data, DS however differs from CA in a manner that implicitly caters for response styles by including parameters for the Likert scale categories in an innovative way. These parameters allow for the detection of frequent (or infrequent) usage of certain ratings since the optimal scores assigned by DS to these parameters depend on how often each rating occurs in the data. The new method builds on this aspect of DS by including monotone spline functions to model the response styles and by allowing for multiple response styles through latent classes.

The literature on response styles (also known as scale-usage bias or heterogeneity) can be traced back at least to the work of Cronbach in the 1940's (e.g. Cronbach, 1941, 1942, 1946, 1950). For an overview of the early work, see for example Rorer (1965). A more recent set of references can be found in Baumgartner and Steenkamp (2001). Krosnick (1999) discuss the origins of response styles as a shift in the procedure whereby a response is formulated; this is also known as satisficing in the literature (e.g. Krosnick, 1991). The use of so-called personal equations with double coding, as known in the French school of CA, is a related method of dealing with differences in the interpretation of rating scales at the respondent level (e.g. Benzécri, 1992; Murtagh, 2005).

The next section focuses on a closer discussion of response styles. Section 2.3 introduces spline functions for modelling response styles, explains the new methodology and details an alternating least squares algorithm for solving an extended version of the dual scaling problem. A simulation study is conducted in Section 2.4 to assess the strengths and weaknesses of the method. Finally, an application (Section 2.5) is presented.

2.2 Overview of Response Styles

It is assumed that the process of formulating a response to a survey item requires the respondent to map a latent opinion, preference or some similar concept to a Likert scale. For example, the respondent may be asked how much she agrees with a certain statement using a scale with categories ranging from "1 – Totally Disagree" to "5 – Totally Agree." During the cognitive process of formulating the answer, the respondent first forms an opinion about the survey item and subsequently needs to decide how to transform or map this opinion to the presented rating scale (see for example Weijters and Baumgartner, 2012). The mathematical properties of this response mapping from the latent to the Likert scale determines whether a respondent exhibits a response style or not.

Specifically, a response style can be defined as a monotone nonlinear response mapping (Van de Velden, 2007). If this transformation is linear, no response style is present. Consequently, once a method is available to estimate response mappings the presence of response styles can be assessed by looking at the curvature properties of the estimated mappings. These steps are carried out in subsequent sections. In the case where Likert scales are used these transformations are step functions, but for the moment it is more intuitive to consider continuous transformations.

Four different response styles are considered here, as depicted in Figure 2.1. This figure shows different possible inverse mappings from the rating supplied by the respondent on the horizontal axis to the respondent's true latent opinion on the vertical axis. The inverse transformations are shown since these must be estimated from the observed data.

The different styles can be characterized by which parts of the latent opinion scale is stretched and which parts are shrunk. These are shown by the rug plots on the respected axes in Figure 2.1. For ease of exposition it is assumed here that the true latent opinion comes from a uniform distribution. The rug on the horizontal axis partitions the axis into intervals of equal length, with each interval receiving a rating on the Likert scale. Here a seven-point scale is employed. The rug on the vertical axis shows the effect that the response style transformation has on the intervals of equal length. Hence these transformations characterize the following four response styles:

- *Acquiescence* (ARS) shrinks the lower part of the latent scale and stretches the upper part indicating that higher ratings are favoured (panel (a));

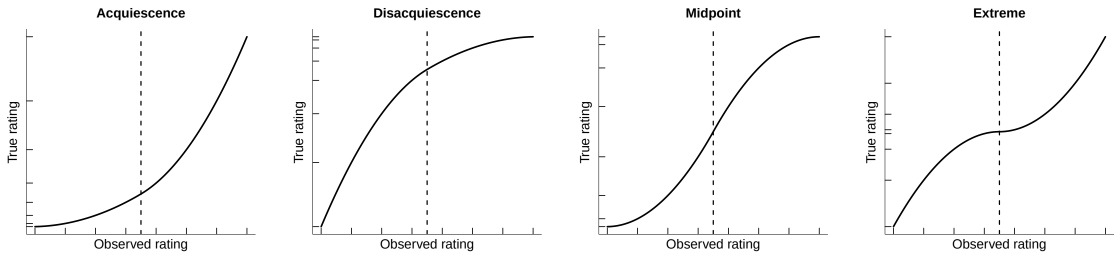
- *Disacquiescence* (DRS) in contrast favours lower ratings by stretching and shrinking the lower and upper parts of the latent scale respectively (panel (b));

- *Midpoint responding* (MRS) reflects a tendency to frequent the middle categories of the rating scale (panel (c)); and

- *Extreme responding* (ERS) in contrast means that the endpoints of the rating scale is used more often than the middle categories (panel (d)).

A critical concept is that the boundaries dividing the latent preference scale into the different rating categories, that is the tick marks on the vertical axes in Figure 2.1, determines which response style is present. If these boundaries are equally spaced, no response style is present. Any significant deviations however give cause to believe that a response style is present.

The methodology outlined in the next section makes use of these boundaries to provide an estimate of the response mappings of groups of individuals.



(a) (b) (c) (d)

Figure 2.1: Examples of (inverse) response style functions mapping the true item content scale (vertical axis) into the observed measurement scale (horizontal axis).

2.3 Methodology

Consider the situation where a set of m objects or survey items are being rated on a q-point Likert scale, enumerated as 1 to q. Due to the ordinality this is sometimes known as successive categories data (Nishisato, 1980, 1994). It is supposed that n individuals are asked to rate the objects according to their preference. Objects may receive equal ratings, and it is assumed that there exists a fixed but unknown preference structure for the set of objects, such as a population mean. Let $\mathbf{X}$ denote the $n \times m$ data matrix. Note that the method detailed below requires all items to use a common rating scale.

The next subsection discusses using dual scaling for analysing successive categories data in general, making use of the method's relationship with correspondence analysis. Monotone quadratic splines for modelling response styles are introduced in Section 2.3.2. Subsequently the dual scaling method is modified to utilise these splines together with latent classes to allow for multiple response styles. An alternating non-negative least squares algorithm is described for fitting the model in Section 2.3.4. Selecting the number of latent response style groups (Section 2.3.5) and creating a data set purged of the effects of response styles (Section 2.3.6) are also discussed.

2.3.1 Dual Scaling of Successive Categories Data

Dual scaling (DS) is a multivariate exploratory statistical technique which is equivalent to correspondence analysis (CA) when analysing contingency tables (Van de Velden, 2000). For such cases it is used to visualise departures from the independence assumption in the two-way contingency table in a low dimensional space, akin to principle components analysis (PCA) for continuous data (Nishisato, 980a; Greenacre, 2007). However, for the successive categories data dealt with here there are important differences.

Both DS and CA deal with non-contingency table data by typically applying the standard procedure to a specific recoding of the data, designed to transform the data into a form that resembles a contingency table (Greenacre, 2007). This recoding requires the original data matrix **X** to be transformed before analysis, and for successive categories data in particular the recoding schemes differ in an important way. The usual CA method uses a doubling of columns (that is, adding an additional column to **X** for each object) to construct scales with "positive" and "negative" poles before applying ordinary CA (see Greenacre, 2007). However Nishisato (1980) proposes the following alternative method. This involves augmenting rating scale category thresholds or boundaries to the data, which increases the number of columns from m to $m + q - 1$, and then converting this to rank-orders. Although Nishisato's original DS formulation focuses on a so-called dominance matrix (see Nishisato, 980a), it has been shown that DS applied to these rank-orders are equivalent to doubling the rows (instead of the columns) of the matrix of rankings before applying CA (Van de Velden, 2000; Torres and Greenacre, 2002).

The method is perhaps best illustrated by an example. Consider transforming the following data matrix **X**, where three objects A, B and C are rated by $n = 4$ respondents on a 5-point Likert scale (thus, $q = 5$). The first step requires augmenting 4 ($= q - 1$) columns to **X**, one column for each of the boundaries between the pairs of adjacent ratings. Let the boundaries be called $b_1, \ldots, b_4$, where b_1 falls between ratings 1 and 2, and so forth up to b_4 which falls between categories 4 and 5. It suffices to assign scores midway between the rating categories to each boundary, to arrive at the augmented data matrix:

$$\mathbf{X} = \begin{array}{c} \begin{array}{ccc} \mathtt{A} & \mathtt{B} & \mathtt{C} \end{array} \\ \begin{pmatrix} 4 & 3 & 1 \\ 2 & 2 & 5 \\ 3 & 2 & 2 \\ 1 & 5 & 4 \end{pmatrix} \end{array} \Rightarrow \mathbf{X}_{aug} = \begin{array}{c} \begin{array}{ccccccc} \mathtt{A} & \mathtt{B} & \mathtt{C} & \mathtt{b}_1 & \mathtt{b}_2 & \mathtt{b}_3 & \mathtt{b}_4 \end{array} \\ \begin{pmatrix} 4 & 3 & 1 & 1.5 & 2.5 & 3.5 & 4.5 \\ 2 & 2 & 5 & 1.5 & 2.5 & 3.5 & 4.5 \\ 3 & 2 & 2 & 1.5 & 2.5 & 3.5 & 4.5 \\ 1 & 5 & 4 & 1.5 & 2.5 & 3.5 & 4.5 \end{pmatrix} \end{array}. \quad (2.1)$$

Secondly, each row is converted to rankings, starting with a lowest rank of 0 and a highest rank of 6 ($= m + q - 2$) in this case. For ties the total ranking assigned to the tied objects are distributed equally. This yields the following result for the example:

$$\mathbf{X}_{aug} \Rightarrow \mathbf{T} = \begin{array}{c} \begin{array}{ccccccc} \mathtt{A} & \mathtt{B} & \mathtt{C} & \mathtt{b}_1 & \mathtt{b}_2 & \mathtt{b}_3 & \mathtt{b}_4 \end{array} \\ \begin{pmatrix} 5 & 3 & 0 & 1 & 2 & 4 & 6 \\ 1.5 & 1.5 & 6 & 0 & 3 & 4 & 5 \\ 4 & 1.5 & 1.5 & 0 & 3 & 5 & 6 \\ 0 & 6 & 4 & 1 & 2 & 3 & 5 \end{pmatrix} \end{array}. \tag{2.2}$$

Note that in general $\mathbf{T}$ has n rows and $m + q - 1$ columns. DS also requires construction of the matrix $\mathbf{S}$ that would have resulted if q was the lowest and not the highest rating of the Likert scale. This is easily achieved as

$$\mathbf{S} = (m + q - 2)\mathbf{1}\mathbf{1}' - \mathbf{T}. \tag{2.3}$$

Using the CA formulation of DS of Van de Velden (2000), a row-doubled ratings matrix $\mathbf{F}_r : 2n \times (m + q - 1)$ is constructed as

$$\mathbf{F}_r = \begin{pmatrix} \mathbf{T} \\ \mathbf{S} \end{pmatrix}. \tag{2.4}$$

This matrix is subjected to CA, which assigns optimal scores in the vectors $\mathbf{a}$ and $\mathbf{b}$ to the rows and columns of $\mathbf{F}_r$ respectively. Since the aim is to assign to the boundaries ordered scores which are sensitive to rating scale use, a one-dimensional solution is used. This assignment is achieved by minimising a least squares criterion $L(\mathbf{a}, \mathbf{b})$ through the singular value decomposition (SVD) (Van de Velden et al., 2009). In the present context L is given by

$$L(\mathbf{a}, \mathbf{b}) = c\|\mathbf{F}_r - \frac{1}{2}(m + q - 2)(\mathbf{1}\mathbf{1}' + \mathbf{a}\mathbf{b}')\|^2 \tag{2.5}$$

where c is a proportionality constant, $\mathbf{1}$ denotes vectors of ones of the appropriate lengths and $\frac{1}{2}(m + q - 2)\mathbf{1}\mathbf{1}'$ centres the rankings in $\mathbf{F}_r$. For identifiability a constraint such as $\|\mathbf{a}\| = 1$ is imposed. The method is discussed in more detail in Section 2.3.3.

Note that an important consequence of the data recoding scheme is that the dual scaling procedure provides coordinates for the boundaries. The effect of the boundaries is to retain the information on how different the original ratings assigned to the objects were before the rankings were constructed. The coding scheme also imposes ordinality on the object and the boundary scores in $\mathbf{b}$ by constructing rankings.

The optimal scores assigned to the boundaries can be used to detect response styles since they estimate the thresholds of the response mapping of the group of respondents, as was discussed

Response style	Lower Curvature	Upper Curvature
No Response Style	None	None
Acquiescence	Convex	Convex
Disacquiescence	Concave	Concave
Extreme Responding	Concave	Convex
Midpoint Responding	Convex	Concave

Table 2.1: Curvature properties of the four response styles.

in relation to Figure 2.1. Intuitively optimal scores assigned to the boundaries work as follows. If a specific rating category j is used very often, the boundaries b_{j-1} and b_j will often receive rankings which differ substantially since the category is often filled. Consequently, the optimal scores assigned will differ significantly, indicating that respondents use the category very often. The same reasoning illustrates that when rating j is used very infrequently, the optimal scores for b_{j-1} and b_j will be very similar. Therefore, when a group of respondents have the same response mapping, the method will be able to tell which type that mapping is.

In Section 2.3.3 latent classes will be introduced for the boundary scores which allows for multiple response styles. First, however, using monotone quadratic splines with the dual scaling method is discussed.

2.3.2 Modelling Response Styles by Monotone Quadratic Splines

From Figure 2.1 it is evident that the four response styles considered can be completely described in terms of its curvature properties. By dividing the horizontal axes into two equal lower and upper parts, the four response styles are characterized by either concavity or convexity in the lower and upper parts of its domain. This is summarised in Table 2.1.

For inferential and response style classification purposes it will prove useful to parameterize the response style transformations considered here. Furthermore, using smooth functions will improve model parsimony and the stability of parameter estimation, as well as facilitate the process of purging the response styles from the data by interpolation (see Section 2.3.6). The family of monotone quadratic splines with a single interior knot is ideal for this purpose as it combines two quadratic polynomial functions in the adjacent intervals of the domain, subject to continuity and differentiability restrictions at the interior knot. These splines are either concave, convex or linear in the lower and upper halves of the domain and therefore reproduce all the curves described in Figure 2.1 and Table 2.1.

The splines have three non-constant basis functions (the so-called I-spline basis) derived by appropriately integrating the basis functions of the M-spline basis (see Ramsay, 1988). A quadratic monotone spline with a single interior knot $t \in [L, U]$ and intercept μ is of the form

$$f(x) = \mu + \sum_{i=1}^{3} \alpha_i M_i(x \mid t). \tag{2.6}$$

In the proposed model $t = L + 0.5(U - L)$ is chosen to lie halfway between the lower and upper limits L and U respectively. Monotonicity requires that $\alpha_i \geq 0$ for $i = 1,2,3$. The basis functions M_1, M_2 and M_3 are constructed to ensure continuity and first-order differentiability at t, and their formulae are as follows (Ramsay, 1988):

$$\begin{aligned}
M_1(x \mid t) &= \begin{cases} \frac{2t(x-L)-(x^2-L^2)}{(t-L)^2}, & \text{if } L \leq x < t; \\ 1, & \text{if } t \leq x \leq U; \end{cases} \\
M_2(x \mid t) &= \begin{cases} \frac{(x-L)^2}{(t-L)(U-L)}, & \text{if } L \leq x < t; \\ \frac{t-L}{U-L} + \frac{2U(x-t)-(x^2-t^2)}{(U-t)(U-L)}, & \text{if } t \leq x \leq U; \end{cases} \\
M_3(x \mid t) &= \begin{cases} 0, & \text{if } L \leq x < t; \\ \frac{(x-t)^2}{(U-t)^2}, & \text{if } t \leq x < U; \end{cases}
\end{aligned} \tag{2.7}$$

Hence (2.6) is simply a linear combination of these three piece-wise quadratic functions with an intercept.

The spline functions are built into the column scores $\mathbf{b}$ in (2.5) by using the $(q-1) \times 4$ design matrix $\mathbf{M}$ to collect the basis functions corresponding to μ, α_1, α_2 and α_3 respectively. The basis functions are evaluated at the midpoints between rating categories, for example at 1.5, 2.5 up to 6.5 for a 7-point Likert scale. Hence $\mathbf{b}$ can be written as

$$\mathbf{b} = \begin{pmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{b}_1 \\ \mathbf{M}\boldsymbol{\alpha} \end{pmatrix} \tag{2.8}$$

with $\mathbf{b}_1$ the m-vector of unrestricted object scores and $\mathbf{b}_2$ the $(q-1)$-vector of spline-restricted boundary scores. The spline parameters are collected in $\boldsymbol{\alpha} = (\mu, \alpha_1, \alpha_2, \alpha_3)'$.

The basis functions M_1, M_2 and M_3 in (2.7), as depicted in Figure 2.2, are piecewise quadratic, with only two of them nonconstant in each of the intervals $[L, t)$ and $[t, U]$. This is convenient because it means the second derivative of f, and hence the curvature, depends only on two parameters in each interval. Rescaling without loss of generality so that $L = 0$ and $U = 1$, the

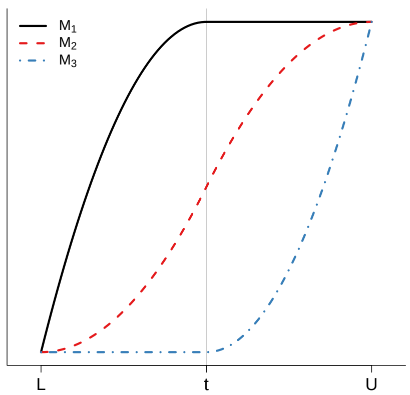


Figure 2.2: The three I-spline basis functions for quadratic monotone splines with a single interior knot t.

curvature of f (not necessarily defined at $t = 1/2$) is given by

$$\frac{d^2}{dx^2} f(x) = \begin{cases} -8\alpha_1 + 4\alpha_2, & \text{if } 0 \leq x < 1/2; \\ -4\alpha_2 + 8\alpha_3, & \text{if } 1/2 < x \leq 1; \end{cases} \tag{2.9}$$

The function $f(x)$ is either convex, concave or linear in a given interval depending on whether the second derivative (2.9) is positive, negative or zero respectively, which does not depend on x. In fact, assuming that α_1 and α_3 are larger than zero, the curvature can be measured solely in terms of the ratios α_2/α_1 and α_2/α_3, referred to henceforth as the curvature parameters. For example, the requirement for convexity in both the lower and upper domain is

$$\frac{d^2}{dx^2} f(x) > 0 \Leftrightarrow \begin{cases} \frac{\alpha_2}{\alpha_1} > 2, & \text{if } L \leq x < t; \\ \frac{\alpha_2}{\alpha_3} < 2, & \text{if } t < x < U. \end{cases} \tag{2.10}$$

When one or both of α_1 and α_3 are zero, one or both of these curvature parameters may be undefined. This can cause problems for its graphical representation, some of which will be shown below. In such cases a continuity adjustment through the addition of a small positive constant to both the numerator and denominator in (2.10) can be useful.

It is possible to rewrite Table 2.1 wholly in terms of the curvature parameters, but more importantly using the curvature parameters it is possible to visualize the curvature of an estimated response style in a single plot. Figure 2.3 illustrates the situation by plotting α_2/α_3 against α_2/α_1, as well as incorporating the response style classification regions derived from Table 2.1. When both curvature parameters equal two, no response style is present. Due to the fact that

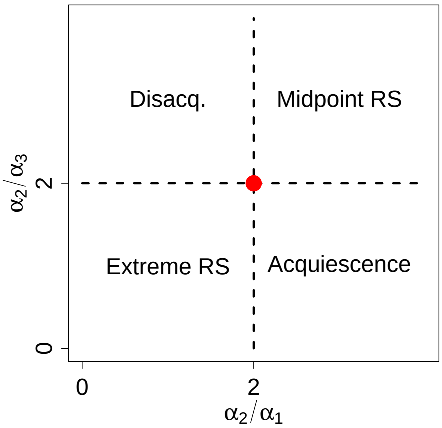


Figure 2.3: Classifying response styles graphically using the curvature properties of monotone quadratic splines.

both curvature parameters has the range $[0, \infty)$, a more symmetric plot is arrived at by using the base-2 logarithmic transform and centring – this is illustrated in Section 2.5.

2.3.3 Dual Scaling Method for Multiple Response Styles

To allow for multiple response styles, suppose that each of the n individuals belongs to one of K response style groups, the exact membership being unknown. Let the $n \times K$ matrix $\mathbf{G}$ contain as columns the group indicator vectors $\{\mathbf{g}_k\}_{k=1}^{K}$, each indicating which individuals belong to that specific group. The column scores $\{\mathbf{b}_k\}_{k=1}^{K}$ are of the same form as $\mathbf{b}$ in Equation (2.8), but are now group-specific by replacing $\mathbf{b}_2$ with $\mathbf{b}_{2k} = \mathbf{M}\boldsymbol{\alpha}_k$. This allows for the different groups to have different response mappings by letting the spline parameters $\boldsymbol{\alpha}_k = (\mu_k, \alpha_{1k}, \alpha_{2k}, \alpha_{3k})'$ vary between groups. The object scores $\mathbf{b}_1$ and the row scores $\mathbf{a}$ remain fixed across all response style groups.

The group membership $\mathbf{G}$ needs to be estimated, together with the $2n$-vector $\mathbf{a}$ of optimal scores for the individuals and the column score vectors $\mathbf{b}_k$ of length $(m + q - 1)$ contained in the K columns of $\mathbf{B}$. It is required for monotonicity that $\alpha_{ik} \geq 0$ for all i and k. The loss function in Equation (2.5) must be adjusted to allow for the multiple response styles as well as for the spline restrictions. This constrained dual scaling method for the detection of response styles can be formulated as

$$\min_{\mathbf{a},\mathbf{B},\mathbf{G}} L(\mathbf{a}, \mathbf{B}, \mathbf{G})$$

$$\text{subject to } \mathbf{b}_k = \begin{pmatrix} \mathbf{b}_1 \\ \mathbf{b}_{2k} \end{pmatrix} \text{ and } \alpha_{ik} \geq 0,\ i = 1, 2, 3,\ k = 1, 2, \ldots, K. \tag{2.11}$$

The adjusted loss function (compare Eq. (2.5)) is

$$L(\mathbf{a},\mathbf{B},\mathbf{G}) = c\|\mathbf{F}_r - \frac{1}{2}(m+q-2)(\mathbf{1}\mathbf{1}' + \sum_{k=1}^{K} \mathbf{D}_{\mathbf{g}_k}\mathbf{a}\mathbf{b}_k')\|^2. \tag{2.12}$$

Again, c is a proportionality constant, and the diagonal matrices $\mathbf{D}_{\mathbf{g}_k}$ are constructed as

$$\mathbf{D}_{\mathbf{g}_k} = \begin{pmatrix} \mathrm{diag}\,(\mathbf{g}_k) & \mathbf{0} \\ \mathbf{0} & \mathrm{diag}\,(\mathbf{g}_k) \end{pmatrix}. \tag{2.13}$$

In this context, diag($\mathbf{x}$) denotes the diagonal matrix with $\mathbf{x}$ on the main diagonal. Through using the $\{\mathbf{D}_{\mathbf{g}_k}\}_{k=1}^{K}$ in (2.12), individuals are associated with the corresponding $\mathbf{b}_k$ for their group. As K increases, the number of parameters in the model increases and consequently the loss function L decreases as well. Therefore, when considering how the value of L changes for different values of K in a scree plot, it is convenient to standardise these values to the unit interval $[0,1]$.

An algorithm for minimising L is discussed in the next section.

2.3.4 An Alternating Nonnegative Least Squares Algorithm

Solving the optimization problem in (2.11) requires finding $\mathbf{a}, \mathbf{B}$ and $\mathbf{G}$ under the appropriate restrictions. The approach discussed here alternates over two steps:

1. The algorithm combines alternating least squares (ALS) and nonnegative least squares (NNLS; Lawson and Hanson, 1974) to approximate the optimal $\mathbf{a}$ and $\mathbf{B}$ for a given group membership matrix $\mathbf{G}$. This involves fixing the value of $\mathbf{a}$, estimating the optimal $\mathbf{B}$ with NNLS, and then updating $\mathbf{a}$ by ordinary least squares (OLS) based on the estimate of $\mathbf{B}$. This ALS process is repeated for a given $\mathbf{G}$ until numerical convergence is observed.

2. For fixed $\mathbf{a}$ and $\mathbf{B}$, $\mathbf{G}$ is updated by a K-means type algorithm given the values determined for $\mathbf{a}$ and $\mathbf{B}$. This step simply allocates each individual sequentially to the group which minimises the loss function.

The algorithm alternates between steps one and two until the loss function L changes by less than a small positive constant. Note that starting values for both $\mathbf{a}$ and $\mathbf{G}$ are required. For the $\boldsymbol{a}$ vector standard normal random numbers are simulated, while random assignment to K groups is used for $\mathbf{G}$. Block-relaxation algorithms such as this is guaranteed to converge monotonically, albeit to a local minimum; therefore multiple random starts are required (De Leeuw, 1994). The related issues of local optima in K-means clustering and categorical principal components analysis are discussed in Hand and Krzanowski (2005) and Van der Kooij (2007, Chapter 2) respectively. In Appendix B an overview of these local optima is given in the context of the empirical example (Section 2.5).

The optimization process is described in more detail in Algorithm 1, with an exposition of its derivation deferred to Appendix A. The formulation is for a single starting configuration of **G**, and needs to be repeated for multiple such configurations. Parameters that need to be specified include n_a, the number of (random) starts used for **a**, the maximum number of iterations maxit_a and maxit_G for the ALS and K-means phases respectively, and also the numerical tolerances $\epsilon_1 > 0$ and $\epsilon_2 > 0$ for these two steps. Note that the spline restrictions are sufficient as normalization constraints in the ALS part of the algorithm, and hence the vector $\boldsymbol{a}$ is only standardized to $||\boldsymbol{a}||^2 = 2n$ after convergence.

To update **G**, the algorithm cycles through all respondents in turn. For the current respondent i, it calculates for each class what the loss function would be if respondent i were assigned to that class, given the current classification of all other respondents. This respondent is then moved to the class with minimum loss (or stays in the same class if this is already the best choice). The algorithm then proceeds to the next respondent $i + 1$, and starts again with respondent 1 once the last respondent is reached. Once a complete pass over all respondents are made where no change in classification occurs, the updating of **G** terminates and the algorithm returns to the ALS updating step.

2.3.5 Selecting the Number of Response Style Groups

To select the number of groups K, a scree plot of the loss function for different values of K can be used. The aim is to choose the smallest K such that larger values do not significantly reduce the loss. This method was introduced by Cattell (1966) and has been widely adopted. The dual scaling method also separates individuals based on the shape of the response transformations and rating frequencies in the groups. This supplementary information can be helpful for choosing K in cases where the scree plot is not conclusive. This will be illustrated in the empirical application of Section 2.5.

2.3.6 Purging Response Styles

Once the estimates $\hat{\mathbf{a}}, \hat{\mathbf{B}}$ and $\hat{\mathbf{G}}$ have been obtained, these can be used to create a version of the original data $\mathbf{X}$ which is purged of response styles. All that is needed is to use the splines estimated for each response style group to assign optimal scores to the rating scale. Then these scores are substituted in $\mathbf{X}$ by replacing every rating with the appropriate optimal score.

Determining the optimal scores of the ratings proceeds by evaluating the splines as before, but now at the ratings themselves and not at the boundaries. This simply requires constructing

Algorithm 1 Alternating Nonnegative Least Squares Algorithm

```
set $i = 0$, $h = 0$ and $n_a$, $\text{maxit}_a$, $\text{maxit}_G$, $\epsilon_1 > 0$ and $\epsilon_2 > 0$
initialise $\mathbf{G}_0$, set $\mathbf{F}_r^* = \mathbf{F}_r - \frac{1}{2}(m+q-2)\mathbf{1}\mathbf{1}'$
while $L_{h-1} - L_h > \epsilon_2$ and $h \leq \text{maxit}_G$ do
    construct $\mathbf{D}_{\mathbf{g}_k}^h$ from $\mathbf{G}_h$ according to Equation (2.13)
    for all $j = 1, 2, \ldots, n_a$ do (iterate over different starts for $\mathbf{a}$)
        if $i = 0$ and $h = 0$, generate a starting configuration $\mathbf{a}_j$ for $\mathbf{a}$
        while $L_{i-1,j} - L_{ij} > \epsilon_1$ and $i \leq \text{maxit}_a$ do
            update (indices i and h are omitted for readability)
                $w_{kj} \leftarrow (\mathbf{a}_j' \mathbf{D}_{\mathbf{g}_k}^h \mathbf{a}_j)^{-1/2}$ for all $k$
                $(\mathbf{v}_{1kj}, \mathbf{v}_{2kj})' \leftarrow \frac{2}{m+q-2} w_{kj} (\mathbf{F}_r^*)' \mathbf{D}_{\mathbf{g}_k}^h \mathbf{a}_j$ for all $k$
                $\mathbf{b}_{1j} \leftarrow (\mathbf{a}_j' \mathbf{a}_j)^{-1} \sum_{k=1}^K w_{kj} \mathbf{v}_{1kj}$
                $\boldsymbol{\alpha}_{kj} \leftarrow \arg\min_{\alpha_{kj}} \|w_{kj}^{-1} \mathbf{M} \boldsymbol{\alpha}_{kj} - v_{2kj}\|^2$ s.t. $\alpha_{1kj}, \alpha_{2kj}, \alpha_{3kj} \geq 0$ for all $k$
                $\mathbf{b}_{2kj} \leftarrow \mathbf{M} \boldsymbol{\alpha}_{kj}$ for all $k$ so that $\mathbf{b}_{kj} = (\mathbf{b}_{1j}, \mathbf{b}_{2kj})'$
                $\mathbf{a}_j \leftarrow \frac{2}{m+q-2} (\sum_{k=1}^K \mathbf{b}_{kj}' \mathbf{b}_{kj} \mathbf{D}_{\mathbf{g}_k}^h)^{-1} \sum_{k=1}^K \mathbf{D}_{\mathbf{g}_k}^h \mathbf{F}_r^* \mathbf{b}_{kj}$
                $i \leftarrow i + 1$
            calculate $L_{ij} = L(\mathbf{a}_j, \mathbf{B}_j, \mathbf{G}_h)$
        end while
    end for
    if $n_a > 1$, set $(\mathbf{a}_1, \mathbf{B}_1) \leftarrow \arg\min_{(\mathbf{a}_j, \mathbf{B}_j)} L_{ij}$ and $n_a \leftarrow 1$
    update $h \leftarrow h + 1$ and $\mathbf{G}_{h-1}$ to $\mathbf{G}_h$ by reassigning each individual to the group which
      minimises $L$
    calculate $L_h = L(\mathbf{a}_1, \mathbf{B}_1, \mathbf{G}_h)$
end while
return $\hat{\mathbf{a}} = \mathbf{a}_1, \hat{\mathbf{B}} = \mathbf{B}_1$ and $\hat{\mathbf{G}} = \mathbf{G}_h$, and repeat for different starting values $\mathbf{G}_0$
```

a design matrix from the spline basis functions evaluated at the rating categories 1 to q, where for categories 1 and q respectively L and U are used in the notation of Section 2.3.2. As before, a single interior knot t at the middle of the domain $[L, U]$ of the splines are assumed. Let this matrix be $\mathbf{M}^*$. The optimal scores are then simply determined as $\mathbf{M}^*\boldsymbol{\alpha}_k$. In Section 2.4.3 a simulation experiment is conducted to assess how accurately this method can reproduce a known underlying correlation structure from contaminated data.

2.4 Simulation Results

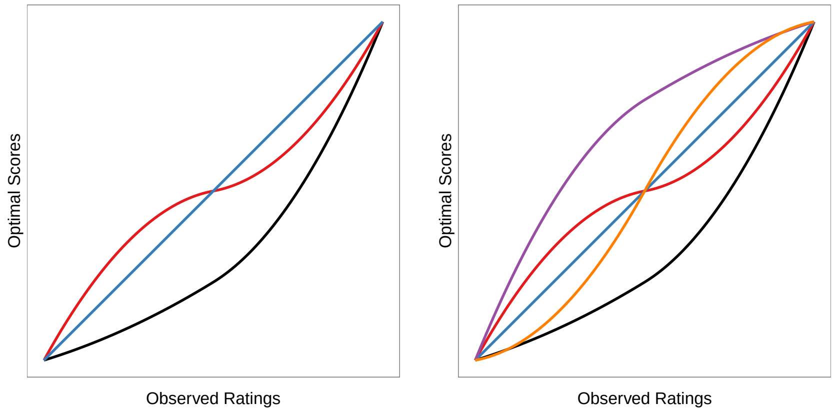
2.4.1 Simulation Model

The simulated data was generated in a three-step procedure. First, the true underlying preference structure for the m objects is obtained by simulating m random numbers from a $U(0, 1)$-distribution. These are gathered into the m-vector $\boldsymbol{\mu}$. Second, individual preferences are generated by simulating n times from each of m truncated normal distributions respectively centred at the elements of $\boldsymbol{\mu}$. The individual preferences are given by $\boldsymbol{\delta}_i = \boldsymbol{\mu} + \boldsymbol{\varepsilon}_i$, with $\boldsymbol{\varepsilon}_i, i = 1, \ldots, n$, representing the individuals deviation from the mean.

Truncation is done at 0 and 1 so that response styles can be defined on the closed interval $[0, 1]$. Note that the use of truncation avoids overflow problems at the lower and upper ends of the response style mapping, and hence improves on the original approach of Van de Velden (2007). The truncated normal draws are done independently and with error variance σ^2, which is an important parameter because it determines how pronounced the multi-modality of the mixture of truncated normals over $[0, 1]$ is. An increase in the value of σ implies that it easier to detect response styles as the actual preference structure plays less of a role in forming the ratings.

The resultant true preferences are randomly divided into different response style groups. Finally, these data are discretized to m categorical variables with q-point Likert-scales, according to the cut points on $[0, 1]$ implied by the chosen K response styles. These response styles are parameterized to come from the family of monotone quadratic splines outlined in Section 2.3.2.

In the simulations, choices must be made regarding the following: the number of objects m, the number of rating categories q, the underlying standard deviation σ, the number of response styles K, as well as their shapes defined by $\boldsymbol{\alpha}_k, k = 1, \ldots, K$, the sample size n and how this is divided among the K groups, namely $n_k, k = 1, \ldots, K$.



(a) $K = 3$ (b) $K = 5$

Figure 2.4: Response styles used in the simulation study. Each curve represents a different style.

2.4.2 Assessing Classification Performance

The first simulation study assesses the classification accuracy of the dual scaling method. It is assumed in this experiment that the number of groups K is known beforehand. For each of the experimental conditions, 50 simulated data sets were constructed and the dual scaling method applied. For each data set estimation was based on 15 random starts for $\mathbf{G}$, and for each of these starts the ALS procedure was initialised from 50 different random configurations for the row scores $\mathbf{a}$.

The 108 experimental conditions consisted of the following. The number of objects m was varied over 10, 20 and 30 items. The rating scales employed were either $q = 5$ or 7-point scales. Sample sizes of $n = 200$, 1000 and 5000 respectively were used. The number of groups K were either 3 or 5. For each of these K, it was assumed that one of the groups has a linear response mapping (that is, a group with no response style). The additional $K - 1$ groups exhibited response styles through nonlinear mappings. For $K = 3$, these additional groups were acquiescence and extreme responding, since Baumgartner and Steenkamp (2001) found that these are most prevalent in survey data. For $K = 5$, groups for disacquiescence and midpoint responding were also added. The corresponding spline functions used to simulate from are shown in Figure 2.4. The sample of n respondents was assigned to the groups by allocating either 20%, 50% or 80% of respondents equally among the $K - 1$ response style groups. These percentages represent the amount of contamination in the simulated data. The remaining percentage of respondents was assigned to the group exhibiting no response style. The latent standard deviation σ was fixed at 0.1 for all experiments.

To assess the classification performance of the method, the adjusted Rand index as well as the percentage correctly classified (the so-called hit rate) were computed. The adjusted Rand index (ARI) of Hubert and Arabie (1985) assesses the similarity between two partitions, adjusted for chance correspondences between these partitions. The upper limit of the ARI is one, and indicates perfect agreement. An ARI of zero indicates that the method does not improve on random assignment, with all positive values indicating an improvement. Negative ARI values are also possible, and indicate poorer performance than random assignment. The ARI is in general lower than the hit rate, and can be considered as a more objective measure of performance.

For each of the 108 experimental conditions, Tables 2.2 and 2.3 show the average values over the 50 simulated data sets. It is apparent that the sample size n does not have a large influence on the ARI and hit rate. The number of groups K is very important for performance when the contamination percentage is low (20%). This is because for $K = 5$ groups the 20% of contaminated data points must be divided into 4 groups instead of 2 when $K = 3$, which results in groups with very low proportions n_k/n of the total sample. The low performance here is somewhat compensated for by using more items, such as $m = 30$, but for $K = 5$ groups even more items are needed. In general, using more items increases the classification accuracy. Using a larger number of rating categories q also increases performance, but mostly so with fewer groups ($K = 3$). The method improves on random assignment – especially in cases with higher response style prevalence and 20 or more items the improvement is substantial.

2.4.3 Recovering Underlying Structure through Data Cleaning

The simulation model of Section 2.4.1 assumes that, given the expected value of the object scores m, the objects are independently distributed as truncated normal distributions. Although the true correlation matrix between the objects thus is the identity matrix $\mathbf{I}$, the observed correlations after the response style contamination is often inflated. To show improvement, the cleaned data derived as in Section 2.3.6 should have correlations resembling independence more closely. A visual example is given in Figure 2.5 for simulated data ($m = 20, K = 3$ similar to the conditions used in Tables 2.2 and 2.3), where the colours indicate the magnitude of the Pearson correlations. It is evident that the response styles artificially inflate the correlations. When $q = 7$, the cleaned data to some extent succeeds in removing the spurious correlations, but when $q = 5$ the situation is not much improved.

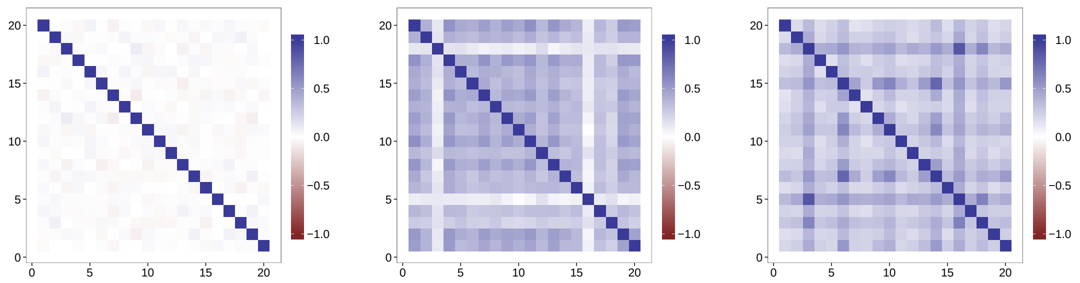
The conditions under which the cleaned data can be expected to provide a better representation of the underlying correlation matrix was studied further through simulations. For the different values of K, n, q, and the proportion of response style contamination used in Section 2.4.2, 50 simulated data sets were constructed and cleaned through the dual scaling method. Here $m = 20$

		q = 5									q = 7								
		n = 200			n = 1000			n = 5000			n = 200			n = 1000			n = 5000		
	RS% m =	10	20	30	10	20	30	10	20	30	10	20	30	10	20	30	10	20	30
	20%	0.28	0.40	0.61	0.30	0.42	0.62	0.29	0.40	0.62	0.31	0.48	0.74	0.29	0.48	0.80	0.30	0.48	0.80
K = 3	50%	0.59	0.80	0.90	0.57	0.80	0.91	0.58	0.80	0.91	0.62	0.85	0.93	0.64	0.86	0.94	0.62	0.85	0.94
	80%	0.73	0.90	0.93	0.72	0.89	0.95	0.75	0.89	0.95	0.75	0.91	0.96	0.76	0.90	0.96	0.76	0.91	0.96
	20%	0.16	0.22	0.33	0.16	0.22	0.34	0.16	0.21	0.34	0.17	0.24	0.35	0.17	0.25	0.36	0.18	0.25	0.36
K = 5	50%	0.42	0.65	0.82	0.42	0.65	0.81	0.42	0.65	0.82	0.44	0.67	0.86	0.44	0.66	0.84	0.44	0.66	0.85
	80%	0.70	0.85	0.93	0.70	0.86	0.93	0.71	0.86	0.93	0.73	0.88	0.94	0.73	0.88	0.95	0.73	0.88	0.95

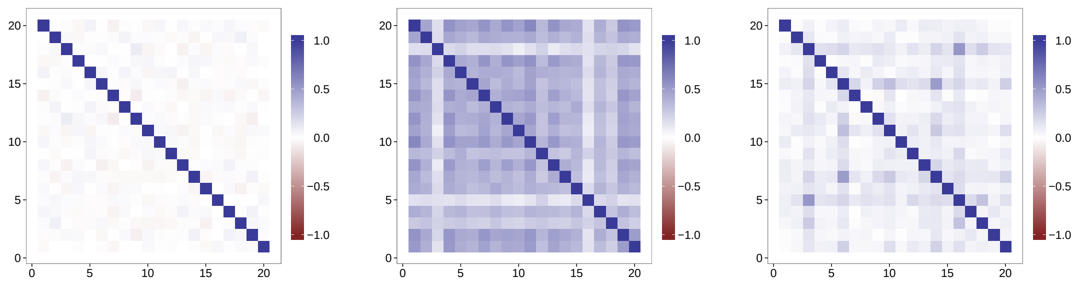
Table 2.2: Average adjusted Rand index for 50 simulations at the different parameter settings.

		q = 5									q = 7								
		n = 200			n = 1000			n = 5000			n = 200			n = 1000			n = 5000		
	RS% m =	10	20	30	10	20	30	10	20	30	10	20	30	10	20	30	10	20	30
	20%	0.66	0.76	0.87	0.67	0.77	0.87	0.67	0.76	0.88	0.69	0.81	0.92	0.67	0.81	0.94	0.68	0.81	0.94
K = 3	50%	0.84	0.93	0.97	0.83	0.93	0.97	0.84	0.93	0.97	0.85	0.95	0.98	0.86	0.95	0.98	0.86	0.95	0.98
	80%	0.87	0.96	0.97	0.87	0.95	0.98	0.89	0.95	0.98	0.88	0.96	0.98	0.89	0.96	0.99	0.89	0.96	0.98
	20%	0.50	0.56	0.68	0.50	0.57	0.70	0.49	0.56	0.69	0.51	0.60	0.70	0.50	0.61	0.71	0.52	0.61	0.72
K = 5	50%	0.72	0.86	0.93	0.71	0.86	0.93	0.71	0.86	0.93	0.73	0.87	0.95	0.74	0.87	0.94	0.74	0.87	0.94
	80%	0.84	0.93	0.97	0.84	0.94	0.97	0.85	0.94	0.97	0.86	0.95	0.98	0.86	0.95	0.98	0.86	0.95	0.98

Table 2.3: Average hit rates for 50 simulations at the different parameter settings.



(a) *True preferences* (b) *Observed ratings* (c) *Cleaned data*



(d) *True preferences* (e) *Observed ratings* (f) *Cleaned data*

Figure 2.5: The effect of response styles on the underlying uncorrelated objects: estimated Pearson correlations before and after contamination, as well as after cleaning the data. The number of rating categories is $q = 5$ for (a) – (c) and $q = 7$ for (d) – (f), with $m = 20$ items in all cases .

was fixed for simplicity. For each of these data sets, the root mean square error (RMSE) between **I** and the empirical Pearson correlation matrix for the contaminated data was calculated, where

$$\mathrm{RMSE}(\mathbf{V},\mathbf{W}) = \sqrt{\sum_i \sum_j (v_{ij} - w_{ij})^2} \tag{2.14}$$

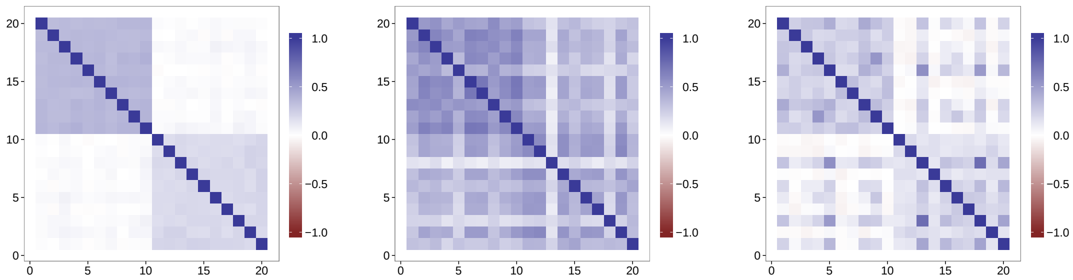
for commensurable matrices **V** and **W**. Similarly, the RMSE comparing **I** with the empirical Pearson correlations of the cleaned data can be computed. A reduction in the RMSE when using the cleaned data as opposed to the contaminated data indicates that the cleaned data has a correlation structure which matches the true correlation structure more closely.

A two-sample Wilcoxon test, also known as the Mann-Whitney test, (e.g. Rice, 2007) was used to test the null hypothesis that the RMSE is equal for the contaminated and cleaned data against the one-sided alternative that the RMSE for the contaminated data is greater than that of the cleaned data. The results are quite clear: when $q = 7$ the null hypothesis is always rejected ($p < 0.001$) in favour of the alternative, whilst when $q = 5$ the null hypothesis cannot be rejected even once (all $p > 0.2$). It can therefore be deduced that when a sufficient number of rating categories q are used, the correlation structure of the cleaned data is more representative of the true underlying structure of the data.

A related question concerns the performance of the method in the presence of a nontrivial correlation structure. To impose such a structure whilst retaining truncated normal marginal distributions for the objects, a copula is used (note that the truncated multivariate normal distribution does not guarantee truncated normal marginals). A copula is a multivariate distribution function $C(u_1, u_2, \ldots, u_m)$ with uniform marginals (Hofert and Mächler, 2011). According to Sklar's theorem (Sklar, 1959; Hofert and Mächler, 2011) a multivariate distribution function F with marginals $\{F_j\}_{j=1}^m$ can be constructed as

$$F(x_1, x_2, \ldots, x_m) = C(F_1(x_1), F_2(x_2), \ldots, F_m(x_m)). \tag{2.15}$$

The marginal truncated normal distributions can be achieved by the inverse probability integral transform. The dependence structure between the variables is solely determined by the copula. Here two independent Clayton copula (Clayton, 1978) functions will be used to impose a correlation structure in terms of Kendall's τ, a well-known measure of rank correlation (see Kendall, 1938; Hofert and Mächler, 2011). The structure induced here for $m = 20$ is as follows: the first 10 objects are correlated with $\tau = 0.2$, independent of the other 10 objects which are correlated with $\tau = 0.35$. These τ values amount to Pearson correlations of approximately $\rho = 0.3$ and $\rho = 0.5$ respectively (an approximate relationship is $\rho \approx \sin(\tau\pi/2)$ - see Kendall and Gibbons (1990)). It is also possible to introduce negative correlations by using $1 - U$ instead



(a) *True preferences* (b) *Observed ratings* (c) *Cleaned data*

Figure 2.6: An example of the correlation structure imposed by the Clayton copula's, in terms of Kendall's τ.

of U in the inverse probability integral transform. In the application here these reversals are made randomly with differing probability γ. The theoretical, observed and cleaned correlations given by Kendall's τ for one such copula is illustrated in Figure 2.6, with $m = 20$ and $q = 7$.

The difference in RMSE can again be used to evaluate the effect of the data cleaning on the correlation structure, now using Kendall's τ since the Clayton copula's use this measure directly. A simulation study was conducted for $m = 20$ objects with the other parameters varying as before. For each combination of the parameters, the RMSE was calculated for 50 randomly generated data sets according to the copula model described above. Then for each data set the constrained dual scaling model was fit as before, and a cleaned data set constructed. The difference in the RMSE for the contaminated data as compared to the cleaned data was recorded.

Table 2.4 presents the average reduction in RMSE as a result of cleaning the data with the dual scaling procedure. As before the two-sample Wilcoxon test was performed. Significant improvements were found in all cases except those printed in italic in Table 2.4. It is apparent that the cleaned data improves the RMSE in all cases, except where both q and K are small and the proportion of contamination is moderate (50%) to large (80%). Except for these circumstances, the constrained dual scaling method improves the estimation of the true correlation structure by removing the response styles effects.

2.4.4 Recovering the Parameters in Principal Components Analysis

It is possible to examine how well the method can recover parameters after the contaminated data have been cleaned of response styles. For simplicity, Principal Components Analysis (PCA) (e.g. Johnson and Wichern, 2002) was used as analysis method, a well-known multivariate dimension reduction technique that seeks to summarize the majority of the variation in the

		q = 5									q = 7								
		$n = 200$			$n = 1000$			$n = 5000$			$n = 200$			$n = 1000$			$n = 5000$		
	RS% $\gamma =$	0.5	0.75	1.0	0.5	0.75	1.0	0.5	0.75	1.0	0.5	0.75	1.0	0.5	0.75	1.0	0.5	0.75	1.0
	20%	0.08	0.09	0.33	0.05	0.03	0.35	*0.02*	0.01	0.36	0.67	0.71	0.45	0.63	0.77	0.44	0.69	0.74	0.48
$K = 3$	50%	*-0.10*	*-0.09*	*-0.07*	*-0.09*	*-0.14*	*-0.24*	*-0.07*	*-0.02*	*-0.15*	0.64	0.70	0.83	0.70	0.70	0.86	0.64	0.69	0.87
	80%	*-0.41*	*-0.37*	*-0.44*	*-0.34*	*-0.41*	*-0.47*	*-0.38*	*-0.43*	*-0.46*	0.60	0.65	0.81	0.64	0.66	0.79	0.61	0.66	0.8
	20%	0.09	0.19	0.50	0.14	0.19	0.55	0.14	0.15	0.54	0.75	0.85	0.47	0.70	0.82	0.48	0.70	0.79	0.49
$K = 5$	50%	0.12	0.15	0.18	0.12	0.14	0.21	0.13	0.14	0.26	0.71	0.75	0.93	0.70	0.76	0.94	0.70	0.76	0.92
	80%	0.10	0.12	0.07	0.07	0.11	0.12	0.08	0.11	0.10	0.70	0.72	0.85	0.68	0.72	0.85	0.68	0.72	0.85

Table 2.4: Average proportional improvement in the RMSE of the cleaned over the contaminated data. A two-sample Wilcoxon test for no difference in RMSE against the alternative hypothesis that the cleaned data significantly reduces the RMSE shows significant improvements ($\alpha = .95$) for all tests *except* those shown in italic print.

		q = 5									q = 7								
		$n = 200$			$n = 1000$			$n = 5000$			$n = 200$			$n = 1000$			$n = 5000$		
	r $m =$	10	20	30	10	20	30	10	20	30	10	20	30	10	20	30	10	20	30
	2	0.13	-0.04	-0.05	0.11	-0.05	-0.06	0.13	-0.03	-0.04	0.61	0.59	0.42	0.60	0.63	0.44	0.60	0.62	0.43
$K = 3$	3	0.06	-0.03	-0.05	0.08	-0.03	-0.04	0.07	-0.02	-0.04	0.54	0.41	0.17	0.59	0.46	0.18	0.60	0.46	0.20
	4	0.04	-0.02	-0.03	0.06	-0.01	-0.03	0.04	-0.03	-0.03	0.50	0.27	0.09	0.50	0.32	0.11	0.53	0.31	0.09
	2	0.10	-0.00	-0.02	0.13	0.00	-0.02	0.10	0.00	-0.02	0.64	0.50	0.20	0.73	0.54	0.21	0.66	0.53	0.20
$K = 5$	3	0.08	0.01	0.00	0.08	0.02	0.00	0.09	0.01	-0.00	0.61	0.30	0.11	0.63	0.28	0.10	0.66	0.32	0.08
	4	0.06	0.03	0.01	0.07	0.03	0.02	0.08	0.03	0.01	0.55	0.19	0.10	0.60	0.24	0.10	0.61	0.21	0.09

Table 2.5: Average proportional improvement in the RMSE when comparing the principal component loadings between the cleaned and contaminated data.

data by a few uncorrelated linear combinations of the original variables (the so-called principal components). Subsequent principal components each account for as much variation in the data as possible, subject to being uncorrelated with the previous components. PCA relies on the eigendecomposition of the covariance (or correlation) matrix, where the eigenvalue-eigenvector pairs give the variance accounted for (VAF) and the linear combination (also known as the principal component loadings) respectively for each component.

The following procedure was used to compare the PCA conducted on the true correlation matrix to those conducted on the correlation matrices of the cleaned and contaminated data respectively. First, a matrix of standard normal random numbers of dimension $m \times r$ is simulated, with r denoting the required rank of the PCA solution. The rows are then standardized to length one; denote this matrix by $\mathbf{L}$. The simulated correlation matrix is then $\mathbf{R} = \mathbf{L}\mathbf{L}'$, with the corresponding covariance matrix assumed to be $\Sigma = \sigma^2\mathbf{R}$. Here σ^2 is the same error variance as assumed in Section 2.4.1. Since the decomposition $\mathbf{R} = \mathbf{L}\mathbf{L}'$ is not unique, the eigendecomposition of $\mathbf{R}$ is used to re-express $\mathbf{R}$ as $\mathbf{R} = \mathbf{L}_r\mathbf{L}_r'$, where $\mathbf{L}_r$ is constructed from the first r eigenvectors and singular values of $\mathbf{R}$.

Second, a population mean vector $\boldsymbol{\mu}$ for the m items is simulated as uniform random numbers. The true underlying data for the respective respondents are then simulated from the multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ. The resultant matrix represents the uncontaminated data. Subsequently, response styles are added to arrive at the contaminated data. The same response styles as in Section 2.4.2 were used, the only difference being that the range $[L, U]$ of the splines was set to be the 1st and 99th percentiles of the sampled values respectively. Any spillovers outside the range of the splines are then added to the lowest or highest rating category. The interior knot t was fixed at the mean of the sampled values.

Finally, the constrained dual scaling method was applied to the contaminated data, assuming that the correct number of response styles K are known and using 15 and 50 random starts for $\mathbf{G}$ and $\boldsymbol{a}$ respectively. Based on this, a cleaned data set was constructed, from which the cleaned empirical correlation matrix, $\hat{\mathbf{R}}_c$ is obtained. Similarly, let $\hat{\mathbf{R}}_o$ be the empirical correlation matrix of the observed (i.e. the contaminated data). To compare the PCA solutions on these correlation matrices to that of $\mathbf{R}$, the decompositions $\hat{\mathbf{R}}_c \approx \mathbf{L}_c\mathbf{L}_c'$ and $\hat{\mathbf{R}}_o \approx \mathbf{L}_o\mathbf{L}_o'$ are constructed as before assuming that the researcher is able to identify the correct rank r of $\mathbf{R}$. The RMSE between $\mathbf{L}_r$ and $\mathbf{L}_c$ is then compared to that between $\mathbf{L}_r$ and $\mathbf{L}_o$ to determine whether the PCA structure of the cleaned data reflect the actual structure better or worse than the contaminated data.

For this simulation study, it was assumed that all groups are of equal size. The total sample size was varied over $n = 200$, 1000 and 5000 respondents as before, with either $K = 3$ or 5 response styles added. Again, either $q = 5$ or 7 response categories were studied, with $m = 10$, 20 or 30 items. The rank of Σ was either $r = 2, 3$ or 4. For each combination of these factors, 100 simulated data sets were analyzed.

The results are shown in Table 2.5, which displays the average relative improvement in the RMSE of the cleaned over the contaminated data. It is evident that the PCA structure is better reflected by the cleaned data when $q = 7$. From the table it can therefore be concluded that rating scales of more than 5 categories are ideal for the method. For rating scales with $q = 5$, marginal improvements are seen only for small numbers of items. It is reassuring that the method does not yield significantly worse result for less refined rating scales such as $q = 5$. The improvement of the method is greatest for small values of m. The number of segments K does not influence performance. Finally, the method performs best for low values of r, which corresponds to simpler underlying structures.

2.5 Application

To illustrate the method in an empirical application, consider data obtained from an anonymous multinational food and beverage conglomerate regarding an investigation of product perceptions for 20 similar products. These include in-house products as well as those of competitors. Data were collected from $n = 268$ panellists, who scored each product on 7 different sensory attributes using a 9-point Likert scale. Each product is rated on all 7 attributes (or, equivalently, items), so that there are 140 items collected in a data matrix with 268 rows and $m = 140$ columns. The Likert scale ranges from 1 (“low”) to 9 (“high”), and hence $q = 9$. Since these products are generally liked by consumers, acquiescence can be expected. The data set is available in coded form as part of the **cds** package (Schoonees, 2015b) for the statistical computing environment R (R Core Team, 2015). This can be obtained online from the Comprehensive R Archive Network (CRAN). The package contains the software used for all computations in the present paper.

The first step is to select K by inspecting the loss function through a scree plot. Consideration is also given to the curvature properties of the splines as well as how well the method separates groups of panellists who exhibit different distributions of rating scale use. It is expected that once spurious clusters are added at least two of the estimated response curves will be very similar, and/or that two groups will on aggregate use the rating scale in a very similar fashion. For each of $K = 1, 2, \ldots, 8$ groups, the algorithm was run from 50 different random starts for the grouping matrix $\mathbf{G}$, where appropriate. Also, 50 random starts for the alternating least squares (ALS) part of the algorithm was used. Appendix B gives insight into the effect of local optima for these data.

Figure 2.7 shows the resulting (rescaled) scree plot. There does not seem to be a clear “elbow” in the plot, although it is apparent that $K = 3, 4$ and 5 are the options requiring closer scrutiny. As K increases beyond 5 not much improvement in the loss function is observed.

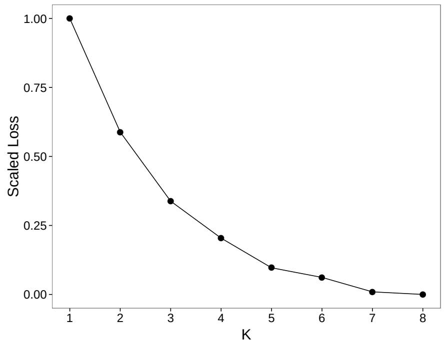


Figure 2.7: Scree plot for the sensory data.

The response mappings for the solutions $K = 1, \ldots, 8$ are displayed in Figure 2.8. In these plots the horizontal axis contains the original rating scale, while the vertical axis denotes the optimal scores assigned to the Likert scale. The area of the bubbles superimposed on the transformation plots indicate how often each rating category is used, aiding in the interpretation. A first observation is that (strictly, almost) all the detected response mappings have the characteristic convex shape of acquiescence. This means that all panellists have a tendency to use positive ratings frequently. The groups differ with respect to the intensity of the acquiescence.

Furthermore, the range of optimal scores that is assigned to each group, namely $\sum_{i=1}^{3} \alpha_{ik}$ in terms of the spline parameters set out in Sections 2.3.2 and 2.3.3, depends on the within-group variability of rating scale use. Groups where individual panellists' rating scale use show more variability from the group's aggregate rating scale use are assigned optimal scores with a wider range. Hence the method treats such groups, i.e. groups containing more individualistic respondents, as more informative as opposed to groups with more uniform response behaviour.

A closer look at the distribution of the rating scale use in the identified groups reveal that all groups in the solutions $K = 3, 4$ and 5 show visually different distributions, except group I and group III when $K = 5$. The relative frequencies with which each rating is used in each of the groups when $K = 5$ are shown in the barplots in Figure 2.9. It is obvious that groups I and III have very similar aggregate behaviour when $K = 5$. This is however not immediately apparent from the spline functions displayed in Figure 2.8, which assign different optimal scores to these groups.

A more formal comparison can also be made by using the Kullback-Leibler divergence (KL; e.g. Lehmann and Casella, 1998) between the distributions of different groups. This is also known as entropy distance and is often employed in the construction of classification trees (e.g. Breiman et al., 1984). It is an asymmetric measure of the dissimilarity between two density functions, the reference density f and another density g, which is defined as $E_f[\log(f(X)/g(X))]$. When

Group	I	II	III	IV	V
I	-	0.158	0.009	0.187	0.234
II	0.161	-	0.138	0.699	0.701
III	0.008	0.134	-	0.224	0.297
IV	0.166	0.606	0.202	-	0.053
V	0.231	0.680	0.317	0.065	-

Table 2.6: The Kullback-Leibler divergence between the groups when $K = 5$, based on the rating scale use per group. The distributions of the groups in the rows are treated as the respective reference distributions, f.

$f = g$, the entropy is zero; otherwise it is positive. For discrete distributions the integral is replaced by a summation. In the present context, let $\hat{f}_1, \ldots, \hat{f}_q$ and $\hat{g}_1, \ldots, \hat{g}_q$ denote the observed proportion of all answers in two different groups that use ratings $1, \ldots, q$ respectively. The observed KL divergence between these groups, with respect to $\hat{f}$, is then $\sum_{h=1}^{q} \hat{f}_h \log(\hat{f}_h/\hat{g}_h)$.

Assessing the pairwise KL divergence for all pairs of groups (and using both f and g as reference) show that indeed the above-mentioned two groups diverge the least among all pairs when $K = 5$ – see Table 2.6. Since the method is designed to detect groups with different aggregate rating scale use it can be concluded that the addition of a fifth group is spurious and therefore $K = 4$ is selected. The findings of Figure 2.9 are therefore supported by this analysis.

Consider the results for $K = 4$ groups. These four groups consist of 67, 71, 61 and 69 panellists respectively. The rating scale usage of these groups are displayed in Figure 2.10, panels (a) – (d). Figure 2.11 displays the optimal scores assigned to the ratings in the different groups as well as their curvature chart. The curvature chart includes an approximate 95% confidence ellipse constructed for the parameter estimates of 5000 data sets simulated under the assumption that no response styles exist. Any group falling outside this band therefore has a significantly nonlinear response mapping and hence a response style.

Group I represents acquiescence as mainly ratings 6 to 9 are used by panellists. There is a slight boundary effect, as also with the other groups, in that category 9 is used less often than category 8. Because the ratings 6 to 9 are frequently used, the optimal scores assigned to these are close to zero. The most meaningful optimal scores are assigned to the lower categories since when these are used it contains more information for this group of panellists. Overall the information provided by this group is low since the range of optimal scores assigned is very narrow. This is because the group members display low variability with respect to their rating scale use. This is evident from Figure 2.10 (e), which plots the frequency with which each rating is used per individual. Group II represents a more extreme acquiescence where categories 7 to 9 are often used. The range of assigned optimal scores, and hence information, is similarly narrow,

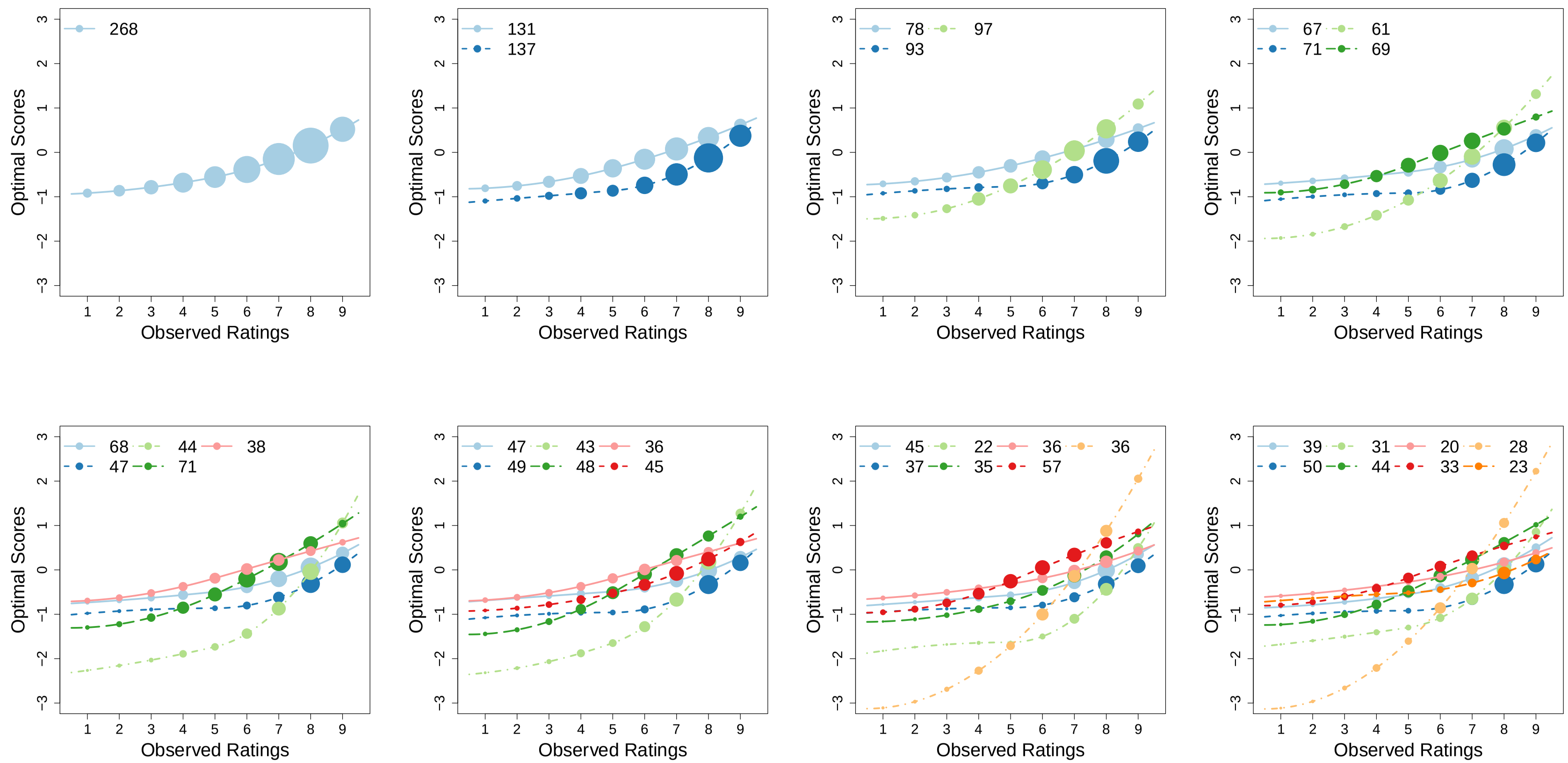
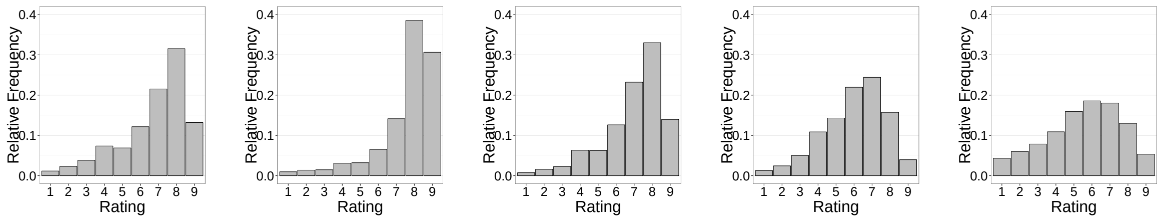


Figure 2.8: The estimated response mappings for $K = 1$ (top left) to 8 (bottom right) groups respectively. The area of the bubbles are proportional to how often that particular rating is used. The group sizes are also shown in a legend. Groups are labelled sequentially; the legend should be read vertically and then horizontally.



(a) *Group I* (b) *Group II* (c) *Group III* (d) *Group IV* (e) *Group V*

Figure 2.9: Relative aggregate frequencies of rating scale use in the identified groups when $K = 5$.

but shifted further to the left since the upper categories are used even more frequently. Since the response mapping is concave in the lower part of the domain there is a slight deviation from acquiescence towards an extreme response style.

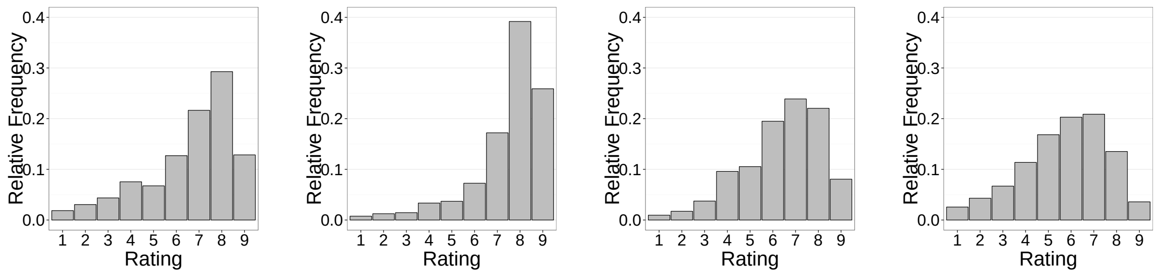
Groups III and IV both exhibit a mix of acquiescence and midpoint responding. This is evident from the relative frequencies in Figure 2.10 and the curvature chart in Figure 2.11 (b). In these groups generally ratings 4 to 8 are preferred. Based on the range of optimal scores assigned to them these consist of the panellists providing the most information. Especially Group III is endowed with the most meaningful spread of optimal scores, and can be seen in Figure 2.10 (g) to exhibit the most within-group variation.

Finally, consider the optimal scores assigned to the items as displayed in Figure 2.12. It is evident that Product R, and to a lesser extent Products N, D, E and F, received the lowest ratings. In contrast, Product P was the best performing one. By using a cleaned data set constructed by replacing the ratings by optimal scores further analyses can be conducted which are less influenced by the presence of the response styles.

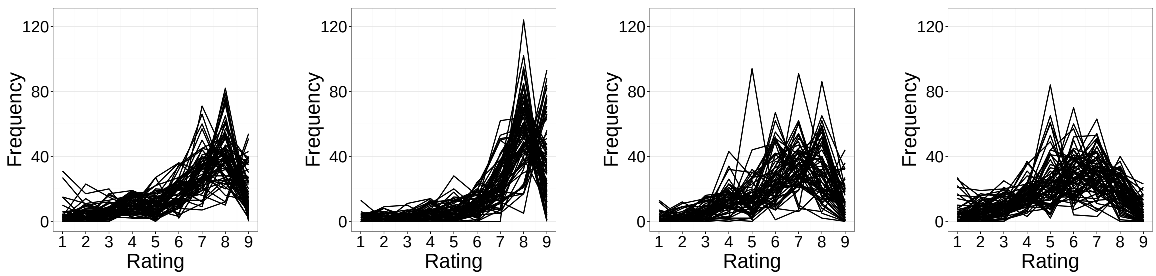
2.6 Conclusions

A method that relies on the properties of dual scaling for successive category data to detect response styles in categorical data was presented. It combines newly suggested spline models for four main types of response styles with the original dual scaling method to construct optimal scores for the boundaries between rating categories. These optimal scores are sensitive to the presence of response styles. The method was adapted to allow for multiple response style groups by utilizing a k-means type procedure, which is combined with a constrained alternating least squares algorithm using nonnegative least squares to fit the model.

Both the ability of the method to detect response styles and the improvement in correlation structure that results from a cleaned data set where ratings are replaced by optimal scores were

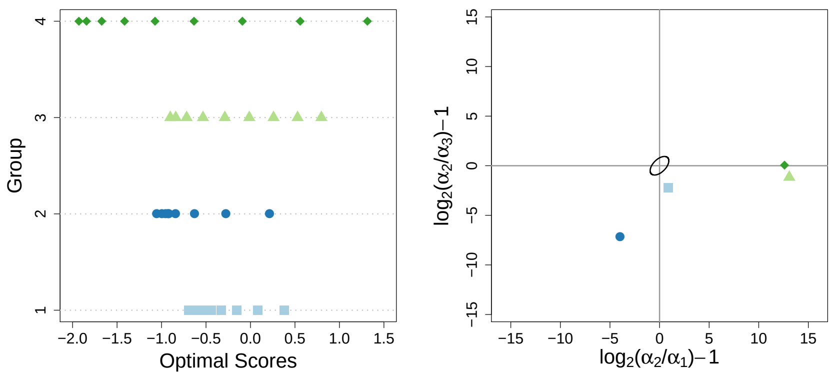


(a) *Group I* (b) *Group II* (c) *Group III* (d) *Group IV*



(e) *Group I* (f) *Group II* (g) *Group III* (h) *Group IV*

Figure 2.10: (a) - (d): Relative frequencies of rating scale use for the chosen solution $K = 4$; and (e) - (h) Variability of rating scale use within these groups, with each line representing a single individual.



(a) (b)

Figure 2.11: (a) Optimal scores assigned to the $K = 4$ response style groups, from rating 1 (left) to rating 9 (right). (b) Curvature plot similar to Figure 2.3 for the four groups, with the axes now transformed to obtain a more symmetrical plot. The ellipse in the centre is an approximate 95% confidence ellipse for no response style.

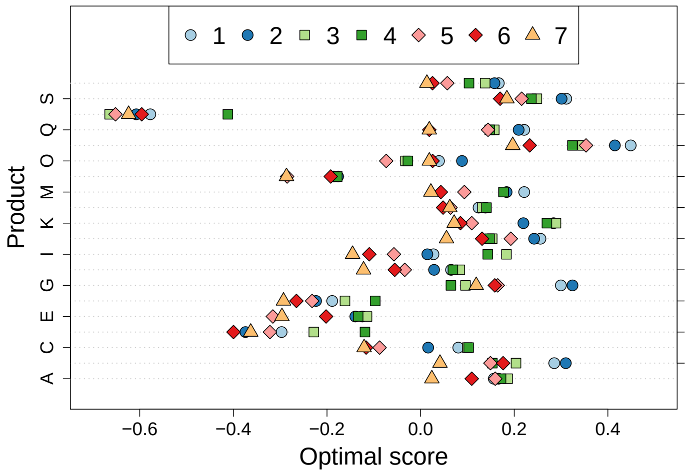


Figure 2.12: Optimal scores for each of the 7 questions, separated by product and with similar items depicted by the same colours.

studied. It was found that using 30 or more items and a rating scale of 7 or more categories yields great improvements in the classification of individuals to different response style groups. When fewer rating categories are used other factors become important, such as the extent to which response styles are present in the data. Also, when using a 7-point scale or more, the resulting cleaned data provide a more accurate description of the true substantial content in the data, after accounting for different response styles. The use of the method to identify respondents who provide similar amounts of information in their responses to a survey was illustrated on an empirical data set.

The number of response style groups to retain was selected on the grounds of a scree plot of the loss function, combined with the distribution of rating scale use in the different response style groups. It remains to be seen whether a more formal selection method can be derived. Other grounds for further research include alternatives for or additional restriction to the spline functions, and whether more freedom is needed by allowing for differences between the m object scores in different groups. An overview an adapted algorithm that can handle missing observations, missing by design or otherwise, is given in Appendix C.

Appendix

Appendix A

Here an overview of the derivation of Algorithm 1 is provided (specifically, steps 9 – 14). Consider expanding the criterion of Equation (2.12), assuming without loss of generality that the proportionality constant $c = 1$:

$$\begin{aligned} L(\boldsymbol{a}, \mathbf{B}, \mathbf{G}) &= \|\mathbf{F}_r^* - \frac{1}{2}(m+q-2)\sum_{k=1}^{K} \mathbf{D}_{\mathbf{g}_k}\mathbf{a}\mathbf{b}_k'\|^2 \\ &= \operatorname{tr}\mathbf{F}_r^{*\prime}\mathbf{F}_r^* + \frac{1}{4}(m+q-2)^2\sum_{k=1}^{K}\boldsymbol{b}_k'\boldsymbol{b}_k\boldsymbol{a}'\mathbf{D}_{\mathbf{g}_k}\boldsymbol{a} - (m+q-2)\sum_{k=1}^{K}\boldsymbol{b}_k'\mathbf{F}_r^{*\prime}\mathbf{D}_{\mathbf{g}_k}\boldsymbol{a}. \end{aligned} \tag{2.16}$$

This derivation uses $\mathbf{F}_r^* = \mathbf{F}_r - \frac{1}{2}(m+q-2)\mathbf{1}\mathbf{1}'$, the fact that $\mathbf{D}_{\mathbf{g}_k}$ is idempotent and that $\mathbf{D}_{\mathbf{g}_k}\mathbf{D}_{\mathbf{g}_l} = \mathbf{0}\ \forall\ k \neq l$, as well as the properties of the matrix trace operator. Note that the first term does not depend on the model parameters and hence are not used in the optimization algorithm.

Now, consider optimizing $\boldsymbol{a}$ and $\mathbf{B}$ when $\mathbf{G}$ is fixed. It follows from Equation (2.16) that, given a starting configuration of $\boldsymbol{a}$, the relevant loss function to be minimized for finding a new $\mathbf{B}$ is proportional to:

$$\begin{aligned} L(\mathbf{B} \mid \boldsymbol{a}, \boldsymbol{G}) &= \sum_{k=1}^{K}\left[\frac{1}{4}(m+q-2)^2\boldsymbol{b}_k'\boldsymbol{b}_k\boldsymbol{a}'\mathbf{D}_{\mathbf{g}_k}\boldsymbol{a} - (m+q-2)\boldsymbol{b}_k'\mathbf{F}_r^{*\prime}\mathbf{D}_{\mathbf{g}_k}\boldsymbol{a}\right] \\ &= \frac{1}{4}(m+q-2)^2\sum_{k=1}^{K}\|(\boldsymbol{a}'\mathbf{D}_{\mathbf{g}_k}\boldsymbol{a})^{1/2}\boldsymbol{b}_k - \frac{2}{m+q-2}(\boldsymbol{a}'\mathbf{D}_{\mathbf{g}_k}\boldsymbol{a})^{-1/2}\mathbf{F}_r^{*\prime}\mathbf{D}_{\mathbf{g}_k}\boldsymbol{a}\|^2 + c_1 \end{aligned} \tag{2.17}$$

where the constant c_1 depends only on K and $\mathbf{F}_r^*$. Hence $\mathbf{B}$, and, more specifically, the parameters $\boldsymbol{b}_1$ and $\boldsymbol{\alpha}_k,, k = 1, 2, \ldots, K$, are updated by minimizing:

$$\sum_{k=1}^{K}\|(\boldsymbol{a}'\mathbf{D}_{\mathbf{g}_k}\boldsymbol{a})^{1/2}\boldsymbol{b}_k - \frac{2}{m+q-2}(\boldsymbol{a}'\mathbf{D}_{\mathbf{g}_k}\boldsymbol{a})^{-1/2}\mathbf{F}_r^{*\prime}\mathbf{D}_{\mathbf{g}_k}\boldsymbol{a}\|^2. \tag{2.18}$$

Now, recall that $\boldsymbol{b}_k = (\boldsymbol{b}_1^{'}, \boldsymbol{b}_{2k}^{'})^{'}$ with $\boldsymbol{b}_{2k} = \mathbf{M}\boldsymbol{\alpha}_k$, so that the relevant parameters in the $\{\boldsymbol{b}_k\}_{k=1}^K$ is $\boldsymbol{b}_1$ and $\{\boldsymbol{\alpha}_k\}_{k=1}^K$. These parameters must therefore be updated using the loss function in Equation (2.18). Let $w_k = (\boldsymbol{a}^{'}\mathbf{D}_{\boldsymbol{g}_k}\boldsymbol{a})^{-1/2}$ and $\frac{2}{m+q-2} w_k \mathbf{F}_r^{*'} \mathbf{D}_{\boldsymbol{g}_k}\boldsymbol{a} = (\boldsymbol{v}_{1k}^{'}, \boldsymbol{v}_{2k}^{'})^{'}$. Since

$$\|(\boldsymbol{x}_1^{'}, \boldsymbol{x}_2^{'})^{'} - (\boldsymbol{y}_1^{'}, \boldsymbol{y}_2^{'})^{'}\|^2 = \|\boldsymbol{x}_1 - \boldsymbol{y}_1\|^2 + \|\boldsymbol{x}_2 - \boldsymbol{y}_2\|^2,$$

it follows that

$$L(\boldsymbol{b}_1, \boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_K \,|\, \boldsymbol{a}, \boldsymbol{G}) = \sum_{k=1}^{K} \|w_k^{-1}\boldsymbol{b}_1 - \boldsymbol{v}_{1k}\|^2 + \sum_{k=1}^{K} \|w_k^{-1}\mathbf{M}\boldsymbol{\alpha}_k - \boldsymbol{v}_{2k}\|^2. \tag{2.19}$$

Therefore $\boldsymbol{b}_1$ can be updated by minimizing the first summation in Equation (2.19) by OLS independently of $\{\boldsymbol{\alpha}_k\}_{k=1}^K$. Since $\alpha_{ik} \geq 0$ for all i and k, the latter vectors are updated for each k by using NNLS to minimize each of the individual elements of the second summation.

Appendix B

Here a short exposition is given of the spread of local optima for the empirical example. Specifically, the variability of the loss function for the 50 random starts of $\mathbf{G}$ is shown in Figure B1. The curves are ordered from $K = 2$ at the top to $K = 8$ at the bottom. It is evident that only a single random start typically produces the best result. In general, the local optima is less stable for larger values of K, as can be expected. It is evident from this example that attention must be paid to the number of random starts used in empirical applications of such algorithms. These results suggests that the "best of 20 random starts" rule often favoured by practitioners of K-means clustering may not suffice (Hand and Krzanowski, 2005); a pragmatic approach is required.

Appendix C

This Appendix gives an overview of how the constrained dual scaling (CDS) algorithm, outlined above, can be used as a building block for an algorithm that extends this procedure so that missing data can be handled. Missing data can occur for various reasons (see Little and Rubin, 2002, for example). Here we are specifically interested in data missing by design, although more traditional forms of missing data can also be handled. For example, in a chocolate sensory study, the overall liking, texture, taste and melt rate of a range of chocolates might be assessed by a consumer panel. Incomplete experimental designs, which expose the consumers only to

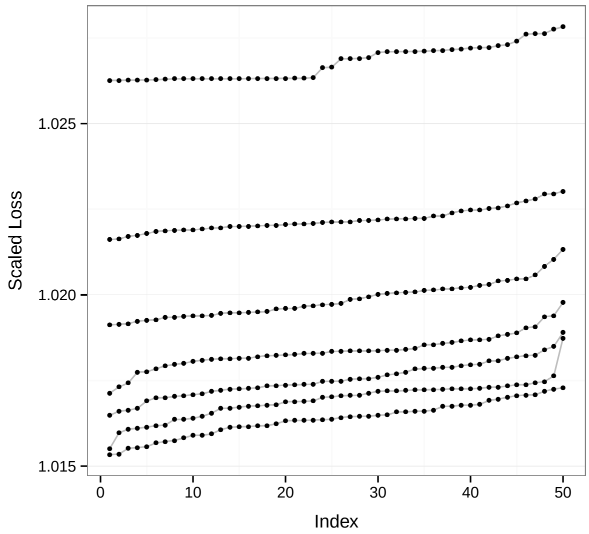


Figure B1: The spread of the loss values (scaled by a constant) for $K = 2, \ldots, 8$ in the empirical example, for 50 different starting configurations of $\mathbf{G}$.

a subset of all products, are often used in such studies. This may be done because of concerns regarding respondent fatigue when all products must be assessed, there may not be enough product samples available for all subjects to assess all products, or the researchers may face budget or time limitations.

It may very well be that the respondents in incomplete studies based on rating scales display different response styles. But the CDS algorithm cannot handle incomplete designs without modification. We therefore briefly present here an incomplete CDS (ICDS) algorithm which embeds the CDS algorithm in a majorization procedure originally proposed by Kiers (1997). This allows for arbitrary incomplete designs to be handled, including conventional missing data.

In the case of incomplete data, the same transformation from $\mathbf{X}$ to $\mathbf{F}_r^*$, as outlined in Section 2.3.1, must be applied. Only the observed products are used when forming rankings. But to ensure that the range of rankings assigned are commensurate across rows (persons), the same number of products must be evaluated by each respondent. This will be the case for randomized block designs, but not when data are missing for other reasons. Although it is unlikely that small differences in the number of products assessed by the different persons will have a large effect, larger differences can have an effect. We assume in the sequel that this is not an issue, as will be the case with most well-designed studies.

For incomplete designs, the matrices **X**, **T**, **S** and consequently also $\mathbf{F}_r^*$ will have missing blocks. A very simple balanced incomplete block design might for example have

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_{11} & \texttt{NA} & \mathbf{X}_{13} & \mathbf{X}_{14} \\ \mathbf{X}_{21} & \mathbf{X}_{22} & \mathbf{X}_{23} & \texttt{NA} \\ \texttt{NA} & \mathbf{X}_{32} & \mathbf{X}_{33} & \mathbf{X}_{34} \\ \mathbf{X}_{41} & \mathbf{X}_{42} & \texttt{NA} & \mathbf{X}_{44} \end{bmatrix}. \tag{2.20}$$

Here $\mathbf{X}_{ij}$ is the matrix that pertains to the answers of the respondents assigned to block i on product j, and `NA` indicates incomplete data. The columns of $\mathbf{X}_{ij}$ correspond to the questions used to assess each of the products. Converting this **X** to $\mathbf{F}_r^*$ results similarly in missing blocks within $\mathbf{F}_r^*$.

We handle missing data by introducing a weight matrix **W** which assigns zero weights to all missing or incomplete observations, and nonnegative weights to all other elements of $\mathbf{F}_r^*$. The nonnegative weights can be either equal to one for all nonmissing observations, or a more refined specification of nonnegative values. In this way, missing data does not play any role in the model fit. The ICDS problem therefore directly modifies (2.5) to seek the minimization of the weighted loss function given by

$$L(\boldsymbol{a}, \{\boldsymbol{b}_k\}, \mathbf{G}) = \left\| \left(\mathbf{F}_r^* - \frac{1}{2}(m+q-2) \sum_{k=1}^{K} \mathbf{D}_{g_k} \boldsymbol{a} \boldsymbol{b}_k' \right) * \mathbf{W} \right\|^2 \tag{2.21}$$

over $\boldsymbol{a}$, $\{\boldsymbol{b}_k\}$ and **G**. Here the $*$ operator denotes the Hadamard (or elementwise) product.

To minimize (2.21), we derive an iterative procedure based on Kiers (1997), which uses the CDS algorithm as building block. Specifically, Kiers (1997) discusses majorization algorithms for the weighted least squares (WLS) problem seeking minimization of

$$h(\mathbf{M}|\mathbf{Z}, \mathbf{W}) = \|(\mathbf{Z} - \mathbf{M}) * \mathbf{W}\|^2 \tag{2.22}$$

$$= \|\mathbf{D}_w(\mathrm{Vec}\,(\mathbf{Z}) - \mathrm{Vec}\,(\mathbf{M}))\|^2, \tag{2.23}$$

over a model **M**, where **W** is a weight matrix and $\mathbf{D}_w = \mathrm{diag}\,(\mathrm{Vec}\,(\mathbf{W}))$. In our context then, $\mathbf{Z} = \mathbf{F}_r^*$ and $\mathbf{M} = \frac{1}{2}(m+q-2)\sum_{k=1}^{K} \mathbf{D}_{g_k} \boldsymbol{a} \boldsymbol{b}_k'$. The idea is to solve this WLS problem by repeatedly solving an ordinary least squares (OLS) problem by standard methods. The OLS problem changes after each iteration by updating the target to which the model is fitted after each iteration. This is done by defining a function $k(\mathbf{M}|\mathbf{M}_l, \mathbf{Z}, \mathbf{W})$ that majorizes $h(\mathbf{M}|\mathbf{Z}, \mathbf{W})$. The majorizing function is a function that is everywhere larger than or equal to h, and equal to h at the support point $\mathbf{M} = \mathbf{M}_l$. Here $\mathbf{M}_l$ is the value of **M** at iteration l. Minimizing the majorizing function k is simpler than minimizing h, and assures that the value of h decreases at each step too.

The majorizing function for the generic criterion (2.22) is of the form (Kiers, 1997)

$$k(\mathbf{M}|\mathbf{M}_l, \mathbf{Z}, \mathbf{W}) = \alpha + w_m^2 ||\mathbf{M}_l + w_m^{-2}\mathbf{W}^{(2)} * \mathbf{Z} - w_m^{-2}\mathbf{W}^{(2)} * \mathbf{M}_l - \mathbf{M}||^2 \tag{2.24}$$

$$= \alpha + w_m^2 ||\widetilde{\mathbf{Z}}_l - \mathbf{M}||^2. \tag{2.25}$$

Here α is a constant, $\mathbf{W}^{(2)} = \mathbf{W} * \mathbf{W}$, and w_m^2 is the largest eigenvalue of $\mathbf{D}_w^2$, which is just the largest squared element of $\mathbf{W}$. Equation (2.25) shows that at step $l + 1$ of the majorization algorithm, a simple least squares problem with working target $\widetilde{\mathbf{Z}}_l = \mathbf{M}_l + w_m^{-2}\mathbf{W}^{(2)} * \mathbf{Z} - w_m^{-2}\mathbf{W}^{(2)} * \mathbf{M}_l$ must be solved to find $\mathbf{M}_{l+1}$. A sequence of monotonically decreasing function values for h is constructed by the following inequality:

$$h(\mathbf{M}_{l+1}|\mathbf{Z}, \mathbf{W}) \leq k(\mathbf{M}_{l+1}|\mathbf{M}_l, \mathbf{Z}, \mathbf{W}) < k(\mathbf{M}_l|\mathbf{M}_l, \mathbf{Z}, \mathbf{W}) = h(\mathbf{M}_l|\mathbf{Z}, \mathbf{W}). \tag{2.26}$$

Monotonic convergence occurs, but only to local minima. Hence several random starts $\mathbf{M}_0$ are required to increase the likelihood of finding the global optima.

We can hence derive the majorizing function for the ICDS criterion (2.21) as

$$\begin{aligned} k(\mathbf{M}|\mathbf{M}_l, \mathbf{F}_r^*, \mathbf{W}) &= \alpha + w_m^2 \left\|\mathbf{M}_l + w_m^{-2}\mathbf{W}^{(2)} * (\mathbf{F}_r^* - \mathbf{M}_l) - \mathbf{M}\right\|^2 \\ &= \alpha + w_m^2 \left\|w_m^{-2}\mathbf{W}^{(2)} * \mathbf{F}_r^* + \mathbf{M}_l * \left(\mathbf{1}\mathbf{1}' - w_m^{-2}\mathbf{W}^{(2)}\right) - \mathbf{M}\right\|^2 \\ &= \alpha + w_m^2 \left\|\widetilde{\mathbf{F}}_{rl}^* - \mathbf{M}\right\|^2. \end{aligned} \tag{2.27}$$

The working target at each iteration is therefore $\widetilde{\mathbf{F}}_{rl}^*$. Finding $\mathbf{M}$ which minimizes (2.27) is achieved by the CDS algorithm. Note though that random starts for $\boldsymbol{a}$ should only be employed once for each random start of $\mathbf{G}$ in the CDS algorithm for the majorization procedure to work. The constant α is inconsequential for optimization and can be safely ignored.

A starting value $\mathbf{M}_0$ can be obtained in at least three ways: (1) by random generation; (2) by setting $\mathbf{M}_0 = \mathbf{0}$; or (3) by using for example a corresponding principal components analysis (PCA) solution. The ICDS algorithm proceeds as follows:

Step 1. Initialize $\mathbf{M}_l$ as $\mathbf{M}_0$ and set $l = 0$.

Step 2. Compute $\widetilde{\mathbf{F}}_{rl}^* = \mathbf{M}_l + w_m^{-2}\mathbf{W}^{(2)} * (\mathbf{F}_r^* - \mathbf{M}_l)$.

Step 3. Find the update $\mathbf{M}_{l+1}$ where $\mathbf{M}_{l+1}$ decreases (or minimizes) $||\widetilde{\mathbf{F}}_{rl}^* - \mathbf{M}||^2$ by CDS. Set $l \leftarrow l + 1$.

Step 4. Repeat Steps 2 and 3 until numerical convergence in (2.21).

There are several options for the implementation of Step 3, as well as some caveats. Random starts for $\boldsymbol{a}$ can only be done in the first CDS iteration since parameters must be updated sequentially between majorization iterations for the majorization procedure to work. Furthermore, it is not necessary to do a full CDS update each time Step 3 is executed – indeed it is recommended not to be greedy in this regard. Therefore we choose to do only a single update of $\boldsymbol{a}, \{\boldsymbol{b}_k\}$ and $\mathbf{G}$ each time Step 3 is repeated. Experimentation shows that using a few random starts for $\boldsymbol{a}$ (such as 5) in the first majorization step can be advantageous without incurring a large computational overhead.

CHAPTER 3

Least-Squares Bilinear Clustering of Three-Way Data

3.1 Introduction

Three-way data appear regularly in research, such as when a number of respondents are asked to rate several objects based on a set of characteristics in a marketing survey. Such data can be collected in a three-way array, with slices along one dimension containing the data matrices for the different individuals. Several models have been formulated for least-squares approximation of such three-way arrays, such as CANDECOMP/PARAFAC (Carroll and Chang, 1970; Harshman, 1970) and TUCKALS3 (Kroonenberg and De Leeuw, 1980). Much research is available on theoretical aspects of these models, nonuniqueness properties, and estimation algorithms (e.g. Kruskal, 1977; Kiers and Krijnen, 1991; Ten Berge and Sidiropoulos, 2002; Faber et al., 2003). However, these models suffer from disadvantages including that they are relatively complicated, hard to fit and that the interpretation of their graphical representations require sound knowledge of the models (e.g. Kiers, 2000; Krijnen et al., 2008).

Here our aim is to simplify the analysis of such data by using a simple model for the two-way matrix slices of the three-way data array, combined with clustering over the third way. The model we elaborate on is the well-known bilinear decomposition of a matrix into an overall mean, row means, column means, and a low-rank decomposition of the remaining row-column interactions (see Gower and Hand, 1996, for example). This can be viewed as a two-way analysis of variance-type decomposition into an overall mean effect, marginal effects and row-column interaction effects, and is also known as a biadditive model.

We gratefully acknowledge the input of Karl Jöreskog, whose question at a seminar in Uppsala sowed the seeds for this research.

The novelty of the least-squares bilinear clustering model proposed in this paper, henceforth referred to as LSBCLUST, is that the clusters over the third way of the array are introduced jointly for each of the terms in the bilinear decomposition. Importantly, however, we show that this clustering can be done separately for each of the terms because of the orthogonality of the terms in the bilinear model (with one exception – see Section 3.2.1). This property greatly reduces the computational cost of the method, and aids interpretation of the results. Additionally, we show how to construct biplots for the interaction effects so that these can be easily interpreted, and how different choices for the identifiability constraints in the bilinear model lead to different submodels.

The main ideas of LSBCLUST are summarized in Figure 3.1. The data array represented by **X** is decomposed into overall means, row means, column means and row-column interaction. Each of these components are modelled clusterwise, with different sets of clusters introduced for each of the components of the decomposition. The clusters are represented by different colours and/or labelled effects in the figure. Within each cluster the effects are determined by modelling the cluster means, and in the case of the interactions low-rank decompositions of the cluster means are used. This enables us to elicit only the most prominent structure in the interactions and improves interpretability by allowing biplots to be constructed. We note that adjusting for different row and column margins pre-analysis is routinely done in correspondence analysis (e.g. Greenacre, 2007).

The problem is formulated as a least-squares loss function in Section 3.2. We also discuss the separability of the different clustering problems in this section. In Section 3.3 we develop an alternating least-squares algorithm for minimizing the loss function and discuss the construction of biplots for the interactions, as well as model selection. Section 3.4 contains two illustrative empirical applications, and Section 3.5 concludes.

3.2 Problem Formulation

Consider as starting point data consisting of J objects that have been rated by N individuals on K attributes. Define the indices i, j and k such that these identify the respondents, objects and attributes respectively. It follows that $i = 1, \dots, N$, $j = 1, \dots, J$ and $k = 1, \dots, K$. For each individual, form the $J \times K$ matrix $\mathbf{X}_i$ where each row j gives the scores of person i for object j on all K attributes. For now, we assume that there are no missing values and that the values in the $\mathbf{X}_i$'s are commensurable, that is, measured on the same scale. The collection of these $\mathbf{X}_i$'s can also be viewed as a three-way array, with the objects, attributes and individuals constituting the three dimensions.

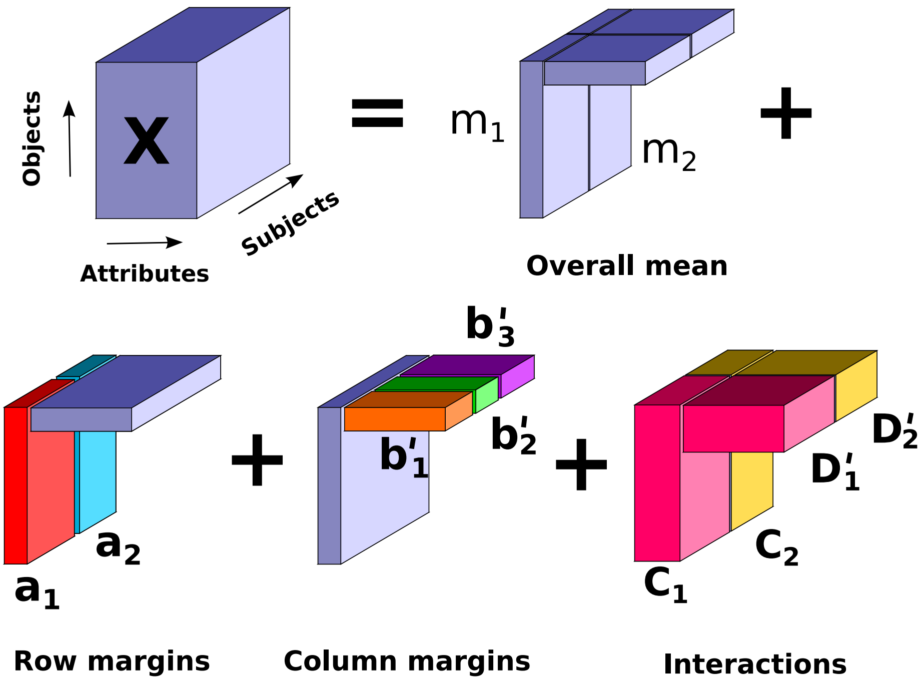


Figure 3.1: A diagram illustrating our LSBCLUST approach to analysing three-way data. Colours and/or labels indicate different clusters, while pale blue indicates constant vectors or matrices. Different clusters are introduced for all four parts of the bilinear decomposition of the two-way matrix slices. The cluster means are modelled within each such cluster, and the clustering and modelling steps are undertaken jointly.

The $\mathbf{X}_i$'s are modelled by a variation of the bilinear (or biadditive) decomposition in the least-squares loss function

$$L(m,\boldsymbol{a},\boldsymbol{b},\mathbf{C},\mathbf{D}) = \sum_{i=1}^{N} \left\| \mathbf{X}_i - \left(m\mathbf{1}_J\mathbf{1}_K' + \boldsymbol{a}\mathbf{1}_K' + \mathbf{1}_J\boldsymbol{b}' + \mathbf{C}\mathbf{D}' \right) \right\|^2 . \tag{3.1}$$

Here m denotes an overall constant or average effect, $\boldsymbol{a}$ the row effects, $\boldsymbol{b}$ the column effects, $\mathbf{CD'}$ a low-rank decomposition of the interaction effects between rows and columns, and $\|\cdot\|$ the Frobenius norm. Also, $\mathbf{1}_K$ denotes the length-K vector of ones. Representing the interaction effects as inner products permit these to be displayed in biplots (Gower and Hand, 1996; Gower et al., 2011). To ensure uniqueness of the model, the usual sum-to-zero constraints $\boldsymbol{a}'\mathbf{1}_J = \boldsymbol{b}'\mathbf{1}_K = 0$ and $\mathbf{1}_J'\mathbf{C} = \mathbf{1}_K'\mathbf{D} = \mathbf{0}$ must be imposed. Additionally, the columns of $\mathbf{C}$ and $\mathbf{D}$ are required to be orthogonal and of equal length (for more information, see Denis and Gower, 1994). Model (3.1) has an analytic solution.

Our main contribution is to embed (3.1) in a general modelling framework by adding several different types of clusters (latent classes or segments) while allowing for a variety of parameter constraints. Different choices for these constraints lead to different submodels. Clusters are introduced to separate the respondents on four different characteristics: the first with respect to the overall average, the second for the row effects, the third for the column effects, and the fourth for the interaction effects. For modelling the interactions, we allow for three options: (a) a common $\mathbf{C}_1$ for representing the rows and a differential $\mathbf{D}_u$ for each interaction cluster indexed by u; (b) a differential $\mathbf{C}_u$ but common $\mathbf{D}_1$ for each interaction cluster; or (c) both $\mathbf{C}_u$ and $\mathbf{D}_u$ are specific to the interaction cluster. Options (a) and (b) are more parsimonious than (c) and are particularly useful for linking the graphical representations (biplots) of the clusterwise interaction effects through common row or column representations. In Section 3.3.3 we consider generalized Procrustes analysis (Gower and Dijksterhuis, 2004) as an interpretative aid for option (c).

Let $\mathbf{G}^{(\mathrm{o})}$ be the $N \times R$ matrix of cluster memberships for the overall constant, which has $g_{ir}^{(\mathrm{o})} = 1$ if person i belongs to cluster r and $g_{ir}^{(\mathrm{o})} = 0$ otherwise ($r = 1, 2, \ldots, R$). Similarly, $\mathbf{G}^{(\mathrm{r})}$ is the $N \times S$ matrix of cluster memberships for the row effects, $\mathbf{G}^{(\mathrm{c})}$ the $N \times T$ matrix of cluster memberships for the column effects, and $\mathbf{G}^{(\mathrm{i})}$ the $N \times U$ matrix of cluster memberships for the interaction effects. Now, by incorporating the clustering, the least-squares loss function becomes

$$\begin{aligned}
&L(\mathbf{G}^{(o)},\mathbf{G}^{(r)},\mathbf{G}^{(c)},\mathbf{G}^{(i)},\boldsymbol{m},\mathbf{A},\mathbf{B},\mathbf{C},\mathbf{D})\\
&\quad=\sum_{i=1}^{N}\sum_{r=1}^{R}\sum_{s=1}^{S}\sum_{t=1}^{T}\sum_{u=1}^{U} g_{ir}^{(o)}g_{is}^{(r)}g_{it}^{(c)}g_{iu}^{(i)}\left\|\mathbf{X}_i-\left(m_r\mathbf{1}_J\mathbf{1}_K'+\boldsymbol{a}_s\mathbf{1}_K'+\mathbf{1}_J\boldsymbol{b}_t'+\mathbf{C}_u\mathbf{D}_u'\right)\right\|^2\\
&\quad=\sum_{i,r,s,t,u} g_{ir}^{(o)}g_{is}^{(r)}g_{it}^{(c)}g_{iu}^{(i)}L(i|r,s,t,u).
\end{aligned}\tag{3.2}$$

Here $\boldsymbol{m}=(m_1,\ldots,m_R)'$, $\mathbf{A}=\begin{bmatrix}\boldsymbol{a}_1 & \cdots & \boldsymbol{a}_S\end{bmatrix}$, $\mathbf{B}=\begin{bmatrix}\boldsymbol{b}_1 & \cdots & \boldsymbol{b}_T\end{bmatrix}$, and $\mathbf{C}'=\begin{bmatrix}\mathbf{C}_1' & \cdots & \mathbf{C}_U'\end{bmatrix}$ and $\mathbf{D}'=\begin{bmatrix}\mathbf{D}_1' & \cdots & \mathbf{D}_U'\end{bmatrix}$. In case options (a) or (b) are used, $\mathbf{C}_u\mathbf{D}_u'$ in (3.2) should be replaced by $\mathbf{C}_1\mathbf{D}_u'$ or $\mathbf{C}_u\mathbf{D}_1'$ respectively.

3.2.1 Separability of Different Cluster Types

Here we show that the joint clustering can be simplified significantly into four separate clustering problems. Define the matrix $\mathbf{J}_J^{(\delta)}$ to be of the form

$$\mathbf{J}_J^{(\delta)}=\mathbf{I}_J-\frac{\delta}{J}\mathbf{1}_J\mathbf{1}_J',\tag{3.3}$$

where $\delta\in\{0,1\}$. Hence when $\delta=1$, $\mathbf{J}_J^{(\delta)}$ is the J-dimensional centring matrix; otherwise it reduces to the identity matrix. The shorthand notation $\mathbf{J}_J=\mathbf{J}_J^{(1)}$ will also be used. Applying a generalization of the well-known double-centring operation to $\mathbf{X}_i$ yields

$$\mathbf{J}_J^{(\delta_1)}\mathbf{X}_i\mathbf{J}_K^{(\delta_2)}=\left(\mathbf{I}_J-\frac{\delta_1}{J}\mathbf{1}_J\mathbf{1}_J'\right)\mathbf{X}_i\left(\mathbf{I}_K-\frac{\delta_2}{K}\mathbf{1}_K\mathbf{1}_K'\right)'.\tag{3.4}$$

Expanding and rearranging the above implies that

$$\mathbf{X}_i=-\delta_1\delta_2\frac{\mathbf{1}_J'\mathbf{X}_i\mathbf{1}_K}{JK}\mathbf{1}_J\mathbf{1}_K'+\frac{\delta_1}{J}\mathbf{1}_J\mathbf{1}_J'\mathbf{X}_i+\frac{\delta_2}{K}\mathbf{X}_i\mathbf{1}_K\mathbf{1}_K'+\mathbf{J}_J^{(\delta_1)}\mathbf{X}_i\mathbf{J}_K^{(\delta_2)}.\tag{3.5}$$

Furthermore, (3.5) can be rewritten as

$$\begin{aligned}
\mathbf{X}_i=&(\delta_1\delta_3+\delta_2\delta_4-\delta_1\delta_2)\frac{\mathbf{1}_J'\mathbf{X}_i\mathbf{1}_K}{JK}\mathbf{1}_J\mathbf{1}_K'+\frac{\delta_1}{J}\mathbf{1}_J\mathbf{1}_J'\mathbf{X}_i\mathbf{J}_K^{(\delta_3)}\\
&+\frac{\delta_2}{K}\mathbf{J}_J^{(\delta_4)}\mathbf{X}_i\mathbf{1}_K\mathbf{1}_K'+\mathbf{J}_J^{(\delta_1)}\mathbf{X}_i\mathbf{J}_K^{(\delta_2)}
\end{aligned}\tag{3.6}$$

by inserting additional centring matrices operating on the row and column means. The binary variables in $\boldsymbol{\delta}'=(\delta_1,\delta_2,\delta_3,\delta_4)$ act as switches which determine which centring constraints are applied and are chosen by the user according to the problem at hand. The choice of these switches determines the biadditive model to be fitted. Although there are 16 different choices for the δ's, there are fewer unique specifications. For example, notice that if $\delta_1 = 0$, the second term on the right-hand side of (3.6) vanishes so that δ_3 does not have any effect. The choices $\delta_3 = 0$ and $\delta_3 = 1$ therefore leads to the same decomposition when $\delta_1 = 0$. A similar relationship exists between δ_2 and δ_4.

By using the same centring approach in the model specification, we can drop the sum-to-zero constraints from the formulation. This is done by redefining the terms in the summation in (3.2) as

$$\begin{aligned} L(i|r,s,t,u) = \Big\| \mathbf{X}_i - \Big(&(\delta_1\delta_3 + \delta_2\delta_4 - \delta_1\delta_2) m_r \mathbf{1}_J \mathbf{1}_K' + \delta_2 \mathbf{J}_J^{(\delta_4)} \boldsymbol{a}_s \mathbf{1}_K' \\ &+ \delta_1 \mathbf{1}_J \boldsymbol{b}_t' \mathbf{J}_K^{(\delta_3)} + \mathbf{J}_J^{(\delta_1)} \mathbf{C}_u \mathbf{D}_u' \mathbf{J}_K^{(\delta_2)} \Big) \Big\|^2 . \end{aligned} \tag{3.7}$$

Note that using $\delta_1 = 1$, $\delta_2 = 1$, $\delta_3 = 1$ or $\delta_4 = 1$ enforces the sum-to-zero constraints on the columns of $\mathbf{C}_u$ or $\mathbf{D}_u$, or on $\boldsymbol{b}_t$ or $\boldsymbol{a}_s$ respectively. For example, estimating the parameters in $\mathbf{J}_J \boldsymbol{a}_s$ is equivalent to estimating $\boldsymbol{a}_s$ subject to $\mathbf{1}_J' \boldsymbol{a}_s = 0$.

We can now associate each of the terms in (3.6) with the corresponding terms in the model (3.7), by substituting (3.6) in (3.7) so that

$$\begin{aligned} L(i|r,s,t,u) = \Big\| &(\delta_1\delta_3 + \delta_2\delta_4 - \delta_1\delta_2) \left(\tfrac{1}{JK} \mathbf{1}_J' \mathbf{X}_i \mathbf{1}_K - m_r \right) \mathbf{1}_J \mathbf{1}_K' \\ &+ \delta_1 \mathbf{1}_J \left(\tfrac{1}{J} \mathbf{1}_J' \mathbf{X}_i - \boldsymbol{b}_t' \right) \mathbf{J}_K^{(\delta_3)} \\ &+ \delta_2 \mathbf{J}_J^{(\delta_4)} \left(\tfrac{1}{K} \mathbf{X}_i \mathbf{1}_K - \boldsymbol{a}_s \right) \mathbf{1}_K' \\ &+ \mathbf{J}_J^{(\delta_1)} \left(\mathbf{X}_i - \mathbf{C}_u \mathbf{D}_u' \right) \mathbf{J}_K^{(\delta_2)} \|^2 . \end{aligned} \tag{3.8}$$

It can be shown (see Appendix A for an overview) that for all choices of $\boldsymbol{\delta}$, with the exception of the case $\boldsymbol{\delta}' = (1,1,0,0)$, the decomposition is orthogonal such that

$$\begin{aligned} L(i|r,s,t,u) &= JK \left\| (\delta_1\delta_3 + \delta_2\delta_4 - \delta_1\delta_2) \left(\tfrac{1}{JK} \mathbf{1}_J' \mathbf{X}_i \mathbf{1}_K - m_r \right) \right\|^2 + J \left\| \delta_1 \left(\tfrac{1}{J} \mathbf{1}_J' \mathbf{X}_i - \boldsymbol{b}_t' \right) \mathbf{J}_K^{(\delta_3)} \right\|^2 \\ &\quad + K \left\| \delta_2 \mathbf{J}_J^{(\delta_4)} \left(\tfrac{1}{K} \mathbf{X}_i \mathbf{1}_K - \boldsymbol{a}_s \right) \right\|^2 + \left\| \mathbf{J}_J^{(\delta_1)} (\mathbf{X}_i - \mathbf{C}_u \mathbf{D}_u') \mathbf{J}_K^{(\delta_2)} \right\|^2 \\ &= L_{(\mathrm{o})}(i|r) + L_{(\mathrm{r})}(i|s) + L_{(\mathrm{c})}(i|t) + L_{(\mathrm{i})}(i|u). \end{aligned} \tag{3.9}$$

This equality follows from the fact that all the cross-products are zero. Furthermore, the orthogonality leads to a profound simplification in the clustering, since now the loss function

(3.2) equals

$$
\begin{aligned}
& L(\mathbf{G}^{(\mathrm{o})}, \mathbf{G}^{(\mathrm{r})}, \mathbf{G}^{(\mathrm{c})}, \mathbf{G}^{(\mathrm{i})}, \boldsymbol{m}, \mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}) \\
&= \sum_{i,r,s,t,u} \left\{ g_{ir}^{(\mathrm{o})} L_{(\mathrm{o})}(i|r) + g_{is}^{(\mathrm{r})} L_{(\mathrm{r})}(i|s) + g_{it}^{(\mathrm{c})} L_{(\mathrm{c})}(i|t) + g_{iu}^{(\mathrm{i})} L_{(\mathrm{i})}(i|u) \right\} \\
&= JK \sum_{i=1}^{N} \sum_{r=1}^{R} g_{ir}^{(\mathrm{o})} \left\| (\delta_1\delta_3 + \delta_2\delta_4 - \delta_1\delta_2) \left(\tfrac{1}{JK} \mathbf{1}_J' \mathbf{X}_i \mathbf{1}_K - m_r \right) \right\|^2 \\
&\quad + K \sum_{i=1}^{N} \sum_{s=1}^{S} g_{is}^{(\mathrm{r})} \left\| \delta_2 \mathbf{J}_J^{(\delta_4)} \left(\tfrac{1}{K} \mathbf{X}_i \mathbf{1}_K - \boldsymbol{a}_s \right) \right\|^2 \\
&\quad + J \sum_{i=1}^{N} \sum_{t=1}^{T} g_{it}^{(\mathrm{c})} \left\| \delta_1 \left(\tfrac{1}{J} \mathbf{1}_J' \mathbf{X}_i - \boldsymbol{b}_t' \right) \mathbf{J}_K^{(\delta_3)} \right\|^2 \\
&\quad + \sum_{i=1}^{N} \sum_{u=1}^{U} g_{iu}^{(\mathrm{i})} \left\| \mathbf{J}_J^{(\delta_1)} \left(\mathbf{X}_i - \mathbf{C}_u \mathbf{D}_u' \right) \mathbf{J}_K^{(\delta_2)} \right\|^2 \\
&= L_{(\mathrm{o})}(\mathbf{G}^{(\mathrm{o})}, \boldsymbol{m}) + L_{(\mathrm{r})}(\mathbf{G}^{(\mathrm{r})}, \mathbf{A}) + L_{(\mathrm{c})}(\mathbf{G}^{(\mathrm{c})}, \mathbf{B}) + L_{(\mathrm{i})}(\mathbf{G}^{(\mathrm{i})}, \mathbf{C}, \mathbf{D}).
\end{aligned} \tag{3.10}
$$

Consequently, the joint clustering reduces to separate clusterings on the overall mean, row margins, column margins and interactions respectively. To see why simplification in (3.10) is possible after utilizing (3.9), consider the summation

$$
\sum_{r=1}^{R} \sum_{s=1}^{S} \sum_{t=1}^{T} \sum_{u=1}^{U} g_{ir}^{(\mathrm{o})} g_{is}^{(\mathrm{r})} g_{it}^{(\mathrm{c})} g_{iu}^{(\mathrm{i})} \left\| \mathbf{J}_J^{(\delta_1)} \left(\mathbf{X}_i - \mathbf{C}_u \mathbf{D}_u' \right) \mathbf{J}_K^{(\delta_2)} \right\|^2 . \tag{3.11}
$$

As the term $\left\| \mathbf{J}_J^{(\delta_1)} \left(\mathbf{X}_i - \mathbf{C}_u \mathbf{D}_u' \right) \mathbf{J}_K^{(\delta_2)} \right\|^2$ only depends on the subscript u, the summations of the other cluster types are equal to one because for each i we have $\sum_{r=1}^{R} g_{ir}^{(\mathrm{r})} = 1$, $\sum_{s=1}^{S} g_{is}^{(\mathrm{c})} = 1$, and $\sum_{u=1}^{U} g_{iu}^{(\mathrm{i})} = 1$. The same holds for the other terms.

Different choices for $\boldsymbol{\delta}$ lead to different model specifications, the choice of which should be guided by the application. The nine different cases are summarized in Table 3.1. Note that Model 6 does not lead to an orthogonal decomposition and is therefore not discussed further. The most general model is Model 1, which amounts to a clusterwise low-rank decomposition of the mean. All the other models are essentially special cases of this model, stemming from specific forms for $\mathbf{C}_u$ or $\mathbf{D}_u$ or both. For example, Model 4 can be written as $\mathbf{C}_u^* \mathbf{D}_u^{*\prime}$ where $\mathbf{C}_u^* = \begin{bmatrix} \mathbf{1}_J & \mathbf{J}_J \mathbf{C}_u \end{bmatrix}$ and $\mathbf{D}_u^* = \begin{bmatrix} \boldsymbol{b}_t & \mathbf{D}_u \end{bmatrix}$. We note that the added clustering adds additional restrictions on the parameters which is not the case in ordinary bilinear models. Besides the choice of $\boldsymbol{\delta}$, the choice of either (a) $\mathbf{C}_1 \mathbf{D}_u'$, (b) $\mathbf{C}_u \mathbf{D}_1'$ or (c) $\mathbf{C}_u \mathbf{D}_u'$ must be made.

Model	δ_1	δ_2	δ_3	δ_4	**Model for $\mathbf{X}_i$**
1	0	0	0	0	$\mathbf{C}_u\mathbf{D}_u^{'}$
2	0	1	0	0	$\boldsymbol{a}_s\mathbf{1}_K^{'} + \mathbf{C}_u\mathbf{D}_u^{'}\mathbf{J}_K$
3	0	1	0	1	$m_r\mathbf{1}_J\mathbf{1}_K^{'} + \mathbf{J}_J\boldsymbol{a}_s\mathbf{1}_K^{'} + \mathbf{C}_u\mathbf{D}_u^{'}\mathbf{J}_K$
4	1	0	0	0	$\mathbf{1}_J\boldsymbol{b}_t^{'} + \mathbf{J}_J\mathbf{C}_u\mathbf{D}_u^{'}$
5	1	0	1	0	$m_r\mathbf{1}_J\mathbf{1}_K^{'} + \mathbf{1}_J\boldsymbol{b}_t^{'}\mathbf{J}_K + \mathbf{J}_J\mathbf{C}_u\mathbf{D}_u^{'}$
6	1	1	0	0	$-m_r\mathbf{1}_J\mathbf{1}_K^{'} + \boldsymbol{a}_s\mathbf{1}_K^{'} + \mathbf{1}_J\boldsymbol{b}_t^{'} + \mathbf{J}_J\mathbf{C}_u\mathbf{D}_u^{'}\mathbf{J}_K$
7	1	1	0	1	$\mathbf{J}_J\boldsymbol{a}_s\mathbf{1}_K^{'} + \mathbf{1}_J\boldsymbol{b}_t^{'} + \mathbf{J}_J\mathbf{C}_u\mathbf{D}_u^{'}\mathbf{J}_K$
8	1	1	1	0	$\boldsymbol{a}_s\mathbf{1}_K^{'} + \mathbf{1}_J\boldsymbol{b}_t^{'}\mathbf{J}_K + \mathbf{J}_J\mathbf{C}_u\mathbf{D}_u^{'}\mathbf{J}_K$
9	1	1	1	1	$m_r\mathbf{1}_J\mathbf{1}_K^{'} + \mathbf{J}_J\boldsymbol{a}_s\mathbf{1}_K^{'} + \mathbf{1}_J\boldsymbol{b}_t^{'}\mathbf{J}_K + \mathbf{J}_J\mathbf{C}_u\mathbf{D}_u^{'}\mathbf{J}_K$

Table 3.1: A summary of the models implied by different choices of $\boldsymbol{\delta}$. Note that Model 6 is not orthogonal and is only included for completeness.

3.3 Algorithm

Due to the form of the loss function (3.10), we can treat each of the components separately. Conveniently, the loss functions $L_{(o)}(\mathbf{G}^{(o)}, \boldsymbol{m})$, $L_{(r)}\left(\mathbf{G}^{(r)}, \mathbf{A}\right)$ and $L_{(c)}\left(\mathbf{G}^{(c)}, \mathbf{B}\right)$ are specific k-means problems (e.g. Everitt et al., 2011), on the rows of (say) the data matrix $\mathbf{Y} : N \times d$. In each of these cases, $\mathbf{Y}$ can be defined as follows:

- For estimating $\mathbf{G}^{(o)}$ and $\boldsymbol{m}$, $\mathbf{Y}$ has a single column ($d = 1$) containing the overall means $\frac{1}{JK}\mathbf{1}_J'\mathbf{X}_i\mathbf{1}_K$ of the $\{\mathbf{X}_i\}$;
- For estimating $\mathbf{G}^{(r)}$ and $\mathbf{A}$, the rows of $\mathbf{Y}$ ($d = J$) consist of the row mean vectors $\frac{1}{K}\mathbf{1}_K'\mathbf{X}_i'$; and
- For estimating $\mathbf{G}^{(c)}$ and $\mathbf{B}$, the rows of $\mathbf{Y}$ ($d = K$) are the column mean vectors $\frac{1}{J}\mathbf{1}_J'\mathbf{X}_i$.

Hence optimizing $L_{(o)}(\mathbf{G}^{(o)}, \boldsymbol{m})$, $L_{(r)}(\mathbf{G}^{(r)}, \mathbf{A})$ and $L_{(c)}(\mathbf{G}^{(c)}, \mathbf{B})$ can resort to standard methods for k-means on the overall mean, row margins and column margins respectively. Also, there are a variety of tools available for selecting R, S and T. We stress that caution is required with respect to local minima in k-means clustering, which can be alleviated by using multiple random starts.

The degrees-of-freedom for estimating the clusterwise overall means are $(\delta_1\delta_3 + \delta_2\delta_4 - \delta_1\delta_2)R$, while the degrees-of-freedom associated with estimating the clusterwise row and column means are $\delta_2 S(J - \delta_4)$ and $\delta_1 T(K - \delta_3)$ respectively. The degrees-of-freedom for the interactions depends on the choice of (a), (b) or (c), and are deferred to the next section.

3.3.1 An Algorithm for the Interaction Clustering

We proceed to formulate a special algorithm for minimizing $L_{(i)}(\mathbf{G}^{(i)}, \mathbf{C}, \mathbf{D})$ based on block-relaxation methods (see for example De Leeuw, 1994). The proposed algorithm iterates over optimizing one set of parameters while keeping the others fixed. Specifically, in step (1) we consider $\mathbf{G}^{(i)}$ fixed and simultaneously update $\mathbf{C}$ and $\mathbf{D}$, while step (2) consists of updating $\mathbf{G}^{(i)}$ while keeping $\mathbf{C}$ and $\mathbf{D}$ fixed at the values obtained in step (1). Finally, steps (1) and (2) are repeated until numerical convergence of the loss function is observed. This algorithm is guaranteed to converge monotonically, but only to local minima. It must be initialized by a starting configuration for $\mathbf{G}^{(i)}$. To increase the likelihood of locating the global minimum, it is advisable to use multiple (random) starting values for $\mathbf{G}^{(i)}$.

We now describe the steps of our algorithm in more detail. To minimize $L_{(\mathrm{i})}(\mathbf{G}^{(\mathrm{i})},\mathbf{C},\mathbf{D})$ over $\mathbf{C}$ and $\mathbf{D}$ for fixed $\mathbf{G}^{(\mathrm{i})}$, it is useful to rewrite it as follows:

$$\begin{aligned} &L_{(\mathrm{i})}(\mathbf{G}^{(\mathrm{i})},\mathbf{C},\mathbf{D}) \\ &\quad= \sum_{i=1}^{N}\sum_{u=1}^{U} g_{iu}^{(\mathrm{i})} \left\|\mathbf{J}_J^{(\delta_1)}\left(\mathbf{X}_i - \overline{\mathbf{X}}_u + \overline{\mathbf{X}}_u - \mathbf{C}_u\mathbf{D}_u'\right)\mathbf{J}_K^{(\delta_2)}\right\|^2 \\ &\quad= \sum_{i=1}^{N}\sum_{u=1}^{U} g_{iu}^{(\mathrm{i})} \left\|\mathbf{J}_J^{(\delta_1)}\left(\mathbf{X}_i - \overline{\mathbf{X}}_u\right)\mathbf{J}_K^{(\delta_2)}\right\|^2 + \sum_{u=1}^{U} N_u \left\|\mathbf{J}_J^{(\delta_1)}\left(\overline{\mathbf{X}}_u - \mathbf{C}_u\mathbf{D}_u'\right)\mathbf{J}_K^{(\delta_2)}\right\|^2. \end{aligned} \tag{3.12}$$

Here $N_u = \sum_{i=1}^{N} g_{iu}^{(\mathrm{i})}$ is the cardinality of cluster u, and $\overline{\mathbf{X}}_u = \frac{1}{N_u}\sum_{i=1}^{N} g_{iu}^{(\mathrm{i})}\mathbf{X}_i$ is the cluster mean. Equality holds as the cross-product equals zero. Note also that (3.12) shows $L_{(\mathrm{i})}(\mathbf{G}^{(\mathrm{i})},\mathbf{C},\mathbf{D})$ to be decomposable into two parts: the first term gives the deviations of the observations from their respective cluster means, while the last term is the modelling part which models the cluster means by a low-rank decomposition $\mathbf{C}_u\mathbf{D}_u'$.

It is evident from (3.12) that estimating the $\mathbf{C}_u$'s and $\mathbf{D}_u$'s only requires minimization of the modelling term. In order to do this, we first make the following definitions:

$$\begin{aligned} \mathbf{D}_* &= \left[\sqrt{N_1}\,\mathbf{D}_1'\mathbf{J}_K^{(\delta_2)} \quad \sqrt{N_2}\,\mathbf{D}_2'\mathbf{J}_K^{(\delta_2)} \quad \cdots \quad \sqrt{N_U}\,\mathbf{D}_U'\mathbf{J}_K^{(\delta_2)}\right]'; \\ \mathbf{C}_* &= \left[\sqrt{N_1}\,\mathbf{C}_1'\mathbf{J}_J^{(\delta_1)} \quad \sqrt{N_2}\,\mathbf{C}_2'\mathbf{J}_J^{(\delta_1)} \quad \cdots \quad \sqrt{N_U}\,\mathbf{C}_U'\mathbf{J}_J^{(\delta_1)}\right]'; \\ \overline{\mathbf{X}}_{(C)} &= \left[\sqrt{N_1}\,\overline{\mathbf{X}}_1\mathbf{J}_K^{(\delta_2)} \quad \sqrt{N_2}\,\overline{\mathbf{X}}_2\mathbf{J}_K^{(\delta_2)} \quad \cdots \quad \sqrt{N_U}\,\overline{\mathbf{X}}_U\mathbf{J}_K^{(\delta_2)}\right]; \\ \overline{\mathbf{X}}_{(R)} &= \left[\sqrt{N_1}\,\overline{\mathbf{X}}_1'\mathbf{J}_J^{(\delta_1)} \quad \sqrt{N_2}\,\overline{\mathbf{X}}_2'\mathbf{J}_J^{(\delta_1)} \quad \cdots \quad \sqrt{N_U}\,\overline{\mathbf{X}}_U'\mathbf{J}_J^{(\delta_1)}\right]'. \end{aligned} \tag{3.13}$$

Distinction must be made between the three cases where (a) $\mathbf{C}_1\mathbf{D}_u'$, (b) $\mathbf{C}_u'\mathbf{D}_1$ or (c) $\mathbf{C}_u'\mathbf{D}_u'$ applies. For a specified rank P, we now consider finding updates for $\mathbf{C}$ and $\mathbf{D}$ in each of these cases.

(a) Suppose that $\mathbf{C}_1\mathbf{D}_u'$ applies. The last term in (3.12) can be rewritten as

$$\left\|\mathbf{J}_J^{(\delta_1)}\overline{\mathbf{X}}_{(C)} - \mathbf{J}_J^{(\delta_1)}\mathbf{C}_1\mathbf{D}_*'\right\|^2. \tag{3.14}$$

Hence by the Eckart-Young theorem (Eckart and Young, 1936), the best rank-P least-squares approximation of $\mathbf{J}_J^{(\delta_1)}\overline{\mathbf{X}}_{(C)}$ is given by the truncated singular value decomposition (SVD) of $\mathbf{J}_J^{(\delta_1)}\overline{\mathbf{X}}_{(C)}$. Consequently we can update $\mathbf{C}$ and $\mathbf{D}$ as follows:

$$\begin{aligned} \mathbf{J}_J^{(\delta_1)}\mathbf{C}_1 &= \mathbf{U}\Gamma^{\alpha}\mathbf{L} \\ \mathbf{D}_* &= \mathbf{V}\Gamma^{1-\alpha}\mathbf{L} \end{aligned} \tag{3.15}$$

where $\mathbf{J}_J^{(\delta_1)}\overline{\mathbf{X}}_{(C)} = \mathbf{U}\Gamma\mathbf{V}'$ is the appropriate SVD and the square matrix $\mathbf{L}$ of dimension $\min\{J, UK\}$ is given by

$$\mathbf{L} = \begin{bmatrix} \mathbf{I}_P & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}. \tag{3.16}$$

Hence multiplication by $\mathbf{L}$ sets all singular values except the first P equal to zero. The parameter $0 \leq \alpha \leq 1$ is typically taken to be 0.5, but can be set by the user to improve the interpretability of the graphical output (see Section 3.4). In this notation Γ^α denotes the diagonal matrix where the diagonal contains the singular values to the power α. The degrees-of-freedom for the interactions are $P(J + UK - P - \delta_1 - \delta_2 U)$ in this case. See Appendix C for a short overview of how to derive this and the other degrees-of-freedom stated in the current section.

(b) In case $\mathbf{C}_u\mathbf{D}_1'$ applies, we can rewrite the last term in (3.12) as

$$\left\|\overline{\mathbf{X}}_{(R)}\mathbf{J}_K^{(\delta_2)} - \mathbf{C}_*\mathbf{D}_1'\mathbf{J}_K^{(\delta_2)}\right\|^2. \tag{3.17}$$

Analogously to (a), the update is based on the SVD $\overline{\mathbf{X}}_{(R)}\mathbf{J}_K^{(\delta_2)} = \mathbf{U}\Gamma\mathbf{V}'$ and is

$$\begin{aligned} \mathbf{C}_* &= \mathbf{U}\Gamma^\alpha\mathbf{L} \\ \mathbf{J}_K^{(\delta_2)}\mathbf{D}_1 &= \mathbf{V}\Gamma^{1-\alpha}\mathbf{L}. \end{aligned} \tag{3.18}$$

Here $\mathbf{L}$ is square with dimensions $\min\{UJ, K\}$. In this case, the degrees-of-freedom associated with the interactions are $P(UJ + K - P - \delta_1 U - \delta_2)$.

(c) Finally, when both $\mathbf{C}_u$ and $\mathbf{D}_u$ are cluster-specific, the update is based on the cluster-wise SVD's $\mathbf{J}_J^{(\delta_1)}\overline{\mathbf{X}}_u\mathbf{J}_K^{(\delta_2)} = \mathbf{U}_u\Gamma_u\mathbf{V}_u'$. We then have

$$\begin{aligned} \mathbf{J}_J^{(\delta_1)}\mathbf{C}_u &= \mathbf{U}_u\Gamma_u^\alpha\mathbf{L} \\ \mathbf{J}_K^{(\delta_2)}\mathbf{D}_u &= \mathbf{V}_u\Gamma_u^{1-\alpha}\mathbf{L}, \end{aligned} \tag{3.19}$$

with $\mathbf{L}$ having dimensions $\min\{J, K\}$. The appropriate degrees-of-freedom are $UP(J + K - P - \delta_1 - \delta_2)$.

Hence the first step of our algorithm conveniently relies only on SVD's. Now in step (2) $\mathbf{G}^{(\mathrm{i})}$ is updated while regarding $\mathbf{C}$ and $\mathbf{D}$ as fixed. The updated $\mathbf{G}^{(\mathrm{i})}$ is constructed by simply assigning each i to the cluster with the closest mean, hence minimizing $L_{(\mathrm{i})}(\mathbf{G}^{(\mathrm{i})}, \mathbf{C}, \mathbf{D})$ for each individual in a greedy manner. This entails setting $\boldsymbol{g}_{ik_i} = 1$ and zero elsewhere, where $\boldsymbol{g}_i$ is the ith row of

$\mathbf{G}^{(\mathrm{i})}$ and k_i is determined as

$$k_i = \arg\min_{u=1}^{U} \left\| \mathbf{J}_J^{(\delta_1)} \left(\mathbf{X}_i - \mathbf{C}_u \mathbf{D}_u' \right) \mathbf{J}_K^{(\delta_2)} \right\|^2 . \tag{3.20}$$

We prefer reporting the minimized value of $L_{(\mathrm{i})}(\mathbf{G}^{(\mathrm{i})}, \mathbf{C}, \mathbf{D})$ after division by $\sum_{i=1}^{N} \left\| \mathbf{J}_J^{(\delta_1)} \mathbf{X}_i \mathbf{J}_K^{(\delta_2)} \right\|^2$, since this standardized value lies within $[0, 1]$.

3.3.2 Fit Diagnostics

If $P \in \{1, 2, 3\}$ dimensions are used, biplots can be used to visualize the relationships between the J objects and K attributes for each of the clusters. Biplots generalize scatterplots of two variables to multiple variables (Gower and Hand, 1996; Gower et al., 2011), and rely on low-rank inner product approximations.

Constructing biplots for the interactions here simply entails plotting the approximation of $\mathbf{J}_J^{(\delta_1)} \overline{\mathbf{X}}_u \mathbf{J}_K^{(\delta_2)}$ for each cluster. For example, for case (c) in (3.19) the object coordinates are given by $\mathbf{U}_u \Gamma_u^{\alpha} \mathbf{L}$, while $\mathbf{V}_u \Gamma_u^{1-\alpha} \mathbf{L}$ provides the coordinates for the attributes in P-dimensional space. The inner products between the pairs of rows in these matrices are rank-P approximations of the corresponding entries in $\mathbf{J}_J^{(\delta_1)} \overline{\mathbf{X}}_u \mathbf{J}_K^{(\delta_2)}$. Similar results are easily obtained for cases (a) and (b). We defer discussion of the interpretation of these biplots to Section 3.4, where empirical examples are examined.

It is possible to construct goodness-of-fit measures for the biplots used to visualize the interactions. These are based on the proportion of variation accounted for by the model. A fit value of one indicates perfect fit: the model captures all the variation in the data. In contrast, low fit values imply that a substantial amount of variation occurs in the subspace orthogonal to that identified by the model. An increase in fit can usually be achieved by increasing P, but even though overall fit is guaranteed to improve with increasing P, the fit of all individual rows and columns will not necessarily improve concurrently. In practice the choice $P = 2$ is the most convenient because the biplots can readily be displayed.

Measures can be defined for the overall fit, the J objects as well as for the K attributes. Again, we must distinguish between the three cases (a), (b) and (c). Case (c), where the applicable model is $\mathbf{J}_J^{(\delta_1)} \mathbf{C}_u \mathbf{D}_u' \mathbf{J}_K^{(\delta_2)}$, is the simplest and therefore discussed first. The overall fit relies on the result

$$\left\| \mathbf{J}_J^{(\delta_1)} \overline{\mathbf{X}}_u \mathbf{J}_K^{(\delta_2)} \right\|^2 = \left\| \mathbf{J}_J^{(\delta_1)} \mathbf{C}_u \mathbf{D}_u' \mathbf{J}_K^{(\delta_2)} \right\|^2 + \left\| \mathbf{J}_J^{(\delta_1)} \left(\overline{\mathbf{X}}_u - \mathbf{C}_u \mathbf{D}_u' \right) \mathbf{J}_K^{(\delta_2)} \right\|^2 , \tag{3.21}$$

which shows that the total sum-of-squares in cluster u can be decomposed into that explained by the model and the residual sum-of-squares. We therefore define the overall quality of fit within cluster u for P dimensions as

$$o_{\text{fit}} = \frac{\|\mathbf{J}_J^{(\delta_1)}\mathbf{C}_u\mathbf{D}_u'\mathbf{J}_K^{(\delta_2)}\|^2}{\|\mathbf{J}_J^{(\delta_1)}\overline{\mathbf{X}}_u\mathbf{J}_K^{(\delta_2)}\|^2} = \frac{\operatorname{tr}\Gamma_u^2\mathbf{L}}{\operatorname{tr}\Gamma_u^2}. \tag{3.22}$$

This is just the proportion of the variation in the cluster mean explained by the model. Here $\operatorname{tr}\mathbf{A}$ denotes the trace of the square matrix $\mathbf{A}$, which is just the sum of its diagonal elements. Diagnostics for the J objects rely on the more general decomposition

$$\begin{aligned}
&\left(\mathbf{J}_J^{(\delta_1)}\overline{\mathbf{X}}_u\mathbf{J}_K^{(\delta_2)}\right)\left(\mathbf{J}_J^{(\delta_1)}\overline{\mathbf{X}}_u\mathbf{J}_K^{(\delta_2)}\right)' \\
&\quad = \left(\mathbf{J}_J^{(\delta_1)}\mathbf{C}_u\mathbf{D}_u'\mathbf{J}_K^{(\delta_2)}\right)\left(\mathbf{J}_J^{(\delta_1)}\mathbf{C}_u\mathbf{D}_u'\mathbf{J}_K^{(\delta_2)}\right)' \\
&\qquad + \left(\mathbf{J}_J^{(\delta_1)}\left(\overline{\mathbf{X}}_u - \mathbf{C}_u\mathbf{D}_u'\right)\mathbf{J}_K^{(\delta_2)}\right)\left(\mathbf{J}_J^{(\delta_1)}\left(\overline{\mathbf{X}}_u - \mathbf{C}_u\mathbf{D}_u'\right)\mathbf{J}_K^{(\delta_2)}\right)'.
\end{aligned} \tag{3.23}$$

Hence the total sum-of-squares for each of the objects can be decomposed orthogonally into the part explained by the model in the first term on the right-hand side of (3.23), and the residual sum-of-squares in the second term. The proportion of the variation explained by each of the rows, also known as sample predictivities (Gower et al., 2011), is therefore given by

$$\begin{aligned}
\boldsymbol{r}_{\text{fit}} &= \left[\operatorname{diag}\mathbf{J}_J^{(\delta_1)}\mathbf{C}_u\mathbf{D}_u'\mathbf{J}_K^{(\delta_2)}\mathbf{D}_u\mathbf{C}_u'\mathbf{J}_J^{(\delta_1)}\right]\left[\operatorname{diag}\mathbf{J}_J^{(\delta_1)}\overline{\mathbf{X}}_u\mathbf{J}_K^{(\delta_2)}\overline{\mathbf{X}}_u'\mathbf{J}_J^{(\delta_1)}\right]^{-1}\mathbf{1}_J \\
&= \left[\operatorname{diag}\mathbf{U}_u\Gamma_u^{2\alpha}\mathbf{L}\mathbf{U}_u'\right]\left[\operatorname{diag}\mathbf{U}_u\Gamma_u^{2\alpha}\mathbf{U}_u'\right]^{-1}\mathbf{1}_J,
\end{aligned} \tag{3.24}$$

with each element bounded on $[0,1]$. In this context, $\operatorname{diag}\mathbf{A}$ denotes the diagonal matrix constructed from the main diagonal of $\mathbf{A}$.

The column fit for case (c) can be defined analogously for each of the K attributes as

$$\begin{aligned}
\boldsymbol{c}_{\text{fit}} &= \left[\operatorname{diag}\mathbf{J}_K^{(\delta_2)}\mathbf{D}_u\mathbf{C}_u'\mathbf{J}_J^{(\delta_1)}\mathbf{C}_u\mathbf{D}_u'\mathbf{J}_K^{(\delta_2)}\right]\left[\operatorname{diag}\mathbf{J}_K^{(\delta_2)}\overline{\mathbf{X}}_u'\mathbf{J}_J^{(\delta_1)}\overline{\mathbf{X}}_u\mathbf{J}_K^{(\delta_2)}\right]^{-1}\mathbf{1}_K \\
&= \left[\operatorname{diag}\mathbf{V}_u\Gamma_u^{2\alpha}\mathbf{L}\mathbf{V}_u'\right]\left[\operatorname{diag}\mathbf{V}_u\Gamma_u^{2\alpha}\mathbf{V}_u'\right]^{-1}\mathbf{1}_K.
\end{aligned} \tag{3.25}$$

These quantities are also known as axis predictivities (Gower et al., 2011). Diagnostics for cases (a) and (b) are deferred to Appendix C.

The loss contribution for person i towards the interactions is defined as

$$L_{(\text{i})}(i) = \sum_{u=1}^{U} g_{iu}^{(\text{i})} L_{(\text{i})}(i|u) = \sum_{u=1}^{U} g_{iu}^{(\text{i})} \left\|\mathbf{J}_J^{(\delta_1)}\left(\mathbf{X}_i - \mathbf{C}_u\mathbf{D}_u'\right)\mathbf{J}_K^{(\delta_2)}\right\|^2. \tag{3.26}$$

This gives an indication of badness-of-fit, and the sum over all persons gives the minimized value of $L_{(\mathrm{i})}\left(\mathbf{G}^{(\mathrm{i})}, \mathbf{C}, \mathbf{D}\right)$. These loss contributions account for possible differences in origin, scale and/or rotation between a person's interactions and the modelled cluster mean $\mathbf{J}_J^{(\delta_1)}\mathbf{C}_u\mathbf{D}_u'\mathbf{J}_K^{(\delta_2)}$. A more informative manner of presenting these loss contributions may be as percentage contributions to $L_{(\mathrm{i})}\left(\mathbf{G}^{(\mathrm{i})}, \mathbf{C}, \mathbf{D}\right)$.

An alternative measure of person fit which is bounded on $[-1, 1]$ is given by

$$p_{\mathrm{fit}}(i) = \sum_{u=1}^{U} g_{iu}^{(\mathrm{i})} \frac{\operatorname{tr} \mathbf{J}_J^{(\delta_1)}\mathbf{X}_i\mathbf{J}_K^{(\delta_2)}\mathbf{D}_u\mathbf{C}_u'}{\left\|\mathbf{J}_J^{(\delta_1)}\mathbf{X}_i\mathbf{J}_K^{(\delta_2)}\right\| \left\|\mathbf{J}_J^{(\delta_1)}\mathbf{C}_u\mathbf{D}_u'\mathbf{J}_K^{(\delta_2)}\right\|}. \tag{3.27}$$

This only takes into account differences in rotation and origin, and high values indicate good fit whilst negative values indicate poor fit. When the origins coincide, the quantity (3.27) can be interpreted as a product-moment correlation coefficient between $\operatorname{Vec} \mathbf{J}_J^{(\delta_1)}\mathbf{X}_i\mathbf{J}_K^{(\delta_2)}$ and $\operatorname{Vec} \mathbf{J}_J^{(\delta_1)}\mathbf{C}_u\mathbf{D}_u'\mathbf{J}_K^{(\delta_2)}$. The notation $\operatorname{Vec} \mathbf{A}$ denotes the vector formed by concatenating the columns of a matrix $\mathbf{A}$ into a single vector.

3.3.3 Biplot Interpretability

When neither the row nor column configurations are fixed across biplots, as is the case in (c), it can aid interpretation to rotate the configurations so that the axes lie more or less in the same direction. For any orthogonal matrix $\mathbf{Q}_u$, it holds for the inner product matrices that $\mathbf{C}_u\mathbf{D}_u' = (\mathbf{C}_u\mathbf{Q}_u)(\mathbf{D}_u\mathbf{Q}_u)'$, and hence these are invariant to orthogonal rotations. The problem of finding orthogonal matrices $\mathbf{Q}_u, u = 1, 2, \ldots, U$, such that either the row or column configurations match each other as closely as possible is known as the generalized orthogonal Procrustes problem (Gower, 1975; Gower and Dijksterhuis, 2004).

Supposing without loss of generality that we use the $\mathbf{J}_J^{(\delta_1)}\mathbf{C}_u$ as axes in the biplots, a typical loss function for this problem is

$$\begin{aligned} L\left(\mathbf{Q}_1, \cdots, \mathbf{Q}_U\right) &= \sum_{u<v}^{U} \|\mathbf{J}_J^{(\delta_1)}\mathbf{C}_u\mathbf{Q}_u - \mathbf{J}_J^{(\delta_1)}\mathbf{C}_v\mathbf{Q}_v\|^2 \\ &= U \sum_{u=1}^{U} \|\mathbf{J}_J^{(\delta_1)}\mathbf{C}_u\mathbf{Q}_u - \mathbf{H}\|^2 \end{aligned} \tag{3.28}$$

where $\mathbf{H} = U^{-1}\sum_{u=1}^{U}\mathbf{J}_J^{(\delta_1)}\mathbf{C}_u\mathbf{Q}_u$. A solution for (3.28) can be obtained through an alternating least-squares (ALS) algorithm; see Gower and Dijksterhuis (2004) for more details.

We note that two types of scalings can be used to make the biplot displays more attractive, namely α and λ scaling. First, since our choice of α does not change the inner product approximations (3.15), (3.18) and (3.19), we are free to choose it such that the resulting biplots are easy to interpret. In our software implementation we use as a heuristic method the value of α which maximizes the minimum Euclidean distance over all row and column points to the origin. Alternatively, the user can choose any other quantile of these distances, such as the median, or specify the desired value of α explicitly.

Second, note that for matrices $\mathbf{A}$ and $\mathbf{B}$ it holds that $\mathbf{AB}' = (\lambda\mathbf{A})(\mathbf{B}'/\lambda)$, so that λ can also be freely chosen. Following Gower et al. (2011), we choose λ such that the average squared Euclidean distances from the two sets of points represented by the rows of the matrices in (3.15), (3.18) and (3.19) to the origin are equal. For case (c) in (3.19) for example, this amounts to choosing

$$\lambda = \left(\frac{J\|\mathbf{V}\Gamma_u^{1-\alpha}\mathbf{L}\|^2}{K\|\mathbf{U}\Gamma_u^{\alpha}\mathbf{L}\|^2}\right)^{1/4} = \left(\frac{J\,\mathrm{tr}\,\Gamma_u^{1-\alpha}\mathbf{L}}{K\,\mathrm{tr}\,\Gamma_u^{\alpha}\mathbf{L}}\right)^{1/4}. \tag{3.29}$$

3.3.4 Model Selection

Some questions remain, including how to select the number of segments for each of the four clustering problems, and what to do when missing values are present in the data. Here and in the next section we give a short description of viable options for dealing with these issues, which are by no means intended to be exhaustive. The applications of Section 3.4 will further illustrate some of the points raised here.

Selection of the number of clusters can be handled separately for each of the four clustering problems. Many criteria have been proposed in the literature, especially in the context of k-means and hierarchical clustering. Naturally, these can be utilized directly for the three k-means subproblems, and most can also be used for the interaction clustering. The simplest approach, and the one we use here for illustration, is probably the scree test (Cattell, 1966). This method involves running the algorithm for several values of k and plotting the loss function against k. The user must then choose a value for k based on this so-called scree plot, such that the chosen k is close to an "elbow" in the plot. This indicates that adding additional groups to the analysis does not significantly increase how well the results describe the data.

A variation of the scree test is the convex hull (CHull) procedure of Ceulemans and Kiers (2006), which uses a measure of model complexity other than k, such as the degrees-of-freedom, for the scree plot. A convex hull is constructed from these points and a point close to the resulting elbow is selected, which represents a trade-off between model complexity and goodness-of-fit. This approach has recently been used successfully in several component analysis contexts (e.g. Schepers et al., 2008; Ceulemans and Kiers, 2009; Ceulemans et al., 2011; Lorenzo-Seva et al., 2011).

Several other approaches have also been proposed in the literature, such as the Caliński-Harabasz criterion (Caliński and Harabasz, 1974), the Krzanowski-Lai criterion (Krzanowski and Lai, 1988), the silhouette plot (Kaufman and Rousseeuw, 1990), the gap statistic (Tibshirani et al., 2001), the jump method (Sugar and James, 2003) and bootstrapping (Dolnicar and Leisch, 2010). Milligan and Cooper (1985), Hardy (1996) and Everitt et al. (2011) provide an assessment of some of these criteria and additional references. Experimenting with such alternative criteria are left for future research.

We note that it is also possible to select the number of clusters for all four subproblems at the same time. This may be easier to achieve with e.g. the CHull procedure than doing model selection separately. This approach can also be employed when the analyst wishes to treat the dimensionality of the interactions P as part of the model selection process.

3.4 Applications

We discuss two empirical examples: the first considers the evaluation of car manufacturers by a Dutch consumer panel, while the second considers the list of values data set (e.g. Van Rosmalen et al., 2010).

3.4.1 Car Manufacturers

This data set consists of 187 persons evaluating 10 car manufacturers on a set of 8 attributes, as collected via an online survey (Bijmolt and van de Velden, 2012). The data is a subset of a larger set collected from panellists within the CentERpanel of Tilburg University in the Netherlands. The sample used was selected to be representative of the population of Dutch households. The car brands and items were presented in random order to each respondent, with the order of the items fixed per respondent. More information regarding the data collection can be found in Bijmolt and van de Velden (2012).

The manufacturers considered are ten international brands, namely Citroën, Fiat, Ford, Opel, Peugeot, Renault, Seat, Toyota, Volkswagen and Volvo. The task respondents were taxed with was to rate each of these brands on 8 different attributes using a 10-point rating scale. For 6 out of the 8 items, namely Affordability, Attractiveness, Safety, Sportiness, Reliability and Features, a score of 10 relates to the most desirable outcome. However, for the items Size and Operating Cost, a score of 10 reflects small cars and those with high operating costs respectively. Consequently, higher ratings on these items indicate more negative assessments.

We fit an LSBCLUST model with $\delta = (1, 1, 1, 1)$ so that the overall means, row means, columns means and interactions are estimated separately. Also, we use $P = 2$ dimensions and fix the coordinates of the 10 car brands across all interaction biplots (case (a)). The number of clusters to use for each of the four components must be determined. As mentioned in Section 3.3.4, this can be done separately for each subproblem. Here we fit LSBCLUST models for 1 to 15 clusters and inspect the resulting scree plots to select R, S, T and U. Based on these plots, we selected $R = 5$, $S = 8$, $T = 6$ and $U = 8$ clusters. We note that these choices are subjective, should take into account the aims of the research and that alternative selection criteria can also be used. The number of random starts used for the interaction and k-means clustering were 100 and 1000 respectively.

The mean ratings for each of the five overall mean clusters are shown in Figure 3.1. Most interesting here are that 8 and 3 persons used very high and very low scores overall respectively (clusters O4 and O5). Inspecting the individuals belonging to Segment O5, we can identify individuals 50, 66 and 85, all of which respond with a rating of 1 to all items on all car brands, save for a single 2 assigned by person 66. These respondents obviously do not provide very interesting information in their answers, but since their corresponding row means, column means and interactions do not differ from the overall mean, we do not have to remove them from our analysis. These person are merely assigned to the row, column and interaction segments containing negligible effects (see below). In a similar vein, of the 8 persons in Segment O4, individuals 40, 47, 151 and 162 exclusively use rating category 10. The remaining persons in this cluster also almost exclusively use high ratings. LSBCLUST has therefore been able to identify the 11 persons in clusters O4 and O5 who provide very little sensible information.

Figure 3.2 displays the means of the eight car brand (row) clusters across all attributes. Effect sizes can be read off on the horizontal axis. Both Segments R7 and R8 consist of single individuals, namely persons 121 and 156 respectively. Person 121 used constant ratings for all car brands, except for Citroën (perhaps indicating that this person owns a Citroën). Specifically, 10's were assigned to Fiat and Toyota, 1's to Opel and Seat and 6's to the other brands. Segment R8 can be explained by only 10's being assigned to Fiat and Toyota and 1's to Peugeot and Volvo by person 156. Segment R6 contains 2.7% of respondents who scores Peugeot on average nearly 2.5 points below average, and rather prefers Fiat and Toyota, albeit with smaller effects (approximately 1.2 and 1.5 above average respectively). Segment R5 (8%) is not attracted to Opel with a negative mean effect of approximately 1.6 rating points, but does like Renault and Citroën somewhat. Persons in Segment R4 (9.6%) mainly dislike Volvo, scoring it 1.6 below average. The remaining segments can be interpreted similarly. Note that the largest cluster (R1, 48.7%) does not contain any large effects, indicating that these consumers do not have strong preferences for any of the brands across all attributes.

The attribute (column) mean effects for all six clusters are displayed in Figure 3.3. There are no singleton clusters here. Segment C6 (7%) consists of respondents who assign high scores

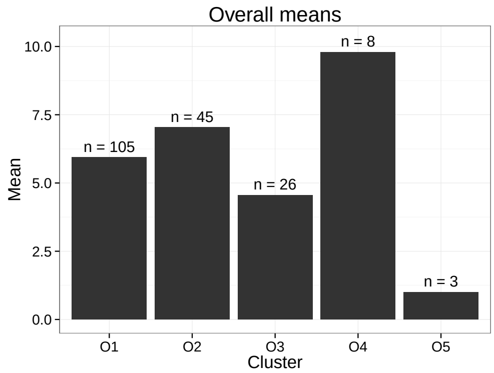


Figure 3.1: Clusterwise overall means detected for the cars data.

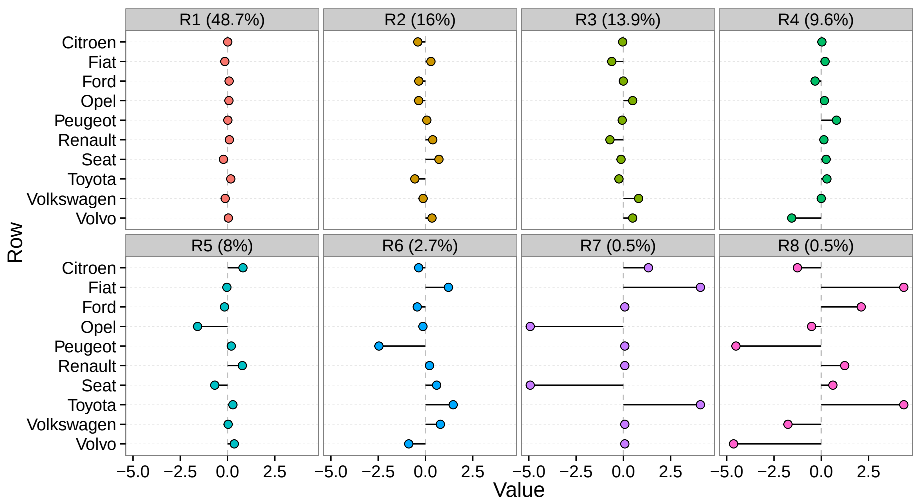


Figure 3.2: Car manufacturer (row) cluster means detected in the cars data. The size of the effects can be read off from the horizontal axis.

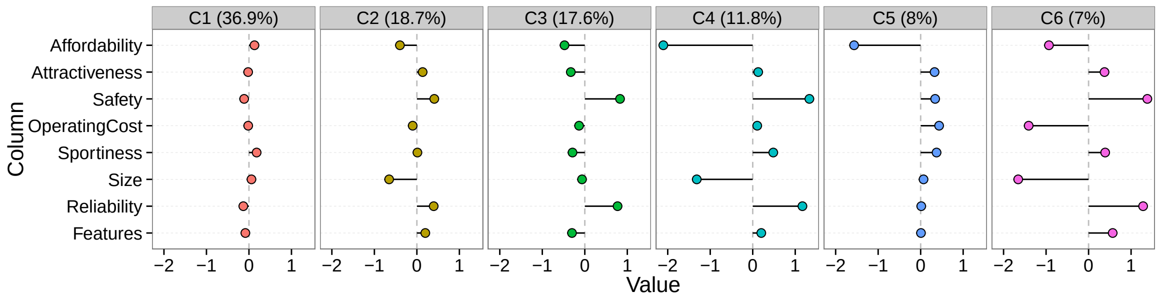


Figure 3.3: Attribute (column) cluster means detected in the cars data.

to Safety and Reliability irrespective of the car brand, with effects of approximately 1.4 and 1.3 respectively. Operating Cost and Size are also assessed positively on average, taking into account that for these items lower scores are better. The effects are approximately −1.4 and −1.7 on these two items. Segment C5 (8%) is somewhat curious in that the large negative effect on Affordability (−1.6) dominates. These respondents give ratings below their average on Affordability, irrespective of the car brand. Affordability also has a large negative effect in Segment C4 (11.8%), together with positive effects for Safety, Size and Reliability. The largest segment, C1 (36.9%), again has negligible effects.

The most interesting results can be found among the interactions, which is where respondents distinguish between different car manufacturers on the measured attributes. Figure 3.4 shows the biplots for the eight interaction segments. The car manufacturers are represented by points, and the attributes by arrows. The labels, points and arrows are shaded according to their goodness-of-fit, with well-fitting points being darker. The locations of the car brands are fixed across all biplots to make them easier to interpret. All car brands except Ford, with a fit of only 0.04, fit reasonably well – see Table 3.1. It is immediately apparent that the French manufacturers (Peugeot, Citroën and Renault) are judged to be similar, while the German brands Opel and Volkswagen are also located close together. The Swedish car manufacturer, Volvo, is somewhat isolated towards the right of the biplots. Fiat and Toyota are judged to be somewhat similar to the French and German brands respectively. Seat in turn are most similar to Toyota. By virtue of its low fit, Ford is hardly visible and lies near the origin.

The fit for the eight attributes vary per segment and is summarized in Table 3.2. Typically only a subset of items fit well in each segment, and only those with a fit larger than 0.5 are adorned with calibrated axes in Figure 3.4. For any manufacturer, the estimated cluster mean effects can be read off from the orthogonal projection of its representing point unto the biplot axes. For example, Volkswagen scores approximately 2 points above that predicted by the overall mean, row mean and column mean on Safety in Segment I8. Also, Volvo score about 3.3 rating points below the overall and marginal effects on Affordability in the same segment. The overall variance accounted for is 82.5% in two dimensions, with 63.2% and 19.2% attributed to dimension 1 and 2 respectively. Hence two dimensions are a reasonable choice.

Again, the largest Segment I1 (31.6%) represents very small effects, indicating that there is a significant proportion of respondents who did not discern between the car manufacturers based on the measured attributes. Predictably, this includes the 11 persons in Segments O4 and O5 discussed above. Segment I2 (16.6%) is characterized by large effects on Affordability, Size, Safety and Operating Cost. In particular, Affordability goes hand-in-hand with smaller cars (i.e. higher scores on Size), while high Operating Cost is strongly associated with increased Safety. Volvo scores well on Safety and Size but badly on Operating Cost and Affordability. On the other end of the scale Seat, Fiat and Toyota are seen as producing more affordable, smaller cars which are less safe to drive.

The 15% of respondents in Segment I3 consider Affordability and Size by far the most important items. These attributes are also highly correlated: more affordable cars are the smaller ones. Safety and Reliability also correlates, and Attractiveness as well as Features, although to a lesser extent, also elicit relatively large effects. Seat scores well on Affordability but low on size, as does Fiat, but these cars are seen as unsafe and unreliable. Fiat is considered much more attractive that Seat, but less than the French cars and Volvo. Volvo are now seen as both safe and reliable, but also as the most attractive brand.

Segment I4 (9.6%) displays large effects on Affordability, Reliability, Safety and Attractiveness. As in Segment I3, cars cannot be affordable as well as safe and reliable at the same time. Attractiveness is more or less orthogonal to these polar opposites. People in this segment consider Fiat to be the most affordable, even more so than Seat. Volvo is followed by Volkswagen in Safety and Reliability but Volkswagen cannot compete with Volvo in terms of Attractiveness. The French cars are seen as the most attractive though.

In segment I5 (9.1%), Affordability again elicits the largest effects, although smaller than in some of the previous segments. In this case, Fiat, Renault and Citroën are seen as the most affordable cars while Volkswagen and Volvo again score badly in this category. Volkswagen, Opel and Toyota are seen as the most attractive. Segment I6 (8.6%) harbours even smaller effects, the largest being with respect to Safety, Affordability and Operating Cost. In the opinions of these respondents, better Affordability means higher Operating Costs and lower Safety. It is particularly interesting to see that these people do not believe that Volvo's are safe, but do think they are affordable. This contrasts with most other segments. Seat and Fiat are considered the safest cars in this case, but they are considered to be expensive too.

Size and Affordability are again highly correlated in segment I7 (5.3%), and responsible for some large effects. These attributes are strongly opposed to Features, Sportiness, Safety and Reliability. Fiat and Seat score best on Size and Affordability, but score low on Features.

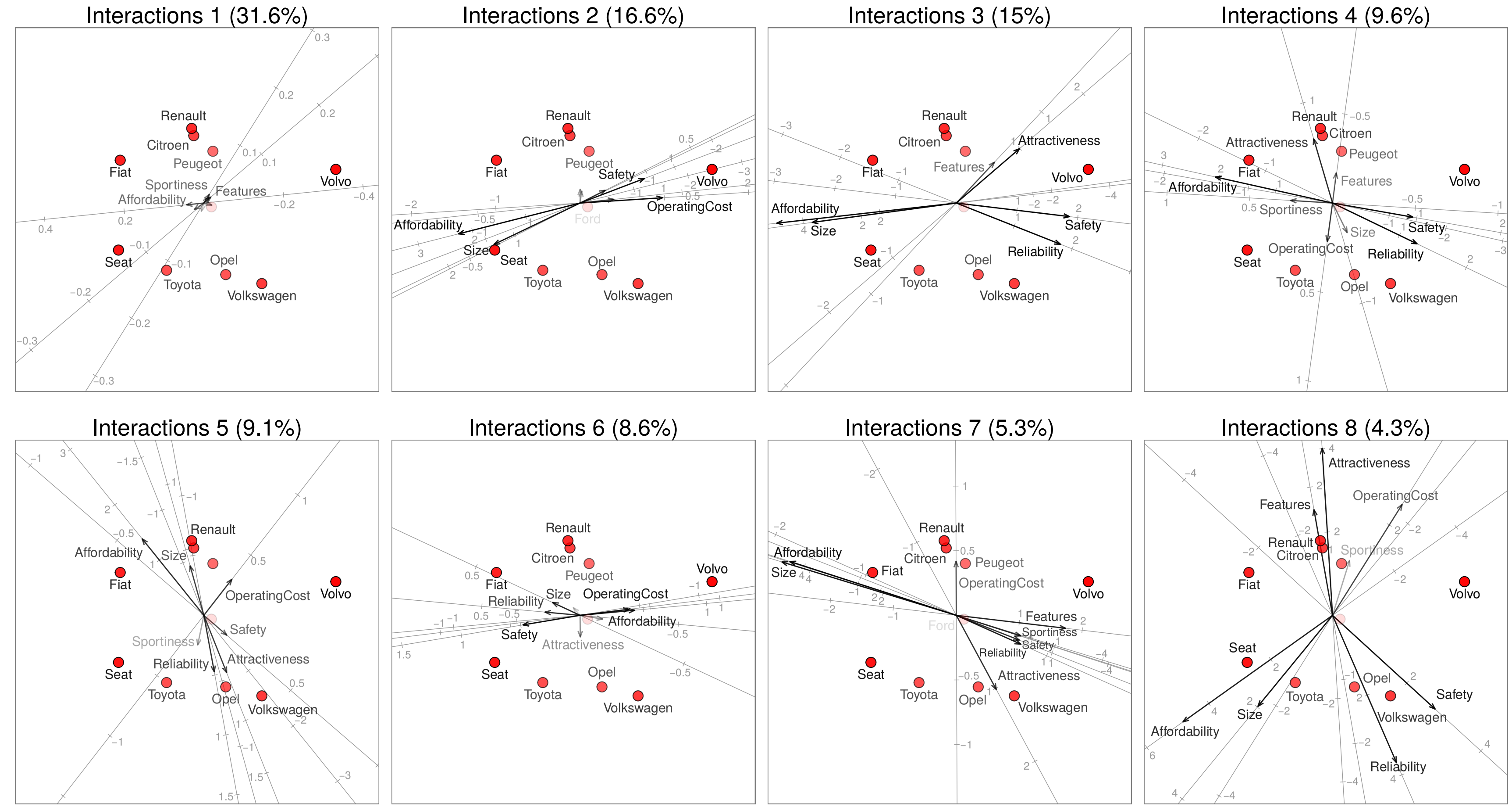


Figure 3.4: Biplots for the interaction clusters detected in the cars data. Each attribute is represented by a vector, and those which fit well also by calibrated axes. The car manufacturers are represented by points, and the orthogonal projections of these points onto the attribute axes give the estimated mean effects. The colours and labels are faded according to how well they fit into the display: solid colours fit well and transparent ones fit badly.

Volkswagen and Opel are seen as the most attractive cars. Finally, the smallest segment (Segment I8, 4.3%) considers Attractiveness and Features to be correlated, as well as Affordability and Size, and Reliability and Safety. The French cars score well on Attractiveness and Features, followed by Fiat and Volvo. Volkswagen are now seen as safer and more reliable than Volvo, with Opel also scoring well in these categories. Seat and Toyota score more than two rating points above the marginal effect on Affordability, but are considered to produce particularly small cars. Volvo and the French manufacturers produce much larger models. Volvo is the only manufacturer to combine attractive cars with high Safety and Reliability, but they are by far the most expensive.

Brand	**All segments**
Citroën	0.72
Fiat	0.87
Ford	0.04
Opel	0.63
Peugeot	0.53
Renault	0.82
Seat	0.92
Toyota	0.66
Volkswagen	0.75
Volvo	0.97

Table 3.1: Brand fit for the cars data across all clusters. Higher values indicate better fit, with a maximum of one and minimum of zero.

	Interaction segment							
Item	**I1**	**I2**	**I3**	**I4**	**I5**	**I6**	**I7**	**I8**
Affordability	0.52	0.96	0.94	0.93	0.72	0.83	0.95	0.80
Attractiveness	0.39	0.25	0.93	0.74	0.62	0.38	0.68	0.83
Safety	0.38	0.91	0.98	0.94	0.51	0.92	0.68	0.90
OperatingCost	0.46	0.94	0.08	0.64	0.66	0.90	0.53	0.57
Sportiness	0.51	0.22	0.03	0.56	0.28	0.35	0.73	0.26
Size	0.15	0.87	0.95	0.44	0.61	0.76	0.87	0.89
Reliability	0.07	0.62	0.97	0.98	0.69	0.71	0.83	0.84
Features	0.60	0.68	0.51	0.45	0.00	0.12	0.77	0.81

Table 3.2: Attribute fit for the cars data, for all eight interaction segments.

3.4.2 List of Values

The list of values (LOV) is a well-known values system, developed in Kahle (1983), which asks respondents to indicate the importance of nine different values as a guiding principle in their lives. These items are: (1) a sense of belonging, (2) excitement, (3) warm relationships with others, (4) self-fulfillment, (5) being well-respected, (6) fun and enjoyment in life, (7) security, (8) self-respect and (9) a sense of accomplishment (Van Rosmalen et al., 2010). A labelled nine-point rating scale with endpoints 'very important' (category 1) to 'not important at all' (category 9) are used for all items. LOV data have been studied in single or multi-country contexts in various studies, including Beatty et al. (1985); Grunert and Scherlorn (1990); Kamakura and Novak (1992); Brunsø et al. (2004); Chryssohoidis and Krystallis (2005); Lee et al. (2007); Sudbury and Simcock (2009).

The particular data set we analyse here originates from a commercial survey performed in 1996 (Van Herk, 2000), and was also analysed by Van Rosmalen et al. (2010). As in many data sets collected using rating scales, there are some concern regarding differences in response style between respondents. Response styles are related to respondents exercising their freedom-of-choice with respect to how to the rating scale is used, irrespective of the item content (e.g. Baumgartner and Steenkamp, 2001). For example, a person exhibiting an extreme response style may choose to use rating categories one and nine a majority of the time, while a person exhibiting midpoint scoring may favour mainly categories four, five and six. Crucially, response styles do not convey anything regarding the preference that a person may have for the items. Therefore response style effects should be accounted for separately in a model so that these effects are not confounded with the substantive information in the data.

Van Rosmalen et al. (2010) develop the so-called latent-class bilinear multinomial logit (LC-BML) model specifically to deal with situations where response styles are a concern. The LC-BML model is a parametric finite mixture of multinomial logit models which models the response to all items jointly. Also, this model simultaneously segments respondents into two types of clusters, namely response style and substantive item segments. Similarly to LSBCLUST, the LC-BML model produces biplots describing the relationship between the values and the rating categories within each item segment. The coordinates of the rating categories are fixed across all biplots. The response styles are modelled as marginal effects for the rating categories.

A nonparametric equivalent of the LC-BML model can be formulated within the LSBCLUST framework. The data array is constructed by transforming each observation into an indicator matrix, with the rows representing the respective rating categories and the columns the value items. Each column contains a single one indicating which rating was used to answer that item. In effect we therefore consider the rating scale as one of the modes in our three-way data set. Choosing $\boldsymbol{\delta} = (0, 1, 0, 0)$ fits a model containing only row or response style effects and interactions. Additionally, we use the option (a) $\mathbf{C}_1\mathbf{D}_u'$ as a model for the interactions so that the coordinates for the rating categories are fixed across biplots. The resulting model is, except for the inclusion of demographic variables in the LC-BML model, equivalent to the LC-BML model. It has the distinct advantage of being much faster to compute, as least-squares estimation and crisp clustering are used instead of maximum likelihood and finite mixture models.

Our particular sample consists of 4514 respondents from five European countries (Great Britain, France, Germany, Italy and Spain). The demographic variables Education (Low or High) and Age (0–25, 25–39, 40–54 or 55+ years) are also available. There are 344 missing values in the data, spread among 141 respondents. In order to emulate the results in Van Rosmalen et al. (2010), we follow their convention of adding a specific rating category labelled 'NA' for the missing value. Also, all respondents who did not answer any item or who answered all items with the same rating were removed from the sample, as well as those who did not disclose either or both of the demographics Education and Age. Van Rosmalen et al. (2010) report that this resulted in a deletion of 8% of the original respondents.

For comparability, we use the same number of clusters as Van Rosmalen et al. (2010): 11 row and 5 interaction segments, with two-dimensional biplots. These were selected for the LC-BML model by using the Bayesian Information Criterion (BIC; Schwarz, 1978). The number of starts used in LSBCLUST for the row and interaction effects were 1000 and 100 respectively. The clusters for the row effects (response styles) are summarized in Figure 3.5. For comparison the response styles detected by LC-BML are given in Figure 3.6 (see also Table 2 in Van Rosmalen et al. (2010)). These have been reordered such that they match more or less those detected by LSBCLUST.

There are definite similarities between the two sets of clusters. Quite clearly segments R1 (18.7%) and R1* (17.5%) correspond quite closely, indicating that slightly more than 1 in every 6 people tend to use rating category 1 ('very important') for 75% of their answers. This is an extreme form of an acquiescent response style, which favours positive responding (Baumgartner and Steenkamp, 2001). Segments R2 (16.6%) and R2* (16.7%) agree closely and indicate that a less severe acquiescent response style is used by approximately 1 in every 6 persons. In addition, clusters R3 (14.1%) and R4 (13.6%) are also versions of acquiescence, but with less focus on category 1 only. We note that the similarity between R3 and R3* (24.2%) are not perfect, but R4 and R4* does correspond well. Indeed R3* was the largest segment detected by the LC-BML model. Among the remaining segments, there are also strong agreement between the LSBCLUST and LC-BML results, except for R8 (5%) and R9 (4.5%). These segments can be interpreted as imperfect forms of midpoint responding. We note that LSBCLUST tend to identify more balanced response styles while the LC-BML model identifies styles which give large points masses to specific ratings.

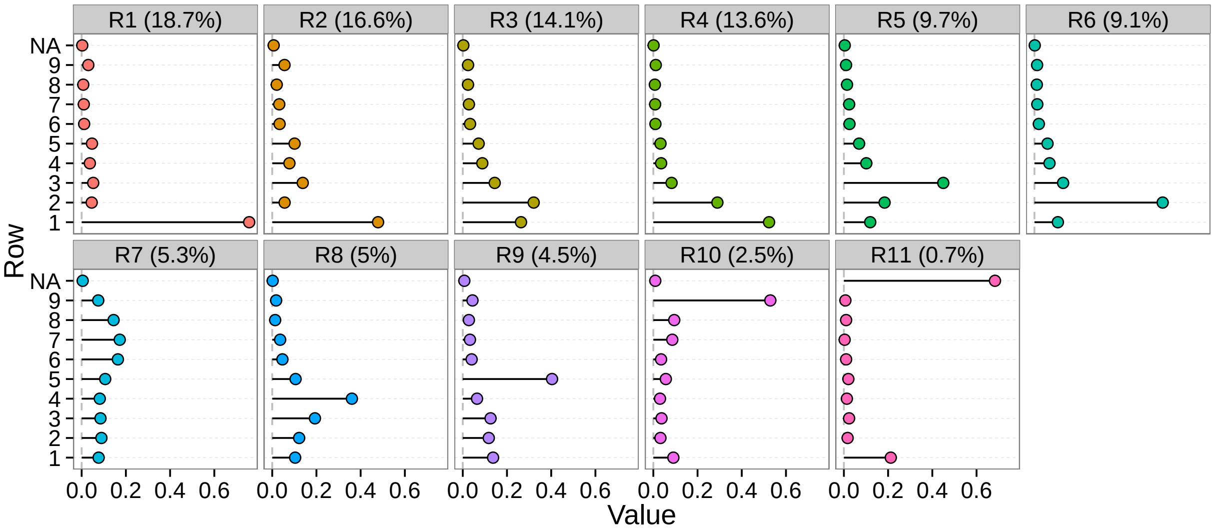


Figure 3.5: Response styles (row main effects) detected by LSBCLUST.

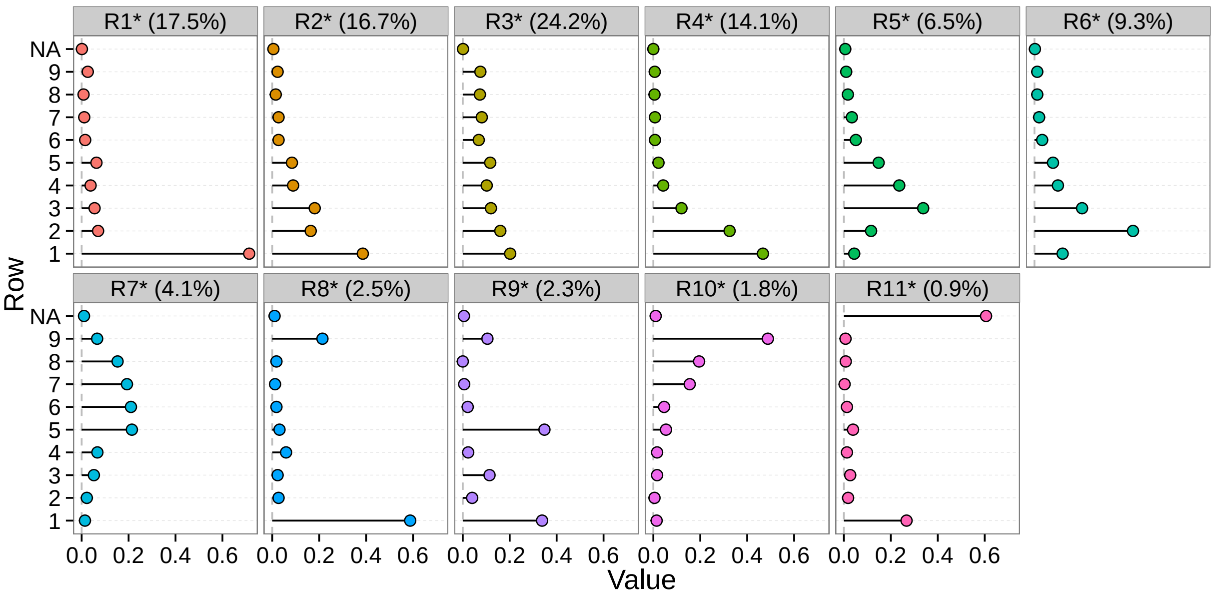


Figure 3.6: Response styles detected by the LC-BML model (see Table 2 in Van Rosmalen et al., 2010). The order of the segments have been adjusted so that it matches those of Figure 3.5 as closely as possible.

The interaction segments detected by LSBCLUST are shown as biplots in Figure 3.7. The overall fit is very high at 99.1% for two dimensions, with 88.5% attributed to the first dimension alone. The goodness-of-fit for the rating categories and value items are given in Table 3.3 and 3.4. All rating categories except the 'NA' category fit very well; likewise the fit for the items are very high in almost all cases.

Looking at the rating categories, the variation on the first dimension is associated almost exclusively with effects on rating category 1. Categories 1 and 2 are involved in the largest effects, but these categories are almost orthogonal. This indicates that a large effect on category 1 does not necessarily translate to a similarly large effect on category 2. The second dimension is associated with effects on category 2 and to a lesser extent categories 4 through 9. The directions of the axes for these latter categories are very similar, indicating that when a respondent uses one of these categories other persons in the same segment are likely to use one of these ratings too. This suggests that we should perhaps consider collapsing these categories into a single one. Effects on the 'NA' category is very small – being very close to the origin it is not labelled in the biplots.

Segment I1 (31.1%) contains respondents who attach most importance to a sense of belonging, warm relationships and self-respect, and find excitement the least important value. The effects in this segment are not particularly large, indicating that these persons do not consider the values to differ very much in importance. Segment I2 (20.8%) boasts much larger effects, with self-respect now considered the most important value followed by security and fun and enjoyment. The least important values in this segment are excitement, being respected and self-fulfilment. In I3 (20.3%), self-fulfilment and self-respect are most valued while belonging and excitement are not considered important. Segment I4 (15.4%) is characterized by the great importance attached to being respected and self-respect, while a sense of belonging, excitement and self-fulfilment are not considered important. Finally, in I5 (12.4%) the most importance is attached to fun and enjoyment, with self-respect also considered important. Least important in this segment is a sense of belonging and excitement.

Drawing comparisons between our segments and those detected by LC-BML in Van Rosmalen et al. (2010) proves to be a difficult task. This is not entirely unexpected as there are substantial differences between the methods. The most important difference is that LSBCLUST models the interactions directly whilst the LC-BML models the response probabilities. Hence the resulting sets of biplots are on different scales.

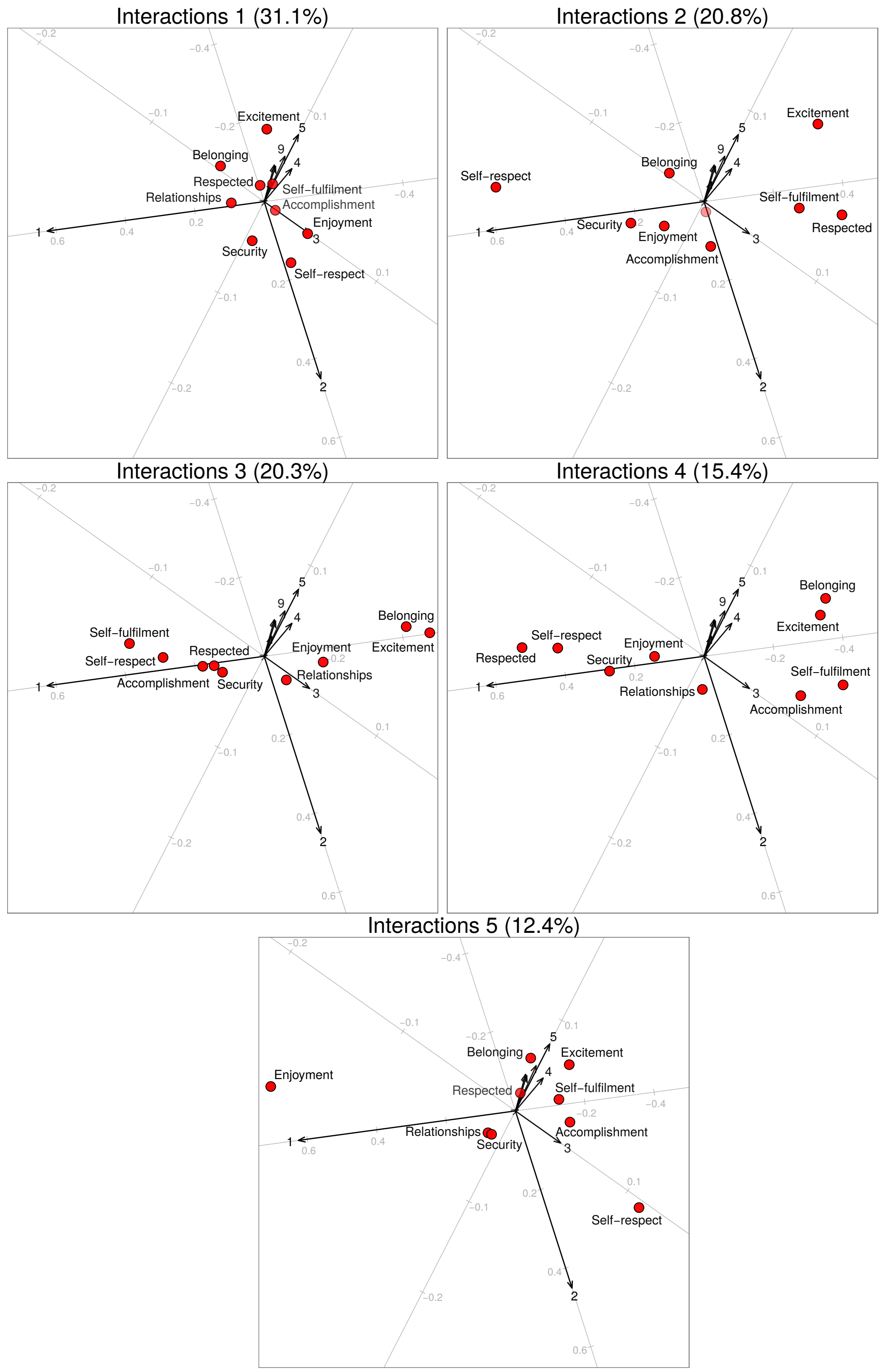


Figure 3.7: Biplots for the interaction segments detected in the LOV data. Rating categories are indicated as vectors and biplot axes, while the LOV items are indicated by labelled points. Points, lines and labels are shaded according to their goodness-of-fit (see Tables 3.3 and 3.4).

	All segments
1	1.00
2	1.00
3	0.90
4	0.94
5	0.98
6	0.91
7	0.92
8	0.85
9	0.86
NA	0.60

Table 3.3: Rating category fit for the LOV data across all interaction segments.

	Interaction segment				
Item	**I1**	**I2**	**I3**	**I4**	**I5**
Belonging	0.97	0.98	1.00	0.98	0.94
Excitement	0.97	0.99	1.00	1.00	0.98
Relationships	0.92	0.35	0.96	0.97	0.96
Self-fulfilment	0.80	0.99	1.00	1.00	0.95
Respected	0.91	0.99	0.98	1.00	0.73
Enjoyment	0.99	0.98	0.95	0.99	1.00
Security	0.97	1.00	0.99	1.00	0.96
Self-respect	0.95	1.00	1.00	1.00	0.98
Accomplishment	0.73	0.99	1.00	0.99	0.96

Table 3.4: Item fit for the LOV data across all interaction segments.

3.5 Conclusions

LSBCLUST is a modelling framework for three-way data, where one of the three ways is clustered over whilst the corresponding matrix slices are approximated by low-rank decompositions. The philosophy is to alternate between finding clusters of similar two-way matrices and doing low-rank matrix decompositions within each cluster. The clustering is done simultaneously with respect to up to four different aspects of these matrix slices, namely the overall mean responses, the row means, the column means, and the row-column interactions. These are the four elements of the biadditive (or bilinear) model used to approximate each of the matrix slices. Which of these terms are included in the model depends on the choice of identifiability constraints, as elegantly parametrized by $\boldsymbol{\delta}$. We show that in eight out of nine unique choices for $\boldsymbol{\delta}$, the combination of the bilinear model and least-squares loss allows the four clustering problems to be treated independently. This important property greatly simplifies the complexity of the clustering problem, which also has positive implications for model selection and the interpretation of the results. The low-rank decompositions of the interaction cluster means lead to readily interpretable biplots which greatly aid in the interpretation of the results.

We argue that LSBCLUST is a useful and natural alternative to more traditional three-way matrix decomposition methods such as PARAFAC/CANDECOMP and TUCKALS3. Whereas these methods can be difficult to interpret, hard to fit and require careful study to understand properly, LSBCLUST uses a combination of well-known statistical methods in k-means cluster analysis, low-rank decompositions of two-way matrices and biplots. Since least-squares loss functions are used, the problems can be treated very efficiently in software. Such software implementing LSBCLUST has been developed in the form of an eponymous R (R Core Team, 2015) package. The package, **lsbclust** (Schoonees, 2015a), is available for download from the Comprehensive R Archive Network (CRAN, `http://cran.r-project.org`).

There are some points that require further research. The treatment of missing values have not been discussed, apart from adding a dedicated catch-all missing category in the LOV example (Section 3.4.2). More refined techniques can be employed, such as iterative majorization (Kiers, 1997), and should be investigated in the future. In terms of model selection, a wide variety of alternatives to the scree test can and should be investigated. There are a number of promising methods out there, including using multiple criteria and taking a vote to determine the most attractive choice. We note that the rank of the low-rank decomposition can also be considered as a model selection step. Furthermore, it would be possible to add case weights to the methodology. An advantage of case weights is that it allows a mechanism for implementing the bootstrap (e.g. Efron and Tibshirani, 1994) to assess the variability of any given solution.

Finally, we have compared the results of LSBCLUST when $\boldsymbol{\delta} = (0, 1, 0, 0)$ and the LC-BML model of Van Rosmalen et al. (2010) in the LOV example (Section 3.4.2). This required reformulating a two-way data set of categorical variables as three-way data by introducing dummy variables for each of the rating categories as the third way. Interestingly, there were obvious similarities between the response styles detected by both methods, but not so much among the interaction segments. The differences in the interaction clusters found can most likely be attributed to some important differences between the two methods: crisp clustering (LSBCLUST) versus probabilistic clustering (LC-BML), least-squares approximation of matrices (LSBCLUST) compared to generalized linear modelling of response probabilities (LC-BML), and the inclusion of covariates in the LC-BML. A definite advantage of LSBCLUST over LC-BML is that the results can be computed in a fraction of the time needed to run the Expectation-Maximization algorithm needed for the latter.

Appendix

Appendix A: Orthogonality

Here we discuss the orthogonality of the decomposition (3.9), and consequently also (3.10). For the decomposition (3.9) to be orthogonal, it must be shown that all six cross-products occurring among the terms in (3.9) are zero. We treat each of these cross-products in turn.

1. For the cross-product between the interaction term and the row term, it holds that

$$\begin{aligned} &\operatorname{tr}\left(\mathbf{J}_J^{(\delta_1)}\left(\mathbf{X}_i-\mathbf{C}_u\mathbf{D}_u'\right)\mathbf{J}_K^{(\delta_2)}\right)\left(\delta_2\mathbf{J}_J^{(\delta_4)}\left(\tfrac{1}{K}\mathbf{X}_i\mathbf{1}_K-\boldsymbol{a}_s\right)\mathbf{1}_K'\right)' \\ &= \delta_2\operatorname{tr}\mathbf{J}_J^{(\delta_1)}\left(\mathbf{X}_i-\mathbf{C}_u\mathbf{D}_u'\right)\mathbf{J}_K^{(\delta_2)}\mathbf{1}_K\left(\tfrac{1}{K}\mathbf{1}_K'\mathbf{X}_i'-\boldsymbol{a}_s'\right)\mathbf{J}_J^{(\delta_4)} \\ &= 0. \end{aligned} \tag{3.30}$$

   The last equality follows since when $\delta_2 = 1$, $\mathbf{J}_K^{(\delta_2)}\mathbf{1}_K = \mathbf{J}_K\mathbf{1}_K = \mathbf{0}$. When $\delta_2 = 0$, the equality is trivial.

2. For the cross-product between the interaction term and the column term, the result is analogous to the above, except that now the equality $\mathbf{1}_J'\mathbf{J}_J = \mathbf{0}'$ is used.

3. For the cross-product between the interaction term and the term for the overall mean, we have that

$$\begin{aligned} &\operatorname{tr}\left(\mathbf{J}_J^{(\delta_1)}\left(\mathbf{X}_i-\mathbf{C}_u\mathbf{D}_u'\right)\mathbf{J}_K^{(\delta_2)}\right)\left((\delta_1\delta_3+\delta_2\delta_4-\delta_1\delta_2)\left(\tfrac{1}{JK}\mathbf{1}_J'\mathbf{X}_i\mathbf{1}_K-e_r\right)\mathbf{1}_J\mathbf{1}_K'\right)' \\ &= (\delta_1\delta_3+\delta_2\delta_4-\delta_1\delta_2)\left(\mathbf{1}_J'\mathbf{J}_J^{(\delta_1)}\left(\mathbf{X}_i-\mathbf{C}_u\mathbf{D}_u'\right)\mathbf{J}_K^{(\delta_2)}\mathbf{1}_K\right)\left(\tfrac{1}{JK}\mathbf{1}_J'\mathbf{X}_i\mathbf{1}_K-e_r\right) \\ &= 0. \end{aligned} \tag{3.31}$$

   This cross-product equals zero whenever at least one of the following is true: $\delta_1 = 1$, $\delta_2 = 1$ or $\delta_1\delta_3+\delta_2\delta_4-\delta_1\delta_2 = 0$. But whenever both $\delta_1 = \delta_2 = 0$, $\delta_1\delta_3+\delta_2\delta_4-\delta_1\delta_2 = 0$ irrespective of δ_3 and δ_4. Hence the cross-product always equals zero.

4. For the cross-product between the row and column terms, we have

$$\begin{aligned}
&\operatorname{tr}\left(\delta_1 \mathbf{1}_J \left(\tfrac{1}{J}\mathbf{1}_J'\mathbf{X}_i - \boldsymbol{b}_t'\right)\mathbf{J}_K^{(\delta_3)}\right)\left(\delta_2 \mathbf{J}_J^{(\delta_4)}\left(\tfrac{1}{K}\mathbf{X}_i\mathbf{1}_K - \boldsymbol{a}_s\right)\mathbf{1}_K'\right)' \\
&= \delta_1\delta_2\left(\tfrac{1}{J}\mathbf{1}_J'\mathbf{X}_i - \boldsymbol{b}_t'\right)\mathbf{J}_K^{(\delta_3)}\mathbf{1}_K\left(\tfrac{1}{K}\mathbf{1}_K'\mathbf{X}_i' - \boldsymbol{a}_s'\right)\mathbf{J}_J^{(\delta_4)}\mathbf{1}_J \\
&= \begin{cases}\left(\tfrac{1}{J}\mathbf{1}_J'\mathbf{X}_i\mathbf{1}_K - \boldsymbol{b}_t'\mathbf{1}_K\right)\left(\tfrac{1}{K}\mathbf{1}_K'\mathbf{X}_i'\mathbf{1}_J - \boldsymbol{a}_s'\mathbf{1}_J\right) & \text{if } \boldsymbol{\delta} = (1,1,0,0)' \\ 0 & \text{otherwise.}\end{cases}
\end{aligned} \tag{3.32}$$

Deducing when the cross-product equals zero uses the same concepts as above, but when $\boldsymbol{\delta} = (1,1,0,0)'$ none of these apply and the cross-product is not necessarily equal to zero.

5. The cross-product between the column term and the term for the overall mean also does not necessarily equal zero. Here we can derive the following:

$$\begin{aligned}
&\operatorname{tr}\left(\delta_1 \mathbf{1}_J \left(\tfrac{1}{J}\mathbf{1}_J'\mathbf{X}_i - \boldsymbol{b}_t'\right)\mathbf{J}_K^{(\delta_3)}\right)\left((\delta_1\delta_3 + \delta_2\delta_4 - \delta_1\delta_2)\left(\tfrac{1}{JK}\mathbf{1}_J'\mathbf{X}_i\mathbf{1}_K - e_r\right)\mathbf{1}_J\mathbf{1}_K'\right)' \\
&= J\delta_1(\delta_1\delta_3 + \delta_2\delta_4 - \delta_1\delta_2)\left(\tfrac{1}{J}\mathbf{1}_J'\mathbf{X}_i - \boldsymbol{b}_t'\right)\mathbf{J}_K^{(\delta_3)}\mathbf{1}_K\left(\tfrac{1}{JK}\mathbf{1}_J'\mathbf{X}_i\mathbf{1}_K - e_r\right) \\
&= \begin{cases}-J\left(\tfrac{1}{J}\mathbf{1}_J'\mathbf{X}_i\mathbf{1}_K - \boldsymbol{b}_t'\mathbf{1}_K\right)\left(\tfrac{1}{JK}\mathbf{1}_J'\mathbf{X}_i\mathbf{1}_K - e_r\right) & \text{if } \boldsymbol{\delta} = (1,1,0,0)' \\ 0 & \text{otherwise.}\end{cases}
\end{aligned} \tag{3.33}$$

The last line follows from the fact that the crossproduct equals zero if $\delta_1 = 0$, if $\delta_1 = 1$ and $\delta_3 = 1$, and when $\delta_1\delta_3 + \delta_2\delta_4 - \delta_1\delta_2 = 0$. Hence consideration must be given to the four cases $\delta_1 = 1, \delta_2 \in \{0,1\}, \delta_3 = 0, \delta_4 \in \{0,1\}$. It is easy to see that $\delta_1\delta_3 + \delta_2\delta_4 - \delta_1\delta_2 = \delta_2\delta_4 - \delta_2 = 0$ in all these cases except when $\boldsymbol{\delta} = (1,1,0,0)'$.

6. Analogously to the above, consider the cross-product between the row term and the term for the overall mean:

$$\begin{aligned}
&\operatorname{tr}\left(\delta_2 \mathbf{J}_J^{(\delta_4)}\left(\tfrac{1}{K}\mathbf{X}_i\mathbf{1}_K - \boldsymbol{a}_s\right)\mathbf{1}_K'\right)\left((\delta_1\delta_3 + \delta_2\delta_4 - \delta_1\delta_2)\left(\tfrac{1}{JK}\mathbf{1}_J'\mathbf{X}_i\mathbf{1}_K - e_r\right)\mathbf{1}_J\mathbf{1}_K'\right)' \\
&= K\delta_2(\delta_1\delta_3 + \delta_2\delta_4 - \delta_1\delta_2)\mathbf{1}_J'\mathbf{J}_J^{(\delta_4)}\left(\tfrac{1}{K}\mathbf{X}_i\mathbf{1}_K - \boldsymbol{a}_s\right)\left(\tfrac{1}{JK}\mathbf{1}_J'\mathbf{X}_i\mathbf{1}_K - e_r\right) \\
&= \begin{cases}-K\left(\tfrac{1}{K}\mathbf{1}_J'\mathbf{X}_i\mathbf{1}_K - \mathbf{1}_J'\boldsymbol{a}_s\right)\left(\tfrac{1}{JK}\mathbf{1}_J'\mathbf{X}_i\mathbf{1}_K - e_r\right) & \text{if } \boldsymbol{\delta} = (1,1,0,0)' \\ 0 & \text{otherwise.}\end{cases}
\end{aligned} \tag{3.34}$$

The expression equals zero when $\delta_2 = 0$, when $\delta_2 = 1$ and $\delta_4 = 1$ or when $\delta_1\delta_3 + \delta_2\delta_4 - \delta_1\delta_2 = 0$. Considering the cases $\delta_1 \in \{0,1\}, \delta_2 = 1, \delta_3 \in \{0,1\}, \delta_4 = 1$ then, it can be seen that $\delta_1\delta_3 + \delta_2\delta_4 - \delta_1\delta_2 = \delta_1\delta_3 - \delta_1 = 0$ except when $\boldsymbol{\delta} = (1,1,0,0)'$.

Consequently the decomposition in (3.9) and (3.10) is valid for all $\boldsymbol{\delta}$, except for $\boldsymbol{\delta} = (1,1,0,0)'$.

Appendix B: Degrees-of-freedom

The degrees-of-freedom (DF) associated with each of the four subproblems in (3.10) are given in Table B1. These are split according to the two sets of parameters that must be estimated in each case: the cluster memberships and the cluster means, or low-rank approximations of these. In this Appendix we give a short overview of how to calculate these quantities.

The number of parameters affiliated with the cluster memberships are somewhat ambiguous to quantify. For each person and C clusters, there are arguably $C - 1$ free parameters to find since this is the number of dummy coded variables needed to encode the clustering. However, only one of these C parameters are nonzero, and equal to one, since we are doing crisp clustering. So using $N(C - 1)$ DF for the cluster memberships is a conservative approach. Recall also that the inclusion of certain types of clusters depends on the choice of $\boldsymbol{\delta}$, and consequently also the number of parameters associated with the clustering.

The DF for the estimation of the mean effects not only depends on $\boldsymbol{\delta}$, but also on the choice between cases (a), (b) and (c). This choice affects the SVD used and hence the length and orthogonality restrictions imposed within the SVD itself. We illustrate the calculation here for case (a) only and leave cases (b) and (c) to be verified by the reader. Firstly, there are P singular values to be estimated in (3.15). Secondly, the estimation of $\mathbf{UL}$ require in total JP parameters, but these are subject to length, orthogonality and, depending on δ_1, sum-to-zero constraints. Each column of $\mathbf{UL}$ must be of length one, orthogonal to the other columns and possibly have a zero sum, giving $P + \binom{P}{2} + \delta_1 P$ restrictions. Finally, $\mathbf{VL}$ has UKP parameters but these are subject to the same length and orthogonality restrictions on the P columns. However, δ_2 now optionally applies sum-to-zero constraints within each block of $\mathbf{D}_*$. Hence there are $P + \binom{P}{2} + \delta_2 UP$ restrictions here.

The DF are therefore given by the total number of parameters $P + JP + UKP$ in $\Gamma\mathbf{L}$, $\mathbf{UL}$ and $\mathbf{VL}$, minus all these constraints.

Appendix C: Biplot Diagnostics

Here we note fit diagnostics for cases (a) and (b) – case (c) has already been treated in Section 3.3.2. For case (a), we again have an orthogonal decomposition, based on (3.15), namely

$$\left(\mathbf{J}_J^{(\delta_1)}\overline{\mathbf{X}}_{(C)}\right)\left(\mathbf{J}_J^{(\delta_1)}\overline{\mathbf{X}}_{(C)}\right)' = \left(\mathbf{J}_J^{(\delta_1)}\mathbf{C}_1\mathbf{D}_*'\right)\left(\mathbf{J}_J^{(\delta_1)}\mathbf{C}_1\mathbf{D}_*'\right)' + \left(\mathbf{J}_J^{(\delta_1)}\overline{\mathbf{X}}_{(C)} - \mathbf{J}_J^{(\delta_1)}\mathbf{C}_1\mathbf{D}_*'\right)\left(\mathbf{J}_J^{(\delta_1)}\overline{\mathbf{X}}_{(C)} - \mathbf{J}_J^{(\delta_1)}\mathbf{C}_1\mathbf{D}_*'\right)'. \quad (3.35)$$

	Cluster membership	**Mean effects**
Overall means	$\delta^* N(R-1)$	$\delta^* R$
Rows	$\delta_2 N(S-1)$	$\delta_2 S(J-\delta_4)$
Colums	$\delta_1 N(T-1)$	$\delta_1 T(K-\delta_3)$
Interactions (a)	$N(U-1)$	$P(J+UK-P-\delta_1-\delta_2 U)$
Interactions (b)	$N(U-1)$	$P(UJ+K-P-\delta_1 U-\delta_2)$
Interactions (c)	$N(U-1)$	$UP(J+K-P-\delta_1-\delta_2)$

Table B1: Degrees-of-freedom associated with each of the four subproblems, split into the cluster membership parameters and estimates of the mean effects. Note that $\delta^* = \delta_1\delta_3 + \delta_2\delta_4 - \delta_1\delta_2$.

The row fit can therefore be defined as

$$\boldsymbol{r}_{\text{fit}} = \left[\operatorname{diag} \mathbf{J}_J^{(\delta_1)} \mathbf{C}_1 \mathbf{D}_*' \mathbf{D}_* \mathbf{C}_1' \mathbf{J}_J^{(\delta_1)}\right] \left[\operatorname{diag} \mathbf{J}_J^{(\delta_1)} \overline{\mathbf{X}}_{(C)} \overline{\mathbf{X}}_{(C)}' \mathbf{J}_J^{(\delta_1)}\right]^{-1} \mathbf{1}_J \tag{3.36}$$

and the column fit as

$$\boldsymbol{c}_{\text{fit}} = \left[\operatorname{diag} \mathbf{D}_* \mathbf{C}_1' \mathbf{J}_J^{(\delta_1)} \mathbf{C}_1 \mathbf{D}_*'\right] \left[\operatorname{diag} \overline{\mathbf{X}}_{(C)}' \mathbf{J}_J^{(\delta_1)} \overline{\mathbf{X}}_{(C)}\right]^{-1} \mathbf{1}_{UK}. \tag{3.37}$$

For case (b), a similar decomposition follows from (3.18) and we have

$$\boldsymbol{r}_{\text{fit}} = \left[\operatorname{diag} \mathbf{C}_* \mathbf{D}_1' \mathbf{J}_K^{(\delta_2)} \mathbf{D}_1 \mathbf{C}_*'\right] \left[\operatorname{diag} \overline{\mathbf{X}}_{(R)} \mathbf{J}_K^{(\delta_2)} \overline{\mathbf{X}}_{(R)}'\right]^{-1} \mathbf{1}_{UJ} \tag{3.38}$$

for the rows, and similarly

$$\boldsymbol{c}_{\text{fit}} = \left[\operatorname{diag} \mathbf{J}_K^{(\delta_2)} \mathbf{D}_1 \mathbf{C}_*' \mathbf{C}_* \mathbf{D}_1 \mathbf{J}_K^{(\delta_2)}\right] \left[\operatorname{diag} \mathbf{J}_K^{(\delta_2)} \overline{\mathbf{X}}_{(R)}' \overline{\mathbf{X}}_{(R)} \mathbf{J}_K^{(\delta_2)}\right]^{-1} \mathbf{1}_K \tag{3.39}$$

for the columns.

CHAPTER 4

Quadratic Majorization of the Rating Scale Model

4.1 Introduction

Item response theory (IRT) is concerned with the construction of joint models for the responses of multiple respondents to a battery of items in a test. The goal of such models is to express the probability of a response as a mathematical function of that respondent's location on a latent trait and the item and rating scale characteristics (Van der Linden and Hambleton, 1997). A typical example of such a trait is mathematical ability. IRT models customarily include person, item and answer category parameters, and are used intensively in test construction, which involves the calibration of the item batteries in trial runs.

The original IRT models were developed for dichotomous items, where a test taker scores one point for each correct answer (e.g. Lord et al., 1968). The focus has since shifted to include models for polytomous items, where multiple answer categories are possible (e.g. Samejima, 1969; Andrich, 1978a; Masters, 1982). Concurrently, IRT models have also found applications in non-traditional settings, such as in the measurement of attitudes, or nonability testing, as it is sometimes known. In such a context, the latent trait can be seen, for example, as a political spectrum with the extremes indicating extreme left or extreme right political tendencies. The latent trait is hence not measured in terms of an ideal point or unfolding model, but on a scale with definite low and high extremes: a so-called cumulative model.

A more functional view of the latent trait is that it is simply a construct which captures all the dependency between answers extracted from the same individual. In nonability testing this may be a more helpful view, as more factors may come into consideration when formulating an answer than in traditional psychometric settings. An important distinction between this context and that of the typical aptitude or ability settings in psychometrics is that an item does not have a correct response. A person must merely map his or her opinion to the set of ratings presented, such as categories ranging from "disagree completely" to "agree completely." In aptitude testing a person may also be expected to answer a mathematical problem with one of a number of possible answers, but in contrast to attitude measurement such items have definite correct answers.

Estimating IRT models can be challenging because of the inclusion of one or more such latent variables. Three prominent maximum likelihood estimation methods have emerged, namely joint maximum likelihood (JML), conditional maximum likelihood (CML) and marginal maximum likelihood (MML) estimation. JML estimation seeks to estimate the latent trait value θ for all persons directly. In contrast, CML estimation uses the properties of Rasch models, a subfamily of IRT models, to maximize the likelihood conditional on the sufficient statistic for the person parameters. This conditional likelihood does not contain the person parameters and can be maximized to obtain estimates of the item and answer category parameters. Finally, MML estimation involves assuming a distribution for the person parameters and maximizing the marginal likelihood function constructed by (numerically) integrating over this distribution.

Here we discuss and illustrate a new algorithm for JML estimation, applied to the rating scale model (RSM), a polytomous Rasch model (Andrich, 1978a). The iterative majorization procedure we derive leads to an iterative least-squares method for obtaining maximum likelihood estimates for the person, item and answer category parameters jointly. The simple form of the updates enables a speedy procedure to be developed which is guaranteed to converge monotonically to the global optimum of the likelihood function. Missing values and case weights can be handled seamlessly. We also show how penalized JML estimation can be implemented by applying a quadratic penalty on the linear predictors. This can be useful for solving the issue of infinite parameter estimates that can occur in models based on the (multinomial) logistic transform, as most IRT models are. Penalized estimation can also lead to better estimates, as Hoerl and Kennard (1970) show in their existence theorem for ridge regression.

It should be noted that MML and CML estimation have traditionally been favoured over JML estimation by practitioners. The reason for this is that the estimates of the person parameters are not consistent as the number of respondents tends towards infinity ($N \to \infty$) for a fixed number of items, J (e.g. Fischer and Molenaar, 1995). Simulation studies have nevertheless shown that the resulting bias can be compensated for by a small adjustment. The JML estimates are consistent when $N \to \infty$, the number of items $J \to \infty$ and $N/J \to \infty$. We note that there is a strong recent trend of favouring estimation methods with superior predictive power over those with sound traditional statistical properties (e.g. Hoerl and Kennard, 1970; Tibshirani, 1996; Zou and Hastie, 2005). Also, there are some cases where the latent trait parameters have to be modelled explicitly, such as in finite mixture models where the complete likelihood is needed (Rost, 1990; Rost and von Davier, 1995; Von Davier and Rost, 1995). In finite mixture models it is insufficient to use CML or MML estimation as a proper likelihood is required for the Expectation-Maximization (EM) algorithm (Frick et al., 2015). Larger sample sizes, both in terms of the number of observations N and especially with respect to the number of items J in test batteries, means that JML estimation can be expected to perform similarly, or even better than the alternatives.

The RSM is introduced in the next section, followed by a discussion of the iterative majorization estimation approach. In Section 4.4 we discuss the application of iterative majorization for the JML estimation of the RSM, and this is followed by empirical illustrations of the method in Section 4.5.

4.2 The Rating Scale Model

The rating scale model (RSM) was introduced by Andrich (1978a) for the situation where an individual i ($i = 1,2,\dots,N$) answers item j from a collection of J items on a common rating scale $1,2,\dots,K$ (see also Andrich (1978b, 1982)). It is a latent trait adjacent-category logit model for ordinal responses (Agresti, 2010), where a single latent dimension is assumed to underlie all J items. The RSM contains three sets of parameters, one set each for characterizing the persons on the latent trait, the items and the rating categories. We denote these parameters as follows. The person parameter θ_i indicates the attitude of person i on the latent trait, β_j the item difficulty for item j, and τ_k the rating category parameter for category k.

The RSM is a special case of the partial credit model (PCM; Masters, 1982), which does not assume that all items are answered on the same rating scale. The PCM using item-specific rating category parameters of the form $\beta_{jk} = \beta_j + \tau_k$, where $k = 1,2,\dots,K_j$. The RSM is thus a PCM model which makes the simplifying assumption that the category scores are equally spaced across different items. It is more appropriate than the PCM in cases where a common rating scale is used, as is the case in many surveys and also in the application in Section 4.5.

Use $Y_{ij} \in \{1,2,\dots,K\}$ to denote the random variable representing the response of person i to item j, and let $\boldsymbol{\gamma}' = (\boldsymbol{\theta}', \boldsymbol{\beta}', \boldsymbol{\tau}')$ contain all the model parameters. The RSM models the adjacent-category logit as

$$\begin{aligned} \operatorname{logit} \mathsf{P}\left(Y_{ij} = k | Y_{ij} \in \{k-1, k\}, \boldsymbol{\gamma}\right) &= \log \frac{\mathsf{P}\left(Y_{ij} = k | \boldsymbol{\gamma}\right)}{\mathsf{P}\left(Y_{ij} = k-1 | \boldsymbol{\gamma}\right)} \\ &= \theta_i - \left(\beta_j + \tau_k\right) \quad k = 2,\dots,K. \end{aligned} \tag{4.1}$$

The probability of choosing category k over category $k-1$ is therefore governed by a (restricted) dichotomous IRT model. Here β_j denotes the overall location of item j on the latent continuum, whereas τ_k denotes the distance between any item's overall latent location to the point where the probability of answering with either rating $k-1$ or k is equal. This point is given by $\beta_j + \tau_k$. The β_j therefore shifts the general location for each item, while the τ_k monitors the spread of the item thresholds across all items. It follows from (4.1) that

$$\log \frac{\mathsf{P}\left(Y_{ij} = k|\boldsymbol{\gamma}\right)}{\mathsf{P}\left(Y_{ij} = 1|\boldsymbol{\gamma}\right)} = \sum_{l=2}^{k} \left(\theta_i - \beta_j - \tau_l\right), \quad k = 2, \ldots, K.$$

Subsequently, the response probability is

$$\mathsf{P}(Y_{ij} = k|\boldsymbol{\gamma}) = \pi_{ijk}(\boldsymbol{\gamma}) = \frac{\exp \sum_{l=2}^{k} \left(\theta_i - \beta_j - \tau_l\right)}{1 + \sum_{l=2}^{K} \exp \sum_{k=2}^{l} \left(\theta_i - \beta_j - \tau_k\right)}, \quad k = 2, \ldots, K. \tag{4.2}$$

Rating category $k = 1$ is used as the so-called reference category; it would also be possible to use category K as reference. We prefer the symmetric but over-determined specification

$$\pi_{ijk}(\boldsymbol{\gamma}) = \frac{\exp \sum_{l=1}^{k} \left(\theta_i - \beta_j - \tau_l\right)}{\sum_{l=1}^{K} \exp \sum_{k=1}^{l} \left(\theta_i - \beta_j - \tau_k\right)}, \quad k = 1, \ldots, K. \tag{4.3}$$

In the traditional formulation (4.2) it is implicitly assumed that $\theta_i - \beta_j - \tau_1 = 0$. We now replace this requirement with the identifiability constraint $\sum_{k=1}^{K} \left(\theta_i - \beta_j - \tau_k\right) = 0$. A cumulative, mean centred form of the linear predictors is more convenient. Particularly, define $\kappa_k = \sum_{l=1}^{k} \tau_l$ such that

$$\begin{aligned} \sum_{l=1}^{k} \left(\theta_i - \beta_j - \tau_k\right) &= k\theta_i - k\beta_j - \kappa_k; \text{ and} \\ \frac{1}{K} \sum_{l=1}^{k} \left(\theta_i - \beta_j - \tau_k\right) &= \overline{K}\theta_i - \overline{K}\beta_j - \overline{\kappa}, \end{aligned}$$

with $\overline{K} = \frac{K+1}{2}$ and $\overline{\kappa} = \frac{1}{K} \sum_{k=1}^{K} \kappa_k$. By dividing both the numerator and denominator by the same term, (4.3) becomes

$$\begin{aligned} \pi_{ijk}(\boldsymbol{\gamma}) &= \frac{\exp\left((k - \overline{K})\theta_i - (k - \overline{K})\beta_j - (\kappa_k - \overline{\kappa})\right)}{\sum_{l=1}^{K} \exp\left((l - \overline{K})\theta_i - (l - \overline{K})\beta_j - (\kappa_l - \overline{\kappa})\right)} \\ &= \frac{\exp \delta_{ijk}}{\sum_{l=1}^{K} \exp \delta_{ijl}}, \quad k = 1, \ldots, K. \end{aligned} \tag{4.4}$$

Since δ_{ijk} is mean centred, it follows trivially that $\sum_{k=1}^{K} \delta_{ijk} = 0$, as required. The reference category is now the virtual mean rating category. Furthermore, it is evident from (4.4) that an adjacent-category logit model is equivalent to a nominal multinomial logit model with the predictors adjusted by an equally-spaced category-specific constant factor of $k - \overline{K}$, a well-known result (Agresti, 2010). For properly identifying these models, the origins of the three sets of parameters must be fixed. We therefore impose zero-sum constraints $\sum_{i=1}^{N} \theta_i = \sum_{j=1}^{J} \beta_j = \sum_{k=1}^{K} \kappa_k = 0$, so that the model has $P = N + J + K - 3$ degrees-of-freedom.

The joint probability for the J responses of person i is arrived at by the conditional independence assumption. This states that all dependence between the answers of the same person is captured by the latent trait θ. If this is the case, and $\boldsymbol{Y}_i$ contains all random variables $Y_{ij},\ j = 1, \ldots, J$, then

$$\mathsf{P}(\boldsymbol{Y}_i = \boldsymbol{y}_i|\boldsymbol{\gamma}) = \prod_{j=1}^{J}\prod_{k=1}^{K} \pi_{ijk}(\boldsymbol{\gamma})^{y_{ijk}}. \tag{4.5}$$

Here y_{ijk} denotes the observed value of the Bernoulli random variable Y_{ijk} indicating whether that response was category k, or otherwise. These are collected in the three-dimensional arrays $\underline{\mathbf{y}}$ and $\underline{\mathbf{Y}}$ of size $N \times J \times K$ respectively. The expression (4.5) defines a probability distribution on all $J \times K$ indicator matrices with the entries $y_{ijk} \in \{0, 1\},\ j = 1, \ldots, J, k = 1, \ldots, K$. Combining this with the assumption of random sampling of respondents' answer vectors, we have the likelihood function

$$L(\boldsymbol{\gamma}|\underline{\mathbf{Y}} = \underline{\mathbf{y}}) = \prod_{i=1}^{N}\prod_{j=1}^{J}\prod_{k=1}^{K} \pi_{ijk}(\boldsymbol{\gamma})^{w_i y_{ijk}}, \tag{4.6}$$

where $w_i \geq 0, i = 1, \ldots, N$, are nonnegative case weights. The likelihood (4.6) defines a probability distribution on all three-way arrays $\underline{\mathbf{Y}}$ of size $N \times J \times K$, with slice i along the first dimension containing the $J \times K$ indicator matrix for person i, weighted by w_i. As a useful extension for handling missing responses, we also assume that $Y_{ijk} = 0$ for all $k = 1, \ldots, K$ when person i did not answer item j. Hence we define $M_{ij} = \sum_{k=1}^{K} Y_{ijk}$ with realizations m_{ij} as the Bernoulli random variables indicating whether person i responded to item j, or otherwise.

Finally, a note on sufficiency. Define the random variable $R_i = \mathbf{1}'\boldsymbol{Y}_i$ as the total score achieved by person i across all J items. It can be shown for the RSM that

$$\begin{aligned}\mathsf{P}(\boldsymbol{Y}_i = \boldsymbol{y}_i|R_i = r_i, \theta_i, \boldsymbol{\beta}, \boldsymbol{\kappa}) &= \frac{\mathsf{P}\left(R_i = r_i|\boldsymbol{Y}_i = \boldsymbol{y}_i, \theta_i, \boldsymbol{\beta}, \boldsymbol{\kappa}\right) \mathsf{P}\left(\boldsymbol{Y}_i = \boldsymbol{y}_i|\theta_i, \boldsymbol{\beta}, \boldsymbol{\kappa}\right)}{\mathsf{P}\left(R_i = r_i|\theta_i, \boldsymbol{\beta}, \boldsymbol{\kappa}\right)} \\ &= \frac{\mathsf{P}\left(\boldsymbol{Y}_i = \boldsymbol{y}_i|\theta_i, \boldsymbol{\beta}, \boldsymbol{\kappa}\right)}{\mathsf{P}\left(R_i = r_i|\theta_i, \boldsymbol{\beta}, \boldsymbol{\kappa}\right)}\end{aligned} \tag{4.7}$$

since $\mathsf{P}\left(R_i = r_i | \boldsymbol{Y}_i = \boldsymbol{y}_i, \theta_i, \boldsymbol{\beta}\right) = 1$, which follows from the dependence of R_i on $\boldsymbol{Y}_i$. Writing $\boldsymbol{y}|r_i$ as shorthand for the set $\left\{\boldsymbol{y} : \mathbf{1}'\boldsymbol{y} = r_i\right\}$, it can be verified that

$$\begin{aligned}
\mathsf{P}(\boldsymbol{Y}_i = \boldsymbol{y}_i | R_i = r_i, \theta_i, \boldsymbol{\beta}, \boldsymbol{\kappa}) &= \frac{\prod_{j=1}^{J} \exp \sum_{k=1}^{y_{ij}} \delta_{ijk}}{\sum_{\boldsymbol{y}|r_i} \prod_{j=1}^{J} \exp \sum_{k=1}^{y_j} \delta_{ijk}} \\
&= \mathsf{P}(\boldsymbol{Y}_i = \boldsymbol{y}_i | R_i = r_i, \boldsymbol{\beta}, \boldsymbol{\kappa}),
\end{aligned} \tag{4.8}$$

which does not depend on θ_i. Hence R_i is a sufficient statistic for θ_i, such that all the information on θ_i in the data is contained in the total score of person i over all items. In CML, the part of the likelihood based only on (4.8) is maximized to estimate $\boldsymbol{\beta}$ and $\boldsymbol{\kappa}$. Sufficient statistics also exist for the item and rating category parameters. In our notation, sufficient statistics for θ_i, β_j and κ_k are

$$\begin{aligned}
T_{\theta_i}(\mathbf{Y}) &= w_i \sum_{j=1}^{J} m_{ij}(Y_{ij} - \overline{K}) \\
T_{\beta_j}(\mathbf{Y}) &= \sum_{i=1}^{N} w_i m_{ij}(Y_{ij} - \overline{K}) \\
T_{\kappa_k}(\mathbf{Y}) &= \sum_{i=1}^{N} \sum_{j=1}^{J} w_i Y_{ijk}.
\end{aligned} \tag{4.9}$$

These can be interpreted as the weighted, centred and observed total score for person i across all items, the weighted, centred and observed score for item j across all persons, and the weighted number of times category k was selected across all persons and items, respectively. We use scaled versions of the observed sufficient statistics as starting values in our algorithm (see Section 4.4.1).

4.3 Iterative Majorization

Iterative majorization algorithms approach the minimization (or maximization) of a complicated function by iteratively minimizing a simpler auxiliary or *majorizing* function (De Leeuw, 2011b; Groenen et al., 1995; De Leeuw, 1994). The majorizing function is chosen such that it is easy to minimize and always has a larger value than the target function, except in the so-called support point where the functions are equal. Minimizing this function leads to a reduction in the value of the original target function. At each step of the iterative procedure the majorizing function is updated and minimized. This guarantees monotone convergence to a local mininum. The family of such majorizing algorithms can be seen as a generalization of the EM algorithm, and they are also known as *optimization transfer* methods (Lange et al., 2000).

Constructing majorizing functions relies on mathematical results such as the Cauchy-Schwartz, Jensen and Popoviciu inequalities, or on Taylor's theorem (De Leeuw, 2011b). Here we will use the latter to derive a procedure also known as quadratic majorization (Böhning and Lindsay, 1988). Consider minimizing a function $f(\boldsymbol{x})$, which will be a negative log-likelihood function in our context. The quadratic Taylor series expansion of f at $\boldsymbol{x}^*$ is

$$f(\boldsymbol{x}) = f(\boldsymbol{x}^*) + (\boldsymbol{x} - \boldsymbol{x}^*)'\partial f(\boldsymbol{x}^*) + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}^*)'\partial^2 f(\boldsymbol{\xi})(\boldsymbol{x} - \boldsymbol{x}^*).$$

Here $\boldsymbol{\xi}$ lies on the line between $\boldsymbol{x}$ and $\boldsymbol{x}^*$, $\partial f(\boldsymbol{x}^*)$ denotes the vector of first partial derivatives of f evaluated at $\boldsymbol{x}^*$ and $\partial^2 f(\boldsymbol{\xi})$ is the Hessian matrix of second partial derivatives of f evaluated at $\boldsymbol{\xi}$. Let $\mathbf{B}$ be an invertible, symmetric matrix such that $\mathbf{B} - \partial^2 f(\boldsymbol{\xi})$ is positive semi-definite. Define

$$g(\boldsymbol{x},\boldsymbol{x}^*) = f(\boldsymbol{x}^*) + (\boldsymbol{x} - \boldsymbol{x}^*)'\partial f(\boldsymbol{x}^*) + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}^*)'\mathbf{B}(\boldsymbol{x} - \boldsymbol{x}^*). \tag{4.10}$$

It follows that

$$\begin{aligned} g(\boldsymbol{x},\boldsymbol{x}) &= f(\boldsymbol{x}) \\ g(\boldsymbol{x},\boldsymbol{x}^*) &\geq f(\boldsymbol{x}) \end{aligned}$$

so that $g(\boldsymbol{x},\boldsymbol{x}^*)$ is said to majorize f with support point $\boldsymbol{x}^*$. Minimizing f can therefore proceed by successively minimizing $g(\boldsymbol{x},\boldsymbol{x}^{(b)})$ over $\boldsymbol{x}$, where $\boldsymbol{x}^{(b)}$ is the current estimate of $\boldsymbol{x}$ at iteration b. A starting configuration $\boldsymbol{x}^{(0)}$ is necessary. This procedure is guaranteed to decrease the value of g in each iteration. It also follows that for a vector-valued function $\boldsymbol{h}$ of the parameters in $\boldsymbol{x}$,

$$\begin{aligned} g(\boldsymbol{h}(\boldsymbol{x}),\boldsymbol{h}(\boldsymbol{x})) &= f(\boldsymbol{h}(\boldsymbol{x})) \\ g(\boldsymbol{h}(\boldsymbol{x}),\boldsymbol{h}(\boldsymbol{x}^*)) &\leq f(\boldsymbol{h}(\boldsymbol{x})), \end{aligned}$$

and therefore g also majorizes f as a function of $\boldsymbol{h}(\boldsymbol{x})$ with support point $\boldsymbol{h}(\boldsymbol{x}^*)$. Simple forms of $\boldsymbol{h}$, such as linear functions used to construct the linear predictors from a set of parameters in generalized linear models, are especially easy to handle. Equation (4.10) can be rewritten as (De Leeuw, 2011b)

$$g(\boldsymbol{x},\boldsymbol{x}^*) = f(\boldsymbol{x}^*) + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{z})'\mathbf{B}(\boldsymbol{x} - \boldsymbol{z}) - \frac{1}{2}\partial f(\boldsymbol{x}^*)\mathbf{B}^{-1}\partial f(\boldsymbol{x}^*) \tag{4.11}$$

where $\boldsymbol{z}$, the current target, is given by

$$\boldsymbol{z} = \boldsymbol{x}^* - \mathbf{B}^{-1}\partial f(\boldsymbol{x}^*). \tag{4.12}$$

Note that only the second term in (4.11) depends on $\boldsymbol{x}$ so that minimizing $g(\boldsymbol{x},\boldsymbol{x}^*)$ over $\boldsymbol{x}$ is a simple least-squares problem, namely

$$\min_{\boldsymbol{x}}(\boldsymbol{x}-z)'\mathbf{B}(\boldsymbol{x}-z). \tag{4.13}$$

Linear functions of the form $\boldsymbol{h}(\boldsymbol{x}) = \mathbf{A}\boldsymbol{x}$ is a special case, where $\mathbf{A}$ is a design matrix. The solution is

$$\begin{aligned}\hat{\boldsymbol{x}} &= (\mathbf{A}'\mathbf{B}\mathbf{A})^{-1}\mathbf{A}'\mathbf{B}z \\ &= (\mathbf{A}'\mathbf{B}\mathbf{A})^{-1}\mathbf{A}'\mathbf{B}(\mathbf{A}\boldsymbol{x}^* - \mathbf{B}^{-1}\partial f(\mathbf{A}\boldsymbol{x}^*)) \\ &= \boldsymbol{x}^* - (\mathbf{A}'\mathbf{B}\mathbf{A})^{-1}\mathbf{A}'\partial f(\mathbf{A}\boldsymbol{x}^*)). \end{aligned} \tag{4.14}$$

The rate of convergence can be shown to be only linear, with a tighter bound $\mathbf{B}$ on the Hessian leading to faster convergence. De Leeuw and Heiser (1980) showed that a simple way of increasing the convergence rate whilst retaining the global convergence property is to use an over-relaxed update of the form

$$z = \boldsymbol{x}^* - 2\mathbf{B}^{-1}\partial f(\boldsymbol{x}^*). \tag{4.15}$$

This makes the update step twice as large as in (4.12), and will reduce the number of iterations required by approximately one half (De Leeuw, 2011b).

In the subsequent section we show how to derive such a majorization algorithm for the RSM.

4.4 Estimating the Rating Scale Model

We seek to minimize the negative log-likelihood of the RSM, which follows from (4.6) as

$$\mathcal{L} = -\sum_{i=1}^{N}\sum_{j=1}^{J}\sum_{k=1}^{K} w_i y_{ijk} \log \pi_{ijk}. \tag{4.16}$$

The dependence of $\mathcal{L}$ on $\boldsymbol{\gamma}$ is suppressed for notational brevity. Before proceeding to derive a majorization algorithm for this function, an optional quadratic penalty term is added for the sum of squared linear predictors $\sum_{k=1}^{K}\delta_{ijk}^2$. This penalty is controlled by the tuning parameter λ_{ij}, which controls the overall size of the linear predictors for a specific person i and item j. The objective function hence becomes

$$\mathcal{L} = -\sum_{i=1}^{N}\sum_{j=1}^{J}\sum_{k=1}^{K} w_i y_{ijk} \log \pi_{ijk} + \sum_{i=1}^{N}\sum_{j=1}^{J} \lambda_{ij} \sum_{k=1}^{K} \delta_{ijk}^2$$
$$= \sum_{i=1}^{N}\sum_{j=1}^{J} \mathcal{L}_{ij}. \tag{4.17}$$

In our formulation, penalizing the δ_{ijk} instead of the individual parameters is computationally convenient. Note though that the dependence of λ_{ij} on person i and item j makes it possible to apply different penalties for different person and item parameters, albeit indirectly through δ_{ijk}. This is not possible for κ_k in this formulation.

Our optimization strategy is to construct a majorizing function for each $\mathcal{L}_{ij}$ separately. The sum of the resulting NJ majorizing functions is then necessarily also a majorizing function of $\mathcal{L}$. To achieve this, the first partial derivatives of $\mathcal{L}_{ij}$ can be derived as

$$\frac{\partial \mathcal{L}_{ij}}{\partial \delta_{ijk}} = w_i \left(\pi_{ijk} m_{ij} - y_{ijk}\right) + 2\lambda_{ij}\delta_{ijk}. \tag{4.18}$$

Here π_{ijk} is the function (4.4) of δ_{ijk}, and m_{ij} are the observed missingness indicators. The entries of the Hessian matrix $\mathbf{H}_{ij} = \partial^2 \mathcal{L}_{ij}$ are

$$w_i m_{ij} \left(\delta^{kl}\pi_{ijk} - \pi_{ijk}\pi_{ijl}\right) + 2\lambda_{ij}\delta^{kl}, \tag{4.19}$$

with δ^{kl} denoting the Kronecker delta.

To find a function that majorizes $\mathcal{L}_{ij}$, we must find a matrix $\mathbf{B}_{ij}$ such that $\mathbf{B}_{ij} - \mathbf{H}_{ij} \geq 0$. Now, any square matrix $\mathbf{A}$ is dominated by $v_{\max}\mathbf{I}$ in the sense that $v_{\max}\mathbf{I} - \mathbf{A} \geq 0$ (positive semi-definite), where $v_{\max}$ is the largest eigenvalue of $\mathbf{A}$. We can derive an upper bound on the largest eigenvalue of $\mathbf{H}_{ij}$ by using the mathematical result that the largest eigenvalue of a square matrix is smaller than any norm of that matrix. This therefore also holds for the maximum absolute column sum norm, which has a simple form for $\mathbf{H}_{ij}$ (De Leeuw, 2011a).

The maximum absolute column sum norm of $\mathbf{H}_{ij}$, with $\mathbf{H}_{ij}(k,l)$ the element of $\mathbf{H}_{ij}$ in row k and column l, is

$$\begin{aligned}
\|\mathbf{H}_{ij}\|_1 &= \max_{k=1}^{K} \sum_{l=1}^{K} \left|\mathbf{H}_{ij}(k,l)\right| \\
&= \max_{k=1}^{K} \left|\mathbf{H}_{ij}(k,k)\right| + \sum_{l\neq k}^{K} \left|\mathbf{H}_{ij}(k,l)\right| \\
&= \max_{k=1}^{K} \underbrace{w_i m_{ij}\pi_{ijk}(1-\pi_{ijk}) + 2\lambda_{ij}}_{\left|\mathbf{H}_{ij}(k,k)\right|} + \underbrace{w_i m_{ij}\pi_{ijk} \sum_{l\neq k}^{K} \pi_{ijl}}_{\sum_{l\neq k}^{K}\left|\mathbf{H}_{ij}(k,l)\right|} \\
&= \max_{k=1}^{K} \left\{2w_i m_{ij}\pi_{ijk}(1-\pi_{ijk}) + 2\lambda_{ij}\right\}, \qquad (4.20)
\end{aligned}$$

since $\sum_{l\neq k}^{K} \pi_{ijl} = 1 - \pi_{ijk}$. The only unknown parameters in (4.20) are the π_{ijk}, for which we know from inspection or Popoviciu's inequality on variances (Popoviciu, 1935) that

$$\pi_{ijk}(1-\pi_{ijk}) \leq \frac{1}{4}.$$

Therefore

$$\mathbf{H}_{ij} \lesssim \left(\frac{1}{2}w_i m_{ij} + 2\lambda_{ij}\right)\mathbf{I}_K = \rho_{ij}\mathbf{I}_K = \mathbf{B}_{ij}. \qquad (4.21)$$

To majorize the complete $\mathcal{L} = \sum_{i=1}^{N}\sum_{j=1}^{J}\mathcal{L}_{ij}$ then, we choose $\mathbf{B}$ to be the $NJK \times NJK$ diagonal matrix with the $\mathbf{B}_{ij}$ $(i = 1,\ldots,N; j = 1,\ldots,J)$ as diagonal blocks.

It can be shown that $\mathcal{L}$ in (4.17) is a convex function in $\boldsymbol{\gamma}$, such that any local minimum $\hat{\boldsymbol{\gamma}} = \min_{\boldsymbol{\gamma}} \mathcal{L}(\boldsymbol{\gamma})$ is also a global minimum. To establish this, we show that $\mathbf{H}_{ij} \gtrsim 0$. It is known that a symmetric diagonally dominant real matrix with nonnegative diagonal entries is positive semi-definite (see Briggs, 2015, for example). Hence it suffices to show that $\mathbf{H}_{ij}$ is diagonally dominant. A matrix is diagonally dominant when the absolute value of the diagonal entry in each column is larger than or equal to the sum of the absolute values of all the other entries in that column. This has in fact already been proven in (4.20), since we have shown there that the diagonal entries $\left|\mathbf{H}_{ij}(k,k)\right| = w_i m_{ij}\pi_{ijk}(1-\pi_{ijk}) + 2\lambda_{ij} \geq \sum_{l\neq k}^{K}\left|\mathbf{H}_{ij}(k,l)\right| = w_i m_{ij}\pi_{ijk}(1-\pi_{ijk})$ for all k. Therefore $\mathcal{L}_{ij}$ is a convex function in the δ_{ijk}, $k = 1,\ldots,K$, which we collect in the three-way array $\underline{\Delta}$ of size $N \times J \times K$. Since affine mappings such as $\mathrm{Vec}\,\underline{\Delta} = \mathbf{X}\boldsymbol{\gamma}$ preserve convexity, $\mathcal{L}_{ij}$ is also convex in $\boldsymbol{\gamma}$ (Boyd and Vandenberghe, 2004). Consequently, being a sum of convex functions, (4.17) is also a convex function. Since majorization algorithms are globally convergent, we are therefore guaranteed to locate the global minimum of (4.17). Some starting configurations may however lead to faster convergence of the algorithm than others. This and other computational specifics are outlined next.

4.4.1 Computational Details

Additional notation is required for a full description of the least-squares updates (4.14) in the majorization algorithm. Let $\boldsymbol{\gamma}' = (\boldsymbol{\theta}', \boldsymbol{\beta}', \boldsymbol{\kappa}')$, with $\boldsymbol{\theta}' = (\theta_1, \ldots, \theta_{N-1})$, $\boldsymbol{\beta}' = \beta_1, \ldots, \beta_{J-1}$ and $\boldsymbol{\kappa} = (\kappa_1, \ldots, \kappa_{K-1})'$ be the vector containing the $P = N + J + K - 3$ free parameters. Define $\underline{\Delta}$ to be the $N \times J \times K$ array with entries δ_{ijk} and denote by $\mathbf{X}$ the $NJK \times P$ design matrix such that $\mathbf{X}\boldsymbol{\gamma} = \operatorname{Vec} \underline{\Delta}$. The least-squares update (4.14) at iteration $b + 1$ now becomes

$$\boldsymbol{\gamma}^{(b+1)} = \boldsymbol{\gamma}^{(b)} - (\mathbf{X}'\mathbf{B}\mathbf{X})^{-1}\mathbf{X}' \operatorname{Vec} \frac{\partial \mathcal{L}}{\partial \underline{\Delta}}, \tag{4.22}$$

where $\frac{\partial \mathcal{L}}{\partial \underline{\Delta}}$ is the three-way array of first derivatives with entries (4.18).

Several types of regression contrasts can be used for constructing $\mathbf{X}$. It is particularly advantageous to choose contrasts such that $\mathbf{X}'\mathbf{B}\mathbf{X}$ is block diagonal, since this implies that the least-squares updates for all parameters separate into independent updates for the three parameter sets. Particularly, $\mathbf{X} = \begin{bmatrix} \mathbf{X}_\theta & \mathbf{X}_\beta & \mathbf{X}_\kappa \end{bmatrix}$ so that if $\mathbf{X}_\theta$, $\mathbf{X}_\beta$ and $\mathbf{X}_\kappa$ are mutually orthogonal, $\mathbf{X}'\mathbf{B}\mathbf{X}$ is block diagonal with entries $\mathbf{X}_\theta'\mathbf{B}\mathbf{X}_\theta$, $\mathbf{X}_\beta'\mathbf{B}\mathbf{X}_\beta$ and $\mathbf{X}_\kappa'\mathbf{B}\mathbf{X}_\kappa$. If this is the case, $(\mathbf{X}'\mathbf{B}\mathbf{X})^{-1}$ is similarly block diagonal with entries $(\mathbf{X}_\theta'\mathbf{B}\mathbf{X}_\theta)^{-1}$, $(\mathbf{X}_\beta'\mathbf{B}\mathbf{X}_\beta)^{-1}$ and $(\mathbf{X}_\kappa'\mathbf{B}\mathbf{X}_\kappa)^{-1}$ respectively. Consequently the parameter updates for $\boldsymbol{\theta}$, $\boldsymbol{\beta}$ and $\boldsymbol{\kappa}$ in (4.22) can be done independently, which has computational advantages.

To ensure separability of the parameter updates, we use contrasts of the form

$$\mathbf{C}_L = \begin{bmatrix} \mathbf{I}_{L-1} \\ -\mathbf{1}_{L-1}' \end{bmatrix} \tag{4.23}$$

as the basis for the design matrix $\mathbf{X}$. Specifically, $\mathbf{X}_\theta$, $\mathbf{X}_\beta$ and $\mathbf{X}_\kappa$ are based on $\mathbf{C}_N$, $\mathbf{C}_J$ and $\mathbf{C}_K$ respectively. This implies that $\theta_N = -\sum_{i=1}^{N-1} \theta_i$, $\beta_J = -\sum_{j=1}^{J-1} \beta_j$ and $\kappa_K = -\sum_{k=1}^{K-1} \kappa_k$ so that the zero-sum constraint on the parameters are naturally enforced. For $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$, the constants $k - \overline{K}$ in the vector $\boldsymbol{k}$ must also be accounted for. Denoting by $\mathbf{R}$ the $N \times J$ matrix with ρ_{ij} as entries, it can be shown that

$$\begin{aligned} \mathbf{X}_\theta'\mathbf{B}\mathbf{X}_\theta &= \|\boldsymbol{k}\|^2 \mathbf{C}_N' \operatorname{diag}(\mathbf{R}\mathbf{1}_J)\mathbf{C}_N \\ \mathbf{X}_\beta'\mathbf{B}\mathbf{X}_\beta &= \|\boldsymbol{k}\|^2 \mathbf{C}_J' \operatorname{diag}(\mathbf{R}'\mathbf{1}_N)\mathbf{C}_J \\ \mathbf{X}_\kappa'\mathbf{B}\mathbf{X}_\kappa &= \mathbf{1}_N'\mathbf{R}\mathbf{1}_J\mathbf{C}_K'\mathbf{C}_K. \end{aligned} \tag{4.24}$$

Inverses for these matrices can be found explicitly via the Sherman-Morrison formula (Sherman and Morrison, 1950), saving costly numerical computations. Furthermore, the crossproduct $\mathbf{X}' \operatorname{Vec} \frac{\partial \mathcal{L}}{\partial \underline{\Delta}}$ performs elementary operations on the array of first derivatives. Denote by $\underline{\mathbf{K}}$ the $N \times J \times K$ array containing the entries of $\boldsymbol{k}$ varying only in the third dimension. Let $\boldsymbol{s}^{(\theta)}$ be the sum of $\underline{\mathbf{K}} \circ \frac{\partial \mathcal{L}}{\partial \underline{\Delta}}$ over the last two dimensions, where $\circ$ denotes the elementwise (or Hadamard) product of the arrays. Similarly, $\boldsymbol{s}^{(\beta)}$ is the sum of $\underline{\mathbf{K}} \circ \frac{\partial \mathcal{L}}{\partial \underline{\Delta}}$ over the first and third dimensions, whereas $\boldsymbol{s}^{(\kappa)}$ is the sum of $\frac{\partial \mathcal{L}}{\partial \underline{\Delta}}$ over the first two dimensions. The crossproducts are then simply

$$
\begin{aligned}
\mathbf{X}_\theta' \frac{\partial \mathcal{L}}{\partial \underline{\Delta}} &= \mathbf{C}_N \boldsymbol{s}^{(\theta)} \\
\mathbf{X}_\beta' \frac{\partial \mathcal{L}}{\partial \underline{\Delta}} &= -\mathbf{C}_J \boldsymbol{s}^{(\beta)} \\
\mathbf{X}_\kappa' \frac{\partial \mathcal{L}}{\partial \underline{\Delta}} &= -\mathbf{C}_K \boldsymbol{s}^{(\kappa)}.
\end{aligned} \tag{4.25}
$$

We can therefore completely circumvent doing expensive linear algebra calculations in the least-squares updates.

Finally, we construct starting configurations for the parameters from their respective sufficient statistics. Good starting configurations will lead to faster convergence. Our general strategy is first to standardize the sufficient statistics (4.9), and then to convert these standardized values to the quantiles of some distribution. For $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$, the sufficient statistics are scaled to the interval $[0, 1]$ using their theoretical minima and maxima. These are then transformed to standard Gaussian quantiles and centred. Finite quantiles are ensured by altering the standardized sufficient statistics for extreme scoring persons and items by a small quantity ϵ. For $\boldsymbol{\kappa}$, the category counts are first converted to proportions, taking into account the case weights and missingness structure. These are then converted to approximate standard Gaussian values via the probit transformation, again ensuring finiteness by adjusting extreme values by a small constant. The resulting starting configurations lead to much faster convergence than random starts.

In the following section we illustrate the use of the algorithm and discuss some key properties of the RSM.

4.5 Applications

4.5.1 Simulated Example

As a first example, consider a simulated data set with $N = 1000$ observations, $J = 15$ items and $K = 7$ rating categories. The true parameter values are all simulated from standard

Gaussian distributions, and the data set is constructed by sampling from the corresponding RSM probabilities (4.4). There are 10 persons who scored either the minimum or maximum number of points on the hypothetical test. These persons present a problem for the estimation algorithm since the theoretical θ for these persons are $-\infty$ and ∞ respectively, a well-known characteristic of logistic models. For these parameters, the log-likelihood function very slowly approaches these extremes. Nonfinite parameter values are associated with extreme sufficient statistics, and can also arise for parameters in $\boldsymbol{\beta}$ and $\boldsymbol{\kappa}$. However in practice this occurs much less frequently than for $\boldsymbol{\theta}$, since it is unlikely that all persons answer an item with the same rating category, or that a rating category is not used in the entire data set. In extreme cases, response styles (Baumgartner and Steenkamp, 2001, for example) may however cause unexpected behaviour.

Here we consider three strategies for dealing with the problem: (a) retain these persons and continue to estimate their θ, (b) remove these persons from the data so that only the remaining θ are estimated, and (c) apply a penalty on the squared linear predictors. Strategy (a) is unsatisfactory because we can expect the algorithm to converge very slowly to the infinite theoretical values. In contrast, strategy (b) assures convergence but does not give estimates of θ for these persons, and moreover does not take the responses of these persons into account when estimating the β and κ. Just because these persons scored as high or as low as possible on this test does not mean that they will never do worse or better on other items. It is quite reasonable that their θ are not as extreme as the data suggests (and indeed we know this is true for our simulated data set). Strategy (c) allows estimates for all parameters to be found, as will be illustrated below.

4.5.1.1 Unregularized Estimation

Consider first our estimates with strategy (a). The algorithm does not converge and terminates when the maximum number of iterations (500) are reached. On this laptop computer (Intel Core i7-3537U 2.00GHz processor with 10 GB RAM) this takes 6.4 seconds with over-relaxation. Figure 4.1 illustrates the problem with the θ parameters by plotting the estimates after each iteration. It is quite clear that the algorithm has not converged for the sole reason that the θ of the so-called extreme scorers are wandering off to infinity. Figure 4.2 compares the final estimates with the actual values. Colours are assigned according to the total score achieved on the test, which is a sufficient statistic for θ (see Section 4.2). The mean squared error (MSE) for the full $\boldsymbol{\theta}$, $\boldsymbol{\beta}$ and $\boldsymbol{\kappa}$ vectors are 0.180, 0.001 and 0.026 respectively. The overall MSE is 0.177.

Estimation with strategy (b), takes only 4.1 seconds and 359 iterations. Convergence is assumed when the relative change in the negative log-likelihood from one iteration to the next is less than 10^{-8}. The MSE for the estimated parameters in $\boldsymbol{\theta}$, $\boldsymbol{\beta}$ and $\boldsymbol{\kappa}$ are 0.127, 0.001 and 0.014

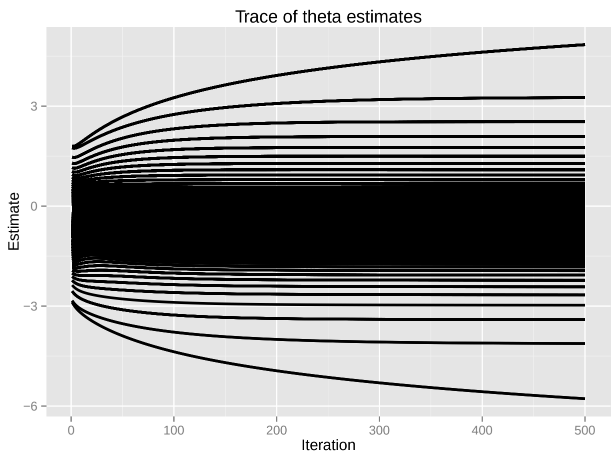


Figure 4.1: The evolution of the estimates of the θ parameters over the iterations of the estimation algorithm.

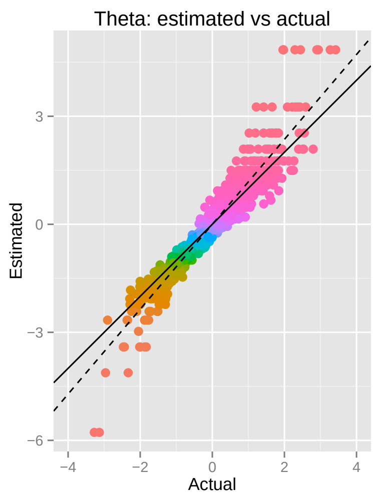


Figure 4.2: A comparison of the actual and estimated θ. Colouration are according to the total score achieved on the hypothetical test, the sufficient statistic for θ. The solid line is the 45 degree line through the origin representing the expected one-to-one relationship, while the dashed line is a fitted regression line.

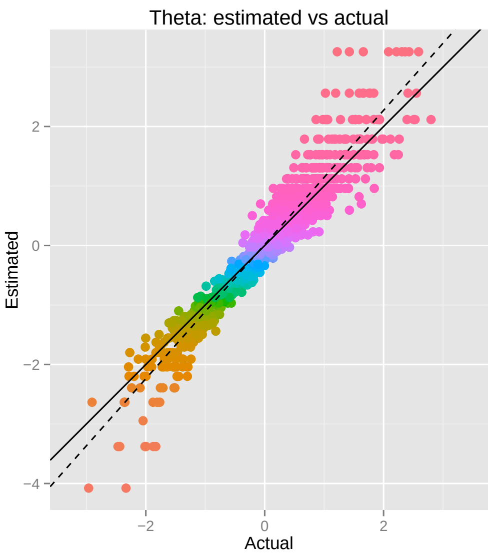


Figure 4.3: A comparison of the actual and estimated θ after removing the extreme scorers.

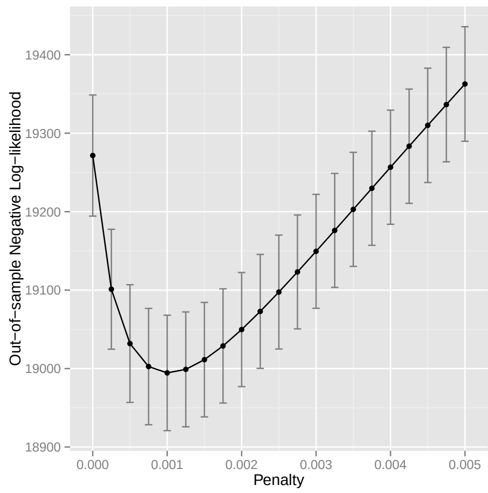


Figure 4.4: Plot of the out-of-sample negative log-likelihood against the penalty parameter λ for the first penalized estimation strategy. Bands showing the extent of the estimated standard errors are also shown.

respectively, and the overall MSE is 0.124. Hence we are doing much better at estimating $\boldsymbol{\theta}$ and $\boldsymbol{\kappa}$ and similarly for $\boldsymbol{\beta}$. The resulting estimates for $\boldsymbol{\theta}$ are shown in Figure 4.3.

4.5.1.2 Regularized Estimation

In strategy (c) we consider estimating all parameters but with $\lambda_{ij} > 0$ in the quadratic penalty term. Two strategies are investigated. In the first of these, we let $\lambda_{ij} = \lambda$ for all i and j in (4.17). In order to select a good value for λ, five-fold cross-validation (CV) is used on the out-of-sample negative log-likelihood. This is assessed over the sequence of candidate λ values ranging from 0 to 0.005, with increments of 0.00025. V-fold CV involves splitting the set of all NJ observations into V folds of roughly equal size, fitting for each fold the model to all observation not in that fold, and determining the contributions of the observations in that fold to the total negative log-likelihood (4.16). This gives an objective estimate of the out-of-sample negative log-likelihood for the entire data set, and is repeated for each candidate value of λ. We can also obtain an estimate of the standard error of the mean out-of-sample loss by calculating the empirical standard deviation of the mean estimated loss over all five folds.

The result of our CV procedure is displayed in Figure 4.4, including bands showing the estimated standard errors. It is evident that nonnegative penalties can lead to smaller out-of-sample negative log-likelihoods. The optimal value $\lambda = 0.0010$ was found to correspond to a loss value of 18994, which is significantly lower than the unpenalized out-of-sample loss of 19272. Fitting the model for this optimal value of λ, convergence is achieved in 1.7 seconds and 145 iterations. The overall MSE is now further reduced to 0.075, with the individual MSE's for

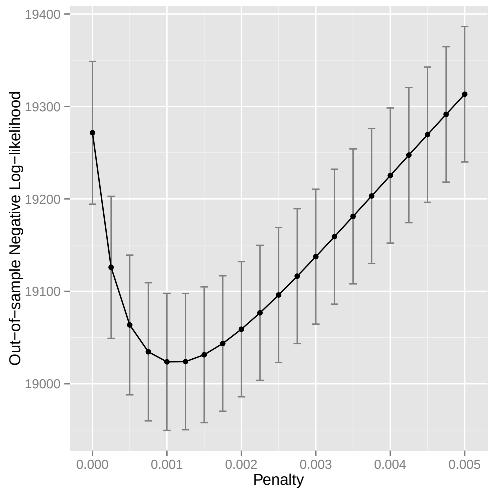


Figure 4.5: The results of CV for the second penalized estimation strategy.

$\boldsymbol{\theta}$, $\boldsymbol{\beta}$ and $\boldsymbol{\kappa}$ being 0.077, 0.0011 and 0.026 respectively. It therefore pays to use the quadratic penalty while retaining all samples in the estimation process, given that a good choice of λ_{ij} can be made. Note however that the MSE for $\boldsymbol{\kappa}$ has increased even though the overall MSE has increased.

As a second regularized approach, we consider letting

$$\lambda_{ij} = \lambda \left| \frac{2t_{\theta_i}(\mathbf{y})}{w_i J(K-1)} \right|, \tag{4.26}$$

where $\lambda > 0$ must be chosen. Here $t_{\theta_i}(\mathbf{y})$ are the observed values of the sufficient statistic for θ_i in (4.9). In (4.26) these values are standardized to $[0, 1]$ in a manner which takes into account the absolute size of the sufficient statistic. Hence extreme values of the sufficient statistic, be it either the minimum or maximum, will invoke the maximum penalty. As before, we select λ via 5-fold CV, with candidate values ranging from 0 to 0.005 with increments of 0.0005. The optimal λ was found to be 0.0010. Fitting the RSM model with these values for λ_{ij} results in a minimum loss of 19024, with convergence in 1.7 seconds and 150 iterations. The overall MSE is now further reduced to 0.073, with the individual MSE's for $\boldsymbol{\theta}$, $\boldsymbol{\beta}$ and $\boldsymbol{\kappa}$ now being 0.075, 0.0007 and 0.017 respectively. Penalizing relative to the sufficient statistics for $\boldsymbol{\theta}$ therefore seems to be the optimal strategy for this simulated example. A comparison of all three sets of parameter estimates with their actual values are given in Figure 4.6.

4.5.2 European Social Survey Data

As a further illustration, we apply our algorithm to 11 personal well-being items taken from wave 6 of the European Social Survey (ESS; www.europeansocialsurvey.org), which was conducted

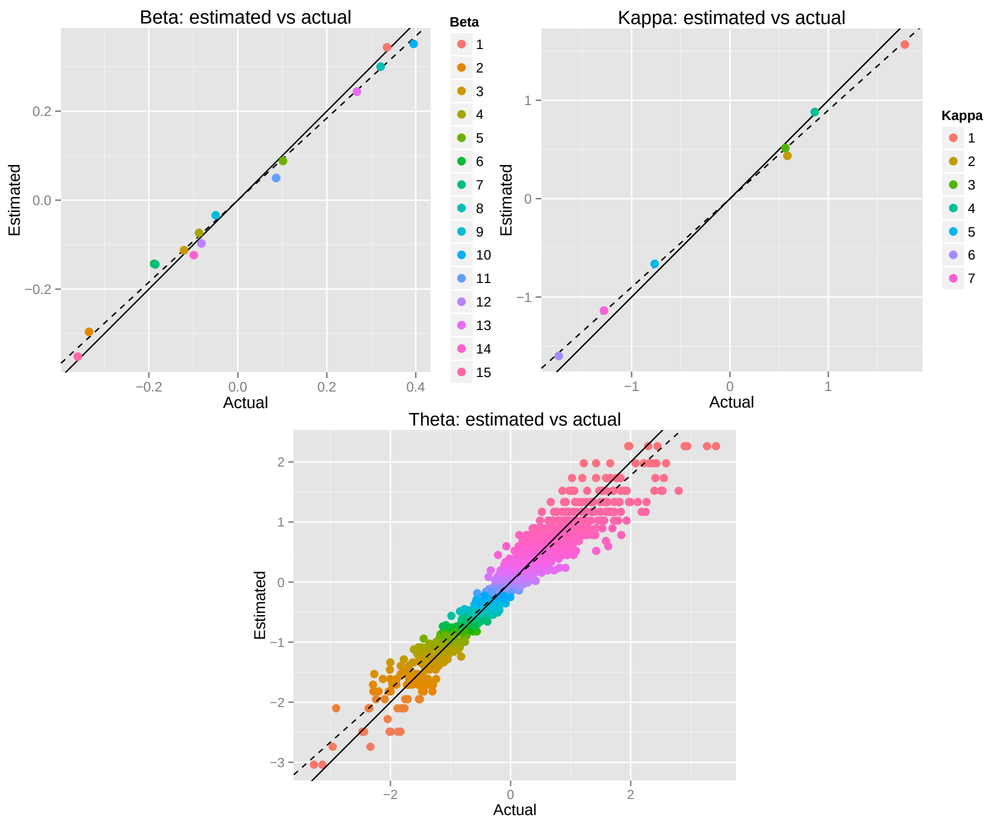


Figure 4.6: A comparison of all three sets of parameter estimates with their actual values for the second penalized estimation strategy.

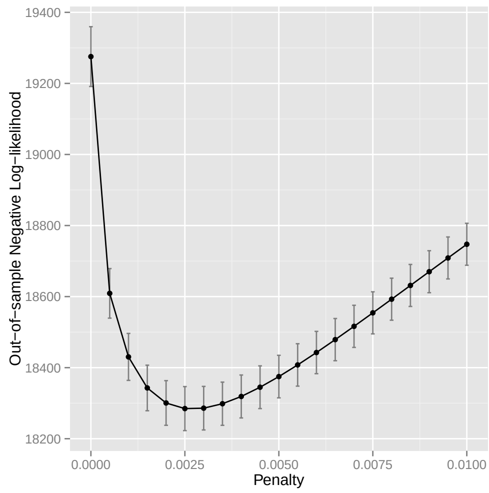


Figure 4.7: Cross-validation results for the ESS data using 5 folds.

in 2012. We use only the Dutch sample of 1845 respondents. The items are summarized in Table 4.1. A rating scale with 4 categories was used, with labels '1 – none or almost none of the time, '2 – some of the time,' '3 – most of the time' and '4 – all or almost all of the time.' We reversed Items 4, 6, 9 and 11 so that lower scores on these items indicate higher personal well-being, in accordance with the other items. Hence for these items the categories are effectively labelled '1 – all or almost all of the time,' '2 – most of the time,' '3 – some of the time' and '4 – none or almost none of the time.' This reversal assumes that the rating scale was interpreted symmetrically by the respondents, which we feel is a reasonable assumption for this simple scale. We apply post-stratification weights as recommended for the ESS, which adjusts for the sampling design as well as for sampling and nonresponse errors.

The penalized approach (4.26) is applied with five-fold CV to select the value of the penalty parameter. The results of this procedure are shown in Figure 4.7, resulting in the optimal choice $\lambda = 0.0025$. Convergence took 77 iterations and 0.8 seconds. The estimates of the item and category parameters are given in Tables 4.1 and 4.2 respectively, and the corresponding category characteristic curves (CCCs) for the 11 items are shown in Figure 4.8. As is characteristic of the RSM, the curves all have the same shape. The item parameters β_j give the overall location of the item on the latent scale θ, whereas $\beta_j + \tau_k$ gives the value of the latent trait for which the probabilities of selecting either category $k-1$ or category k are equal for item j – see (4.1). Higher estimates of β_j imply that θ must be comparably high before a person is likely to select category 3 or 4 for that item, and vice versa for lower estimates of β_j.

Observe that the rating categories are indeed ordered. A larger θ is associated with higher probabilities for the higher rating categories. The relatively large estimates for τ_3 and τ_4 indicate that these categories are in general associated with persons with very low personal well-being.

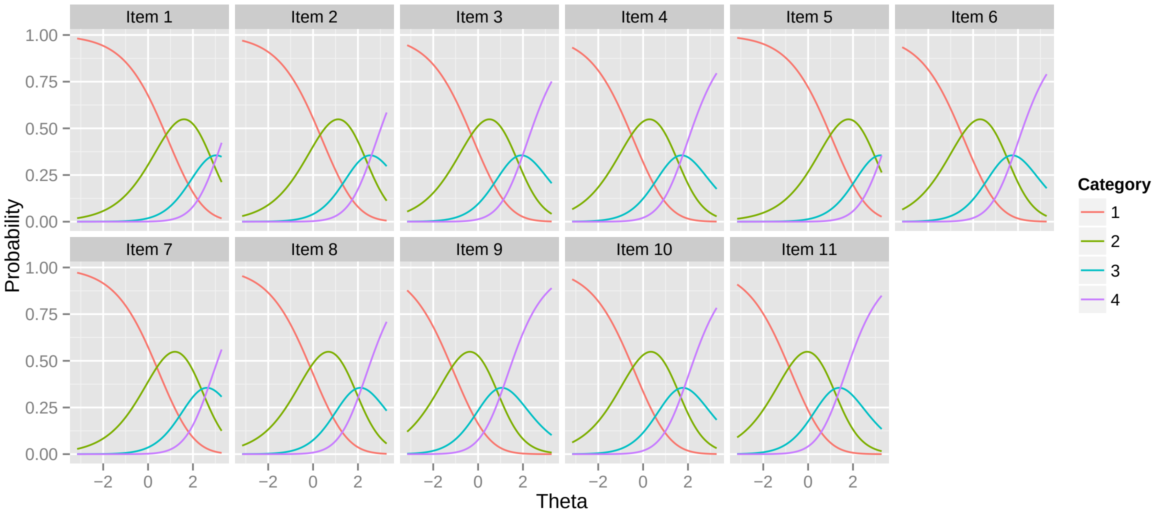


Figure 4.8: The category characteristic curves (CCCs) of the 11 ESS items as fitted by the RSM. The curves show the estimated probability of choosing each category as a function of the latent trait.

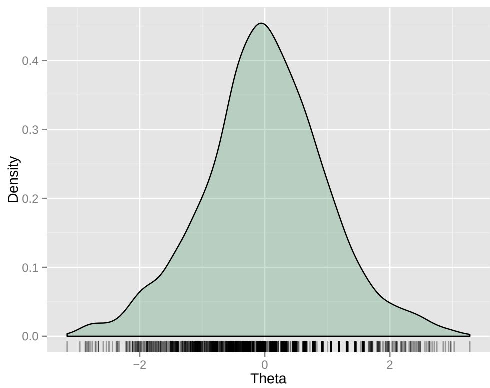


Figure 4.9: A kernel density estimate of the distribution of the estimated θ. The individual estimates are indicated on the x-axis.

Item	Description	$\hat{\beta}_j$
	Please indicate how much of the time during the past week...	
1	...you felt depressed?	0.93
2	...you felt that everything you did was an effort?	0.44
3	...your sleep was restless?	−0.17
4*	...you were happy?	−0.39
5	...you felt lonely?	1.12
6*	...you enjoyed life?	−0.36
7	...you felt sad?	0.52
8	...you could not get going?	0.01
9*	...you had a lot of energy?	−1.04
10	...you felt anxious?	−0.33
11*	...you felt calm and peaceful?	−0.72

Table 4.1: The 11 ESS personal well-being items included in the Dutch sample. In our analysis, items 4, 6, 9 and 11 were reversed so that lower scores indicates higher personal well-being, in correspondence with the other items. The estimated item parameters are given in the third column.

k	κ_k	τ_k
1	−1.37	−1.37
2	−1.50	−0.13
3	0.36	1.86
4	2.51	2.16

Table 4.2: The parameter estimates for the category parameters κ, and the corresponding τ estimates.

For a specific person i and item j, the estimated odds of choosing category 2 over category 1 is $\exp(\theta_i - \beta_j)\exp(-\tau_2 + \tau_1) = 0.29\exp(\theta_i - \beta_j)$. Similarly, the estimated odds of choosing category 3 over category 2, and of choosing category 4 over category 3, are $0.14\exp(\theta_i - \beta_j)$ and $0.74\exp(\theta_i - \beta_j)$ respectively.

The item parameters indicate that high scores on Items 1 and 5 are associated with extremely low values of personal well-being since these items' parameter estimates are especially high. Frequently feeling depressed and lonely therefore are strong indicators of very low levels of personal well-being. On the opposite end of the scale, Items 9 and 11 set those with extremely high personal well-being apart from others. Hence frequently feeling energetic, calm and peaceful contributes strongly to high levels of personal well-being.

The distribution of the estimated θ are summarized in Figure 4.9. There are 41 persons who obtained the minimum score across all items and hence received large negative scores on the latent trait – these are the persons with the highest possible personal well-being. Theoretically, these persons would receive a θ estimate of $-\infty$, but the penalty term in the log-likelihood has enabled a finite estimate to be found. However, this estimate will depend wholly on the penalty used. No person attained the maximum possible score.

4.6 Conclusions

In this paper we develop a novel algorithm for JML estimation of the RSM. It is shown how iterative majorization can be used to construct an iterative least-squares algorithm for maximum likelihood estimation of essentially any model that can be written in terms of the multinomial logit transform, as outlined by (De Leeuw, 2011b). The method can therefore also be applied to e.g. the PCM, and the simple form of the parameter updates has the potential to enable restricted versions of these models to be fit with ease. This can be especially advantageous in the PCM, which contains many more parameters than the RSM, and we envisage investigating this in future work. Another strength of the algorithm is the ease with which missing responses can be handled.

We showed how a quadratic penalty can be included to penalize the size of the linear predictors in the model. It would also be possible to extend this to penalties on individual parameters, but together with the investigation of lasso-type penalties (Tibshirani, 1996), this is left for future work. A cross-validation procedure on the out-of-sample negative log-likelihood was shown to work well for selecting the value of the penalization parameter. Furthermore, a more refined approach which penalizes relative to the scaled value of the sufficient statistics for the person parameters was found to perform even better on a simulated example. This procedure also worked well in our empirical application. We note that in the presence of a penalty term we chose not to calculate standard errors for the parameter estimates since a penalty necessarily reduces the variance of estimates at the cost of a potential increase in the estimation bias. Hence reporting only standard errors would be likely to overstate the accuracy of estimation: an estimate of the induced bias would also be needed.

Computationally, the efficiency of the algorithm is greatly increased by the separability of the parameter updates for the three sets of parameters. This separability enables us the find the inverses of the relevant matrices explicitly, saving a substantial amount of computations. Similarly, the cross-product matrices can be shown to perform simple summations over three-way arrays, which negates the need for matrix products in the updates completely. We feel that JML estimation combined with regularization is an attractive alternative method to CML and MML estimation.

CHAPTER 5

Competing for the Same Value Segments: Insight into the Volatile Dutch Political Landscape

5.1 Introduction

Recent years have seen substantial changes to political landscapes in many countries (Van der Meer et al., 2015), including the Netherlands. Major political parties such as the CDA (Christian Democrats) have lost ground to new parties, most notably the LPF (Pim Fortuyn List) and the populist PVV (Party for Freedom), which launched in 2002 and 2006 respectively. Concurrently, the proportion of the electorate choosing to abstain from voting is on the rise, possibly due to a growing separation between voters and political parties as the latter fail to align their political agendas with voters' values.

Human values are deeply ingrained in a people's psyche and direct their behaviour, especially abstract behaviour such as voting in elections (Schwartz et al., 2010; Caprara et al., 2006; Barnea and Schwartz, 1998). Political parties appeal to political values such as safety and security, freedom of speech and concern for one's fellow man that closely resemble human values such as security, self-direction and universalism (Schwartz et al., 2013). It is therefore to be expected that people will be more likely to vote for the political party that best reflects their own values. Previous research has focused on relationships between human values and choosing whether to vote or abstain (Caprara et al., 2012), the left-right dimension in politics (Thorisdottir et al., 2007), and value preferences of people voting for specific political parties (Barnea and Schwartz, 1998).

This chapter was co-authored with Hester van Herk, Patrick J.F. Groenen and Joost van Rosmalen.

This study reports on an empirical investigation into the relationship between human values and political preference in the Netherlands. All political parties and values are considered simultaneously, using data collected for five waves of the European Social Survey. Schwartz (1992)'s Human Values Scale, comprising of 21 survey items measured on 6-point Likert (1932) scales, is used to assess individuals' values. Self-reported voting behaviour in national elections is also available. These data span the years 2002 until 2010, giving insight into concurrent changes in voting behaviour and human values. Whilst until now research has focused on values separately (Schwartz et al., 2010; Barnea and Schwartz, 1998), we interpret combinations of values that together influence voting behaviour. Such combinations of values are termed 'value segments' hereafter.

The established method of measuring human values is by using rating scales. This requires correcting for response styles (Schwartz, 2007), since comparing responses between different survey participants should account for the heterogeneous use of rating scales pervasive in such studies (e.g. Van Vaerenbergh and Thomas, 2013; Baumgartner and Steenkamp, 2001). Only after disentangling these response styles from substantive content can valid comparisons be made. However, few studies correct for response styles as it is difficult and content-related information might be removed in the process (Schoonees et al., 2015b). Recently, more advanced alternatives to the commonly used response style indices advocated by Baumgartner and Steenkamp (2001) have been developed. Specifically, we use the latent-class bilinear multinomial logit (LC-BML) model of Van Rosmalen et al. (2010), which allows us to estimate value segments while at the same time correcting for differences in response style. This correction results in more valid value segments.

The questions we pose in this study are as follows: How many content-based value segments are there in the Dutch population and how are these value segments related to voting behaviour? How stable is the relationship between the value segments and the political parties over time? What effect do individual differences in response style have on content-based value segments? We start by giving an overview of political dimensions and cultural value theory applied to Dutch politics, a short summary of important recent events affecting voting, and an overview of our research propositions. Then we will further describe the LC-BML model and present our results. We end with a discussion and show the relevance of our method to study political change in other multi-party political systems.

5.2 Dutch Politics and Political Orientation

Dutch politics have seen a number of important events during the first decade of the 21st century. Populist parties such as Liveable Netherlands (LN), the LPF and the PVV have emerged, placing great emphasis on issues surrounding ethnic groups, immigration, Islam and asylum seekers. Especially the PVV has successfully elevated immigration to a salient issue, whilst also harbouring strong anti-European sentiments. They are particularly concerned that any benefits likely to be brought on by the expansion of the European Union will be offset by problems associated with increased immigration, a belief also shared with the Socialist Party (SP). An overview of Dutch political parties, past and present, is given in Table 5.1.

Other events, such as the September 11 attacks in the United States in 2001, the assassination of the LPF leader Pim Fortuyn shortly before the general election in 2002 (May 6), and the murder of film director Theo van Gogh[1] (November 2, 2004) have strengthened the call for increased security across society. Fundamental shifts in the population's most pressing issues have occurred too (Aarts and Thomassen, 2008). Before 1998, unemployment was a key issue but it hardly featured as a factor for voters in the first decade of this century. The predominant issues were found by Aarts and Thomassen (2008) to be minorities and asylum, followed by health care, law and order as well as security. All of these issues feature prominently in the manifestos of the newly emerging populist parties.

A useful shorthand classification of the orientation of political parties is the customary distinction between left and right. In countries or regions with a two-party system, it can also be an accurate reflection of the political situation. However, in countries which have multi-party systems, such as the Netherlands, the left-right political spectrum is often too simplistic. Therefore a second dimension is sometimes added to distinguish between authoritarian and libertarian parties (Evans et al., 1996).

Aarts and Thomassen (2008) identify three dimensions: left versus right, authoritarian versus libertarian and religious versus secular. In their context, left (right) indicates opposition to (support for) differences in social equality between people and support for (opposition to) a strong role for government in society. Here the right is primarily opposed to the state having a strong influence on the economy: it considers private enterprise important and accepts inequality in society. Authoritarian versus libertarian primarily relates to dealing with people from other cultures and law enforcement. On the libertarian side, people are open to other cultures and non-conformist practices such as abortion or euthanasia, whereas authoritarians are more traditional, less open to foreign cultures and consider law enforcement to be important. Finally, the religion dimension concerns the role that the church should play in society on moral issues. Aarts and Thomassen (2008) show that the authoritarian versus libertarian dimension has become increasingly important in deciding which party to vote for between 1989 and 2006.

[1]Van Gogh, critical of aspects of Islam in his work, was murdered by radical Dutch Muslim Mohammed Bouyeri, sparking a number of retaliatory incidents.

Specific to the Netherlands, the religious factor distinguishes the strictly religious SGP (Reformed Political Party), the CU (Christian Union) and to a lesser extent the CDA (Christian Democrats) from the other parties. The libertarian versus authoritarian factor distinguishes the D66 (Democrats 1966) and the GL (Green Left) from the other parties, and from the liberal VVD (People's Party for Freedom and Democracy) in particular. The left-right factor distinguishes GL, the SP and to a lesser extent the PvdA (Labour Party) from the VVD and the new populist parties, such as the PVV. Each of these parties has political values that might be related to human values deemed important by the electorate (Schwartz et al., 2013).

For a long time there have been three main political parties: CDA, PvdA and VVD. Together these parties have enjoyed over 60% of the votes. Since 2008, other parties have become larger, such as the SP (9.8% in 2010) and the PVV (15%). The growth of these latter parties came at the expense of the three long-established parties. In the 2010 elections, five parties received about 80% of the votes; the political landscape became more diverse.

5.3 Human Values

Human values are abstract and context-free, in contrast to attitudes which are more closely related to various life domains. The seminal framework of Schwartz (1992) comprises ten fundamental value domains. These value domains can each be defined based on their central goal: power, achievement, hedonism, stimulation, self-direction, universalism, benevolence, tradition, conformity, and security (Schwartz and Rubel, 2005). Table 5.1 provides a brief summary of these domains.

Schwartz represented these ten value domains in a circular structure – see Figure 5.1. Domains with adjoining positions are closely related while opposing positions indicate incompatible values. For example, power and achievement are complementary but generally incompatible with universalism and benevolence. Hence the theory predicts that persons who attach high value to universalism and benevolence will find power and achievement much less important, and vice versa. A further example of incongruent values are conformity and stimulation; however, stimulation is compatible with the adjacent self-direction because people who enjoy challenges are also more likely to be creative and investigative.

These ten individual value domains can be merged into four different higher-order domains (Schwartz, 1992), namely self-enhancement, self-transcendence, conservation and openness-to-change. This is also depicted in Figure 5.1. Fontaine et al. (2008) further distinguished person-focused (self-enhancement and openness-to-change), socially-oriented (self-transcendence and conservation), protection (power and security) and growth (self-direction and universalism). For example, growth is associated with independent thought and action, and people who find these

Abbr.	Name	Left–Right	Authoritarian–Libertarian–Religious	Description
VVD	Party for Freedom and Democracy	Right (Conservatives)	Authoritarian	Right-wing liberal party emphasizing freedom, self-initiative and small government. Also harbours progressive views on ethical matters, for example.
PvdA	Labour Party	Left		Progressive, traditional social democratic party which is in favour of an active government.
PVV	Party for Freedom	New right		Nationalistic populist party with both conservative, liberal, rightist and leftist views. Against immigrants, Islam, European integration and broad leftist thinking in general, such as subsidies and development.
CDA	Christian Democrats	Centre	Religious	A centrist party that unifies three Christian parties, two Protestant and one Catholic. Conservative with an emphasis on family values and taking care of the needy.
SP	Socialist Party	Left		Left-wing socialist and Eurosceptic protest party with an emphasis on blue collar workers.
D66	Democrats 1966		Libertarian	Reformist social liberal party with a social democratic touch.
GL	Green Left	Left	Libertarian	Small social democratic left-wing party placing emphasis on sustainability and nature. A fusion of four small left-wing part parties, including pacifist and communist parties.
CU	Christian Union		Religious	Christian democrats with stronger Christian values than the CDA and more social democratic policies for nonreligious items. Progressive views on social and ecological matters.
SGP	Reformed Political Party		Religious (strict)	Small right-wing conservative Protestant Christian party with a very stable electorate.
PvdD	Party for the Animals	New left	Libertarian	Small social democratic party with emphasis on animal welfare, animal rights and sustainability.
TON	Proud of the Netherlands	New right	Authoritarian	Populist right-wing conservative party.
LPF	Pim Fortuyn List	New right	Authoritarian	Former populist right-wing conservative party emphasizing problems with immigration.
LN	Liveable Netherlands	New right	Authoritarian	Former small populist national party.

Table 5.1: Explanation of current and past Dutch political parties and their abbreviations, based on the authors' own views and those expressed by the political scientists at www.parlement.com (Dutch only). Current parties are ordered from largest to smallest with respect to the proportion of votes received in the national election of 2010. We indicate the political orientations of these parties where there is a clear classification.

Value	Code	Description
Power	PO	Social status and prestige, control or dominance over people and things.
Achievement	AC	Demonstrating personal success by displaying competences considered to be socially valuable.
Hedonism	HE	Enjoyment and sensual self-reward.
Stimulation	ST	Doing exciting, new and challenging things.
Self-direction	SD	Thinking independently and opting for action, being creative and investigative.
Universalism	UN	Showing understanding, and appreciating, tolerating and protecting all people and the natural world.
Benevolence	BE	Protecting and improving the well-being of people with whom one has frequent personal contact.
Tradition	TR	Respect for, involvement in and acceptance of ideas that traditional culture or religion offers people.
Conformity	CO	Refraining from action, tendencies and impulses that can disrupt or hurt others and which conflict with social standards and expectations.
Security	SE	Safety, security, harmony and stability of the community, relationships and one's self.

Table 5.1: An overview of the ten fundamental value domains of Schwartz and Rubel (2005).

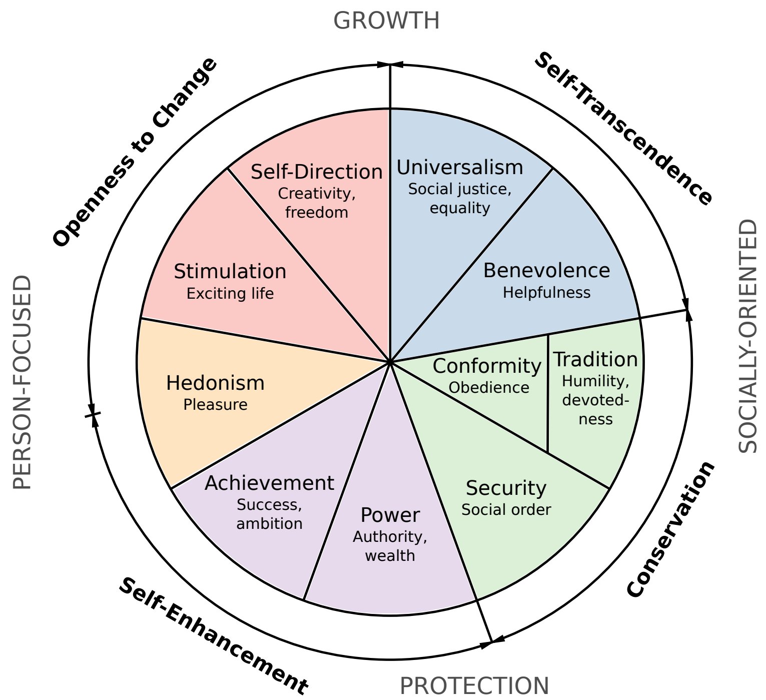


Figure 5.1: Schwartz's value circumplex.

values important will therefore accept a diverse society. Those who emphasize protection will in contrast aim to maintain the power that they have and expect good security.

Various studies have used individual values or their higher-order domains to predict attitudes and behaviour. Those who score high on openness-to-change were found to be more innovative (Steenkamp et al., 1999), less religious (Schwartz and Huismans, 1995) and more likely to vote left (Thorisdottir et al., 2007). The opposite applies to people who score high on conservation. Those who score high on achievement and power are more materialistic (Burroughs and Rindfleisch, 2002) and those who score high on self-transcendence are generally more environmentally aware (Schultz et al., 2005). Human values have also been used to link specific political parties to individual values (Barnea and Schwartz, 1998; Caprara et al., 2006; Schwartz et al., 2010). Conservative values were found to be more important to those who voted for right-wing parties, and universalism more important to those who voted left.

Although benevolence has been shown to be the most important value in general, the relative importance of values differ among people (Schwartz and Bardi, 2001). This hierarchy may depend on demographic characteristics and personal circumstance. In addition, some people draw a firm distinction between values – finding some to be very important while explicitly rejecting others – whereas others consider most values to be more or less equally important. These differences result in several distinct groups of people sharing the same value preferences. In this research, we aim to identify segments of people with such similar value profiles and to explore the role these profiles play in determining which political party an individual votes for. We also provide insight into the value profiles of people who did not vote. To find valid segments, a correction for differences in rating scale use will be made (Van Vaerenbergh and Thomas, 2013).

5.4 Values and Voting

Values are linked to political orientation and determine the political choices that individuals make (Barnea and Schwartz, 1998; Thorisdottir et al., 2007). People tend to vote for the party which best reflects their own values. While group values are relatively stable over time (Schwartz, 2006), the emergence and disappearance of political parties in the Netherlands in recent years suggest that for some groups of voters their best-fitting political parties have changed. Most people aim to achieve congruence between their values and attitudes on the one hand and behaviour on the other (Edwards and Cable, 2009).

Left-wing parties emphasize equality, tolerance and a strong governmental role. These first two issues are congruent with the human values universalism and self-direction (Schwartz et al., 2010) and, for more traditional left-wing parties, environment-related security (Thorisdottir et al., 2007). We therefore expect that a segment considering these values relatively more important than other segments will vote for left-wing parties such as the PvdA and to a lesser extent GL; voters of SP are expected to emphasize security values more than the other left-wing parties. In contrast, right-wing parties consider inequality in society, maintenance of the status quo, and economic freedom important. These issues are associated with the values power, achievement and to a lesser extent self-direction (Thorisdottir et al., 2007; Caprara et al., 2006). We expect that a segment prioritising these values will vote for a right-wing party such as the VVD.

Libertarian parties emphasize equal opportunities and taking initiative, which is compatible with universalism, self-direction and stimulation but incompatible with security and conformity. People who consider the values universalism and self-direction and to a lesser extent benevolence essential have a positive stance toward immigration (Davidov et al., 2008; Schwartz et al., 2010); accepting immigrants is specifically associated with libertarian parties. We therefore expect that there is a segment considering benevolence, universalism, self-direction and stimulation of particular importance, which predominantly votes for parties such as GL, D66 and the PvdD. Authoritarian parties emphasize maintenance of the status quo, inequality in society and strict rules, which relates to power, security and tradition but not to universalism and self-direction. Such a segment would be anticipated to vote for parties such as the conservative VVD and the populist LPF and PVV. The latter parties also emphasize the negative side of immigration, which is compatible with security, tradition and conformity. Finally, religious parties consider the role of the church in society of key importance, relating to traditional (gender) roles, conformity to rules, and taking care of people. Religiousness is therefore related to human values in the conservation domain (Schwartz and Huismans, 1995). Specifically, it is expected that there is a segment highlighting conformity, tradition, and security as highly important and self-direction, stimulation, and hedonism as unimportant. This segment would tend to vote for the religious parties SGP, CU and to a lesser extent the CDA. They would consider benevolence to be more important than universalism as benevolence emphasizes the traditional in-group more.

Summarizing, we expect to find a small number of segments differing in value importance, which we link to voting behaviour. We expect this relation to exist for people with clear value profiles; however, the value profiles are derived solely from the data and therefore the segments are not guaranteed to adhere to the Schwartz theory of compatibilities and incompatibilities among values.

5.5 The LC-BML Model

To determine the value segments free from response styles, we use the latent-class bilinear multinomial logit model (LC-BML model; Van Rosmalen et al., 2010). This model is a multivariate generalization of the standard multinomial (or baseline-category) logit model for multinomial responses (e.g. Agresti, 2002), which in turn is the multicategory extension of logistic regression. The multinomial logit model describes the probability that a respondent answers a single item with a given rating. The LC-BML model combines such models for multiple items in order to describe the responses of a respondent on all items simultaneously.

Furthermore, the model allows for respondents to be segmented into two types of segments (or latent classes), namely value and response style segments. This is done by allowing the model parameters to vary between segments. The response style segments correct for the differences in response styles, while the value segments indicate differences in the importance attached to the respective values. Hence each person is assigned simultaneously to both a response style segment and a value segment by the LC-BML model. Another feature is the use of a bilinear decomposition of the model parameters so that the effects can be summarized parsimoniously with the help of biplots (e.g. Gower et al., 2011).

Describing the LC-BML model in more detail requires some notation. Let Y_{ij} be the random variable denoting the response of respondent i $(i = 1, \ldots, N)$ to item j. Index by k $(k = 1, \ldots, K)$ the common rating scale used for all J items. Missing responses are included by adding a dedicated 'Missing' category to the rating scale. Let there be R and S latent classes for the response style and value segments respectively. Supposing that π_{rs} is the prior probability that any respondent belongs to segments r and s, it follows that

$$\mathsf{P}(Y_{ij} = k) = \sum_{r=1}^{R} \sum_{s=1}^{S} \pi_{rs} \mathsf{P}(Y_{ij} = k \mid r, s). \tag{5.1}$$

In the LC-BML model, the segment-specific probabilities follow multinomial logit models such that

$$\mathsf{P}(Y_{ij} = k \mid r, s) = \frac{\exp\left(\eta_{ijk \mid r,s}\right)}{\sum_{k=1}^{K} \exp\left(\eta_{ijk \mid r,s}\right)}, \tag{5.2}$$

where $\eta_{ijk \mid r,s}$ is a segment-specific linear predictor. The basic form of the linear predictor is

$$\eta_{ijk \mid r,s} = \alpha_{k \mid r} + \sum_{l=1}^{L} \boldsymbol{\beta}'_{kl} \boldsymbol{x}_{il} + \gamma_{jk \mid s}. \tag{5.3}$$

Here $\alpha_{k \mid r}$ is the attractiveness of rating k in response style segment r, $\boldsymbol{x}_{il}$ is an indicator vector indicating which category of the (discretized) socio-demographic variable l $(l = 1, \ldots, L)$ person i belongs to, $\boldsymbol{\beta}'_{kl}$ is the transpose of the vector of effects for this socio-demographic variable on category k, and $\gamma_{jk \mid s}$ is the effect of rating k on item j in value segment s. Finally, a bilinear decomposition is applied to the parameters in (5.3), a description of which is deferred to the

Appendix. This decomposition serves two purposes, namely to reduce the number of parameters to be estimated and to allows for all effects to be interpreted graphically in biplots. Note that the dimensionality P of this decomposition determines the dimensionality of these biplots.

For a given choice of the number of segments R and S, and dimensionality P, the LC-BML model is estimated by maximum likelihood via the Expectation-Maximization (EM; Dempster et al., 1977) algorithm. The likelihood contribution of person i is given by

$$\sum_{r=1}^{R}\sum_{s=1}^{S}\pi_{rs}\prod_{j=1}^{J}\prod_{k=1}^{K}\mathsf{P}(Y_{ij}=k\,|\,r,s)^{I(y_{ij}=k)}, \tag{5.4}$$

with y_{ij} being the realized value of Y_{ij} and $I(\cdot)$ the indicator function. An important by-product of the estimation algorithm is the estimated posterior probabilities of each person belonging to each of the $R \times S$ latent classes, which can be calculated as

$$\pi_{rs}(i) = \frac{\pi_{rs}\prod_{j=1}^{J}\prod_{k=1}^{K}\mathsf{P}(Y_{ij}=k\,|\,r,s)^{I(y_{ij}=k)}}{\sum_{r=1}^{R}\sum_{s=1}^{S}\pi_{rs}\prod_{j=1}^{J}\prod_{k=1}^{K}\mathsf{P}(Y_{ij}=k\,|\,r,s)^{I(y_{ij}=k)}}. \tag{5.5}$$

This gives a posterior measure of class membership for all persons. These can for instance be aggregated over the respondents' self-reported voting behaviour to establish how people within each segment voted. Information criteria, such as the Akaike Information Criterion (AIC; Akaike, 1974) or Schwarz's Bayesian Information Criterion (BIC; Schwarz, 1978) can be used to select the number of segments R and S, as well as the dimensionality P. An alternative is the graphical CHull procedure of Ceulemans et al. (2011). We opt for a combination of the BIC and the CHull procedure. The BIC has been shown to work well in conjunction with the LC-BML model (Van Rosmalen et al., 2010).

5.6 European Social Survey Data

This study focuses on five rounds of the European Social Survey (ESS), namely those of 2002, 2004, 2006, 2008 and 2010. In each round, the ESS included the 21 PVQ-based value items (Portrait Values Questionnaire; Schwartz and Bardi, 2001; Schwartz, 2007). These statements are gender-specific and operationalized by posing statements such as 'Thinking up new ideas and being creative is important for her. She likes to do things in her own original way.' The responses to each statement were based on a 6-point Likert scale, ranging from 1 ('Very much like me') to 6 ('Not like me at all'). More information and average ratings can be found in Table 5.1. We control for the socio-demographics Gender, Education Years and Age. Since discrete covariates are required for the LC-BML model, both Education Years and Age were discretized into three categories. For years of education, the categories were low (1 to 10 years; comparable to ISCED[2] levels 1 and 2), intermediate (11 to 15 years; ISCED levels 3 – 5) and high (16+ years; ISCED level 6 or higher). Age was divided into the categories 15 to 34 years, 35 to 59 years and 60 to 96 years.

A total of 9607 respondents from the Netherlands were available across the 5 ESS waves. Respondents with missing values for the abovementioned socio-demographic variables were removed from the analysis (85 persons), together with respondents who used the same rating to answer all 21 value items (11 persons). After removing these respondents, 9511 observations were used in the analysis. The numbers of respondents per ESS wave were 2323, 1854, 1845, 1713 and 1776 respectively. Note that although sampling weights are applicable in the ESS, it was not possible to apply them in the LC-BML analysis due to software limitations. We do however apply post-stratification weights in summary statistics, which corrects for the sampling design and unit nonresponse.

In the ESS, respondents were also asked to indicate whether they voted in the most recent elections for the Second Chamber of Parliament[3], and, if so, for which political party. For the 2002 survey, this concerned the Dutch elections of 6 May 2002, for 2004 the elections of 22 January 2003, for both 2006[4] and 2008 the elections of 22 November 2006 and for 2010 the elections of 9 June 2010. The official election results are given in Table A1 in the Appendix.

Voting is not mandatory in the Netherlands. In recent elections, roughly 20% to 25% of the electorate chose not to vote. The weighted proportion of eligible respondents in our sample who indicated that they voted is slightly higher at 86.4%, 82.4%, 83.3%, 86.1% and 84.4% for the 5 ESS waves respectively. We include an explicit group for persons who chose not to vote in our analyses of the reported voting behaviour. Besides these individuals, approximately 7.1% of the respondents were not eligible to vote, mostly on account of being younger than 18 at the time of the election. Moreover, the ESS asked respondents which political parties they voted for. The disclosure rates for those who indicated that they voted were quite high at 97.9%, 96.6%, 95.1%, 96.4% and 96.0% respectively.

Since the results of the relevant elections are known at the population level (see Table A1), we use these to recalibrate the post-stratification weights from the ESS before analyzing the voting behaviour. These recalibrated weights are constructed so that the proportion of votes for all competing parties as well as the proportion of nonvoters in the observed sample match

[2]See http://www.uis.unesco.org/Education/Documents/isced-2011-en.pdf.

[3]This is the main legislative body of the Netherlands.

[4]Data collection for the third ESS wave started in September 2006, before the election took place in November. Respondents interviewed before the election were asked who they intended to vote for. After the election they were asked who they actually voted for. Roughly half the sample were interviewed before the election. In the questionnaire, LN was still given as a voting option. However, LN disbanded before the 2006 elections, hence we do not consider LN when interpreting voting behaviour after 2004.

the official election results as closely as possible. Iterative proportional fitting, also known as raking, was used to construct these weights separately for each wave of the ESS (see Chapter 8 of Lohr, 1999, for example). These weights are subsequently used whenever voting behaviour is considered. This recalibration procedure also adjusts for the respondents not eligible to vote in the elections, as well as the small proportion who did vote but either chose not to disclose for which party they voted, or could not recall which party they supported. This is done by increasing the weights of the remaining respondents such that the total weight for the eligible voters equals that of the entire sample.

5.7 Results

A mean hierarchy of human values often exists in society. Schwartz and Bardi (2001) identified pan-cultural norms, in other words norms that apply generally in every society. According to their findings, people across the world consider values such as benevolence, self-direction and universalism to be the most important; power, tradition and stimulation are considered the least important. This is also true in the Netherlands, as the weighted mean ratings reported in Table 5.1 shows. Dutch people consider values such as self-direction (SD1; average score 2.11), universalism (UN1; 2.11) and benevolence (BE2; 2.16) to be important. Power (PO2; 3.40) and stimulation (ST2; 3.72) are considered to be relatively unimportant. The least important is power (PO1; 4.18). This ordering of the average responses is stable over time: the minimum Spearman rank correlation of the average rating score per item between the various ESS rounds is 0.98. We use the LC-BML model to find more refined subsets of respondents who exhibit different value hierarchies.

5.7.1 Model Selection

In order to select the most appropriate model, we fitted the LC-BML model for $R = 1,2,\ldots,20$ response style segments, $S = 1,2,\ldots,12$ value segments, and $P = 1,2$ dimensions. In total 480 different models were considered; however 9 models did not converge in the allotted number of EM iterations (10 000) and were discarded. A numerical convergence criterion of 10^{-5} was used. The EM algorithm is however only guaranteed to find a local optimum of the likelihood function. We therefore estimated each of these models for 20 different random starts to increase our chances of finding the global optimum. Only the start which resulted in the highest value of the likelihood function is retained.

We plotted the maximized log-likelihood values against the model degrees-of-freedom for the 471 models, as in Figure 5.1. The convex hull enclosing the cloud of points can then be

Value	Code	Item	Description	Average Rating
Benevolence	BE1	It's very important to him to help the people around him. He wants to care for their well-being.	Help others	2.27
	BE2	It is important to her to be loyal to her friends. She wants to devote herself to people close to her.	Loyalty	2.16
Universalism	UN1	He thinks it is important that every person in the world should be treated equally. He believes everyone should have equal opportunities in life.	Equality	2.11
	UN2	It is important to him to listen to people who are different from him. Even when he disagrees with them, he still wants to understand them.	Understand others	2.43
	UN3	She strongly believes that people should care for nature. Looking after the environment is important to her.	Care for nature	2.31
Self-Direction	SD1	It is important to her to make her own decisions about what she does. She likes to be free and not depend on others.	Independent	2.11
	SD2	Thinking up new ideas and being creative is important to her. She likes to do things in her own original way.	Creative	2.46
Stimulation	ST1	He likes surprises and is always looking for new things to do. He thinks it is important to do lots of different things in life.	Look for new things	2.85
	ST2	He looks for adventures and likes to take risks. He wants to have an exciting life.	Excitement	3.72
Hedonism	HE1	Having a good time is important to her. She likes to spoil herself.	Have a good time	3.04
	HE2	He seeks every chance he can to have fun. It is important to him to do things that give him pleasure.	Have fun	2.52
Achievement	AC1	It's important to her to show her abilities. She wants people to admire what she does.	Be admired	3.27
	AC2	Being very successful is important to him. He hopes people will recognise his achievements.	Be successful	3.26
Power	PO1	It is important to her to be rich. She wants to have a lot of money and expensive things.	Be rich	4.18
	PO2	It is important to her to get respect from others. She wants people to do what she says.	Get respect	3.40
Security	SE1	It is important to him to live in secure surroundings. He avoids anything that might endanger his safety.	Security	2.80
	SE2	It is important to her that the government ensures her safety against all threats. She wants the state to be strong so it can defend its citizens.	Strong government	2.73
Conformity	CO1	He believes that people should do what they're told. He thinks people should follow rules at all times, even when no-one is watching.	Follow rules	2.95
	CO2	It is important to her always to behave properly. She wants to avoid doing anything people would say is wrong.	Behave properly	2.85
Tradition	TR1	Tradition is important to her. She tries to follow the customs handed down by her religion or her family.	Tradition	2.86
	TR2	It is important to him to be humble and modest. He tries not to draw attention to himself.	Modesty	3.29

Table 5.1: The 21 value items from the ESS and their weighted average ratings for the Dutch sample. The unadjusted post-stratification weights supplied in the ESS were used to calculate the weighted averages.

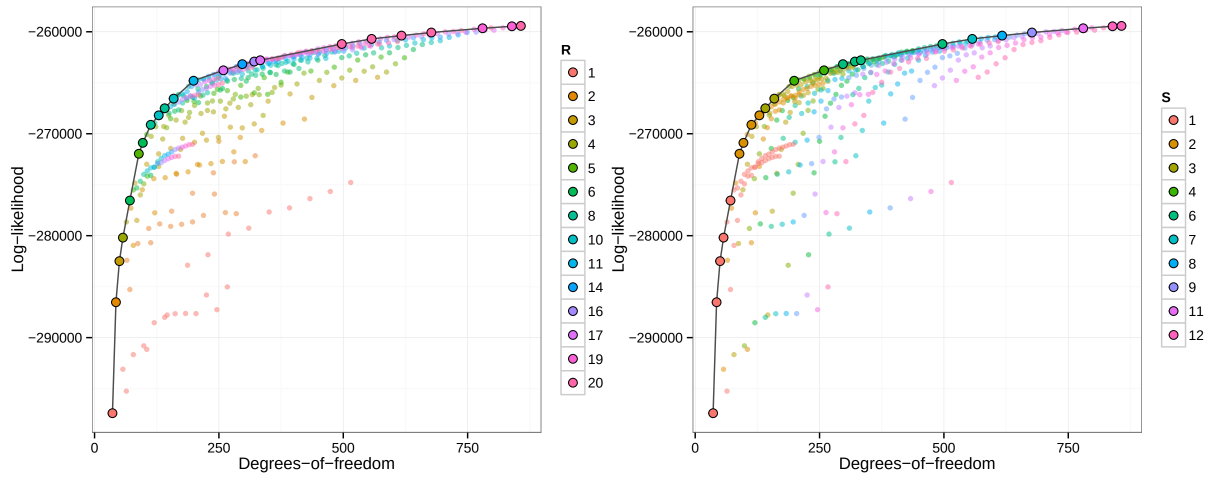


Figure 5.1: The maximized log-likelihood plotted against degrees-of-freedom for the models considered. Colours are assigned according to the number of response styles R and item segments S in the left and right panels respectively. The convex hull enclosing the cloud of points is also shown.

determined. We reduced the model selection problem by considering only the models that lie on this hull, similar to the CHull procedure of Ceulemans et al. (2011). These models can be considered to present a good trade-off between complexity and data fit. The BIC values for the 23 models that lie on this hull are summarized in Table 5.2. Although it has only the fourth lowest BIC value, we selected the model with $R = 20$ response styles, $S = 7$ value segments and $P = 2$ dimensions, since we found this model to be more interpretable than the more complex models with slightly better BIC values. We now proceed to describe and interpret the results from this model.

5.7.2 Value Segments

As shown in Table 5.3, the sizes of the value segments are relatively stable over time. This is expected as values at a higher aggregation level hardly change (Schwartz, 2006). Segment 4 has grown somewhat over the years from 12.8% to 16.3%, whilst Segments 3 and 5 have decreased in size from 16.1% and 16.0% to 12.3% and 13.2% respectively. Overall, Segment 7 is estimated to contain only 1.1% of the population, while the other segments range in size from 14.3% to 22.0%. Table 5.4 shows the distribution of the demographic variables across the segments. These distributions are calculated by summing the post-stratification weights of all respondents who fall in a specific category (e.g. all males). Segment membership is incorporated by weighting the post-stratification weights with the estimates of each respondent's posterior segment-membership probabilities in (5.5). The segments differ with respect to socio-demographic characteristics.

R	S	P	**DF**	**BIC**	R	S	P	**DF**	**BIC**
20	9	2	677	526 346	10	3	1	159	534 569
20	8	2	617	526 402	8	3	1	141	536 296
19	11	2	780	526 459	10	2	1	129	537 576
20	**7**	**2**	**557**	**526 501**	8	2	1	113	539 285
19	12	2	839	526 607	6	2	1	97	542 677
20	12	2	857	526 699	5	2	1	89	544 748
20	6	2	497	526 948	6	1	1	71	553 751
17	6	1	333	528 639	4	1	1	57	560 906
16	6	1	321	528 782	3	1	1	50	565 456
14	6	1	297	529 069	2	1	1	43	573 453
17	4	1	259	529 938	1	1	1	36	595 156
11	4	1	199	531 425					

Table 5.2: A summary of the 23 models lying on the convex hull considered in the final model selection step. Given are the number of segments R and S and dimensionality P, the number of degrees-of-freedom (DF), and the respective BIC values. The models are ordered from low to high BIC. The selected model is shown in boldface.

	ESS Wave					
Segment	**2002**	**2004**	**2006**	**2008**	**2010**	**Overall**
1: Mainstream	21.4	22.2	22.5	21.2	22.7	22.0
2: Security Seekers	18.3	18.8	18.2	18.5	19.3	18.6
3: Traditionalists	16.1	15.2	13.6	14.9	12.3	14.5
4: Universalists	12.8	13.1	16.8	16.8	16.3	15.0
5: Indifferent	16.0	15.7	13.2	14.1	13.2	14.5
6: Entrepreneurs	14.3	13.9	15.0	13.4	14.7	14.3
7: Uninterested	1.1	1.1	0.8	1.1	1.6	1.1

Table 5.3: Segment sizes per year (in percentage), using unadjusted post-stratification weights.

	Gender		Education Years			Age			
Segment	**Male**	**Female**	**1–10**	**11–15**	**16–30**	**15–34**	**35–59**	**60–96**	**Size**
1: Mainstream	38.9	61.1	28.4	50.1	21.5	18.6	52.0	29.5	22.0
2: Security Seekers	52.4	47.6	22.2	48.9	28.8	30.2	50.0	19.8	18.6
3: Traditionalists	43.7	56.3	40.6	44.7	14.6	12.8	45.3	42.0	14.5
4: Universalists	49.3	50.7	14.0	47.9	38.1	34.2	52.2	13.6	15.0
5: Indifferent	53.4	46.6	35.2	46.6	18.3	33.9	41.7	24.4	14.5
6: Entrepreneurs	63.3	36.7	15.1	51.6	33.3	62.0	33.6	4.4	14.3
7: Uninterested	42.7	57.3	29.7	53.9	16.4	26.3	41.1	32.6	1.1
Average	49.1	50.9	26.5	49.1	24.5	31.1	45.1	23.8	100.0

Table 5.4: The distribution of the socio-demographic variables across value segments, and the segment sizes. All values are expressed as percentages per row and variable, except for the segment sizes. Post-stratification weights are applied.

Segment 3 has a large group of people who are lower educated females, aged primarily above 60. In Segment 6, the majority is highly educated young males, while Segment 4 includes higher educated people, both male and female, who are mostly aged 35–59. Segment 5 includes more people with low education levels than average.

These seven value segments are based on different value priorities. To help interpret these priorities, all the response categories and values are represented jointly in a two-dimensional space using biplots (see Figures 5.2 – 5.4). The positions of the value items vary across all value segments, but the locations of the response categories remain the same. The magnitude of the effect of each item on the probability to endorse a specific rating category is determined by the inner product (also known as the scalar or dot product) between the position of the item in that segment and the vector representing the rating. This can be determined by multiplying the length of the projection of the item point onto the rating vector by the length of the rating vector. Simply put, items which have large positive projections are associated with large probabilities of endorsing that rating, and vice versa. For example, in Segment 1 the item PO1 has a very large positive projection on rating category 6 ("Not like me at all"), which implies that in Segment 1 people do not value this item highly. In contrast, UN1 has a large positive projection on rating category 1 ("Very much like me"), implying that this segment values this aspect of universalism most highly.

We first interpret each value segment, and then relate these to voting behaviour in the next section. To ease interpretation, names are assigned to the segments:

Segment 1: Mainstream. The 'Mainstream' group forms the largest segment comprising roughly 22.0% of the sample. From Figure 5.2, we see that this segment attaches great importance to values universalism (UN1: equal treatment and opportunities for everyone; and UN2: understanding others), benevolence (BE2: loyalty and devotion to friends and family; and BE1: helping others and caring about their well-being), and self-direction (SD1: being independent). Values such as power (PO1: being rich; and PO2: being respected by others), achievement (AC2: being successful; AC1: and being admired) as well as stimulation (ST2: living an exciting life), are considered to be completely unimportant. This is mostly in line with the pan-cultural hierarchy of Schwartz and Bardi (2001). We might consider this segment to represent the modal person in the Netherlands.

Segment 2: Security Seekers. The 'Security Seekers' (18.6%) do not have any pronounced preference for or aversion to any particular values. They consider all values to be of approximately equal importance. However, in comparison to the other segments, they consider security (SE1 and SE2) relatively more important. Further, benevolence (BE2: being loyal to friends and family) and self-direction (SD1: independence) are considered important. As in most other segments, they consider power (PO1: being rich) and stimulation (ST2: living an exciting life) to be relatively unimportant. Interestingly, tradition (TR2: modesty) is also considered to be relatively unimportant.

Segment 3: Traditionalists. Compared to the other segments, the 'Traditionalists' (14.5%) emphasize the conservation values (security, conformity and tradition) far more than the other segments. People in this segment consider it to be especially important that the area in which they live is safe (SE1: security) and that traditions are respected (TR1: tradition). Contrary to many of the other segments, they consider the item gaining new experiences (ST1: stimulation) to be of only minor importance. The key motivation for people in this segment are conservation and socially-oriented values.

Segment 4: Universalists. The fourth value segment, the 'Universalists' (15.0%), places relatively great emphasis on self-direction (SD1: independence; and SD2: creativity), universal-

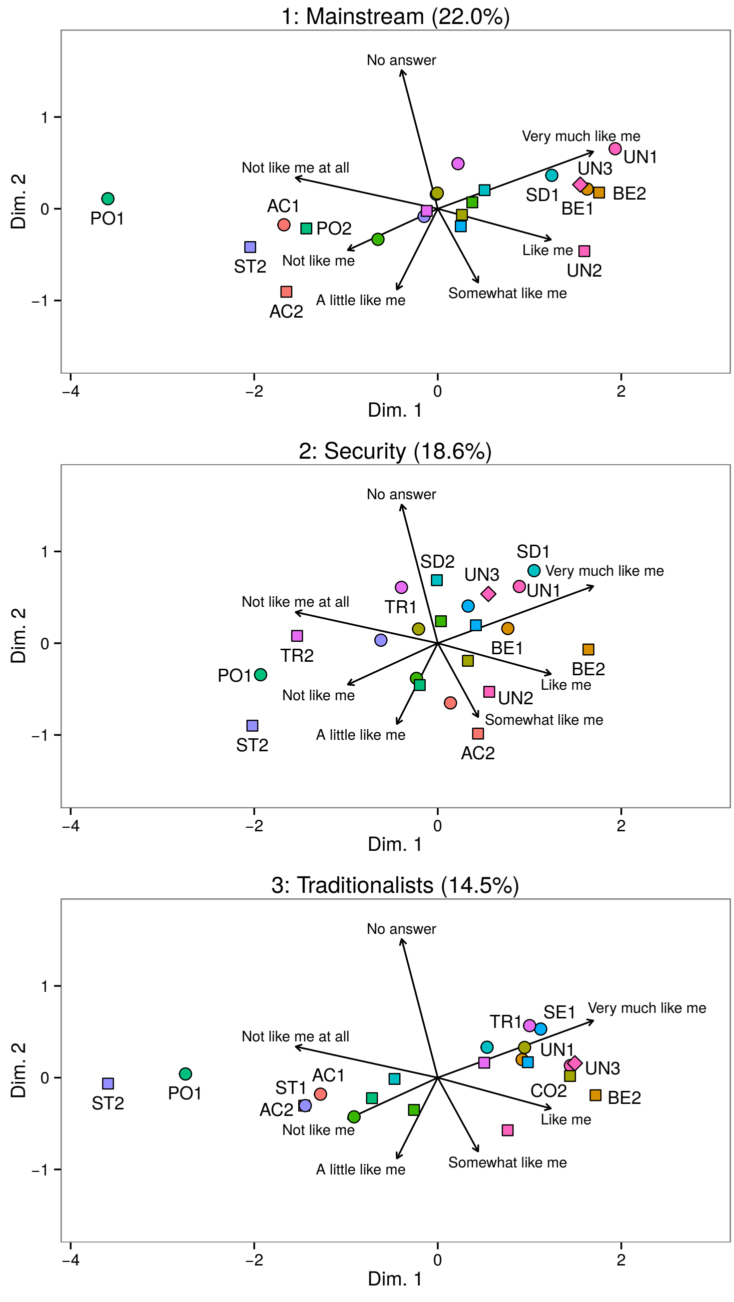


Figure 5.2: Biplots for value segments 1 – 3. The same colours and symbols apply in all plots, as explained in the legend in Figure 5.4.

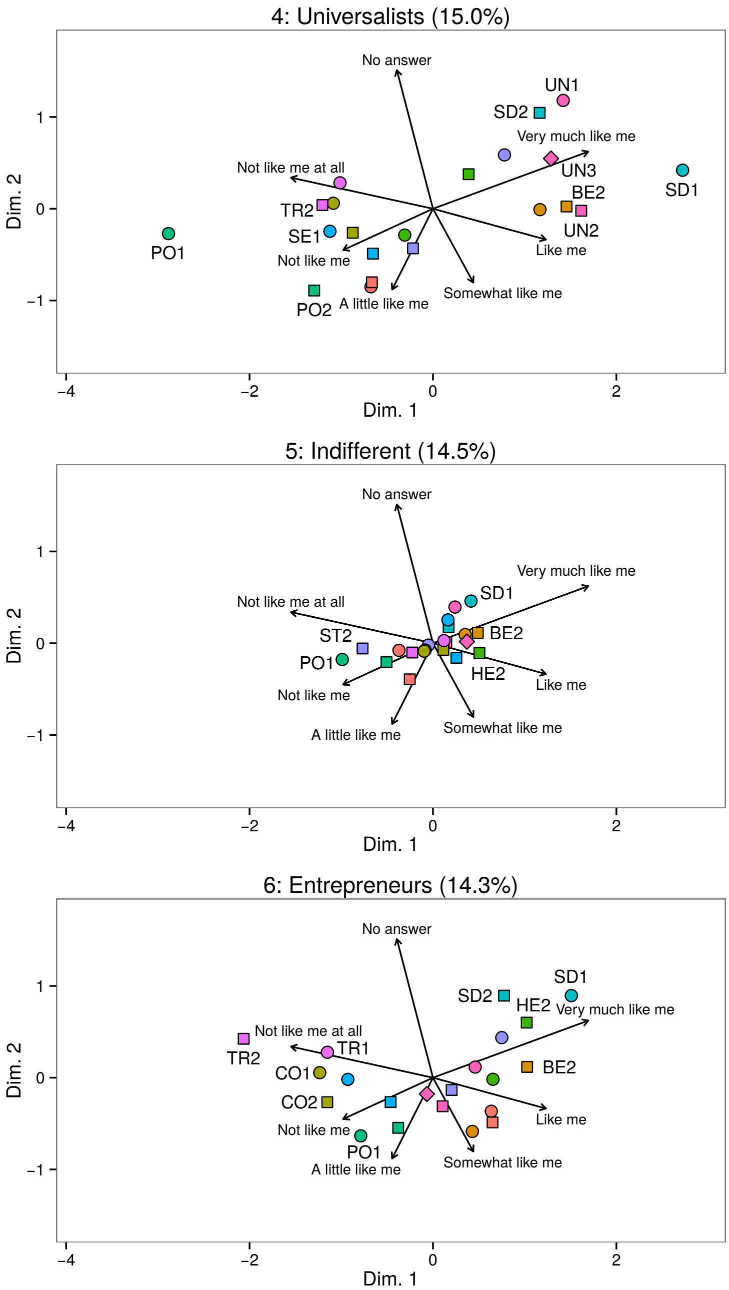


Figure 5.3: Biplots for value segments 4 – 6. The same colours and symbols apply in all plots, as explained in the legend in Figure 5.4.

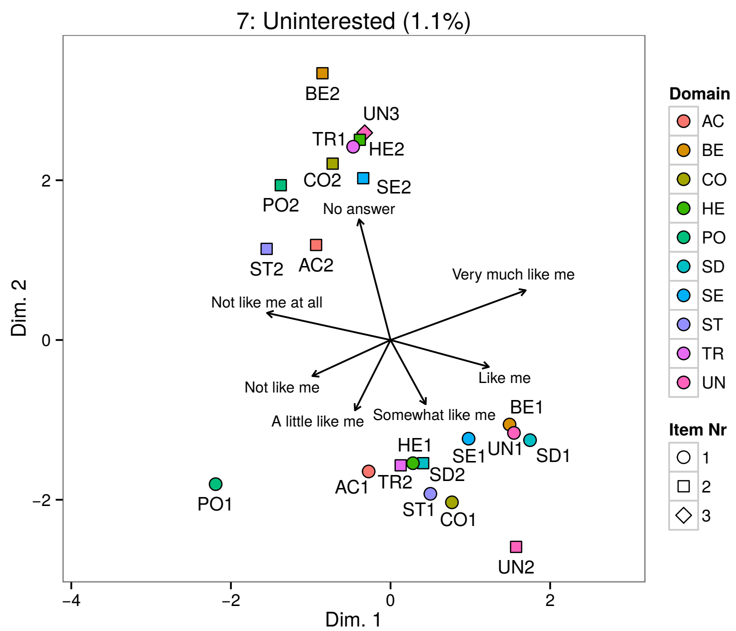


Figure 5.4: Biplot for value segment 7.

ism (UN1 to UN3), and also benevolence (BE2: loyalty to friends and family). Values in the conservation domain such as tradition (TR) and conformity (CO) are not considered important. Values representing growth are the motivation of people in this segment.

Segment 5: Indifferent. This segment, simply named 'Indifferent,' consists of individuals who hardly differentiated between the values. They comprise roughly 14.5% of the sample. All values are considered of average importance. Self-direction (SD1: being independent) and benevolence (BE2: being loyal to family and friends) are the values considered relatively most important to them.

Segment 6: Entrepreneurs. The 'Entrepreneurs' (14.3%) consider the values self-direction (SD1 and SD2), benevolence (BE2: loyalty to friends and family), and hedonism (HE2: having fun) to be extremely important. Universalism is considered relatively less important. Relative to the other segments, power and achievement are more important, while tradition and conformity are considered less important. Person-focused values are the key motivation for the people in this segment.

Segment 7: Uninterested. The smallest segment consists of 1.1% of individuals who are seemingly not interested in completing the entire value scale. This may have been an issue with the data collection or recording. In the ESS questionnaire, the values were divided over two pages and the respondents might have missed the second page. As is evident from the biplot in Figure 5.4, roughly half the items have been completed (the items in the lower half of the plot), while responses to the remaining nine items are mostly missing. This corresponds to the order in which the items appeared in the questionnaire.

5.7.3 Value Segments and Political Parties

To relate the value segments to voting over time we performed a correspondence analysis (CA; e.g. Hoffman and Franke, 1986; Greenacre, 2007). CA can be used to plot the value segments together with the political parties for which they voted. This makes it easier to interpret our results. For this analysis, a cross-tabulation of the individual posterior probabilities in Equation (5.5) was made with the value segments in the columns and the political parties, split into the five waves of the ESS, in the rows. The CA was used to visualize the links between the rows and columns in this table simultaneously (see Figure 5.5). The total inertia in the table can be effectively shown in two dimensions (78.6% of the inertia is explained in two dimensions).

Figure 5.5 shows the column-principal map of the CA (Greenacre, 2007), where each of the political parties are labelled. In addition, each year is indicated by a different symbol. Symbols for the same party are connected by lines. The black triangles represent the seven value segments. The darker the points, the better the explained inertia (fit) of the points – see Greenacre (2007)

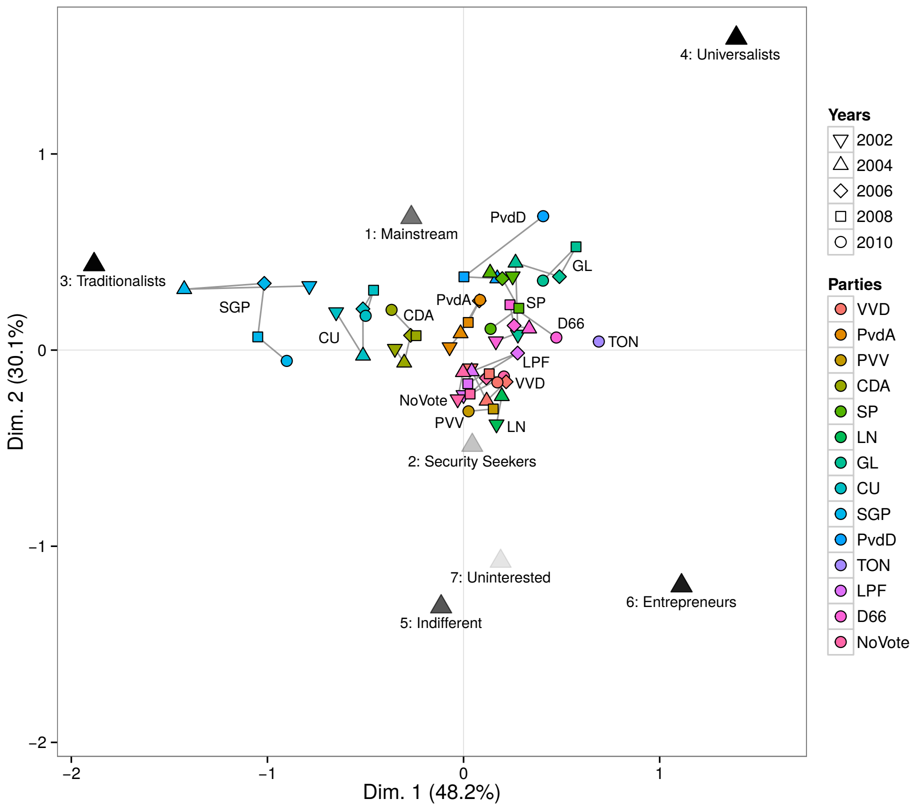


Figure 5.5: Correspondence analysis for the segments and reported votes, using the adjusted poststratification weights applied to the posterior probabilities of Equation (5.5). The symbols for the segments are faded according to the explained inertia so that darker points fit better. No shading was done for the political parties.

for more information on inertia. Across the five waves, three value segments are quite distinct, namely the Traditionalists, the Universalists, and the Entrepreneurs. The Mainstream segment is located in between the Traditionalists and the Universalists and lie more towards the centre of the graph. The Indifferent and Uninterested segments lie close together; however, the Uninterested segment fits less well into the CA solution, suggesting possible measurement issues in this segment. Last, the Security Seekers also fit less well into the display and should therefore not be strongly interpreted.

The first dimension divides the Traditionalists from the Universalists and Entrepreneurs. In value terms this is a distinction between conservation and openness-to-change. The second dimension is a distinction between Universalists, and to a lesser extent Traditionalists, and the Entrepreneurs. This is a distinction between an emphasis on self-transcendence versus self-enhancement. The first dimension distinguishes the religious parties (SGP, CU and CDA) from the non-religious parties (e.g., GL and D66); this meets our expectation that segments considering conservation and self-transcendence important more often vote for religious parties.

The religious parties lie in the upper left quadrant, strongly associated with the Traditionalist segment. They are ordered from the most traditional SGP to the more moderate CDA, and clearly distinguished from the libertarian and authoritarian parties towards the right of the graph. The libertarian parties, namely the PvdD, D66, GL and PvdA, are located alongside each other to the top right quadrant of the figure. The more extreme libertarian parties, namely the GL and PvdD, are located in the upper right corner, whereas the more centrist PvdA are located towards the origin. The authoritarian parties form a tight cluster slightly below the centre. These parties (LPF, PVV, VVD and LN) have tight historical connections, as for instance the LPF was formed by a former member of LN, and the PVV leader similarly was a former member of parliament for the VVD. Members of the Indifferent segment as well as the Entrepreneurs are the most avid supporters of the newly established authoritarian parties. In addition to voting for the VVD, in the elections in 2002 and 2003 these segments voted relatively often for LPF and LN, and since 2006 for the populist PVV. For the Entrepreneurs, universalism is relatively unimportant compared to the other segments. This is congruent with their preference for authoritarian parties, and our expectations. Individuals emphasizing conservation values tend to vote more often for religious parties, especially the SGP and CU. This is in line with our expectations.

As expected, the Universalist segment, which opposes the protection and conservation values and emphasizes growth and openness-to-change values, votes for libertarian parties such as GL.

Figure 5.6: The proportion of ratings used in each response style segment.

Figure 5.7: The distribution of each response style across all value segments.

5.7.4 Response Styles

The value segments analyzed above are adjusted for response styles through the inclusion of the response style segments in the model. Figure 5.6 gives the proportion of responses in each response style segment that was attributed to each of the ratings. Colours are used to highlight larger values. It is evident that a variety of response styles have been detected. The largest styles (1 – 3) concentrate on using ratings one through five, but very few sixes. Response style 4 focuses on using ratings two and three, while response style 5 uses the breadth of the rating scale. Styles 9 and 10 can be described as midpoint scoring, while style 16 comprise extreme scoring focusing on categories 1 and 6. Response styles 18 and 20 contain a disproportionate number of missing values. These results show that there is indeed a lot of heterogeneity with respect to rating scale use.

Figure 5.7 shows the association between the response style and value segments. The percentages in each column sums to one, showing which segments each response style is associated with. Note that some response style segments are often strongly associated with a single value segment. This is especially true for the Indifferent and Uninterested segments. Specifically, the Uninterested segment associates strongly with response styles 18 and 20 (missing values), and the Indifferent segment with response styles 13 – 16 and 19. The latter include some well-known styles (see Baumgartner and Steenkamp, 2001, for example), such as response range (13), acquiescence (14 and 15), extreme responding (16) and disacquiescence (19).

5.8 Discussion and Conclusion

There is a clear link between the seven value segments based on Schwartz' values and people's voting behaviour for political parties. Our approach using the LC-BML model, followed by a correspondence analysis, clearly reveals the three factors identified in the literature according to which political parties in the Netherlands may be categorized (Aarts and Thomassen, 2008). We can see religious parties (SGP, CU and CDA), left-leaning (GL, SP, PvdD) versus right-leaning parties (VVD and PVV) and libertarian (D66, GL) versus more liberal and traditional parties. The division between right and left applied widely in the literature is therefore too simplistic to effectively describe multi-party political systems, such as the Dutch system. Importantly, our method can be applied in other multi-party contexts too.

This study distinguishes itself from previous studies on values and voting in multiple ways. First, we focus on segments instead of considering the whole population as one group. Second, we relate value segments to voting for specific political parties. With our approach we are not only able to confirm established relationships between single values and political orientation, such as correlation between the universalism and a leftist orientation (e.g, Caprara et al., 2006), but also to determine which combinations of value items are related to voting behaviour. Third, we correct for response styles, which is a neglected issue in research on values and voting. It has however long been acknowledged as an issue in public opinion research (for example, Peter and Valkenburg, 2011; Greenleaf, 1992; Alwin and Krosnick, 1985; Bachman and O'Malley, 1984; Cunningham et al., 1977; O'Neill, 1967). Fourth, our study is the first that shows the relationship between values and voting behaviour over a period longer than a decade; the longitudinal study by Schwartz et al. (2010) covered one month.

Fifth, we treat all 21 value items separately, instead of reducing the items to the 10 values they set out to measure (Piurko et al., 2011; Caprara et al., 2006), or to their respective value domains, such as self-transcendence and conservation (Barnea and Schwartz, 1998). The value segments not only differ with respect to the importance they attach to values, but also with respect to specific items. For example, in most segments the two items measuring power are far apart, as are the items measuring universalism. The average for power and stimulation, as would be used when analysing value domains only, would have masked that a similar importance is attached to two different value items PO1 (rich) and ST2 (adventure). The LC-BML model accepts correlated items, allowing us to treat items belonging to the same value separately.

The relations between the value segments and voting behaviour are quite stable over time. For example, the Traditionalists tend to vote for the religious parties (SGP and CU) relatively frequently, whereas the Entrepreneurs vote more frequently for the libertarian parties, such as D66 and VVD. A remarkable segment is the Indifferent segment. This segment either tends not to vote, or, when voting, likely votes for the newly emerging populist parties. Such behaviour is compatible with the results of the Dutch panel study by Van der Meer et al. (2012). However, the reason for their (non)-voting behaviour might be either substantive or methodological. One possible substantive reason is value incongruence with the existing parties, as suggested by Caprara et al. (2012). Another possibility is that populist parties are more engaging to lower educated citizens than the established parties (Hakhverdian et al., 2012). Disinterest or lack of cognitive capacity when answering the survey items, resulting in a response style, might be a methodological explanation.

The LC-BML model identified 20 different response style segments, providing further empirical evidence of the prevalence of response styles in rating scale data. Importantly, our results show that the value and response style segments are not independent. Two problematic value segments, namely the Indifferent and Disinterested segments, are closely associated with specific response styles. Together, these comprise 15.7% of the sample. Hence the answers of a significant number of respondents are mainly driven by specific response styles. In the other segments we might consider the way in which people use the rating scale a communication style. With a communication style, no adjustment, or simple adjustments, may suffice (He and Van De Vijver, 2015). Our study indicates that taking into account response styles in value measurement is important, since ignoring response styles can lead to segments that differ only with respect to rating scale use, and not in value preference. Our study also suggests that the response styles present in empirical data are not limited to one specific style such as extreme responding (for example, Liu et al., 2015; De Jong et al., 2008): people use many different response styles that all might invalidate our findings.

Appendix

Appendix

The official national election results are shown in Table A1. We now describe the bilinear decomposition of the parameters used in the LC-BML model. In order to reduce the number of parameters in (5.3), Van Rosmalen et al. (2010) introduce bilinear decompositions which also make it possible to display all effects in P-dimensional graphs. Let $\mathbf{B}_l$ be the matrix with

	Election			
Party	**2002**	**2003**	**2006**	**2010**
VVD	15.4	17.9	14.7	20.5
PvdA	15.1	27.3	21.2	19.6
PVV	–	–	5.9	15.4
CDA	27.9	28.6	26.5	13.6
SP	5.9	6.3	16.6	9.8
D66	5.1	4.1	2.0	6.9
GL	7.0	5.1	4.6	6.7
CU	2.5	2.1	4.0	3.2
SGP	1.7	1.6	1.6	1.7
PvdD	–	0.5	1.8	1.3
TON	–	–	–	0.6
LPF	17.0	5.7	0.2	–
LN	1.6	0.4	–	–
Other	0.7	0.4	1.0	0.5
Subtotal	9 501 152	9 654 475	9 838 683	9 416 001
Blank/Invalid	14 074	12 127	16 315	26 976
Total	9 515 226	9 666 602	9 854 998	9 442 977
Electorate	12 035 935	12 076 711	12 264 503	12 524 152
Turnout	79.1%	80.0%	80.4%	75.4%

Table A1: The percentage of votes won by the different political parties in the four Dutch elections, as well as the total number of votes cast and the size of the electorate. (Source: www.verkiezingsuitslagen.nl)

$\boldsymbol{\beta}_{1l}, \ldots, \boldsymbol{\beta}_{Kl}$ as rows, and gather the $\gamma_{jk\,|\,s}$ in the $J \times K$ matrix Γ_s. The bilinear restrictions are imposed by requiring that

$$\mathbf{B}_l = \mathbf{F}\mathbf{G}_l' \text{ and } \Gamma_s' = \mathbf{F}\mathbf{H}_s', \tag{5.6}$$

where $\mathbf{F}$, $\mathbf{G}_l$ and $\mathbf{H}_s$ has P columns. Typically, the dimensionality of the graphical representations is chosen to be $P = 1$, 2 or 3 so that it can be displayed easily. The matrix $\mathbf{F}$ contains the coordinates of the K rating categories, $\mathbf{G}_l$ the coordinates for the categories of socio-demographic variable l, and $\mathbf{H}_s$ the coordinates of the J items in value segment s in P-dimensional space. Under these bilinear restrictions, (5.3) becomes

$$\eta_{ijk\,|\,r,s} = \alpha_{k\,|\,r} + \sum_{l=1}^{L} \sum_{p=1}^{P} f_{kp} \boldsymbol{g}_{lp}' \boldsymbol{x}_{il} + \sum_{p=1}^{P} f_{kp} h_{jp\,|\,s}, \tag{5.7}$$

with f_{kp} and $h_{tp\,|\,s}$ being the elements of $\mathbf{F}$ and $\mathbf{H}_s$ respectively, and $\boldsymbol{g}_{lp}$ the pth column of $\mathbf{G}_l$. Besides these bilinear restrictions, several identifiability constraints must be imposed on the parameters in (5.7) – details are given in Van Rosmalen et al. (2010).

CHAPTER 6

Conclusions

This dissertation provides new methods for analyzing categorical data while accounting for the presence of response styles. We also uncover overwhelming empirical evidence that response styles are widespread in such data. Researchers should therefore strive to use appropriate methods, such as those proposed here, whenever analyzing such data. The mantra that data analysis cannot account for bad data collection still applies, but in the absence of proper design methods to negate the effects of the response styles, allowing for this type of variation in the analysis of the data is necessary. Nonetheless, training in these types of methods are limited, mostly because no consensus has been reached as to what constitutes a good method. Another limit is also the statistical expertise of social science researchers who typically conduct surveys. Perhaps in time more objective ways of measuring opinion can be developed, for example as a results of the efforts of the neuroscience community, but it will likely be a while before such methods are widely available to researchers.

A problem with all methods that account for response styles is, of course, that response styles are latent and hard, if not impossible, to measure. Evaluating methods are consequently hard, apart from in idealized simulated situations. Typically, these studies show parameter recovery when simulating from the model specification itself, which, although very interesting, is of limited use when arguing for applying the method to actual data. Ideally the added value of 'response style methods' should be established by showing it's added value in predicting future data. This remains firmly in the domain of future research directions, but such studies would be invaluable in evaluating the growing number of proposed methods out there.

Several avenues for further research stem from the studies in this dissertation. One of these are to use the algorithm in Chapter 4 as the basis for a finite mixture model which can be used for studying the relation between differential item functioning (DIF) and response style bias in psychometric models. DIF and response styles are both sources of measurement invariance, but while DIF is a popular research topic in the psychometric literature, response styles have not received much attention. It is very likely that the effects of these two phenomena are confounded in current statistical assessments. In traditional IRT models, such as aptitude testing, this may not be a problem, but in recent years IRT models have been used for nonability testing as well, such as for analyzing the types of surveys found in this monograph. It is not well understood how the impact from response styles and DIF differ, and if and how they can be separated.

Chapter 3 provides a nonparametric and much more computationally efficient implementation of the LC-BML model as a special case. It would be interesting to directly compare the results of these two methods on the same data sets, and to assess how and why the results differ. The computational difficulties of the LC-BML model makes it very hard to apply in practice, but the results can be readily interpreted. Another feature that will be added to the LSBCLUST framework, is the ability to handle case weights. Modern survey research almost always include weights which should be accounted for in the analysis.

The LC-BML model assumes that the data is measured on a nominal level. Much more can be done when assuming an ordinal measurement level, including reducing the complexity of the model by using the ordering information. An interesting alternative in categorical data pertains to how associations are measured. Often, so-called polychoric correlations are used, which is just the correlation from an assumed underlying multinormal distribution. Correllation is a very simple method of measuring association. For example, specific associations often exist between rating categories on different variables. Alternative association measures can be much more informative, and can be incorporated into a model via copula functions. Work on such a model is already in an advanced stage, and shows promising results. There are many more possibilities, however.

The final chapter, Chapter 5, provides a framework that can easily be applied in other contexts, including other countries in the European Social Survey. We can also substitute the LSBCLUST framework for the LC-BML model to reduce the computational burden.

Bibliography

Aarts, K. and Thomassen, J. (2008). Dutch Voters and the Changing Party Space 1989–2006. *Acta Politica*, 43(2):203–234.

Agresti, A. (2002). *Categorical Data Analysis*. John Wiley & Sons, Hoboken, second edition.

Agresti, A. (2010). *Analysis of Ordinal Categorical Data*. John Wiley & Sons, Hoboken, second edition.

Akaike, H. (1974). A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control*, 19(6):716–723.

Alwin, D. F. and Krosnick, J. A. (1985). The Measurement of Values in Surveys: A Comparison of Ratings and Rankings. *Public Opinion Quarterly*, 49(4):535–552.

Andrich, D. (1978a). A Rating Formulation for Ordered Response Categories. *Psychometrika*, 43(4):561–573.

Andrich, D. (1978b). Application of a Psychometric Rating Model to Ordered Categories Which are Scored with Successive Integers. *Applied Psychological Measurement*, 2(4):581–594.

Andrich, D. (1982). An Extension of the Rasch Model for Ratings Providing both Location and Dispersion Parameters. *Psychometrika*, 47(1):105–113.

Bachman, J. G. and O'Malley, P. M. (1984). Yea-saying, Nay-saying, and Going to Extremes: Black-white Differences in Response Styles. *Public Opinion Quarterly*, 48(2):491–509.

Barnea, M. F. and Schwartz, S. H. (1998). Values and Voting. *Political Psychology*, 19(1):17–40.

Baumgartner, H. and Steenkamp, J.-B. E. (2001). Response Styles in Marketing Research: A Cross-National Investigation. *Journal of Marketing Research*, 38(2):143–156.

Beatty, S. E., Kahle, L. R., Homer, P., and Misra, S. (1985). Alternative Measurement Approaches to Consumer Values: The List of Values and the Rokeach Value Survey. *Psychology & Marketing*, 2(3):181–200.

Benzécri, J.-P. (1992). *Correspondence Analysis Handbook*. CRC Press, Boca Raton.

Bijmolt, T. H. and van de Velden, M. (2012). Multiattribute Perceptual Mapping with Idiosyncratic Brand and Attribute Sets. *Marketing Letters*, 23(3):585–601.

Böhning, D. and Lindsay, B. G. (1988). Monotonicity of Quadratic-Approximation Algorithms. *Annals of the Institute of Statistical Mathematics*, 40(4):641–663.

Boyd, S. and Vandenberghe, L. (2004). *Convex Optimization*. Cambridge University Press, Cambridge.

Breiman, L., Friedman, J., Stone, C. J., and Olshen, R. A. (1984). *Classification and Regression Trees*. CRC Press, Boca Raton.

Briggs, K. (2015). Diagonally Dominant Matrix. From MathWorld—A Wolfram Web Resource, created by Eric W. Weisstein. http://mathworld.wolfram.com/DiagonallyDominantMatrix.html. Last visited on 25/8/2015.

Brunsø, K., Scholderer, J., and Grunert, K. G. (2004). Closing the Gap Between Values and Behavior: A Means–End Theory of Lifestyle. *Journal of Business Research*, 57(6):665–670.

Burroughs, J. E. and Rindfleisch, A. (2002). Materialism and Well-Being: A Conflicting Values Perspective. *Journal of Consumer Research*, 29(3):348–370.

Caliński, T. and Harabasz, J. (1974). A Dendrite Method for Cluster Analysis. *Communications in Statistics - Theory and Methods*, 3(1):1–27.

Caprara, G. V., Schwartz, S. H., Capanna, C., Vecchione, M., and Barbaranelli, C. (2006). Personality and Politics: Values, Traits, and Political Choice. *Political Psychology*, 27(1):1–28.

Caprara, G. V., Vecchione, M., and Schwartz, S. H. (2012). Why People Do Not Vote: The Role of Personal Values. *European Psychologist*, 17(4):266.

Carroll, J. D. and Chang, J.-J. (1970). Analysis of Individual Differences in Multidimensional Scaling via an N-way Generalization of Eckart-Young Decomposition. *Psychometrika*, 35(3):283–319.

Cattell, R. B. (1966). The Scree Test for the Number of Factors. *Multivariate Behavioral Research*, 1(2):245–276.

Ceulemans, E. and Kiers, H. A. (2006). Selecting Among Three-Mode Principal Component Models of Different Types and Complexities: A Numerical Convex Hull Based Method. *British Journal of Mathematical and Statistical Psychology*, 59(1):133–150.

Ceulemans, E. and Kiers, H. A. (2009). Discriminating Between Strong and Weak Structures in Three-Mode Principal Component Analysis. *British Journal of Mathematical and Statistical Psychology*, 62(3):601–620.

Ceulemans, E., Timmerman, M. E., and Kiers, H. A. (2011). The CHull Procedure for Selecting Among Multilevel Component Solutions. *Chemometrics and Intelligent Laboratory Systems*, 106(1):12–20.

Chryssohoidis, G. M. and 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*, 16(7):585–599.

Clayton, D. G. (1978). A Model for Association in Bivariate Life Tables and its Application in Epidemiological Studies of Familial Tendency in Chronic Disease Incidence. *Biometrika*, 65(1):141–151.

Cronbach, L. J. (1941). An Experimental Comparison of the Multiple True-False and Multiple Multiple-Choice Tests. *Journal of Educational Psychology*, 32(7):533.

Cronbach, L. J. (1942). Studies of Acquiescence as a Factor in the True-False Test. *Journal of Educational Psychology*, 33(6):401.

Cronbach, L. J. (1946). Response Sets and Test Validity. *Educational and Psychological Measurement*, 6(4):475–494.

Cronbach, L. J. (1950). Further Evidence on Response Sets and Test Design. *Educational and Psychological Measurement*.

Cunningham, W. H., Cunningham, I. C., and Green, R. T. (1977). The Ipsative Process to Reduce Response Set Bias. *Public Opinion Quarterly*, 41(3):379–384.

Davidov, E., Schmidt, P., and Schwartz, S. H. (2008). Bringing Values Back in the Adequacy of the European Social Survey to Measure Values in 20 Countries. *Public Opinion Quarterly*, 72(3):420–445.

De Jong, M. G., Steenkamp, J.-B. E., Fox, J.-P., and 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. (1994). Block-Relaxation Algorithms in Statistics. In Bock, H.-H., Lenski, W., and Richter, M. M., editors, *Information Systems and Data Analysis*, Studies in Classification, Data Analysis, and Knowledge Organization, pages 308–324. Springer, Berlin.

De Leeuw, J. (2011a). Gifi Goes Logistic. Technical report, Department of Statistics, University of California, Los Angeles.

De Leeuw, J. (2011b). Quadratic and Cubic Majorization. Technical report, Department of Statistics, University of California, Los Angeles.

De Leeuw, J. and Heiser, W. J. (1980). Multidimensional Scaling with Restrictions on the Configuration. *Multivariate Analysis*, 5:501–522.

De Jong, M. G. and Steenkamp, J. B. E. M. (2010). Finite Mixture Multilevel Multidimensional Ordinal IRT Models for Large Scale Cross-Cultural Research. *Psychometrika*, 75(1):3–32.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 39(1):1–38.

Denis, J.-B. and Gower, J. C. (1994). Asymptotic Covariances for the Parameters of Biadditive Models. *Utilitas Mathematica*, 46:193–205.

Dolnicar, S. and Leisch, F. (2010). Evaluation of Structure and Reproducibility of Cluster Solutions Using the Bootstrap. *Marketing Letters*, 21(1):83–101.

Eckart, C. and Young, G. (1936). The Approximation of One Matrix by Another of Lower Rank. *Psychometrika*, 1(3):211–218.

Edwards, J. R. and Cable, D. M. (2009). The Value of Value Congruence. *Journal of Applied Psychology*, 94(3):654.

Efron, B. and Tibshirani, R. J. (1994). *An Introduction to the Bootstrap*. CRC Press, Boca Raton.

Evans, G., Heath, A., and Lalljee, M. (1996). Measuring Left-Right and Libertarian-Authoritarian Values in the British Electorate. *British Journal of Sociology*, 47(1):93–112.

Everitt, B., Landau, S., Leese, M., and Stahl, D. (2011). *Cluster Analysis*. John Wiley & Sons, Chichester, fifth edition.

Faber, N. K. M., Bro, R., and Hopke, P. K. (2003). Recent Developments in CANDECOMP/PARAFAC Algorithms: a Critical Review. *Chemometrics and Intelligent Laboratory Systems*, 65(1):119–137.

Fischer, G. H. and Molenaar, I. W. (1995). *Rasch Models: Foundations, Recent Developments, and Applications*. Springer-Verlag, New York.

Fontaine, J. R., Poortinga, Y. H., Delbeke, L., and Schwartz, S. H. (2008). Structural Equivalence of the Values Domain Across Cultures Distinguishing Sampling Fluctuations From Meaningful Variation. *Journal of Cross-Cultural Psychology*, 39(4):345–365.

Frick, H., Strobl, C., and Zeileis, A. (2015). Rasch Mixture Models for DIF Detection: A Comparison of Old and New Score Specifications. *Educational and Psychological Measurement*, 75(2):208–234.

Gower, J. C. (1975). Generalized Procrustes Analysis. *Psychometrika*, 40(1):33–51.

Gower, J. C. and Dijksterhuis, G. B. (2004). *Procrustes Problems*. Oxford University Press, Oxford.

Gower, J. C., Gardner-Lubbe, S., and Le Roux, N. J. (2011). *Understanding Biplots*. John Wiley & Sons, Hoboken.

Gower, J. C. and Hand, D. J. (1996). *Biplots*. Chapman & Hall, London.

Greenacre, M. (2007). *Correspondence Analysis in Practice*. Chapman & Hall/CRC, Boca Raton.

Greenleaf, E. A. (1992). Measuring Extreme Response Style. *Public Opinion Quarterly*, 56(3):328–351.

Groenen, P. J. F., Mathar, R., and Heiser, W. J. (1995). The Majorization Approach to Multidimensional Scaling for Minkowski Distances. *Journal of Classification*, 12(1):3–19.

Grunert, S. C. and Scherlorn, G. (1990). Consumer Values in West Germany Underlying Dimensions and Cross-Cultural Comparison With North America. *Journal of Business Research*, 20(2):97–107.

Hakhverdian, A., van der Brug, W., and de Vries, C. (2012). The Emergence of a 'Diploma Democracy'? The Political Education Gap in the Netherlands, 1971–2010. *Acta Politica*, 47(3):229–247.

Hand, D. J. and Krzanowski, W. J. (2005). Optimising K-Means Clustering Results with Standard Software Packages. *Computational Statistics & Data Analysis*, 49(4):969–973.

Hardy, A. (1996). On the Number of Clusters. *Computational Statistics & Data Analysis*, 23(1):83–96.

Harshman, R. A. (1970). Foundations of the Parafac Procedure: Models and Conditions for an Explanatory Multimodal Factor Analysis. *UCLA Working Papers in Phonetics*, 16:1–84.

He, J. and Van De Vijver, F. J. (2015). Effects of a General Response Style on Cross-Cultural Comparisons: Evidence from the Teaching and Learning International Survey. *Public Opinion Quarterly*, 79(S1):267–290.

Hoerl, A. E. and Kennard, R. W. (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, 12(1):55–67.

Hofert, M. and Mächler, M. (2011). Nested Archimedean Copulas Meet R: The **nacopula** Package. *Journal of Statistical Software*, 39(9):1–20.

Hoffman, D. L. and Franke, G. R. (1986). Correspondence Analysis: Graphical Representation of Categorical Data in Marketing Research. *Journal of Marketing Research*, 23(3):213–227.

Hubert, L. and Arabie, P. (1985). Comparing Partitions. *Journal of Classification*, 2(1):193–218.

Johnson, R. and Wichern, D. (2002). *Applied Multivariate Statistical Analysis*. Prentice Hall, Upper Saddle River, fifth edition.

Johnson, T. R. (2003). On the Use of Heterogeneous Thresholds Ordinal Regression Models to Account for Individual Differences in Response Style. *Psychometrika*, 68(4):563–583.

Kahle, L. R. (1983). *Social Values and Social Change: Adaptation to Life in America*. Praeger Publishing, Westport.

Kamakura, W. A. and Novak, T. P. (1992). Value-System Segmentation: Exploring the Meaning of LOV. *Journal of Consumer Research*, 19(1):119–132.

Kaufman, L. and Rousseeuw, P. J. (1990). *Finding Groups in Data: An Introduction to Cluster Analysis*. John Wiley & Sons, Hoboken.

Kendall, M. G. (1938). A New Measure of Rank Correlation. *Biometrika*, 30(1/2):81–93.

Kendall, M. G. and Gibbons, J. D. (1990). *Rank Correlation Methods*. Edward Arnold, London.

Kiers, H. A. (2000). Some Procedures for Displaying Results from Three-Way Methods. *Journal of Chemometrics*, 14(3):151–170.

Kiers, H. A. and Krijnen, W. P. (1991). An Efficient Algorithm for PARAFAC of Three-Way Data with Large Numbers of Observation Units. *Psychometrika*, 56(1):147–152.

Kiers, H. A. L. (1997). Weighted Least Squares Fitting Using Ordinary Least Squares Algorithms. *Psychometrika*, 62(2):251–266.

Krijnen, W. P., Dijkstra, T. K., and Stegeman, A. (2008). On the Non-Existence of Optimal Solutions and the Occurrence of Degeneracy in the Candecomp/Parafac Model. *Psychometrika*, 73(3):431–439.

Kroonenberg, P. M. and De Leeuw, J. (1980). Principal Component Analysis of Three-Mode Data by Means of Alternating Least Squares Algorithms. *Psychometrika*, 45(1):69–97.

Krosnick, J. A. (1991). Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys. *Applied Cognitive Psychology*, 5(3):213–236.

Krosnick, J. A. (1999). Survey Research. *Annual Review of Psychology*, 50(1):537–567.

Kruskal, J. B. (1977). Three-way Arrays: Rank and Uniqueness of Trilinear Decompositions, with Application to Arithmetic Complexity and Statistics. *Linear Algebra and Its Applications*, 18(2):95–138.

Krzanowski, W. J. and Lai, Y. (1988). A Criterion for Determining the Number of Groups in a Data Set Using Sum-of-Squares Clustering. *Biometrics*, 44(1):23–34.

Lange, K., Hunter, D. R., and Yang, I. (2000). Optimization Transfer using Surrogate Objective Functions. *Journal of Computational and Graphical Statistics*, 9(1):1–20.

Lawson, C. L. and Hanson, R. J. (1974). *Solving Least Squares Problems*. Prentice Hall, Upper Saddle River.

Lee, J. A., Soutar, G. N., and Louviere, J. (2007). Measuring Values Using Best-Worst Scaling: The LOV Example. *Psychology & Marketing*, 24(12):1043–1058.

Lehmann, E. L. and Casella, G. (1998). *Theory of Point Estimation*. Springer, New York.

Likert, R. (1932). A Technique for the Measurement of Attitudes. *Archives of Psychology*, 22(140):1–55.

Little, R. J. and Rubin, D. B. (2002). *Statistical Analysis with Missing Data*. John Wiley & Sons, Hoboken, second edition.

Liu, M., Lee, S., and Conrad, F. G. (2015). Comparing Extreme Response Styles Between Agree-Disagree and Item-Specific Scales. *Public Opinion Quarterly*.

Lohr, S. L. (1999). *Sampling: Design and Analysis*. Duxbury Press, Boston.

Lord, F. M., Novick, M. R., and Birnbaum, A. (1968). *Statistical Theories of Mental Test Scores*. Addison-Wesley, Oxford.

Lorenzo-Seva, U., Timmerman, M. E., and Kiers, H. A. (2011). The Hull Method for Selecting the Number of Common Factors. *Multivariate Behavioral Research*, 46(2):340–364.

Masters, G. N. (1982). A Rasch Model for Partial Credit Scoring. *Psychometrika*, 47(2):149–174.

Milligan, G. W. and Cooper, M. C. (1985). An Examination of Procedures for Determining the Number of Clusters in a Data Set. *Psychometrika*, 50(2):159–179.

Murtagh, F. (2005). *Correspondence Analysis and Data Coding with Java and R*. Chapman & Hall/CRC, Boca Raton.

Nishisato, S. (1980). Dual Scaling of Successive Categories Data. *Japanese Psychological Research*, 22(3):134–143.

Nishisato, S. (1980a). *Analysis of Categorical Data: Dual Scaling and its Applications*. University of Toronto Press, Toronto.

Nishisato, S. (1994). *Elements of Dual Scaling: An Introduction to Practical Data Analysis*. L. Erlbaum Associates, Hillsdale.

O'Neill, H. W. (1967). Response Style Influence in Public Opinion Surveys. *Public Opinion Quarterly*, 31(1):95–102.

Paulhus, D. L. (1991). Measurement and Control of Response Bias. In Robinson, J. P., Shaver, P. R., and Wrightsman, L. S., editors, *Measures of Personality and Social Psychological Attitudes*, volume 1, pages 17–59. Academic Press, San Diego.

Peter, J. and Valkenburg, P. M. (2011). The Impact of "Forgiving" Introductions on the Reporting of Sensitive Behavior in Surveys The Role of Social Desirability Response Style and Developmental Status. *Public Opinion Quarterly*, 75(4):779–787.

Piurko, Y., Schwartz, S. H., and Davidov, E. (2011). Basic Personal Values and the Meaning of Left-Right Political Orientations in 20 Countries. *Political Psychology*, 32(4):537–561.

Popoviciu, T. (1935). Sur les équations Algébriques Ayant Toutes Leurs Racines Réelles. *Mathematica*, 9:129–145.

R Core Team (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ramsay, J. O. (1988). Monotone Regression Splines in Action. *Statistical Science*, 3(4):425–441.

Rice, J. A. (2007). *Mathematical Statistics and Data Analysis*. Duxbury, Belmont, third edition.

Rorer, L. G. (1965). The Great Response-Style Myth. *Psychological Bulletin*, 63(3):129.

Rost, J. (1990). Rasch Models in Latent Classes: An Integration of Two Approaches to Item Analysis. *Applied Psychological Measurement*, 14(3):271–282.

Rost, J. and von Davier, M. (1995). Mixture Distribution Rasch Models. In Fischer, G. W. and Molenaar, I. W., editors, *Rasch Models: Foundations, Recent Developments, and Applications*, pages 257–268. Springer-Verlag, New York.

Samejima, F. (1969). Estimation of Latent Ability Using a Response Pattern of Graded Scores. *Psychometrika Monograph Supplement*, 34(4 Pt. 2).

Schepers, J., Ceulemans, E., and Van Mechelen, I. (2008). Selecting Among Multi-Mode Partitioning Models of Different Complexities: A Comparison of Four Model Selection Criteria. *Journal of Classification*, 25(1):67–85.

Schoonees, P. C. (2015a). ***lsbclust**: Least-Squares Bilinear Clustering for Three-Way Data*. R package version 1.0.3.

Schoonees, P. C. (2015b). ***cds**: Constrained Dual Scaling for Detecting Response Styles*. R package version 1.0.2.

Schoonees, P. C., Groenen, P. J. F., and van de Velden, M. (2015a). Least-Squares Bilinear Clustering of Three-Way Data. Submitted.

Schoonees, P. C., van de Velden, M., and Groenen, P. J. (2015b). Constrained Dual Scaling for Detecting Response Styles in Categorical Data. *Psychometrika*.

Schultz, P. W., Gouveia, V. V., Cameron, L. D., Tankha, G., Schmuck, P., and Franěk, M. (2005). Values and Their Relationship to Environmental Concern and Conservation Behavior. *Journal of Cross-Cultural Psychology*, 36(4):457–475.

Schwartz, S. H. (1992). Universals in the Content and Structure of Values: Theoretical Advances and Empirical Tests in 20 Countries. *Advances in Experimental Social Psychology*, 25(1):1–65.

Schwartz, S. H. (2006). A Theory of Cultural Value Orientations: Explication and Applications. *Comparative Sociology*, 5(2):137–182.

Schwartz, S. H. (2007). Value Orientations: Measurement, Antecedents and Consequences Across Nations. In Jowell, R., Roberts, C., Fitzgerald, R., and Eva, G., editors, *Measuring Attitudes Cross-Nationally: Lessons from the European Social Survey*, pages 169–203. Sage Publications, London.

Schwartz, S. H. and Bardi, A. (2001). Value Hierarchies Across Cultures Taking a Similarities Perspective. *Journal of Cross-Cultural Psychology*, 32(3):268–290.

Schwartz, S. H., Caprara, G. V., and Vecchione, M. (2010). Basic Personal Values, Core Political Values, and Voting: A Longitudinal Analysis. *Political Psychology*, 31(3):421–452.

Schwartz, S. H., Caprara, G. V., Vecchione, M., Bain, P., Bianchi, G., Caprara, M. G., Cieciuch, J., Kirmanoglu, H., Baslevent, C., Lönnqvist, J.-E., et al. (2013). Basic Personal Values Underlie and Give Coherence to Political Values: A Cross National Study in 15 Countries. *Political Behavior*, 36(4):1–32.

Schwartz, S. H. and Huismans, S. (1995). Value Priorities and Religiosity in Four Western Religions. *Social Psychology Quarterly*, 58(2):88–107.

Schwartz, S. H. and Rubel, T. (2005). Sex Differences in Value Priorities: Cross-Cultural and Multimethod Studies. *Journal of Personality and Social Psychology*, 89(6):1010.

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, 6(2):461–464.

Sherman, J. and Morrison, W. J. (1950). Adjustment of an Inverse Matrix Corresponding to a Change in One Element of a Given Matrix. *The Annals of Mathematical Statistics*, 21(1):124–127.

Sklar, A. (1959). Fonctions de Répartition à n Dimensions et Leurs Marges. *Publications de l'institut de Statistique de l'Université de Paris*, 8(1):11.

Steenkamp, J.-B. E., ter Hofstede, F., and Wedel, M. (1999). A Cross-National Investigation into the Individual and National Cultural Antecedents of Consumer Innovativeness. *Journal of Marketing*, 63(2):55–69.

Sudbury, L. and Simcock, P. (2009). Understanding Older Consumers Through Cognitive Age and the List of Values: A UK-Based Perspective. *Psychology & Marketing*, 26(1):22–38.

Sugar, C. A. and James, G. M. (2003). Finding the Number of Clusters in a Dataset. *Journal of the American Statistical Association*, 98(463).

Ten Berge, J. M. and Sidiropoulos, N. D. (2002). On Uniqueness in CANDECOMP/PARAFAC. *Psychometrika*, 67(3):399–409.

Thorisdottir, H., Jost, J. T., Liviatan, I., and Shrout, P. E. (2007). Psychological Needs and Values Underlying Left-Right Political Orientation: Cross-National Evidence from Eastern and Western Europe. *Public Opinion Quarterly*, 71(2):175–203.

Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 58(1):267–288.

Tibshirani, R., Walther, G., and Hastie, T. (2001). Estimating the Number of Clusters in a Data Set via the Gap Statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2):411–423.

Torres, A. and Greenacre, M. (2002). Dual Scaling and Correspondence Analysis of Preferences, Paired Comparisons and Ratings. *International Journal of Research in Marketing*, 19(4):401–405.

Van der Kooij, A. J. (2007). *Prediction Accuracy and Stability of Regression with Optimal Scaling Transformations*. PhD thesis, Leiden University.

Van der Linden, W. J. and Hambleton, R. K. (1997). Item Response Theory: Brief History, Common Models, and Extensions. In Van der Linden, W. J. and Hambleton, R. K., editors, *Handbook of Modern Item Response Theory*, pages 1–28. Springer, New York.

Van der Meer, T. W., Lubbe, R., Van Elsas, E., Elff, M., and van der Brug, W. (2012). Bounded Volatility in the Dutch Electoral Battlefield: A Panel Study on the Structure of Changing Vote Intentions in the Netherlands During 2006–2010. *Acta Politica*, 47(4):333–355.

Van der Meer, T. W., van Elsas, E., Lubbe, R., and van der Brug, W. (2015). Are Volatile Voters Erratic, Whimsical or Seriously Picky? A Panel Study of 58 Waves into the Nature of Electoral Volatility (The Netherlands 2006–2010). *Party Politics*, 21(1):100–114.

Van Herk, H. (2000). *Equivalence in a Cross-National Context: Methodological & Empirical Issues in Marketing Research*. PhD thesis, Tilburg University.

Van Rosmalen, J., Van Herk, H., and Groenen, P. J. (2010). Identifying Response Styles: A Latent-Class Bilinear Multinomial Logit Model. *Journal of Marketing Research*, 47(1):157–172.

Van Vaerenbergh, Y. and Thomas, T. D. (2013). Response Styles in Survey Research: A Literature Review of Antecedents, Consequences, and Remedies. *International Journal of Public Opinion Research*, 25(2):195–217.

Van de Velden, M. (2000). Dual Scaling and Correspondence Analysis of Of Rank Order Data. In Heijmans, R. D. H., Pollock, D. S. G., and Satorra, A., editors, *Innovations in Multivariate Statistical Analysis*, volume 36, pages 87–99. Kluwer Academic Publishers, Dordrecht.

Van de Velden, M. (2007). Detecting Response Styles by Using Dual Scaling of Successive Categories. Econometric Institute Report, Erasmus University Rotterdam, Rotterdam.

Van de Velden, M., Groenen, P. J. F., and Poblome, J. (2009). Seriation by Constrained Correspondence Analysis: A Simulation Study. *Computational Statistics & Data Analysis*, 53(8):3129–3138.

Von Davier, M. and Rost, J. (1995). Polytomous Mixed Rasch Models. In Fischer, G. W. and Molenaar, I. W., editors, *Rasch Models: Foundations, Recent Developments, and Applications*, pages 371–379. Springer, New York.

Weijters, B. and Baumgartner, H. (2012). Misresponse to Reversed and Negated Items in Surveys: A Review. *Journal of Marketing Research*, 49(5):737–747.

Zou, H. and Hastie, T. (2005). Regularization and Variable Selection via the Elastic Net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2):301–320.

Summary

Rating scale data are widely used for measurement in the social sciences, business and beyond. For example, ratings scales are often utilized for deducing consumer perceptions of films, measuring job satisfaction and evaluating employee performance. Yet using a rating scale is subjective. A CEO pressed for time may choose to use only the extreme rating categories as a means of approving or disapproving the statements put to him. In contrast, a trained panellist being paid to fill out the questionnaire may use a wider range of ratings. Hence, even if they have similar opinions, they would likely respond with different ratings, implying that a rating category potentially carry different meanings for different respondents. Typical statistical analyses of rating data, however, do not account for such individual differences in response styles. Ignoring these discrepancies are dangerous since doing so can lead to erroneous conclusions.

This dissertation makes methodological and empirical contributions towards modelling rating scale data while accounting for differences in response styles. The general approach is to identify individuals in the data which exhibit similar response styles, and to extract substantive information only within such groups. These elements naturally lead to the synthesis of cluster analysis and dimensionality reduction methods. In order to identify these response styles, responses to multiple survey questions are used to assess within-subject rating scale usage. Both non-parametric and parametric approaches are formulated and studied, and accompanying open-source software implementations are made available. The added value of using the developed algorithms, and therefore accounting for response style differences in data analyses, is illustrated by applying these to empirical data. Applications range from sensometrics and brand studies, to psychology and political science.

Afrikaanse Samevatting

(Summary in Afrikaans)

Meting met beoordelings- of graderingskale kom algemeen voor in die sosiale en bestuurswetenskappe, asook in verskeie ander velde, waar dit byvoorbeeld gebruik word om films te beoordeel, werksbevrediging te meet of werknemersprestasie te assesseer. Tog is die gebruik van graderingskale subjektief: ‘n Hoof uitvoerende beampte onder tydsdruk kan byvoorbeeld besluit om slegs die eindpunte van die skaal te gebruik om sy goed- of afkeuring vir die gestelde vrae uit te druk, terwyl ‘n persoon wat vergoeding ontvang vir die voltooiing van die vraelys ‘n wyer keuse van kategorieë kan gebruik. Verskille in hierdie sogenaamde antwoordstyle beteken dat antwoordkategorieë verskillende betekenisse vir verskillende persone kan bedra. Tipiese statistiese analises van graderingsdata maak egter nie voorsiening vir sulke individuele verskille nie. Indien geen aanpassings gemaak word vir verskillende antwoordstyle nie, word die gevaar geloop dat ongeldige gevolgtrekkings gemaak kan word.

Hierdie proefskrif dra by tot die wetenskap deur die ontwikkeling van nuwe statistiese metodes vir die hantering van verskille in antwoordstyle in die analise van graderingsdata. Empiriese gevallestudies lewer ook verdere insig in die voorkoms en effek van antwoordstyle. Die algemene benadering van die nuwe metodes is om eerstens groepe individue te identifiseer wat soortgelyke antwoordstyle openbaar, en om dan beduidende informasie slegs binne sulke groepe individue te ontbloot. Hierdie elemente lei tot die natuurlike kombinasie van trosanalise en dimensiereduksiemetodes. Ten einde antwoordstyle te indentifiseer, word die antwoorde van ‘n spesifieke persoon oor meerdere vrae geanaliseer. Beide parametriese en nie-parametriese metodes word geformuleer en bestudeer, en meegaande sagteware word vrylik beskikbaar gestel. Die toegevoegde waarde wat hierdie metodes oplewer, dit wil sê deur te korrigeer vir verskillende antwoordstyle, word geïllustreer aan die hand van empiriese analises wat oor verskeie velde strek, insluitende sensometrie, handelsmerkstudies en politieke wetenskap.

Nederlandse Samenvatting

(Summary in Dutch)

Opinieschalen worden veel gebruikt voor het doen van metingen in de sociale wetenschappen, in het bedrijfsleven, en in andere sectoren. Voorbeelden hiervan zijn het evalueren van reacties van consumenten op films, het meten van de arbeidstevredenheid, en het evalueren van de prestaties van werknemers. Het gebruik van een opinieschaal is echter altijd subjectief: Een CEO onder tijdsdruk kan ervoor kiezen om alleen de extreme categorieën te gebruiken om de voorgelegde stellingen goed te keuren of af te keuren, terwijl een getraind panellid die wordt betaald voor het invullen van de vragenlijst een groter bereik van categorieën kan gebruiken. Standaard statistische analyses van de opinieschaal-gegevens kunnen niet goed omgaan met individuele verschillen in responsstijlen. Het gevaar van het negeren van variaties in responsstijlen die ontstaan door verschillende interpretaties van de aangeboden antwoordschalen, is dat foutieve conclusies getrokken kunnen worden uit de verzamelde gegevens.

Dit proefschrift levert methodologische en empirische bijdragen aan het modelleren van opinieschaal-gegevens door rekening te houden met verschillen in responsstijlen. De algemene aanpak is om individuen die gelijksoortige responsstijlen vertonen te identificeren en om inhoudelijke informatie alleen te extraheren binnen dergelijke groepen. Deze aspecten leiden op natuurlijke wijze tot een synthese van clusteranalyse en dimensionaliteitsreductie methoden. Om responsstijlen te identificeren worden de antwoorden op meerdere enquêtevragen gebruikt zodat het mogeljik is individueelspecifiek gebruik van opinieschalen op te sporen. In dit proefschrift worden zowel niet-parametrische en parametrische benaderingen geformuleerd en bestudeerd. De bijbehorende open-source software-implementaties zijn openbaar beschikbaar. De toegevoegde waarde van het gebruik van de ontwikkelde algoritmes, en daarmee ook het nut van corrigeren voor responsstijl-verschillen in data-analyses, wordt geïllustreerd door het toepassen van de voorgestelde methoden op gegevens, afkomstig uit verschillende onderzoeksgebieden reikend van zintuigstudies en merkstudies tot psychologie en politieke wetenschappen.

About the Author

Pieter Schoonees obtained his Bachelor's degree in Actuarial Science from the University of Stellenbosch, South Africa, in 2007. He also completed postgraduate Honours Bachelor degrees in Actuarial Science (2008) and Mathematical Statistics (2009, *cum laude*), before completing his Master's degree in Mathematical Statistics (*cum laude*) in early 2011, also in Stellenbosch.

Pieter joined the Erasmus Research Institute of Management in 2011 as a PhD candidate, working on statistical models. In April 2015, he joined the Department of Marketing Management at the Rotterdam School of Management, Erasmus University as a tenure-track post-doctoral researcher. His research interests include computationally intensive statistical models, both parametric and non-parametric, statistical software development, and statistical (or machine) learning. Recently, he is interested in applying such methodology to marketing and neuroscientific problems.

His work has been published in *Psychometrika*, will appear in the *Journal of Statistical Software*, and is under review at other top journals. He has presented at international forums such as the International Conference on Computational Statistics, the International Federation of Classification Societies Conference, the ISMS Marketing Science Conference and the Annual Meeting of the Psychometric Society. Pieter was a visiting researcher at the Department of Psychology of the Universität Zürich in the fall of 2014.

Erasmus Research Institute of Management (ERIM)

ERIM Ph.D. Series Research in Management

The ERIM PhD Series contains PhD dissertations in the field of Research in Management defended at Erasmus University Rotterdam and supervised by senior researchers affiliated to the Erasmus Research Institute of Management (ERIM). All dissertations in the ERIM PhD Series are available in full text through the ERIM Electronic Series Portal: http://repub.eur.nl/pub. ERIM is the joint research institute of the Rotterdam School of Management (RSM) and the Erasmus School of Economics at the Erasmus University Rotterdam (EUR).

Dissertations Last Five Years

Abbink, E.J., *Crew Management in Passenger Rail Transport*, Promotor(s): Prof.dr. L.G. Kroon & Prof.dr. A.P.M. Wagelmans, EPS-2014-325-LIS, http://repub.eur.nl/pub/76927

Acar, O.A., *Crowdsourcing for Innovation: Unpacking Motivational, Knowledge and Relational Mechanisms of Innovative Behavior in Crowdsourcing Platforms*, Promotor(s): Prof.dr.ir. J.C.M. van den Ende, EPS-2014-321-LIS, http://repub.eur.nl/pub/76076

Acciaro, M., *Bundling Strategies in Global Supply Chains*, Promotor(s): Prof.dr. H.E. Haralambides, EPS-2010-197-LIS, http://repub.eur.nl/pub/19742

Akin Ates, M., *Purchasing and Supply Management at the Purchase Category Level: Strategy, Structure and Performance*, Promotor(s): Prof.dr. J.Y.F. Wynstra & Dr. E.M. van Raaij, EPS-2014-300-LIS, http://repub.eur.nl/pub/50283

Akpinar, E., *Consumer Information Sharing*, Promotor(s): Prof.dr.ir. A. Smidts, EPS-2013-297-MKT, http://repub.eur.nl/pub/50140

Alexander, L., *People, Politics, and Innovation: A Process Perspective*, Promotor(s): Prof.dr. H.G. Barkema & Prof.dr. D.L. van Knippenberg, EPS-2014-331-S&E, http://repub.eur.nl/pub/77209

Alexiev, A.S., *Exploratory Innovation: The Role of Organizational and Top Management Team Social Capital*, Promotor(s): Prof.dr.ing. F.A.J. van den Bosch & Prof.dr. H.W. Volberda, EPS-2010-208-STR, http:

//repub.eur.nl/pub/20632

Almeida e Santos Nogueira, R.J. de, *Conditional Density Models Integrating Fuzzy and Probabilistic Representations of Uncertainty*, Promotor(s): Prof.dr.ir. U. Kaymak & Prof.dr. J.M.C. Sousa, EPS-2014-310-LIS, http://repub.eur.nl/pub/51560

Bannouh, K., *Measuring and Forecasting Financial Market Volatility Using High-Frequency Data*, Promotor(s): Prof.dr. D.J.C. van Dijk, EPS-2013-273-F&A, http://repub.eur.nl/pub/38240

Ben-Menahem, S.M., *Strategic Timing and Proactiveness of Organizations*, Promotor(s): Prof.dr. H.W. Volberda & Prof.dr.ing. F.A.J. van den Bosch, EPS-2013-278-S&E, http://repub.eur.nl/pub/39128

Benning, T.M., *A Consumer Perspective on Flexibility in Health Care: Priority Access Pricing and Customized Care*, Promotor(s): Prof.dr.ir. B.G.C. Dellaert, EPS-2011-241-MKT, http://repub.eur.nl/pub/23670

Berg, W.E. van den, *Understanding Salesforce Behavior Using Genetic Association Studies*, Promotor(s): Prof.dr. W.J.M.I. Verbeke, EPS-2014-311-MKT, http://repub.eur.nl/pub/51440

Betancourt, N.E., *Typical Atypicality: Formal and Informal Institutional Conformity, Deviance, and Dynamics*, Promotor(s): Prof.dr. B. Krug, EPS-2012-262-ORG, http://repub.eur.nl/pub/32345

Bezemer, P.J., *Diffusion of Corporate Governance Beliefs: Board Independence and the Emergence of a Shareholder Value Orientation in the Netherlands*, Promotor(s): Prof.dr.ing. F.A.J. van den Bosch & Prof.dr. H.W. Volberda, EPS-2010-192-STR, http://repub.eur.nl/pub/18458

Binken, J.L.G., *System Markets: Indirect Network Effects in Action, or Inaction?*, Promotor(s): Prof.dr. S. Stremersch, EPS-2010-213-MKT, http://repub.eur.nl/pub/21186

Bliek, R. de, *Empirical Studies on the Economic Impact of Trust*, Promotor(s): Prof.dr. J. Veenman & Prof.dr. Ph.H.B.F. Franses, EPS-2015-324-ORG, http://repub.eur.nl/pub/78159

Blitz, D.C., *Benchmarking Benchmarks*, Promotor(s): Prof.dr. A.G.Z. Kemna & Prof.dr. W.F.C. Verschoor, EPS-2011-225-F&A, http://repub.eur.nl/pub/22624

Boons, M., *Working Together Alone in the Online Crowd: The Effects of Social Motivations and Individual Knowledge Backgrounds on the Participation and Performance of Members of Online Crowdsourcing Platforms*, Promotor(s): Prof.dr. H.G. Barkema & Dr. D.A. Stam, EPS-2014-306-S&E, http://repub.eur.nl/pub/50711

Borst, W.A.M., *Understanding Crowdsourcing: Effects of Motivation and Rewards on Participation and Performance in Voluntary Online Activities*, Promotor(s): Prof.dr.ir. J.C.M. van den Ende & Prof.dr.ir. H.W.G.M. van Heck, EPS-2010-221-LIS, http://repub.eur.nl/pub/21914

Brazys, J., *Aggregated Marcoeconomic News and Price Discovery*, Promotor(s): Prof.dr. W.F.C. Verschoor, EPS-2015-351-F&A, http://repub.eur.nl/pub/78243

Budiono, D.P., *The Analysis of Mutual Fund Performance: Evidence from U.S. Equity Mutual Funds*, Promotor(s): Prof.dr. M.J.C.M. Verbeek & Dr.ir. M.P.E. Martens, EPS-2010-185-F&A, http://repub.eur.nl/pub/18126

Burger, M.J., *Structure and Cooptition in Urban Networks*, Promotor(s): Prof.dr. G.A. van der Knaap & Prof.dr. H.R. Commandeur, EPS-2011-243-ORG, http://repub.eur.nl/pub/26178

Byington, E., *Exploring Coworker Relationships: Antecedents and Dimensions of Interpersonal Fit, Coworker Satisfaction, and Relational Models*, Promotor(s): Prof.dr. D.L. van Knippenberg, EPS-2013-292-ORG, http://repub.eur.nl/pub/41508

Camacho, N.M., *Health and Marketing: Essays on Physician and Patient Decision- Making*, Promotor(s): Prof.dr. S. Stremersch, EPS-2011-237-MKT, http://repub.eur.nl/pub/23604

Cancurtaran, P., *Essays on Accelerated Product Development*, Promotor(s): Prof.dr. F. Langerak & Prof.dr.ir. G.H. van Bruggen, EPS-2014-317-MKT, http://repub.eur.nl/pub/76074

Caron, E.A.M., *Explanation of Exceptional Values in Multi-Dimensional Business Databases*, Promotor(s): Prof.dr.ir. H.A.M. Daniels & Prof.dr. G.W.J. Hendrikse, EPS-2013-296-LIS, http://repub.eur.nl/pub/50005

Carvalho, L. de, *Knowledge Locations in Cities: Emergence and Development Dynamics*, Promotor(s): Prof.dr. L. Berg, EPS-2013-274-S&E, http://repub.eur.nl/pub/38449

Carvalho de Mesquita Ferreira, L., *Attention Mosaics: Studies of Organizational Attention*, Promotor(s): Prof.dr. P.P.M.A.R. Heugens & Prof.dr. J. van Oosterhout, EPS-2010-205-ORG, http://repub.eur.nl/pub/19882

Cox, R.H.G.M., *To Own, to Finance, and to Insure - Residential Real Estate Revealed*, Promotor(s): Prof.dr. D. Brounen, EPS-2013-290-F&A, http://repub.eur.nl/pub/40964

Defilippi Angeldonis, E.F., *Access Regulation for Naturally Monopolistic Port Terminals: Lessons from Regulated Network Industries*, Promotor(s): Prof.dr. H.E. Haralambides, EPS-2010-204-LIS, http://repub.eur.nl/pub/19881

Deichmann, D., *Idea Management: Perspectives from Leadership, Learning, and Network Theory*, Promotor(s): Prof.dr.ir. J.C.M. van den Ende, EPS-2012-255-ORG, http://repub.eur.nl/pub/31174

Deng, W., *Social Capital and Diversification of Cooperatives*, Promotor(s): Prof.dr. G.W.J. Hendrikse, EPS-2015-341-ORG, http://repub.eur.nl/pub/77449

Desmet, P.T.M.,*In Money We Trust? Trust Repair and the Psychology of Financial Compensations*, Promotor(s): Prof.dr. D. de Cremer, EPS-2011-232-ORG, http://repub.eur.nl/pub/23268

Dietvorst, R.C., *Neural Mechanisms Underlying Social Intelligence and their Relationship with the Performance of Sales Managers*, Promotor(s): Prof.dr. W.J.M.I. Verbeke, EPS-2010-215-MKT, http://repub.eur.nl/pub/21188

Dollevoet, T.A.B., *Delay Management and Dispatching in Railways*, Promotor(s): Prof.dr. A.P.M. Wagelmans, EPS-2013-272-LIS, http://repub.eur.nl/pub/38241

Doorn, S. van, *Managing Entrepreneurial Orientation*, Promotor(s): Prof.dr. J.J.P. Jansen, Prof.dr.ing. F.A.J. van den Bosch, & Prof.dr. H.W. Volberda, EPS-2012-258- STR, http://repub.eur.nl/pub/32166

Douwens-Zonneveld, M.G., *Animal Spirits and Extreme Confidence: No Guts, No Glory?*, Promotor(s): Prof.dr. W.F.C. Verschoor, EPS-2012-257-F&A, http://repub.eur.nl/pub/31914

Duca, E., *The Impact of Investor Demand on Security Offerings*, Promotor(s): Prof.dr. A. de Jong, EPS-2011-240-F&A, http://repub.eur.nl/pub/26041

Duursema, H., *Strategic Leadership: Moving Beyond the Leader-Follower Dyad*, Promotor(s): Prof.dr. R.J.M. van Tulder, EPS-2013-279-ORG, http://repub.eur.nl/pub/39129

Eck, N.J. van, *Methodological Advances in Bibliometric Mapping of Science*, Promotor(s): Prof.dr.ir. R. Dekker, EPS-2011-247-LIS, http://repub.eur.nl/pub/26509

Ellen, S. ter, *Measurement, Dynamics, and Implications of Heterogeneous Beliefs in Financial Markets*, Promotor(s): Prof.dr. W.F.C. Verschoor, EPS-2015-343-F&A, http://repub.eur.nl/pub/78191

Eskenazi, P.I., *The Accountable Animal*, Promotor(s): Prof.dr. F.G.H. Hartmann, EPS-2015-355-F&A, http://repub.eur.nl/pub/78300

Essen, M. van, *An Institution-Based View of Ownership*, Promotor(s): Prof.dr. J. van Oosterhout & Prof.dr. G.M.H. Mertens, EPS-2011-226-ORG, http://repub.eur.nl/pub/22643

Evangelidis, I., *Preference Construction Under Prominence*, Promotor(s): Prof.dr. S.M.J. van Osselaer, EPS-2015-340-MKT, http://repub.eur.nl/pub/78202

Feng, L., *Motivation, Coordination and Cognition in Cooperatives*, Promotor(s): Prof.dr. G.W.J. Hendrikse, EPS-2010-220-ORG, http://repub.eur.nl/pub/21680

Fourne, S.P., *Managing Organizational Tensions: A Multi-Level Perspective on Exploration, Exploitation and Ambidexterity*, Promotor(s): Prof.dr. J.J.P. Jansen & Prof.dr. S.J. Magala, EPS-2014-318-S&E, http://repub.eur.nl/pub/76075

Gharehgozli, A.H., *Developing New Methods for Efficient Container Stacking Operations*, Promotor(s): Prof.dr.ir. M.B.M. de Koster, EPS-2012-269-LIS, http://repub.eur.nl/pub/37779

Gils, S. van, *Morality in Interactions: On the Display of Moral Behavior by Leaders and Employees*, Promotor(s): Prof.dr. D.L. van Knippenberg, EPS-2012-270-ORG, http://repub.eur.nl/pub/38027

Ginkel-Bieshaar, M.N.G. van, *The Impact of Abstract Versus Concrete Product Communications on Consumer Decision-Making Processes*, Promotor(s): Prof.dr.ir. B.G.C. Dellaert, EPS-2012-256-MKT, http://repub.eur.nl/pub/31913

Gkougkousi, X., *Empirical Studies in Financial Accounting*, Promotor(s): Prof.dr. G.M.H. Mertens & Prof.dr. E. Peek, EPS-2012-264-F&A, http://repub.eur.nl/pub/37170

Glorie, K.M., *Clearing Barter Exchange Markets: Kidney Exchange and Beyond*, Promotor(s): Prof.dr. A.P.M. Wagelmans & Prof.dr. J.J. van de Klundert, EPS-2014-329-LIS, http://repub.eur.nl/pub/77183

Hakimi, N.A., *Leader Empowering Behaviour: The Leader's Perspective*, Promotor(s): Prof.dr. D.L. van Knippenberg, EPS-2010-184-ORG, http://repub.eur.nl/pub/17701

Hensmans, M., *A Republican Settlement Theory of the Firm: Applied to Retail Banks in England and the Netherlands (1830-2007)*, Promotor(s): Prof.dr. A. Jolink & Prof.dr. S.J. Magala, EPS-2010-193-ORG, http://repub.eur.nl/pub/19494

Hernandez-Mireles, C., *Marketing Modeling for New Products*, Promotor(s): Prof.dr. Ph.H.B.F. Franses, EPS-2010-202-MKT, http://repub.eur.nl/pub/19878

Heyde Fernandes, D. von der, *The Functions and Dysfunctions of Reminders*, Promotor(s): Prof.dr. S.M.J. van Osselaer, EPS-2013-295-MKT, http://repub.eur.nl/pub/41514

Heyden, M.L.M., *Essays on Upper Echelons & Strategic Renewal: A Multilevel Contingency Approach*, Promotor(s): Prof.dr.ing. F.A.J. van den Bosch & Prof.dr. H.W. Volberda, EPS-2012-259-STR, http://repub.eur.nl/pub/32167

Hoever, I.J., *Diversity and Creativity*, Promotor(s): Prof.dr. D.L. van Knippenberg, EPS-2012-267-ORG, http://repub.eur.nl/pub/37392

Hogenboom, F.P., *Automated Detection of Financial Events in News Text*, Promotor(s): Prof.dr.ir. U. Kaymak & Prof.dr. F.M.G. de Jong, EPS-2014-326-LIS, http://repub.eur.nl/pub/77237

Hoogendoorn, B., *Social Entrepreneurship in the Modern Economy: Warm Glow, Cold Feet*, Promotor(s): Prof.dr. H.P.G. Pennings & Prof.dr. A.R. Thurik, EPS-2011-246-STR, http://repub.eur.nl/pub/26447

Hoogervorst, N., *On the Psychology of Displaying Ethical Leadership: A Behavioral Ethics Approach*, Promotor(s): Prof.dr. D. de Cremer & Dr. M. van Dijke, EPS-2011-244-ORG, http://repub.eur.nl/pub/26228

Hout, D.H. van, *Measuring Meaningful Differences: Sensory Testing Based Decision Making in an Industrial Context; Applications of Signal Detection Theory and Thurstonian Modelling*, Promotor(s): Prof.dr. P.J.F. Groenen & Prof.dr. G.B. Dijksterhuis, EPS-2014-304-MKT, http://repub.eur.nl/pub/50387

Houwelingen, G.G. van, *Something to Rely On*, Promotor(s): Prof.dr. D. de Cremer & Prof.dr. M.H. van Dijke, EPS-2014-335-ORG, http://repub.eur.nl/pub/77320

Huang, X., *An Analysis of Occupational Pension Provision: From Evaluation to Redesign*, Promotor(s): Prof.dr. M.J.C.M. Verbeek, EPS-2010-196-F&A, http://repub.eur.nl/pub/19674

Hurk, E. van der, *Passengers, Information, and Disruptions*, Promotor(s): Prof.dr. L.G. Kroon & Prof.mr.dr. P.H.M. Vervest, EPS-2015-345-LIS, http://repub.eur.nl/pub/78275

Hytonen, K.A., *Context Effects in Valuation, Judgment and Choice: a Neuroscientific Approach*, Promotor(s): Prof.dr.ir. A. Smidts, EPS-2011-252-MKT, http://repub.eur.nl/pub/30668

Iseger, P. den, *Fourier and Laplace Transform Inversion with Applications in Finance*, Promotor(s): Prof.dr.ir. R. Dekker, EPS-2014-322-LIS, http://repub.eur.nl/pub/76954

Jaarsveld, W.L. van, *Maintenance Centered Service Parts Inventory Control*, Promotor(s): Prof.dr.ir. R. Dekker, EPS-2013-288-LIS, http://repub.eur.nl/pub/39933

Jalil, M.N., *Customer Information Driven After Sales Service Management: Lessons from Spare Parts Logistics*, Promotor(s): Prof.dr. L.G. Kroon, EPS-2011-222-LIS, http://repub.eur.nl/pub/22156

Kagie, M., *Advances in Online Shopping Interfaces: Product Catalog Maps and Recommender Systems*, Promotor(s): Prof.dr. P.J.F. Groenen, EPS-2010-195-MKT, http://repub.eur.nl/pub/19532

Kappe, E.R., *The Effectiveness of Pharmaceutical Marketing*, Promotor(s): Prof.dr. S. Stremersch, EPS-2011-239-MKT, http://repub.eur.nl/pub/23610

Karreman, B., *Financial Services and Emerging Markets*, Promotor(s): Prof.dr. G.A. van der Knaap & Prof.dr. H.P.G. Pennings, EPS-2011-223-ORG, http://repub.eur.nl/pub/22280

Khanagha, S., *Dynamic Capabilities for Managing Emerging Technologies*, Promotor(s): Prof.dr. H.W. Volberda, EPS-2014-339-S&E, http://repub.eur.nl/pub/77319

Kil, J., *Acquisitions Through a Behavioral and Real Options Lens*, Promotor(s): Prof.dr. H.T.J. Smit, EPS-2013-298-F&A, http://repub.eur.nl/pub/50142

Klooster, E. van 't, *Travel to Learn: The Influence of Cultural Distance on Competence Development in Educational Travel*, Promotor(s): Prof.dr. F.M. Go & Prof.dr. P.J. van Baalen, EPS-2014-312-MKT, http://repub.eur.nl/pub/51462

Koendjbiharie, S.R., *The Information-Based View on Business Network Performance: Revealing the Performance of Interorganizational Networks*, Promotor(s): Prof.dr.ir. H.W.G.M. van Heck & Prof.mr.dr. P.H.M. Vervest, EPS-2014-315-LIS, http://repub.eur.nl/pub/51751

Koning, M., *The Financial Reporting Environment: The Role of the Media, Regulators and Auditors*, Promotor(s): Prof.dr. G.M.H. Mertens & Prof.dr. P.G.J. Roosenboom, EPS-2014-330-F&A, http://repub.eur.nl/pub/77154

Konter, D.J., *Crossing Borders with HRM: an Inquiry of the Influence of Contextual Differences in the Adoption and Effectiveness of HRM*, Promotor(s): Prof.dr. J. Paauwe & Dr. L.H. Hoeksema, EPS-2014-305-ORG, http://repub.eur.nl/pub/50388

Korkmaz, E., *Bridging Models and Business: Understanding Heterogeneity in Hidden Drivers of Customer Purchase Behavior*, Promotor(s): Prof.dr. S.L. van de Velde & Prof.dr. D. Fok, EPS-2014-316-LIS, http://repub.eur.nl/pub/76008

Kroezen, J.J., *The Renewal of Mature Industries: An Examination of the Revival of the Dutch Beer Brewing Industry*, Promotor(s): Prof.dr. P.P.M.A.R. Heugens, EPS-2014-333-S&E, http://repub.eur.nl/pub/77042

Kysucky, V., *Access to Finance in a Cross-Country Context*, Promotor(s): Prof.dr. L. Norden, EPS-2015-350-F&A, http://repub.eur.nl/pub/78225

Lam, K.Y., *Reliability and Rankings*, Promotor(s): Prof.dr. Ph.H.B.F. Franses, EPS-2011-230-MKT, http://repub.eur.nl/pub/22977

Lander, M.W., *Profits or Professionalism? On Designing Professional Service Firms*, Promotor(s): Prof.dr. J. van Oosterhout & Prof.dr. P.P.M.A.R. Heugens, EPS-2012-253-ORG, http://repub.eur.nl/pub/30682

Langhe, B. de, *Contingencies: Learning Numerical and Emotional Associations in an Uncertain World*, Promotor(s): Prof.dr.ir. B. Wierenga & Prof.dr. S.M.J. van Osselaer, EPS-2011-236-MKT, http://repub.eur.nl/pub/23504

Larco Martinelli, J.A., *Incorporating Worker-Specific Factors in Operations Management Models*, Promotor(s): Prof.dr.ir. J. Dul & Prof.dr.ir. M.B.M. de Koster, EPS-2010-217- LIS, http://repub.eur.nl/pub/21527

Leunissen, J.M., *All Apologies: On the Willingness of Perpetrators to Apologize*, Promotor(s): Prof.dr. D. de Cremer & Dr. M. van Dijke, EPS-2014-301-ORG, http://repub.eur.nl/pub/50318

Li, D., *Supply Chain Contracting for After-Sales Service and Product Support*, Promotor(s): Prof.dr.ir. M.B.M. de Koster, EPS-2015-347-LIS, http://repub.eur.nl/pub/78526

Li, Z., *Irrationality: What, Why and How*, Promotor(s): Prof.dr. H. Bleichrodt, Prof.dr. P.P. Wakker, & Prof.dr. K.I.M. Rohde, EPS-2014-338-MKT, http://repub.eur.nl/pub/77205

Liang, Q.X., *Governance, CEO Identity, and Quality Provision of Farmer Cooperatives*, Promotor(s): Prof.dr. G.W.J. Hendrikse, EPS-2013-281-ORG, http://repub.eur.nl/pub/39253

Liket, K., *Why 'Doing Good' Is Not Good Enough: Essays on Social Impact Measurement*, Promotor(s): Prof.dr. H.R. Commandeur & Dr. K.E.H. Maas, EPS-2014-307-STR, http://repub.eur.nl/pub/51130

Loos, M.J.H.M. van der, *Molecular Genetics and Hormones: New Frontiers in Entrepreneurship Research*, Promotor(s): Prof.dr. A.R. Thurik, Prof.dr. P.J.F. Groenen, & Prof.dr. A. Hofman, EPS-2013-287-S&E, http://repub.eur.nl/pub/40081

Lovric, M., *Behavioral Finance and Agent-Based Artificial Markets*, Promotor(s): Prof.dr. J. Spronk & Prof.dr.ir. U. Kaymak, EPS-2011-229-F&A, http://repub.eur.nl/pub/22814

Lu, Y., *Data-Driven Decision Making in Auction Markets*, Promotor(s): Prof.dr.ir. H.W.G.M. van Heck & Prof.dr. W. Ketter, EPS-2014-314-LIS, http://repub.eur.nl/pub/51543

Manders, B., *Implementation and Impact of ISO 9001*, Promotor(s): Prof.dr. K. Blind, EPS-2014-337-LIS, http://repub.eur.nl/pub/77412

Markwat, T.D., *Extreme Dependence in Asset Markets Around the Globe*, Promotor(s): Prof.dr. D.J.C. van Dijk, EPS-2011-227-F&A, http://repub.eur.nl/pub/22744

Mees, H., *Changing Fortunes: How China's Boom Caused the Financial Crisis*, Promotor(s): Prof.dr. Ph.H.B.F. Franses, EPS-2012-266-MKT, http://repub.eur.nl/pub/34930

Meuer, J., *Configurations of Inter-Firm Relations in Management Innovation: A Study in China's Biopharmaceutical Industry*, Promotor(s): Prof.dr. B. Krug, EPS-2011-228-ORG, http://repub.eur.nl/pub/22745

Micheli, M.R., *Business Model Innovation: A Journey Across Managers' Attention and Inter-Organizational Networks*, Promotor(s): Prof.dr. J.J.P. Jansen, EPS-2015-344-S&E, http://repub.eur.nl/pub/78241

Mihalache, O.R., *Stimulating Firm Innovativeness: Probing the Interrelations Between Managerial and Organizational Determinants*, Promotor(s): Prof.dr. J.J.P. Jansen, Prof.dr.ing. F.A.J. van den Bosch, & Prof.dr. H.W. Volberda, EPS-2012-260-S&E, http://repub.eur.nl/pub/32343

Milea, V., *News Analytics for Financial Decision Support*, Promotor(s): Prof.dr.ir. U. Kaymak, EPS-2013-275-LIS, http://repub.eur.nl/pub/38673

Naumovska, I., *Socially Situated Financial Markets: A Neo-Behavioral Perspective on Firms, Investors and Practices*, Promotor(s): Prof.dr. P.P.M.A.R. Heugens & Prof.dr. A. de Jong, EPS-2014-319-S&E, http://repub.eur.nl/pub/76084

Nielsen, L.K., *Rolling Stock Rescheduling in Passenger Railways: Applications in Short Term Planning and in Disruption Management*, Promotor(s): Prof.dr. L.G. Kroon, EPS-2011-224-LIS, http://repub.eur.nl/pub/22444

Nijdam, M.H., *Leader Firms: The Value of Companies for the Competitiveness of the Rotterdam Seaport Cluster*, Promotor(s): Prof.dr. R.J.M. van Tulder, EPS-2010-216-ORG, http://repub.eur.nl/pub/21405

Noordegraaf-Eelens, L.H.J., *Contested Communication; A Critical Analysis of Central Bank Speech*, Promotor(s): Prof.dr. Ph.H.B.F. Franses, Prof.dr. J. de Mul, & Prof.dr. D.J.C. van Dijk, EPS-2010-209-MKT, http://repub.eur.nl/pub/21061

Nuijten, A.L.P., *Deaf Effect for Risk Warnings: A Causal Examination Applied to Information Systems Projects*, Promotor(s): Prof.dr. G.J. van der Pijl, Prof.dr. H.R. Commandeur & Prof.dr. M. Keil, EPS-2012-263-S&E, http://repub.eur.nl/pub/34928

Oosterhout, M. van, *Business Agility and Information Technology in Service Organizations*, Promotor(s): Prof.dr.ir. H.W.G.M. van Heck, EPS-2010-198-LIS, http://repub.eur.nl/pub/19805

Osadchiy, S.E., *The Dynamics of Formal Organization: Essays on Bureaucracy and Formal Rules*, Promotor(s): Prof.dr. P.P.M.A.R. Heugens, EPS-2011-231-ORG, http://repub.eur.nl/pub/23250

Otgaar, A.H.J., *Industrial Tourism: Where the Public Meets the Private*, Promotor(s): Prof.dr. L. Berg, EPS-2010-219-ORG, http://repub.eur.nl/pub/21585

Ozdemir, M.N., Project-level Governance, Monetary Incentives, and Performance in Strategic R&D Alliances, Promotor(s): Prof.dr.ir. J.C.M. van den Ende, EPS-2011-235-LIS, http://repub.eur.nl/pub/23550

Peers, Y., *Econometric Advances in Diffusion Models*, Promotor(s): Prof.dr. Ph.H.B.F. Franses, EPS-2011-251-MKT, http://repub.eur.nl/pub/30586

Peters, M., *Machine Learning Algorithms for Smart Electricity Markets*, Promotor(s): Prof.dr. W. Ketter, EPS-2014-332-LIS, http://repub.eur.nl/pub/77413

Pince, C., *Advances in Inventory Management: Dynamic Models*, Promotor(s): Prof.dr.ir. R. Dekker, EPS-2010-199-LIS, http://repub.eur.nl/pub/19867

Porck, J., *No Team Is an Island: An Integrative View of Strategic Consensus Between Groups*, Promotor(s): Prof.dr. P.J.F. Groenen & Prof.dr. D.L. van Knippenberg, EPS-2013-299-ORG, http://repub.eur.nl/pub/50141

Porras Prado, M., *The Long and Short Side of Real Estate, Real Estate Stocks, and Equity*, Promotor(s): Prof.dr. M.J.C.M. Verbeek, EPS-2012-254-F&A, http://repub.eur.nl/pub/30848

Poruthiyil, P.V., *Steering Through: How Organizations Negotiate Permanent Uncertainty and Unresolvable Choices*, Promotor(s): Prof.dr. P.P.M.A.R. Heugens & Prof.dr. S.J. Magala, EPS-2011-245-ORG, http://repub.eur.nl/

pub/26392

Potthoff, D., *Railway Crew Rescheduling: Novel Approaches and Extensions*, Promotor(s): Prof.dr. A.P.M. Wagelmans & Prof.dr. L.G. Kroon, EPS-2010-210-LIS, http://repub.eur.nl/pub/21084

Pourakbar, M., *End-Of-Life Inventory Decisions of Service Parts*, Promotor(s): Prof.dr.ir. R. Dekker, EPS-2011-249-LIS, http://repub.eur.nl/pub/30584

Pronker, E.S., *Innovation Paradox in Vaccine Target Selection*, Promotor(s): Prof.dr. H.J.H.M. Claassen & Prof.dr. H.R. Commandeur, EPS-2013-282-S&E, http://repub.eur.nl/pub/39654

Pruijssers, J.K., *An Organizational Perspective on Auditor Conduct*, Promotor(s): Prof.dr. J. van Oosterhout & Prof.dr. P.P.M.A.R. Heugens, EPS-2015-342-S&E, http://repub.eur.nl/pub/78192

Retel Helmrich, M.J., *Green Lot-Sizing*, Promotor(s): Prof.dr. A.P.M. Wagelmans, EPS-2013-291-LIS, http://repub.eur.nl/pub/41330

Rietveld, N., *Essays on the Intersection of Economics and Biology*, Promotor(s): Prof.dr. A.R. Thurik, Prof.dr. Ph.D. Koellinger, Prof.dr. P.J.F. Groenen, & Prof.dr. A. Hofman, EPS-2014-320-S&E, http://repub.eur.nl/pub/76907

Rijsenbilt, J.A., *CEO Narcissism: Measurement and Impact*, Promotor(s): Prof.dr. A.G.Z. Kemna & Prof.dr. H.R. Commandeur, EPS-2011-238-STR, http://repub.eur.nl/pub/23554

Roelofsen, E.M., *The Role of Analyst Conference Calls in Capital Markets*, Promotor(s): Prof.dr. G.M.H. Mertens & Prof.dr. L.G. van der Tas, EPS-2010-190-F&A, http://repub.eur.nl/pub/18013

Roza-van Vuren, M.W., *The Relationship Between Offshoring Strategies and Firm Performance: Impact of Innovation, Absorptive Capacity and Firm Size*, Promotor(s): Prof.dr. H.W. Volberda & Prof.dr.ing. F.A.J. van den Bosch, EPS-2011-214-STR, http://repub.eur.nl/pub/22155

Rubbaniy, G., *Investment Behaviour of Institutional Investors*, Promotor(s): Prof.dr. W.F.C. Verschoor, EPS-2013-284-F&A, http://repub.eur.nl/pub/40068

Schellekens, G.A.C., *Language Abstraction in Word of Mouth*, Promotor(s): Prof.dr.ir. A. Smidts, EPS-2010-218-MKT, http://repub.eur.nl/pub/21580

Schoonees, P.C., *Methods for Modelling Response Styles*, Promotor(s): Prof.dr. P.J.F. Groenen, EPS-2015-348-MKT, http://repub.eur.nl/pub/

Shahzad, K., *Credit Rating Agencies, Financial Regulations and the Capital Markets*, Promotor(s): Prof.dr. G.M.H. Mertens, EPS-2013-283-F&A, http://repub.eur.nl/pub/39655

Sotgiu, F., *Not All Promotions Are Made Equal: From the Effects of a Price War to Crosschain Cannibalization*, Promotor(s): Prof.dr. M.G. Dekimpe & Prof.dr.ir. B. Wierenga, EPS-2010-203-MKT, http://repub.eur.nl/pub/19714

Sousa, M.J.C. de, *Servant Leadership to the Test: New Perspectives and Insights*, Promotor(s): Prof.dr. D.L. van Knippenberg & Dr. D. van Dierendonck, EPS-2014-313-ORG, http://repub.eur.nl/pub/51537

Spliet, R., *Vehicle Routing with Uncertain Demand*, Promotor(s): Prof.dr.ir. R. Dekker, EPS-2013-293-LIS, http://repub.eur.nl/pub/41513

Srour, F.J., *Dissecting Drayage: An Examination of Structure, Information, and Control in Drayage Operations*, Promotor(s): Prof.dr. S.L. van de Velde, EPS-2010-186-LIS, http://repub.eur.nl/pub/18231

Staadt, J.L., *Leading Public Housing Organisation in a Problematic Situation: A Critical Soft Systems Methodology Approach*, Promotor(s): Prof.dr. S.J. Magala, EPS-2014-308-ORG, http://repub.eur.nl/pub/50712

Stallen, M., *Social Context Effects on Decision-Making: A Neurobiological Approach*, Promotor(s): Prof.dr.ir. A. Smidts, EPS-2013-285-MKT, http://repub.eur.nl/pub/39931

Tarakci, M., *Behavioral Strategy: Strategic Consensus, Power and Networks*, Promotor(s): Prof.dr. D.L. van Knippenberg & Prof.dr. P.J.F. Groenen, EPS-2013-280-ORG, http://repub.eur.nl/pub/39130

Teixeira de Vasconcelos, M., *Agency Costs, Firm Value, and Corporate Investment*, Promotor(s): Prof.dr. P.G.J. Roosenboom, EPS-2012-265-F&A, http://repub.eur.nl/pub/37265

Tempelaar, M.P., *Organizing for Ambidexterity: Studies on the Pursuit of Exploration and Exploitation Through Differentiation, Integration, Contextual and Individual Attributes*, Promotor(s): Prof.dr.ing. F.A.J. van den Bosch & Prof.dr. H.W. Volberda, EPS-2010- 191-STR, http://repub.eur.nl/pub/18457

Tiwari, V., *Transition Process and Performance in IT Outsourcing: Evidence from a Field Study and Laboratory Experiments*, Promotor(s): Prof.dr.ir. H.W.G.M. van Heck & Prof.mr.dr. P.H.M. Vervest, EPS-2010-201-LIS, http://repub.eur.nl/pub/19868

Troster, C., *Nationality Heterogeneity and Interpersonal Relationships at Work*, Promotor(s): Prof.dr. D.L. van Knippenberg, EPS-2011-233-ORG, http://repub.eur.nl/pub/23298

Tsekouras, D., *No Pain No Gain: The Beneficial Role of Consumer Effort in Decision-Making*, Promotor(s): Prof.dr.ir. B.G.C. Dellaert, EPS-2012-268-MKT, http://repub.eur.nl/pub/37542

Tuijl, E. van, *Upgrading Across Organisational and Geographical Configurations*, Promotor(s): Prof.dr. L. van den Berg, EPS-2015-349-S&E, http://repub.eur.nl/pub/78224

Tuncdogan, A., *Decision Making and Behavioral Strategy: The Role of Regulatory Focus in Corporate Inno-*

vation Processes, Promotor(s): Prof.dr.ing. F.A.J. van den Bosch, Prof.dr. H.W. Volberda, & Prof.dr. T.J.M. Mom, EPS-2014-334-S&E, http://repub.eur.nl/pub/76978

Tzioti, S., *Let Me Give You a Piece of Advice: Empirical Papers About Advice Taking in Marketing*, Promotor(s): Prof.dr. S.M.J. van Osselaer & Prof.dr.ir. B. Wierenga, EPS- 2010-211-MKT, http://repub.eur.nl/pub/21149

Uijl, S. den, *The Emergence of De-Facto Standards*, Promotor(s): Prof.dr. K. Blind, EPS-2014-328-LIS, http://repub.eur.nl/pub/77382

Vaccaro, I.G., *Management Innovation: Studies on the Role of Internal Change Agents*, Promotor(s): Prof.dr.ing. F.A.J. van den Bosch, Prof.dr. H.W. Volberda, & Prof.dr. J.J.P. Jansen, EPS-2010-212-STR, http://repub.eur.nl/pub/21150

Vagias, D., *Liquidity, Investors and International Capital Markets*, Promotor(s): Prof.dr. M.A. van Dijk, EPS-2013-294-F&A, http://repub.eur.nl/pub/41511

Veelenturf, L.P., *Disruption Management in Passenger Railways: Models for Timetable, Rolling Stock and Crew Rescheduling*, Promotor(s): Prof.dr. L.G. Kroon, EPS-2014-327- LIS, http://repub.eur.nl/pub/77155

Venus, M., *Demystifying Visionary Leadership: In Search of the Essence of Effective Vision Communication*, Promotor(s): Prof.dr. D.L. van Knippenberg, EPS-2013-289-ORG, http://repub.eur.nl/pub/40079

Verheijen, H.J.J., *Vendor-Buyer Coordination in Supply Chains*, Promotor(s): Prof.dr.ir. J.A.E.E. van Nunen, EPS-2010-194-LIS, http://repub.eur.nl/pub/19594

Visser, V.A., *Leader Affect and Leadership Effectiveness: How Leader Affective Displays Influence Follower Outcomes*, Promotor(s): Prof.dr. D.L. van Knippenberg, EPS-2013-286-ORG, http://repub.eur.nl/pub/40076

Vlam, A.J., *Customer First? The Relationship between Advisors and Consumers of Financial Products*, Promotor(s): Prof.dr. Ph.H.B.F. Franses, EPS-2011-250-MKT, http://repub.eur.nl/pub/30585

Waard, E.J. de, *Engaging Environmental Turbulence: Organizational Determinants for Repetitive Quick and Adequate Responses*, Promotor(s): Prof.dr. H.W. Volberda & Prof.dr. J. Soeters, EPS-2010-189-STR, http://repub.eur.nl/pub/18012

Waltman, L., *Computational and Game-Theoretic Approaches for Modeling Bounded Rationality*, Promotor(s): Prof.dr.ir. R. Dekker & Prof.dr.ir. U. Kaymak, EPS-2011-248-LIS, http://repub.eur.nl/pub/26564

Wang, T., *Essays in Banking and Corporate Finance*, Promotor(s): Prof.dr. L. Norden & Prof.dr. P.G.J. Roosenboom, EPS-2015-352-F&A, http://repub.eur.nl/pub/78301

Wang, Y., *Information Content of Mutual Fund Portfolio Disclosure*, Promotor(s): Prof.dr. M.J.C.M. Verbeek, EPS-2011-242-F&A, http://repub.eur.nl/pub/26066

Wang, Y., *Corporate Reputation Management: Reaching Out to Financial Stakeholders*, Promotor(s): Prof.dr. C.B.M. van Riel, EPS-2013-271-ORG, http://repub.eur.nl/pub/38675

Weenen, T.C., *On the Origin and Development of the Medical Nutrition Industry*, Promotor(s): Prof.dr. H.R. Commandeur & Prof.dr. H.J.H.M. Claassen, EPS-2014-309-S&E, http://repub.eur.nl/pub/51134

Wolfswinkel, M., *Corporate Governance, Firm Risk and Shareholder Value*, Promotor(s): Prof.dr. A. de Jong, EPS-2013-277-F&A, http://repub.eur.nl/pub/39127

Wubben, M.J.J., *Social Functions of Emotions in Social Dilemmas*, Promotor(s): Prof.dr. D. de Cremer & Prof.dr. E. van Dijk, EPS-2010-187-ORG, http://repub.eur.nl/pub/18228

Xu, Y., *Empirical Essays on the Stock Returns, Risk Management, and Liquidity Creation of Banks*, Promotor(s): Prof.dr. M.J.C.M. Verbeek, EPS-2010-188-F&A, http://repub.eur.nl/pub/18125

Yang, S., *Information Aggregation Efficiency of Prediction Markets*, Promotor(s): Prof.dr.ir. H.W.G.M. van Heck, EPS-2014-323-LIS, http://repub.eur.nl/pub/77184

Zaerpour, N., *Efficient Management of Compact Storage Systems*, Promotor(s): Prof.dr.ir. M.B.M. de Koster, EPS-2013-276-LIS, http://repub.eur.nl/pub/38766

Zhang, D., *Essays in Executive Compensation*, Promotor(s): Prof.dr. I. Dittmann, EPS- 2012-261-F&A, http://repub.eur.nl/pub/32344

Zhang, X., *Scheduling with Time Lags*, Promotor(s): Prof.dr. S.L. van de Velde, EPS- 2010-206-LIS, http://repub.eur.nl/pub/19928

Zhou, H., *Knowledge, Entrepreneurship and Performance: Evidence from Country-Level and Firm-Level Studies*, Promotor(s): Prof.dr. A.R. Thurik & Prof.dr. L.M. Uhlaner, EPS- 2010-207-ORG, http://repub.eur.nl/pub/20634

Zwan, P.W. van der, *The Entrepreneurial Process: An International Analysis of Entry and Exit*, Promotor(s): Prof.dr. A.R. Thurik & Prof.dr. P.J.F. Groenen, EPS-2011-234-ORG, http://repub.eur.nl/pub/23422