KAR YIN LAM

Reliability and Rankings

Reliability and Rankings

Reliability and Rankings

Betrouwbaarheid en Rankings

Thesis

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

Prof.dr. H.G. Schmidt

and in accordance with the decision of the Doctorate Board.

The public defence shall be held on
Thursday 14 April 2011 at 13:30 hrs

by

Kar Yin Lam
born in Zeist.

Doctoral Committee

Promoter:	Prof.dr. P.H.B.F. Franses
Other members:	Prof.dr. P.J.F. Groenen Prof.dr. J. de Leeuw Prof.dr. J.J. Louviere
Copromotor:	Dr. A.J. Koning

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://hdl.handle.net/1765/1

ERIM PhD Series in Research in Management, 230
ERIM reference number: EPS-2011-230-MKT
ISBN 978-90-5892-272-4

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

Acknowledgements

All things come to an end and the time has come to write the final words of this thesis. I have been thinking about this moment a lot, not because I was hankering for it but because I wish to express my gratitude to the people who supported me during my PhD journey.

I am deeply indebted to my promotor Philip Hans Franses and my copromotor Alex Koning. Philip Hans, thank you for giving me the opportunity to start the PhD. Your everlasting enthusiasm and optimism pulled me through the difficult moments in the PhD. In your extremely busy agenda, there was always time to discuss my research. Alex, it is hard to express how much gratitude I owe you. You have done so much more than I could ever ask for from a supervisor. Over the years, we have spent more and more time on non-work related talks and drinking coffee. I am impressed by your barista skills.

I would like to thank the members of the doctoral committee, Patrick Groenen, Jan de Leeuw, and Jordan Louviere, for making time available to evaluate this thesis. Your comments and feedback were very useful and highly appreciated.

This thesis would not have been possible without the support of ERIM and the Econometric Institute. I thank the staff of both institutes for their help and kind support throughout the years. Besides, I want to mention Joost for sharing his LaTeX template and Iris for her help. I would also like to thank Adriana Gabor for her encouragements and the pleasant collaboration on teaching.

Warm thanks to my fellow PhD students. Arco, I was very lucky to have you as my officemate. I really enjoyed the nice time we had during the whole PhD period. Hans, I can't thank you enough for all your help and advices. Peter, you were always in for a never-ending chat. Eelco and Yuri, I was happy to have your accompany at the conferences. Mathijn, thanks for the nice lunches. I also thank Carlos, Joris, Karim, Remi, Ron, Sjoerd, Sven, Tim, Twan, and Wei.

A big thank you to my friends. Wei Yie and Wei Ying, for your sincere interest and fun outside the PhD life. Wei Yie, your visits to the university and the too long coffee

breaks, lunches and discussions really meant a lot to me. Agatha, since the first day at university you helped me get through the difficult times when I was struggling with the studies. Also, special thanks to Jake for his input and contribution.

Finally, I wish to thank my family for their interest, patience and support during my PhD. Karli, thank you for pretending to understand when I told you about the great results and the not so great results. My deepest thank goes to my parents for their unconditional trust and support.

Kar Yin Lam
Rotterdam, January 2011

Contents

Chapter 1

Introduction

1.1 Motivation

Questionnaires are an important way to gather information about large populations for both qualitative and quantitative research. Hence, the value of a good questionnaire design and the quality of questionnaire data cannot be emphasized enough. This thesis discusses some aspects of the statistical analysis of measurement data obtained via questionnaires.

The quality of the questionnaire data depends on the precision of the measurements. Typically, a questionnaire is composed of multiple items, where each item is an independent attempt to measure the same underlying construct of interest. For example, multiple statements independently measure consumer satisfaction. A measured test score can be decomposed into an unobserved true score and a random error (Lord & Novick, 1968). The amount of agreement between the observed test score and the unobserved true score is an indicator of the reliability of the measurements of the items. Since the true score cannot be observed and only the test score is directly observable, we are unable to directly estimate the reliability of a test. Fortunately, we may use the dependence between the individual item scores in order to indirectly evaluate the reliability of a test.

The most widely used indirect measure for assessing scale reliability is Cronbach's coefficient alpha (Cronbach, 1951). In special cases coefficient alpha coincides with reliability, but actually it should be regarded as a lower bound of reliability (see Novick & Lewis, 1967). One can question why one should use a lower bound and not maximize the reliability coefficient (see for instance Li, Rosenthal, & Rubin, 1996). Several methods allow for improving reliability by using a weighted sum of the individual item scores as a test score rather than an unweighted sum (see for example Knott & Bartholomew, 1993;

Li, 1997; Yuan & Bentler, 2002; Raykov & Penev, 2006). The optimally weighted test score is referred to as the maximal reliability of the test (Lindsey, 1996).

A crucial issue concerns the evaluation of the value of each reliability measure that is, which values are high and which are low. For example, the adequacy of an obtained alpha is often based on Nunnally's recommendations (Nunnally & Bernstein, 1994) regarding minimally acceptable reliability, either in support or criticism. In much marketing research, rules of thumb are used, like "more than 0.8" is good. This practice is of course not very rigorous. Particularly, nothing is known about the accuracy of the measure, and moreover no judgements can be made about the significance of differences between obtained reliability measures. Cortina (1993) mentions that there seems to be no real metric for judging the adequacy of the statistic. The answer to this is to provide confidence intervals for the test statistic. The additional information included in the interval allows for a more critical assessment of the statistic. In this thesis we stress the importance of confidence intervals accompanying single measures for summarizing the parameters of a model.

In the first part of this thesis we focus on maximizing scale reliability. We derive the asymptotic distribution of maximal reliability measures to construct confidence intervals in order to assess the adequacy of the measure. We design various simulation experiments to demonstrate the adequacy of these results, and we also illustrate the merits of our findings using actual survey data. The results can lead to better designs of questionnaires, which in turn lead to more precise survey outcomes.

The second part of this thesis proposes methodologies to perform statistical analysis of stated consumer preferences measured as rankings data, especially in the context of conjoint measurements. Metric measurements such as rating and matching may be less reliable due to respondents' limited ability to accurately report degrees of preferences (Ben-Akiva, Morikawa, & Shiroishi, 1992). That is, respondents are able to prefer alternative A over B, however they are rarely able to express how much more they prefer A over B. Rankings are scale-free measures that are often used in marketing to evaluate products, brands, and services. Other examples are rankings of academic journals, business schools and so on. Rankings are easy to understand and easy to collect. However, task difficulty increases substantially with the number of alternatives to be ranked. It has been widely recognized that individuals face difficulties, or even become dissatisfied, when having to compare too many choice options. Iyengar and Lepper (2000) in their famous experiment showed that too much choice can be de-motivating for consumers. In a marketing context, Boatwright and Nunes (2001) demonstrated that a reduction of the

assortment is in fact felt as beneficial to consumers, a finding which has been supported by Chernev (2003) and Gourville and Soman (2005), among others. In another context, DeShazo and Fermo (2002) showed that consumers experience task complexity when choice options are plenty, see also Sandor and Franses (2009). In sum, consumers may find it difficult to rank preferences across too large an amount of alternatives. The problem of task complexity can be alleviated by asking respondents to evaluate all alternatives but only to give preference rankings for a subset of the alternatives. We shall refer to such a ranking as a partial ranking.

Partial rankings are essential in practical conjoint analysis to collect data efficiently to relieve respondents' task burden. Moreover, as in Louviere, Henser, and Swait (2000), one does not need complete rankings to obtain extra preference information. When collecting conjoint measurement data, individuals often need to compare many objects on many attributes with each many levels. This is a difficult task for various reasons mentioned above. Consider a consumer who is asked to compare products and the product with the highest utility is appreciated most. However, the utilities are latent in the sense that the consumer is unable to express them, but nevertheless influence his/her behavior.

There are at least two advantages of our approach to analyze partial rankings. First, our statistical models allow for the efficient use of partial rankings to collect preference data. Partial rankings may also be dictated by the experimental design of the conjoint study. Second, given that a partial ranking task amounts to a smaller burden for respondents than a complete ranking task, they may be more motivated to complete the task and as such the quality of the obtained data may improve. This certainly will help marketers to identify and target consumers by understanding their preference behavior, and to implement a more efficient and optimal marketing strategy.

1.2 Outline

This thesis can be divided into two parts. The first part consist of Chapter 2 and Chapter 3, and focuses on the reliability of questionnaires. Chapter 2 is based on Lam, Koning, and Franses (2009) and Chapter 3 is joint work with Alex Koning and Philip Hans Franses. In these chapters we bridge the gap between on the one side the very restrictive parallel model and at the other side the extremely permissive model, without any assumptions made on the underlying structure of items' variance. The motivation lies in discovering the structure of the covariance matrix of the items. Ignoring the structure in the covariance matrix will neglect valuable information. Without any

assumptions, nothing can be said about the true score variance and error variance of the items as these parameters are not estimated. However, these unique variances for each of the item components are particularly interesting to improve measurement scales and contain valuable information. One should always prefer a model which provides the most information. We discuss the issue of estimating the maximal reliability in the tau-equivalent and the congeneric model and acquire insights in how corresponding confidence intervals can be improved. We point out possible anti-conservatism of existing confidence intervals for maximal reliability, which may give the impression that we have estimated with sufficient accuracy. Our simulation experiments show that this indeed occurs. We derive new confidence intervals for the maximal reliability and compare them to existing intervals by means of coverage curve analysis. It turns out that confidence intervals where variance stabilization techniques are applied, are less prone to Type II error. This interval has a coverage for the true value which is approximately equal to the confidence level. We advocate the use of variance stabilization also in more complicated models to improve confidence intervals.

The second part of the thesis consists of the remaining chapters and focuses on the analysis of partial rankings. All cases that may arise in the assignment of ranks, such as the existence of ties and missing values, are incorporated. In Chapter 4 we propose a tool which examines whether the overall preference rankings differ from randomness. Moreover, we introduce a multiple comparison procedure to examine which alternatives actually differ. In a preliminary analysis of preference rankings when there are too many alternatives, the proposed methodology reduces the amount of alternatives in an easy and understandable way, such that irrelevant alternatives can be removed from further analysis. This chapter is partly based on Lam, Koning, and Franses (2010a).

In subsequent chapters, we analyze rankings data in the context of conjoint measurement analysis. As there is only a finite number of possible rankings, the rankings have a discrete distribution. In principle, standard methods for analyzing discrete data apply here, see Marden (1995, p. 140). However, probability models for rankings become very complex, as the computation of each ranking probability usually requires high-dimensional integration when the number of stimuli becomes large. In the literature an abundance of ranking models is available, see Critchlow, Fligner, and Verducci (1991, Section 3) and also Marden (1995, Chapter 5). In Chapter 5, we show that any Thurstone order statistics model may be approximated by a simpler model and this allows for a unified approach. We adapt general ranking models to conjoint experiments by introducing a

linear model which allows for modeling the dependence of the rankings on the stimulus characteristics. Chapter 5 is joint work with Alex Koning and Philip Hans Franses.

In marketing and other social sciences it is generally assumed that individuals are heterogeneous in their choice and preference behavior. To adequately model this heterogeneity, in Chapter 6 we present a finite mixture ranking model for analyzing full and partial ranking data in the presence of respondent heterogeneity. We show that our model is able to extract sufficient preference information from partial ranking data to take respondents' heterogeneity into account. Chapter 6 is joint work with Alex Koning and Philip Hans Franses.

All chapters contain various illustrations to demonstrate how the methods work. In the final chapter of this thesis we conclude with an overview of our main findings in the individual chapters, and a discussion of the implications, limitations and recommendations for further research.

Part I

Reliability

Chapter 2

Confidence Intervals for Maximal Reliability in Tau-Equivalent Models*

Abstract

Subjective probabilities play an important role in marketing research, for example where individuals rate the likelihood that they will purchase a new to develop product. The tau-equivalent model can describe the joint behavior of multiple test items measuring the same subjective probability. In this paper we stress the use of confidence intervals to assess reliability, as this allows for a more critical assessment of the items as measurement instruments. To improve the reliability one can use a weighted sum as the outcome of the test rather than an unweighted sum. In principle, the weights may be chosen so as to obtain maximal reliability. We propose two new confidence intervals for the maximal reliability in the tau-equivalent model and we compare these two new intervals to intervals derived earlier in Yuan and Bentler (2002) and Raykov and Penev (2006). The comparison involves coverage curves, a methodology that is new in the field of reliability. The existing Yuan-Bentler and Raykov-Penev intervals are shown to overestimate the maximal reliability, whereas one of our proposed intervals, the stable interval, performs very well. This stable interval hardly shows any bias, and has a coverage for the true value which is approximately equal to the confidence level.

2.1 Introduction and motivation

Marketing researchers often conduct tests to learn about individuals' preferences, opinion and attitudes. Opinions may often be expressed by subjective probabilities. A person's

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

subjective probability is the degree of belief a person holds regarding a statement or an event, that is, it is a person's personal judgment about how likely a particular statement is or how likely a particular event is to occur. Subjective probabilities play a role in marketing, for example when potential consumers are asked how likely it is that they will purchase a new to develop product.

Most subjective probability scales are similar to psychological measurement scales, see Wallsten and Budescu (1983). In line with classical test theory, in subjective probability theory it is assumed that the observed test score U can be decomposed into an unobserved true score T and a random error score.

The reliability of the test is defined to be the amount of agreement between U and T, as captured by the squared correlation, see Lord and Novick (1968). Since, T cannot be observed, we are unable to estimate the reliability of a test directly. Therefore, a test is composed of multiple items. Each item score Y_j is an independent attempt to measure the same construct of interest T. The observed test score U is usually the unweighted sum of the individual item scores Y_j. We may use the dependence between the individual item scores in order to evaluate the reliability of a test indirectly.

Cronbach's coefficient alpha, see Cronbach (1951), the most widely used indirect technique for assessing reliability, compares the variance of the test score with the sum of the variances of the individual item scores. Cronbach's alpha should be regarded as a lower bound of the reliability of a test, and in certain special situations it coincides with reliability, see Novick and Lewis (1967).

Despite its popularity, the interpretation of Cronbach's alpha in practice is quite arbitrary. Nunnally's thresholds, see Nunnally and Bernstein (1994) are often taken as recommendations regarding minimally acceptable reliability, although one may argue that it is rather subjective to compare alpha to an arbitrary threshold.

Moreover, such an approach does not take into account the accuracy of the estimated alpha. In this paper we advocate the use of confidence intervals to assess reliability, as the additional information included in the interval allows for a more critical assessment of the statistic.

In the literature we find two types of intervals for Cronbach's alpha. The first type is derived under the so-called parallel model. In the parallel model each of the item scores Y_j is assumed to be the sum of the true score T and a measurement error ϵ_j with population mean zero and common population variance ψ. Moreover, the random variables $T, \epsilon_1, \ldots, \epsilon_k$ are assumed to be independent. As a consequence, the parallel model imposes strong restrictions on the population covariance matrix of the item scores. The

statistical theory needed to derive confidence intervals for Cronbach's alpha in the parallel model appeared in Kristof (1963); Feldt (1965), but it remained relatively unnoticed.

Van Zyl, Neudecker, and Nel (2000) draws new attention for confidence intervals for Cronbach's alpha, see Iacobucci and Duhachek (2003); Duhachek, Coughlan, and Iacobucci (2005), and leads to a second type of confidence intervals for Cronbach's alpha. These intervals are derived under the so-called saturated model, which does not impose any restriction on the population covariance matrix of the item scores.

It should be stressed that between the extremely restrictive parallel model on the one hand, and the extremely permissive saturated model on the other hand, other models exist. As both extreme cases have their problems, in this paper we restrict ourselves to the so-called tau-equivalent model. The tau-equivalent model can be obtained from the parallel model by relaxing the assumption that the measurement errors ϵ_j have common population variance ψ. Thus, the respective population variances $\psi_1, \ldots, \psi_k$ of the measurement errors $\epsilon_1, \ldots, \epsilon_k$ may differ. In other words, the tau-equivalent model allows nonhomogeneous error variances.

By measuring the probability that a same event will occur with different methods, one can say something about the quality of the measured probability. Hence, the tau-equivalent model can describe the joint behavior of multiple test items measuring the same subjective probability. In case of subjective probabilities, a tau-equivalent model is more plausible than the extremely restrictive parallel model and also more informative than the extremely permissive saturated model.

In the parallel model as well as in the tau-equivalent model, Cronbach's alpha coincides with reliability. By using a weighted sum W rather than an unweighted sum U as the outcome of the test reliability can be improved. One may show that choosing the weight for Y_j equal to $1/\psi_j$ yields a test score $W = \sum_{j=1}^{k} Y_j/\psi_j$ which has maximal reliability, which is

$$\varphi = 1 - \frac{1}{1 + \sum_{j=1}^{k} \lambda^2/\psi_j},$$

where λ^2 is the true score variance. The quantity φ is referred to as the maximal reliability of $Y_1, \ldots, Y_k$.

Remark that under the parallel model the optimal weights have a common value $1/\psi$, which implies that the unweighted sum U yields maximal reliability. Thus, in the parallel model maximal reliability coincides with reliability and Cronbach's alpha.

In this paper we aim to derive confidence intervals for maximal reliability in the tau-equivalent model. In Yuan and Bentler (2002) and Raykov and Penev (2006) different expressions for the standard deviation of the maximum likelihood estimator $\hat{\varphi}$ of the

maximal reliability φ are given, from which confidence intervals for φ are readily derived. However, as they propose an unstable variance this may lead to anti-conservative confidence intervals. We propose new methods to derive confidence intervals and compare two new derived intervals to the Yuan-Bentler and Raykov-Penev intervals.

The structure of this paper is as follows. In Section 2.2 we present the four confidence intervals, and show that their coverage is asymptotically equal to the requested confidence level. In Section 2.3 we apply the four confidence intervals to real data involving measures of subjective probability. Section 2.4 discusses the results of an extensive simulation experiment. In Section 2.5 conclusions are drawn. Proofs are relegated to the Appendix.

2.2 Confidence intervals for the tau-equivalent model

The maximum likelihood estimator of φ is given by

$$\hat{\varphi} = 1 - \frac{1}{1+\hat{\zeta}},$$

with

$$\hat{\zeta} = \sum_{j=1}^{k} \hat{\lambda}^2/\hat{\psi}_j,$$

where $\hat{\psi}_j$ and $\hat{\lambda}^2$ are the maximum likelihood estimators of ψ_j and λ in the tau-equivalent model. The confidence intervals for φ involve

$$s^2 = 2 + 2\frac{\hat{\zeta}+1}{\hat{\zeta}-1}\frac{\hat{Q}}{1+\hat{Q}},$$

with

$$\hat{Q} = \frac{\hat{\zeta}-1}{\hat{\zeta}^2}\sum_{j=1}^{k}\frac{\left\{\hat{\lambda}^2/\hat{\psi}_j\right\}^2}{1+\hat{\zeta}-2\hat{\lambda}^2/\hat{\psi}_j},$$

and let n denote the sample size. Choose a confidence level $0 < \gamma < 1$, and determine $z_{(1-\gamma)/2}$ so such that

$$P\left(-z_{(1-\gamma)/2} < Z < z_{(1-\gamma)/2}\right) = \gamma.$$

with Z denoting a standard normal random variable. Next, we present several asymptotic $100\gamma\%$ confidence intervals for the maximal reliability coefficient φ.

2.2.1 Available results

In Yuan and Bentler (2002) a confidence interval for maximal reliability is proposed in general terms. According to Raykov and Penev (2006) it is "(a) rather laborious and tedious in routine behavioral research in need of interval estimation of maximal reliability, (b) involves taking by the researcher of multiple partial derivatives of this reliability coefficient with respect to model parameters, (c) has the inconvenient property that the number of these derivatives increases with increasing length k of the initial composite of interest (as could be repeatedly the case when one is involved in scale development and revision), and (d) can be viewed as based on a first-order approximation of maximal reliability as a function of model parameters". However, by taking the computations in Yuan and Bentler (2002) for the tau-equivalent model, the problem of taking multiple partial derivatives is eliminated. This yields that the endpoints of the Yuan-Bentler interval are given by

$$\hat{\varphi} \pm z_{(1-\gamma)/2}\,(1-\hat{\varphi})\,\frac{s}{\sqrt{n}} \tag{2.1}$$

Raykov and Penev (2006) proposed a "second-order" alternative to the Yuan-Bentler interval. In the tau-equivalent model, the endpoints of the Raykov-Penev interval are given by

$$\hat{\varphi} \pm z_{(1-\gamma)/2}\,(1-\hat{\varphi})\,\frac{s}{\sqrt{n}}\sqrt{1+2\frac{s^2}{n}}. \tag{2.2}$$

Note that the Yuan-Bentler interval is always contained in the Raykov-Penev interval.

2.2.2 The unstable interval

If n tends to infinity, it is shown in the Appendix that

$$\sqrt{n}\frac{\hat{\zeta}-\zeta}{(1+\zeta)\sqrt{2+2\frac{\zeta+1}{\zeta-1}\frac{Q}{1+Q}}} \rightarrow_d N(0,1)\,, \tag{2.3}$$

with

$$\zeta = \sum_{j=1}^{k} \lambda/\psi_j, \tag{2.4}$$

$$Q = \frac{\zeta-1}{\zeta^2}\sum_{j=1}^{k}\frac{\{\lambda/\psi_j\}^2}{1+\zeta-2\lambda^2/\psi_j}. \tag{2.5}$$

Note that s is a consistent estimator of $\sqrt{2+2\frac{\zeta+1}{\zeta-1}\frac{Q}{1+Q}}$. As

$$\varphi = 1 - \frac{1}{1+\zeta},$$

it follows from (2.3) that

$$\lim_{n\to\infty} P\left(-z_{(1-\gamma)/2}\frac{s}{(1-\hat{\varphi})\sqrt{n}} < \hat{\zeta} - \zeta < z_{(1-\gamma)/2}\frac{s}{(1-\hat{\varphi})\sqrt{n}}\right) = \gamma,$$

which implies that

$$\hat{\zeta} \pm z_{(1-\gamma)/2}\frac{s}{(1-\hat{\varphi})\sqrt{n}}$$

are the endpoints of an asymptotic $100\gamma\%$ confidence interval for ζ. As φ is a monotone function of ζ, we obtain that

$$\begin{aligned} 1 - \frac{1}{1+\hat{\zeta} \pm z_{(1-\gamma)/2}\frac{s}{(1-\hat{\varphi})\sqrt{n}}} &= 1 - \frac{1}{\frac{1}{(1-\hat{\varphi})} \pm z_{(1-\gamma)/2}\frac{s}{(1-\hat{\varphi})\sqrt{n}}} \\ &= 1 - \frac{1-\hat{\varphi}}{1 \pm z_{(1-\gamma)/2}\frac{s}{\sqrt{n}}} \end{aligned} \tag{2.6}$$

are the endpoints of an asymptotic $100\gamma\%$ confidence interval for φ. We shall refer to (2.6) as the unstable interval.

The unstable interval is related to the Yuan-Bentler interval in (2.1) and the Raykov-Penev interval in (2.2). Note that the derivative of $1-(1+\zeta)^{-1}$ with respect to ζ equals $(1+\zeta)^{-2}$. Combining (2.3) with the delta method yields that

$$\sqrt{n}\frac{\hat{\varphi}-\varphi}{(1-\varphi)\sqrt{2+2\frac{\zeta+1}{\zeta-1}\frac{Q}{1+Q}}} \to_d N(0,1).$$

It follows that

$$\lim_{n\to\infty} P\left(-z_{(1-\gamma)/2}(1-\hat{\varphi})\frac{s}{\sqrt{n}} < \hat{\varphi} - \varphi < z_{(1-\gamma)/2}(1-\hat{\varphi})\frac{s}{\sqrt{n}}\right) = \gamma,$$

which implies that

$$\hat{\varphi} \pm z_{(1-\gamma)/2}(1-\hat{\varphi})\frac{s}{\sqrt{n}}$$

are the endpoints of an asymptotic $100\gamma\%$ confidence interval for φ. This interval coincides with the Yuan-Bentler interval.

To obtain an asymptotic confidence interval for $\hat{\varphi} - \varphi$ in Raykov and Penev (2006), properties of the normal distribution and the second order approximation

$$\hat{\varphi} - \varphi = \frac{\hat{\zeta} - \zeta}{(1+\zeta)^2} - \frac{\left(\hat{\zeta} - \zeta\right)^2}{(1+\zeta)^3},$$

are needed. This yields

$$\lim_{n\to\infty} P\left(-z_{(1-\gamma)/2}\,(1-\hat{\varphi})\,\frac{s}{\sqrt{n}}\sqrt{1+2\frac{s^2}{n}} < \hat{\varphi} - \varphi < z_{(1-\gamma)/2}\,(1-\hat{\varphi})\,\frac{s}{\sqrt{n}}\sqrt{1+2\frac{s^2}{n}}\right) = \gamma,$$

which implies that

$$\hat{\varphi} \pm z_{(1-\gamma)/2}\,(1-\hat{\varphi})\,\frac{s}{\sqrt{n}}\sqrt{1+2\frac{s^2}{n}}$$

are the endpoints of an asymptotic $100\gamma\%$ confidence interval for φ. This interval coincides with the Raykov-Penev interval.

2.2.3 The stable interval

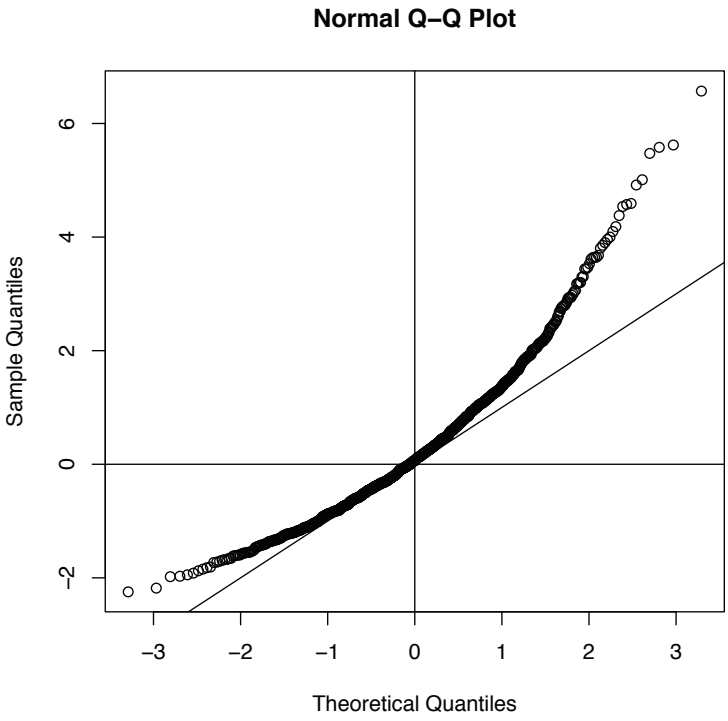
Although, the left hand side of (2.3) converges in distribution to a standard normal random variable as the number of observations n becomes large, it may well have a very different distribution for a given value of n. Figure 2.1 illustrates that this is indeed the case.

To motivate our remedy for this problem, we assume that the items of the test are drawn at random from a large test battery. That is, the values $\psi_1, \ldots, \psi_k$ are drawn independently from the distribution of some non-negative random variable, say E. In addition, we assume that the variance of E^{-1} is finite. It follows by (2.4) and (2.5) that the quantities ζ and Q depend on the number k of items drawn. In the appendix it is shown that

$$\lim_{k\to\infty} \frac{\zeta+1}{\zeta-1}\frac{Q}{1+Q} = 0. \tag{2.7}$$

It follows by (2.3) that for large k the asymptotic variance of $\sqrt{n}\left(\hat{\zeta} - \zeta\right)$ is approximated by $2\,(1+\zeta)^2$. Thus, the variance of $\hat{\zeta}$ depends on ζ, and this does not disappear when k tends to infinity. That is, the variance of $\hat{\zeta}$ is not stable. As the variance is proportional to $(1+\zeta)^2$ for k sufficiently large, the theory of variance stabilizing transformations suggests the use of $\ln\left(1+\hat{\zeta}\right)$ for constructing confidence intervals.

Figure 2.1: Normal probability plot of simulated values of the left hand side of (2.3) with $k = 6$, $\varphi = 0.6$, $n = 25$. The plot approaches the line with intercept 0 and slope 1 only in the center, there is a marked deviation in the tails.



It can be shown that

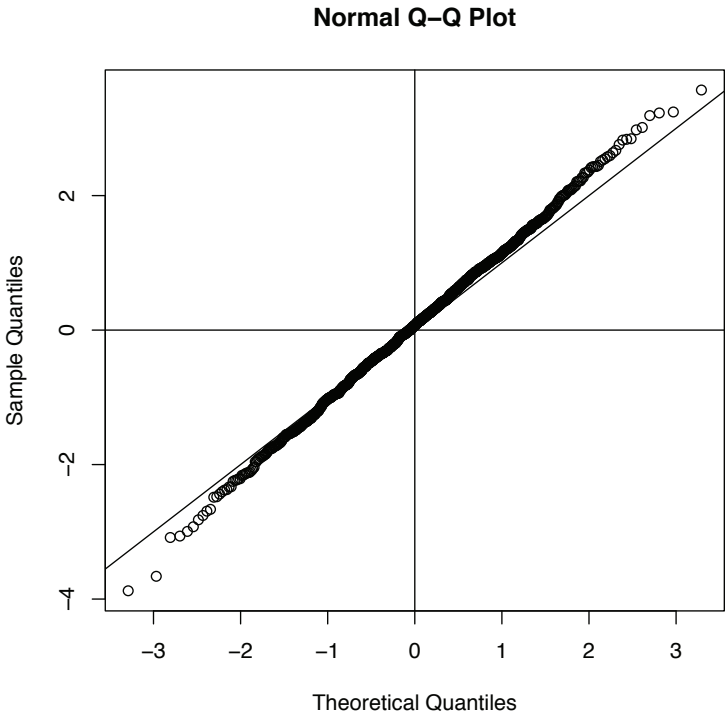
$$\sqrt{n}\frac{\ln\left(1+\hat{\zeta}\right)-\ln\left(1+\zeta\right)}{\sqrt{2+2\frac{\zeta+1}{\zeta-1}\frac{Q}{1+Q}}}\rightarrow_d N\left(0,1\right). \tag{2.8}$$

Figure 2.2 illustrates that the standard normal distribution provides a far better fit to the distribution of the left hand side of (2.8)) than to the distribution of the left hand side of (2.3).

It follows from (2.8) that

$$\lim_{n\rightarrow\infty} P\left(-z_{(1-\gamma)/2}\frac{s}{\sqrt{n}}<\ln\left(\frac{1+\hat{\zeta}}{1+\zeta}\right)<z_{(1-\gamma)/2}\frac{s}{\sqrt{n}}\right)=\gamma,$$

Figure 2.2: Normal probability plot of simulated values of the left hand side of (2.8), with $k = 6$, $\varphi = 0.6$, $n = 25$. The plot approaches the line with intercept 0 and slope 1, even in the tails.



which implies that

$$\left(1+\hat{\zeta}\right)\exp\left\{\pm z_{(1-\gamma)/2}\frac{s}{\sqrt{n}}\right\}-1$$

are the endpoints of an asymptotic $100\gamma\%$ confidence interval for ζ. As φ is a monotone function of ζ, we obtain that

$$1-\frac{\exp\left\{\pm z_{(1-\gamma)/2}\frac{s}{\sqrt{n}}\right\}}{1+\hat{\zeta}}=1-(1-\hat{\varphi})\exp\left\{\pm z_{(1-\gamma)/2}\frac{s}{\sqrt{n}}\right\} \tag{2.9}$$

are the endpoints of an asymptotic $100\gamma\%$ stable confidence interval for φ. This stable confidence interval is the interval which we wish to advocate in the present paper.

2.3 Illustration

In Wallsten and Budescu (1983) it is assumed that measures of subjective uncertainty can be written as the sum of two independent random variables, a fixed true measure and a variable error. Thus, if the subjective probability of a given event is measured by different methods, then a tau-equivalent model seems plausible.

In Ofir and Reddy (1996) the psychometric properties of three measures of subjective uncertainty are investigated. The 117 respondents were asked to express the subjective probability of the event "The stock market will rise during 1991 by at least 10%" on each of the following measurement scales:

StckP a seven-point "probabl" rating scale with categories Highly Improbable, Improbable, Somewhat Improbable, Equally Probable, Somewhat Probable, Probable and Highly Probable;

StckL a seven-point "likelihood" rating scale with categories Very Unlikely, Unlikely, Somewhat Unlikely, Equal Likelihood, Somewhat Likely, Likely, Very Likely.

StckC a subjective probability scale ranging from zero to one hundred.

The rating scales *StckP* and *StckL* were transformed to 0-100 scales by using the transformation $100(x-1)/6$, where x is the value on the seven-point scale.

Table 2.1 reports the covariance matrix of these three measures of the subjective probability of the event "The stock market will rise during 1991 by at least 10%." Indeed, the tau-equivalent model fits the data [$\chi^2 = 3.161$, df $= 2$, $P = 0.206$]. However, the parallel model is clearly rejected [$\chi^2 = 14.679$, df $= 4$, $P = 0.005$]. In the tau-equivalent model, the estimated true variance is 541.563, and the estimated error variances are 72.454, 67.065 and 162.962. Thus,

$$\frac{StckP}{72.454} + \frac{StckL}{67.065} + \frac{StckC}{162.962}$$

is an estimate of the weighted composite of *StckP*, *StckL* and *StckC* yielding maximal reliability.

The corresponding 95% confidence intervals for maximal reliability φ are found in Table 2.2. Note that the Yuan-Bentler and the Raykov-Penev intervals are symmetric around $\hat{\varphi}$, whereas the stable and the unstable interval are not. Moreover, the Raykov-Penev interval contains the Yuan-Bentler interval. The lower endpoint of the unstable interval is smaller than the lower endpoint of the stable interval, which in turn is smaller

Table 2.1: Covariance matrix of a zero-to-hundred subjective probability scale and two seven point rating scales measuring the subjective probability of the event "The stock market will rise during 1991 by at least 10%", compiled from Appendix A in Ofir & Reddy (1996). There are 117 respondents.

	StckP	*StckL*	*StckC*
StckP	638.790		
StckL	562.214	620.501	
StckC	509.735	501.765	619.956

Table 2.2: Asymptotic 95% confidence intervals for maximal reliability φ, derived from the data in Table 2.1. The estimator $\hat{\varphi}$ takes the value 0.9496805.

Interval	*Lower*	*Upper*
Unstable	0.926	0.962
YuanBentler	0.933	0.966
RaykovPenev	0.934	0.967
Stable	0.930	0.964

than the lower endpoint of the Raykov-Penev interval. Similarly, the upper endpoint of the unstable interval is smaller than the upper endpoint of the stable interval, which in turn is smaller than the upper endpoint of the Raykov-Penev interval.

2.4 Simulation results

The above illustration for actual data showed that there are differences between confidence intervals, but these do not necessarily have to be very large. To further our understanding of the differences between the four confidence bounds, we rely on simulation, to be discussed in this section.

In the simulations, we let k take the values $2, 3, 4, 5, 10, 15$, and n take the values $25, 50, 100, 200, 400$. We expect that the simulation results largely depend on the quantities μ_1 and μ_2, where

$$\mu_1 = \frac{1}{k}\sum_{j=1}^{k}\frac{1}{\psi_j}, \quad \mu_2 = \frac{1}{k}\sum_{j=1}^{k}\left(\frac{1}{\psi_j} - \mu_1\right)^2. \tag{2.10}$$

In order to be able to confirm this expectation, we generate the ψ_i's using various patterns. In particular, we choose

$$\frac{1}{\psi_j} = a + bg\left(\frac{j}{k+1}\right),$$

where g is one of the three following functions:

$$g_1(s) = s - \tfrac{1}{2},$$

$$g_2(s) = s^2 - \tfrac{1}{3},$$

$$g_3(s) = \begin{cases} -1 & \text{for } s < \frac{1}{2}, \\ +1 & \text{for } s \geq \frac{1}{2}, \end{cases}.$$

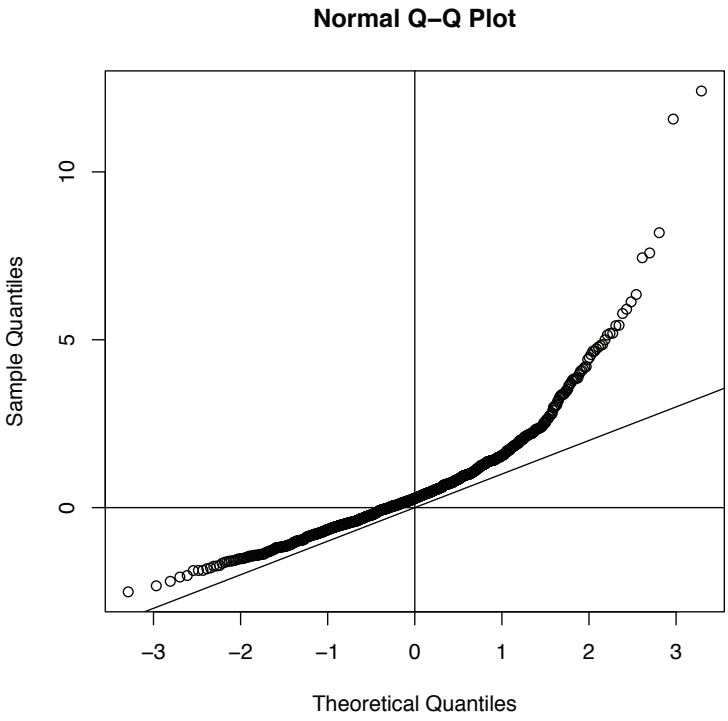
Here a and b are chosen so as to satisfy (2.10) for given μ_1 and μ_2. Throughout the simulations we set λ equal to 1. As $\mu_1 = (1/k)\sum_{j=1}^{k} 1/\psi_j$, we may show that $\mu_1 = (k\lambda^2)^{-1}\varphi/(1-\varphi)$. In our simulations we set μ_1 equal to values which correspond to maximal reliability 0.60, 0.75, 0.90 and 0.95. Finally, we set μ_2 equal to $\frac{1}{2}\mu_1^2$, μ_1^2 and $2\mu_1^2$.

First, we investigate the extent in which the standard normal approximations (2.8) and (2.3) are valid for small to moderate sample sizes n. As Figure 2.3 illustrates, these approximations are truly asymptotic in nature when the number of test items k is equal to 2. When $k = 3$, the approximations are reasonably accurate, except for the sample size $n = 25$. Recall that the standard normal approximations (2.8) and (2.3) provide the probabilistic basis for all confidence intervals discussed in this paper. Hence, in the remainder of this section we shall restrict ourselves to situations in which $k \geq 3$ and $n \geq 50$.

Next, we examine the simulated coverage curves of the four intervals. The coverage of a confidence interval for a given hypothetical value of φ is defined as the probability that this hypothetical value is contained in the interval. The coverage curve is the curve that the coverage follows as the hypothetical value of φ ranges through an interval of possible values of φ. The ideal shape of the coverage curve is as follows:

- if the hypothetical value of φ equals the true value, then the coverage should be equal to the confidence level;
- if the hypothetical value of φ differs from the true value, then the coverage should be as low as possible. In particular, the coverage should be lower than the confidence level.

Figure 2.3: Normal probability plot of simulated values of the left hand side of (2.8), with $k = 2$, $\varphi = 0.95$, $n = 400$. Although the sample size n is large, the plot clearly deviates from the line with intercept 0 and slope 1.



There are various ways in which a coverage curve may deviate from the ideal shape. In our examination of the performance of the confidence interval, we shall in particular consider the unbiasedness of a confidence interval. A confidence interval is called unbiased if for every hypothetical value of φ the coverage of the interval does not exceed the coverage for the true value of φ.

If an interval is unbiased, the next issue to consider is whether the coverage for the true value equals the confidence level. If this coverage is larger than the confidence interval, the unbiased confidence interval is called conservative; if this coverage is smaller than the confidence interval, the unbiased confidence interval is called anti-conservative. Being conservative is considered less harmful than being anti-conservative.

Statistical testing theory yields an alternative way to interpret coverage curves. Note that for every confidence interval for φ, there exists a related statistical test of the null hypothesis $H_0 : \varphi = \varphi_0$; this test does not reject the null hypothesis if the

hypothetical value φ_0 lies inside the confidence interval, and rejects the null hypothesis if the hypothetical value φ_0 is outside the confidence interval. Now, if the true value of φ differs from the hypothetical value φ_0, then the coverage of the hypothetical value φ_0 is equal to the probability of a type II error of the related statistical test of $H_0 : \varphi = \varphi_0$. Thus, by subtracting the coverage curve from 1, we in fact obtain the power curve of the related test. In fact, the description of the ideal shape of the coverage curve given above is a direct translation of generally accepted rules involving the ideal shape of the power curve.

Evaluating the coverage not only under the null hypothesis, but also under the alternative hypothesis provides a much more comprehensive view of the behavior of the various confidence intervals.

The simulated coverage curves depend on the sample size n, the true value of the maximal reliability φ, the number of items k and on the quantities μ_1 and μ_2.

- For fixed values of n, φ, k, μ_1 and μ_2, there is little difference between the coverage curves.

- The sample size n has a positive effect on the performance of all confidence intervals, see Figure 2.4. That is, the confidence intervals perform better for large n.

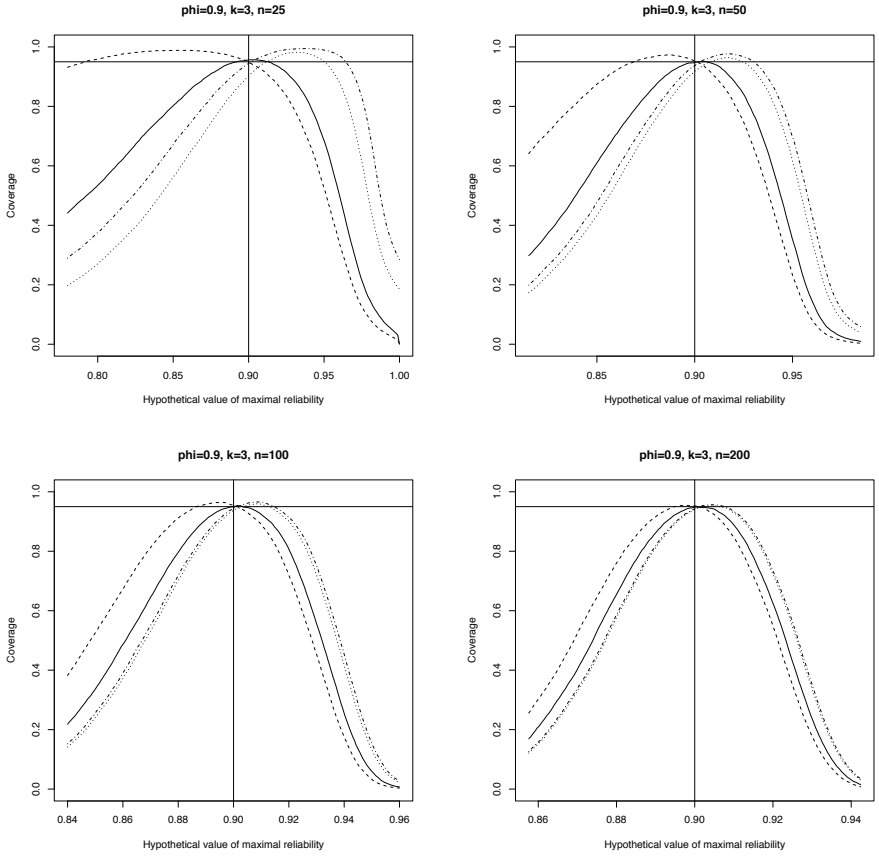
- The true value of the maximal reliability has a positive effect on the coverage of all confidence intervals, see Figure 2.5. That is, the coverage of the confidence intervals becomes higher when the true value of φ approaches 1. If the true value of the maximal reliability is small, the confidence intervals are anti-conservative. When the true value of φ approaches 1, the coverage of the true value increases. As a result, the anti-conservatism diminishes in most situations. However, in some situations (especially $k = 3$) the anti-conservatism turns into conservatism.

- The number of items k has a positive effect on the performance of all confidence intervals, see Figure 2.6. That is, the confidence intervals perform better for large k.

The simulated coverage curves yield the following general findings with respect to the differences in performance between the four confidence intervals.

- There is little difference between the Yuan-Bentler and the Raykov-Penev interval. Both intervals show a positive bias, and thus overestimate the the true maximal reliability. This bias is more marked when the true value of maximal reliability is low.

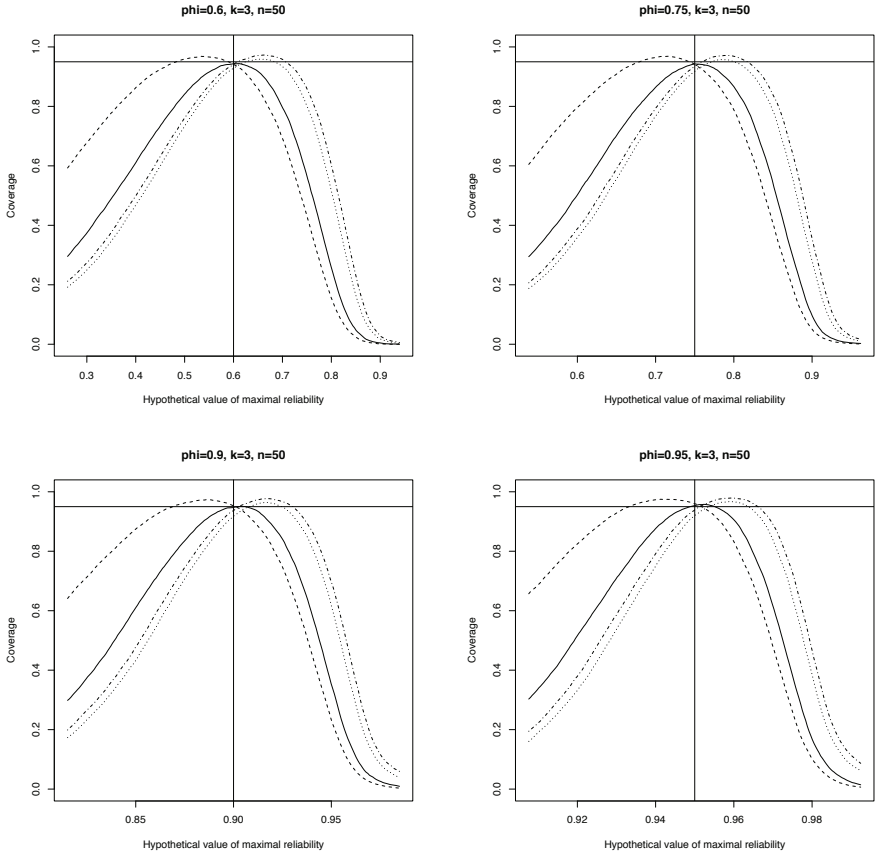
Figure 2.4: Coverage curves of stable interval (solid line), unstable interval (dashed line), Yuan-Bentler interval (dotted line) and Raykov-Penev interval (dotted-dashed line) for $\varphi = 0.90$, $k = 3$ and $n = 25, 50, 100, 200$.



- The unstable interval shows a negative bias, and thus underestimates the true maximal reliability.
- Except for extreme conditions (that is, a combination of a small sample size, a small number of items k and a high true value of the maximal variability), the stable interval hardly shows any bias, and has a coverage for the true value which is approximately equal to the confidence level.

Finally, we remark that we could not have detected the positive bias of the Yuan-Bentler and the Raykov-Penev without the construction of coverage curves, that is, eval-

Figure 2.5: Coverage curves of stable interval (solid line), unstable interval (dashed line), Yuan-Bentler interval (dotted line) and Raykov-Penev interval (dotted-dashed line) for $\varphi = 0.60, 0.75, 0.90, 0.95$, $k = 3$ and $n = 50$.

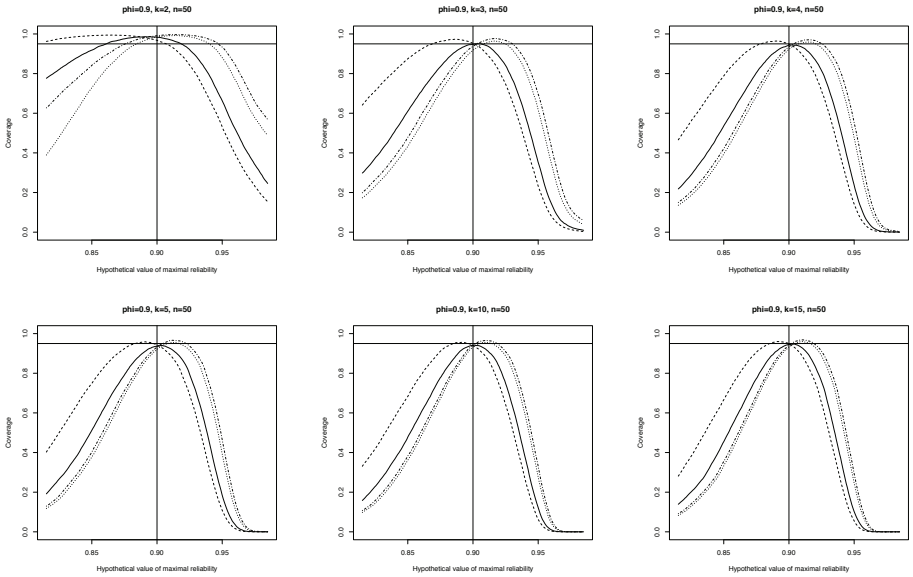


uating the coverage under the null hypothesis as well as under the alternative hypothesis. We highly recommend the use of coverage curves in other studies on confidence intervals.

2.5 Conclusion

We have shown that a tau-equivalent model is plausible when measuring subjective probabilities, which play an important role in marketing research. To improve the reliability of the test we use a weighted sum of individual item scores rather than an

Figure 2.6: Coverage curves of stable interval (solid line), unstable interval (dashed line), Yuan-Bentler interval (dotted line) and Raykov-Penev interval (dotted-dashed line) for $\varphi = 0.90$, $k = 2, 3, 4, 5, 10, 15$ and $n = 50$.



unweighted sum. In principle, the weights may be chosen so as to obtain maximal reliability, and these optimal weights may be estimated from the data.

We stressed the use of confidence intervals rather than point estimators to assess reliability, as the additional information included in the interval allows for a more critical assessment of the quality of the items as measurement instruments of the underlying subjective probability.

We have derived two new confidence intervals for maximal reliability in the tau-equivalent model. We motivated that the confidence intervals as derived in Yuan and Bentler (2002) and Raykov and Penev (2006) may be anti-conservative as they do not have a stable variance.

To compare the confidence intervals, we introduced coverage curves, a methodology that seems new in the field of reliability. That is, we have not only considered the coverage of the true maximal reliability, but the coverage of hypothetical values which differ from the true maximal reliability as well.

It turns out that the Yuan-Bentler and the Raykov-Penev intervals are closely related to each other. In fact, the Yuan-Bentler interval is always contained in the alternative

interval proposed in Raykov and Penev (2006). Interestingly, the similarity in behavior of the two intervals is in contrast with the remarks in Raykov and Penev (2006) with regard to the Yuan-Bentler interval. Moreover, it seems that the additional complexity of the Raykov-Penev interval does not pay off, as a clear advantage of using this interval over the Yuan-Bentler is lacking. Both intervals show a positive bias, which is more marked for low values of the true maximal reliability. This might be due to a ceiling effect that cuts down the positive biases of these two intervals when the true value of reliability is high.

We have also examined the performance of the two new intervals proposed in this article. Though the unstable interval shows a considerable negative bias, the stable interval performs considerably well. Except for extreme conditions the stable interval hardly shows any bias, and has a coverage for the true value which is approximately equal to the confidence level. This shows the advantage of the use of a stabilization technique in constructing confidence intervals. In further applications in marketing we therefore recommend the use of this new stable confidence interval.

The advantage of stabilization should not only hold in the tau-equivalent model. This raises the issue whether variance stabilization is of use in the parallel model as well. In the parallel model an unstable interval is given in Van Zyl et al. (2000), see also Iacobucci and Duhachek (2003). It would be interesting to compare those intervals using coverage curves. In addition, our results suggest that variance stabilization techniques might also improve confidence intervals for maximal reliability in more general models as congeneric models and factor analysis models.

2.A Proofs

Proof of (2.3)**.** The tau-equivalent model implies that the population covariance matrix of $Y_1, \ldots, Y_k$ is given by

$$\lambda^2 \boldsymbol{\iota}\boldsymbol{\iota}^\top + \boldsymbol{\Psi},$$

where $\boldsymbol{\iota}$ is the k-dimensional vector with each element having value 1, and $\boldsymbol{\Psi}$ the diagonal matrix with diagonal elements $\psi_1, \ldots, \psi_k$.

Let $\boldsymbol{\theta}$ denote the $k+1$-dimensional parameter vector $(\psi_1, \ldots, \psi_k, \lambda)^\top$, and let $\hat{\boldsymbol{\theta}}$ denote the maximum likelihood estimator of $\boldsymbol{\theta}$. It follows from the theory of covariance structures, see Browne (1984); Kano, Berkane, and Bentler (1990), that

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right) \to_d N_k\left(\mathbf{0}, \boldsymbol{\Omega}\right),$$

with

$$\boldsymbol{\Omega} = \begin{pmatrix} \boldsymbol{\Omega}_{\psi\psi} & \boldsymbol{\Omega}_{\psi\lambda} \\ \boldsymbol{\Omega}_{\lambda\psi} & \boldsymbol{\Omega}_{\lambda\lambda} \end{pmatrix}.$$

Here, the $1 \times k$ matrix $\boldsymbol{\Omega}_{\psi\lambda}$ and the $k \times 1$ matrix $\boldsymbol{\Omega}_{\lambda\psi}$ are given by

$$\boldsymbol{\Omega}_{\psi\lambda} = \boldsymbol{\Omega}_{\lambda\psi}{}^\top = -\boldsymbol{\Omega}_{\psi\psi}\left(\sum_{j=1}^{k}\left(1/\psi_j\right)\right)^{-2}\boldsymbol{\Psi}^{-2}\boldsymbol{\iota},$$

the 1×1 matrix $\boldsymbol{\Omega}_{\lambda\lambda}$ is given by

$$\boldsymbol{\Omega}_{\lambda\lambda} = 2\left(\frac{1+\zeta}{\sum_{j=1}^{k}\left(1/\psi_j\right)}\right)^2 + \left(\sum_{j=1}^{k}\left(1/\psi_j\right)\right)^{-4}\boldsymbol{\iota}^\top\boldsymbol{\Psi}^{-2}\boldsymbol{\Omega}_{\psi\psi}\boldsymbol{\Psi}^{-2}\boldsymbol{\iota},$$

and the $k \times k$ matrix $\boldsymbol{\Omega}_{\psi\psi}$ is given by

$$\boldsymbol{\Omega}_{\psi\psi} = 2\boldsymbol{\Psi}\left(\mathbf{D} + \mathbf{v}\mathbf{v}^\top\right)^{-1}\boldsymbol{\Psi},$$

with

$$\mathbf{D} = \left(\mathbf{I} - \frac{2\lambda^2}{1+\zeta}\boldsymbol{\Psi}^{-1}\right), \qquad \mathbf{v} = \frac{\lambda}{\zeta}\sqrt{\frac{\zeta-1}{\zeta+1}}\boldsymbol{\Psi}^{-1}\boldsymbol{\iota}. \tag{2.11}$$

The Sherman-Morrison-Woodbury formula, see Hager (1989), yields

$$\left(\mathbf{D} + \mathbf{v}\mathbf{v}^\top\right)^{-1} = \mathbf{D}^{-1} - \left(1 + \mathbf{v}^\top\mathbf{D}^{-1}\mathbf{v}\right)^{-1}\mathbf{D}^{-1}\mathbf{v}\mathbf{v}^\top\mathbf{D}^{-1}.$$

It follows that

$$\mathbf{v}^\top \mathbf{D}^{-1}\mathbf{v} = \frac{\zeta - 1}{\zeta^2} \sum_{j=1}^{k} \frac{\{\lambda/\psi_j\}^2}{1 + \zeta - 2\lambda^2/\psi_j} = Q. \tag{2.12}$$

Now, note that we may view ζ as a function h of $\psi_1, \ldots, \psi_k, \lambda$. Define the vector $\dot{h}$ as $(\dot{h}_\psi^\top, \dot{h}_\lambda^\top)^\top$, where

$$\dot{h}_\psi = \begin{pmatrix} \frac{\partial h}{\partial \psi_1} \\ \frac{\partial h}{\partial \psi_2} \\ \vdots \\ \frac{\partial h}{\partial \psi_k} \end{pmatrix} = \begin{pmatrix} -\lambda \left(\psi_1\right)^{-2} \\ -\lambda \left(\psi_2\right)^{-2} \\ \vdots \\ -\lambda \left(\psi_k\right)^{-2} \end{pmatrix} = -\lambda \boldsymbol{\Psi}^{-2} \boldsymbol{\iota},$$

$$\dot{h}_\lambda = \frac{\partial h}{\partial \lambda^2} = \sum_{j=1}^{k} \frac{1}{\psi_j}.$$

The delta method yields

$$\sqrt{n}\left(\hat{\zeta} - \zeta\right) \to_d N\left(0, \dot{h}^\top \boldsymbol{\Omega} \dot{h}\right), \tag{2.13}$$

with

$$\dot{h}^\top \boldsymbol{\Omega} \dot{h} = (1 + \zeta)^2 \left\{ 2 + 2\frac{\zeta + 1}{\zeta - 1} \frac{\mathbf{v}^\top \mathbf{D}^{-1}\mathbf{v}}{1 + \mathbf{v}^\top \mathbf{D}^{-1}\mathbf{v}} \right\}, \tag{2.14}$$

where $\mathbf{D}$ and $\mathbf{v}$ are defined in (2.11). The combination of (2.12), (2.13) and (2.14) yields (2.3). □

Proof of (2.7)**.** We have that the values $1/\psi_1, \ldots, 1/\psi_k$ are drawn independently from the distribution of the non-negative random variable E^{-1}, which has finite variance. Hence, the limits

$$m_1 = \lim_{k \to \infty} \frac{1}{k} \sum_{j=1}^{k} \frac{1}{\psi_j}, \tag{2.15}$$

$$m_2 = \lim_{k \to \infty} \frac{1}{k} \sum_{j=1}^{k} \left\{ \frac{1}{\psi_j} - \frac{1}{k} \sum_{j=1}^{k} \frac{1}{\psi_j} \right\}^2 \tag{2.16}$$

exist, and are finite. Moreover, one may show

$$\lim_{k \to \infty} \max_{j=1,\ldots,k} \frac{1/\psi_j}{\sum_{j=1}^{k} 1/\psi_j} = 0. \tag{2.17}$$

It immediately follows from (2.15) that

$$\lim_{k\to\infty} \frac{1}{k}\zeta = m_1.$$

Moreover, it follows from (2.15)–(2.17) that

$$\lim_{k\to\infty} \frac{1}{k} \sum_{j=1}^{k} \frac{\{\lambda/\psi_j\}^2}{1+\zeta-2\lambda^2/\psi_j} = m_2 + (m_1)^2.$$

Now, (2.7) readily follows.

□

Chapter 3

Confidence Intervals for Maximal Reliability in Congeneric Models

Abstract

In this paper we discuss the issue of estimating the maximal reliability in the congeneric model. The aim is to acquire insight in how confidence intervals for the maximal reliability can be improved. We point out possible anti-conservatism of existing confidence intervals for maximal reliability. Two new confidence intervals for the maximal reliability are derived and compared to intervals derived in Yuan and Bentler (2002) and Raykov and Penev (2006). We introduce coverage curve analysis to compare confidence intervals. Simulation experiments show that among the two new confidence intervals, the new stable interval is less prone to Type II error than the existing candidates. This stable interval has a coverage for the true value which is approximately equal to the confidence level. This raises the issue whether variance stabilization is of use in more complicated models as well to improve confidence intervals.

3.1 Introduction and motivation

Measurements play an important role in research and much has been written about scale development. However, this is not the case for scale reliability maximization and especially the accuracy of the reliability estimation in general.

The precision of measurements is affected by the portion of random error. In classical test theory it is assumed that observed uncertainty is given by the sum of the true score and a random error. Typically, a test is composed of multiple items, where each item is an independent attempt to measure the same construct of interest T; that is, the measurement errors corresponding to the items are assumed to be independent. The outcome of the test U is usually the unweighted sum of the individual item outcomes Y_j.

We shall refer to T, U and Y_j as the true score, the test score, and the j^{th} item score, respectively. Without loss of generality we assume that the true score, the test score, and the item score all have zero expectation.

The amount of agreement between the test score U and the true score T, as captured by the squared correlation between U and T, is defined to be the reliability of the test, see Lord and Novick (1968). However, as T cannot be observed, we are unable to estimate the reliability of a test directly. Fortunately, we may use the dependence between the individual item scores in order to evaluate the reliability of a test indirectly.

The most widely used indirect technique for assessing reliability, Cronbach's coefficient alpha (see Cronbach, 1951), compares the variance of the test score with the unweighted sum of the variances of the individual item scores. Actually, coefficient alpha is a lower bound estimate of reliability if the assumption of essential tau equivalency is violated, (see Novick & Lewis, 1967). However, if the assumption of uncorrelated errors is violated, coefficient alpha may yield spuriously high estimates of reliability, see S. B. Green and Herschberger (2000); S. B. Green and Yang (2009) and also Sijtsma (2009). It is important to measure with high reliability.

Several methods allow for improving reliability by using a weighted sum W rather than an unweighted sum U as test score for instance, (see Knott & Bartholomew, 1993; Li, 1997; Yuan & Bentler, 2002; Raykov & Penev, 2006). We shall denote the reliability of the optimally weighted test score by φ, and we shall refer to it as the maximal reliability of $Y_1, \ldots, Y_k$.

The least restrictive linear model involving a one-dimensional true score, is the congeneric model. The congeneric model assumes that each of the item scores Y_j is the sum of $\lambda_j T$ and a measurement error ϵ_j with population mean zero and population variance ψ_j depending on j. The coefficients λ_j and ψ_j are unknown. To avoid identification problems, we assume without loss of generality that the variance of T is equal to 1.

One may show that the optimal weight for Y_j equals λ_j/ψ_j to obtain maximal reliability. This yields a test score $W = \sum_{j=1}^{k} \lambda_j Y_j/\psi_j$ which has maximal reliability defined as

$$\varphi = 1 - \frac{1}{1 + \sum_{j=1}^{k} \lambda_j^2/\psi_j}. \tag{3.1}$$

One may show that these optimal weights are proportional to the weights used to compute regression scores or Bartlett scores in the factor analysis model (see Lawley & Maxwell, 1971, Section 8.2 and 8.10).

In Yuan and Bentler (2002) and Raykov and Penev (2006) asymptotic confidence intervals for φ are derived by using the maximum likelihood estimator $\hat{\varphi}$ as pivotal

quantity. Different expressions for the standard deviation of $\hat{\varphi}$ lead to different confidence intervals.

However, the pivotal quantity $\hat{\varphi}$ does not have a stable variance; that is, a variance not depending on φ, see Hoaglin, Mosteller, and Tukey (1983). It is well-known that using an unstable variance may well yield anti-conservative confidence intervals. Anti-conservative confidence intervals have an attained coverage less than the required coverage and hence may falsely give the impression that φ is estimated with sufficient accuracy.

In this paper we propose new methods to establish confidence bounds for the maximal reliability (3.1). We advocate the use of a variance stabilizing transformation technique in constructing confidence intervals in the congeneric model and compare this interval with unstablized intervals to which the Yuan-Bentler and Raykov-Penev intervals also belong. We use coverage curve methodology to compare the performance of the intervals.

The structure of this paper is as follows. First, we present four confidence intervals, and show that their coverage is asymptotically equal to the requested confidence level. Proofs are relegated to the Appendix. Second, we apply the four confidence intervals to real data. Third, the results of an extensive simulation experiment are discussed. Finally, conclusions of this study are drawn.

3.2 Confidence intervals for the congeneric model

Let $\hat{\psi}_j$ and $\hat{\lambda}_j$ be the maximum likelihood estimators of ψ_j and λ_j. Then, the maximum likelihood estimator of φ is given by

$$\hat{\varphi} = 1 - \frac{1}{1+\hat{\zeta}}, \tag{3.2}$$

with

$$\hat{\zeta} = \sum_{j=1}^{k} \hat{\lambda}_j^2/\hat{\psi}_j. \tag{3.3}$$

Choose $0 < \gamma < 1$, and determine $z_{(1-\gamma)/2}$ so as to satisfy

$$P\left(-z_{(1-\gamma)/2} < Z < z_{(1-\gamma)/2}\right) = \gamma.$$

We next present several asymptotic $100\gamma\%$ confidence intervals for the maximal reliability coefficient φ. The confidence intervals for φ involve

$$\hat{s}^2 = 2 + 2\frac{\hat{Q}}{1+\hat{Q}}, \tag{3.4}$$

with

$$\hat{Q} = \frac{1}{\hat{\zeta}} \sum_{j=1}^{k} \frac{\left\{\hat{\lambda}_j^2/\hat{\psi}_j\right\}^2}{\hat{\zeta} - 2\hat{\lambda}_j/\hat{\psi}_j}.$$

Let n denote the sample size.

3.2.1 Available intervals

The first interval is proposed in general terms in Yuan and Bentler (2002). By taking their computations for the congeneric model, this yields that the endpoints of the Yuan-Bentler interval are given by

$$\hat{\varphi} \pm z_{(1-\gamma)/2}\,(1-\hat{\varphi})\,\frac{\hat{s}}{\sqrt{n}} \tag{3.5}$$

In Raykov and Penev (2006) a "second-order" alternative to the Yuan-Bentler interval is proposed. Here, the endpoints of the Raykov-Penev interval are given by

$$\hat{\varphi} \pm z_{(1-\gamma)/2}\,(1-\hat{\varphi})\,\frac{\hat{s}}{\sqrt{n}}\sqrt{1+2\frac{\hat{s}^2}{n}}. \tag{3.6}$$

Note that the Yuan-Bentler interval is always contained in the Raykov-Penev interval.

3.2.2 The unstable interval

The first new interval is related to the Yuan-Bentler interval and Raykov-Penev interval. In the Appendix it is shown that

$$\sqrt{n}\frac{\hat{\zeta} - \zeta}{(1+\zeta)\sqrt{2+2\frac{Q}{1+Q}}} \rightarrow_d N(0,1), \tag{3.7}$$

with

$$\zeta = \sum_{j=1}^{k} \lambda_j/\psi_j, \tag{3.8}$$

and

$$Q = \frac{1}{\zeta} \sum_{j=1}^{k} \frac{\left\{\lambda_j^2/\psi_j\right\}^2}{\zeta - 2\lambda_j/\psi_j}. \tag{3.9}$$

Note that $\hat{s}$ is a consistent estimator of $\sqrt{2 + 2\frac{Q}{1+Q}}$. We shall refer to the left hand side side of (3.7) as the unstable asymptotic pivotal quantity.

As the derivative of $1 - (1+\zeta)^{-1}$ with respect to ζ equals $(1+\zeta)^{-2}$, combining (3.7) with the delta method yields that

$$\sqrt{n} \frac{\hat{\varphi} - \varphi}{(1-\varphi)\sqrt{2 + 2\frac{Q}{1+Q}}} \rightarrow_d N(0,1). \tag{3.10}$$

It follows that

$$\lim_{n\to\infty} P\left(-z_{(1-\gamma)/2}(1-\hat{\varphi})\frac{\hat{s}}{\sqrt{n}} < \hat{\varphi} - \varphi < z_{(1-\gamma)/2}(1-\hat{\varphi})\frac{\hat{s}}{\sqrt{n}}\right) = \gamma,$$

which implies that

$$\hat{\varphi} \pm z_{(1-\gamma)/2}(1-\hat{\varphi})\frac{\hat{s}}{\sqrt{n}}$$

are the endpoints of an asymptotic $100\gamma\%$ confidence interval for φ. This interval coincides with the Yuan-Bentler interval in (3.5).

In Raykov and Penev (2006) properties of the normal distribution and the second order approximation

$$\hat{\varphi} - \varphi = \frac{\hat{\zeta} - \zeta}{(1+\zeta)^2} - \frac{\left(\hat{\zeta} - \zeta\right)^2}{(1+\zeta)^3},$$

are used to obtain an asymptotic confidence interval for $\hat{\varphi} - \varphi$. This yields

$$\lim_{n\to\infty} P\left(-z_{(1-\gamma)/2}(1-\hat{\varphi})\frac{\hat{s}}{\sqrt{n}}\sqrt{1+2\frac{\hat{s}^2}{n}} < \hat{\varphi} - \varphi < z_{(1-\gamma)/2}(1-\hat{\varphi})\frac{\hat{s}}{\sqrt{n}}\sqrt{1+2\frac{\hat{s}^2}{n}}\right) = \gamma,$$

which implies that

$$\hat{\varphi} \pm z_{(1-\gamma)/2}(1-\hat{\varphi})\frac{\hat{s}}{\sqrt{n}}\sqrt{1+2\frac{\hat{s}^2}{n}}$$

are the endpoints of an asymptotic $100\gamma\%$ confidence interval for φ. This interval coincides with the Raykov-Penev interval in (3.6).

We now turn to our two new confidence intervals for φ. From (3.1) and (3.8), it follows from (3.7) that

$$\lim_{n\to\infty} P\left(-z_{(1-\gamma)/2}\frac{\hat{s}}{(1-\hat{\varphi})\sqrt{n}} < \hat{\zeta}-\zeta < z_{(1-\gamma)/2}\frac{\hat{s}}{(1-\hat{\varphi})\sqrt{n}}\right) = \gamma,$$

which implies that

$$\hat{\zeta} \pm z_{(1-\gamma)/2}\frac{\hat{s}}{(1-\hat{\varphi})\sqrt{n}}$$

are the endpoints of an asymptotic $100\gamma\%$ confidence interval for ζ. As φ is a monotone function of ζ, we obtain that

$$\begin{aligned} 1-\frac{1}{1+\hat{\zeta}\pm z_{(1-\gamma)/2}\frac{\hat{s}}{(1-\hat{\varphi})\sqrt{n}}} &= 1-\frac{1}{\frac{1}{(1-\hat{\varphi})}\pm z_{(1-\gamma)/2}\frac{\hat{s}}{(1-\hat{\varphi})\sqrt{n}}} \\ &= 1-\frac{1-\hat{\varphi}}{1\pm z_{(1-\gamma)/2}\frac{\hat{s}}{\sqrt{n}}} \end{aligned} \tag{3.11}$$

are the endpoints of an asymptotic $100\gamma\%$ confidence interval for φ.

3.2.3 The stable interval

We believe that there is a serious drawback to the direct use of (3.7) in constructing confidence intervals for φ. Although we know that the left hand side of (3.7) converges in distribution to a standard normal random variable as the number of observations n becomes large, the convergence may well be slow.

In order to motivate our remedy for this problem, we assume in a first instance that the items of the test are drawn at random from a large test battery. That is, the values $\psi_1, \ldots, \psi_k$ are drawn independently from the distribution of some non-negative random variable, say E. In addition, we shall assume that the variance of E^{-1} is finite. The quantities ζ and Q depend on the number k of items drawn. One may show that

$$\lim_{k\to\infty}\left(\frac{\zeta+1}{\zeta}\right)^2\frac{Q}{1+Q} = 0, \tag{3.12}$$

It follows by (3.7) that the asymptotic variance of $\sqrt{n}\left(\hat{\zeta}-\zeta\right)$ is approximated by $2\left(1+\zeta\right)^2$ for large k. Thus, the variance of $\hat{\zeta}$ depends on ζ, and this dependence does not disappear when k tends to infinity. That is, the variance of $\hat{\zeta}$ is not stable. As the variance is proportional to $\left(1+\zeta\right)^2$ for k sufficiently large, the theory of variance stabilizing

transformations suggests the use of $\ln\left(1+\hat{\zeta}\right)$ for constructing confidence intervals (see e.g. Hoaglin et al., 1983). It can be shown that

$$\sqrt{n}\frac{\ln\left(1+\hat{\zeta}\right)-\ln\left(1+\zeta\right)}{\sqrt{2+2\frac{Q}{1+Q}}}\rightarrow_d N\left(0,1\right). \tag{3.13}$$

We shall refer to the left hand side side of (3.13) as the stable asymptotic pivotal quantity.

It follows from (3.13) that

$$\lim_{n\rightarrow\infty} P\left(-z_{(1-\gamma)/2}\frac{\hat{s}}{\sqrt{n}} < \ln\left(\frac{1+\hat{\zeta}}{1+\zeta}\right) < z_{(1-\gamma)/2}\frac{\hat{s}}{\sqrt{n}}\right)=\gamma,$$

which implies that

$$\left(1+\hat{\zeta}\right)\exp\left\{\pm z_{(1-\gamma)/2}\frac{\hat{s}}{\sqrt{n}}\right\}-1$$

are the endpoints of an asymptotic $100\gamma\%$ confidence interval for ζ. As φ is a monotone function of ζ, we obtain that

$$1-\frac{\exp\left\{\pm z_{(1-\gamma)/2}\frac{\hat{s}}{\sqrt{n}}\right\}}{1+\hat{\zeta}}=1-\left(1-\hat{\varphi}\right)\exp\left\{\pm z_{(1-\gamma)/2}\frac{\hat{s}}{\sqrt{n}}\right\} \tag{3.14}$$

are the endpoints of an asymptotic $100\gamma\%$ confidence interval for φ.

3.3 Illustration

We will illustrate the above interval estimation for maximal reliability with real data. Three hundred respondents have expressed their agreement on a five point scale consisting of six items. The items are statements concerning actions to be taken to reduce the effect of the financial crisis on the respondent's company.

The congeneric model fits the data well [$\chi^2 = 6.049$, df $= 9$, $P = 0.735$]. The estimator for the maximal reliability $\hat{\varphi}$ takes the value 0.8770 and the corresponding 95% confidence intervals for the maximal reliability φ are found in Table 3.1.

Note that the Yuan-Bentler and the Raykov-Penev intervals are symmetric around $\hat{\varphi}$, whereas the new interval 2 and the new interval 1 are not. Moreover, the Raykov-Penev interval contains the Yuan-Bentler interval. The lower endpoint of the new interval 1 is smaller than the lower endpoint of the stable interval, which in turn is smaller than

Table 3.1: Asymptotic 95% confidence intervals for maximal reliability φ. The estimator $\hat{\varphi}$ takes the value 0.8770.

Interval	*Lower*	*Upper*
Unstable	0.833	0.903
YuanBentler	0.844	0.910
RaykovPenev	0.844	0.910
Stable	0.840	0.906

the lower endpoint of the Raykov-Penev interval. Similarly, the upper endpoint of the new interval 1 is smaller than the upper endpoint of the stable interval, which in turn is smaller than the upper endpoint of the Raykov-Penev interval.

Due to the relatively large sample size, the differences between the confidence intervals in Table 1 are not very large. To further our understanding of the differences between the four confidence intervals, we rely on simulations, to be discussed in the next section.

3.4 Simulation results

In the simulations, we let k take the values $3, 5, 8, 10, 15, 20$, and n take the values $25, 50, 100, 200, 400$. The number of replications in the simulations is 10000. We define the quantities μ_1 and μ_2 as

$$\mu_1 = \frac{1}{k}\sum_{j=1}^{k}\frac{\lambda_j^2}{\psi_j}, \quad \mu_2 = \frac{1}{k}\sum_{j=1}^{k}\left(\frac{\lambda_j^2}{\psi_j} - \mu_1\right)^2. \tag{3.15}$$

Note that we may also write $\mu_1 = k^{-1}\varphi/(1-\varphi)$. In our simulations we set μ_1 equal to values which correspond to the maximal reliability 0.60, 0.75, 0.90 and 0.95. Next, we set μ_2 equal to $0.05\mu_1^2$, $0.1\mu_1^2$ and $0.15\mu_1^2$.

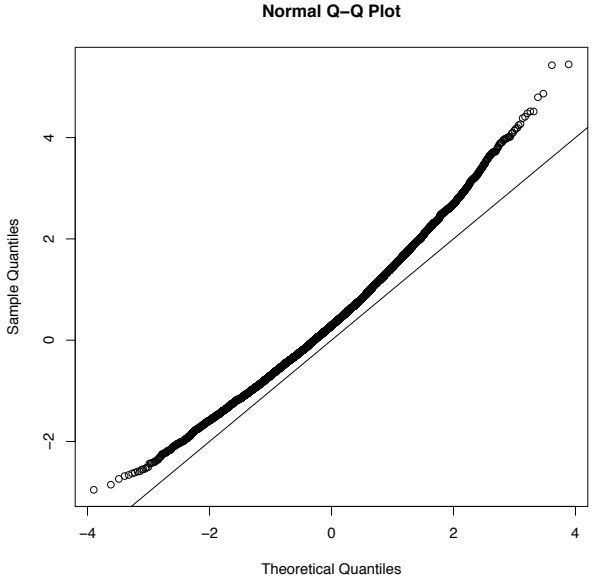
We expect that the simulation results largely depend on the quantities μ_1 and μ_2. In order to be able to confirm this expectation, we generate the λ_i^2/ψ_i's using various patterns. In particular, we choose

$$\frac{\lambda_i^2}{\psi_i} = a + bg\left(\frac{i}{k+1}\right),$$

where $g(x)$ is one of the three following functions:

$$g_1(x) = x - \tfrac{1}{2},$$

Figure 3.1: Normal probability plot of simulated values of the left hand side of (3.7) with $k = 5$, $\varphi = 0.9$, $n = 100$. The plot approaches the line with intercept 0 and slope 1 only in the center, there is a marked deviation in the tails.



$$g_2(x) = x^2 - \tfrac{1}{3},$$

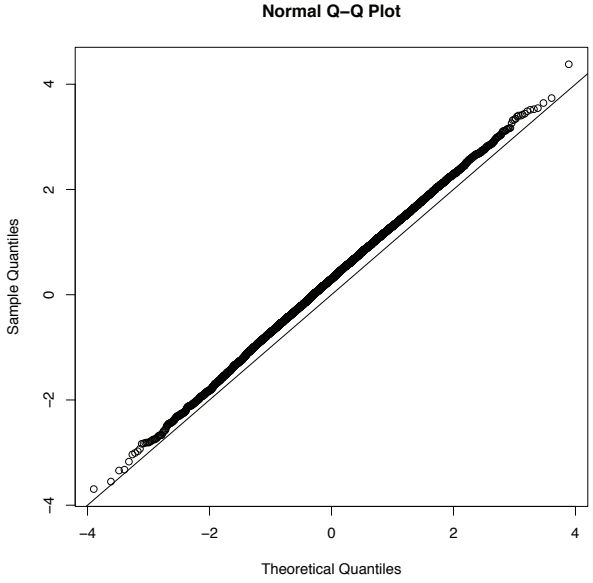
$$g_3(x) = \begin{cases} -1 & \text{for } x < \tfrac{1}{2}, \\ +1 & \text{for } x \geq \tfrac{1}{2}, \end{cases} .$$

Here a and b are chosen so as to satisfy (3.15) for given μ_1 and μ_2.

The parameters in the congeneric model may be estimated by any statistical program capable of performing either exploratory factor analysis or confirmatory factor analysis.

First, we investigate the extent in which the standard normal approximations (3.13) and (3.7) are valid. It turns out that these approximations are troublesome when the number of test items k is small. When $k = 5$, the approximations are reasonably accurate, except when both n and φ are small. Recall that the standard normal approximations (3.13) and (3.7) provide the probabilistic basis for all confidence intervals discussed in this paper. Hence, in the remainder of this section we shall restrict ourselves to situations in which $k \geq 5$ and $n \geq 50$. As is to be expected, the distribution of the stable asymptotic pivotal quantity is closer to normal than the distribution of the unstable asymptotic pivotal quantity as illustrated by Figures 3.1 and 3.2.

Figure 3.2: Normal probability plot of simulated values of the left hand side of (3.13), with $k = 5$, $\varphi = 0.9$, $n = 100$. The plot approaches the line with intercept 0 and slope 1, even in the tails.



We have also examined the validity of the limiting approximation when item scores are discrete. It turns out that for reasonable n, it is already possible for $k \geq 5$ to achieve the normal approximation. However, φ should not be too near to one, say $\varphi \leq 0.95$.

Hence, though we assume a standard normal distribution, it is sufficient that the pivotal quantity (3.13) converges to a standard normal distribution. As long as this is the case, the confidence intervals derived from this pivotal quantity remain asymptotically valid; consequently, the assumption that the data follow a multivariate normal distribution may be relaxed somewhat.

Next, we examine the simulated coverage curves of the four intervals. The coverage of a confidence interval for a given hypothetical value of φ is defined as the probability that this hypothetical value is contained in the interval. The coverage curve is the curve that the coverage follows as the hypothetical value of φ ranges through an interval of possible values of φ. The ideal shape of the coverage curve is as follows:

- if the hypothetical value of φ equals the true value, then the coverage should be equal to the confidence level;

- if the hypothetical value of φ differs from the true value, then the coverage should be as low as possible. In particular, the coverage should be lower than the confidence level.

There are various ways in which a coverage curve may deviate from the ideal shape. In our examination of the performance of the confidence interval, we shall in particular consider the unbiasedness of a confidence interval. A confidence interval is called unbiased if for every hypothetical value of φ the coverage of the interval does not exceed the coverage for the true value of φ.

If an interval is unbiased, the next issue to consider is whether the coverage for the true value of φ equals the confidence level. If this coverage is larger than the confidence level, the unbiased confidence interval is called conservative; if this coverage is smaller than the confidence level, the unbiased confidence interval is called anti-conservative. Being conservative is considered less harmful than being anti-conservative.

Statistical testing theory yields an alternative way to interpret coverage curves. Note that for every confidence interval for φ, there exists a related statistical test of the null hypothesis $H_0 : \varphi = \varphi_0$; this test does not reject the null hypothesis if the hypothetical value φ_0 lies inside the confidence interval, and rejects the null hypothesis if the hypothetical value φ_0 is outside the confidence interval. Now, if the true value of φ differs from the hypothetical value φ_0, then the coverage of the hypothetical value φ_0 is equal to the probability of a type II error of the related statistical test of $H_0 : \varphi = \varphi_0$. Thus, by subtracting the coverage curve from 1, we in fact obtain the power curve of the related test. In fact, the description of the ideal shape of the coverage curve given above is a direct translation of generally accepted rules involving the ideal shape of the power curve.

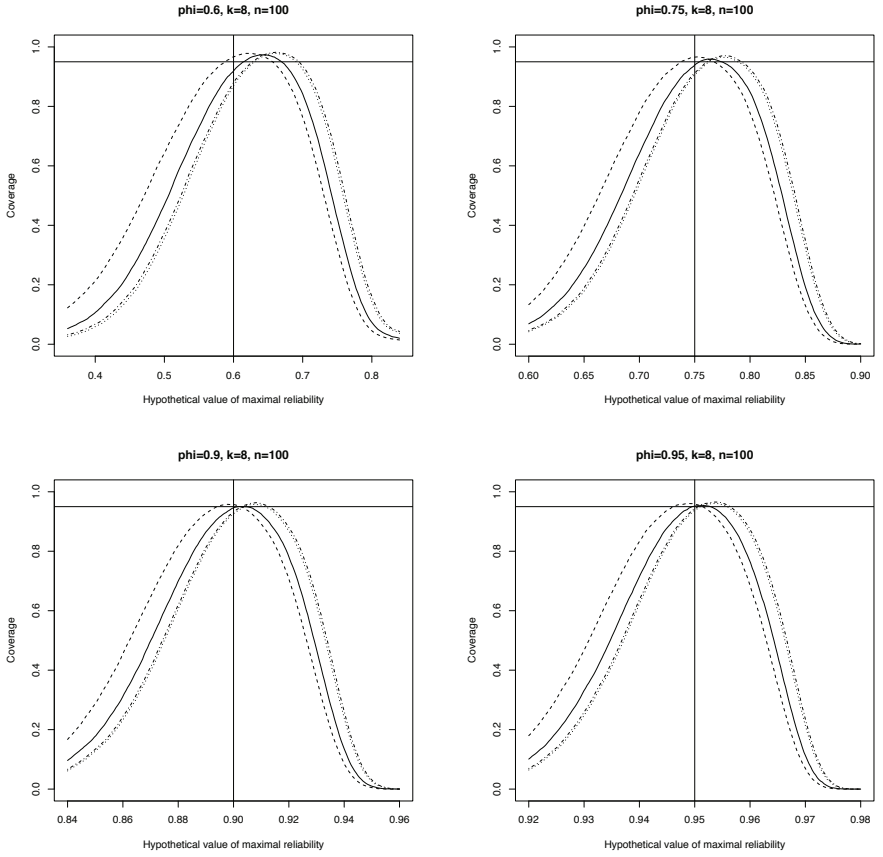
Evaluating the coverage not only under the null hypothesis, but also under the alternative hypothesis provides a much more comprehensive view of the behavior of the various confidence intervals.

As it is not possible to include all generated figures, we have only included the most striking ones. All relevant figures are available upon request.

The simulated coverage curves depend on the sample size n, the true value of the maximal reliability φ, the number of items k and on the quantities μ_1 and μ_2. It turns out that μ_2 has little effect, and implies that the intervals do not vary too much. As μ_1 is determined by φ, only n, φ and k are of interest. Their effects are as follows.

- The sample size n has a positive effect on the performance of all confidence intervals. That is, the coverage curve approaches the ideal shape for large n.

Figure 3.3: Coverage curves of stable interval (solid line), unstable interval (dashed line), Yuan-Bentler interval (dotted line) and Raykov-Penev interval (dotted-dashed line) for $\varphi = 0.60, 0.75, 0.90, 0.95$, $k = 8$ and $n = 100$.



- The true value of the maximal reliability φ has a positive effect on the coverage of all confidence intervals, see Figure 3.3. When the true value of φ approaches 1, the coverage of the true value increases.

- The number of items k has a positive effect on the performance of all confidence intervals. That is, the coverage curve approaches the ideal shape for large k.

The simulated coverage curves yield the following general findings with respect to the differences in performance between the four confidence intervals.

- There is little difference between the Yuan-Bentler and the Raykov-Penev interval. Both intervals show a positive bias, and thus overestimate the true maximal reliability. This bias is more marked when the true value of the maximal reliability φ is low.

- The new interval 1 shows a negative bias, and thus underestimates the true maximal reliability. This bias is more marked when the true value of φ approaches 1. The negative bias turns into a positive bias when the true value of φ is becomes lower. However, the coverage of the true value of φ is still larger than the confidence level.

- The stable interval shows some positive bias for low values of the true value of φ, but this bias is less than the bias of the other intervals. Except for extreme conditions (that is, a combination of a small sample size n, a small number of items k and a low true value of the maximal reliability φ), the stable interval performs well, and has a coverage for the true value φ which is approximately equal to the confidence level.

Finally, we remark that we could not have detected the positive bias of the Yuan-Bentler and the Raykov-Penev without the construction of coverage curves, that is, evaluating the coverage under the null hypothesis as well as under the alternative hypothesis. We highly recommend the use of coverage curves in other studies on confidence intervals as it has shown to be more informative.

3.5 Conclusion

We discussed the issue of estimating the maximal reliability in the congeneric model. We motivated that existing confidence intervals for maximal reliability as derived in Yuan and Bentler (2002) and Raykov and Penev (2006) are possibly anti-conservative as they are derived by using $\hat{\varphi}$ as pivotal quantity, which does not have a stable variance. Our simulations show that anti-conservatism of these intervals indeed occurs. Anti-conservative confidence intervals have an attained coverage less than the required coverage and hence may falsely give the impression that φ is estimated with sufficient accuracy. Therefore, we advocate the use of a variance stabilizing transformation technique in constructing confidence intervals for φ.

We have derived two new confidence intervals for maximal reliability and compared the performance of these intervals with intervals proposed in Yuan and Bentler (2002) and Raykov and Penev (2006). To compare these intervals, we have introduced coverage

curves. That is, we have not only considered the coverage of the true maximal reliability, but the coverage of hypothetical values which differ from the true maximal reliability as well. We would highly recommend the use of coverage curves analysis to compare those intervals as it has shown to be more informative.

It turns out that the Yuan-Bentler and the Raykov-Penev intervals are closely related to each other. In fact, the Yuan-Bentler interval is always contained in the Raykov-Penev interval. The similarity in behavior of the two intervals is in contrast with the remarks in Raykov and Penev (2006) with regard to the Yuan-Bentler interval. According to Raykov and Penev (2006) it is "(a) rather laborious and tedious in routine behavioral research in need of interval estimation of maximal reliability, (b) involves taking by the researcher of multiple partial derivatives of this reliability coefficient with respect to model parameters, (c) has the inconvenient property that the number of these derivatives increases with increasing length k of the initial composite of interest (as could be repeatedly the case when one is involved in scale development and revision), and (d) can be viewed as based on a first-order approximation of maximal reliability as a function of model parameters." By taking the computations in Yuan and Bentler (2002) one step further, one can simply eliminate the problem of taking multiple partial derivatives. Moreover, it seems that the additional complexity of the alternative interval proposed in Raykov and Penev (2006) does not pay off, as a clear advantage of using this interval over the Yuan-Bentler is lacking. Both intervals overestimate the true maximal reliability, which is more marked for low values of the true maximal reliability. This might be due to a ceiling effect that cuts down the positive biases of these two intervals when the true value of φ is high.

We have also examined the performance of the two new intervals proposed in this article. The new interval 1, which is also unstable, underestimates the true value of the maximal reliability considerable when the true value of φ approaches 1 and overestimates when the true value of φ becomes lower. The coverage for the true value of φ of this interval is larger than the confidence level.

Except for a combination of a small sample size, a small number of items and a low true value of the maximal reliability, the stable interval performs well, and has a coverage for the true value of φ which is approximately equal to the confidence level. This shows the advantage of the use of a stabilization technique in constructing confidence intervals.

Observe that the congeneric model is in fact a factor analysis model with a single factor. This raises the issue whether variance stabilization transformation technique is of use in more complicated models as well, for example multiple factor analysis models. The

obtained results give us insight in how we could apply variance stabilization to improve confidence intervals for the maximal reliability in these models in further research.

3.A Proofs

Proof of (3.7)**.** The congeneric model implies that the population covariance matrix of $Y_1, \ldots, Y_k$ is given by

$$\boldsymbol{\lambda}\boldsymbol{\lambda}^\top + \boldsymbol{\Psi},$$

where $\boldsymbol{\lambda}$ is the k-dimensional vector with elements $\lambda_1, \ldots, \lambda_k$, and $\boldsymbol{\Psi}$ the diagonal matrix with diagonal elements $\psi_1, \ldots, \psi_k$.

Let $\boldsymbol{\theta}$ denote the $2k$-dimensional parameter vector $(\psi_1, \ldots, \psi_k, \lambda_1, \ldots, \lambda_k)^\top$, and let $\hat{\boldsymbol{\theta}}$ denote the maximum likelihood estimator of $\boldsymbol{\theta}$. It follows from the theory of covariance structures (see Browne, 1984; Kano et al., 1990) that

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right) \rightarrow_d N_{2k}\left(\mathbf{0}, \boldsymbol{\Omega}\right), \tag{3.16}$$

with

$$\boldsymbol{\Omega} = \begin{pmatrix} \boldsymbol{\Omega}_{\psi\psi} & \boldsymbol{\Omega}_{\psi\lambda} \\ \boldsymbol{\Omega}_{\lambda\psi} & \boldsymbol{\Omega}_{\lambda\lambda} \end{pmatrix},$$

where

$$\boldsymbol{\Omega}_{\psi\lambda} = \boldsymbol{\Omega}_{\lambda\psi}{}^\top = -\frac{1}{\zeta}\boldsymbol{\Omega}_{\psi\psi}\boldsymbol{\Upsilon}\left(\mathbf{I} - \frac{\boldsymbol{\lambda}\boldsymbol{\lambda}^\top\boldsymbol{\Psi}^{-1}}{1+\zeta}\right)\left\{\boldsymbol{\Psi} - \frac{1}{2}\frac{1-\zeta}{\zeta}\boldsymbol{\lambda}\boldsymbol{\lambda}^\top\right\},$$

$$\begin{aligned} \boldsymbol{\Omega}_{\lambda\lambda} &= \frac{1+\zeta}{\zeta}\left\{\boldsymbol{\Psi} - \frac{1}{2}\frac{1-\zeta}{\zeta}\boldsymbol{\lambda}\boldsymbol{\lambda}^\top\right\} \\ &+ \frac{1}{\zeta^2}\left\{\boldsymbol{\Psi} - \frac{1}{2}\frac{1-\zeta}{\zeta}\boldsymbol{\lambda}\boldsymbol{\lambda}^\top\right\}\left(\mathbf{I} - \frac{\boldsymbol{\Psi}\boldsymbol{\lambda}\boldsymbol{\lambda}^\top}{1+\zeta}\right)\boldsymbol{\Upsilon}\boldsymbol{\Omega}_{\psi\psi} \\ &\times \boldsymbol{\Upsilon}\left(\mathbf{I} - \frac{\boldsymbol{\lambda}\boldsymbol{\lambda}^\top\boldsymbol{\Psi}^{-1}}{1+\zeta}\right)\left\{\boldsymbol{\Psi} - \frac{1}{2}\frac{1-\zeta}{\zeta}\boldsymbol{\lambda}\boldsymbol{\lambda}^\top\right\}, \end{aligned}$$

and

$$\boldsymbol{\Omega}_{\psi\psi} = 2\boldsymbol{\Psi}\left(\mathbf{D} + \mathbf{v}\mathbf{v}^\top\right)^{-1}\boldsymbol{\Psi},$$

with

$$\mathbf{D} = \left(\mathbf{I} - 2\frac{\boldsymbol{\Psi}\boldsymbol{\Upsilon}\boldsymbol{\Psi}\boldsymbol{\Upsilon}\boldsymbol{\Psi}}{1+\zeta}\right), \qquad \mathbf{v} = \frac{\boldsymbol{\Psi}\boldsymbol{\Upsilon}\boldsymbol{\lambda}}{1+\zeta}, \tag{3.17}$$

where $\boldsymbol{\Upsilon}$ is the $k \times k$-dimensional diagonal matrix with $\upsilon_{ii} = \lambda_i/\psi_i^2$ as i^{th} diagonal element.

The Sherman-Morrison-Woodbury formula (see Hager, 1989) yields

$$\left(\mathbf{D} + \mathbf{v}\mathbf{v}^\top\right)^{-1} = \mathbf{D}^{-1} - \left(1 + \mathbf{v}^\top\mathbf{D}^{-1}\mathbf{v}\right)^{-1}\mathbf{D}^{-1}\mathbf{v}\mathbf{v}^\top\mathbf{D}^{-1}.$$

It follows that

$$\mathbf{v}^\top \mathbf{D}^{-1}\mathbf{v} = \frac{1}{\zeta}\sum_{j=1}^{k}\frac{\left\{\lambda_j^2/\psi_j\right\}^2}{\zeta - 2\lambda_j^2/\psi_j} = Q. \tag{3.18}$$

Now, note that we may view ζ as a function h of $\psi_1, \ldots, \psi_k, \lambda_1, \ldots, \lambda_k$. Define the vector $\dot{h}$ as $(\dot{h}_\psi^\top, \dot{h}_\lambda^\top)^\top$, where

$$\dot{h}_\psi = \begin{pmatrix} \frac{\partial h}{\partial \psi_1} \\ \frac{\partial h}{\partial \psi_2} \\ \vdots \\ \frac{\partial h}{\partial \psi_k} \end{pmatrix} = \begin{pmatrix} -\lambda_1^2/\psi_1^2 \\ -\lambda_2^2/\psi_2^2 \\ \vdots \\ -\lambda_k^2/\psi_k^2 \end{pmatrix} = -\boldsymbol{\Upsilon}\boldsymbol{\lambda},$$

$$\dot{h}_\lambda = \begin{pmatrix} 2\lambda_1/\psi_1 \\ 2\lambda_2/\psi_2 \\ \vdots \\ 2\lambda_k/\psi_k \end{pmatrix} = 2\boldsymbol{\Psi}^{-1}\boldsymbol{\lambda}.$$

The delta method yields

$$\sqrt{n}\left(\hat{\zeta} - \zeta\right) \to_d N\left(0, \dot{h}^\top \boldsymbol{\Omega}\dot{h}\right), \tag{3.19}$$

with

$$\dot{h}^\top \boldsymbol{\Omega}\dot{h} = (1+\zeta)^2\left\{2 + 2\frac{\mathbf{v}^\top\mathbf{D}^{-1}\mathbf{v}}{1+\mathbf{v}^\top\mathbf{D}^{-1}\mathbf{v}}\right\}, \tag{3.20}$$

where $\mathbf{D}$ and $\mathbf{v}$ are defined in (3.17). The combination of (3.18), (3.19) and (3.20) yields (3.7). □

Proof of (3.12)**.** We have that the values $\lambda_1^2/\psi_1, \ldots, \lambda_k^2/\psi_k$ are drawn independently from the distribution of the non-negative random variable E^{-1}, which has finite variance. Hence, the limits

$$m_1 = \lim_{k\to\infty}\frac{1}{k}\sum_{j=1}^{k}\frac{\lambda_j^2}{\psi_j}, \tag{3.21}$$

$$m_2 = \lim_{k\to\infty}\frac{1}{k}\sum_{j=1}^{k}\left\{\lambda_j^2/\psi_j - \frac{1}{k}\sum_{j=1}^{k}\frac{\lambda_j^2}{\psi_j}\right\}^2 \tag{3.22}$$

exist, and are finite. Moreover, one may show

$$\lim_{k\to\infty}\max_{i=1,\ldots,k}\frac{\lambda_i^2/\psi_i}{\sum_{j=1}^{k}1/\psi_j} = 0. \tag{3.23}$$

Condition (3.23) implies that

$$\lim_{k\to\infty} \mathbf{v}^\top \mathbf{D}^{-1}\mathbf{v} = \lim_{k\to\infty} \frac{1}{k}\left(\frac{1}{\frac{1}{k}\zeta}\right)^2 \frac{1}{k}\sum_{j=1}^{k}\left\{\lambda_j^2/\psi_j\right\}^2,$$

and thus it follows by (3.21) and (3.22) that

$$\lim_{k\to\infty} \mathbf{v}^\top \mathbf{D}^{-1}\mathbf{v} = \left\{\lim_{k\to\infty} \frac{1}{k}\right\}\left\{\lim_{k\to\infty}\left(\frac{1}{\frac{1}{k}\zeta}\right)^2\right\}\left\{\lim_{k\to\infty} \frac{1}{k}\sum_{j=1}^{k}\left\{\lambda_j^2/\psi_j\right\}^2\right\} = 0.$$

This concludes the proof of (3.12). □

Part II

Rankings

Chapter 4

Preliminary Analysis of Preference Rankings when there are Too Many Alternatives*

Abstract

We derive the score test for Best-Worst ranking data when there are too many alternatives. Moreover, we introduce a multiple comparison procedure to examine differences across alternatives when the null hypothesis is rejected. This methodology has not yet been applied in a marketing context. We show that the same results can also be obtained by a nonparametric approach, which in addition enhances interpretation. Moreover, we show that our proposed nonparametric test and the score test are equivalent, which implies that the far more simpler nonparametric approach is as good as the likelihood approach, as the score test has optimality characteristics. We extend the nonparametric approach to general ranking schemes, which can be easily done in contrast to the likelihood approach. In a preliminary analysis of preference rankings when there are too many alternatives, the proposed methodology reduces the amount of alternatives in an easy and understandable way, such that irrelevant alternatives can be removed from further analysis.

4.1 Introduction and motivation

Stated preferences are respondents' expressed preferences in hypothetical situations. The main argument against stated preference data is what respondents say they will do may be different from what they actually do. However, discrepancies between stated and actual behavior are prone to occur in a hypothetical situation which differs from a real choice context: such discrepancies may depend on the clarity of the questionnaire and

*This chapter is partly based on Lam et al. (2010a).

the setting of the experiment. Ben-Akiva et al. (1992) mentions several advantages of stated preference data over revealed preference data: choice sets can be prespecified, ranges of attributes can be extended, multicollinearity among attributes can be avoided, attributes that are not easily quantified, such as safety, reliability, and availability, can be incorporated. Moreover, in Louviere et al. (2000), a number of reasons is given why we should be interested in stated preferences. In practice, managers often need to estimate market response to new product with new attributes or features. Revealed preference data are generally limited to helping us understand preferences within existing markets, whereas stated preference data provide insights into future markets and are more useful for forecasting changes in behavior. Such applications have no revealed preference data on which they can rely and one has little alternative than to use stated preference data. Inferences about and predictions of real market behavior based on stated preferences are indistinguishable from revealed preferences.

Stated consumer preferences can be measured by rankings of alternatives. Metric measurements such as rating and matching may be less reliable due to respondents' limited ability to accurately report degrees of preferences (Ben-Akiva et al., 1992). Rankings are easy to collect and they are easy to interpret.

However, task difficulty increases substantially with the number of alternatives to be ranked. It has been widely recognized that individuals face difficulties, or even become dissatisfied, when having to compare too many choice options. Iyengar and Lepper (2000) in their famous experiment showed that too much choice can be demotivating for consumers. In a marketing context, Boatwright and Nunes (2001) demonstrated that a reduction of the assortment is in fact felt as beneficial to consumers, a finding which has been supported by Chernev (2003) and Gourville and Soman (2005), among others. In another context, DeShazo and Fermo (2002) showed that consumers experience task complexity when choice options are plenty, see also Sandor and Franses (2009). In sum, consumers may find it difficult to rank preferences across too large an amount of alternatives.

The problem of task complexity can be alleviated by asking respondents to evaluate all alternatives but only to give preference rankings for a subset of the alternatives. Moreover, in Louviere et al. (2000), one does not need complete rankings to obtain extra preference information.

One partial ranking that is gaining in popularity is Best-Worst ranking, developed by Louviere and Woodworth (1991a) as a multiple choice extension of Thurstone (1927) method of paired comparisons. Best-Worst ranking assumes that respondents evaluate

all possible pairs of alternatives within a displayed (sub)set and choose that pair of alternatives that reflects maximum difference in preference or importance on an underlying latent dimension.

Best-Worst ranking is a special case of what we call a ranking scheme. Ranked preference data allows considerably more information to be gathered from a given survey observation than is gathered from simply the most and least preferred alternatives (Hausman & Ruud, 1987). In this paper we present the analysis of more extensive partial ranking schemes. As individuals are well able to discriminate among alternatives in the extremes, we could efficiently obtain more preference information by extending the rank data of these extreme alternatives rather than limited to the best and worst alternatives. That is what we call a partial ranking scheme: a predefined data generating process where we ask respondents to indicate and rank their most preferred alternatives and their least preferred alternatives. The method can easily extended by asking respondents to rank a subset of the alternatives in the middle part of their preferences. However, various studies observed that inconsistency occurs in the middle ranks (e.g. Ben-Akiva et al., 1992).

We propose an appropriate test statistic to test the observed preference rankings. That is, we want to test the null hypothesis

> H_0: There are no differences across the alternatives. Each arrangement of the C ranks is equally likely,

against the alternative hypothesis

> H_a: At least one alternative tends to yield a higher ranking than at least one other alternative.

If there is statistical evidence of such a difference, the question is of course which alternatives it concerns. To this end, we will present a multiple comparison procedure to reach conclusions about differences between all pairs of alternatives. To our knowledge this is new in the field of marketing.

The advantage of multiple comparisons is that it helps a marketer to decide which alternatives to include in subsequent more detailed study. A marketer can analyze a large amount of alternatives in a preliminary analysis and alternatives which are not relevant can now be removed before further analysis is done.

The outline of our paper is as follows. In Section 4.2 and Section 4.3 we outline the statistical methodology from two different statistical paradigms: likelihood and nonparametric theory. In Section 4.4 we illustrate the equivalence of both methodologies. In Section 4.5 we outline the nonparametric theory in a general context and illustrate its

relevance and ease of use in Section 4.6. Section 4.7 concludes with a few avenues for further research.

4.2 Likelihood approach to Best-Worst ranking

Suppose each of n respondents is asked to indicate the most as well as the least preferred alternatives out of C alternatives. Assume that the probability that a random respondent simultaneously selects alternative s as the most preferred, and alternative t as the least preferred is

$$p_{st} = \frac{\exp\{\delta_{st}\}}{\sum_{s'} \sum_{t' \neq s'} \exp\{\delta_{s't'}\}},$$

where $\delta_{st} = \mu_s - \mu_t$ and μ_c is the average preference for alternative c, see also Finn and Louviere (1992). Set $\mu_C = 0$ to avoid identification issues, and note that p_{st} is now a function of the remaining parameters $\mu_1, \mu_2, \ldots, \mu_{C-1}$. We obtain

$$\frac{\partial \ln p_{st}}{\partial \mu_c} = \left\{1_{\{s=c\}} - p_{c\cdot}\right\} - \left\{1_{\{t=c\}} - p_{\cdot c}\right\},$$

where $1_{\{\cdot\}}$ denotes an indicator function that is 1 if the argument is true and 0 otherwise, and where $p_{c\cdot}$ denotes $\sum_{t' \neq c} p_{ct'}$ and $p_{\cdot c}$ denotes $\sum_{s' \neq c} p_{s'c}$.

Let o_{st} denote the number of respondents indicating alternative s as the most preferred, and alternative t as the least preferred. Then, the log-likelihood is given by

$$\ln L\left(\mu_1, \mu_2, \ldots, \mu_{C-1}\right) = \sum_{st \in \mathcal{K}} o_{st} \ln p_{st}. \tag{4.1}$$

We obtain

$$\begin{aligned} \frac{\partial}{\partial \mu_c} \ln L\left(\mu_1, \mu_2, \ldots, \mu_{C-1}\right) &= \sum_{st \in \mathcal{K}} o_{st} \frac{\partial \ln p_{st}}{\partial \mu_c} \\ &= \left\{o_{c\cdot} - n p_{c\cdot}\right\} - \left\{o_{\cdot c} - n p_{\cdot c}\right\}, \end{aligned} \tag{4.2}$$

where $o_{c\cdot}$ denotes $\sum_{t' \neq c} o_{ct'}$ and $o_{\cdot c}$ denotes $\sum_{s' \neq c} o_{s'c}$. Further, we obtain

$$\begin{aligned} \frac{\partial^2}{\partial \mu_c \partial \mu_{c'}} \ln L\left(\mu_1, \mu_2, \ldots, \mu_{C-1}\right) &= \frac{\partial}{\partial \mu_{c'}} \left\{o_{c\cdot} - n p_{c\cdot}\right\} - \left\{o_{\cdot c} - n p_{\cdot c}\right\} \\ &= -n \left\{\left(p_{c\cdot} + p_{\cdot c}\right) 1_{\{c=c'\}} - \left(p_{cc'} + p_{c'c}\right) 1_{\{c \neq c'\}} - \left(p_{c\cdot} - p_{\cdot c}\right)\left(p_{c'\cdot} - p_{\cdot c'}\right)\right\}. \end{aligned} \tag{4.3}$$

4.2.1 Overall test

To test the null hypothesis of no differences across the alternatives, we compute the score test, see Rao (1948). The score test is also known as the Lagrange Multiplier test, see Aitchison and Silvey (1958) and Buse (1982). Under the usual regularity conditions, the score test is asymptotically optimal, see Lindsey (1996, p. 300). The score test is commonly used as a specification test. Here we shall use it to test the specification of a model under which each arrangement of the ranks is equally likely.

Let $\mathbf{U}$ denote the score vector, that is the $(C-1)$-dimensional vector with elements (4.2). The score test is defined as

$$LM = \mathbf{U}_0^T \mathbf{J}_0^{-1} \mathbf{U}_0, \tag{4.4}$$

where $\mathbf{U}_0$ and $\mathbf{J}_0$ are respectively the score vector and the information matrix evaluated under H_0. Under H_0 (4.4) has approximately a chi-square distribution with $C-1$ degrees of freedom.

Under H_0 we have

$$p_{c\cdot} = p_{\cdot c} = \frac{1}{C}, \quad p_{cc'} = p_{c'c} = \frac{1}{C(C-1)},$$

and hence the cth element of $\mathbf{U}_0$ is given by

$$u_{0,c} = o_{c\cdot} - o_{\cdot c}.$$

Moreover, one may show that

$$\mathbf{J}_0 = \frac{2n}{C-1}\left(\text{diag}\,(\iota) - \frac{1}{C}\iota\iota^T\right), \tag{4.5}$$

where $\text{diag}(\iota)$ denotes a diagonal matrix with elements one on the diagonal and ι is a $(C-1)$-dimensional vector of ones.

According to the Sherman-Morrison-Woodbury formula, see Hager (1989), we may write

$$\mathbf{J}_0^{-1} = \frac{C-1}{2n}\left(\text{diag}\,(\iota) + \iota\iota^T\right),$$

and hence we may rewrite (4.4) as

$$LM = \frac{C-1}{2n}\sum_{c=1}^{C}\left(o_{c\cdot} - o_{\cdot c}\right)^2. \tag{4.6}$$

That is, the score test in this case is proportional to the sum of squared differences between the number of times an alternative is chosen as most preferred minus the number of times it is chosen as least preferred over all C alternatives.

4.2.2 Multiple comparisons

When the null hypothesis is rejected, there is sufficient evidence to conclude that there exist differences across the alternatives. However, we do not know which alternatives differ, so we present a multiple comparison procedure to examine all $C(C-1)/2$ pairwise comparisons.

The multiple comparison procedure could be interpreted as a hypothesis test for each pairwise comparison, that rejects the subhypothesis $H_{0,cc'}$ of no difference between alternative c and alternative c' if and only if

$$|u_{0,c} - u_{0,c'}| \geq r^{\alpha}_{C,n}, \tag{4.7}$$

where the critical constant $r^{\alpha}_{C,n}$ is chosen to make the type I error rate equal to α. That is, $r^{\alpha}_{C,n}$ is the largest constant such that

$$P_{H_0}\left(\max u_{0,c} - \min u_{0,c'} \geq r^{\alpha}_{C,n}\right) \leq \alpha.$$

This implies that when H_0 is true all $C(C-1)/2$ inequalities in (4.7) fail to exceed the critical constant $r^{\alpha}_{C,n}$ with probability α. Hence, multiple comparisons are only of interest if the global null hypothesis H_0 is rejected. When the global null hypothesis H_0 is not rejected, it is generally agreed that all hypotheses implied by that hypothesis must also be considered as not rejected.

Define $u_{0,C} = o_{C\cdot} - o_{\cdot C}$, it follows from (4.5) that the covariance matrix of $u_{0,c} - u_{0,c'}$ is the same as the covariance matrix of $Z_c - Z_{c'}$, where $Z_1, \ldots, Z_k$ are independent random variables with zero means and variances $2n/(C-1)$. Hence, the asymptotic distribution of

$$\frac{\max_{c,c'} |u_{0,c} - u_{0,c'}|}{\sqrt{2n/(C-1)}}$$

coincides with the distribution of the range of C independent standard normal random variables. When n is large, the critical constant $r^{\alpha}_{C,n}$ can be approximated by

$$r^{\alpha}_{C,n} \approx q^{\alpha}_{C} \sqrt{\frac{2n}{C-1}}, \tag{4.8}$$

where q_C^α is the upper α percentile point of the range of C independent standard normal random variables. The percentile points q_C^α can be found in Harter (1960).

Multiple comparisons plot

To visualize the multiple comparisons (4.7), we construct an interval Q_c for each alternative c centered at $u_{0,c}$ with length $r_{C,n}^\alpha$ and endpoints

$$u_{0,c} \pm \frac{r_{C,n}^\alpha}{2}. \tag{4.9}$$

The intervals $Q_1, \ldots, Q_C$ are simultaneously displayed in a plot. When intervals Q_c and $Q_{c'}$ do not overlap, the distance between $u_{0,c}$ and $u_{0,c'}$ exceeds $r_{C,n}^\alpha$ and hence (4.7) should be rejected, yielding the conclusion that there is a significant difference between alternative c and alternative c'.

Homogeneous subsets

Based on the multiple comparisons we can identify subsets of alternatives by cluster analysis. The corresponding distance matrix summarizes the hypothesis tests (4.7) by zero's (non rejection) and ones (rejection). It is well known that such a distance matrix of zero's and ones could lead to multiple solutions in complete linkage clustering. Therefore, when alternatives c and c' are not significantly different, we multiply $|u_{0,c} - u_{0,c'}|$ by ϵ and add this in the distance matrix. As adding $|u_{0,c} - u_{0,c'}|\,\epsilon$ may not disguise the multiple comparisons results, ϵ should be chosen small enough like 0.001.

4.3 Nonparametric approach to Best-Worst ranking

The same results can be obtained by a nonparametric approach, which utilizes the Best-Worst ranking directly, and connects to the well-known Friedman statistic. The nonparametric approach also allows us to further improve the interpretation of the multiple comparisons plot as the average rank is easier to understand than the score.

For each of n respondent, assign rank value 1 to the most preferred alternative s and rank value C to the least preferred alternative t. Denote the observed rankings of respondent i for alternative c by x_{ic}. Note that for each respondent we have three groups of equal observations. Under H_0 all rank permutations within each group are equally likely.

Table 4.1: Computation of weighted and adjusted ranks for a hypothetical Best-Worst dataset for one respondent.

Alternative	x_{ic}	$R(x_{i1},\ldots,x_{iC})$						$\bar{r}_{ic}$	a_{ic}
A	-	2	2	3	3	4	4	18/6=3	0
B	-	3	4	2	4	2	3	18/6=3	0
C	1	1	1	1	1	1	1	6/6=1	-2
D	5	5	5	5	5	5	5	30/6=5	2
E	-	4	3	4	2	3	2	18/6=3	0

Let $R(x_{i1},\ldots,x_{iC})$ be the set of all possible rankings consistent with $x_{i1},\ldots,x_{iC}$ given the observed rankings for respondent i. The weighted rank $\bar{r}_{ic}$ assigned to x_{ic} is defined as the average of all possible ranks within the set $R(x_{i1},\ldots,x_{iC})$.

In Table 4.1 we illustrate the computation of $\bar{r}_{ic}$. This respondent indicates alternative C as most preferred and alternative D as least preferred from a set of $C = 5$ alternatives. The set of all possible rankings $R(x_{i1},\ldots,x_{iC})$ is listed in the third column. The weighted rank $\bar{r}_{ic}$ for alternative c is the average of all its possible ranks and is listed in the fourth column.

4.3.1 Overall test

The test statistic for H_0 is based on sums of adjusted ranks. The adjusted rank a_{ic} is obtained by subtracting the expected rank $(C+1)/2$ under H_0 from the weighted rank $\bar{r}_{ic}$. In our example, the expected rank under H_0 is 3 and the adjusted ranks a_{ic} are listed in the last column of Table 4.1. The adjusted ranks can be summarized in a vector a_+, by summing all adjusted ranks for each alternative over the n respondents, that is, element c of vector a_+ is given by $a_{+c} = \sum_i a_{ic}$. As a_{+c} equals the number of times that alternative c is chosen as most preferred minus the number of times that it is chosen as least preferred multiplied by $-(C-1)/2$, it follows that

$$a_{+c} = -\left(\frac{C-1}{2}\right) u_{0,c}. \tag{4.10}$$

The random vectors $(a_{i1},\ldots,a_{iC})$ are independent under H_0 and the covariance matrix V_+ of the vector a_+ is thus equal to the sum of the individual covariance matrices V_i. The

covariance matrix V_i is defined in Wittkowski (1988) as

$$V_i = A_0^2 \left(\text{diag}\ (\iota) - \frac{\iota \iota^T}{C} \right), \tag{4.11}$$

where ι is a C-dimensional vector of ones, diag (ι) denotes a diagonal matrix with elements one on the diagonal and A_0^2 denotes the individual variance under H_0. In Best-Worst ranking we have $A_0^2 = (C-1)/2$ and it follows that the individual covariance matrix (4.11) is given by the diagonal elements $(C-1)^2/2C$ and the off-diagonal elements $(1-C)/2C$. Note that V_i depends only on C and as a consequence, V_i is the same for each respondent. Hence, (4.11) implies $V_+ = nV_i$.

The test statistic is now computed along standard lines as

$$W = a'_+ \left(V_+\right)^- a_+, \tag{4.12}$$

where $(V_+)^-$ denotes a generalized inverse of V_+, that is, any matrix which satisfies $V_+ (V_+)^- V_+ = V_+$. Under H_0 and for large n, W has approximately a chi-square distribution with $C-1$ degrees of freedom. As already suggested by (4.10), test (4.12) and the score test (4.6) are equivalent.

4.3.2 Multiple comparisons

As in the likelihood approach, we present a multiple comparison procedure to examine which alternatives differ in preference rankings when H_0 is rejected.

To enhance interpretability we will use the average weighted ranks $\bar{r}_c = \sum_i \bar{r}_{ic}/n$, rather than the average adjusted ranks $\bar{a}_c = \sum_i a_{ic}/n$. Note that

$$\bar{r}_c = \bar{a}_c + \frac{C+1}{2},$$

and thus $|\bar{a}_c - \bar{a}_{c'}| = |\bar{r}_c - \bar{r}_{c'}|$. The subhypothesis $H_{0,cc'}$ of no difference between alternative c and alternative c' is given by

$$|\bar{r}_c - \bar{r}_{c'}| \geq r^{\alpha}_{C,n}, \tag{4.13}$$

where the critical constant $r^{\alpha}_{C,n}$ is defined as before.

Under H_0, one may show that the covariance matrix of $\bar{a}_c - \bar{a}_{c'}$ coincides with the covariance matrix of the differences $Z_c - Z_{c'}$ where $Z_c, Z_{c'}$ are independent random variables with mean zero and variance $(C-1)/2n$. Hence, when n is large, $r^{\alpha}_{C,n}$ can

be approximated by

$$r_{C,n}^{\alpha} \approx q_C^{\alpha} \sqrt{\frac{C-1}{2n}}, \tag{4.14}$$

where q_C^{α} is defined as before. The intervals $Q_1, \ldots, Q_C$ are thus given by endpoints

$$\bar{r}_c \pm \frac{r_{C,n}^{\alpha}}{2}, \tag{4.15}$$

for alternative c.

Note that the critical value (4.14) is proportional to (4.8) and consequently, (4.13) must yield the same (non-)rejections as in (4.7). As the cluster analysis is based on the results of the multiple comparisons, it should also yield the same homogeneous subsets of alternatives. The weighted rank lends out to a better interpretation than the score, for example, we may interpret results of the cluster analysis as subsets of alternatives with the same within-cluster rank.

Moreover, remark that we have assigned rank value 1 to the most preferred alternative, whereas in the likelihood approach, the most preferred alternative has the highest score. Hence, the intervals should be at the mirror positions of the ones obtained in the likelihood approach.

4.4 Illustration Best-Worst ranking

To illustrate and compare the two different approaches, we analyze a Best-Worst data set concerning winter sports holiday packages obtained from $n = 169$ respondents.

4.4.1 Data

Before performing the BW ranking task, we introduce the respondents to the following hypothetical situation.

> *Suppose you have won an voucher worth EUR 750,- which you can spend on a winter sports holiday. This winter sports holiday must take place between 1 December 2008 and 31 March 2009. Assume that you are able to plan two week holidays during this period. This holds also for your travel companions. You may decide when and where to go. Please note that the voucher is strictly personal and can only be used for your part of the holiday and accommodation costs. If the total costs are less then EUR 750,- will be given the remainder as a voucher for a winter sports holiday next year.*

Table 4.2: List of $C = 8$ winter sports holiday destinations.

Destination	a_{+c}	$\bar{r}_c$
1. Valmorel (France)	87.5	5.018
2. Val Thorens (France)	-283.5	2.822
3. Mayrhofen (Austria)	31.5	4.686
4. Gerlos (Austria)	-49.0	4.210
5. Kirchberg (Austria)	45.5	4.769
6. Livigno (Italy)	196.0	5.660
7. Alpe d'Huez (France)	-24.5	4.355
8. Pas de la Casa (Andorra)	-3.5	4.479

These details are provided to standardize the decision situation across all respondents, see Rewtrakunphaiboon and Oppewal (2008), and controls for budget, holiday period and the right to take decisions by the respondent. Rebates were available to discourage the respondent to choose the most expensive holiday package to use the entire voucher at once. Respondents are then presented a list of eight winter sports holiday packages to spend their voucher. Each package description contains information about the period of holiday, price, general information about the destination, information about accommodation and the prices of ski and snowboard lessons and material rent. Respondents indicate the best package option and worst package option from this list. The winter sports holiday destinations are listed in Table 4.2.

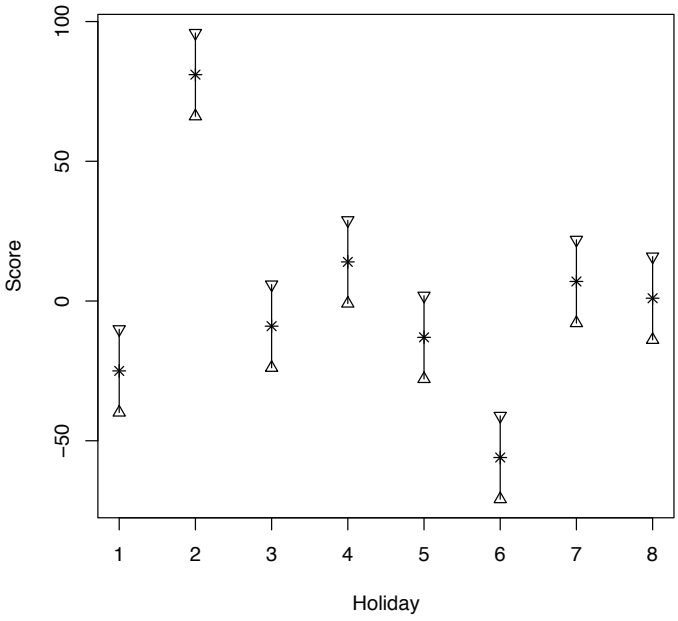
4.4.2 Likelihood approach

Our null hypothesis of no differences across the winter sports holidays is clearly rejected as the score test (4.6) takes the value 224.041 with corresponding p-value is 0. As H_0 is rejected, the question remains which winter sports holidays differ. For $\alpha = 0.05$ and $C = 8$, we obtain $q_{\alpha,C} = 4.286$ and $r_{\alpha,C,n}$ can approximated by (4.8) and takes the value 29.783.

We summarize the multiple comparisons (4.7) in a plot, which simultaneously displayed the intervals (4.9). We observe in Figure 4.1 that the most preferred winter sports holiday is Val Thorens in France as it has the highest score, i.e. Val Thorens in France has been observed most often as most preferred winter sports holiday after subtracted the times that it was observed as least preferred.

Next, we perform a complete linkage clustering (see e.g. Lattin, Carroll, & Green, 2003) based on the multiple comparisons as described. The corresponding dendrogram is

Figure 4.1: Rank plot with $\alpha = 0.05$ of $C = 8$ winter sports holiday packages.

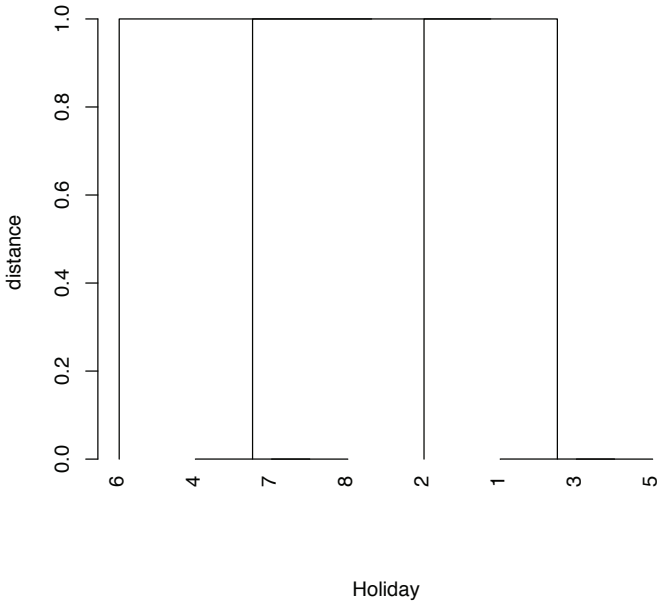


displayed in Figure 4.2. The cluster analysis suggests four main clusters. The first cluster contains the most preferred holiday: Val Thorens in France. A second cluster contains the least preferred holiday: Livigno in Italy. A third cluster contains the holidays: Gerlos (Austria), Alpe d'Huez (France) and Pas de la Case (Andorra). The last cluster contains the holidays: Valmorel (France), Mayrhofen (Austria) and Kirchberg (Austria).

4.4.3 Nonparametric approach

Note that for each respondent we obtain three groups of equal observations. Remark that all rank permutations within each group are equally likely under H_0 and hence, the weighted ranks are $\bar{r}_{ic} = 1$ for the most preferred holiday, $\bar{r}_{ic} = 8$ for the least preferred holiday and $\bar{r}_{ic} = 4.5$ else. Adjusted ranks are obtained by subtracting the expected rank 4.5 from the weighted ranks. Furthermore, $A_0^2 = 3.5$ and V_i is given by 3.0625 on the diagonal and -0.4375 on the off-diagonal elements. The sum adjusted

Figure 4.2: Dendrogram from cluster analysis based on multiple comparisons of $C = 8$ winter sports holiday packages.

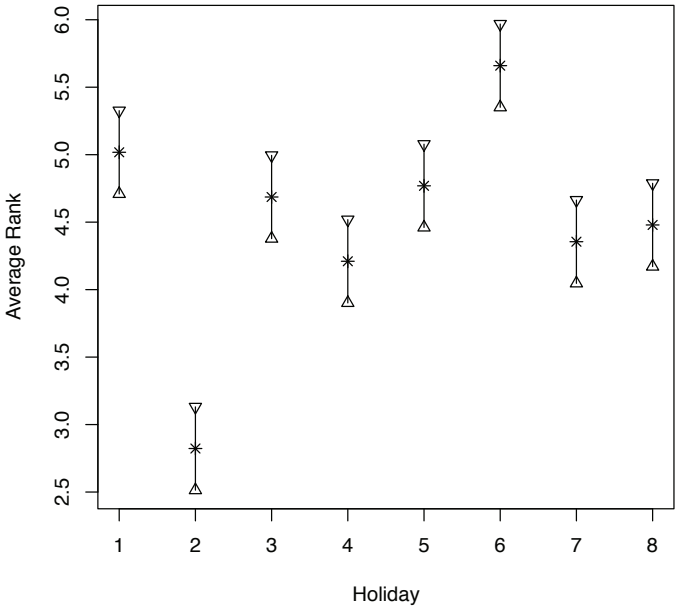


ranks over all respondents for each destination is given in the second column of Table 4.2. The conditional covariance matrix V_+ is obtained by summing all individual covariance matrices V_i.

The test statistic (4.12) takes the value $W = 224.041$ with corresponding p-value is 0 and thus we reject H_0. Remark that this value is exactly the same as in the likelihood approach.

The critical constant $r^{\alpha}_{C,n}$ in our multiple comparison procedure can be approximated by (4.14) and $r^{\alpha}_{C,n} = 0.617$. The average weighted ranks $\bar{r}_c$ are given in the last column of Table 4.2. Figure 4.3 visualizes the hypothesis tests (4.13). Remark that this plot is similar to the plot obtained by the likelihood approach. The intervals (4.15) are on the opposite of the intervals (4.9). Moreover, note that the scale in this plot is better interpretable as it is just the average weighted rank $\bar{r}_c$.

Figure 4.3: Rank plot with $\alpha = 0.05$ of $C = 8$ winter sports holiday packages.



Remark that (4.13) and (4.7) are equivalent tests and thus yield the same multiple comparisons results. Hence, we obtain the same homogeneous subsets in the corresponding cluster analysis.

4.5 Nonparametric approach to other ranking schemes

In contrast to the likelihood approach, the nonparametric approach is easily extended to other ranking schemes, which we will discuss in this section.

Suppose there are C alternatives, with $c = 1, \ldots, C$, and that there are n respondents, with $i = 1, \ldots, n$, who are asked to indicate their preferences according to some predefined partial ranking scheme. In such a partial ranking scheme respondents indicate their most preferred S alternatives and their least preferred T alternatives. Examples of a partial preference ranking scheme are:

- Ask respondents to indicate the most preferred alternative and the least preferred alternative, and assign rank value 1 and rank value C respectively. This scheme yields data corresponding to Best-Worst ranking experiments, which we have discussed before.

- Ask respondents to indicate a subset of S most preferred alternatives and rank this subset, where rank value 1 is assigned to the most preferred alternative. Note that each respondent select S most preferred alternatives, but each respondent can have a different subset of S alternatives.

Typically, a predefined partial ranking scheme differs from a full ranking scheme in the sense that it generates structured groups of equal observations. Note that all rank permutations within each group are equally likely under H_0.

Denote the observed rankings of respondent i for alternative c by x_{ic}. Let $R(x_{i1}, \ldots, x_{iC})$ be the set of all possible rankings consistent with $x_{i1}, \ldots, x_{iC}$ given the observed rankings for respondent i. The weighted rank $\bar{r}_{ii}$ assigned to x_{ii} is defined as the average of all possible ranks within the set $R(x_{i1}, \ldots, x_{iC})$.

4.5.1 Overall test

In general, the individual variance A_0^2 is given by

$$A_0^2 = \frac{C(C+1)}{12} \times \left(1 - \frac{\sum_{g=1}^{G} \left(W_g^3 - W_g\right)}{C^3 - C}\right), \tag{4.16}$$

where W_g is the number of equal observations in group g. Remark that each respondent has to assign ranks according to a predefined ranking scheme and hence (4.16) is the same for all respondents. Consequently, (4.11) implies that the diagonal elements of V_+ are given by $nA_0^2(C-1)/C$ and the off-diagonal elements by $-nA_0^2/C$.

The test statistic is given as in (4.12).

4.5.2 Multiple comparisons

To determine the critical value $r_{C,n}^{\alpha}$, one may show that the covariance matrix of $\bar{a}_c - \bar{a}_c$ coincides with the covariance matrix of the differences $Z_c - Z_{c'}$ where $Z_c, Z_{c'}$ are independent random variables with mean zero and variance A_0^2/n. Hence, the asymptotic distribution of

$$\frac{\max_{c,c'} |\bar{a}_c - \bar{a}_{c'}|}{\sqrt{A_0^2/n}}$$

coincides with the distribution of the range of C independent standard normal random variables. When n is large, $r^{\alpha}_{C,n}$ can be approximated by

$$r^{\alpha}_{C,n} \approx q^{\alpha}_{C}\sqrt{\frac{A_0^2}{n}}, \tag{4.17}$$

where q^{α}_{C} is the upper α percentile point of the range of C independent standard normal random variables.

The multiple comparisons tests are as defined in (4.13) and the intervals $Q_1, \ldots, Q_C$ are given by the endpoints (4.15).

4.6 Illustration

In this section we illustrate our statistical methodology to analyze preference rankings where individuals are asked to rank a subset of alternatives according to some predefined partial ranking scheme.

4.6.1 Top ranked alternatives

Ranked preference data from individuals allows considerably more information to be gathered from a given survey observation than is gathered from simply the most and least preferred alternatives (Hausman & Ruud, 1987). The possibility exists that an individual will pay more attention to his/her top choice or top few choices rather than carefully ranking all alternatives. Hausman and Ruud (1987) results indicate that the top ranked alternatives are done more carefully than the lower ranked choices. Thus, respondents may pay less attention to ranking inferior alternatives (see also Ben-Akiva et al., 1992).

Hence, suppose there are C alternatives, with $c = 1, \ldots, C$, and that there are n respondents, with $i = 1, \ldots, n$, who are asked to indicate and rank their top S alternatives, where the most preferred alternative is assigned rank value 1. Note that each respondent select S alternatives, but each respondent can have a different subset of S alternatives.

In this ranking scheme, the individual variance (4.16) under H_0 is given by

$$A_0^2 = \frac{C(C+1)}{12}\left(1 - \frac{C-S-1}{C-1}\frac{C-S}{C}\frac{C-S+1}{C+1}\right).$$

Table 4.3: List of $C = 10$ blockbuster movies in Dutch cinema theatres in 2007, ranked in the first column according to the total size of audience.

Name movie	a_{+c}	$\bar{r}_c$
1. Pirates of the Caribbean: At World's End	-79.5	4.645
2. Harry Potter and the order of the Phoenix	1.0	5.511
3. Alles is Liefde	40.5	5.936
4. Shrek the Third	-6.0	5.435
5. Mr. Bean's Holiday	150.0	7.113
6. Ratatouille	1.5	5.516
7. Ocean's Thirteen	-66.5	4.785
8. Spider-Man 3	76.5	6.323
9. Transformers	5.5	5.559
10. The Bourne Ultimatum	-123.0	4.177

4.6.2 Data

We illustrate our proposed methodology with data of $n = 93$ individuals who are asked to evaluate a list of $C = 10$ blockbuster movies in Dutch cinema theatres in 2007. The movies are listed in Table 4.3.

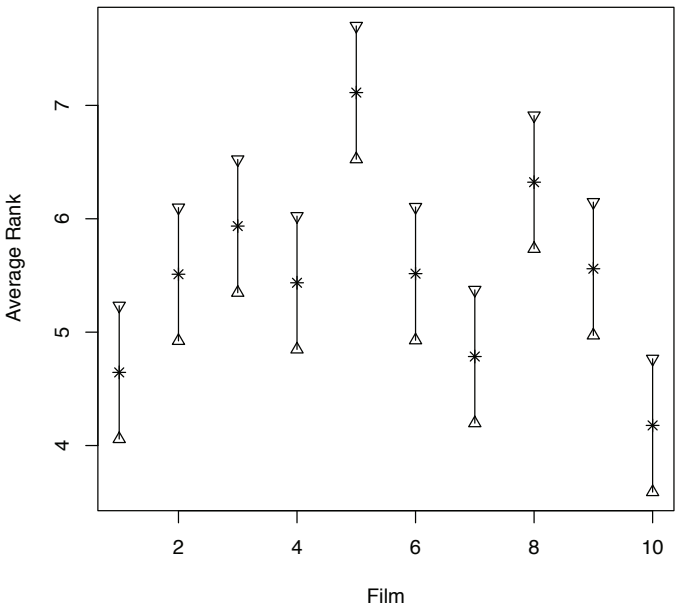
Respondents are asked to indicate and rank their top $S = 4$ movies and the weighted rank $\bar{r}_{ic}$ for movies outside this subset is 7.5. Adjusted ranks a_{ic} are obtained by subtracting the expected rank 5.5, from the weighted ranks. The individual variance $A_0^2 = 65/9$ and the individual covariance matrix V_i is given by $65/10$ on the diagonal and $-65/90$ on the off-diagonal elements. The sum of adjusted ranks over all respondents for each movie is given in the second column of Table 4.3. The conditional covariance matrix V_+ is obtained by summing all individual covariance matrices V_i.

4.6.3 Results

Our null hypothesis of no differences between the movies is clearly rejected as the test statistic (4.12) takes the value $W = 83.276$ with corresponding p-value is 0.

As H_0 is rejected, we apply hypothesis tests (4.13) for each pairwise comparison. For $\alpha = 0.05$ and $C = 10$, we obtain $q_C^\alpha = 4.474$. The critical constant $r_{C,n}^\alpha$ can be approximated by (4.14) and $r_{C,n}^\alpha = 1.247$. The average weighted ranks $\bar{r}_c$ are given in the last column of Table 4.3. The multiple comparisons are summarized in Figure 4.4. We observe in Figure 4.4 that the most favorite movie is "The Bourne Ultimatum". The movies "Pirates of the Caribbean: At World's End" and "Ocean's Thirteen", have overlap

Figure 4.4: Rank plot with $\alpha = 0.05$ of $C = 10$ blockbuster movies in Dutch cinema theatres in 2007.



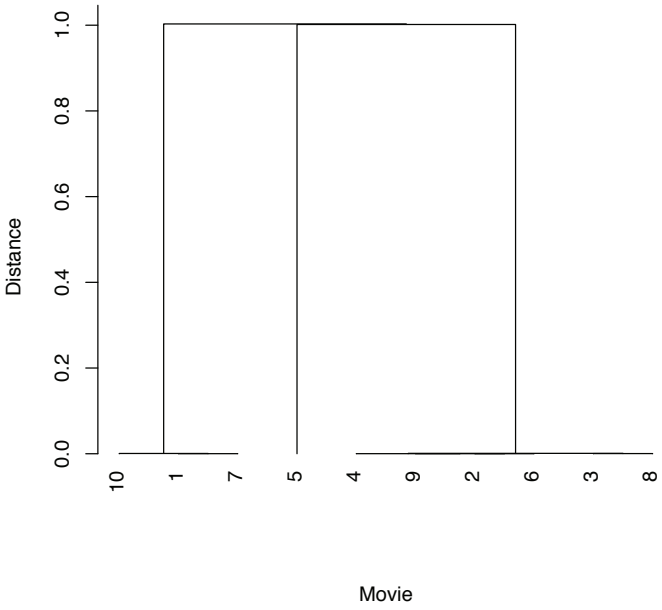
with the interval of the movie "The Bourne Ultimatum" and hence are not significantly ranked lower.

In Figure 4.5 we observe that the cluster analysis suggests three main clusters. Cluster 1 contains the most favorite movies: "The Bourne Ultimatum", "Pirates of the Caribbean: At World's End" and "Ocean's Thirteen". Cluster 2 contains the least favorite movie (5): "Mr. Bean's Holiday" and the last cluster contains all other movies. In sum, there seems to be just three clusters of movies with the same within-cluster rank.

4.7 Conclusion

We have derived the score test for Best-Worst ranking data. The score test has optimality characteristics and is commonly used as a specification test. Here, we have used it to test the specification of a model in which each arrangement of the ranks is equally likely.

Figure 4.5: Dendrogram from cluster analysis based on multiple comparisons of $C = 10$ blockbuster movies in Dutch cinema theatres in 2007.



Moreover, we introduced a multiple comparison procedure to examine differences across alternatives when the null hypothesis is rejected. To the best of our knowledge, this methodology has not yet been applied in a marketing context.

We have shown that the same results can also be obtained by a nonparametric approach, which enabled us to enhance interpretation of the results. Moreover, we showed that our proposed nonparametric test and the score test are equivalent, which implies that the far more simpler nonparametric approach is as good as the likelihood approach. In addition, the nonparametric approach also allows for a better interpretation of the multiple comparisons results, as (the average) ranks are easier to understand than the scores.

An advantage of the nonparametric approach over the likelihood approach is that the nonparametric approach can be easily extended to general ranking schemes. This allows

us to efficiently gather more preference information than gathered from the most and least preferred alternatives.

Remark that our proposed methodology should be considered in the context of a preliminary analysis of preference rankings when there are too many alternatives. As our purpose is to give directions for further analysis, we do not focus on estimating parameters. The null hypothesis just serves as a starting point from where the model is improved.

Consider the situation where a marketer has such a large amount of alternatives that it is hard to decide beforehand which alternatives are of interest to include in the experiment. One may suggest to apply a partial profile design to overcome this. However, it is well known in practice that with such a design respondents are inconsistent, in the sense that their ranking fails to meet the transitivity criterium.

A preliminary analysis with our methodology, reduces the amount of alternatives in an easy and understandable way, such that irrelevant alternatives can be removed from further analysis.

Another practical application is, for example, in conjoint analysis, where the number of alternatives grows exponentially as the number of attributes increases. Thus, when there are many attributes to deal with, the technique can be used to exclude irrelevant attributes.

Chapter 5

Ranking Models in Conjoint Analysis

Abstract

In this paper we consider the estimation of probabilistic ranking models in the context of conjoint experiments. By using approximate rather than exact ranking probabilities, we avoid the computation of high-dimensional integrals. We extend the approximation technique proposed by Henery (1981) in the context of the Thurstone-Mosteller-Daniels model to any Thurstone order statistics model and show that our approach allows for a unified approach. Moreover, our approach also allows for the analysis of partial rankings, which are an essential part of practical conjoint analysis in collecting data efficiently as a means of relieving the respondents' task burden. We apply our approach to the reanalysis of the career preference data set described in Maydeu-Olivares and Böckenholt (2005) and to a holiday preferences data set.

5.1 Introduction

Ranking stimuli is a simple means of measuring preferences. Metric measurements such as rating and matching may be less reliable due to respondents' limited ability to accurately report degrees of preferences, see P. E. Green and Srinivasan (1978); Ben-Akiva et al. (1992). Two complementary approaches to modeling preferences among multi-attribute stimuli can be distinguished: conjoint measurement and conjoint analysis. Conjoint measurement, see Luce and Tukey (1964); Krantz and Tversky (1971); Barron (1977), investigates whether there are interval scales for each of the attributes which can be combined according to some prespecified functional form and which are consistent with the preferences. Conjoint analysis, see Kruskal (1965); Johnson (1974); P. E. Green and Srinivasan (1978), aims at estimating interval scales assuming that a specific functional form applies.

Although conjoint analysis originated more than forty years ago, it continues to attract active interest as a research field, see P. E. Green, Krieger, and Wind (2001); Bradlow (2005). According to Hauser and Rao (2004), full-profile analysis remains the most common form of conjoint analysis and has the advantage that the respondent evaluates each profile holistically and in the context of all attributes. Its weakness is that the respondent's burden grows dramatically with the number of stimuli that must be ranked or rated.

In full-profile conjoint experiments, each respondent evaluates and ranks a set/subset of stimuli, where each stimulus is defined as a specific combination of attributes levels. As the number of possible rankings is finite, the rankings have a discrete distribution. In principle, standard methods for analyzing discrete data apply here, see Marden (1995, p. 140). Unfortunately, probability models for rankings become rather complex, as the computation of each ranking probability usually requires high-dimensional integration when the number of stimuli becomes large. Earlier approaches for analyzing conjoint experiments thus avoid the use of probability models for rankings by resorting to multi-dimensional scaling techniques to derive respondent preferences (for details, see P. E. Green and Rao (1971)). Recently, there is however a renewed interest in the modeling and estimation of ranking models, see Maydeu-Olivares (1999); Maydeu-Olivares and Böckenholt (2005); Böckenholt (2006); Maydeu-Olivares and Hernández (2007).

In this paper we consider the estimation of probabilistic ranking models in the context of full-profile conjoint experiments. We reduce the complexity of probabilistic ranking models considerably by using approximate rather than exact ranking probabilities. In the literature an abundance of ranking models is available, but we concentrate on Thurstone order statistics models, to be introduced in the next section. In Henery (1981), a simpler model approximated the Thurstone-Mosteller-Daniels model. We show that any Thurstone order statistics model may be approximated by such a "Henery model". This allows for a unified approach.

Moreover, our approach also allows for the analysis of partial rankings. Methods which result in partial rankings are essential in practical conjoint analysis to collect data efficiently to relieve respondents' task burden. A specific partial ranking method gaining in popularity is Best-Worst ranking, see Louviere and Woodworth (1991b); Finn and Louviere (1992). In this specific case, respondents are instructed to select only the best and the worst stimulus. Detailed discussions of the Best-Worst ranking method are found in Marley and Louviere (2005); Chrzan and Golovashkina (2006); Flynn, Louviere, Peters, and Coast (2007). Partial ranking methods may also be dictated by the experimental

design of the conjoint study. For example, each respondent may be asked to rank only a strict subset of all stimuli used in the study.

The structure of our paper is as follows. First, we discuss ranking models and the approximation of ranking probabilities. Then, we adapt general ranking models to conjoint experiments by introducing a linear model which allows for modeling the dependence of the rankings on the stimulus characteristics. Next, we illustrate our methodology on the career preference data set described in Maydeu-Olivares and Böckenholt (2005) and compare our results to those obtained in that paper. Further, we analyze a holiday preferences data set, which provides a realistic multi-attribute example and demonstrates the use of partial rankings. Finally, we conclude with suggestions for further research. Most technical issues are relegated to the appendix.

5.2 Methods

5.2.1 Ranking probabilities

In this section, we consider a single respondent who lists all stimuli, $1, 2, \ldots, C$, in order of preference, with the most preferred stimulus listed first. For each stimulus c in $\{1, 2, \ldots, C\}$, we define the rank $\pi_{(c)}$ of c as the position of c within this ordering. For example, $\pi_{(3)} = 7$ indicates that stimulus 3 is listed in the 7^{th} place in order of preference. We shall refer to $\pi = \left(\pi_{(1)}, \pi_{(2)}, \ldots, \pi_{(C)}\right)$ as a full ranking.

Observe that in a full ranking for each rank r there exists exactly one stimulus c such that $\pi_{(c)} = r$. We shall denote this stimulus by $\pi_{(r)}^{-1}$. For example, $\pi_{(7)}^{-1} = 3$ denotes that stimulus 3 is listed in 7^{th} place in order of preference. Remark that we now may express the ordering as $\pi^{-1} = \left(\pi_{(1)}^{-1}, \pi_{(2)}^{-1}, \ldots, \pi_{(C)}^{-1}\right)$.

We assume that the probability p_π of actually obtaining π as a full ranking depends on a C-dimensional linear predictor vector $\boldsymbol{\eta} = (\eta_1, \eta_2, \ldots, \eta_C)^\top$, that is,

$$p_\pi = p\left(\pi \mid \boldsymbol{\eta}\right). \tag{5.1}$$

A so-called ranking model specifies the exact nature of the dependence of p_π on $\boldsymbol{\eta}$.

A common issue in the analysis of rankings is the handling of ties. A tie means that the same rank is assigned to multiple stimuli because the respondent is enable to differentiate between two or more stimuli. Ties may also occur due to requirements imposed by the research design. It has been widely recognized that respondents may find it difficult to compare too many choice options. This can be solved by asking respondents to rank only

a subset of stimuli. For instance, in Best-Worst ranking, a respondent is instructed to select only the best and the worst stimulus. By definition, all the other stimuli are then tied.

Another common issue in the analysis of rankings is the handling of missing values. A missing value means that no rank was assigned to a stimulus and may occur due to requirements imposed by the research design. For example, to alleviate respondents task complexity, respondents could be shown only a subset of all stimuli. Missing values differ from ties in that a missing value could have been assigned any rank r in $\{1, 2, \ldots, C\}$.

A ranking containing ties or missing values in fact corresponds to a partial ordering of the stimuli rather than to a full ordering. Hence, we shall refer to such a ranking as a partial ranking. Observe that for each partial ranking ϖ there exists a set $\mathcal{S}_\varpi$ of all full rankings which do not contradict the partial ordering implied by ϖ. Thus, we may assign the probability

$$p_\varpi = p\left(\varpi \mid \boldsymbol{\eta}\right) = \sum_{\pi \in \mathcal{S}_\varpi} p\left(\pi \mid \boldsymbol{\eta}\right) \tag{5.2}$$

to the partial ranking ϖ.

5.2.2 Random utility models

There is a bewildering number of ranking probability models. For complete overviews, see Critchlow et al. (1991, Section 3) and Marden (1995, Chapter 5). In this subsection, we focus on random utility models, see Thurstone (1927); Luce and Suppes (1965); Böckenholt (2006). Random utility models assume that the rank of stimulus c among the stimuli $1, 2 \ldots, C$ is in fact equal to the rank of a random variable Y_c among the random variables $Y_1, Y_2, \ldots, Y_C$. Here, $Y_1, Y_2, \ldots, Y_C$ are random variables having some joint continuous distribution. It follows that

$$p_\pi = P\left(Y_{c_1} < Y_{c_2} < \ldots < Y_{c_C}\right) \tag{5.3}$$

for an ordering $\pi^{-1} = (c_1, c_2, \ldots, c_C)$. Observe that

$$p_\pi = \int_{-\infty}^{\infty} \int_{y_{c_1}}^{\infty} \int_{y_{c_2}}^{\infty} \cdots \int_{y_{c_{C-1}}}^{\infty} f\left(y_{c_1}, y_{c_2}, \ldots, y_{c_C}\right) dy_{c_C} \cdots dy_{c_2} dy_{c_1}, \tag{5.4}$$

where $f\left(y_{c_1}, y_{c_2}, \ldots, y_{c_C}\right)$ denotes the joint density of $Y_{c_1}, Y_{c_2}, \ldots, Y_{c_C}$.

An obvious choice of $f\left(y_{c_1}, y_{c_2}, \ldots, y_{c_C}\right)$ is the multivariate normal density, see Thurstone (1927, p. 285). In addition, we may assume that the corresponding covariance

matrix obeys some structure, see Thurstone (1927, p. 286). In particular, Maydeu-Olivares (1999); Maydeu-Olivares and Böckenholt (2005); Böckenholt (2006) connect to the theory of covariance structures proposed in Browne (1982). Note that the multivariate normal density only allows linear dependence. For other forms of dependence, see Joe (1997).

Thurstone order statistics models, see Critchlow et al. (1991, p. 298), are the most popular random utility models in practice. These models assume that $Y_1, Y_2, \ldots, Y_C$ are independent random variables with distributions from the same family with density $g(Y, \eta)$. Under this assumption, $Y_{c_1}, Y_{c_2}, \ldots, Y_{c_C}$ have joint density

$$f\left(y_{c_1}, y_{c_2}, \ldots, y_{c_C}\right) = g\left(y_{c_1}; \eta_{c_1}\right) g\left(y_{c_2}; \eta_{c_2}\right) \cdots g\left(y_{c_C}; \eta_{c_C}\right), \tag{5.5}$$

where the parameters $\eta_{c_1}, \ldots, \eta_{c_C}$ are allowed to vary. The independence assumption is reasonable when $\eta_{c_1}, \ldots, \eta_{c_C}$ depend on the attributes of the stimuli under consideration, and these attributes are able to fully "explain" the dependence between $\eta_{c_1}, \ldots, \eta_{c_C}$. Combining (5.4) and (5.5) yields

$$p_\pi = \int_{-\infty}^{\infty} \int_{y_{c_1}}^{\infty} \cdots \int_{y_{c_{C-1}}}^{\infty} g\left(y_{c_1}; \eta_{c_1}\right) g\left(y_{c_2}; \eta_{c_2}\right) \cdots g\left(y_{c_C}; \eta_{c_C}\right) dy_{c_C} \cdots dy_{c_2} dy_{c_1}. \tag{5.6}$$

Thurstone order statistics models are sometimes called independent random utility models, see Luce and Suppes (1965) and Yellott (1977, p. 142).

Thurstone models, see Critchlow et al. (1991, p. 299), are special cases of Thurstone order statistics models and assume that η is a location parameter, that is, $g(Y, \eta)$ takes the form $g(Y - \eta)$. Special Thurstone models are obtained by making further distributional assumptions.

- The Thurstone-Mosteller-Daniels model, see Critchlow et al. (1991, p. 295), assumes that g is a standard normal density. An alternative name for this model is Thurstone's Case V model, see Yellott (1977, p. 111). However, the alternative name ignores the fact that "Thurstone's original method considers only paired comparisons of the stimuli, but Daniels (1950) extends the method to experiments where the data are full orderings of the C stimuli", see Critchlow et al. (1991, p. 298).

- The Luce model, see Critchlow et al. (1991, p. 300), assumes that g is a Gumbel density. The Luce model is sometimes called Plackett's (1975) first order model.

5.2.3 Approximate probabilities

The multiple integral on the right-hand side of (5.4) is usually evaluated by means of numerical integration. Unfortunately, this approach is not feasible when the number of stimuli becomes large.

Working in the Thurstone-Mosteller-Daniels model, the probability p_π is approximated in Henery (1981) by means of a first order Taylor expansion around $\eta_1 = \eta_2 = \ldots = \eta_C = \eta_0$, where η_0 can take any value. Below, we extend Henery's approach to any model in which $Y_1, Y_2, \ldots, Y_C$ have joint density of the form (5.5).

Introduce

$$\phi_0(y) = \left.\frac{\partial \ln g(y;\eta)}{\partial \eta}\right|_{\eta=\eta_0} = \frac{1}{g(y;\eta_0)} \left.\frac{\partial g(y;\eta)}{\partial \eta}\right|_{\eta=\eta_0}. \tag{5.7}$$

We shall refer to $\phi_0(y)$ as the score function. The score function is well-known in mathematical statistics, especially in likelihood theory and the theory of rank tests. In particular, our definition (5.7) corresponds with Hájek and Šidák (1967, Equation I.2.4.4). An important property of the score function is that it transforms the random variable Y into a random variable $\phi_0(Y)$ with expected value

$$\mathcal{E}\phi_0(Y) = \int \left.\frac{\partial g(y;\eta)}{\partial \eta}\right|_{\eta=\eta_0} dy = 0. \tag{5.8}$$

Here $\mathcal{E}$ denotes the expectation in the general model, that is, without any restrictions on $\eta_1, \eta_2, \ldots, \eta_C$.

Denote the expected score of the r^{th} order statistic $Y_{r:C}$ by

$$q_{r:C} = \mathcal{E}_0\phi_0(Y_{r:C}), \tag{5.9}$$

where $\mathcal{E}_0$ denotes the expectation under the condition that $\eta_1 = \eta_2 = \ldots = \eta_C = \eta_0$. When this condition holds, we show in the Appendix that a first order Taylor expansion in $(\eta_0, \eta_0, \cdots, \eta_0)$ yields

$$p_\pi \approx p_* + p_* \sum_{r=1}^{C} q_{r:C}(\eta_{c_r} - \eta_0) = p_* \left(1 + \sum_{r=1}^{C} q_{r:C}\eta_{c_r}\right) \tag{5.10}$$

for $\pi = (c_1, c_2, \ldots, c_C)$. Here,

$$p_* = \frac{1}{C!}. \tag{5.11}$$

denotes the average full ranking probability (there are $C!$ possible full rankings). Note that $p_\pi = p_*$ if $\eta_1, \eta_2, \ldots, \eta_C$ are all equal, irrespective of their common value η_0.

For a given full ranking π, let $\mathbf{q}_\pi$ denote the C-dimensional vector containing the $\pi_{(c)}^{th}$ expected score $q_{\pi_{(c)}:C}$ as c^{th} element. For example, for $\pi_{(3)} = 7$, the 3^{th} element of vector $\mathbf{q}_\pi$ is the 7^{th} expected score $q_{7:C}$. As $\mathbf{q}_\pi$ is central in deriving an approximation to the probability p_π, we shall refer to $\mathbf{q}_\pi$ as the expected score vector belonging to π. We may now write (5.10) as

$$p_\pi \approx p_* \left(1 + \mathbf{q}_\pi^\top \boldsymbol{\eta}\right). \tag{5.12}$$

The right hand side of (5.12) is not necessarily positive and hence does not necessarily define a valid probability model. However, when all η_c's are sufficiently close to each other, a first order Taylor expansion of the exponential function yields that $\exp\left\{\mathbf{q}_\pi^\top \boldsymbol{\eta}\right\}$ may be approximated by $1 + \mathbf{q}_\pi^\top \boldsymbol{\eta}$. It follows that

$$p_\pi \approx \frac{\exp\left\{\mathbf{q}_\pi^\top \boldsymbol{\eta}\right\}}{\sum_{\pi'} \exp\left\{\mathbf{q}_{\pi'}^\top \boldsymbol{\eta}\right\}}. \tag{5.13}$$

The probabilities on the right hand side of (5.13) are all positive and add up to one, and thus define a probability model with respect to the rankings.

Above we have shown that expected scores allow the approximation of any Thurstone model. For example, one may show that in the Thurstone-Mosteller-Daniels model $q_{r:C}$ coincides with a normal score, that is,

$$q_{r:C} = \mathcal{E} Y_{r:C}, \tag{5.14}$$

where $Y_{r:C}$ denotes the r^{th} order statistic corresponding to the random sample $Y_1, Y_2, \ldots, Y_C$ drawn from a standard normal distribution. In Harter (1961) all normal scores for $C = 400$ are given.

Another Thurstone model is the Luce model, also known as Plackett's first order model, and one may show that in this model

$$q_{r:C} = 1 + \digamma(r) - \digamma(C+1), \tag{5.15}$$

where $\digamma$ is the digamma function, see Abramowitz and Stegun (1964, Section 6.3). Refer to the Appendix for a proof of (5.15).

5.2.4 Incorporating attribute values

In the previous subsections, we have seen how a ranking model translates the predictor vector $\boldsymbol{\eta}$ into a probability distribution on rankings. We now focus on the question how the attributes of the stimuli influence $\boldsymbol{\eta}$. Assume that the stimuli are adequately described by means of M attributes. Each attribute takes a limited number of values, which we call levels. Each stimulus may be viewed as a specific combination of the levels of the attributes. Let x_{cm} denote the value that attribute m takes for stimulus c.

In order to be able to perform a statistical analysis of conjoint experimental data, we have to specify the construction of the predictor vectors. We assume that $\boldsymbol{\eta} = (\eta_1, \eta_2, \ldots, \eta_C)^\top$ is given by

$$\eta_c = \beta_1 x_{c1} + \beta_2 x_{c2} + \ldots + \beta_M x_{cM} = \sum_{m=1}^{M} \beta_m x_{cm}, \tag{5.16}$$

where $\beta_1, \beta_2, \ldots, \beta_M$ are unknown coefficients. We may write $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$, where $\boldsymbol{\beta}$ is the M dimensional coefficient vector $(\beta_1, \beta_2, \ldots, \beta_M)^\top$ and $\mathbf{X}$ is the $C \times M$ matrix which contains the value x_{cm} in its (c, m) location. We shall refer to $\mathbf{X}$ as the plan matrix of the conjoint experiment.

We may rewrite (5.13) as

$$p_\pi \approx \frac{\exp\left\{\mathbf{q}_\pi^\top \mathbf{X}\boldsymbol{\beta}\right\}}{\sum_{\pi'} \exp\left\{\mathbf{q}_{\pi'}^\top \mathbf{X}\boldsymbol{\beta}\right\}}. \tag{5.17}$$

Observe that $\mathbf{q}_\pi^\top \mathbf{X}$ is in fact a weighted average of the columns of the plan matrix, where the weights are completely determined by the preferences in π. Combining (5.2) and (5.17) now yields

$$p_\varpi \approx \frac{\sum_{\pi \in \mathcal{S}_\varpi} \exp\left\{\mathbf{q}_\pi^\top \mathbf{X}\boldsymbol{\beta}\right\}}{\sum_{\pi'} \exp\left\{\mathbf{q}_{\pi'}^\top \mathbf{X}\boldsymbol{\beta}\right\}}. \tag{5.18}$$

Note that (5.17) is actually a special case of (5.18). Although in the next section we shall focus on partial rankings, full rankings are of course implicitly covered.

5.2.5 Approximate log-likelihood

In principle, the maximum likelihood estimator $\hat{\boldsymbol{\beta}}$ of the parameter vector $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_M)^\top$ may be obtained via maximization of the log-likelihood in the Thurstone model. As mentioned earlier, computing p_{π_i} requires the numerical evaluation of the C-dimensional integral (5.4) and hence, is not feasible when the number of stimuli becomes large.

Fortunately, we may approximate p_{π_i} by (5.18) and thus we may estimate $\boldsymbol{\beta}$ by maximizing the corresponding approximate log-likelihood

$$\ln L(\boldsymbol{\beta}) = \sum_{i=1}^{n} \ln \left(\sum_{\pi \in \mathcal{S}_{\varpi_i}} \exp\left\{ \mathbf{q}_{\pi}^{\top} \mathbf{X} \boldsymbol{\beta} \right\} \right) - n \ln \left(\sum_{\pi'} \exp\left\{ \mathbf{q}_{\pi'}^{\top} \mathbf{X} \boldsymbol{\beta} \right\} \right), \tag{5.19}$$

where the rankings $\varpi_1, \varpi_2, \ldots, \varpi_n$ are independently obtained from n different respondents. Note that the standard likelihood theory applies as we have shown that (5.18) is a probability model itself.

In particular, in case of full rankings we have that the log-likelihood (5.19) simplifies to

$$\ln L(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left(\mathbf{q}_{\pi_i}^{\top} \mathbf{X} \boldsymbol{\beta} \right) - n \ln \left(\sum_{\pi'} \exp\left\{ \mathbf{q}_{\pi'}^{\top} \mathbf{X} \boldsymbol{\beta} \right\} \right). \tag{5.20}$$

Standard iterative methods for finding an estimator $\hat{\boldsymbol{\beta}}$ maximizing the log-likelihood (5.19) require the first order derivatives of (5.19) with respect to $\boldsymbol{\beta}$ and possibly the second order derivatives as well. The computation of these are given in the Appendix.

5.3 Results

In this section, we apply our technique to two data sets. First, we analyze the career preference data set described in Maydeu-Olivares and Böckenholt (2005) and compare our estimates to the estimates obtained under the Thurstone-Mosteller-Daniels model in that paper. Next, we analyze a holiday preferences data set, which provides a realistic multi-attribute example.

We should first make a general remark. The smaller its rank, the more preferred is a stimulus. Hence, when interpreting the estimation results, we should always take into account that our preference measure is inversely related to preference. Hence, a positive coefficient indicates that higher levels lead to a higher, i.e. worse, ranking. Consequently, a negative coefficient indicates that higher levels lead to a lower, i.e. more preferred, ranking. The strength of preference is reflected in the absolute value of the coefficient. In addition, a positive coefficient value does not necessarily mean that the respective attribute is rejected, but that it is less preferable than the reference level.

Table 5.1: Plan matrix **X** for career area preferences.

	Academic	*Clinical*	*Educational*
Academic	1	0	0
Clinical	0	1	0
Educational	0	0	1
Industrial	0	0	0

5.3.1 Career area preferences

In Maydeu-Olivares and Böckenholt (2005) career preferences among undergraduate psychology students from a Spanish university were investigated. A sample of 57 psychology students were asked to rank their preferences for four broad psychology career areas, these are *Academic*, *Clinical*, *Educational* and *Industrial*.

The plan matrix **X** contains the dummy variables for the levels *Academic*, *Clinical* and *Educational* of the attribute and is given in Table 5.1. Note that the plan matrix lacks a column containing only 1's, indicating that the intercept has been omitted from the model. This was enforced by (5.25), which shows that removing the intercept from the model does not change the approximate ranking probabilities, and hence that including the intercept brings an unwanted indeterminacy to the model. When using the plan matrix in Table 5.1, we should interpret each of the coefficients for the levels *Academic*, *Clinical* and *Educational* relative to the reference level *Industrial*.

In Table 5.2 the estimated coefficients are reported. Our estimates differ slightly from Maydeu-Olivares and Böckenholt (2005), but the conclusions are the same. Remark that the estimated coefficients are inversely related to preference. The estimated coefficient for *Clinical* differs significantly from zero and is negative, which means that the clinical career area is more preferred than the industrial career area. As the estimated coefficient for *Educational* does not significantly differ from zero, we do not have sufficient statistical evidence to conclude that the educational career area is more preferred than the industrial career area. The academic career area is least preferred, as this coefficient differs significantly from zero and is positive.

In the remainder of this subsection, we use the career data to shed further light on some aspects of the proposed technique. We start by considering the first student in the data set, who orders the career areas as follows: *Clinical*, *Industrial*, *Educational*, *Academic*. The corresponding ranking π_1 is 4132. Rearranging the normal scores (which are -1.0294, -0.2970, 0.2970 and 1.0294 for sample size 4) according to the preferences of

Table 5.2: Estimated coefficients for career area preferences.

Career Area	*Coefficient*	*S.E.*	*p-value*
Academic	1.110	0.276	0.000
Clinical	-1.106	0.269	0.000
Educational	-0.336	0.243	0.166

student 1 yields

$$\mathbf{q}_{\pi_1} = \begin{pmatrix} q_4 \\ q_1 \\ q_3 \\ q_2 \end{pmatrix} = \begin{pmatrix} 1.0294 \\ -1.0294 \\ -0.2970 \\ 0.2970 \end{pmatrix},$$

and hence

	$\mathbf{q}_{\pi_1}^\top \mathbf{X}$
Academic	1.0294
Clinical	-1.0294
Educational	-0.2970

Table 5.3 lists the frequency among the 57 psychology students, the estimated probability according to the Thurstone-Mosteller-Daniels model (using the estimated parameters in Maydeu-Olivares and Böckenholt (2005)), and the estimated probability according to the proposed technique for each possible full ranking of the career areas. At first sight, the estimated probabilities seem close to each other, which is confirmed by the PP-plot in Figure 5.1. This suggests that the approximation (5.13) works well in practice.

However, not all estimated probabilities in Table 5.3 are in agreement with the data. In particular, the probability of ranking 4213 is seriously overestimated. According to the Thurstone-Mosteller-Daniels model, the probability of 4213 is over 11 percent, whereas the empirical probability 1/57 of 4213 is below 2 percent. The PP-plot in Figure 5.2 gives a full display of the discrepancies between estimated Thurstone-Mosteller-Daniels probabilities and empirical probabilities. The estimated probabilities are relatively large for small small empirical probabilities and relatively small for large empirical probabilities. This suggests that some other Thurstone order statistics model may be more appropriate.

Table 5.3: Counts and estimated probabilities for career area preferences. MB: probabilities estimated according to the Thurstone-Mosteller-Daniels model; LKF: probabilities estimated according to the proposed technique.

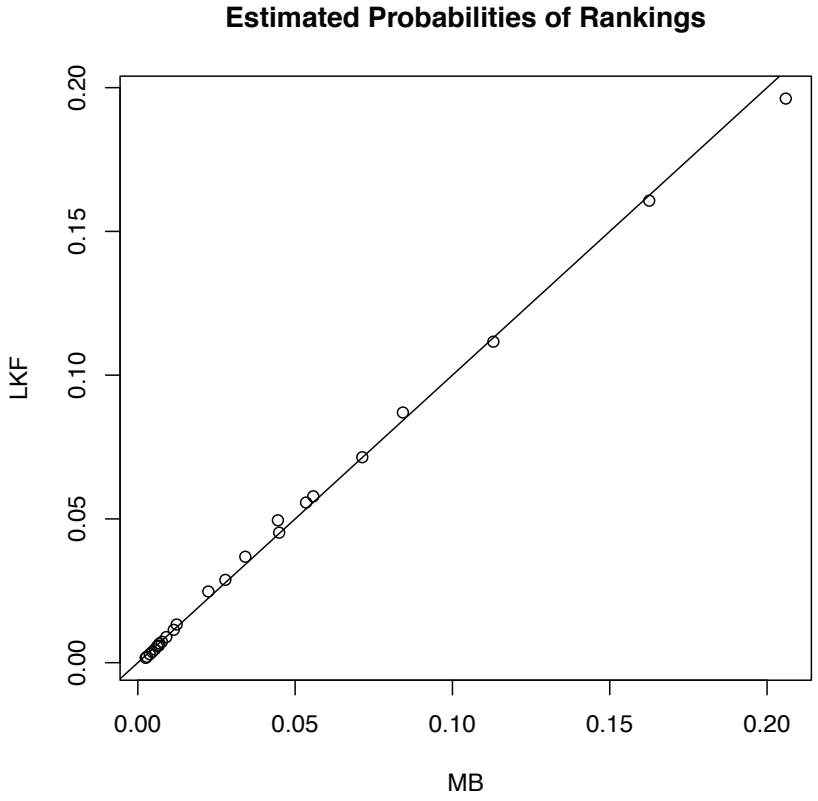
Ordering	*Ranking*	*Count*	*MB*	*LKF*
ACEI	1234	0	0.0077	0.0073
ACIE	1243	1	0.0061	0.0057
AECI	1324	1	0.0054	0.0046
AEIC	1423	0	0.0028	0.0020
AICE	1342	0	0.0038	0.0029
AIEC	1432	0	0.0025	0.0017
CAEI	2134	0	0.0342	0.0369
CAIE	2143	1	0.0278	0.0288
CEAI	3124	6	0.0843	0.0870
CEIA	4123	12	0.2059	0.1962
CIAE	3142	3	0.0535	0.0557
CIEA	4132	11	0.1625	0.1607
EACI	2314	0	0.0123	0.0133
EAIC	2413	1	0.0067	0.0059
ECAI	3214	6	0.0445	0.0495
ECIA	4213	1	0.1130	0.1117
EIAC	3412	1	0.0114	0.0114
EICA	4312	4	0.0558	0.0579
IACE	2341	0	0.0069	0.0066
IAEC	2431	0	0.0046	0.0038
ICAE	3241	1	0.0224	0.0248
ICEA	4231	6	0.0713	0.0715
IEAC	3421	1	0.0090	0.0089
IECA	4321	1	0.0449	0.0452

5.3.2 Winter sports holiday preferences

Typically, conjoint analysis involve more than one attribute. Therefore, we have collected data concerning winter sports holidays, where each winter sports holiday is described by the multiple attributes; that is, country, period of holiday, duration of holiday and size of the ski area. As each attribute can take a limited number of values, each holiday may be viewed as a specific combination of levels of the attributes. Each of $n = 169$ respondents is asked to indicate the most preferred alternative and the least preferred alternative, yielding Best-Worst ranking data.

As before, we omit the intercept to avoid indeterminacy and use dummy variables to obtain the plan matrix $\mathbf{X}$. To show that our approach indeed allows for a unified approach,

Figure 5.1: Estimated probability according to the proposed technique versus estimated probability according to the Thurstone-Mosteller-Daniels model.

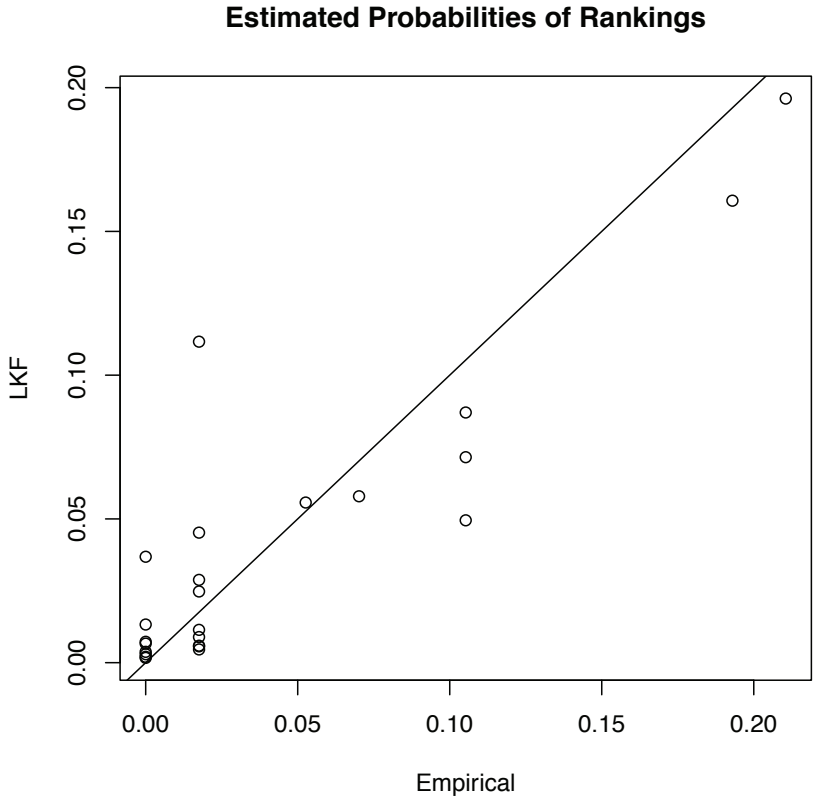


we will present result for (our approximate versions of) the Thurstone-Mosteller-Daniels model as well as the Luce model.

Results Thurstone-Mosteller-Daniels model

Table 5.4 shows the estimated coefficients. The estimated coefficients for Austria and Italy differ significantly from zero and are positive, and thus France is preferred over Austria and Italy as winter sports holiday location. The estimated coefficient for Andorra is not significant, hence there is insufficient statistical evidence to conclude that there exists a difference in preference between Andorra and France. The estimated coefficient for

Figure 5.2: Estimated probability according to the proposed technique versus empirical probability.



February is not significant, thus there is insufficient statistical evidence to conclude that the holiday period matters. The estimated coefficient for "ten-days" differs significantly from zero and is positive, and thus eight-days holidays are preferred over ten-days holidays. The estimated coefficients for "average" and "large" (as the size of the ski area) differ significantly from zero and are negative, and thus average and large ski areas are preferred over small ski areas. Moreover, the results suggest large ski areas are preferred over average ski areas, yielding the general conclusion that the larger the size, the more preferred is the ski area.

Table 5.4: Estimated coefficients in the Thurstone-Mosteller-Daniels model. Reference levels are marked by an asterisk.

Attribute	*Level*	*Coefficient*	*S.E.*	*p-value*
Country	Austria	1.619	0.188	0.000
	Italy	0.918	0.239	0.000
	Andorra	0.047	0.237	0.843
	France*	0	-	-
Period	February	-0.087	0.171	0.611
	January*	0	-	-
Duration	ten-days	0.573	0.175	0.001
	eight-days*	0	-	-
Ski Area	average	-0.660	0.172	0.000
	large	-1.941	0.258	0.000
	small*	0	-	-

Table 5.5: Testing the effects of attributes in the Thurstone-Mosteller-Daniels model by means of likelihood ratio tests.

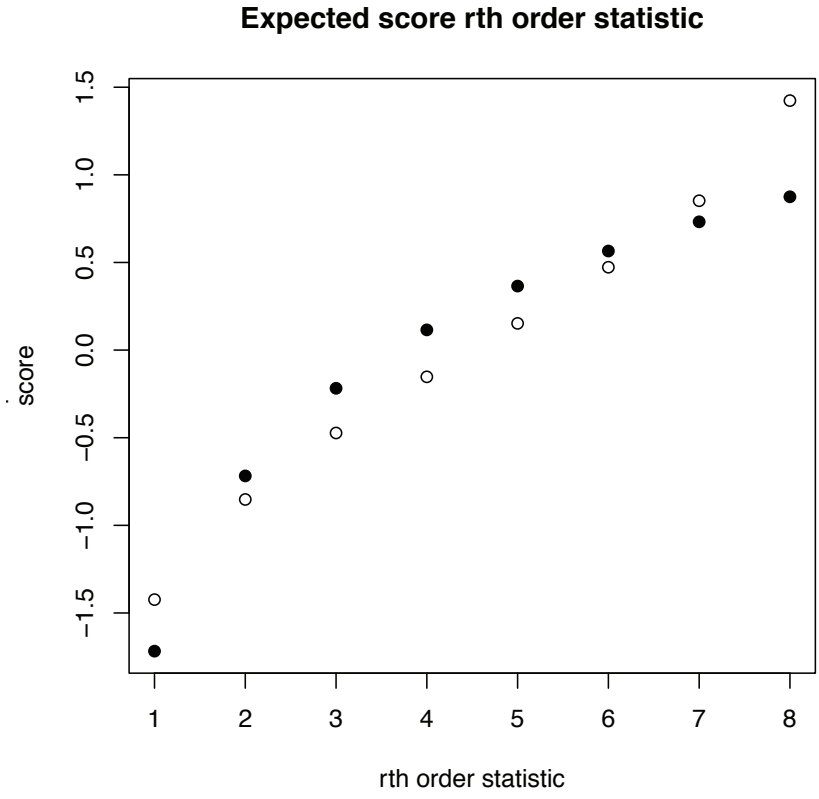
Attribute	*Chi-square*	*df*	*p-value*
Country	114.259	3	0.000
Period	0.258	1	0.611
Duration	11.083	1	0.001
Ski Area	65.402	2	0.000

Alternatively, we can examine the effect of each attribute by means of likelihood ratio tests. The results are given in Table 5.5. One can observe in this table that the attribute *Period* has no significant effect on the preference of the respondents and may be omitted from the model. On the other hand, the attributes *Country*, *Duration*, and *Ski Area* do have a significant effect on preference and should remain in the model.

Results Luce model

The Thurstone-Mosteller- Daniels model puts the same emphasis on the lower ranks as on the higher ranks. In the Luce model more emphasis is put on the lower ranks, that is, on higher preferences. This is visualized in Figure 5.3 by plotting the expected scores for both models. One can observe the symmetry around zero for expected scores belonging to

Figure 5.3: Expected score of r^{th} order statistic drawn from a sample of size 8. The open dots are Thurstone-Mosteller-Daniels scores (5.14) and the solid dots are Luce scores (5.15).



the Thurstone-Mosteller-Daniels model. In contrast, expected scores for highly preferred stimuli receive more emphasis in the Luce model.

Table 5.6 shows the estimated coefficients in the Luce model. The estimates are only slightly different from those in Table 5.4 and the conclusions remain the same. The effect of each attribute are reported in Table 5.7. Again, we arrive at the same conclusion. The attribute *Period* may be omitted from the model, but the other attributes should remain in the model.

Table 5.6: Estimated coefficients in the Luce model. Reference levels are marked by an asterisk.

Attribute	*Level*	*Coefficient*	*S.E.*	*p-value*
Country	Austria	1.403	0.188	0.000
	Italy	0.767	0.330	0.020
	Andorra	-0.152	0.269	0.571
	France*	0	-	-
Period	February	-0.181	0.219	0.408
	January*	0	-	-
Duration	ten-days	0.646	0.199	0.001
	eight-days*	0	-	-
Ski Area	average	-0.855	0.228	0.000
	large	-1.871	0.311	0.000
	small*	0	-	-

Table 5.7: Testing the effects of attributes in the Luce model by means of likelihood ratio tests.

Attribute	*Chi-square*	*df*	*p-value*
Country	104.645	3	0.000
Period	0.691	1	0.708
Duration	11.885	1	0.003
Ski Area	45.003	2	0.000

5.4 Discussion

Preferences may be measured in a simple way by means of rankings. It has been well known that task difficulty increases substantially with the number of stimuli to be ranked. Partial rankings reduce task complexity for respondents. Efficient partial ranking methods, such as Best-Worst ranking, have become popular in practical conjoint analysis. This requires new methods to analyze these partial rankings data. The technique proposed in this paper is not limited to Best-Worst ranking data, as it is in fact able to handle any partial ranking data.

In the context of conjoint experiments, we have shown in this paper how any Thurstone order statistics model could be approximated by extending the technique proposed by Henery (1981) for the Thurstone-Mosteller-Daniels model. Computing approximate rather than exact ranking probabilities leads to a considerable reduction in complexity. A further advantage is that the approximate ranking models allow a unified approach with respect to estimation and testing.

In future research, the approximate ranking model will be further developed. The current model does not take respondents' heterogeneity into account, which makes it less suitable for marketing applications. Using finite mixture models is a popular approach of accommodating respondents' heterogeneity. A finite mixture version of the model could be instrumental in identifying "market segments", that is, groups of respondents who appreciate the attributes in a homogeneous way.

Moreover, there are limits to the flexibility of the current model, as it assumes that the expected scores are fully specified. We may go beyond these limits by introducing unknown additional parameters which influence the shape of the density g. As a consequence, the expected scores will depend on the unknown shape parameters as well and therefore the estimation procedure described in this paper will no longer suffice. The extended model would allow tests with respect to the shape of the density g, and hence it may be used to answer fundamental questions such as "do the random utilities $Y_{c_1}, Y_{c_2}, \ldots, Y_{c_C}$ follow a normal distribution, as is assumed in the Thurstone-Mosteller-Daniels model?" or "do the random utilities $Y_{c_1}, Y_{c_2}, \ldots, Y_{c_C}$ follow a Gumbel distribution, as is assumed in the Luce model?". The answers to these questions may provide further insight in the way respondents build preferences.

5.A Derivation of the approximate probability (5.10)

Define

$$\phi(y;\eta) = \frac{\partial \ln g(y;\eta)}{\partial \eta} = \frac{1}{g(y;\eta)} \frac{\partial g(y;\eta)}{\partial \eta}, \tag{5.21}$$

and remark that

$$\frac{\partial g(y;\eta)}{\partial \eta} = \phi(y;\eta) g(y;\eta).$$

Let $g_0(y)$ and $\phi_0(y)$ denote $g(y;\eta_0)$ and $\phi(y;\eta_0)$, respectively. Recall that $\phi_0(y)$ is introduced in (5.7) as the score function.

As

$$\begin{aligned}
\frac{\partial p_\pi}{\partial \eta_{c_r}} &= \int_{-\infty}^{\infty} \int_{y_{c_1}}^{\infty} \cdots \int_{y_{c_{C-1}}}^{\infty} g\left(y_{c_1};\eta_{c_1}\right) g\left(y_{c_2};\eta_{c_2}\right) \\
&\quad \cdots g\left(y_{c_{r-1}};\eta_{c_{r-1}}\right) \frac{\partial g(y_{c_r};\eta_{c_r})}{\partial \eta_{c_r}} g\left(y_{c_{r+1}};\eta_{c_{r+1}}\right) \\
&\quad \cdots g\left(y_{c_C};\eta_{c_C}\right) dy_{c_C} \cdots dy_{c_2} dy_{c_1} \\
&= \int_{-\infty}^{\infty} \int_{y_{c_1}}^{\infty} \cdots \int_{y_{c_{C-1}}}^{\infty} g\left(y_{c_1};\eta_{c_1}\right) g\left(y_{c_2};\eta_{c_2}\right) \\
&\quad \cdots g\left(y_{c_{r-1}};\eta_{c_{r-1}}\right) \phi\left(y_{c_r};\eta_{c_r}\right) g\left(y_{c_r};\eta_{c_r}\right) g\left(y_{c_{r+1}};\eta_{c_{r+1}}\right) \\
&\quad \cdots g\left(y_{c_C};\eta_{c_C}\right) dy_{c_C} \cdots dy_{c_2} dy_{c_1},
\end{aligned} \tag{5.22}$$

we obtain

$$\begin{aligned}
\left.\frac{\partial p_\pi}{\partial \eta_{c_r}}\right|_{\eta_1=\eta_2=\ldots=\eta_C=\eta_0} &= \int_{-\infty}^{\infty} \int_{y_{1:C}}^{\infty} \cdots \int_{y_{C-1:C}}^{\infty} \phi_0\left(y_{r:C}\right) \\
&\quad g_0\left(y_{1:C}\right) g_0\left(y_{2:C}\right) \cdots g_0\left(y_{C:C}\right) dy_{C:C} \cdots dy_{2:C} dy_{1:C}.
\end{aligned} \tag{5.23}$$

Let $Y_{1:C} < Y_{2:C} < \cdots < Y_{C:C}$ be the order statistics of a random sample of size C from a density g_0. Recall that the joint density of $Y_{1:C}, Y_{2:C}, \cdots, Y_{C:C}$ equals

$$C! g_0\left(y_{1:C}\right) g_0\left(y_{2:C}\right) \cdots g_0\left(y_{C:C}\right)$$

for $y_{1:C} < y_{2:C} < \cdots < y_{C:C}$. It follows that

$$\left.\frac{\partial p_\pi}{\partial \eta_{c_r}}\right|_{\eta_1=\eta_2=\ldots=\eta_C=\eta_0} = \frac{q_{r:C}}{C!} = p_* q_{r:C}, \quad \text{with} \quad q_{r:C} = \mathcal{E}\phi_0\left(Y_{r:C}\right), \tag{5.24}$$

and p_* given by (5.11). Since $\sum_{r=1}^{C} \phi_0(Y_{r:C})$ coincides with $\sum_{r=1}^{C} \phi_0(Y_r)$, (5.8) implies

$$\sum_{r=1}^{C} q_{r:C} = \sum_{r=1}^{C} \mathcal{E}\phi_0(Y_{r:C}) = \mathcal{E}\sum_{r=1}^{C} \phi_0(Y_{r:C}) = \mathcal{E}\sum_{r=1}^{C} \phi_0(Y_r) = \sum_{r=1}^{C} \mathcal{E}\phi_0(Y_r) = 0. \quad (5.25)$$

When all η_c's are close to η_0, a first order Taylor expansion in $(\eta_0, \eta_0, \cdots, \eta_0)$ yields

$$p_\pi \approx p_* + p_* \sum_{r=1}^{C} q_{r:C}(\eta_{c_r} - \eta_0) = p_* \left(1 + \sum_{r=1}^{C} q_{r:C}\eta_{c_r}\right) \quad (5.26)$$

for $\pi = (c_1, c_2, \ldots, c_C)$. The equality follows from (5.25).

5.B Derivation of the first and second order derivatives of (5.19)

Standard iterative methods for finding an estimator $\hat{\boldsymbol{\beta}}$ maximizing the log-likelihood (5.19) require the first order derivatives of (5.19) with respect to $\boldsymbol{\beta}$ and possibly the second order derivatives as well. Write p_ϖ as $\sum_{\pi\in\mathcal{S}_\varpi} s_\pi / \sum_{\pi'} s_{\pi'}$ with $s_\pi = \exp\left\{\mathbf{q}_\pi^\top \mathbf{X}\boldsymbol{\beta}\right\}$. As $(\partial/\partial\boldsymbol{\beta})s_\pi = s_\pi \mathbf{X}^\top \mathbf{q}_\pi$, it follows that

$$\begin{aligned}
\frac{\partial p_\varpi}{\partial \boldsymbol{\beta}} &= \frac{\partial}{\partial \boldsymbol{\beta}} \frac{\sum_{\pi\in\mathcal{S}_\varpi} s_\pi}{\sum_{\pi'} s_{\pi'}} \\
&= \frac{\sum_{\pi\in\mathcal{S}_\varpi} s_\pi \mathbf{X}^\top \mathbf{q}_\pi}{\sum_{\pi'} s_{\pi'}} - \frac{\sum_{\pi\in\mathcal{S}_\varpi} s_\pi}{\sum_{\pi'} s_{\pi'}} \cdot \frac{\sum_{\pi} s_\pi \mathbf{X}^\top \mathbf{q}_\pi}{\sum_{\pi'} s_{\pi'}} \\
&= \mathbf{X}^\top \left(\sum_{\pi\in\mathcal{S}_\varpi} p_\pi \mathbf{q}_\pi - p_\varpi \sum_{\pi'} p_{\pi'} \mathbf{q}_{\pi'} \right), \qquad (5.27)
\end{aligned}$$

and in particular,

$$\frac{\partial p_\pi}{\partial \boldsymbol{\beta}} = \mathbf{X}^\top p_\pi \left(\mathbf{q}_\pi - \sum_{\pi'} p_{\pi'} \mathbf{q}_{\pi'} \right). \qquad (5.28)$$

For any set $\mathcal{S}$ of rankings, (5.28) yields

$$\begin{aligned}
\frac{\partial}{\partial \boldsymbol{\beta}} \frac{\sum_{\pi\in\mathcal{S}} p_\pi \mathbf{q}_\pi^\top}{\sum_{\pi\in\mathcal{S}} p_\pi} &= \frac{\sum_{\pi\in\mathcal{S}} \frac{\partial p_\pi}{\partial \boldsymbol{\beta}} \mathbf{q}_\pi^\top}{\sum_{\pi\in\mathcal{S}} p_\pi} - \frac{\sum_{\pi\in\mathcal{S}} \frac{\partial p_\pi}{\partial \boldsymbol{\beta}}}{\sum_{\pi\in\mathcal{S}} p_\pi} \cdot \frac{\sum_{\pi\in\mathcal{S}} p_\pi \mathbf{q}_\pi^\top}{\sum_{\pi\in\mathcal{S}} p_\pi} \\
&= \frac{\sum_{\pi\in\mathcal{S}} \mathbf{X}^\top p_\pi \left(\mathbf{q}_\pi - \sum_{\pi'} p_{\pi'} \mathbf{q}_{\pi'}\right) \mathbf{q}_\pi^\top}{\sum_{\pi\in\mathcal{S}} p_\pi} \\
&\quad - \frac{\sum_{\pi\in\mathcal{S}} \mathbf{X}^\top p_\pi \left(\mathbf{q}_\pi - \sum_{\pi'} p_{\pi'} \mathbf{q}_{\pi'}\right)}{\sum_{\pi\in\mathcal{S}} p_\pi} \cdot \frac{\sum_{\pi\in\mathcal{S}} p_\pi \mathbf{q}_\pi^\top}{\sum_{\pi\in\mathcal{S}} p_\pi} \\
&= \mathbf{X}^\top \left(\frac{\sum_{\pi\in\mathcal{S}} p_\pi \mathbf{q}_\pi \mathbf{q}_\pi^\top}{\sum_{\pi\in\mathcal{S}} p_\pi} - \frac{\sum_{\pi\in\mathcal{S}} p_\pi \mathbf{q}_\pi}{\sum_{\pi\in\mathcal{S}} p_\pi} \cdot \frac{\sum_{\pi\in\mathcal{S}} p_\pi \mathbf{q}_\pi^\top}{\sum_{\pi\in\mathcal{S}} p_\pi} \right), \qquad (5.29)
\end{aligned}$$

and in particular,

$$\frac{\partial}{\partial \boldsymbol{\beta}} \sum_\pi p_\pi \mathbf{q}_\pi^\top = \mathbf{X}^\top \left\{ \sum_\pi p_\pi \mathbf{q}_\pi \mathbf{q}_\pi^\top - \left(\sum_\pi p_\pi \mathbf{q}_\pi \right) \left(\sum_\pi p_\pi \mathbf{q}_\pi^\top \right) \right\}, \qquad (5.30)$$

It now follows from (5.27) that

$$\begin{aligned}\frac{\partial \ln L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} &= \sum_{i=1}^{n} \frac{\partial}{\partial \boldsymbol{\beta}} \ln p_{\varpi_i} = \sum_{i=1}^{n} \frac{\frac{\partial}{\partial \boldsymbol{\beta}} p_{\varpi_i}}{p_{\varpi_i}} \\ &= \sum_{i=1}^{n} \mathbf{X}^\top \left(\frac{\sum_{\pi \in \mathcal{S}_{\varpi_i}} p_\pi \mathbf{q}_\pi}{\sum_{\pi \in \mathcal{S}_{\varpi_i}} p_\pi} - \sum_{\pi} p_\pi \mathbf{q}_\pi \right) \\ &= \mathbf{X}^\top \left(\sum_{i=1}^{n} \frac{\sum_{\pi \in \mathcal{S}_{\varpi_i}} p_\pi \mathbf{q}_\pi}{\sum_{\pi \in \mathcal{S}_{\varpi_i}} p_\pi} - n \sum_{\pi} p_\pi \mathbf{q}_\pi \right). \end{aligned} \tag{5.31}$$

Similarly, it follows from (5.29) and (5.30) that

$$\begin{aligned}\frac{\partial^2 \ln L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^2} &= \frac{\partial}{\partial \boldsymbol{\beta}} \left(\sum_{i=1}^{n} \frac{\sum_{\pi \in \mathcal{S}_{\varpi_i}} p_\pi \mathbf{q}_\pi^\top}{\sum_{\pi \in \mathcal{S}_{\varpi_i}} p_\pi} - n \sum_{\pi} p_\pi \mathbf{q}_\pi^\top \right) \mathbf{X} \\ &= \left(\sum_{i=1}^{n} \frac{\partial}{\partial \boldsymbol{\beta}} \frac{\sum_{\pi \in \mathcal{S}_{\varpi_i}} p_\pi \mathbf{q}_\pi^\top}{\sum_{\pi \in \mathcal{S}_{\varpi_i}} p_\pi} - n \frac{\partial}{\partial \boldsymbol{\beta}} \sum_{\pi} p_\pi \mathbf{q}_\pi^\top \right) \mathbf{X} \\ &= \mathbf{X}^\top \left(\sum_{i=1}^{n} \left\{ \frac{\sum_{\pi \in \mathcal{S}_{\varpi_i}} p_\pi \mathbf{q}_\pi \mathbf{q}_\pi^\top}{\sum_{\pi \in \mathcal{S}_{\varpi_i}} p_\pi} - \frac{\sum_{\pi \in \mathcal{S}_{\varpi_i}} p_\pi \mathbf{q}_\pi}{\sum_{\pi \in \mathcal{S}_{\varpi_i}} p_\pi} \cdot \frac{\sum_{\pi \in \mathcal{S}_{\varpi_i}} p_\pi \mathbf{q}_\pi^\top}{\sum_{\pi \in \mathcal{S}_{\varpi_i}} p_\pi} \right\} \right. \\ &\quad \left. -n \left\{ \sum_{\pi} p_\pi \mathbf{q}_\pi \mathbf{q}_\pi^\top - \left(\sum_{\pi} p_\pi \mathbf{q}_\pi \right) \left(\sum_{\pi} p_\pi \mathbf{q}_\pi^\top \right) \right\} \right) \mathbf{X}. \end{aligned} \tag{5.32}$$

In particular, in case of full rankings we have that the log-likelihood (5.19) simplifies to (5.20). Consequently, (5.31) becomes

$$\frac{\partial \ln L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \mathbf{X}^\top \left(\sum_{i=1}^{n} \mathbf{q}_{\pi_i}^\top - n \sum_{\pi} p_\pi \mathbf{q}_\pi^\top \right),$$

and (5.32) becomes

$$\frac{\partial^2 \ln L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^2} = \mathbf{X}^\top \left(-n \left\{ \sum_{\pi} p_\pi \mathbf{q}_\pi \mathbf{q}_\pi^\top - \left(\sum_{\pi} p_\pi \mathbf{q}_\pi \right) \left(\sum_{\pi} p_\pi \mathbf{q}_\pi^\top \right) \right\} \right) \mathbf{X}.$$

5.C Derivation of the Luce expected score (5.15)

The Luce model is the Thurstone model obtained by setting $g(y, \eta)$ equal to

$$\exp(-(y-\eta))\exp(-\exp(-(y-\eta))), \quad -\infty < y < \infty, \tag{5.33}$$

the Gumbel density with location η and scale 1. The score function $\phi_0(y)$, see (5.7), becomes

$$1 - \exp(-(y)).$$

One may show that the quantile function $G^{-1}(u)$ (that is, the inverse of the cumulative distribution function $\exp(-\exp(-(y-\eta)))$ belonging to (5.33) is given by $\eta - \ln(-\ln(u))$ for $0 < u < 1$. As the r^{th} order statistic $Y_{r:C}$ has the same probability distribution (and thus the same expectation) as $G^{-1}(U_{r:C})$, where $U_{r:C}$ is the r^{th} order statistic from a random sample drawn from the uniform distribution on the unit interval $(0, 1)$. Since the density of $U_{r:C}$ is

$$\frac{C!}{(r-1)!(C-r)!} u^{r-1}(1-u)^{C-r}, \quad 0 < u < 1,$$

and

$$\phi_0\left(G^{-1}(u)\right) = 1 + \ln u, \quad 0 < u < 1,$$

it follows that the expected score $q_{r:C}$, see (5.9), may be expressed as

$$\begin{aligned}
&\int_0^1 (1+\ln u) \cdot \frac{C!}{(r-1)!(C-r)!} u^{r-1}(1-u)^{C-r}\, du \\
&= 1 + \int_0^1 (\ln u) \cdot \frac{C!}{(r-1)!(C-r)!} u^{r-1} \sum_{i=0}^{C-r} \frac{(C-r)!}{i!(C-r-i)!}(-u)^i du \\
&= 1 + F(r) - F(C+1),
\end{aligned}$$

where the first identity follows from the fact that densities integrate to 1 over their support and the second identity may be verified with standard symbolic algebra computer software.

Chapter 6

Finite Mixture Ranking Models

Abstract

We present a finite mixture ranking model for analyzing full and partial ranking data in the presence of respondent heterogeneity. Such data typically appear in marketing and other social sciences. We illustrate our methodology with two real data sets. The first considers banking preferences in small and medium enterprises and the second focuses on the valuation of health states.

6.1 Introduction

In marketing it is generally assumed that consumers are heterogeneous in their choice and preference behavior. To adequately model this heterogeneity, finite mixture models have become popular, see for example Dillon and Kumar (1994); DeSarbo et al. (1996) and Wedel and Kamakura (2000). There is a large variety of situations for which finite mixture models have been developed. In this paper, we assume that observations are drawn from a population consisting of several homogeneous subpopulations. However, it is usually not known to which subpopulation a particular observation belongs. This is called Type 1 sampling in Redner and Walker (1984) and is perhaps the most natural situation where the finite mixture model is of interest. Alternatively, we may represent Type 1 sampling using a discrete latent variable. Each level of this discrete latent variable then corresponds to a subpopulation, and vice versa. In Type 1 sampling, the data are not available for each segment distribution separately, but only for the overall mixture distribution. The interest is usually in estimating the parameters of a finite mixture model which are the mixing proportions and the parameters of the segment distributions. Here, mixing proportions refer to the unknown relative frequencies of the subpopulations and segment distributions refer to the distributions within a subpopulation. For a review of

finite mixture models, see for example Everitt and Hand (1981); Titterington, Smith, and Makov (1985); Lindsay (1993) and McLachlan and Peel (2000).

Rankings are widely used to express preferences of respondents for different attributes within some choice set. In marketing for example, Best-Worst ranking in conjoint analysis developed by Louviere and Woodworth (1991a) and Finn and Louviere (1992), is gaining in popularity (see for example Chrzan and Golovashkina (2006) for an empirical application).

Probability models for rankings are not easy to analyze, as the computation of each ranking probability usually requires high-dimensional integration when the number of attributes is large. Therefore, for a long time, the use of probability models for rankings has been avoided in the statistical analysis of conjoint experiments.

Recently there is a renewed interest in the modeling and estimation of ranking models, see Marley and Louviere (2005); Maydeu-Olivares and Böckenholt (2005) and Böckenholt (2006). In Lam, Koning, and Franses (2010b) the complexity of probabilistic ranking models is considerably reduced by using approximate rather than exact ranking probabilities. Moreover, their approach also allows for the analysis of partial rankings, which are an essential part of practical conjoint analysis in collecting data efficiently as a means of relieving the respondents' task burden. For example in Best-Worst ranking, respondents are instructed to select only the best and the worst choice set. Partial rankings may also be dictated by the experimental design of the conjoint study (see for example Vermeulen, Goos, & Vandebroek, 2010).

In this paper we increase the usefulness of ranking models in conjoint experiments by incorporating respondents' heterogeneity via Type 1 sampling. We will assume that a heterogeneous population of respondents may be partitioned into a small number of "unobserved" homogeneous subpopulations. Within each subpopulation, a common ranking model holds while the ranking model parameters are allowed to vary across the subpopulations.

There are at least two advantages of our approach. First, our statistical models allow for the efficient use of rankings/partial rankings to collect preference data. Second, as they incorporate respondent heterogeneity, our models are able to find segments in a heterogeneous context, and hence facilitate the development of better targeted marketing strategies.

The structure of the paper is as follows. In Section 6.2 we discuss finite mixture ranking models and how to model the dependence of rankings on stimulus characteristics. Moreover, we explain how we can take respondents' heterogeneity into account in these

models. In Section 6.3 we illustrate our methodology with real data sets. One data set considers banking preferences in small and medium size enterprises and the second focuses on the valuation of health conditions. Finally, in Section 6.4 we conclude with suggestions for further research. Most technical issues are relegated to the Appendix A.

6.2 Mixture ranking models

In this section we will explicitly specify the mixture ranking model and implement the EM algorithm. We start with a brief summary of approximate ranking models as described in Lam et al. (2010b).

6.2.1 Approximate ranking models

We consider a conjoint experiment where each respondent lists all alternative choices, $1, 2, \ldots, C$, in order of preference, with the most preferred choice listed first. For each alternative c in $\{1, 2, \ldots, C\}$, we define the rank $\pi_{(c)}$ of c as the position of c within this ordering. For example, $\pi_{(3)} = 7$ indicates that choice alternative 3 is listed in the 7^{th} place in order of preference. We shall refer to $\pi = \left(\pi_{(1)}, \pi_{(2)}, \ldots, \pi_{(C)}\right)$ as a full ranking. A ranking, which contains ties or missing values should be considered as a partial ordering of the choices rather than a full ordering, and we shall refer to it as a partial ranking and denote it as ϖ.

We assume that for each respondent the probability of actually obtaining π as a full ranking depends on a C-dimensional linear predictor vector $\boldsymbol{\eta} = (\eta_1, \eta_2, \ldots, \eta_C)^\top$. A ranking model specifies the exact nature of the dependence. We have to specify the construction of the predictor vectors. Assume that the choices are adequately described by means of M attributes. Each choice alternative may be viewed as a specific combination of attribute levels. Let x_{cm} denote the value that attribute m takes for alternative c. We assume that $\boldsymbol{\eta} = (\eta_1, \eta_2, \ldots, \eta_C)^\top$ is given by

$$\eta_c = \beta_1 x_{c1} + \beta_2 x_{c2} + \ldots + \beta_M x_{cM} = \sum_{m=1}^{M} \beta_m x_{cm}, \tag{6.1}$$

where $\beta_1, \beta_2, \ldots, \beta_M$ are unknown coefficients. We may write $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$, where $\boldsymbol{\beta}$ is the M dimensional coefficient vector $(\beta_1, \beta_2, \ldots, \beta_M)^\top$ and $\mathbf{X}$ is the $C \times M$ matrix which contains the value x_{cm} in its (c, m) location. $\mathbf{X}$ is usually referred to as the plan matrix of the conjoint experiment.

In Lam et al. (2010b) it is shown that the probability of actually obtaining the full ranking π may be approximated by

$$f(\pi \mid \boldsymbol{\eta}) \approx \frac{\exp\left\{\mathbf{q}^\top(\pi)\mathbf{X}\boldsymbol{\beta}\right\}}{\sum_{\pi'} \exp\left\{\mathbf{q}^\top(\pi')\mathbf{X}\boldsymbol{\beta}\right\}}, \tag{6.2}$$

where $\mathbf{q}(\pi)$ is the expected score vector corresponding to the full ranking π. This C-dimensional vector contains the $\pi^{th}_{(c)}$ expected score as c^{th} element and is central in deriving an approximation to the probability (6.2). The probabilities on the right hand side of (6.2) are all positive and add up to one, and thus define a probability model with respect to the rankings. Observe that for each partial ranking ϖ there exists a set $\mathcal{S}_\varpi$ of all full rankings which do not contradict the partial ordering implied by ϖ. Thus, we may assign the probability

$$f(\varpi \mid \boldsymbol{\eta}) = \sum_{\pi \in \mathcal{S}_\varpi} f(\pi \mid \boldsymbol{\eta}) \approx \frac{\sum_{\pi \in \mathcal{S}_\varpi} \exp\left\{\mathbf{q}^\top(\pi)\mathbf{X}\boldsymbol{\beta}\right\}}{\sum_{\pi'} \exp\left\{\mathbf{q}^\top(\pi')\mathbf{X}\boldsymbol{\beta}\right\}} \tag{6.3}$$

to the partial ranking ϖ.

6.2.2 Mixture models

Recall from the Introduction that we may represent Type I sampling using a latent discrete variable. Denote this variable by Z and assume its levels are $1, \ldots, J$ with corresponding probabilities $\alpha_1, \ldots, \alpha_J$. The distribution of Z is called the mixing distribution and the probabilities $\alpha_1, \ldots, \alpha_J$ are called mixture proportions. The conditional distribution of the rankings depends on a parameter vector, which in turn depends on the level taken by Z. Denote the value of the parameter vector corresponding to level j by $\boldsymbol{\beta}_j$ for $j = 1, \ldots, J$. Although we cannot observe the discrete latent variable directly, we are able to estimate the mixing proportions $\alpha_1, \ldots, \alpha_J$ as they occur as parameters in the distribution of the observed data.

We shall assume, conditional on the level of the discrete latent variable, that the distribution of the rankings belongs to a parametric family of distributions, described by a common density as in (6.3).

If we assume that the respondents are heterogeneous in their preference behavior, the specific density function underlying the observed preference rankings of individuals in each segment is given by the approximate ranking model (6.3). It follows that the distribution

of the *observed* ranking/partial ranking is described by the density

$$f(\varpi;\boldsymbol{\theta}) = \sum_{j=1}^{J} \alpha_j f_j(\varpi;\boldsymbol{\beta}_j) = \sum_{j=1}^{J} \frac{\sum_{\pi\in\mathcal{S}_\varpi} \alpha_j \exp\left\{\mathbf{q}_j^\top(\pi)\mathbf{X}\boldsymbol{\beta}_j\right\}}{\sum_{\pi'} \exp\left\{\mathbf{q}_j^\top(\pi')\mathbf{X}\boldsymbol{\beta}_j\right\}}. \tag{6.4}$$

where $\boldsymbol{\theta}$ denotes the unknown parameter vector $(\alpha_1,\ldots,\alpha_J,\boldsymbol{\beta}_1^T,\ldots,\boldsymbol{\beta}_J^T)^\top$.

6.2.3 EM algorithm

To estimate $\boldsymbol{\theta}$, we shall make use of maximum likelihood via the EM algorithm, see Dempster, Laird, and Rubin (1977); Redner and Walker (1984). Each iteration of the EM algorithm consists of two steps. In the (hth) E-step, the function $Q\left(\boldsymbol{\theta}';\boldsymbol{\theta}^h\right)$ is calculated by taking the expectation of the complete log-likelihood given the current estimate of $\boldsymbol{\theta}$ and is defined here as

$$\begin{aligned} Q\left(\boldsymbol{\theta}';\boldsymbol{\theta}^h\right) &= \sum_{i=1}^{n}\sum_{j=1}^{J} \ln\left(\alpha_j\right) p(j;\varpi_i;\boldsymbol{\theta}^h) \\ &+ \sum_{i=1}^{n}\sum_{j=1}^{J} \ln\left(f(\varpi_i,\boldsymbol{\beta}_j)\right) p(j;\varpi_i;\boldsymbol{\theta}^h), \end{aligned} \tag{6.5}$$

where

$$\begin{aligned} p(j;\varpi_i,\boldsymbol{\theta}^h) &= \frac{\alpha_j^h f_j(\varpi_i;\boldsymbol{\beta}_j^h)}{f(\varpi_i;\boldsymbol{\theta}^h)} = \frac{\alpha_j^h \sum_{\pi\in\mathcal{S}_{\varpi_i}} f_j(\pi;\boldsymbol{\beta}_j^h)}{\sum_{t=1}^{J}\alpha_t^h f_t(\varpi_i;\boldsymbol{\beta}_t^h)} = \\ &= \frac{\sum_{\pi\in\mathcal{S}_{\varpi_i}} \alpha_j^h \exp\left\{\mathbf{q}_j^\top(\pi)\mathbf{X}\boldsymbol{\beta}_j^h\right\}}{\sum_{\pi'}\exp\left\{\mathbf{q}_j^\top(\pi')\mathbf{X}\boldsymbol{\beta}_j^h\right\}} \Bigg/ \sum_{t=1}^{J} \frac{\sum_{\pi\in\mathcal{S}_{\varpi_i}} \alpha_t^h \exp\left\{\mathbf{q}_t^\top(\pi)\mathbf{X}\boldsymbol{\beta}_t^h\right\}}{\sum_{\pi'}\exp\left\{\mathbf{q}_t^\top(\pi')\mathbf{X}\boldsymbol{\beta}_t^h\right\}}. \end{aligned} \tag{6.6}$$

denotes the posterior probability that respondent i with observed preference ranking ϖ_i belongs to segment j. These posterior probabilities are used to classify the individuals into segments and are very important in the estimation of the parameters for each segment. When more information is available from each respondent i, the segment-level estimates will be more accurate.

In the M-step, $Q\left(\boldsymbol{\theta}';\boldsymbol{\theta}^h\right)$ is maximized. The estimates of α_j^{h+1} and $\boldsymbol{\beta}_j^{h+1}$ are based on the posterior probabilities obtained in the E-step. If we assume that the parameters $\alpha_1\ldots,\alpha_J$ and $\boldsymbol{\beta}_1,\ldots,\boldsymbol{\beta}_J$ are unrelated, then the two terms on the right hand side of

(6.5) may be maximized separately. The contribution of individual i to the likelihood is weighted by the posterior probabilities $p(j; \varpi_i, \boldsymbol{\theta}^h)$.

The estimation of the mixing proportions has a unique solution and can be computed as

$$\alpha_j^{h+1} = \sum_{i=1}^{n} p(j; \varpi_i, \boldsymbol{\theta}^h)/n. \tag{6.7}$$

Thus, the estimates of the prior probabilities are simply the averages of the posterior probabilities in each segment.

If $\boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_J$ are mutually independent parameters, then the maximization problem of the segment parameters separates further into J segment problems, each of which involves only one of the parameters $\boldsymbol{\beta}_j$

$$\boldsymbol{\beta}_j^{h+1} \in \arg\max_{\boldsymbol{\beta}_j} \sum_{i=1}^{n} \ln\left(f(\varpi_i, \boldsymbol{\beta}_j)\right) p(j; \varpi_i; \boldsymbol{\theta}^h), \tag{6.8}$$

for $j = 1, ..., J$. Note that both the segment problems and the maximization problem for the proportions alone have the nice property that they can be regarded as weighted maximum likelihood estimation problems involving sums of logarithms weighted by these posterior probabilities. Hence, the estimates of the segment parameters $\boldsymbol{\beta}_j^{h+1}$ are obtained by maximizing

$$\sum_{i=1}^{n} \left(\ln \sum_{\pi \in \mathcal{S}_{\varpi_i}} \exp\left\{\mathbf{q}_j^\top(\pi)\mathbf{X}\boldsymbol{\beta}_j\right\} - \ln \sum_{\pi'} \exp\left\{\mathbf{q}_j^\top(\pi')\mathbf{X}\boldsymbol{\beta}_j\right\} \right) p(j; \varpi_i, \boldsymbol{\theta}^h). \tag{6.9}$$

Note that $p(j; \varpi_i, \boldsymbol{\theta}^h)$ depends on h and not on $h+1$. The computation of the first and second order derivatives are given in Appendix 6.A. These derivatives are instrumental in finding the parameter estimates and their standard deviations.

6.2.4 Determining the number of segments

The actual number of segments is usually unknown and must be inferred from the data. Determining the number of segments is an important but yet not completely resolved problem. In this paper we choose the approach involving the minimization of model selection criteria to "estimate" the number of segments. Model selection criteria attempt to balance the fit of a model (as indicated by $\ln\ \hat{L} = \ln L\left(\hat{\boldsymbol{\theta}}\right)$) and the complexity of the model (as indicated by the number k of "free" parameters). The primary example is the Akaike information criterion $\text{AIC} = -2\ln\ \hat{L} + 2k$, see Akaike (1973).

AIC is known to favor models with too many parameters. The use of the Bayesian information criterion BIC $= -2\ln\ \hat{L} + k\ln n$, see Schwarz (1978), and the consistent Akaike information criterion CAIC $= -2\ln\ \hat{L} + k\,(1+\log n)$, see Bozdogan (1987), is advocated among practitioners; for instance, see DeSarbo, Wang, and Blanchard (2010). Both BIC and CAIC impose an additional sample size penalty on the likelihood, and are more conservative than the AIC statistic in that they tend to favor more parsimonious models (that is, models with fewer segments).

Other procedures and criteria for the selection of the number of segments and model selection heuristics also exist, see for eaxmple Wedel and Kamakura (2000) and Susko (2003).

6.3 Illustrations

In this section we demonstrate how respondents' heterogeneity can be modeled by our finite mixture ranking model. We will apply our proposed procedure to two data sets. The first illustration is a very elementary conjoint experiment and helps the reader to understand the ranking model and shows the effectiveness of partial rankings. The second illustration is a real conjoint experiment as stimuli are described by combinations of attribute levels.

6.3.1 Banking preference data

In this data set, banking preferences of SMEs (small and medium size enterprizes) in the Netherlands are collected. Respondents are the individuals responsible for daily banking issues in their company. Respondents are asked to indicate their preference for each of the banks in a list of eight banks. The obtained rankings are not limited to full rankings, that is, ties are allowed. A total of 616 respondents has participated in the survey.

The most preferred bank is assigned rank 1. Consequently, the smaller its rank, the more preferred is a bank, and the preference measure is inversely related to preference. Hence, when interpreting the estimation results, a positive coefficient indicates that higher levels lead to a higher ranking, that is lower preference. A negative coefficient indicates that higher levels lead to a lower ranking, that is, more preference. The strength of preference is reflected in the absolute value of the coefficient.

We have estimated mixture ranking models with different numbers of segments. The fit and the values of the various information criteria are reported in Table 6.1. We select the model with two segments, as suggested by both BIC and CAIC. In Table 6.2 the

Table 6.1: Model selection criteria for different numbers of segments. Boldface is the most preferred model.

Segments	log-likelihood	df	AIC	BIC	CAIC
1	4793.610	6	4805.610	4832.149	4838.149
2	**4360.574**	**13**	**4386.574**	**4444.076**	**4457.076**
3	4318.450	20	4358.450	4446.915	4466.915
4	4322.342	27	4376.342	4495.770	4522.770
5	4281.725	34	4349.725	4500.115	4534.115
6	4250.734	41	4332.734	4514.088	4555.088

Table 6.2: Parameter estimates of the two segments model.

	Segment 1	S.E.	p-value	Segment 2	S.E.	p-value
probability	0.690	0.035	0.000	0.310	0.033	0.000
Bank						
ASN	4.820	0.444	0.000	-0.198	0.168	0.239
Fortis	1.921	0.187	0.000	0.610	0.188	0.001
ING	-0.339	0.127	0.008	-0.350	0.167	0.036
Rabobank	-0.926	0.130	0.000	-0.488	0.167	0.003
SNS	1.711	0.177	0.000	0.401	0.178	0.024
Triodos	5.731	0.526	0.000	-0.496	0.180	0.006

estimated coefficients are reported. We have set ABN AMRO bank as the reference level. A positive coefficient value does not necessarily mean that the respective bank is rejected, but that it is less preferable than the reference bank ABN AMRO bank. The segment probability for the first segment is about seventy percent. Note that the segment probabilities estimates are both significant. Moreover, all parameter estimates differ significantly from the reference bank. In this first segment the banks Rabobank and ING are more preferred than the others. Furthermore, we see that the banks ASN and Triodos are clearly less preferred.

In the second segment, we can again conclude that Rabobank and ING are the most preferred banks. However, note that in this segment the only other two banks which have also negative coefficients are ASN and Triodos bank (though the coefficient estimate for ASN is not significant). Hence, in this segment we conclude the opposite, namely that idealistic banks are more preferred than regular banks.

Table 6.3: Multiple comparison test shows that there is no difference between ING, Rabobank and ABN-Amro over the two segments.

Bank	Estimate	S.E.	p-value
ASN	-5.019	0.440	0.000
Fortis	-1.311	0.288	0.000
ING	-0.011	0.221	1.000
Rabobank	0.438	0.223	0.228
SNS	-1.311	0.271	0.000
Triodos	-6.227	0.475	0.000

Now that we have revealed two different segments, it is interesting to examine whether the coefficients (i.e. preference behavior) differs across the segments. The Chi-square test statistic yields the value 204.199 (df= 6 and p-value zero) and thus attribute Bank is clearly significant. Multiple comparisons (see Table 6.3) shows that this is indeed caused by the banks ASN, Triodos, SNS and Fortis with respect to ABN AMRO bank. That is, there is no difference between ING, Rabobank and ABN AMRO over the two components, at least according to the CFOs of 616 SMEs.

6.3.2 Health conditions data

For many subjective health conditions, Thurstone scaling and its derivatives may be an attractive methodology to arrive at quantitative measures to do statistical testing to be used in for example health research, economic evaluations, and disease modeling studies. The data used here originate from Krabbe (2008) and are from a Dutch EuroQol-5D (EQ-5D) valuation study conducted during the summer of 2003. The EQ-5D classification describes health status according to five attributes: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Each attribute has 3 levels i.e., 1-"no problems", 2-"some problems", 3-"severe problems". Health condition descriptions are constructed by taking 1 level for each attribute (for example, 11111 represents the best health condition). In our analysis, we selected the eleven most "informative" health conditions. The respondents ranked the health conditions from best to worst where they were free to place multiple health conditions in the same position (that is, we allow for ties).

Again, we vary the number of segments. The fit and the values of the corresponding information criteria are reported in Table 6.4. As suggested by both BIC and CAIC, we select the model with two segments. In Table 6.5 the estimated coefficients are reported

Table 6.4: Model selection criteria for different numbers of components. Boldface is the most preferred model.

Segments	log-likelihood	df	AIC	BIC	CAIC
1	5005.305	10	5025.305	5058.871	5068.871
2	**4931.273**	**21**	**4973.273**	**5043.761**	**5064.761**

Table 6.5: Parameter estimates of the two components model.

	Segment 1	S.E.	p-value	Segment 2	S.E.	p-value
probability	0.672	0.045	0.000	0.328	0.019	0.000
Health condition						
2-mobility	-0.041	0.217	0.852	-0.006	0.866	0.994
3-mobility	2.735	0.215	0.000	9.863	0.077	0.000
2-selfcare	0.976	0.225	0.000	0.636	0.833	0.445
3-selfcare	1.896	0.221	0.000	3.331	0.860	0.000
2-activities	-0.817	0.256	0.001	-1.399	0.931	0.133
3-activities	-0.244	0.307	0.426	-0.524	1.287	0.684
2-pain	0.556	0.157	0.000	1.865	0.602	0.002
3-pain	1.432	0.209	0.000	2.262	0.687	0.001
2-anxiety	1.226	0.209	0.000	2.764	0.757	0.000
3-anxiety	1.016	0.149	0.000	5.406	0.697	0.000

with regard to the reference level, which we have set at the first level ("no problems") for each attribute. The mixing proportions are respectively 67.2% and 32.8% for the two segments. The first notable finding is that the sign for the estimated coefficients is the same for both segments. This means that the respondents in the distinct segments do not differ in their valuation of the attributes of the health conditions. However, the second finding is that the estimated coefficients in the second segment are larger in the second segment or that the coefficient is not significant. Note that the strength of preference is reflected in the absolute value of the coefficient. Hence, there is a stronger discomfort for each attribute in the second segment. Note especially the insignificance of the second level of the attributes mobility and self-care in this segment. We can conclude that respondents in the two subpopulations do not differ in their valuation of health condition attributes but they do differ in the significance of the level of these attributes.

Once we have distinguished two segments, we again examine whether the attributes differ across the segments. The results of the Chi-square tests and corresponding multiple

Table 6.6: Test on the homogeneity of the attributes. The Chi-square tests show that only the attributes mobility and anxiety have a significant heterogeneous effect. That is, the effects differ across the segments.

Attribute	Chi-square	df	p-value
mobility	1091.285	2	0.000
selfcare	6.160	2	0.046
activity	1.139	2	0.566
pain	4.507	2	0.105
anxiety	207.124	2	0.000

Table 6.7: Multiple comparison to test the heterogeneity effect of the attribute mobility.

Attribute	Estimate	S.E.	p-value
mobility2	-0.035	0.961	0.999
mobility3	-7.128	0.235	0.000

comparisons tests are reported in Table 6.6 to Table 6.9. Attribute mobility is clearly significant over the segments. In particular, the coefficients for the level "severe problems" of this attribute differs significantly across the segments (see Table 6.7). Attribute self-care is slightly significant, however the multiple comparisons show that the coefficients do not differ significantly across the segments. As the Chi-square test statistic for attributes activities and pain is not significant, there is no sufficient evidence that respondents are heterogenous in their appreciation of these attributes. The next question that follows on this finding is whether the attributes activities and pain have effect. The Chi-square tests show significance for both attributes. In particular the coefficient for the first level of attribute activity in the first segment is significant (2-activities=-0.817, s.e.=0.256, p-value=0.004). In addition, all coefficients for the attribute pain are significant, see Table 6.9. Finally, the Chi-square test shows that attribute anxiety is clearly significant and multiple comparisons show that both level coefficients differ significantly across the segments (see Table 6.8).

From the above findings, we may conclude that, although the respondents in the distinct segments do not differ in their valuation of the attributes of the health conditions, there is a stronger discomfort in the second segment which is caused by the attributes mobility and anxiety.

Table 6.8: Multiple comparison to test the heterogeneity effect of the attribute anxiety.

Attribute	Estimate	S.E.	p-value
anxiety2	-1.539	0.817	0.089
anxiety3	-4.390	0.727	0.000

Table 6.9: Multiple comparison to test the heterogeneity effect of the attribute pain.

Attribute	Estimate	S.E.	p-value
pain21	0.556	0.157	0.000
pain31	1.432	0.209	0.000
pain22	1.865	0.602	0.000
pain32	2.262	0.687	0.000

6.4 Conclusion

Conjoint analysis continues to attract active interest as a research field. Recent developments have reduced complexity when analyzing rank data in conjoint experiments. Rankings are an easy and efficient way to collect preference information. In particular, partial rankings are essential in practical conjoint analysis to collect data efficiently to relieve respondents' task burden. Moreover, one does not need complete rankings to obtain extra preference information.

Respondents may be heterogeneous and therefore we proposed a finite mixture ranking probability model which takes respondents' heterogeneity into account when modeling preferences.

We demonstrated our methodology with two empirical data sets and we have shown that our model is able to successfully discover distinct segments. Hence, partial rankings yield sufficient preference information to take respondents' heterogeneity into account. Given that a partial ranking task amounts to a smaller burden for respondents than a complete ranking task, this certainly will help marketers to identify and target consumers by understanding their preference behavior, and, implement a more efficient and optimal marketing strategy.

Possible topics for further research are the optimal design of the plan matrix to reduce the variance of the estimated parameters, and to increase the flexibility of the current model by introducing additional parameters which influence the shape of the density to provide further insight in the way respondents build preferences.

6.A Derivation of the first and second order derivatives of (6.9)

Note that $p(j;\varpi_i,\boldsymbol{\theta}^h)$ depends on h and not on $h+1$. Hence, the first order derivative with respect to $\boldsymbol{\beta}_j$ yields

$$\mathbf{X}^\top\left(\sum_{i=1}^{n}\frac{\sum_{\pi\in\mathcal{S}_{\varpi_i}}f_j(\pi;\boldsymbol{\beta}_j)\mathbf{q}_j(\pi)}{\sum_{\pi\in\mathcal{S}_{\varpi_i}}f_j(\pi;\boldsymbol{\beta}_j)}p(j;\varpi_i,\boldsymbol{\theta}^h)-\sum_{\pi'}f_j(\pi';\boldsymbol{\beta}_j)\mathbf{q}_j(\pi')\sum_{i=1}^{n}p(j;\varpi_i,\boldsymbol{\theta}^h)\right), \tag{6.10}$$

and the second order derivative with respect to $\boldsymbol{\beta}_j$ yields

$$\begin{aligned}&\mathbf{X}^\top\left(\sum_{i=1}^{n}p(j;\pi_i,\boldsymbol{\theta}^h)\times\right.\\&\left\{\frac{\sum_{\pi\in\mathcal{S}_{\varpi_i}}f_j(\pi;\boldsymbol{\beta}_j)\mathbf{q}_j(\pi)\mathbf{q}_j^\top(\pi)}{\sum_{\pi\in\mathcal{S}_{\varpi_i}}f_j(\pi;\boldsymbol{\beta}_j)}-\frac{\sum_{\pi\in\mathcal{S}_{\varpi_i}}f_j(\pi;\boldsymbol{\beta}_j)\mathbf{q}_j(\pi)}{\sum_{\pi\in\mathcal{S}_{\varpi_i}}f_j(\pi;\boldsymbol{\beta}_j)}\cdot\frac{\sum_{\pi\in\mathcal{S}_{\varpi_i}}f_j(\pi;\boldsymbol{\beta}_j)\mathbf{q}_j^\top(\pi)}{\sum_{\pi\in\mathcal{S}_{\varpi_i}}f_j(\pi;\boldsymbol{\beta}_j)}\right\}\\&-\sum_{i=1}^{n}p(j;\pi_i,\boldsymbol{\theta}^h)\times\\&\left.\left\{\sum_{\pi}f_j(\pi;\boldsymbol{\beta}_j)\mathbf{q}_j(\pi)\mathbf{q}_j^\top(\pi)-\left(\sum_{\pi}f_j(\pi;\boldsymbol{\beta}_j)\mathbf{q}_j(\pi)\right)\left(\sum_{\pi}f_j(\pi;\boldsymbol{\beta}_j)\mathbf{q}_j^\top(\pi)\right)\right\}\right)\mathbf{X},\end{aligned} \tag{6.11}$$

see also Lam et al. (2010b).

In case of full rankings, the first order derivative with respect to $\boldsymbol{\beta}_j$ simplifies to

$$\mathbf{X}^\top\left(\sum_{i=1}^{n}\mathbf{q}_j(\pi_i)p(j;\pi_i,\boldsymbol{\theta}^h)-\sum_{\pi'}f_j(\pi';\boldsymbol{\beta}_j)\mathbf{q}_j(\pi')\sum_{i=1}^{n}p(j;\pi_i,\boldsymbol{\theta}^h)\right), \tag{6.12}$$

and the second order derivative with respect to $\boldsymbol{\beta}_j$ yields

$$\begin{aligned}&-\sum_{i=1}^{n}p(j;\pi_i,\boldsymbol{\theta}^h)\times\\&\mathbf{X}^\top\left(\sum_{\pi'}f_j(\pi';\boldsymbol{\beta}_j)\mathbf{q}_j(\pi')\mathbf{q}_j^\top(\pi')-\left(\sum_{\pi'}f_j(\pi';\boldsymbol{\beta}_j)\mathbf{q}_j(\pi')\right)\left(\sum_{\pi'}f_j(\pi';\boldsymbol{\beta}_j)\mathbf{q}_j^\top(\pi')\right)\right)\mathbf{X}.\end{aligned} \tag{6.13}$$

Chapter 7

Summary and Discussion

In this chapter, we first give a summary of the main findings in the individual chapters followed by a discussion of the limitations and some recommendations for further research.

7.1 Summary of main findings

In Chapter 2 and Chapter 3 we derived new confidence intervals for maximal reliability measures. Simulation experiments showed that existing confidence intervals for maximal reliability can be anti-conservative. Anti-conservative confidence intervals have an attained coverage less than the required coverage and hence may falsely give the impression that maximal reliability is estimated with sufficient accuracy. We proposed to apply a variance stabilizing transformation technique when constructing confidence intervals. We introduced coverage curve analysis as a new methodology to compare confidence intervals, which has shown to be more informative. That is, we have not only considered the coverage of the true maximal reliability, but also the coverage of hypothetical values which differ from the true maximal reliability. Our new derived stable confidence interval performs well, and has a coverage for the true value which is approximately equal to the confidence level. We stress the use of confidence intervals accompanying single measures that summarize the parameters to assess the adequacy of the measure. The results may help to improve questionnaires by yielding more precise measurements.

In part two of this thesis we focused on the analysis of rankings data. As respondents face less difficulties when completing a partial ranking, they may be more motivated to complete the task and as such the quality of the obtained data may improve. In Chapter 4, we derived a nonparametric approach to analyze partial rankings. Moreover, we showed that our proposed nonparametric test and the score test are equivalent, which implies

that the far more simpler nonparametric approach is as good as the likelihood approach. In addition, the nonparametric approach also allows for a better interpretation of the multiple comparisons results.

In Chapter 5 we demonstrated how partial rankings data can be used in conjoint analysis. More specifically, we estimated probabilistic ranking models in the context of conjoint experiments. We showed that general ranking models can be applied to conjoint experiments by introducing a linear model which allows for modeling the dependence of the rankings on the stimulus characteristics. This method allows for a more efficient way to collect data to reduce respondents' task complexity which is essential in practical conjoint analysis to improve data quality.

As in marketing it is generally assumed that individuals are heterogeneous in their choice and preference behavior, we presented in Chapter 6 a finite mixture variant of the ranking model in Chapter 5 to adequately model this heterogeneity. We showed that our model is able to extract sufficient preference information from partial rankings data to take into account respondents' heterogeneity. This certainly will help marketers to identify and target consumers by understanding their preference behavior, and, implement a more optimal marketing strategy in an efficient way.

7.2 Limitations and further research

Below we will discuss some limitations and issues that may be improved in further research.

Variance stabilization transformation technique is a well-established tool. For a statistic which variance depending on the unknown parameters, the variance of the transformed statistic does not depend on the unknown parameters if a variance stabilization transformation exists. Although the pivotal quantity presented in Chapter 2 and Chapter 3 is asymptotically stable when the number of items tends to infinity, the variance still depends on the unknown parameter. For most psychological measurements, where only a few or best items are available or used, this assumption is hard to meet. When it is not a real variance stabilization transformation, the empirical results obtained may depend on the specific conditions chosen. For example, when the items are parallel, the transformed statistic is close to having a constant variance. Actually, when measurements are parallel, there does exist a variance stabilization transformation and also an exact confidence interval using the F-distribution (see Koning & Franses, 2003). We do not

expect an exact variance stabilization transformation to exist when items are not parallel or the data are not normally distributed.

In the second part of the thesis, we focused on Thurstone order statistics models, which were introduced in Chapter 5. Thurstone order statistics models are specializations of random utility models, which assume that the random utilities $Y_1, Y_2, \ldots, Y_C$ are independent variables. Some interesting random utilities models do not make this assumption, and unfortunately fall outside our framework but may be incorporated in future research. For instance, one may assume that the random utilities follow a multivariate normal distribution with a covariance matrix which is not diagonal. Indeed, this multivariate normal model is already present in Case I and Case II in Thurstone (1927). One may consider general dependence by allowing any covariance matrix, or one may restrict the dependence by assuming that the covariance matrix should exhibit a certain structure. That is, the distribution of the random utilities $Y_1, Y_2, \ldots, Y_C$ should obey a covariance structure model, see Browne (1982). In Maydeu-Olivares and Böckenholt (2005), the distribution of the random utilities $Y_1, Y_2, \ldots, Y_C$ is given by a factor analysis model. Factor analysis models are special covariance structure models. The multivariate normal distribution is quite restrictive with respect to dependence, as only linear dependence is allowed. In McFadden (1977, 1978), see also McFadden (2001, p. 358), a class of random utility models is proposed in which the marginal distributions of the random utilities are Gumbel distributions, and which are able to induce the well-known nested multinomial logit choice model. The nested multinomial logit choice model is often used to avoid the restrictive independence of irrelevant alternatives assumption (IIA) which is part of the standard multinomial logit choice model. As the standard logit choice model derives from a Thurstone order statistics model with independent Gumbel random utilities, it follows that allowing dependence may be quite useful.

We also note that almost all current choice models and limited dependent variable models (including structural equation models) assume that error variances are constant, although a few allow for non-constant error variances for alternatives, as in the heterogeneous error multinomial logit models. More generally, however, there can be a distribution of error variances across the sample, and failure to capture this distribution can lead to seriously biased and incorrect models. This limitation should be a subject for future research.

Considerably more information is obtained in conjoint choice experiments by collecting partial rankings rather than single best choices, or best and worst choices. Future research interest lies in the design of the conjoint experiment to select the profiles to be included in

the experiment. To create optimal designs for these experiments is a difficult task as the total number of profiles depends on the number of attributes and levels and this increases exponentially. Vermeulen et al. (2010) showed how the optimal experimental design theory can be used to set up experiments for best-worst choices, and they compared the resulting designs with various alternatives from the literature in terms of the optimality criterion and prediction accuracy. In the context of partial rankings, the question is then raised how much additional information can be obtained in a conjoint choice experiment from an additional choice in each choice set.

During this thesis we encountered some numerical issues. The computation time increases exponentially with the number of stimuli in the conjoint experiments as we need to compute the sum over all possible rankings of these stimuli. Although we believe that in the near future this is not a limitation anymore as computers becoming faster, for now it does limit the possible number of stimuli that can be included in the conjoint experiment. In this thesis we have computed all sums over each individual. However, one can consider to jointly compute these sums at the same time. A possible solution would be to perform parallel computations, in particular a cloud computing environment may assist to overcome computation time limits.

Another issue is whether the optimization procedures we have applied in our estimation methods, especially in the EM algorithm, are the most suitable methods. It is well-known that the convergence of the EM algorithm might be slow in some cases. In addition, the maximization in the M-step might be misleading as it is not guaranteed that it will converge to a global maximum. However, as the EM algorithm is widely used, improvements on this technique may also be incorporated when modeling rankings data.

Nederlandse Samenvatting (Summary in Dutch)

Inleiding en motivatie

Vragenlijsten zijn een belangrijke manier om informatie te verzamelen over een populatie voor zowel kwalitatief als kwantitatief onderzoek. De meerwaarde van een goede vragenlijst alsmede de kwaliteit van de verkregen data kan niet genoeg benadrukt worden. Dit proefschrift bespreekt bepaalde aspecten van de statistische analyse van data verkregen aan de hand van vragenlijsten.

De kwaliteit van de data hangt af van de nauwkeurigheid van de metingen. Een vragenlijst bestaat uit meerdere items, waarbij groepen van items onafhankelijk hetzelfde onderliggende concept kunnen meten. Het concept zelf kan echter niet direct gemeten worden. Een concept zou bijvoorbeeld klanttevredenheid kunnen zijn en de vragenlijst bevat dan meerdere stellingen die elk onafhankelijk (een aspect van) de klanttevredenheid meten. De uitkomst van de vragenlijst, in het jargon "test score", wordt als benadering gehanteerd voor de "echte" meetwaarde van het concept, met als het gevolg dat we de test score kunnen zie als de optelsom van een "echte" score en een willekeurige meetfout (Lord & Novick, 1968). In het voorbeeld hierboven wordt de uitkomst van de vragenlijst gezien als een optelsom van de "echte" klanttevredenheid plus een meetfout. De mate van consistentie tussen de gemeten test score en de echte score is een indicator van de betrouwbaarheid van de vragenlijst. Echter, de echte score is niet observeerbaar en dus is het niet mogelijk om de betrouwbaarheid van een vragenlijst direct te schatten. Aan de hand van de afzonderlijke item scores is het echter wel mogelijk om de betrouwbaarheid indirect te evalueren.

De meest gebruikte maatstaaf om de betrouwbaarheid van een vragenlijst uit te drukken is Cronbach's coefficiënt alpha (Cronbach, 1951). In specifieke gevallen is coefficiënt alpha gelijk aan de betrouwbaarheid, maar feitelijk is het een ondergrens van

de betrouwbaarheid (zie Novick & Lewis, 1967) en kan deze gemaximaliseerd worden, (zie bijvoorbeeld Li et al., 1996). Coefficiënt alpha is een ongewogen som van de individuele item scores en door een gewogen som te gebruiken, kan de betrouwbaarheidscoefficiënt verbeterd worden (zie bijvoorbeeld Knott & Bartholomew, 1993; Li, 1997; Yuan & Bentler, 2002; Raykov & Penev, 2006). De optimaal gewogen test score wordt de maximale betrouwbaarheid van de test genoemd (Lindsey, 1996).

Essentieel is de evaluatie van de betrouwbaarheidscoefficiënt, oftewel, welke waarden zijn hoog en welke zijn laag. De evaluatie van een verkregen alpha is veelal gebaseerd op de aanbevelingen van Nunnally (Nunnally & Bernstein, 1994) met betrekking tot de minimaal accepteerbare betrouwbaarheid. In veel marketing onderzoeken, worden vuistregels gebruikt als "hoger dan 0.8" is goed. Deze methodiek is uiteraard verre van zorgvuldig. We kunnen immers geen uitspraken doen over de nauwkeurigheid van de meting en over de significante verschillen tussen betrouwbaarheidscoefficiënten onderling. Cortina (1993) wijst op het ontbreken van een metriek om de nauwkeurigheid van de toetsingsgrootheid te beoordelen. Het berekenen van betrouwbaarheidsintervallen voor de toetsingsgrootheid verhelpt dit probleem. De extra informatie in het interval maakt een meer kritische evaluatie van de toetsingsgrootheid mogelijk. We benadrukken dus het belang van betrouwbaarheidsintervallen voor statistische grootheden die de parameters van het model samenvatten.

Het eerste deel van dit proefschrift richt zich op de betrouwbaarheid van vragenlijsten. We bespreken de maximale betrouwbaarheidscoefficiënt in verschillende modellen en we construeren bijbehorende betrouwbaarheidsintervallen, zodat de nauwkeurigheid van de grootheid getoetst kan worden. De bevindingen kunnen leiden tot een betere opzet van vragenlijsten en deze leiden weer tot nauwkeurigere uitkomsten van vragenlijsten.

In het tweede gedeelte van dit proefschrift ontwikkelen we methoden om statistische analyses uit te voeren op data gebaseerd op rankings (rangordes). Rankings worden vaak gebruikt in de marketing om bijvoorbeeld producten, merken, en services te evalueren. Andere voorbeelden zijn rankings van wetenschappelijke tijdschriften, universiteiten en onderzoeksscholen. Rankings zijn makkelijk te begrijpen en eenvoudig te verzamelen. Bovendien zijn rankings meer betrouwbaar dan metrische metingen, omdat respondenten vaak niet in staat zijn om de mate van voorkeur nauwkeurig uit te drukken die deze metingen vereisen (Ben-Akiva et al., 1992). Respondenten zijn goed in staat om alternatief A boven alternatief B te prefereren, maar zij zijn zelden in staat om uit te drukken hoeveel precies zij A prefereren boven B. Echter, de taak om alternatieven te rangschikken naar voorkeur wordt al snel ingewikkeld naarmate het aantal alternatieven toeneemt. Het is

welbekend dat respondenten het moeilijk vinden, of zelfs teleurgesteld worden, indien zij te veel keuzemogelijkheden moeten vergelijken. Iyengar en Lepper (2000) laten zien dat te veel keuzes demotiverend kunnen zijn voor consumenten. Boatwright en Nunes (2001) demonstreren dat een reductie van het assortiment juist als prettig wordt ervaren door consumenten, een bevinding die wordt onderschreven door onder andere Chernev (2003) en Gourville en Soman (2005). DeShazo en Fermo (2002) laten zien dat consumenten taak complexiteit ervaren als keuzemogelijkheden in overvloed zijn, zie ook Sandor en Franses (2009). Kortom, consumenten vinden het moeilijk om voorkeuren te rangschikken als het aantal alternatieven groot is. Het probleem van taak-complexiteit kan verlicht worden door respondenten te vragen om slechts een deelverzameling van alle alternatieven te rangschikken naar voorkeur. We zullen deze rangschikking een partiële ranking noemen.

Partiële rankings zijn essentieel in praktische conjunct analyse. Conjunct meten is gebaseerd op de veronderstelling dat consumenten een product zien als een bundel van productkenmerken die ook wel attributen worden genoemd. Productkenmerken zijn bijvoorbeeld merk en prijs. Aan respondenten worden omschrijvingen van een bepaald product voorgelegd. Deze omschrijvingen worden profielen genoemd en bevatten informatie met betrekking tot de productkenmerken. De niveaus van de attributen worden systematisch gevarieerd, zodat de profielen verschillende varianten van het product representeren. Aan de respondent wordt vervolgens gevraagd om alle profielen naar voorkeur te rangschikken. Profielen bestaan vaak uit veel attributen. Dit leidt al snel tot een groot aantal mogelijke profielen. Het rangschikken van te veel profielen is, zoals eerder genoemd, een complexe taak. Door het gebruik van partiële rankings in de conjunct analyse, wordt de data efficiënter verzameld en de inspanning van de respondent verlicht. Bovendien geven volledige rankings niet per se extra informatie over preferenties, zie Louviere et al. (2000). Stel dat aan een consument gevraagd is om verschillende producten te vergelijken en het product met het hoogste nut is het meest gewaardeerd. Aan elk attribuut wordt nut toegekend en elk attribuut levert afzonderlijk een bijdrage aan het totale nut van het product. Hieruit kan worden afgeleid hoe verschillende attributen tegen elkaar worden afgewogen en hoe belangrijk deze attributen zijn. Nut is echter latent in de zin dat de consument niet in staat is om deze getalsmatig uit te drukken, maar desalniettemin kan het zijn/haar gedrag beïnvloeden.

Er zijn minstens twee voordelen van onze benadering om partiële rankings te analyseren. Ten eerste, onze statistische modellen maken het mogelijk om op een efficiënte manier preferentie data te verzamelen door het gebruik van partiële rankings. Partiële rankings kunnen ook opgelegd zijn door het experimentele ontwerp van de conjunct studie.

Ten tweede, wordt een partiële ranking als minder belastend ervaren door respondenten en zullen zij meer gemotiveerd zijn, waardoor de kwaliteit van de verkregen data beter is. Dit helpt marketeers om consumenten te identificeren door het begrijpen van hun preferentie gedrag, en een efficiëntere en optimale marketing strategie te implementeren.

Samenvatting

In hoofdstuk 2 en hoofdstuk 3, bespreken we de maximale betrouwbaarheidscoefficiënt in respectievelijk het tau-equivalente model en het congenerieke model, en we leiden de bijbehorende asymptotische verdeling af om inzicht te krijgen in de verbetering van de betrouwbaarheidsintervallen. De motivatie voor deze modellen is het blootleggen van de structuur van de onderliggende covariantie matrix van de items. Hiermee overbruggen we het gat tussen enerzijds het erg restrictieve parallelle model en anderzijds het extreem tolerante model, zonder enige veronderstellingen voor de variantie structuur. Het negeren van de structuur in de covariantie matrix leidt tot de verwaarlozing van waardevolle informatie. Zonder enige veronderstellingen kan er niets gezegd worden over de "true score" variantie en "error" variantie van de items, omdat deze parameters niet geschat worden. Deze unieke varianties voor elke item component zijn echter bijzonder interessant om meetschalen te verbeteren en bevatten waardevolle informatie. Men moet immers altijd een model prefereren dat de meeste informatie verschaft.

Aan de hand van simulatie experimenten tonen we aan dat bestaande betrouwbaarheidsintervallen voor de maximale betrouwbaarheidscoefficiënt anti-conservatief zijn en dus ten onrechte de indruk wekken dat met voldoende nauwkeurigheid geschat is. We vergelijken onze nieuwe betrouwbaardheidsintervallen met bestaande intervallen aan de hand van "coverage curve" analyse. Hieruit blijkt dat variantie stabilisatie technieken ervoor zorgen dat betrouwbaarheidsintervallen minder gevoelig zijn voor Type II fouten. Deze intervallen hebben een coverage voor de "echte" waarde die bij benadering gelijk is aan het betrouwbaardheidsniveau. We raden het gebruik van variantie stabilisatie technieken dan ook aan bij meer gecompliceerde modellen om betrouwbaarheidsintervallen te verbeteren. Hoofdstuk 2 is gebaseerd op Lam et al. (2009).

In het tweede gedeelte van dit proefschrift bespreken we de analyse van rankings en in het bijzonder van partiële rankings. In hoofdstuk 4, presenteren we een methodiek om te toetsen of de verkregen rankings significant verschillen van een willekeurige rangschikking. We introduceren meervoudige vergelijking procedures om verder te toesten welke rankings dan significant van elkaar verschillen. We presenteren een niet-parametrische benadering

voor het analyseren van partiële rankings. Bovendien laten we zien dat onze niet-parametrische test en de score test equivalent zijn. Dit impliceert dat de veel simpelere niet-parametrische benadering even goed is als de "likelihood" benadering. Een bijkomend voordeel van de niet-parametrische benadering is dat de resultaten van een meervoudige vergelijking beter interpreteerbaar zijn. De methodiek vermindert de hoeveelheid aan keuzes in een verkennende analyse op een makkelijke en begrijpbare manier, zodat irrelevante alternatieven verwijderd kunnen worden uit verdere analyses. Dit hoofdstuk is gedeeltelijk gebaseerd op Lam et al. (2010a).

In hoofdstuk 5 presenteren we een rankingsmodel in de context van conjunct analyse. Aangezien er slechts een eindig aantal mogelijke rangschikkingen zijn, hebben de rankings een discrete verdeling. In principe kunnen standaard methoden voor het analyseren van discrete data hier toegepast worden, zie Marden (1995, p. 140). Echter, kansmodellen voor rankings worden al snel complex naarmate het aantal te vergelijken objecten groter wordt. In de literatuur is er een overvloed aan rankingsmodellen beschikbaar, zie Critchlow et al. (1991, Section 3) en ook Marden (1995, Chapter 5), maar we concentreren ons op Thurstone "order statistics" modellen. We laten zien dat elk Thurstone order statistics model benaderd kan worden door een simpeler model en dit maakt een algemene benadering mogelijk. We analyseren partiële rankings in de context van conjunct experimenten door een lineair model te introduceren waarin de rankings afhankelijk zijn van de attributen.

In hoofdstuk 6 presenteren we een "finite mixture" model voor het modelleren van heterogeniteit. In marketing en andere sociale wetenschappen wordt in het algemeen aangenomen dat individuen heterogeen zijn in hun keuzes en voorkeuren. Om dit te modelleren worden heterogene responsparameters toegestaan. In de context van conjunct analyse, kunnen consumenten bijvoorbeeld verschillen in prijsgevoeligheid en dus kunnen zij anders reageren op prijsveranderingen. De waarde van de responsparameter voor het attribuut prijs zal dan variëren tussen consumenten. Het model dat we presenteren is in staat om voldoende informatie met betrekking tot preferenties te verkrijgen uit partiële rankings om deze heterogeniteit nauwkeurig te modelleren.

Conclusies en discussie

Hieronder beschrijven we de voornaamste conclusies, aandachtspunten en suggesties voor vervolgonderzoek.

De variantie van een toetsingsgrootheid hangt na een variantie stabilisatie transformatie niet meer af van de onbekende parameters. Hoewel de spilgrootheid gepresenteerd

in hoofdstuk 2 en hoofdstuk 3 asymptotisch stabiel is als het aantal items naar oneindig gaat, hangt de variantie nog steeds af van de onbekende parameters. Voor de meeste psychologische metingen, waar slechts een paar of slechts de beste items beschikbaar zijn of gebruikt worden, gaat deze veronderstelling niet op. Als het geen echte variantie stabilisatie transformatie is, hangen de verkregen empirische resultaten af van de specifieke condities. Als de items bijvoorbeeld parallel zijn, heeft de getransformeerde grootheid is de variantie vrijwel constant. Om precies te zijn, als metingen parallel zijn, bestaan er reeds variantie stabilisatie transformaties en ook een exact betrouwbaarheidsinterval gebaseerd op de F-verdeling (zie Koning & Franses, 2003). We verwachten niet dat er een exacte variantie stabilisatie transformatie bestaat indien items niet parallel zijn of als de data niet normaal verdeeld zijn.

In hoofdstuk 5 hebben we Thurstone order statistics modellen geïntroduceerd. Thurstone order statistics modellen zijn verfijningen van random utility modellen. Deze modellen veronderstellen dat de random utilities $Y_1, Y_2, \ldots, Y_C$ onafhankelijke variabelen zijn. Bepaalde "random utilities" modellen maken deze veronderstelling niet, maar vallen helaas buiten het kader van dit proefschrift. Het is zeker interessant om dit op te nemen in vervolgonderzoek. Veronderstel bijvoorbeeld dat de random utilities een multivariate normale verdeling volgen met een covariantie matrix die niet diagonaal is. Inderdaad, dit multivariate normale model dient zich reeds aan in Case I en Case II in Thurstone (1927). Door het toestaan van iedere covariantie matrix wordt algemene afhankelijkheid verondersteld en door de veronderstelling dat de covariantie matrix een zekere structuur moet vertonen wordt de afhankelijkheid beperkt. Met andere woorden, de verdeling van de random utilities $Y_1, Y_2, \ldots, Y_C$ zijn onderworpen aan een covariantie structuur model, zie Browne (1982). In Maydeu-Olivares en Böckenholt (2005), is de verdeling van de random utilities $Y_1, Y_2, \ldots, Y_C$ gegeven door een factor analyse model. Factor analyse modellen zijn speciale covariantie structuur modellen. De multivariate normale verdeling is nogal restrictief wat betreft afhankelijkheid, aangezien slechts lineaire afhankelijkheid is toegestaan. In McFadden (1977, 1978), zie ook McFadden (2001, p. 358), wordt een klasse van random utility modellen gepresenteerd waarin de marginale verdelingen van de random utilities Gumbel verdelingen zijn. Het welbekende geneste multinomiale logit model is hiervan afgeleid en wordt vaak gebruikt om de restrictieve veronderstelling van onafhankelijkheid van irrelevant alternatieven (IIA) te vermijden, die wel geldt in het standaard multinomiale logit model. Het standaard logit model is afgeleid van een Thurstone order statistics model met onafhankelijke Gumbel random utilities en het toestaan van afhankelijkheid kan dus wellicht interessant zijn.

We merken ook op dat bijna alle huidige keuze modellen en modellen waar de afhankelijke variabele gelimiteerd is (inclusief structurele vergelijkingen modellen) veronderstellen dat de error varianties constant zijn, hoewel enkele modellen niet-constante error varianties voor de alternatieven toestaan, zoals in de multinomiale logit modellen met heterogene error. Echter, in het algemeen kunnen de error varianties in de steekproef een verdeling aannemen, en het niet meenemen van deze verdeling kan leiden tot incorrecte modellen. Dit onderwerp dient meegenomen te worden in vervolgonderzoek.

Door het gebruik van partiële rankings in conjunct experimenten kan aanzienlijk meer informatie verkregen worden dan door enkel de beste keuze, of de beste en slechtste keuzes, te vragen aan respondenten. Om de profielen dusdanig te selecteren in een conjunct experiment opdat informatie op een efficiënte manier verkregen wordt, is vervolgonderzoek naar de opzet van het experiment nuttig. Een optimale opzet creëren is een lastige opdracht aangezien het totaal aantal profielen afhangt van het aantal attributen en niveaus, en dit exponentieel toeneemt. Vermeulen et al. (2010) vergelijken opzetten van conjunct experimenten waar respondenten hun beste en slechtste keuzes aangeven met alternatieven uit de literatuur in termen van optimaliteitscriterium en voorspelkracht. In de context van partiële rankings is het interessant om te onderzoeken hoeveel extra informatie verkregen kan worden in een conjunct keuze experiment uit een extra keuze in elke keuzeset.

Tot slot noemen we enkele numerieke kwesties waarmee wij geconfronteerd werden. De rekentijd neemt exponentieel toe met het aantal te vergelijken objecten in de conjunct experiment, omdat de som over alle mogelijke rangschikkingen van de objecten berekend moet worden. Het aantal objecten dat meegenomen kan worden in de conjunct analyse is voor nu beperkt, maar we verwachten dat in de nabije toekomst dit geen wezenlijke beperking meer zal zijn omdat de rekensnelheid van computers alsmaar toeneemt. De huidige rekentijd kan verkort worden door berekeningen parallel uit te voeren, en in het bijzonder kan een “cloud computing environment” hierbij helpen.

Het is bekend dat de convergentie van het EM algoritme traag kan zijn. Bovendien kan de maximalisatie in de M-stap misleidend zijn, want er is geen garantie dat het zal convergeren naar een globaal maximum. Het is de bespreking waard of de optimalisatie procedures, die wij hebben toegepast in onze schattingsmethoden, in het bijzonder in het EM algoritme, de meest geschikte methoden zijn. Echter, gezien het gebruik van EM algoritme populair is, zullen verbeteringen van deze techniek wellicht ook meegenomen worden tijdens het modelleren van rankings data.

References

Abramowitz, M., & Stegun, I. A. (1964). *Handbook of mathematical functions with formulas, graphs, and mathematical tables* (Vol. 55). Washington, D.C.: U.S. Government Printing Office.

Aitchison, J., & Silvey, S. D. (1958). Estimation of parameters subject to restraints. *Annals of Mathematical Statistics*, *29*(3), 813–828.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *Second International Symposium on Information Theory (Tsahkadsor, 1971)* (pp. 267–281). Budapest: Akadémiai Kiadó.

Barron, F. H. (1977). Axiomatic conjoint measurement. *Decision Sciences*, *8*(3), 548–559. (1540–5915)

Ben-Akiva, M., Morikawa, T., & Shiroishi, F. (1992). Analysis of the reliability of preference ranking data. *Journal of Business Research*, *24*(2), 149–164.

Boatwright, P., & Nunes, J. (2001). Reducing assortment: An attribute-based approach. *Journal of Marketing*, *65*(3), 50–63.

Böckenholt, U. (2006). Thurstonian-based analyses: Past, present, and future utilities. *Psychometrika*, *71*(4), 615–629.

Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika*, *52*(3), 345–370.

Bradlow, E. T. (2005). Current issues and a 'wish list' for conjoint analysis. *Applied Stochastic Models in Business and Industry*, *21*, 319–323.

Browne, M. W. (1982). Covariance structure analysis. In D. M. Hawkins (Ed.), *Topics in applied multivariate analysis* (pp. 72–141). Cambridge: Cambridge University Press.

Browne, M. W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *The British Journal of Mathematical and Statistical Psychology*, *37*(1), 62–83.

Buse, A. (1982). The likelihood ratio, Wald, and Lagrange multiplier tests: An expository note. *The American Statistician*, *36*(3), 153–157.

Chernev, A. (2003). When more is less and less is more: The role of ideal point availability and assortment in consumer choice. *Journal of Consumer Research*, *30*(2), 170–183.

Chrzan, K., & Golovashkina, N. (2006). An empirical test of six stated importance measures. *International Journal of Market Research*, *48*(6), 717–740.

Cortina, J. M. (1993). What is coefficient alpha? an examination of theory and applications. *Journal of Applied Psychology*, *78*(1), 98–104.

Critchlow, D. E., Fligner, M. A., & Verducci, J. S. (1991). Probability models on rankings. *Journal of Mathematical Psychology*, *35*(3), 294–318.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297–334.

Daniels, H. E. (1950). Rank correlation and population models. *Journal of the Royal Statistical Society B*, *12*(2), 171–191.

Dempster, A. P., Laird, N. M., & Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society B*, *39*(1), 1–38.

DeSarbo, W. S., Ansari, A., Chintagunta, P., Himmelberg, C., Jedidi, K., Johnson, R., et al. (1996). Representing heterogeneity in consumer response models. *Marketing Letters*, *8*(3), 335–348.

DeSarbo, W. S., Wang, Q., & Blanchard, S. J. (2010). Exploring intra-industry competitive heterogeneity: The identification of latent competitive groups. *Journal of Modelling in Management*, *5*(2), 94–123.

DeShazo, J. R., & Fermo, G. (2002). Designing choice sets for stated preference methods: The effects of complexity on choice consistency. *Journal of Environmental Economics and Management*, *44*, 123–143.

Dillon, W. R., & Kumar, A. (1994). Latent structure and other mixture models in marketing: an integrative survey and overview. In R. P. Bagozzi (Ed.), *Advanced methods of marketing research* (pp. 352–388). Cambridge: Blackwell.

Duhachek, A., Coughlan, A. T., & Iacobucci, D. (2005). Results on the standard error of the coefficient alpha index of reliability. *Marketing Science*, *24*(2), 294–301.

Everitt, B. S., & Hand, D. J. (1981). *Finite mixture distributions*. London: Chapman and Hall.

Feldt, L. S. (1965). The approximate sampling distribution of Kuder-Richardson reliability coefficient twenty. *Psychometrika*, *30*, 357–370.

Finn, A., & Louviere, J. (1992). Determining the appropriate response to evidence of public concern: The case of food safety. *Journal of Public Policy and Marketing*, *11*, 12–25.

Flynn, T. N., Louviere, J. J., Peters, T. J., & Coast, J. (2007). Best-worst scaling: What it can do for health care research and how to do it. *Journal of Health Economics*, *26*(1), 171–189.

Gourville, J. T., & Soman, D. (2005). Overchoice and assortment type: When and why variety backfires. *Marketing Science*, *24*, 382–395.

Green, P. E., Krieger, A. M., & Wind, Y. J. (2001). Thirty years of conjoint analysis: Reflections and prospects. *Interfaces*, *31*(3(2)), 56–73.

Green, P. E., & Rao, V. R. (1971). Conjoint measurement for quantifying judgmental data. *Journal of Marketing Research*, *8*(3), 355–363.

Green, P. E., & Srinivasan, V. (1978). Conjoint analysis in consumer research: Issues and outlook. *The Journal of Consumer Research*, *5*(2), 103–123.

Green, S. B., & Herschberger, S. L. (2000). Correlated errors in true score models and their effect on coefficient alpha. *Structural Equation Modeling*, *7*(2), 251–270.

Green, S. B., & Yang, Y. (2009). Commentary on coefficient alpha: A cautionary tale. *Psychometrika*, *74*, 121–135.

Hager, W. W. (1989). Updating the inverse of a matrix. *SIAM Review*, *31*(2), 221–239.

Hájek, J., & Šidák, Z. (1967). *Theory of rank tests*. New York: Academic Press.

Harter, H. L. (1960). Tables of range and studentized range. *The Annals of Mathematical Statistics*, *31*, 1122–1147.

Harter, H. L. (1961). Expected values of normal order statistics. *Biometrika*, *48*(1/2), 151–165.

Hauser, J. R., & Rao, V. R. (2004). Conjoint analysis, related modeling, and applications. In P. E. Green & Y. J. Wind (Eds.), *Marketing research and modeling: progress and prospects: a tribute to paul e. green* (p. 141-168). Dordrecht: Kluwer Academic Publishers.

Hausman, J. A., & Ruud, P. A. (1987). Specifying and testing econometric models for rank-ordered data. *Journal of Econometrics*, *34*, 83–104.

Henery, R. J. (1981). Permutation probabilities as models for horse races. *Journal of the Royal Statistical Society B*, *43*(1), 86–91.

Hoaglin, D. C., Mosteller, F., & Tukey, J. W. (1983). *Understanding robust and exploratory data analysis*. John Wiley & Sons, Inc.

Iacobucci, D., & Duhachek, A. (2003). Advancing alpha: measuring reliability with confidence. *Journal of Consumer Psychology*, *13*(4), 478–487.

Iyengar, S. S., & Lepper, M. R. (2000). When choice is de-motivating: Can one desire too much of a good thing? *Journal of Personality and Social Psychology*, *79*, 995–1006.

Joe, H. (1997). *Multivariate models and dependence concepts* (Vol. 73). London: Chapman & Hall.

Johnson, R. M. (1974). Trade-off analysis of consumer values. *Journal of Marketing Research*, *11*(2), 121–127.

Kano, Y., Berkane, M., & Bentler, P. M. (1990). Covariance structure analysis with heterogeneous kurtosis parameters. *Biometrika*, *77*(3), 575–585.

Knott, M., & Bartholomew, D. J. (1993). Constructing measures with maximum reliability. *Psychometrika*, *58*, 331–338.

Koning, A. J., & Franses, P. H. (2003). *Confidence intervals for Cronbach's coefficient alpha values* (Report Series Research in Management No. ERS-2003-041-MK). ERIM.

Krabbe, P. F. M. (2008). Thurstone scaling as a measurement method to quantify subjective health outcomes. *Medical Care*, *46*(4), 357–365.

Krantz, D. H., & Tversky, A. (1971). Conjoint measurement analysis of composition rules in psychology. *Psychological Review*, *78*(2), 151–169.

Kristof, W. (1963). The statistical theory of stepped-up reliability when a test has been divided into several equivalent parts. *Psychometrika*, *28*, 221–238.

Kruskal, J. B. (1965). Analysis of factorial experiments by estimating monotone transformations of the data. *Journal of the Royal Statistical Society B*, *27*(2), 251–263.

Lam, K. Y., Koning, A. J., & Franses, P. H. (2009). Confidence intervals for maximal reliability in tau-equivalent models. *Statistica Neerlandica*, *63*(4), 490–507.

Lam, K. Y., Koning, A. J., & Franses, P. H. (2010a). Analyzing preference rankings when there are too many alternatives. In A. Fink, B. Lausen, W. Seidel, & A. Ultsch (Eds.), *Advances in data analysis, data handling and business intelligence* (p. 553-562). Springer Berlin Heidelberg.

Lam, K. Y., Koning, A. J., & Franses, P. H. (2010b). *Ranking models in conjoint analysis* (Research Paper No. EI 2010-51). Econometric Institute.

Lattin, J., Carroll, J. D., & Green, P. E. (2003). *Analyzing multivariate data*. Canada: Brooks/Cole.

Lawley, D. N., & Maxwell, A. E. (1971). *Factor analysis as a statistical method*. New York: American Elsevier Publishing Company.

Li, H. (1997). A unifying expression for the maximal reliability of a linear composite. *Psychometrika*, *62*, 245–249.

Li, H., Rosenthal, R., & Rubin, D. (1996). Reliability of measurement in psycholog: From spearman-brown to maximal reliability. *Psychological Methods*, *5*, 98–107.

Lindsay, B. G. (1993). Mixture models: Theory, geometry and applications. In *Regional conference series in probability and statistics* (Vol. 5). Institute of Mathematical Statistics and American Statistical Association.

Lindsey, J. K. (1996). *Parametric statistical inference*. Oxford University Press.

Lord, F. M., & Novick, M. R. (1968). *Statististical theories of mental test scores. with contributions by Allan Birnbaum*. Reading, MA.: Addison-Wesley.

Louviere, J. J., Henser, D. A., & Swait, J. D. (2000). *Stated choice methods*. Cambridge University Press.

Louviere, J. J., & Woodworth, G. G. (1991a). *Best-worst scaling: A model for the largest difference judgments* (Working Paper). University of Alberta.

Louviere, J. J., & Woodworth, G. G. (1991b). *Best-worst scaling: a model for the largest difference judgments* (Working paper). Edmonton: University of Alberta.

Luce, R. D., & Suppes, P. (1965). Preference, utility, and subjective probability. In *Handbook of mathematical psychology* (Vol. III, pp. 235–306). New York: Wiley.

Luce, R. D., & Tukey, J. W. (1964). Simultaneous conjoint measurement: A new type of fundamental measurement. *Journal of Mathematical Psychology*, *1*(1), 1-27.

Marden, J. I. (1995). *Analyzing and modeling rank data* (Vol. 64). London: Chapman & Hall.

Marley, A., & Louviere, J. (2005). Some probabilistic models of best, worst and best-worst choices. *Journal of Mathematical Psychology*, *49*, 464–480.

Maydeu-Olivares, A. (1999). Thurstonian modeling of ranking data via mean and covariance structure analysis. *Psychometrika*, *64*(3), 325-340.

Maydeu-Olivares, A., & Böckenholt, U. (2005, Sep). Structural equation modeling of paired-comparison and ranking data. *Psychological Methods*, *10*(3), 285–304.

Maydeu-Olivares, A., & Hernández, A. (2007). Identification and small sample estimation of Thurstone's unrestricted model for paired comparisons data. *Multivariate Behavioral Research*, *42*(2), 323–347.

McFadden, D. (1977). *Modelling the choice of residential location* (Discussion paper No. 477). New Haven, Connecticut: Yale University. (Cowles foundation for research

in economics)

McFadden, D. (1978). Modelling the choice of residential location. In A. Karlqvist (Ed.), *Spatial interaction theory and planning models* (pp. 75–96). Amsterdam: North-Holland.

McFadden, D. (2001). Economic choices. *The American Economic Review*, *91*(3), 351–378.

McLachlan, G. J., & Peel, D. (2000). *Finite mixture models*. New York: Wiley.

Novick, M. R., & Lewis, C. (1967). Coefficient alpha and the reliability of composite measurements. *Psychometrika*, *32*(1), 1–13.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory, 3rd edition*. New York: McGraw-Hill.

Ofir, C., & Reddy, S. K. (1996). Measurement errors in probability judgments. *Management Science*, *42*(9), 1308–1325.

Plackett, R. L. (1975). The analysis of permutations. *Applied Statistics*, *24*(2), 193–202.

Rao, C. R. (1948). Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. *Proceedings of the Cambridge Philosophical Society*, *44*, 50–57.

Raykov, T., & Penev, S. (2006). A direct method for obtaining approximate standard error and confidence interval of maximal reliability for composites with congeneric measures. *Multivariate Behavioral Research*, *41*(1), 15–28.

Redner, R. A., & Walker, H. F. (1984). Mixture densities, maximum likelihood and the EM algorithm. *SIAM Review*, *26*, 195–239.

Rewtrakunphaiboon, W., & Oppewal, H. (2008). Effects of package holiday information presentation on destination choice. *Journal of Travel Research*, *47*, 127–136.

Sandor, Z., & Franses, P. H. (2009). Consumer price evaluations through choice experiments. *Journal of Applied Econometrics*, *24*(3), 517–535.

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

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, *74*, 107–120.

Susko, E. (2003). Weighted tests of homogeneity for testing the number of components in a mixture. *Computational Statistics and Data Analysis*, *41*, 367–378.

Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review*, *4*, 273–286.

Titterington, D. M., Smith, A. F. M., & Makov, U. E. (1985). *Statistical analysis of finite mixture distributions*. Chichester: Willey Series.

Van Zyl, J. M., Neudecker, H., & Nel, D. G. (2000). On the distribution of the maximum likelihood estimator of Cronbach's alpha. *Psychometrika*, *65*(3), 271–280.

Vermeulen, B., Goos, P., & Vandebroek, M. (2010). Obtaining more information from conjoint experiments by best-worst choices. *Computation Statistics and Data Analysis*, *54*, 1426–1433.

Wallsten, T. S., & Budescu, D. V. (1983). Encoding subjective probabilities: A psychological and psychometric review. *Management Science*, *29*(2), 151–173.

Wedel, M., & Kamakura, W. A. (2000). *Market segmentation: Conceptual and methodological foundations* (Second ed.). Dordrecht: Kluwer Academic Publishers.

Wittkowski, K. M. (1988). Friedman-type statistics and consistent multiple comparisons for unbalanced designs with missing data. *Journal of the American Statistical Association*, *83*, 1163–1170.

Yellott, J. I., Jr. (1977). The relationship between Luce's choice axiom, Thurstone's theory of comparative judgement, and the double exponential distribution. *Journal of Mathematical Psychology*, *15*, 109–144.

Yuan, K., & Bentler, P. M. (2002). On robustness of the normal-theory based asymptotic distributions of three reliability coefficient estimates. *Psychometrika*, *67*(2), 251–259.

Curriculum Vitae

Kar Yin Lam (1983) graduated in Econometrics at Erasmus University Rotterdam in 2006. In the same year, she joined the Erasmus Research Institute of Management to carry out her doctoral research. Her research interests are marketing modeling, psychometrics, and econometrics. She presented her research at various international conferences.

ERASMUS RESEARCH INSTITUTE OF MANAGEMENT (ERIM)

ERIM PH.D. SERIES

RESEARCH IN MANAGEMENT

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

Acciaro, M., *Bundling Strategies in Global Supply Chains,* Promotor: Prof.dr. H.E. Haralambides, EPS-2010-197-LIS, ISBN: 978-90-5892-240-3, http://hdl.handle.net/1765/1

Agatz, N.A.H., *Demand Management in E-Fulfillment*, Promotor: Prof.dr.ir. J.A.E.E. van Nunen, EPS-2009-163-LIS, ISBN: 978-90-5892-200-7, http://hdl.handle.net/1765/15425

Alexiev, A., *Exploratory Innovation: The Role of Organizational and Top Management Team Social Capital,* Promotors: Prof.dr. F.A.J. van den Bosch & Prof.dr. H.W. Volberda, EPS-2010-208-STR, ISBN: 978-90-5892-249-6, http://hdl.handle.net/1765/1

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

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

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

Asperen, E. van, *Essays on Port, Container, and Bulk Chemical Logistics Optimization*, Promotor: Prof.dr.ir. R. Dekker, EPS-2009-181-LIS, ISBN: 978-90-5892-222-9, http://hdl.handle.net/1765/1

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

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

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

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

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

Bezemer, P.J., *Diffusion of Corporate Governance Beliefs: Board Independence and the Emergence of a Shareholder Value Orientation in the Netherlands,* Promotors: Prof.dr.ing. F.A.J. van den Bosch & Prof.dr. H.W. Volberda, EPS-2009-192-STR, ISBN: 978-90-5892-232-8, http://hdl.handle.net/1765/1

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

Binken, J.L.G., *System Markets: Indirect Network Effects in Action, or Inaction?,* Promotor: Prof.dr. S. Stremersch, EPS-2010-213-MKT, ISBN: 978-90-5892-260-1, http://hdl.handle.net/1765/21186

Blitz, D.C., *Benchmarking Benchmarks,* Promotors: Prof.dr. A.G.Z. Kemna & Prof.dr. W.F.C. Verschoor, EPS-2011-225-F&A, ISBN: 978-90-5892-268-7, http://hdl.handle.net/1765/1

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

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

Boer, C.A., *Distributed Simulation in Industry,* Promotors: Prof.dr. A. de Bruin & Prof.dr.ir. A. Verbraeck, EPS-2005-065-LIS, ISBN: 90-5892-093-3, http://hdl.handle.net/1765/6925

Boer, N.I., *Knowledge Sharing within Organizations: A situated and Relational Perspective,* Promotor: Prof.dr. K. Kumar, EPS-2005-060-LIS, ISBN: 90-5892-086-0, http://hdl.handle.net/1765/6770

Boer-Sorbán, K., *Agent-Based Simulation of Financial Markets: A modular, Continuous-Time Approach,* Promotor: Prof.dr. A. de Bruin, EPS-2008-119-LIS, ISBN: 90-5892-155-0, http://hdl.handle.net/1765/10870

Boon, C.T., *HRM and Fit: Survival of the Fittest!?,* Promotors: Prof.dr. J. Paauwe & Prof.dr. D.N. den Hartog, EPS-2008-129-ORG, ISBN: 978-90-5892-162-8, http://hdl.handle.net/1765/12606

Borst, W.A.M., *Understanding Crowdsourcing: Effects of Motivation and Rewards on Participation and Performance in Voluntary Online Activities,* Promotors: Prof.dr.ir. J.C.M. van den Ende & Prof.dr.ir. H.W.G.M. van Heck, EPS-2010-221-LIS, ISBN: 978-90-5892-262-5, http://hdl.handle.net/1765/ 21914

Braun, E., *City Marketing: Towards an Integrated Approach,* Promotor: Prof.dr. L. van den Berg, EPS-2008-142-MKT, ISBN: 978-90-5892-180-2, http://hdl.handle.net/1765/13694

Brito, M.P. de, *Managing Reverse Logistics or Reversing Logistics Management?* Promotors: Prof.dr.ir. R. Dekker & Prof.dr. M. B. M. de Koster, EPS-2004-035-LIS, ISBN: 90-5892-058-5, http://hdl.handle.net/1765/1132

Brohm, R., *Polycentric Order in Organizations: A Dialogue between Michael Polanyi and IT-Consultants on Knowledge, Morality, and Organization,* Promotors: Prof.dr. G. W. J. Hendrikse & Prof.dr. H. K. Letiche, EPS-2005-063-ORG, ISBN: 90-5892-095-X, http://hdl.handle.net/1765/6911

Brumme, W.-H., *Manufacturing Capability Switching in the High-Tech Electronics Technology Life Cycle,* Promotors: Prof.dr.ir. J.A.E.E. van Nunen & Prof.dr.ir. L.N. Van Wassenhove, EPS-2008-126-LIS, ISBN: 978-90-5892-150-5, http://hdl.handle.net/1765/12103

Budiono, D.P., *The Analysis of Mutual Fund Performance: Evidence from U.S. Equity Mutual Funds,* Promotor: Prof.dr.ir. M.J.C.M. Verbeek, EPS-2010-185-F&A, ISBN: 978-90-5892-224-3, http://hdl.handle.net/1765/1

Burgers, J.H., *Managing Corporate Venturing: Multilevel Studies on Project Autonomy, Integration, Knowledge Relatedness, and Phases in the New Business Development Process,* Promotors: Prof.dr.ir. F.A.J. Van den Bosch & Prof.dr. H.W. Volberda, EPS-2008-136-STR, ISBN: 978-90-5892-174-1, http://hdl.handle.net/1765/13484

Campbell, R.A.J., *Rethinking Risk in International Financial Markets*, Promotor: Prof.dr. C.G. Koedijk, EPS-2001-005-F&A, ISBN: 90-5892-008-9, http://hdl.handle.net/1765/306

Carvalho de Mesquita Ferreira, L., *Attention Mosaics: Studies of Organizational Attention*, Promotors: Prof.dr. P.M.A.R. Heugens & Prof.dr. J. van Oosterhout, EPS-2010-205-ORG, ISBN: 978-90-5892-242-7, http://hdl.handle.net/1765/1

Chen, C.-M., *Evaluation and Design of Supply Chain Operations Using DEA*, Promotor: Prof.dr. J.A.E.E. van Nunen, EPS-2009-172-LIS, ISBN: 978-90-5892-209-0, http://hdl.handle.net/1765/1

Chen, H., *Individual Mobile Communication Services and Tariffs*, Promotor: Prof.dr. L.F.J.M. Pau, EPS-2008-123-LIS, ISBN: 90-5892-158-1, http://hdl.handle.net/1765/11141

Chen, Y., *Labour Flexibility in China's Companies: An Empirical Study*, Promotors: Prof.dr. A. Buitendam & Prof.dr. B. Krug, EPS-2001-006-ORG, ISBN: 90-5892-012-7, http://hdl.handle.net/1765/307

Damen, F.J.A., *Taking the Lead: The Role of Affect in Leadership Effectiveness*, Promotor: Prof.dr. D.L. van Knippenberg, EPS-2007-107-ORG, http://hdl.handle.net/1765/10282

Daniševská, P., *Empirical Studies on Financial Intermediation and Corporate Policies*, Promotor: Prof.dr. C.G. Koedijk, EPS-2004-044-F&A, ISBN: 90-5892-070-4, http://hdl.handle.net/1765/1518

Defilippi Angeldonis, E.F., *Access Regulation for Naturally Monopolistic Port Terminals: Lessons from Regulated Network Industries,* Promotor: Prof.dr. H.E. Haralambides, EPS-2010-204-LIS, ISBN: 978-90-5892-245-8, http://hdl.handle.net/1765/1

Delporte-Vermeiren, D.J.E., *Improving the Flexibility and Profitability of ICT-enabled Business Networks: An Assessment Method and Tool*, Promotors: Prof. mr. dr. P.H.M. Vervest & Prof.dr.ir. H.W.G.M. van Heck, EPS-2003-020-LIS, ISBN: 90-5892-040-2, http://hdl.handle.net/1765/359

Derwall, J.M.M., *The Economic Virtues of SRI and CSR*, Promotor: Prof.dr. C.G. Koedijk, EPS-2007-101-F&A, ISBN: 90-5892-132-8, http://hdl.handle.net/1765/8986

Diepen, M. van, *Dynamics and Competition in Charitable Giving*, Promotor: Prof.dr. Ph.H.B.F. Franses, EPS-2009-159-MKT, ISBN: 978-90-5892-188-8, http://hdl.handle.net/1765/14526

Dietvorst, R.C., *Neural Mechanisms Underlying Social Intelligence and Their Relationship with the Performance of Sales Managers,* Promotor: Prof.dr. W.J.M.I. Verbeke, EPS-2010-215-MKT, ISBN: 978-90-5892-257-1, http://hdl.handle.net/1765/21188

Dietz, H.M.S., *Managing (Sales)People towards Perfomance: HR Strategy, Leadership & Teamwork*, Promotor: Prof.dr. G.W.J. Hendrikse, EPS-2009-168-ORG, ISBN: 978-90-5892-210-6, http://hdl.handle.net/1765/1

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

Eijk, A.R. van der, *Behind Networks: Knowledge Transfer, Favor Exchange and Performance*, Promotors: Prof.dr. S.L. van de Velde & Prof.dr.drs. W.A. Dolfsma, EPS-2009-161-LIS, ISBN: 978-90-5892-190-1, http://hdl.handle.net/1765/14613

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

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

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

Feng, L., *Motivation, Coordination and Cognition in Cooperatives,* Promotor: Prof.dr. G.W.J. Hendrikse, EPS-2010-220-ORG, ISBN: 90-5892-261-8, http://hdl.handle.net/1765/21680

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

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

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

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

Gertsen, H.F.M., *Riding a Tiger without Being Eaten: How Companies and Analysts Tame Financial Restatements and Influence Corporate Reputation*, Promotor: Prof.dr. C.B.M. van Riel, EPS-2009-171-ORG, ISBN: 90-5892-214-4, http://hdl.handle.net/1765/1

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

Gijsbers, G.W., *Agricultural Innovation in Asia: Drivers, Paradigms and Performance,* Promotor: Prof.dr. R.J.M. van Tulder, EPS-2009-156-ORG, ISBN: 978-90-5892-191-8, http://hdl.handle.net/1765/14524

Gong, Y., *Stochastic Modelling and Analysis of Warehouse Operations*, Promotors: Prof.dr. M.B.M. de Koster & Prof.dr. S.L. van de Velde, EPS-2009-180-LIS, ISBN: 978-90-5892-219-9, http://hdl.handle.net/1765/1

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

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

Greeven, M.J., *Innovation in an Uncertain Institutional Environment: Private Software Entrepreneurs in Hangzhou, China*, Promotor: Prof.dr. B. Krug, EPS-2009-164-ORG, ISBN: 978-90-5892-202-1, http://hdl.handle.net/1765/15426

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

Guenster, N.K., *Investment Strategies Based on Social Responsibility and Bubbles*, Promotor: Prof.dr. C.G. Koedijk, EPS-2008-175-F&A, ISBN: 978-90-5892-206-9, http://hdl.handle.net/1765/1

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

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

Hakimi, N.A, *Leader Empowering Behaviour: The Leader's Perspective: Understanding the Motivation behind Leader Empowering Behaviour*, Promotor: Prof.dr. D.L. van Knippenberg, EPS-2010-184-ORG, http://hdl.handle.net/1765/1

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

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

Hensmans, M., *A Republican Settlement Theory of the Firm: Applied to Retail Banks in England and the Netherlands (1830-2007)*, Promotors: Prof.dr. A. Jolink& Prof.dr. S.J. Magala, EPS-2010-193-ORG, ISBN 90-5892-235-9, http://hdl.handle.net/1765/1

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

Hernandez Mireles, C., *Marketing Modeling for New Products,* Promotor: Prof.dr. P.H. Franses, EPS-2010-202-MKT, ISBN 90-5892-237-3, http://hdl.handle.net/1765/1

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

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

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

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

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

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

Huang, X., *An Analysis of Occupational Pension Provision: From Evaluation to Redesigh,* Promotors: Prof.dr. M.J.C.M. Verbeek & Prof.dr. R.J. Mahieu, EPS-2010-196-F&A, ISBN: 90-5892-239-7, http://hdl.handle.net/1765/1

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

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

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

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

Jalil, M.N., *Customer Information Driven After Sales Service Management: Lessons from Spare Parts Logistics,* Promotor: Prof.dr. L.G. Kroon, EPS-2011-222-LIS, ISBN: 978-90-5892-264-9, http://hdl.handle.net/1765/22156

Jansen, J.J.P., *Ambidextrous Organizations*, Promotors: Prof.dr.ir. F.A.J. Van den Bosch & Prof.dr. H.W. Volberda, EPS-2005-055-STR, ISBN: 90-5892-081-X, http://hdl.handle.net/1765/6774

Jaspers, F.P.H., *Organizing Systemic Innovation*, Promotor: Prof.dr.ir. J.C.M. van den Ende, EPS-2009-160-ORG, ISBN: 978-90-5892-197-), http://hdl.handle.net/1765/14974

Jennen, M.G.J., *Empirical Essays on Office Market Dynamics*, Promotors: Prof.dr. C.G. Koedijk & Prof.dr. D. Brounen, EPS-2008-140-F&A, ISBN: 978-90-5892-176-5, http://hdl.handle.net/1765/13692

Jiang, T., *Capital Structure Determinants and Governance Structure Variety in Franchising*, Promotors: Prof.dr. G. Hendrikse & Prof.dr. A. de Jong, EPS-2009-158-F&A, ISBN: 978-90-5892-199-4, http://hdl.handle.net/1765/14975

Jiao, T., *Essays in Financial Accounting*, Promotor: Prof.dr. G.M.H. Mertens, EPS-2009-176-F&A, ISBN: 978-90-5892-211-3, http://hdl.handle.net/1765/1

Jong, C. de, *Dealing with Derivatives: Studies on the Role, Informational Content and Pricing of Financial Derivatives*, Promotor: Prof.dr. C.G. Koedijk, EPS-2003-023-F&A, ISBN: 90-5892-043-7, http://hdl.handle.net/1765/1043

Kaa, G. van, *Standard Battles for Complex Systems: Empirical Research on the Home Network,* Promotors: Prof.dr.ir. J. van den Ende & Prof.dr.ir. H.W.G.M. van Heck, EPS-2009-166-ORG, ISBN: 978-90-5892-205-2, http://hdl.handle.net/1765/1

Kagie, M., *Advances in Online Shopping Interfaces: Product Catalog Maps and Recommender Systems,* Promotor: Prof.dr. P.J.F. Groenen, EPS-2010-195-MKT, ISBN: 978-90-5892-233-5, http://hdl.handle.net/1765/1

Karreman, B., *Financial Services and Emerging Markets,* Promotors: Prof.dr. G.A. van der Knaap & Prof.dr. H.P.G. Pennings, EPS-2011-223-ORG, ISBN: 978-90-5892-266-3, http://hdl.handle.net/1765/1

Keizer, A.B., *The Changing Logic of Japanese Employment Practices: A Firm-Level Analysis of Four Industries,* Promotors: Prof.dr. J.A. Stam & Prof.dr. J.P.M. Groenewegen, EPS-2005-057-ORG, ISBN: 90-5892-087-9, http://hdl.handle.net/1765/6667

Kijkuit, R.C., *Social Networks in the Front End: The Organizational Life of an Idea,* Promotor: Prof.dr. B. Nooteboom, EPS-2007-104-ORG, ISBN: 90-5892-137-6, http://hdl.handle.net/1765/10074

Kippers, J., *Empirical Studies on Cash Payments,* Promotor: Prof.dr. Ph.H.B.F. Franses, EPS-2004-043-F&A, ISBN: 90-5892-069-0, http://hdl.handle.net/1765/1520

Klein, M.H., *Poverty Alleviation through Sustainable Strategic Business Models: Essays on Poverty Alleviation as a Business Strategy,* Promotor: Prof.dr. H.R. Commandeur, EPS-2008-135-STR, ISBN: 978-90-5892-168-0, http://hdl.handle.net/1765/13482

Knapp, S., *The Econometrics of Maritime Safety: Recommendations to Enhance Safety at Sea*, Promotor: Prof.dr. Ph.H.B.F. Franses, EPS-2007-096-ORG, ISBN: 90-5892-127-1, http://hdl.handle.net/1765/7913

Kole, E., *On Crises, Crashes and Comovements*, Promotors: Prof.dr. C.G. Koedijk & Prof.dr. M.J.C.M. Verbeek, EPS-2006-083-F&A, ISBN: 90-5892-114-X, http://hdl.handle.net/1765/7829

Kooij-de Bode, J.M., *Distributed Information and Group Decision-Making: Effects of Diversity and Affect*, Promotor: Prof.dr. D.L. van Knippenberg, EPS-2007-115-ORG, http://hdl.handle.net/1765/10722

Koppius, O.R., *Information Architecture and Electronic Market Performance*, Promotors: Prof.dr. P.H.M. Vervest & Prof.dr.ir. H.W.G.M. van Heck, EPS-2002-013-LIS, ISBN: 90-5892-023-2, http://hdl.handle.net/1765/921

Kotlarsky, J., *Management of Globally Distributed Component-Based Software Development Projects,* Promotor: Prof.dr. K. Kumar, EPS-2005-059-LIS, ISBN: 90-5892-088-7, http://hdl.handle.net/1765/6772

Krauth, E.I., *Real-Time Planning Support: A Task-Technology Fit Perspective,* Promotors: Prof.dr. S.L. van de Velde & Prof.dr. J. van Hillegersberg, EPS-2008-155-LIS, ISBN: 978-90-5892-193-2, http://hdl.handle.net/1765/14521

Kuilman, J., *The Re-Emergence of Foreign Banks in Shanghai: An Ecological Analysis,* Promotor: Prof.dr. B. Krug, EPS-2005-066-ORG, ISBN: 90-5892-096-8, http://hdl.handle.net/1765/6926

Kwee, Z., *Investigating Three Key Principles of Sustained Strategic Renewal: A Longitudinal Study of Long-Lived Firms,* Promotors: Prof.dr.ir. F.A.J. Van den Bosch & Prof.dr. H.W. Volberda, EPS-2009-174-STR, ISBN: 90-5892-212-0, http://hdl.handle.net/1765/1

Langen, P.W. de, *The Performance of Seaport Clusters: A Framework to Analyze Cluster Performance and an Application to the Seaport Clusters of Durban, Rotterdam and the Lower Mississippi*, Promotors: Prof.dr. B. Nooteboom & Prof. drs. H.W.H. Welters, EPS-2004-034-LIS, ISBN: 90-5892-056-9, http://hdl.handle.net/1765/1133

Larco Martinelli, J.A., *Incorporating Worker-Specific Factors in Operations Management Models,* Promotors: Prof.dr.ir. J. Dul & Prof.dr. M.B.M. de Koster, EPS-2010-217-LIS, ISBN: 90-5892-263-2, http://hdl.handle.net/1765/21527

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

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

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

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

Li, T., *Informedness and Customer-Centric Revenue Management,* Promotors: Prof.dr. P.H.M. Vervest & Prof.dr.ir. H.W.G.M. van Heck, EPS-2009-146-LIS, ISBN: 978-90-5892-195-6, http://hdl.handle.net/1765/14525

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

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

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

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

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

Maas, K.E.G., *Corporate Social Performance: From Output Measurement to Impact Measurement,* Promotor: Prof.dr. H.R. Commandeur, EPS-2009-182-STR, ISBN: 978-90-5892-225-0, http://hdl.handle.net/1765/1

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

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

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

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

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

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

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

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

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

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

Moonen, J.M., *Multi-Agent Systems for Transportation Planning and Coordination*, Promotors: Prof.dr. J. van Hillegersberg & Prof.dr. S.L. van de Velde, EPS-2009-177-LIS, ISBN: 978-90-5892-216-8, http://hdl.handle.net/1

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

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

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

Nederveen Pieterse, A., *Goal Orientation in Teams: The Role of Diversity*, Promotor: Prof.dr. D.L. van Knippenberg, EPS-2009-162-ORG, http://hdl.handle.net/1765/15240

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

Niesten, E.M.M.I., *Regulation, Governance and Adaptation: Governance Transformations in the Dutch and French Liberalizing Electricity Industries*, Promotors: Prof.dr. A. Jolink & Prof.dr. J.P.M. Groenewegen, EPS-2009-170-ORG, ISBN: 978-90-5892-208-3, http://hdl.handle.net/1765/1

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

Nijdam, M.H., *Leader Firms: The Value of Companies for the Competitiveness of the Rotterdam Seaport Cluster,* Promotor: Prof.dr. R.J.M. van Tulder, EPS-2010-216-ORG, ISBN: 978-90-5892-256-4, http://hdl.handle.net/1765/21405

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

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

Noordegraaf-Eelens, L.H.J., *Contested Communication: A Critical Analysis of Central Bank Speech,* Promotor: Prof.dr. Ph.H.B.F. Franses, EPS-2010-209-MKT, ISBN: 978-90-5892-254-0, http://hdl.handle.net/1765/21061

Nuijten, I., *Servant Leadership: Paradox or Diamond in the Rough? A Multidimensional Measure and Empirical Evidence*, Promotor: Prof.dr. D.L. van Knippenberg, EPS-2009-183-ORG, http://hdl.handle.net/1765/1

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

Oosterhout, M., van, *Business Agility and Information Technology in Service Organizations,* Promotor: Prof,dr.ir. H.W.G.M. van Heck, EPS-2010-198-LIS, ISBN: 90-5092-236-6, http://hdl.handle.net/1765/1

Oostrum, J.M., van, *Applying Mathematical Models to Surgical Patient Planning*, Promotor: Prof.dr. A.P.M. Wagelmans, EPS-2009-179-LIS, ISBN: 978-90-5892-217-5, http://hdl.handle.net/1765/1

Otgaar, A.H.J., *Industrial Tourism: Where the Public Meets the Private,* Promotor: Prof.dr. L. van den Berg, EPS-2010-219-ORG, ISBN: 90-5892-259-5, http://hdl.handle.net/1765/21585

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

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

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

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

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

Pince, C., *Advances in Inventory Management: Dynamic Models,* Promotor: Prof.dr.ir. R. Dekker, EPS-2010-199-LIS, ISBN: 978-90-5892-243-4, http://hdl.handle.net/1765/1

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

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

Potthoff, D., *Railway Crew Rescheduling: Novel Approaches and Extensions,* Promotors: Prof.dr. A.P.M. Wagelmans & Prof.dr. L.G. Kroon, EPS-2010-210-LIS, ISBN: 90-5892-250-2, http://hdl.handle.net/1765/21084

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

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

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

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

Quariguasi Frota Neto, J., *Eco-efficient Supply Chains for Electrical and Electronic Products*, Promotors: Prof.dr.ir. J.A.E.E. van Nunen & Prof.dr.ir. H.W.G.M. van Heck, EPS-2008-152-LIS, ISBN: 978-90-5892-192-5, http://hdl.handle.net/1765/14785

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

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

Roelofsen, E.M., *The Role of Analyst Conference Calls in Capital Markets*, Promotors: Prof.dr. G.M.H. Mertens & Prof.dr. L.G. van der Tas RA, EPS-2010-190-F&A, ISBN: 978-90-5892-228-1, http://hdl.handle.net/1765/1

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

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

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

Rosmalen, J. van, *Segmentation and Dimension Reduction: Exploratory and Model-Based Approaches*, Promotor: Prof.dr. P.J.F. Groenen, EPS-2009-165-MKT, ISBN: 978-90-5892-201-4, http://hdl.handle.net/1765/15536

Roza, M.W., *The Relationship between Offshoring Strategies and Firm Performance: Impact of Innovation, Absorptive Capacity and Firm Size,* Promotors: Prof.dr. H.W. Volberda & Prof.dr.ing. F.A.J. van den Bosch, EPS-2011-214-STR, ISBN: 978-90-5892-265-6, http://hdl.handle.net/1765/22155

Rus, D., *The Dark Side of Leadership: Exploring the Psychology of Leader Self-serving Behavior*, Promotor: Prof.dr. D.L. van Knippenberg, EPS-2009-178-ORG, http://hdl.handle.net/1765/1

Samii, R., *Leveraging Logistics Partnerships: Lessons from Humanitarian Organizations,* Promotors: Prof.dr.ir. J.A.E.E. van Nunen & Prof.dr.ir. L.N. Van Wassenhove, EPS-2008-153-LIS, ISBN: 978-90-5892-186-4, http://hdl.handle.net/1765/14519

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

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

Schellekens, G.A.C., *Language Abstraction in Word of Mouth,* Promotor: Prof.dr.ir. A. Smidts, EPS-2010-218-MKT, ISBN: 978-90-5892-252-6, http://hdl.handle.net/1765/21580

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

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

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

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

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

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

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

Sotgiu, F., *Not All Promotions are Made Equal: From the Effects of a Price War to Cross-chain Cannibalization,* Promotors: Prof.dr. M.G. Dekimpe & Prof.dr.ir. B. Wierenga, EPS-2010-203-MKT, ISBN: 978-90-5892-238-0, http://hdl.handle.net/1765/1

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

Srour, F.J., *Dissecting Drayage: An Examination of Structure, Information, and Control in Drayage Operations*, Promotor: Prof.dr. S.L. van de Velde, EPS-2010-186-LIS, http://hdl.handle.net/1765/1

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

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

Sweldens, S.T.L.R., *Evaluative Conditioning 2.0: Direct versus Associative Transfer of Affect to Brands,* Promotor: Prof.dr. S.M.J. van Osselaer, EPS-2009-167-MKT, ISBN: 978-90-5892-204-5, http://hdl.handle.net/1765/1

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

Tempelaar, M.P., *Organizing for Ambidexterity: Studies on the Pursuit of Exploration and Exploitation through Differentiation, Integration, Contextual and Individual Attributes,* Promotors: Prof.dr.ing. F.A.J. van den Bosch & Prof.dr. H.W. Volberda, EPS-2010-191-STR, ISBN: 978-90-5892-231-1, http://hdl.handle.net/1765/1

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

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

Tiwari, V., *Transition Process and Performance in IT Outsourcing: Evidence from a Field Study and Laboratory Experiments,* Promotors: Prof.dr.ir. H.W.G.M. van Heck & Prof.dr. P.H.M. Vervest, EPS-2010-201-LIS, ISBN: 978-90-5892-241-0, http://hdl.handle.net/1765/1

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

Tzioti, S., *Let Me Give You a Piece of Advice: Empirical Papers about Advice Taking in Marketing,* Promotors: Prof.dr. S.M.J. van Osselaer & Prof.dr.ir. B. Wierenga, EPS-2010-211-MKT, ISBN: 978-90-5892-251-9, hdl.handle.net/1765/21149

Vaccaro, I.G., *Management Innovation: Studies on the Role of Internal Change Agents,* Promotors: Prof.dr. F.A.J. van den Bosch & Prof.dr. H.W. Volberda, EPS-2010-212-STR, ISBN: 978-90-5892-253-3, hdl.handle.net/1765/21150

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

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

Verheijen, H.J.J., *Vendor-Buyer Coordination in Supply Chains,* Promotor: Prof.dr.ir. J.A.E.E. van Nunen, EPS-2010-194-LIS, ISBN: 90-5892-234-2, http://hdl.handle.net/1765/1

Verheul, I., *Is There a (Fe)male Approach? Understanding Gender Differences in Entrepreneurship*, Promotor: Prof.dr. A.R. Thurik, EPS-2005-054-ORG, ISBN: 90-5892-080-1, http://hdl.handle.net/1765/2005

Verwijmeren, P., *Empirical Essays on Debt, Equity, and Convertible Securities*, Promotors: Prof.dr. A. de Jong & Prof.dr. M.J.C.M. Verbeek, EPS-2009-154-F&A, ISBN: 978-90-5892-187-1, http://hdl.handle.net/1765/14312

Vis, I.F.A., *Planning and Control Concepts for Material Handling Systems*, Promotors: Prof.dr. M.B.M. de Koster & Prof.dr.ir. R. Dekker, EPS-2002-014-LIS, ISBN: 90-5892-021-6, http://hdl.handle.net/1765/866

Vlaar, P.W.L., *Making Sense of Formalization in Interorganizational Relationships: Beyond Coordination and Control,* Promotors: Prof.dr.ir. F.A.J. Van den Bosch & Prof.dr. H.W. Volberda, EPS-2006-075-STR, ISBN 90-5892-103-4, http://hdl.handle.net/1765/7326

Vliet, P. van, *Downside Risk and Empirical Asset Pricing*, Promotor: Prof.dr. G.T. Post, EPS-2004-049-F&A, ISBN: 90-5892-07-55, http://hdl.handle.net/1765/1819

Vlist, P. van der, *Synchronizing the Retail Supply Chain*, Promotors: Prof.dr.ir. J.A.E.E. van Nunen & Prof.dr. A.G. de Kok, EPS-2007-110-LIS, ISBN: 90-5892-142-0, http://hdl.handle.net/1765/10418

Vries-van Ketel E. de, *How Assortment Variety Affects Assortment Attractiveness: A Consumer Perspective*, Promotors: Prof.dr. G.H. van Bruggen & Prof.dr.ir. A. Smidts, EPS-2006-072-MKT, ISBN: 90-5892-101-8, http://hdl.handle.net/1765/7193

Vromans, M.J.C.M., *Reliability of Railway Systems*, Promotors: Prof.dr. L.G. Kroon, Prof.dr.ir. R. Dekker & Prof.dr.ir. J.A.E.E. van Nunen, EPS-2005-062-LIS, ISBN: 90-5892-089-5, http://hdl.handle.net/1765/6773

Vroomen, B.L.K., *The Effects of the Internet, Recommendation Quality and Decision Strategies on Consumer Choice*, Promotor: Prof.dr. Ph.H.B.F. Franses, EPS-2006-090-MKT, ISBN: 90-5892-122-0, http://hdl.handle.net/1765/8067

Waal, T. de, *Processing of Erroneous and Unsafe Data*, Promotor: Prof.dr.ir. R. Dekker, EPS-2003-024-LIS, ISBN: 90-5892-045-3, http://hdl.handle.net/1765/870

Waard, E.J. de, *Engaging Environmental Turbulence: Organizational Determinants for Repetitive Quick and Adequate Responses*, Promotors: Prof.dr. H.W. Volberda & Prof.dr. J. Soeters, EPS-2010-189-STR, ISBN: 978-90-5892-229-8, http://hdl.handle.net/1765/1

Wall, R.S., *Netscape: Cities and Global Corporate Networks,* Promotor: Prof.dr. G.A. van der Knaap, EPS-2009-169-ORG, ISBN: 978-90-5892-207-6, http://hdl.handle.net/1765/1

Watkins Fassler, K., *Macroeconomic Crisis and Firm Performance,* Promotors: Prof.dr. J. Spronk & Prof.dr. D.J. van Dijk, EPS-2007-103-F&A, ISBN: 90-5892-138-3, http://hdl.handle.net/1765/10065

Weerdt, N.P. van der, *Organizational Flexibility for Hypercompetitive Markets: Empirical Evidence of the Composition and Context Specificity of Dynamic Capabilities and Organization Design Parameters,* Promotor: Prof.dr. H.W. Volberda, EPS-2009-173-STR, ISBN: 978-90-5892-215-1, http://hdl.handle.net/1765/1

Wennekers, A.R.M., *Entrepreneurship at Country Level: Economic and Non-Economic Determinants,* Promotor: Prof.dr. A.R. Thurik, EPS-2006-81-ORG, ISBN: 90-5892-115-8, http://hdl.handle.net/1765/7982

Wielemaker, M.W., *Managing Initiatives: A Synthesis of the Conditioning and Knowledge-Creating View*, Promotors: Prof.dr. H.W. Volberda & Prof.dr. C.W.F. Baden-Fuller, EPS-2003-28-STR, ISBN: 90-5892-050-X, http://hdl.handle.net/1765/1042

Wijk, R.A.J.L. van, *Organizing Knowledge in Internal Networks: A Multilevel Study*, Promotor: Prof.dr.ir. F.A.J. van den Bosch, EPS-2003-021-STR, ISBN: 90-5892-039-9, http://hdl.handle.net/1765/347

Wolters, M.J.J., *The Business of Modularity and the Modularity of Business*, Promotors: Prof. mr. dr. P.H.M. Vervest & Prof.dr.ir. H.W.G.M. van Heck, EPS-2002-011-LIS, ISBN: 90-5892-020-8, http://hdl.handle.net/1765/920

Wubben, M.J.J., *Social Functions of Emotions in Social Dilemmas*, Promotors: Prof.dr. D. De Cremer & Prof.dr. E. van Dijk, EPS-2009-187-ORG, http://hdl.handle.net/1765/1

Xu, Y., *Empirical Essays on the Stock Returns, Risk Management, and Liquidity Creation of Banks*, Promotor: Prof.dr. M.J.C.M. Verbeek, EPS-2010-188-F&A, http://hdl.handle.net/1765/1

Yang, J., *Towards the Restructuring and Co-ordination Mechanisms for the Architecture of Chinese Transport Logistics*, Promotor: Prof.dr. H.E. Harlambides, EPS-2009-157-LIS, ISBN: 978-90-5892-198-7, http://hdl.handle.net/1765/14527

Yu, M., *Enhancing Warehouse Perfromance by Efficient Order Picking*, Promotor: Prof.dr. M.B.M. de Koster, EPS-2008-139-LIS, ISBN: 978-90-5892-167-3, http://hdl.handle.net/1765/13691

Zhang, X., *Strategizing of Foreign Firms in China: An Institution-based Perspective*, Promotor: Prof.dr. B. Krug, EPS-2007-114-ORG, ISBN: 90-5892-147-5, http://hdl.handle.net/1765/10721

Zhang, X., *Scheduling with Time Lags,* Promotor: Prof.dr. S.L. van de Velde, EPS-2010-206-LIS, ISBN: 978-90-5892-244-1, http://hdl.handle.net/1765/1

Zhou, H., *Knowledge, Entrepreneurship and Performance: Evidence from Country-level and Firm-level Studies,* Promotors: Prof.dr. A.R. Thurik & Prof.dr. L.M. Uhlaner, EPS-2010-207-ORG, ISBN: 90-5892-248-9, http://hdl.handle.net/1765/1

Zhu, Z., *Essays on China's Tax System*, Promotors: Prof.dr. B. Krug & Prof.dr. G.W.J. Hendrikse, EPS-2007-112-ORG, ISBN: 90-5892-144-4, http://hdl.handle.net/1765/10502

Zwart, G.J. de, *Empirical Studies on Financial Markets: Private Equity, Corporate Bonds and Emerging Markets*, Promotors: Prof.dr. M.J.C.M. Verbeek & Prof.dr. D.J.C. van Dijk, EPS-2008-131-F&A, ISBN: 978-90-5892-163-5, http://hdl.handle.net/1765/12703

RELIABILITY AND RANKINGS

Questionnaires are an important way to gather information about large populations for both qualitative and quantitative research. Hence, the value of a good questionnaire design and the quality of questionnaire data cannot be emphasized enough. This thesis discusses some aspects of the statistical analysis of measurement data obtained via questionnaires.

In the first part of this thesis we focus on maximizing scale reliability. We derive the asymptotic distribution of maximal reliability measures to construct confidence intervals in order to assess the adequacy of the measure. We stress the use of confidence intervals accompanying single measures that summarize the parameters to assess the adequacy of the measure. The results can lead to better designs of questionnaires, which in turn lead to more precise survey outcomes.

The second part of this thesis proposes methodologies to perform statistical analysis of stated consumer preferences measured as rankings data, especially in the context of conjoint measurements. Our statistical models allow for the efficient use of partial rankings to collect preference data. As a partial rankings task amount to a smaller burden for respondents than a complete ranking task, they may be more motivated to complete the task and as such the quality of the obtained data may improve. Moreover, we show that our model is able to extract sufficient preference information from partial rankings data to take into account respondents' heterogeneity in their choice and preference behavior, which is generally assumed in marketing. This certainly will help marketers to identify and target consumers by understanding their preference behavior, and to implement a more efficient and optimal marketing strategy.

ERIM

The Erasmus Research Institute of Management (ERIM) is the Research School (Onderzoekschool) in the field of management of the Erasmus University Rotterdam. The founding participants of ERIM are Rotterdam School of Management (RSM), and the Erasmus School of Economics (ESE). ERIM was founded in 1999 and is officially accredited by the Royal Netherlands Academy of Arts and Sciences (KNAW). The research undertaken by ERIM is focused on the management of the firm in its environment, its intra- and interfirm relations, and its business processes in their interdependent connections.

The objective of ERIM is to carry out first rate research in management, and to offer an advanced doctoral programme in Research in Management. Within ERIM, over three hundred senior researchers and PhD candidates are active in the different research programmes. From a variety of academic backgrounds and expertises, the ERIM community is united in striving for excellence and working at the forefront of creating new business knowledge.

Erasmus Research Institute of Management - ERIM

ERIM PhD Series

Research in Management

Erasmus Research Institute of Management - **ERIM**
Rotterdam School of Management (RSM)
Erasmus School of Economics (ESE)
P.O. Box 1738, 3000 DR Rotterdam
The Netherlands

Tel. +31 10 408 11 82
Fax +31 10 408 96 40
E-mail info@erim.eur.nl
Internet www.erim.eur.nl

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