ATHENS UNIVERSITY
OF ECONOMICS AND BUSINESS

DEPARTMENT OF STATISTICS

POSTGRADUATE PROGRAM

PARAMETER CONSTRAINTS IN LATENT TRAIT MODELS UNDER THE ITEM RESPONSE THEORY APPROACH

By

Spyridoula I. Tsonaka

A THESIS
Submitted to the Department of Statistics
of the Athens University of Economics and Business
in partial fulfilment of the requirements for
the degree of Master of Science in Statistics

Athens, Greece
2004

ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

DEPARTMENT OF STATISTICS

POSTGRADUATE PROGRAM

PARAMETER CONSTRAINTS IN LATENT TRAIT MODELS UNDER THE ITEM RESPONSE THEORY APPROACH

By

Spyridoula I. Tsonaka

A THESIS

Submitted to the Department of Statistics
of the Athens University of Economics and Business
in partial fulfilment of the requirements for
the degree of Master of Science in Statistics

Athens, Greece
September 2003

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ

ΠΕΡΙΟΡΙΣΜΟΙ ΣΤΙΣ ΠΑΡΑΜΕΤΡΟΥΣ ΜΟΝΤΕΛΩΝ ΛΑΝΘΑΝΟΥΣΩΝ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ ΥΠΟ ΤΗΝ ΠΡΟΣΕΓΓΙΣΗ ΤΗΣ ΘΕΩΡΙΑΣ ΑΠΑΝΤΗΣΕΩΝ ΣΕ ΞΕΧΩΡΙΣΤΑ ΑΝΤΙΚΕΙΜΕΝΑ

Σπυριδούλα Ι. Τσονάκα

ΔΙΑΤΡΙΒΗ

Που υποβλήθηκε στο Τμήμα Στατιστικής
του Οικονομικού Πανεπιστημίου Αθηνών
ως μέρος των απαιτήσεων για την απόκτηση
Μεταπτυχιακού Διπλώματος Ειδίκευσης στη Στατιστική

Αθήνα
Σεπτέμβριος 2003

ATHENS UNIVERSITY
OF ECONOMICS AND BUSINESS
DEPARTMENT OF STATISTICS

A Thesis submitted in partial fulfillment of

the requirements for the degree of

Master of Science

PARAMETER CONSTRAINTS IN LATENT TRAIT MODELS UNDER THE ITEM RESPONSE THEORY APPROACH

Spyridoula I. Tsonaka

Approved by the Graduate Committee

E. Moustaki	E. A. Erosheva	M. Knott
Assistant Professor	Research Assistant Professor	Professor
Thesis Supervisor		Members of the Commitee

Athens, June 2004

Michael Zazanis, Associate Professor
Director of the Graduate Program

Acknowledgements

The author of this dissertation gratefully acknowledges the support and the contribution of many people to the fulfilment of the dissertation.

First of all, I would like to thank my parents and my sister for their trust and continuous support during all these years of my studies. In addition, I would like to express my gratefulness to all of my instructors in the Athens University of Economics and Business, who for five continuous years have exposed me to the entire gamut of statistical courses and helped me to delve into the science of Statistics. Besides, I would like to thank the Graduate Program Committee for giving me the chance to attend a small number of courses during my M.Sc. studies and in the meantime get involved with this dissertation that needs to be at publication level.

Besides, I feel grateful to the supervisor of this dissertation, Dr. Irini Moustaki, Assistant Professor at AUEB, who trusted me and whose ideas, comments and valuable suggestions have contributed to the fulfilment of my dissertation.

Finally, I would like to thank my colleague Dimitris Rizopoulos whose assistance and the fruitful cooperation that we had, have had a direct impact on the current organization of this dissertation. I would also like to thank him for kindly offering me his library `ltm` in order to fit the unconstrained latent trait models.

ABSTRACT

Spyridoula I. Tsonaka

Parameter Constraints in Latent Trait Models under the Item Response Theory approach

September 2003

This dissertation is aimed to introducing the concept of parameter constraints in latent trait models under the Item Response Theory approach. Such a concept has been broadly investigated under a competing approach to Item Response Theory (IRT), the Structural Equation Modelling (SEM) approach. However, the difficulties of fitting latent variable models with categorical responses, such as the latent trait models, and thus of fitting latent trait models with parameter constraints, under SEM have contributed to investigating this topic under the IRT approach. Among the different types of parameter constraints that have been studied in SEM approach, this dissertation considers only equality and fixed value parameter constraints, that is, assumptions that some of the unknown parameters of a latent trait model are equal to each other and to a fixed value respectively.

ΠΕΡΙΛΗΨΗ

Σπυριδούλα Ι. Τσονάκα

Περιορισμοί στις Παραμέτρους Μοντέλων Λανθάνουσων Χαρακτηριστικών υπό την Προσέγγιση της Θεωρίας Απαντήσεων σε Ξεχωριστά Αντικείμενα

Σεπτέμβριος 2003

Σκοπός της παρούσας διατριβής είναι η επιβολή περιορισμών στις παραμέτρους Μοντέλων Λανθάνουσων Χαρακτηριστικών (Latent Trait Models) υπό την προσέγγιση της Θεωρίας Απαντήσεων σε ξεχωριστά Αντικείμενα (Item Response Theory). Η έννοια αυτή έχει μελετηθεί εκτενώς υπό την προσέγγιση της Μοντελοποίησης Δομικών Εξισώσεων (Structural Equation Modelling). Ωστόσο, οι δυσκολίες που αντιμετωπίζει η δεύτερη προσέγγιση κατά την προσαρμογή Μοντέλων Λανθάνουσων Μεταβλητών (Latent Variable Models) σε κατηγορικά δεδομένα, όπως είναι και τα Μοντέλα Λανθάνουσων Χαρακτηριστικών, αλλά και κατά την προσαρμογή Μοντέλων Λανθάνουσων Χαρακτηριστικών με περιορισμούς στις παραμέτρους ενθάρρυνε τη μελέτη του ιδίου θέματος υπό την προσέγγιση της Θεωρίας Απαντήσεων σε ξεχωριστά Αντικείμενα. Από τα διάφορα είδη των περιορισμών στις παραμέτρους που έχουν μελετηθεί υπό την προσέγγιση της Μοντελοποίησης Δομικών Εξισώσεων σε αυτή τη διατριβή θα ασχοληθούμε μόνο με περιορισμούς που αναφέρονται σε ισότητες κάποιων παραμέτρων μεταξύ τους ή σε ισότητες των παραμέτρων με κάποια σταθερή τιμή.

Contents

1 Introduction 1

1.1 Latent Variables and Latent Variable Models 2

1.1.1 The necessity of Latent Variables 2

1.1.2 Different Types of Latent Variable models 4

1.2 Parameter Constraints in Latent Variable Models 5

1.2.1 Different Types of Parameter Constraints 6

1.2.2 The necessity of parameter constraints 7

1.3 Two Competing Approaches in Fitting Latent Variable Models 9

1.3.1 Item Response Theory - Basic Ideas 10

2 A Theoretical Framework 17

2.1 Introduction 17

2.2 General Linear Latent Variable Model 18

2.3 Latent Trait Model for Binary Responses 21

2.3.1 General Definition 21

2.3.2 Estimation 22

2.3.3 Quadrature Methods 27

2.4 Equality and Fixed value Constraints in Latent Trait Models . 29

2.4.1 Specification of the equality and fixed value constraints 30
2.4.2 Estimation Procedure 31
2.5 Factor Scores 44
2.6 Sampling Properties of the Maximum Likelihood Estimators . 45
2.7 Goodness-of-fit 46

3 Estimation Procedure-Optimization Algorithms **49**
3.1 Introduction 49
3.2 The Expectation-Maximization Algorithm 50
3.2.1 Introduction 50
3.2.2 General Definition of the EM Algorithm 52
3.2.3 EM Gradient Algorithm 53
3.2.4 The EM Gradient Algorithm in Latent Trait Models . 54
3.3 Quasi-Newton Methods 55
3.3.1 Introduction 55
3.3.2 BFGS Algorithm 56
3.3.3 Backtracking 58
3.4 Description of the `ltm.con` function 59

4 Applications **63**
4.1 Introduction 63
4.2 Latent Trait Model for Race data 64
4.2.1 Descriptive Analysis 65
4.2.2 One-Factor Unconstrained LTM 67
4.2.3 One-Factor constrained LTM 73
4.3 Latent Trait Model for Law School Admission Test data . . . 78

4.3.1 Descriptive Analysis 79
4.3.2 One-Factor Unconstrained LTM 81
4.3.3 One-Factor constrained LTM 84

List of Tables

4.1 Observed frequencies of response patterns, Race data 66
4.2 Percentages of expressing or not racial prejudice for the observed items, Race data 66
4.3 Pairwise associations between observed variables, Race data . 67
4.4 Parameter estimates and standard errors for the unconstrained one-factor LTM, Race data 69
4.5 Goodness-of-fit measures for the unconstrained one-factor model, Race data 70
4.6 Observed and expected frequencies of the response patterns, Race data 71
4.7 Parameter estimates and standard errors for the constrained one-factor LTM, Race data 74
4.8 Goodness-of-fit measures for the constrained one-factor model. Race data 74
4.9 Observed, expected frequencies and scores of the response patterns for the constrained one-factor LTM, Race data 75
4.10 Comparison between the constrained and the unconstrained model, Race data 75

4.11 Observed frequencies of response patterns, LSAT data 80
4.12 Percentages of answering correctly or wrongly to the observed items, LSAT data 80
4.13 Pairwise associations between the observed variables, LSAT data 82
4.14 Parameter estimates and standard errors for the unconstrained one-factor LTM, LSAT data 82
4.15 Goodness-of-fit measures for the unconstrained one-factor model, LSAT data 84
4.16 Observed and expected frequencies of the response patterns, LSAT data 85
4.17 Parameter estimates and standard errors for the constrained one-factor LTM, LSAT data 88
4.18 Goodness-of-fit measures for the constrained one-factor model, LSAT data 89
4.19 Observed, expected frequencies and scores of the response patterns for the constrained one-factor LTM, LSAT data 89
4.20 Comparison between the constrained and the unconstrained model 91

List of Figures

4.1 Item Characteristic Curves for the four items of the unconstrained latent trait model, Race data 72
4.2 Item Characteristic Curves for the four items of the constrained latent trait model, Race data 77
4.3 Item Characteristic Curves for the five items of the unconstrained latent trait model, LSAT data 86
4.4 Item Characteristic Curves for the five items of the constrained latent trait model, LSAT data 91

Chapter 1

Introduction

Latent quantities appear in many areas; for example in psychology we often talk about intelligence and verbal ability, in sociology we find ambition and racial prejudice and in economics, economic expectation. It is virtually impossible to talk about social phenomena without invoking such hypothetical variables. In this introductory chapter, we will define the latent variables and the methods used to measure them. In particular, in this chapter we will firstly deal with the latent variables, the latent variable models and their applications. Secondly, we will briefly describe the two competing approaches to fitting latent variable models. Finally, we will give the different types of parameter constraints that are usually used.

1.1 Latent Variables and Latent Variable Models

Latent variable models constitute an important tool for the analysis of multivariate data and are closely related to the standard regression models. A regression model expresses the relationship between one or more dependent variables and one or more independent variables, whereas in latent variable models the regression relationship is between manifest variables and the unobserved-latent variables. A distinguishing feature of the latent variable models is that of inverting the regression relationships to tell us about the latent variables when the manifest variables are given. Since we can never observe the latent variables, we only learn about this relationship indirectly.

1.1.1 The necessity of Latent Variables

One reason for introducing latent variables is the presence of certain constructs in the social and behavioral sciences that are not well defined such as social class, public opinion, quality of life or business confidence. Such concepts are referred to as latent variables or factors, since they are not directly observable even in the population and for which there exists no operational method for direct measurement. Although latent variables are not observable, certain of their effects on measurable variables are observable, and hence subject to study. A simple example of a latent variable is the concept of racial prejudice, which is not possible to be measured directly. However, being a racist inevitably affects one's opinions and behavior. That is, a racist is bound not to count members of a particular race among his/her friends and acquaintances, to approve or disapprove of a particular piece of government legislation, etc. Thus, by observing such measurable variables, which are regarded as indicators of the racial prejudice, one can "measure" the unobservable racial prejudice.

The second reason for considering latent variables is to explain the structure in a set of correlated observed variables. In fact, one of the major achievements in the behavioral sciences has been the development of methods to assess and explain the structure in a set of correlated, observed variables, in terms of a small number of latent variables. In practice, one chooses a variety of indicators, which can be measured, such as answers to a set of yes/no questions, and then attempts to extract what is common to them. A simple and familiar example that illustrates this considers the association between the responses to questions in Section VI of the Law School Admission Tests. Here there can be a third variable, which might account for this spurious relationship, the test takers' special trait that this section wishes to measure. Besides, it is well known that the responses to these questions are all associated with that trait. Thus, the association between the responses in the presence of the test taker's trait is vanished and we therefore conclude that the latter variable is responsible for such an association (Bartholomew and Knott, 1999; Muthén, 1978).

Another reason for considering latent variables is to reduce dimensionality. That is, the information conveyed in the interrelationships among the observed variables can be sufficiently expressed in terms of a smaller set of variables, the latent variables. Such a procedure is very useful, especially in social surveys that generate much information, because our ability to explore the structure in a huge data set will be in this way much improved. Otherwise, statistical methods that summarize the data by looking at the frequency distributions of responses and by providing summary measures such as percentages and correlation coefficients should be used. However, in the case of many response variables it may be difficult to "see" any pattern in their interrelationships by means of such summary statistics.

1.1.2 Different Types of Latent Variable models

The benefits of introducing latent variables in a multivariate analysis can be obtained by means of fitting latent variable models. According to the nature of both manifest and latent variables, there are different models available for assessing whether the observed relationships between a set of manifest variables may be accounted for by a small number of latent variables or not. Among these models are the factor analysis, latent profile analysis, latent class analysis and latent trait analysis models. Factor analysis models refer to the models that relate normal continuous manifest variables to normal continuous latent variables, latent profile analysis models refer to the models that relate continuous manifest variables to categorical latent variables, and latent class analysis models refer to the models that relate categorical manifest variables to categorical latent variables.

In this dissertation, we are mainly concerned with latent trait models, that is, models that relate binary observed variables to continuous latent variables. The development of these models is attributed to the educational testing where binary variables of getting some item right or wrong are supposed to be indicators of some 'traits' that the test takers possess.

1.2 Parameter Constraints in Latent Variable Models

In this section we will present an extremely common concept in the area of latent variable models, that of parameter constraints. By the term parameter constraints we mean that the parameters in the latent variable model, that is the intercepts and the factor loadings, are constrained to be equal to or be greater or less than a fixed value or other parameters according to some linear or non-linear function. However, these are not the only parameters in a latent variable model that need to be constrained, that is correlation coefficients may need to be constrained when correlated latent variables are considered. This concept of parameter constraints has been broadly developed under the SEM approach and discussed in relevant literature such as Lee (1980), McDonald (1980), Bentler and Weeks (1980), Lee and Tsui (1982), Bentler and Lee (1983), Rindskoph (1983 and 1984), etc. However, this issue of parameter constraints has been studied under the IRT approach only in terms of the latent class models (Goodman, 1974 (a), 1974 (b), Wright and Stone, 1979, Clogg and Goodman, 1985, Formann, 1985, van de Pol and Langeheine, 1990, Mooijaart and Heijden, 1992) whereas constrained latent trait models have not been studied until this dissertation.

In this section, we will firstly present the different types of constraints that have been developed in the literature when analyzing latent variable models. Secondly, we will give the reasons why the imposition of parameter constraints is necessary in fitting latent variable models.

1.2.1 Different Types of Parameter Constraints

Parameter constraints are common in latent variable models. In the literature two types of parameter constraints are used, the linear and the non-linear constraints. The linear constraints imply that a parameter is a linear function of other parameters, while the non-linear constraints imply that this function is non-linear.

As far as the linear constraints are concerned, we distinguish: a) the fixed value constraints, that is a parameter equals to a fixed value including zero (e.g., $\alpha = 0$), b) the equality constraints, that is a parameter equals to another or other parameter(s) (e.g., $\alpha_1 = \alpha_2$), c) the negative constraints, that is a parameter equals to the negative of another or other parameter(s) (e.g., $\alpha_1 = -\alpha_2$), d) the proportionality constraints, that is the ratio of two parameters equals to a fixed value (e.g., $\frac{\alpha_1}{\alpha_2} = k$) and e) the additive constraints, that is the sum of two or more parameters equals to a fixed value (e.g., $\alpha_1 + \alpha_2 = k$) (for $k = 0$, we have the negative constraints). A brief discussion of the zero, fixed-value and equality parameter constraints can be found, apart from this dissertation, in Mooijaart and Heijden (1992), Rindskoph (1984), etc.

As far as the non-linear constraints are concerned, we distinguish two main classifications the inequality and equality non-linear constraints. The inequality non-linear constraints imply that the parameter may be greater or less than a fixed value, including zero, (e.g., $\alpha_1 > 0$, $\alpha_1 \leq 0$, $\alpha_1 > k$, $\alpha_1 \leq k$, or greater or less than another parameter, (e.g., $\alpha_1 > \alpha_2$, $\alpha_1 \leq \alpha_2$) (Lee, 1980, Rindskoph, 1984). The equality non-linear constraints imply that a parameter equals to a non-linear function of another or other parameter(s), (e.g., $\alpha_1 = \alpha_2 \cdot \alpha_3$, $\alpha_1 = \alpha_2^2$). Provided all these types of parameter constraints, one may be curious about the cases in which such constraints are needed.

1.2.2 The necessity of parameter constraints

Parameter constraints are mainly used in the confirmatory phase of an analysis in which the researcher based on knowledge of the theory, empirical research or both, he or she postulates relations between the observed measures and the underlying factors a priori and then tests this hypothesized structure statistically. For example, suppose a researcher develops a new instrument designed to measure five facets of physical self-concept (e.g., Health, Sport Competence, Physical Appearance, Coordination, Body Strength). Based on prior knowledge the researcher would allow all sport competence self-concept items to be free to load on that factor, but restricted to have zero loadings on the remaining factors (zero linear constraints) (Byrne, 1998). The model would then be evaluated by statistical means to determine the adequacy of its goodness of fit to the sample data. As another illustration of prior knowledge of the effect of latent variables we can refer to Kenny and Cohen (1980) where a situation in over-time analysis is described in which the effect of measures on the pre-test versus the post-test increases by a proportional constant (proportionality constraints).

In addition, parameter constraints are considered in identification issues. The statistical identification issue focuses on whether or not there is a unique set of parameters consistent with the data. When the statistical identification does not hold, it means that the model contains insufficient information for the purpose of attaining a determinate solution of parameter estimation. Such a situation is usually detected by the large standard errors and high correlations between parameter estimates. Indeed, the imposition of constraints on particular parameters can sometimes be beneficial in helping the researcher to attain an over-identified model. In general, the problem of identification cannot be easily checked. However, an option is to use different starting values in separate analysis and if the model is identified, then the estimates should be identical.

Besides, the parameter constraints are regularly used in a second step of the analysis. In the first step, an unconstrained latent variable model is fitted and the statistical significance of some parameters is studied. In the second, the non-significant parameters are set equal to zero and thus the number of observations to be estimated is reduced (zero linear constraints). The goodness-of-fit of the constrained model is estimated and if it is good we conclude that the constrained model represents an adequate fit to the data. The same procedure is used when the parameter estimates are close and the equality between them is checked (equality constraints).

Finally, parameter constraints are common in multi-group analysis in which it is perfectly fine to specify equality constraints across groups between parameters of the same type. To test such constraints is often the main reason to estimate a multi-group model. In particular, we are mainly interested in within - between items and simultaneous estimation of different sets of constraints. For example, if one looks at the effects of mother and father on daughters and sons, one might want to force as a constraint that a parent has proportionally more influence on a same- than an opposite-gender child.

1.3 Two Competing Approaches in Fitting Latent Variable Models

There are two approaches in fitting latent variable models that have been developed the Structural Equation Modelling and Item Response Theory approach. The main difference between them is the estimation procedure in fitting latent variable models. In SEM the estimation procedure is based on the minimization of the distance between the observed and the estimated covariance matrix of the model, while in IRT the estimation procedure is based on the maximization of the likelihood function.

Another difference between these two approaches lies in the estimation of the latent trait model, that is the model in which the observed variables are binary and the latent variables continuous. According to the SEM approach the observed binary variables are generated by a set of underlying latent continuous variables. The parameters of the latent trait model are estimated by two or three-stage estimation methods. In the first stage first-order statistics, such as thresholds, means and variances, are estimated by maximum likelihood. In the second stage, second-order statistics, such as polychoric correlations, are estimated by conditional maximum likelihood given the estimates of the first stage. In the third stage, the parameters of the structural part of the model are estimated using a generalized least-squares or weighted least squares method based on the asymptotic covariance matrix of the polychoric correlations. According to the IRT approach, all the model parameters are estimated simultaneously by means of the maximum likelihood estimation method and there is no need to assume that the binary variables are a manifestation of other underlying variables.

The above differences between the two approaches to the estimation of latent trait models do also exist in the estimation of constrained latent trait models, that is, models under the imposition of equality and fixed value parameter constraints. Under the IRT approach the estimation of the constrained latent trait model is done straightforwardly by maximizing the log-likelihood function with respect to the different parameters. The software that has been developed to estimate constrained latent trait models in IRT approach is the function `ltm.con` that will be presented in Chapter 3, while in SEM approach the LISREL 8 program (Jöreskog and Sörbom, 1989, 1993b) is used.

Provided that the topic of our interest is studied under the IRT approach, in the proceeding of this section we will represent some of the fundamental principles of this approach.

1.3.1 Item Response Theory - Basic Ideas

Nowadays, Item Response Theory is used commonly by the largest testing companies in the United States and Europe for the design of tests, test assembly, test scaling and calibration, construction of test item banks, investigation of test item bias, and other common procedures in the test development process. Item Response Theory rests on two basic postulates: (a) the performance of an examinee on a test item can be predicted (or explained) by a set of factors called traits, latent traits, or abilities; and (b) the relationship between examinees' item performance and the set of traits underlying item performance can be described by a monotonically increasing function called an item characteristic function or item characteristic curve (ICC). This function specifies that examinees with higher scores on the traits have higher expected probabilities for answering the item correctly than examinees with lower scores on the traits. In fact, it provides the probability of examinees answering an item correctly for examinees at different points on the ability scale.

Many possible item response models exist, differing in the mathematical form of the item characteristic function and/or the number of parameters specified in the model. All IRT models contain one or more parameters describing the item and one or more parameters describing the examinee. The first step in any IRT application is to estimate these parameters. Besides, a given item response model may or may not be appropriate for a particular set of test data: that is, the model may not adequately predict or explain the data. Thus, it is essential to assess the fit of the model to the data.

When a given IRT model fits the test data of interest, several desirable features are obtained. Examinee ability estimates are not test dependent, and item indices are not group-dependent. Ability estimates obtained from different sets of items will be the same (except from measurement errors), and item parameter estimates obtained in different groups of examinees will be the same (except from measurement errors). In item response theory, item and ability parameters are said to be invariant. The property of invariance of an item and ability parameters is obtained by incorporating information about the items into the ability-estimation process and by incorporating information about the examinee's abilities into the item-parameter-estimation process.

IRT model Assumptions

The mathematical models employed in IRT specify that an examinee's probability of answering a given item correctly depends on the examinee's ability or abilities and the characteristics of the item. IRT models include a set of assumptions about the data to which the model is applied. Although the viability of assumptions cannot be determined directly, some indirect evidence can be collected and assessed as well.

An assumption common to the IRT models most widely used is that the items that make up the test measure only one ability. This is called the assumption of unidimensionality. Other assumptions made in all IRT models is the local independence and that the item characteristic function specified reflects the true relationship among the unobservable variables (abilities) and observable variables (item responses).

Unidimensionality. As stated above, a common assumption of IRT models is that only one ability is measured by a set of items in a test. This assumption cannot be strictly met because several cognitive, personality, and test-taking factors always affect test performance, at least to some extent. These factors might include level of motivation, test anxiety, ability to work quickly, tendency to guess when in doubt about answers, and cognitive skills in addition to the dominant one measured by the set of test items. What is required for the unidimensionality assumption to be met adequately by a set of data is the presence of a dominant component or factor that influences test performance. This dominant component or factor is referred to as the ability measured by the test; it should be noted, however, that ability is not necessarily inherent or unchangeable. Ability scores may be expected to change over time because of learning, forgetting, and other factors.

In fact, there has been a lot of concern with testing for unidimensionality. Such a testing can be done by finding the best way of detecting the effect of additional latent variables. This has been investigated by Holland (1981), Rosenbaum (1984), Holland and Rosenbaum (1985) and Stout (1987, 1990). Finally, item response models in which a single dominant ability is presumed sufficient to explain or account for examinee performance are referred to as unidimensional models. Models in which it is assumed that more than one ability is necessary to account for examinee test performance are referred to as multidimensional.

Local Independence. The assumption of local independence in an IRT model implies that the abilities specified in the model are the only factors influencing examinees' responses to test items. In other words, when the abilities influencing test performance are held constant, examinee's responses to any pair of items are statistically independent.

To state the definition of local independence more formally, let $\mathbf{y}$ be the complete set of abilities assumed to influence the performance of an examinee on the test. Let x_i be the response of a randomly chosen examinee to item i $(i = 1, ..., p)$. Let $P(x_i \mid \mathbf{y})$ denote the probability of the response of a randomly chosen examinee with abilities $\mathbf{y}$; $P(x_i = 1 \mid \mathbf{y})$ denotes the probability of a correct response, and $P(x_i = 0 \mid \mathbf{y})$ denotes the probability of an incorrect response. The property of local independence can be stated mathematically in the following way:

$$g(\mathbf{x} \mid \mathbf{y}) = \prod_{i=1}^{p} g_i (x_i \mid \mathbf{y}) \tag{1.1}$$

The property of local independence (or conditional independence) means that for a given examinee (or all examinees at a given ability value) the probability of a response pattern on a set of items is equal to the product of probabilities associated with the examinee's responses to the individual items. In other words, if the correlations among the x's are induced by a set of latent variables $\mathbf{y}$ then when all y's are accounted for, the x's will be uncorrelated if all the y's are held fixed. If this were not so the set of y's would not be complete and we should have to add at least one more. Thus q must be chosen so that (1.1) holds. Such an assumption cannot be tested empirically because there is no way in which $\mathbf{y}$ can be held fixed and therefore no way in which the independence can be checked. In addition, this assumption implies that $\mathbf{y}$ is sufficient to explain the dependencies among the x's. We do not assume that (1.1) holds; a key part of our analysis is directed to discovering the smallest q for which such a representation is adequate.

Invariance Property. The property of invariance of item and ability parameters implies that the parameters that characterize an item do not depend on the ability distribution of the examinees and the parameter that characterizes an examinee does not depend on the set of test items. Invariance only holds when the fit of the model to the data is exact in the population. The goal of item response theory is to provide both invariant item statistics and ability estimates. These features will be obtained when there is a reasonable fit between the chosen model and the data set. Through the estimation process, items and persons are placed on the ability scale in such a way that there is as close a relationship as possible between the expected examinee probability parameters and the actual probabilities of performance for examinees positioned at each ability level. Item parameter estimates and examinee ability estimates are revised continually until maximum agreement possible is obtained between predictions based on the ability and item parameter estimates and the actual test data.

Chapter 2

A Theoretical Framework

2.1 Introduction

Binary responses are extremely common, especially in the social sciences. Individuals can be recorded as agreeing or disagreeing with some position or as getting some item in an educational test right or wrong. Such binary variables are often supposed to be indicators of more fundamental attitudes or abilities and it is in these circumstances that latent variable modelling is relevant. In this chapter we shall consider the theoretical framework within which the latent trait models lie. That is, we will consider models in which binary responses are related to continuous latent variables that reveal some traits of prime interest.

Latent trait models constitute a special case of the General Linear Latent Variable Models and for this reason we will start the representation of the theoretical framework in this chapter by introducing this general class of latent variable models. Then, we will introduce the basic ideas related to latent trait models and represent the theoretical solution to the estimation of a latent trait model. Afterwards, the equality and fixed value constraints of the parameters will be defined in a latent trait model and the constrained latent trait model will be estimated in terms of the unconstrained model. That is, it will be shown how the parameter estimates of the constrained model are related to the parameter estimates of the unconstrained latent trait model. Then, the scoring methods and the most commonly used measures of goodness-of-fit of the models will be described. Finally, the sampling properties of the maximum likelihood estimators will be given.

2.2 General Linear Latent Variable Model

As we have already noted, there are two sorts of variables to be considered in latent variable models, the manifest and the latent variables. The manifest variables are the variables that can be directly observed and in the following analysis they will be denoted by x. A collection of p manifest variables will form a column vector $\mathbf{x}$, that is, $\mathbf{x} = (x_1, x_2, \ldots, x_p)^T$. Latent variables, that is, the variables that cannot be observed directly will be denoted by y and q such variables will form a column vector $\mathbf{y}$. In practice, q will be much smaller than p.

A latent variable model consists of two parts, the prior distribution of the latent variables and the set of conditional distributions of the manifest variables x_i given the latent variables. As far as the prior distribution of the latent variables is concerned, it is given by the density function $h(\mathbf{y})$ (and it has been shown that its choice does not affect the results). The conditional distributions of manifest variables x_i given the latent variables, denoted by $g_i(x_i \mid \mathbf{y})$ $(i = 1, 2, ..., p)$ (the subscript i on g implies that the form of the distribution can vary with i), are supposed to belong to the exponential family of distributions of the form:

$$
\begin{aligned}
g_i(x_i \mid \mathbf{y}) &= \exp\left\{ \frac{x_i\theta_i - b_i(\theta_i)}{a(\varphi_i)} + c_i(x_i, \varphi_i) \right\} \\
\mu_i(\mathbf{y}) &= E(x_i \mid \mathbf{y}) \\
s_i(\mu_i(\mathbf{y})) &= \eta_i \qquad (2.1)
\end{aligned}
$$

where θ_i is the canonical parameter which is some function of $\mathbf{y}$, ϕ_i is the scale parameter, η_i is the linear predictor $\eta_i = \alpha_{i0} + \sum_{j=1}^{q} \alpha_{ij} y_j$ and s_i is the link function. The simplest assumption about the form of this function, under the assumption of the canonical link, is to suppose that it is a linear function, in which case we have

$$
\theta_i = \alpha_{i0} + \alpha_{i1} y_1 + \alpha_{i2} y_2 + ... + \alpha_{iq} y_q \qquad (2.2)
$$

This is the general linear latent variable model. The term "linear", refers to its linearity in the α's. As we have already stated in the introductory chapter, the number of the latent variables q must be chosen such that the assumption of conditional or local independence (1.1) holds.

As only $\mathbf{x}$ can be observed, any inference must be based on the joint distribution whose density may be expressed as

$$
f(\mathbf{x}) = \int_{\Re y} h(\mathbf{y})\, g(\mathbf{x} \mid \mathbf{y})\, \partial \mathbf{y} \qquad (2.3)
$$

where $h(\mathbf{y})$ is the prior distribution of $\mathbf{y}$, $g(\mathbf{x} \mid \mathbf{y})$ is the conditional distribution of $\mathbf{x}$ given $\mathbf{y}$ and $\Re_y$ is the range space of $\mathbf{y}$. Under the assumption

(1.1), equation (2.3) takes the form

$$f(\mathbf{x}) = \int_{\Re y} h(\mathbf{y}) \prod_{i=1}^{p} g_i(x_i \mid \mathbf{y})\,\partial\mathbf{y} \tag{2.4}$$

for some q, h and $\{g_i\}$. Our main interest is in what can be known about $\mathbf{y}$ after $\mathbf{x}$ has been observed. This information is conveyed by the conditional density

$$h(\mathbf{y} \mid \mathbf{x}) = h(\mathbf{y})\, g(\mathbf{x} \mid \mathbf{y}) / f(\mathbf{x}) \tag{2.5}$$

In order to find $h(\mathbf{y} \mid \mathbf{x})$ we need to know both h and g, but all that we can estimate is f. It is obvious that h and g are not uniquely determined by equation (2.3) and thus, at this level of generality, we cannot obtain a complete specification of $h(\mathbf{y} \mid \mathbf{x})$. The posterior distribution (2.5) under the assumption of local independence (1.1) takes the form

$$h(\mathbf{y} \mid \mathbf{x}) = h(\mathbf{y}) \prod_{i=1}^{p} g_i(x_i \mid \mathbf{y}) / f(\mathbf{x}) \tag{2.6}$$

Substituting for $g_i(x_i \mid \mathbf{y})$ as given by (2.2), we find

$$h(\mathbf{y} \mid \mathbf{x}) = \frac{h(\mathbf{y})\, exp\{\sum_{i=1}^{p}\left(\frac{x_i\theta_i - b_i(\theta_i)}{\alpha(\varphi_i)} + c_i(x_i, \varphi_i)\right)\}}{\int_{\Re_y} h(\mathbf{y})\, exp\{\sum_{i=1}^{p}\left(\frac{x_i\theta_i - b_i(\theta_i)}{\alpha(\varphi_i)} + c_i(x_i, \varphi_i)\right)\}\partial\mathbf{y}} \tag{2.7}$$

The marginal distributions of the x_i's for the individual m are:

$$f(\mathbf{x}_m) = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} h(\mathbf{y}) \prod_{i=1}^{p} g_i(x_{mi} \mid \mathbf{y})\,\partial y_1 \ldots \partial y_q \tag{2.8}$$

For a random sample of size n the log-likelihood by means of (2.8) is written as:

$$L = \sum_{m=1}^{n} \log f(\mathbf{x}_m) = \sum_{m=1}^{n} log \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} h(\mathbf{y})\, g(\mathbf{x}_m \mid \mathbf{y})\,\partial y_1 \ldots \partial y_q \tag{2.9}$$

$$= \sum_{m=1}^{n} \log \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} h(\mathbf{y}) \prod_{i=1}^{p} g_i(x_{mi} \mid \mathbf{y})\, \partial y_1 \ldots \partial y_q \tag{2.10}$$

2.3 Latent Trait Model for Binary Responses

2.3.1 General Definition

Suppose that x_i is a Bernoulli random variable arising from putting a proposition to people and inviting them to agree or disagree. If the answers are coded 1 (agree) or 0 (disagree) and the associations among the x's are induced by a set of q latent variables, $\mathbf{y}$, the conditional distribution of x_i given the latent variables for the individual m may be written as

$$g(x_{im} \mid \mathbf{y}) = \pi_i^{x_{mi}} (1-\pi_i)^{1-x_{mi}} = e^{\{x_{im} \log\left(\frac{\pi_i}{1-\pi_i}\right) + \log(1-\pi_i)\}} \tag{2.11}$$

where $\pi_i = P(x_i = 1 \mid \mathbf{y})$ is the probability of agreeing to item x_i, $i = 1, 2, \ldots, p$. Comparing (2.11) with (2.2) we see that $\theta_i = \text{logit}(\pi_i) = \log\{\frac{\pi_i}{(1-\pi_i)}\}$, $b(\theta_i) = \log(1 + exp(\theta_i))$, $\alpha(\varphi_i) = 1$ and $c_i(x_i, \varphi_i) = 0$. We assumé that $g_i(x_{mi} \mid \mathbf{y})$ for all the items must be of the same type, that is, of the Bernoulli distribution. The latent variables, y_j, are taken to be standard normal and $h(\mathbf{y})$ denotes the multivariate standard normal density, with correlation matrix $\mathbf{I}$.

The general linear latent variable model (2.3) takes the form:

$$\text{logit}(\pi_i) = \alpha_{i0} + \sum_{j=1}^{q} \alpha_{ij} y_j \quad (i = 1, 2, \ldots, p) \tag{2.12}$$

If the dependence among the x's is wholly explained by a vector $\mathbf{y}$ of latent variables they may be regarded as mutually independent random variables with

$$P\{x_i = 1 \mid \mathbf{y}\} = \pi_i(\mathbf{y}), \; x_i = 0, 1, \; i = 1, \ldots, p$$

where $\pi_i(\mathbf{y})$ is called the response function and since it is a probability, $0 \leq \pi_i(\mathbf{y}) \leq 1$. In test theory, where $\mathbf{y}$ is usually a scalar, $\pi_i(\mathbf{y})$ is known as the item response function or the characteristic curve. In that context y represents an ability of some kind, in which case one would expect $\pi_i(\mathbf{y})$ to be a monotonic function.

The parameter α_{i0} is sometimes called the 'intercept' because of its role in the linear plot of logit (π_i) against $\mathbf{y}$. An alternative parameterization uses

$$\pi_i(0) = \pi_i = \frac{1}{1 + e^{-\alpha_{i0}}} \tag{2.13}$$

This is the probability of a positive response from an individual with $y = 0$, that is, for someone at the median point of the latent scale. The parameter α_{i1} governs the steepness of the curve; in educational testing it is known as the discrimination parameter because the bigger α_{i1} the easier it will be to discriminate between a pair of individuals a given distance apart on the latent scale. This is because the greater the difference between the value of $\pi_i(\mathbf{y})$ for the two individuals, the more likely it is that they will give a different response.

2.3.2 Estimation

The parameters of the latent trait model are estimated by means of the maximum likelihood based on the joint distribution of the manifest variables. Under the assumption of local independence the joint distribution of the manifest variables is given by (2.8). For a random sample of size n the log-likelihood is written as:

$$
\begin{aligned}
L &= \sum_{m=1}^{n} \log f(\mathbf{x}_m) = \sum_{m=1}^{n} \log \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} h(\mathbf{y})\, g(\mathbf{x}_m \mid \mathbf{y})\, \partial y_1 \ldots \partial y_q \\
&= \sum_{m=1}^{n} \log \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} h(\mathbf{y}) \prod_{i=1}^{p} g_i(x_{mi} \mid \mathbf{y})\, \partial y_1 \ldots \partial y_q \\
&= \sum_{m=1}^{n} \log \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} h(\mathbf{y}) \prod_{i=1}^{p} exp\{x_{mi}\theta_i - b_i(\theta_i)\} \partial y_1 \ldots \partial y_q \qquad (2.14)
\end{aligned}
$$

The unknown parameters are in θ_i. We differentiate the log-likelihood given in (2.14) with respect to the model parameters α_{ij}, $(i = 1, 2, \ldots, p)$, $(j = 0, 1, \ldots, q)$. Finding partial derivatives, we have

$$
\begin{aligned}
\frac{\partial L}{\partial \alpha_{ij}} &= \frac{\partial}{\partial \alpha_{ij}} \sum_{m=1}^{n} \log f(\mathbf{x}_m) = \sum_{m=1}^{n} \frac{1}{f(\mathbf{x}_m)} \frac{\partial}{\partial \alpha_{ij}} f(\mathbf{x}_m) \\
&= \sum_{m=1}^{n} \frac{1}{f(\mathbf{x}_m)} \frac{\partial}{\partial \alpha_{ij}} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} h(\mathbf{y}) \prod_{i=1}^{p} g(x_{mi} \mid \mathbf{y})\, \partial y_1 \ldots \partial y_q \\
&= \sum_{m=1}^{n} \frac{1}{f(\mathbf{x}_m)} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} h(\mathbf{y}) \frac{\partial}{\partial \alpha_{ij}} \prod_{i=1}^{p} g(x_{mi} \mid \mathbf{y})\, \partial y_1 \ldots \partial y_q \\
&= \sum_{m=1}^{n} \frac{1}{f(\mathbf{x}_m)} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} h(\mathbf{y}) \frac{\partial}{\partial \alpha_{ij}} \prod_{i=1}^{p} e^{\{x_{mi}\theta_i - b_i(\theta_i)\}} \partial y_1 \ldots \partial y_q \qquad (2.15)
\end{aligned}
$$

The last equation in (2.15) is due to (2.11) and the integral can be approximated by means of a quadrature method common in statistical applications, that is, the Gauss-Hermite quadrature method. This method will be presented in Section (2.3.3). According to the Gauss-Hermite Quadrature method, the latent variables are treated as discrete, with the values y_1, y_2, ..., y_k, that is, the abscissas, having prior probabilities $h(y_1)$, $h(y_2)$, ...,

$h(y_k)$. Thus, we get

$$\frac{\partial L}{\partial \alpha_{ij}} = \sum_{m=1}^{n} \frac{1}{f(\mathbf{x}_m)} \sum_{t_1=1}^{k} \cdots \sum_{t_q=1}^{k} h(\mathbf{y}_t) \frac{\partial}{\partial \alpha_{ij}} \left[\prod_{i=1}^{p} e^{\{x_{mi}\theta_i - b_i(\theta_i)\}} \right] \tag{2.16}$$

Then we will calculate the term $\frac{\partial}{\partial \alpha_{ij}} \left[\prod_{i=1}^{p} e^{\{x_{mi}\theta_i - b_i(\theta_i)\}} \right]$ separately and substitute it in (2.16):

$$\begin{aligned}
\frac{\partial}{\partial \alpha_{ij}} \left[\prod_{i=1}^{p} e^{\{x_{mi}\theta_i - b_i(\theta_i)\}} \right] &= \left[\prod_{i=1}^{p} e^{\{x_{mi}\theta_i - b_i(\theta_i)\}} \right] \frac{\partial}{\partial \alpha_{ij}} \{x_{mi}\theta_i - b_i(\theta_i)\} \\
&= \left[\prod_{i=1}^{p} e^{\{x_{mi}\theta_i - b_i(\theta_i)\}} \right] \frac{\partial}{\partial \theta_i} \{x_{mi}\theta_i - b_i(\theta_i)\} \frac{\partial \theta_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \alpha_{ij}} \\
&= g(\mathbf{x}_m \mid \mathbf{y}_t) \cdot \{x_{mi} - b_i'(\theta_i)\} \cdot \frac{\partial \theta_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \alpha_{ij}} \qquad (2.17)
\end{aligned}$$

Thus, by substituting (2.17) in (2.16) we get:

$$\begin{aligned}
\frac{\partial L}{\partial \alpha_{ij}} &= \sum_{t_1=1}^{k} \cdots \sum_{t_q=1}^{k} h(\mathbf{y}_t) \left[\sum_{m=1}^{n} x_{mi} \frac{g(\mathbf{x}_m \mid \mathbf{y}_t)}{f(\mathbf{x}_m)} - \sum_{m=1}^{n} \frac{g(\mathbf{x}_m \mid \mathbf{y}_t)}{f(\mathbf{x}_m)} b_i'(\theta_i) \right] \frac{\partial \theta_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \alpha_{ij}} \\
&= \sum_{t_1=1}^{k} \cdots \sum_{t_q=1}^{k} \left[r_{it} - N_t b_i'(\theta_i) \right] \frac{\partial \theta_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \alpha_{ij}} \qquad (2.18)
\end{aligned}$$

where

$$r_{it} = h(\mathbf{y}_t) \cdot \sum_{m=1}^{n} x_{mi} \frac{g(\mathbf{x}_m \mid \mathbf{y}_t)}{f(\mathbf{x}_m)} = \sum_{m=1}^{n} x_{mi} h(\mathbf{y}_t \mid \mathbf{x}_m) \tag{2.19}$$

$$N_t = h(\mathbf{y}_t) \cdot \sum_{m=1}^{n} \frac{g(\mathbf{x}_m \mid \mathbf{y}_t)}{f(\mathbf{x}_m)} = \sum_{m=1}^{n} h(\mathbf{y}_t \mid \mathbf{x}_m) \tag{2.20}$$

$$\frac{\partial \theta_i}{\partial \mu_i} = \left[\frac{\partial \mu_i}{\partial \theta_i} \right]^{-1} = \frac{1}{b''(\theta_i)}, \mu_i = b'(\theta_i)$$

$$\frac{\partial \mu_i}{\partial \eta_i} = \left[\frac{\partial \eta_i}{\partial \mu_i} \right]^{-1} = \frac{1}{s'(\mu_i)}, \eta_i = s(\mu_i)$$

$$\frac{\partial \eta_i}{\partial a_{ij}} = \frac{\partial}{\partial \alpha_{ij}} \left(\alpha_{i0} + \sum_{j=1}^{q} \alpha_{ij} y_j \right)$$

In addition, due to model (2.2) we have that

$$\theta_i = \log \frac{\pi_i}{1-\pi_i}$$

$$b(\theta_i) = \log \left(1 + exp\left(\theta_i\right)\right)$$

$$E(x_i) = b'(\theta_i) = exp\left(\frac{exp(\theta_i)}{1+exp(\theta_i)}\right) = \pi_i = \mu_i$$

$$Var(x_i) = b''(\theta_i) = \frac{exp(\theta_i)}{\left(1+exp(\theta_i)\right)^2} = \pi_i \cdot (1-\pi_i)$$

$$V(\mu) = \mu \cdot (1-\mu)$$

$$s'(\mu_i) = \left(\log\left(\frac{\mu_i}{1-\mu_i}\right)\right)' = \frac{1}{\mu_i \cdot (1-\mu_i)}$$

Thus the partial derivatives are given below:

$$\begin{aligned} \frac{\partial L}{\partial \alpha_{ij}} &= \sum_{t_1=1}^{k} \cdots \sum_{t_q=1}^{k} \left[r_{it} - N_t \pi_i\right] \cdot \frac{\partial}{\partial \alpha_{ij}} \left(\alpha_{i0} + \sum_{j=1}^{q} \alpha_{ij} y_j \right) \\ &= \sum_{t_1=1}^{k} \cdots \sum_{t_q=1}^{k} \left[r_{it} - N_t \pi_i\right] \cdot y_j^l \qquad (2.21) \end{aligned}$$

where

- $l = 0$ when we are estimating the intercept of the item i
- $l = 1$ when we are estimating the coefficient of the latent variable j

According to (2.21) the vector of the partial derivatives $\mathbf{\Pi}$ for the latent trait model is:

$$\mathbf{\Pi} = \left[\begin{array}{c} \sum\limits_{t_1=1}^{k} \cdots \sum\limits_{t_q=1}^{k} [r_{1t} - N_t \cdot \pi_1] \\ \sum\limits_{t_1=1}^{k} \cdots \sum\limits_{t_q=1}^{k} [r_{1t} - N_t \cdot \pi_1] \cdot y_1 \\ \cdots \\ \sum\limits_{t_1=1}^{k} \cdots \sum\limits_{t_q=1}^{k} [r_{1t} - N_t \cdot \pi_1] \cdot y_q \\ \cdots \\ \sum\limits_{t_1=1}^{k} \cdots \sum\limits_{t_q=1}^{k} [r_{pt} - N_t \cdot \pi_p] \\ \sum\limits_{t_1=1}^{k} \cdots \sum\limits_{t_q=1}^{k} [r_{pt} - N_t \cdot \pi_p] \cdot y_1 \\ \cdots \\ \sum\limits_{t_1=1}^{k} \cdots \sum\limits_{t_q=1}^{k} [r_{pt} - N_t \cdot \pi_p] \cdot y_q \end{array}\right] \begin{array}{l} \alpha_{10} \\ \alpha_{11} \\ \\ \\ \\ \\ \\ \\ \end{array} \qquad (2.22)$$

The above $\mathbf{\Pi}$ vector is a $[p \cdot (q+1)] \times 1$ vector in which each element is equal to $\sum\limits_{t_1=1}^{k} \cdots \sum\limits_{t_q=1}^{k} [r_{it} - N_t \cdot \pi_i] \cdot y_j^l$, where $l = 0, 1$, $i = 1, 2, \ldots, p$, $j = 1, 2, \ldots, q$.

The quantity $h(\mathbf{y}_t \mid \mathbf{x}_m)$ is the probability that an individual with response vector $\mathbf{x}_m$ is located at y_t; $N_t = \sum\limits_{m=1}^{n} h(\mathbf{y}_t \mid \mathbf{x}_m)$ is thus the expected number of individuals at y_t. By a similar argument r_{it} is the expected number of those predicted to be at y_t who will respond positively.

The unknown model parameters $\boldsymbol{\alpha}$ are estimated by solving the non-linear equations

2.3.3 Quadrature Methods

This subsection is concerned with a dominant method of modern quadrature in many statistical applications, Gauss - Hermite quadrature. Quadrature, in general, refers to the numerical integration of a function that does not have closed form antiderivative. The goal is to attain a given level of precision with the fewest possible function evaluations. The crucial factors that control the difficulty of this problem are the dimensionality, and the smoothness of the function. Any method for numerically approximating $\int_{\Re} f(\mathbf{x})\,\partial\mathbf{x}$ relies on evaluating f on a finite set of points, called the abscissas or quadrature points, and then processing these evaluations somehow to produce an approximation to the integral. Usually the processing involves some form of weighted average.

Gauss-Hermite Quadrature. Gaussian quadrature is ideal for integration against standard probability densities such as the normal or gamma. This method works well for good integrands, such as low-degree polynomials. In addition, it has the merit of applying quadrature rules to points located in the main mass of the integrand, an aspect of prime importance. The basic form of a Gaussian quadrature rule is

$$\int_{\alpha}^{b} W(\mathbf{x})\, f(\mathbf{x})\, \partial\mathbf{x} = \sum_{j=1}^{n} w_j f(x_j)$$

where both the weights w_j and the abscissas x_j are determined to give an accurate approximation. Besides, Gaussian quadrature also has the virtue of handling infinite domains of integration gracefully (Korner, 1988, Powell, 1981, Press et. al., 1992) by means of the Gauss - Hermite quadrature. As far as the Gauss - Hermite quadrature is concerned, the weight function is proportional to a normal density. Thus a doubly infinite integral is put into the form of the Gauss-Hermite integral, through

$$\int_{-\infty}^{+\infty} g(x)\,\partial x = \int_{-\infty}^{+\infty} \exp\left(-x^2\right) f(x)\,\partial x$$

where $f(x) = g(x) \cdot exp(x^2)$. However, if $g(x)$ is concentrated about a point far from 0, or if the spread in $g(x)$ is quite different from that of the weight function $exp(-x^2)$, then applying Gauss-Hermite quadrature directly can give a very poor approximation, because the abscissas in the quadrature rule will not be located where most of the mass of $g(x)$ is located and thus a transformation is necessary.

For multi-dimensional integrals, as in the univariate case, Gauss-Hermite quadrature is most useful when the integrand is fairly concentrated about a single mode. It is again important to make a change of variables to center the quadrature rule near the mode. Near the mode this can be done by first locating the mode of the integrand, and then calculating (or approximating) the second derivative matrix of the log of the integrand at the mode. The transformation is then based on the mode and the inverse of the second derivative matrix. In general, suppose the integral is of the form

$$\int q(\boldsymbol{\theta})\, exp\left[h(\boldsymbol{\theta})\right] \partial\boldsymbol{\theta} \tag{2.23}$$

where $\boldsymbol{\theta}$ is p-dimensional and the region of integration is essentially all of $\Re^p$. Let $\hat{\boldsymbol{\theta}}$ be the mode of h, let $H = -\partial^2 h\left(\hat{\boldsymbol{\theta}}\right)/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'$, and let $B \cdot B' = H$ be the Cholesky factorization of H. If $\boldsymbol{\theta}$ is thought of as a random quantity with log-density given by $h(\boldsymbol{\theta})$, then the first order normal approximation to this distribution is the $N\left(\hat{\boldsymbol{\theta}}, H^{-1}\right)$ distribution. To use a Gauss-Hermite product rule, a function proportional to a $N\left(0, 2^{-1}I\right)$ density would be preferred, since the Gauss-Hermite weight function is proportional to this density. Setting $\alpha = 2^{-1/2}B\left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\right)$ will transform $\boldsymbol{\theta}$ to a variable that does have an approximate $N\left(0, 2^{-1}I\right)$ distribution. With this transformation (2.23) becomes

$$2^{p/2}\left|B\right|^{-1}\int \exp\left(-\boldsymbol{\alpha}'\boldsymbol{\alpha}\right) f\left(\boldsymbol{\alpha}\right)\partial\boldsymbol{\alpha} \tag{2.24}$$

where

$$f\left(\boldsymbol{\alpha}\right) = q\cdot\left(2^{1/2}B^{-1}\boldsymbol{\alpha} + \hat{\boldsymbol{\theta}}\right)\cdot \exp\left(h\left(2^{1/2}B^{-1}\boldsymbol{\alpha} + \hat{\boldsymbol{\theta}}\right) + \boldsymbol{\alpha}'\boldsymbol{\alpha}\right)$$

The advantage is that to first order, $exp\left(h\left(2^{1/2}B^{-1}\boldsymbol{\alpha} + \hat{\boldsymbol{\theta}}\right) + \boldsymbol{\alpha}'\boldsymbol{\alpha}\right)$ should be constant. (2.24) can be approximated using the n-point Gauss-Hermite product rule with abscissas $x_1, \ldots, x_n$ and weights $w_1, \ldots, w_n$ giving

$$2^{p/2}\left|B\right|^{-1}\sum_{i_1=1}^{n}\cdots\sum_{i_p=1}^{n} w_{i1}\ldots w_{ip} f\left(x_{i1}, \ldots, x_{ip}\right)$$

2.4 Equality and Fixed value Constraints in Latent Trait Models

In Section 1.2 we introduced the concept of parameter constraints in terms of the different types of constraints that have been developed in literature and the representation of situations in which such constraints are necessary. In this section we will present the way in which the latent trait model is modified and estimated under the imposition of parameter constraints. The parameter constraints that are considered in this dissertation are only the equality and fixed value constraints. At first we will introduce the equality and the fixed value constraints in the latent trait models and then represent the estimation procedure.

2.4.1 Specification of the equality and fixed value constraints

This subsection deals with the specification of the equality constraints in the latent trait model. In the latent trait model (2.12) we may be interested in setting the factor loadings within an item or between items to be equal or to equal to a fixed value. Let us suppose the following form of the latent trait model under study.

$$\begin{aligned} \operatorname{logit}(\pi_1) &= \alpha_{10} + \sum_{j=1}^{q} \alpha_{1j} y_j \\ &\dots \qquad\qquad j = 1, \dots, q \\ \operatorname{logit}(\pi_p) &= \alpha_{p0} + \sum_{j=1}^{q} \alpha_{pj} y_j \end{aligned}$$

In the above model we may be interested in checking whether the model that adopts one or more sets of equality or fixed value constraints fits adequately the data. That is, we may be interested in testing simultaneously assumptions of the form:

$$\alpha_{51} = \alpha_{52} \text{ (equality constraints within item 5)}$$

or/and

$$\alpha_{11} = \alpha_{72} = \alpha_{61} \text{ (equality constraints between items } 1, 7, 6)$$

or/and

$$\alpha_{12} = \alpha_{13} = \alpha_{62} \text{ (equality constraints within item 1 and between items 1, 6)}$$

or/and

$$\alpha_{31} = \alpha_{42} = 0 \text{ (fixed value constraints)}$$

This implies that the number of the unknown parameters is restricted under the various constraints. The unknown parameters in a latent trait model, where no constraints are imposed, include the p intercepts and the $p \times q$ loadings (the coefficients of the latent variables), that is, there are $p \times q + p$ unknown parameters. If we impose c sets of equality constraints only that each of these sets contains x parameters, then the number of the different parameters in the latent trait model that need to be estimated is $(p \times q + p) - \sum_{i=1}^{c} (x_i - 1)$. If we impose c sets of equality constraints that each of these sets contains x parameters, and d sets of fixed value constraints that each of these sets contains y parameters, then the number of the different parameters in the latent trait model that need to be estimated is $(p \times q + p) - \sum_{i=1}^{c} (x_i - 1) - \sum_{i=1}^{d} y_i$.

2.4.2 Estimation Procedure

The estimation procedure that will be presented below is aimed to estimate the different unknown parameters α_{ij}, by means of the maximum likelihood method. In cases where fixed value parameter constraints are imposed, the estimation procedure accounts for such constraints by means of the optimization algorithms that are presented in Chapter 3. The definitions of the conditional distribution of the manifest variables given the latent variables, the joint distribution of the manifest variables, the posterior distribution of the latent variables given the manifest variables and the log-likelihood function under the imposition of constraints are the same with those given for the latent trait model where no parameter constraints have been imposed. The differences lie in the specification of the partial derivatives that are necessary for the estimation procedure. As we have proved earlier the partial derivatives when there are not any constraints in the factor loadings are given below:

$$\frac{\partial L}{\partial \alpha_{ij}} = \sum_{t_1=1}^{k} \cdots \sum_{t_q=1}^{k} \left[r_{it} - N_t \pi_i \right] \cdot y_j^l$$

When constraints are imposed in the parameters in the latent trait model, we have to make some additional remarks in (2.21) of the partial derivatives. Thus (2.17) under the parameter constraints takes the form:

$$\begin{aligned} \frac{\partial}{\partial \alpha_{ij}} \left[\prod_{i=1}^{p} e^{(x_i \theta_i - b_i(\theta_i))} \right] &= \left[\prod_{i=1}^{p} e^{(x_i \theta_i - b_i(\theta_i))} \right] \frac{\partial}{\partial \alpha_{ij}} \sum_{s=1}^{\{d \epsilon i\}} \left(x_s \theta_s - b_s \left(\theta_s \right) \right) \\ &= g\left(\mathbf{x}_m \mid \mathbf{y}_t \right) \sum_{s=1}^{d \epsilon i} \frac{\partial}{\partial \theta_s} \left(x_s \theta_s - b_s \left(\theta_s \right) \right) \frac{\partial \theta_s}{\partial \mu_s} \frac{\partial \mu_s}{\partial \eta_s} \frac{\partial \eta_s}{\partial \alpha_{ij}} \\ &= g\left(\mathbf{x}_m \mid \mathbf{y}_t \right) \sum_{s=1}^{d \epsilon i} \left(x_s \theta_s - b_s' \left(\theta_s \right) \right) \frac{\partial \theta_s}{\partial \mu_s} \frac{\partial \mu_s}{\partial \eta_s} \frac{\partial \eta_s}{\partial \alpha_{ij}} \end{aligned} \quad (2.25)$$

The index s indicates the items in which the parameter to be estimated, α_{ij}, is located. The index s takes the discrete values d that correspond to the items $i = 1, 2, \ldots, p$ to which the parameter α_{ij} can be traced. For instance, for the set of parameter constraints $\alpha_{11} = \alpha_{52} = \alpha_{61}$, the parameter α_{11} that is the loading of the first factor in the first item is set equal to the loading of the second factor in the fifth item and the loading of the first factor in the sixth item. Thus, the index d in the partial derivative (2.25) with respect to the parameter α_{11} takes the values $1, 5, 6$, that is, the partial derivative with respect to α_{11} consists of three sums. The same happens in the partial derivatives with respect to the parameters α_{52} and α_{61}.

Thus, by substituting (2.25) in (2.18) we get:

$$\begin{aligned}\frac{\partial L}{\partial \alpha_{ij}} &= \sum_{t_1=1}^{k} \cdots \sum_{t_q=1}^{k} h(\mathbf{y}_t) \sum_{s=1}^{d\epsilon i} \left[\sum_{m=1}^{n} x_{sm} \frac{g(\mathbf{x}_m \mid \mathbf{y}_t)}{f(\mathbf{x}_m)} - \sum_{m=1}^{n} \frac{g(\mathbf{x}_m \mid \mathbf{y}_t)}{f(\mathbf{x}_m)} b_s'(\theta_s) \right] \frac{\partial \theta_s}{\partial \mu_s} \frac{\partial \mu_s}{\partial \eta_s} \frac{\partial \eta_s}{\partial \alpha_{ij}} \\ &= \sum_{t_1=1}^{k} \cdots \sum_{t_q=1}^{k} \sum_{s=1}^{d\epsilon i} \left[r_{st} - N_t b_s'(\theta_s) \right] \frac{\partial \theta_s}{\partial \mu_s} \frac{\partial \mu_s}{\partial \eta_s} \frac{\partial \eta_s}{\partial \alpha_{ij}} \end{aligned} \qquad (2.26)$$

where

$$r_{st} = h(\mathbf{y}_t) \cdot \sum_{m=1}^{n} x_{sm} \frac{g(\mathbf{x}_m \mid \mathbf{y}_t)}{f(\mathbf{x}_m)} = \sum_{m=1}^{n} x_{sm} h(\mathbf{y}_t \mid \mathbf{x}_m)$$

$$N_t = h(\mathbf{y}_t) \cdot \sum_{m=1}^{n} x_{sm} \frac{g(\mathbf{x}_m \mid \mathbf{y}_t)}{f(\mathbf{x}_m)} = \sum_{m=1}^{n} h(\mathbf{y}_t \mid \mathbf{x}_m)$$

$$\frac{\partial \theta_s}{\partial \mu_s} = \left[\frac{\partial \mu_s}{\partial \theta_s} \right]^{-1} = \frac{1}{b''(\theta_s)}, \mu_s = b'(\theta_s)$$

$$\frac{\partial \mu_s}{\partial \eta_{si}} = \left[\frac{\partial \eta_s}{\partial \mu_s} \right]^{-1} = \frac{1}{s'(\mu_s)}, \eta_s = s(\mu_s)$$

$$\frac{\partial \eta_s}{\partial a_{ij}} = \frac{\partial}{\partial \alpha_{ij}} \left(\alpha_{s0} + \sum_{j=1}^{q} \alpha_{sj} y_j \right)$$

Thus the partial derivatives of the latent trait model under the constraints are given below:

$$\frac{\partial L}{\partial \alpha_{ij}} = \sum_{t_1=1}^{k} \cdots \sum_{t_q=1}^{k} \sum_{s=1}^{d\epsilon i} \left[r_{st} - N_t b_s'(\theta_s) \right] \cdot \left(\sum_{j=0}^{q} y_j \right) \qquad (2.27)$$

where $y_0 = 1$. The above form of the partial derivatives implies that:

- If we have imposed constraints within the item i then the partial derivative takes the form $\frac{\partial L}{\partial \alpha_{ij}} = \sum_{t_1=1}^{k} \cdots \sum_{t_q=1}^{k} \left[r_{it} - N_t b_i'(\theta_i) \right] \cdot \left(\sum_{j=0}^{q} y_j \right)$

- If we have imposed constraints within item i and between the items, then the partial derivative takes the form $\frac{\partial L}{\partial \alpha_{ij}} = \sum_{t_1=1}^{k} \ldots \sum_{t_q=1}^{k} \sum_{s=1}^{d\epsilon i} \left[r_{st} - N_t b_s' (\theta_s) \right] \cdot \left(\sum_{j=0}^{q} y_j \right)$

Thus, the vector of the partial derivatives for the "constrained" latent trait model $\mathbf{\Pi}_C$ can be expressed in terms of the vector of the partial derivatives of the "unconstrained" model $\mathbf{\Pi}$, given by (2.22). This can be done by means of another indicator matrix $\mathbf{A}$. The matrix $\mathbf{A}$ is called indicator because it indicates the location of the equality parameter constraints in the latent trait model. In fact, the matrix $\mathbf{A}$ is a $[p \cdot (q+1)] \times [p \cdot (q+1)]$ matrix with elements 0 and 1. The 1s are used to indicate the parameters α_{ij} and which of them appear to be equal to which parameters of the p items and q latent variables. In order to make clear how this matrix is constructed we will introduce a simple and "meaningless" example in which we assume that the correlations between three items x_1, x_2, x_3 with $x_i = 0, 1$ can be accounted for by two independent metrical latent variables. Thus, the latent trait model is of the form:

$$\begin{aligned} \text{logit}\,(\pi_1) &= \alpha_{10} + \alpha_{11} y_1 + \alpha_{12} y_2 \\ \text{logit}\,(\pi_2) &= \alpha_{20} + \alpha_{21} y_1 + \alpha_{22} y_2 \\ \text{logit}\,(\pi_3) &= \alpha_{30} + \alpha_{31} y_1 + \alpha_{32} y_2 \end{aligned} \tag{2.28}$$

Let us assume that in the above model we are interested in checking whether the model that adopts two sets of equality constraints fits adequately the data. That is, we are interested in testing simultaneously the following assumptions:

$$\alpha_{11} = \alpha_{12} = \alpha_{21} \tag{2.29}$$

$$\alpha_{22} = \alpha_{32}$$

Firstly, we will represent the matrix **A** in which the 0s and 1s denote the positions of the parameters' constraints. Thus, the matrix **A** of the above model is:

$$\mathbf{A} = \left[\begin{array}{cccccccccc} & \alpha_{10} & \alpha_{11} & \alpha_{12} & \alpha_{20} & \alpha_{21} & \alpha_{22} & \alpha_{30} & \alpha_{31} & \alpha_{32} \\ \alpha_{10} & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \alpha_{11} & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ \alpha_{12} & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ \alpha_{20} & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ \alpha_{21} & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ \alpha_{22} & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ \alpha_{30} & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ \alpha_{31} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ \alpha_{32} & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \end{array}\right]$$

The matrix **A** is a 9×9 matrix, that is, it contains as many elements as are the parameters in the latent trait model under study. In front of the first column and the first row of the matrix we have written the parameters to be estimated. The logic by which we complete this matrix is the following. The first row corresponds to the parameter α_{10}. Each 1 element of this row will indicate to which parameters α_{10} equals and each 0 element of this row will indicate to which parameters α_{10} does not equal. Thus, provided that the parameter α_{10} does not equal to any of the other parameters of the model we complete only the first element of the first row by 1, indicating that the parameter α_{10} equals to itself. The second row corresponds to the parameter α_{11}. According to the constraints (2.30), the parameter α_{11} equals to the parameters α_{12} and α_{21}. Thus, we will complete the second, the third and the fifth element of the second row (these elements correspond to the parameters α_{11}, α_{12} and α_{21}) by 1. The same procedure is followed for all the other parameters. Indeed, the elements of the main diagonal will always be 1.

If we consider model (2.28) without the imposition of constraints the vector of the partial derivatives (2.22) takes the form:

$$\Pi = \begin{bmatrix} \sum\limits_{t_1=1}^{k} \cdots \sum\limits_{t_q=1}^{k} [r_{1t} - N_t \cdot \pi_1] \\ \sum\limits_{t_1=1}^{k} \cdots \sum\limits_{t_q=1}^{k} [r_{1t} - N_t \cdot \pi_1] \cdot y_1 \\ \sum\limits_{t_1=1}^{k} \cdots \sum\limits_{t_q=1}^{k} [r_{1t} - N_t \cdot \pi_1] \cdot y_2 \\ \sum\limits_{t_1=1}^{k} \cdots \sum\limits_{t_q=1}^{k} [r_{2t} - N_t \cdot \pi_2] \\ \sum\limits_{t_1=1}^{k} \cdots \sum\limits_{t_q=1}^{k} [r_{2t} - N_t \cdot \pi_2] \cdot y_1 \\ \sum\limits_{t_1=1}^{k} \cdots \sum\limits_{t_q=1}^{k} [r_{2t} - N_t \cdot \pi_2] \cdot y_2 \\ \sum\limits_{t_1=1}^{k} \cdots \sum\limits_{t_q=1}^{k} [r_{3t} - N_t \cdot \pi_3] \\ \sum\limits_{t_1=1}^{k} \cdots \sum\limits_{t_q=1}^{k} [r_{3t} - N_t \cdot \pi_3] \cdot y_1 \\ \sum\limits_{t_1=1}^{k} \cdots \sum\limits_{t_q=1}^{k} [r_{3t} - N_t \cdot \pi_3] \cdot y_2 \end{bmatrix}$$

However, under the imposition of equality parameter constraints the vector of the partial derivatives Π_C for model (2.28) is of the form:

$$\Pi_C = \frac{\partial L}{\partial \alpha_{ij}} = \begin{bmatrix} \frac{\partial L}{\partial \alpha_{10}} & \frac{\partial L}{\partial \alpha_{11}} & \frac{\partial L}{\partial \alpha_{12}} & \frac{\partial L}{\partial \alpha_{20}} & \frac{\partial L}{\partial \alpha_{21}} & \frac{\partial L}{\partial \alpha_{22}} & \frac{\partial L}{\partial \alpha_{30}} & \frac{\partial L}{\partial \alpha_{31}} & \frac{\partial L}{\partial \alpha_{32}} \end{bmatrix}^T =$$

$$
= \begin{bmatrix}
\sum_{t_1=1}^{k} \cdots \sum_{t_q=1}^{k} [r_{1t} - N_t \cdot \pi_1] \\
\sum_{t_1=1}^{k} \cdots \sum_{t_q=1}^{k} [r_{1t} - N_t \cdot \pi_1] \cdot (y_1 + y_2) + \sum_{t_1=1}^{k} \cdots \sum_{t_q=1}^{k} [r_{2t} - N_t \cdot \pi_2] \cdot y_1 \\
\sum_{t_1=1}^{k} \cdots \sum_{t_q=1}^{k} [r_{1t} - N_t \cdot \pi_1] \cdot (y_1 + y_2) + \sum_{t_1=1}^{k} \cdots \sum_{t_q=1}^{k} [r_{2t} - N_t \cdot \pi_2] \cdot y_1 \\
\sum_{t_1=1}^{k} \cdots \sum_{t_q=1}^{k} [r_{2t} - N_t \cdot \pi_2] \\
\sum_{t_1=1}^{k} \cdots \sum_{t_q=1}^{k} [r_{1t} - N_t \cdot \pi_1] \cdot (y_1 + y_2) + \sum_{t_1=1}^{k} \cdots \sum_{t_q=1}^{k} [r_{2t} - N_t \cdot \pi_2] \cdot y_1 \\
\sum_{t_1=1}^{k} \cdots \sum_{t_q=1}^{k} [r_{2t} - N_t \cdot \pi_2] \cdot y_2 + \sum_{t_1=1}^{k} \cdots \sum_{t_q=1}^{k} [r_{3t} - N_t \cdot \pi_3] \cdot y_2 \\
\sum_{t_1=1}^{k} \cdots \sum_{t_q=1}^{k} [r_{3t} - N_t \cdot \pi_3] \\
\sum_{t_1=1}^{k} \cdots \sum_{t_q=1}^{k} [r_{3t} - N_t \cdot \pi_3] \cdot y_1 \\
\sum_{t_1=1}^{k} \cdots \sum_{t_q=1}^{k} [r_{2t} - N_t \cdot \pi_2] \cdot y_2 + \sum_{t_1=1}^{k} \cdots \sum_{t_q=1}^{k} [r_{3t} - N_t \cdot \pi_3] \cdot y_2
\end{bmatrix}
$$

We observe that, the vector of the partial derivatives can be written by means of the product of two previously defined matrices: $\mathbf{A}$, $\mathbf{\Pi}$. That is,

$$\mathbf{\Pi}_C = \mathbf{\Pi} \cdot \mathbf{A}.$$

The unknown model parameters $\boldsymbol{\alpha}$ are estimated by solving the non-linear equations $\frac{\partial L}{\partial \alpha_{ij}} = 0$. Since there is not closed form solution, we can use the Fisher Scoring algorithm. Thus we will present the Fisher Information matrix for the unknown parameters, that is, the expected value of the minus second derivative of the likelihood function at the ML point.

We have already proved that the partial derivatives of the latent trait model under constraints are given below:

$$\frac{\partial L}{\partial \alpha_{ij}} = \sum_{t_1=1}^{k} \cdots \sum_{t_q=1}^{k} \sum_{s=1}^{dei} \left[r_{st} - N_t \cdot b_s^{'}(\theta_s) \right] \cdot \left(\sum_{j=0}^{q} y_j \right)$$

Each element of the Fisher Information matrix is given as follows:

$$
\begin{aligned}
E\left(-\frac{\partial^2 L}{\partial\alpha_{ij}\partial\alpha_{vw}}\right) &= E\left(-\frac{\partial}{\partial\alpha_{vw}}\frac{\partial L}{\partial\alpha_{ij}}\sum_{t_1=1}^{k}\cdots\sum_{t_q=1}^{k}\sum_{s=1}^{d\epsilon i}\left[r_{st}-N_t\cdot b_s^{'}(\theta_s)\right]\cdot\left(\sum_{j=0}^{q}y_j\right)\right)\\
&= E\left(-\sum_{t_1=1}^{k}\cdots\sum_{t_q=1}^{k}\frac{\partial}{\partial\alpha_{vw}}\sum_{s=1}^{d\epsilon i}\left[r_{st}-N_t\cdot b_s^{'}(\theta_s)\right]\cdot\left(\sum_{j=0}^{q}y_j\right)\right)\\
&= \sum_{t_1=1}^{k}\cdots\sum_{t_q=1}^{k}\sum_{s=1}^{d\epsilon i}N_t\cdot\frac{\partial}{\partial\alpha_{vw}}b_s^{'}(\theta_s)\cdot\left(\sum_{j=0}^{q}y_j\right)\\
&= \sum_{t_1=1}^{k}\cdots\sum_{t_q=1}^{k}\sum_{s=1}^{d\epsilon i}N_t\cdot\frac{\partial\mu_s}{\partial\eta_s}\cdot\frac{\partial\eta_s}{\partial\alpha_{vw}}\cdot\left(\sum_{j=0}^{q}y_j\right)\\
&= \sum_{t_1=1}^{k}\cdots\sum_{t_q=1}^{k}\sum_{s=1}^{d\epsilon i}N_t\cdot\frac{\partial\mu_s}{\partial\eta_s}\cdot\left(\sum_{j=0}^{q}y_j\right)\cdot\frac{\partial\eta_s}{\partial\alpha_{vw}} \qquad (2.30)
\end{aligned}
$$

where

$$
N_t\cdot\frac{\partial\mu_s}{\partial\eta_s}\cdot\left(\sum_{j=0}^{q}y_j\right)\cdot\frac{\partial\eta_s}{\partial\alpha_{vw}}=\begin{cases}N_t\pi_s(1-\pi_s)\left(\sum_{j=0}^{q}y_j\right)\left(\sum_{w=0}^{q}y_w\right), & \text{if } \eta_s=f(\alpha_s)\\ 0, & \text{otherwise}\end{cases}
$$

If we again consider model (2.28) without the imposition of constraints the matrix of the expected minus second order derivatives $\mathbf{I}(i)$ for each item i is:

$$
\mathbf{I}(1)=\begin{bmatrix}\sum_{t_1\ldots t_q} w_1 & \sum_{t_1\ldots t_q} w_1 z_{1t} & \sum_{t_1\ldots t_q} w_1 z_{2t}\\ \sum_{t_1\ldots t_q} w_1 z_{1t} & \sum_{t_1\ldots t_q} w_1 z_{1t} z_{1t} & \sum_{t_1\ldots t_q} w_1 z_{1t} z_{2t}\\ \sum_{t_1\ldots t_q} w_1 z_{2t} & \sum_{t_1\ldots t_q} w_1 z_{2t} z_{1t} & \sum_{t_1\ldots t_q} w_1 z_{2t} z_{2t}\end{bmatrix}
$$

$$
\mathbf{I}(2)=\begin{bmatrix}\sum_{t_1\ldots t_q} w_2 & \sum_{t_1\ldots t_q} w_2 z_{1t} & \sum_{t_1\ldots t_q} w_2 z_{2t}\\ \sum_{t_1\ldots t_q} w_2 z_{1t} & \sum_{t_1\ldots t_q} w_2 z_{1t} z_{1t} & \sum_{t_1\ldots t_q} w_2 z_{1t} z_{2t}\\ \sum_{t_1\ldots t_q} w_2 z_{2t} & \sum_{t_1\ldots t_q} w_2 z_{2t} z_{1t} & \sum_{t_1\ldots t_q} w_2 z_{2t} z_{2t}\end{bmatrix}
$$

$$
\mathbf{I}(3) = \begin{bmatrix}
\sum\limits_{t_1 \ldots t_q} w_3 & \sum\limits_{t_1 \ldots t_q} w_3 z_{1t} & \sum\limits_{t_1 \ldots t_q} w_3 z_{2t} \\
\sum\limits_{t_1 \ldots t_q} w_3 z_{1t} & \sum\limits_{t_1 \ldots t_q} w_3 z_{1t} z_{1t} & \sum\limits_{t_1 \ldots t_q} w_3 z_{1t} z_{2t} \\
\sum\limits_{t_1 \ldots t_q} w_3 z_{2t} & \sum\limits_{t_1 \ldots t_q} w_3 z_{2t} z_{1t} & \sum\limits_{t_1 \ldots t_q} w_3 z_{2t} z_{2t}
\end{bmatrix}
$$

where $w_i = \pi_i \cdot (\pi_i - 1)$ and $i = 1, 2, 3$.

However, by considering model (2.28) under the imposition of the constraints (2.30) its Fisher Information $\mathbf{I}_F^C$ matrix is:

$$
\mathbf{I}_C^F = \begin{bmatrix} \mathbf{I}_{C_1}^F & \mathbf{I}_{C_2}^F & \mathbf{I}_{C_3}^F \end{bmatrix}
$$

where

$$
\mathbf{I}_{C_1}^F = \begin{bmatrix}
\sum\limits_{t_1 \ldots t_q} w_1 & \sum\limits_{t_1 \ldots t_q} w_1 (z_{1t} + z_{2t}) & \sum\limits_{t_1 \ldots t_q} w_1 (z_{1t} + z_{2t}) \\
\sum\limits_{t_1 \ldots t_q} w_1 (z_{1t} + z_{2t}) & \sum\limits_{t_1 \ldots t_q} w_1 (z_{1t} + z_{2t})^2 + \sum\limits_{t_1 \ldots t_q} w_2 z_{1t}^2 & \sum\limits_{t_1 \ldots t_q} w_1 (z_{1t} + z_{2t})^2 + \sum\limits_{t_1 \ldots t_q} w_2 z_{1t}^2 \\
\sum\limits_{t_1 \ldots t_q} w_1 (z_{1t} + z_{2t}) & \sum\limits_{t_1 \ldots t_q} w_1 (z_{1t} + z_{2t})^2 + \sum\limits_{t_1 \ldots t_q} w_2 z_{1t}^2 & \sum\limits_{t_1 \ldots t_q} w_1 (z_{1t} + z_{2t})^2 + \sum\limits_{t_1 \ldots t_q} w_2 z_{1t}^2 \\
0 & \sum\limits_{t_1 \ldots t_q} w_2 z_{1t} & \sum\limits_{t_1 \ldots t_q} w_2 z_{1t} \\
\sum\limits_{t_1 \ldots t_q} w_1 (z_{1t} + z_{2t}) & \sum\limits_{t_1 \ldots t_q} w_1 (z_{1t} + z_{2t})^2 + \sum\limits_{t_1 \ldots t_q} w_2 z_{1t}^2 & \sum\limits_{t_1 \ldots t_q} w_1 (z_{1t} + z_{2t})^2 + \sum\limits_{t_1 \ldots t_q} w_2 z_{1t}^2 \\
0 & \sum\limits_{t_1 \ldots t_q} w_2 z_{1t} z_{2t} & \sum\limits_{t_1 \ldots t_q} w_2 z_{1t} z_{2t} \\
0 & 0 & 0 \\
0 & 0 & 0 \\
0 & \sum\limits_{t_1 \ldots t_q} w_2 z_{1t} z_{2t} & \sum\limits_{t_1 \ldots t_q} w_2 z_{1t} z_{2t}
\end{bmatrix}
$$

$$
\mathbf{I}^F_{C_2} = \begin{bmatrix}
0 & \sum_{t_1 \dots t_q} w_1(z_{1t}+z_{2t}) & 0 \\
\sum_{t_1 \dots t_q} w_2 z_{1t} & \sum_{t_1 \dots t_q} w_1(z_{1t}+z_{2t})^2 + \sum_{t_1 \dots t_q} w_2 z_{1t}^2 & \sum_{t_1 \dots t_q} w_2 z_{1t} z_{2t} \\
\sum_{t_1 \dots t_q} w_2 z_{1t} & \sum_{t_1 \dots t_q} w_1(z_{1t}+z_{2t})^2 + \sum_{t_1 \dots t_q} w_2 z_{1t}^2 & \sum_{t_1 \dots t_q} w_2 z_{1t} z_{2t} \\
\sum_{t_1 \dots t_q} w_2 & \sum_{t_1 \dots t_q} w_2 z_{1t} & \sum_{t_1 \dots t_q} w_2 z_{2t} \\
\sum_{t_1 \dots t_q} w_2 z_{1t} & \sum_{t_1 \dots t_q} w_1(z_{1t}+z_{2t})^2 + \sum_{t_1 \dots t_q} w_2 z_{1t}^2 & \sum_{t_1 \dots t_q} w_2 z_{1t} z_{2t} \\
\sum_{t_1 \dots t_q} w_2 z_{2t} & \sum_{t_1 \dots t_q} w_2 z_{1t} z_{2t} & \sum_{t_1 \dots t_q} w_2 z_{2t}^2 + \sum_{t_1 \dots t_q} w_3 z_{2t}^2 \\
0 & 0 & \sum_{t_1 \dots t_q} w_3 z_{2t} \\
0 & 0 & \sum_{t_1 \dots t_q} w_3 z_{1t} z_{2t} \\
\sum_{t_1 \dots t_q} w_2 z_{2t} & \sum_{t_1 \dots t_q} w_2 z_{1t} z_{2t} & \sum_{t_1 \dots t_q} w_2 z_{2t}^2 + \sum_{t_1 \dots t_q} w_3 z_{2t}^2
\end{bmatrix}
$$

$$
\mathbf{I}^F_{C_3} = \begin{bmatrix}
0 & 0 & 0 \\
0 & 0 & \sum_{t_1 \dots t_q} w_2 z_{1t} z_{2t} \\
0 & 0 & \sum_{t_1 \dots t_q} w_2 z_{1t} z_{2t} \\
0 & 0 & \sum_{t_1 \dots t_q} w_2 z_{2t} \\
0 & 0 & \sum_{t_1 \dots t_q} w_2 z_{1t} z_{2t} \\
\sum_{t_1 \dots t_q} w_3 z_{2t} & \sum_{t_1 \dots t_q} w_3 z_{1t} z_{2t} & \sum_{t_1 \dots t_q} w_2 z_{2t}^2 + \sum_{t_1 \dots t_q} w_3 z_{2t}^2 \\
\sum_{t_1 \dots t_q} w_3 & \sum_{t_1 \dots t_q} w_3 z_{1t} & \sum_{t_1 \dots t_q} w_3 z_{2t} \\
\sum_{t_1 \dots t_q} w_3 z_{1t} & \sum_{t_1 \dots t_q} w_3 z_{1t}^2 & \sum_{t_1 \dots t_q} w_3 z_{1t} z_{2t} \\
\sum_{t_1 \dots t_q} w_3 z_{2t} & \sum_{t_1 \dots t_q} w_3 z_{1t} z_{2t} & \sum_{t_1 \dots t_q} w_2 z_{2t}^2 + \sum_{t_1 \dots t_q} w_3 z_{2t}^2
\end{bmatrix} \tag{2.31}
$$

The above Fisher Information matrix is a 9×9 matrix in which some elements are equal, due to the equality parameter constraints. Thus by excluding the rows that are equal, the Fisher Information matrix turns into a 6×6 matrix according to the number of the different parameters, that is 6. By carefully observing the Fisher Information matrix of the constrained model, we conclude that each of its elements can be expressed in terms of the elements of the minus Fisher Information matrix for each item i of the unconstrained model:

Thus the above matrix can be written as:

$$\mathbf{I}_C^F = \begin{bmatrix} \mathbf{I}_{C_1}^F & \mathbf{I}_{C_2}^F & \mathbf{I}_{C_3}^F & \mathbf{I}_{C_4}^F \end{bmatrix}$$

where

$$\mathbf{I}_{C_1}^F = \begin{bmatrix} \mathbf{I}(1)_{11} & \mathbf{I}(1)_{12} + \mathbf{I}(1)_{13} \\ \mathbf{I}(1)_{21} + \mathbf{I}(1)_{31} & 2 \cdot \mathbf{I}(1)_{22} + \mathbf{I}(1)_{33} + 2 \cdot \mathbf{I}(1)_{32} \\ \mathbf{I}(1)_{21} + \mathbf{I}(1)_{31} & 2 \cdot \mathbf{I}(1)_{22} + \mathbf{I}(1)_{33} + 2 \cdot \mathbf{I}(1)_{32} \\ 0 & \mathbf{I}(2)_{12} \\ \mathbf{I}(1)_{21} + \mathbf{I}(1)_{31} & 2 \cdot \mathbf{I}(1)_{22} + \mathbf{I}(1)_{33} + 2 \cdot \mathbf{I}(1)_{32} \\ 0 & \mathbf{I}(2)_{32} \\ 0 & 0 \\ 0 & 0 \\ 0 & \mathbf{I}(2)_{32} \end{bmatrix}$$

$$\mathbf{I}^{F}_{C_2} = \begin{bmatrix} \mathbf{I}(1)_{12} + \mathbf{I}(1)_{13} & 0 \\ 2 \cdot \mathbf{I}(1)_{22} + \mathbf{I}(1)_{33} + 2 \cdot \mathbf{I}(1)_{32} & \mathbf{I}(2)_{21} \\ 2 \cdot \mathbf{I}(1)_{22} + \mathbf{I}(1)_{33} + 2 \cdot \mathbf{I}(1)_{32} & \mathbf{I}(2)_{21} \\ \mathbf{I}(2)_{12} & \mathbf{I}(2)_{11} \\ 2 \cdot \mathbf{I}(1)_{22} + \mathbf{I}(1)_{33} + 2 \cdot \mathbf{I}(1)_{32} & \mathbf{I}(2)_{21} \\ \mathbf{I}(2)_{32} & \mathbf{I}(2)_{31} \\ 0 & 0 \\ 0 & 0 \\ \mathbf{I}(2)_{32} & \mathbf{I}(2)_{31} \end{bmatrix}$$

$$\mathbf{I}^{F}_{C_3} = \begin{bmatrix} \mathbf{I}(1)_{12} + \mathbf{I}(1)_{13} & 0 \\ 2 \cdot \mathbf{I}(1)_{22} + \mathbf{I}(1)_{33} + 2 \cdot \mathbf{I}(1)_{32} & \mathbf{I}(2)_{23} \\ 2 \cdot \mathbf{I}(1)_{22} + \mathbf{I}(1)_{33} + 2 \cdot \mathbf{I}(1)_{32} & \mathbf{I}(2)_{23} \\ \mathbf{I}(2)_{12} & \mathbf{I}(2)_{13} \\ 2 \cdot \mathbf{I}(1)_{22} + \mathbf{I}(1)_{33} + 2 \cdot \mathbf{I}(1)_{32} & \mathbf{I}(2)_{23} \\ \mathbf{I}(2)_{32} & \mathbf{I}(3)_{33} + \mathbf{I}(2)_{33} \\ 0 & \mathbf{I}(3)_{13} \\ 0 & \mathbf{I}(3)_{23} \\ \mathbf{I}(2)_{32} & \mathbf{I}(2)_{33} + \mathbf{I}(3)_{33} \end{bmatrix}$$

$$
\mathbf{I}^F_{C_4} = \begin{bmatrix}
0 & 0 & 0 \\
0 & 0 & \mathbf{I}(2)_{23} \\
0 & 0 & \mathbf{I}(2)_{23} \\
0 & 0 & \mathbf{I}(2)_{13} \\
0 & 0 & \mathbf{I}(2)_{23} \\
\mathbf{I}(3)_{31} & \mathbf{I}(3)_{32} & \mathbf{I}(3)_{33} + \mathbf{I}(2)_{33} \\
\mathbf{I}(3)_{11} & \mathbf{I}(3)_{12} & \mathbf{I}(3)_{13} \\
\mathbf{I}(3)_{21} & \mathbf{I}(3)_{22} & \mathbf{I}(3)_{23} \\
\mathbf{I}(3)_{31} & \mathbf{I}(3)_{32} & \mathbf{I}(2)_{33} + \mathbf{I}(3)_{33}
\end{bmatrix}
$$

where $\mathbf{I}(1)_{11}$ is the element of the first row and first column of the matrix of the expected minus second order partial derivatives for the first item.

The Fisher Information matrix of the latent trait model in which we have imposed constraints $\mathbf{I}^F_C$ is:

$$
\mathbf{I}^F_C = \mathbf{A}^T \cdot \mathbf{B}_F \cdot \mathbf{A} \qquad (2.32)
$$

where

- $\mathbf{A}$ is the constraints matrix
- $\mathbf{A}^T$ is the transpose of the matrix $\mathbf{A}$.
- $\mathbf{B}_F$ is a block diagonal matrix with blocks the matrices of the expected minus second order derivatives of the unconstrained model for each item (2.31)

2.5 Factor Scores

As we have already stated in the introductory chapter one of the purposes of fitting latent variable models is the assigning of the individuals on the latent dimensions according to their response pattern. In the literature several scoring methods for several latent variable models have been proposed (Bartholomew 1980, 1981, 1984 and Knott and Albanese, 1993) while in this dissertation we will use the scoring method presented in Moustaki and Knott (2000). Thus the assigning of the individuals on the latent dimensions can be obtained by means of the posterior distribution of the latent variables given the observed responses given by (2.4) for the general linear latent variable model. According to Moustaki and Knott (2000) this posterior distribution can also be written as follows:

$$h(z \mid x) = \frac{\exp \sum_{j=1}^{q} C_j(x) z_j g(0 \mid \mathbf{z}) h(\mathbf{z})}{f(0) \prod_{j=1}^{q} M_{z|0}(C_j(x))} \tag{2.33}$$

where $C_j(x) = \sum_{i=1}^{p} \alpha_{ij} x_i$ are the *Component Scores* of the jth factor and $M_{z|0}$ is the moment generating function of the conditional distribution of the latent variable z given a zero response on all items. This posterior distribution conveys all the information about the latent variables which underlies the response variables $\mathbf{x}$.

From (2.33) it is evident that the posterior distribution of $\mathbf{y}$ depends on $\mathbf{x}$ only through the q dimensional vector $\mathbf{C}' = (C_1, C_2, \ldots, C_q)$. As a result, instead of calculating the posterior distribution $h(\mathbf{y} \mid \mathbf{x})$ in order to locate the individuals on the latent dimensions we can do this by calculating the component scores $C_j(x)$ for each factor and response pattern. An important feature of this scoring method is that it does not depend on the distribution of the latent variables.

2.6 Sampling Properties of the Maximum Likelihood Estimators

An attractive feature of the Maximum Likelihood Estimation Method is the asymptotic behavior of the maximum likelihood estimators $\hat{\boldsymbol{\alpha}}$. Thus, as $n \rightarrow \infty$, the distribution of $\hat{\boldsymbol{\alpha}}$ tends to a multivariate normal distribution with mean vector $\boldsymbol{\alpha}$ and variance-covariance matrix $\mathbf{I}(\boldsymbol{\alpha})^{-1}$.

These sampling properties of the maximum likelihood estimators can be very useful in providing measures of precision of the estimates, that is the standard errors of the estimates. For the maximum likelihood estimator $\hat{\boldsymbol{\alpha}}$, the standard error is given by

$$s.e.\left(\hat{\alpha}_i\right) = \left[\mathbf{I}\left(\hat{\boldsymbol{\alpha}}\right)^{-1}\right]_{ii}^{1/2}$$

Thus, in order to compute the standard errors of the estimated parameters it is necessary to compute the Fisher Information matrix. However, this is impracticable when there are many items and thus more tractable alternatives are needed. In Bartholomew and Knott (1999) an approximation to the Fisher Information matrix is given:

$$\mathbf{I}\left(\hat{\boldsymbol{\alpha}}\right) = \left\{\sum_{m=1}^{n} \frac{1}{f\left(\mathbf{x}_m\right)^2} \frac{\partial f\left(\mathbf{x}_m\right)}{\partial \alpha_j} \frac{\partial f\left(\mathbf{x}_m\right)}{\partial \alpha_k}\right\}$$

In this dissertation, an approximation to the Fisher Information matrix is obtained after the algorithm has converged. Thus, we use the forward difference approximation to the second order partial derivatives in order to estimate the Fisher Information matrix and then obtain the standard errors of the estimated parameters.

2.7 Goodness-of-fit

The goodness-of-fit of the latent variable models in general can be checked in many different ways (Bartholomew and Knott, 1999) and in this section we will present the tests that will be used in this dissertation in order to assess the goodness-of-fit of the fitted latent trait models.

Among the different goodness-of-fit measures the adequacy of the estimated models will be checked by means of the Pearson chi-squared test and the log-likelihood ratio test based on all the possible response patterns. In addition, in order to check the adequacy of the fitted models and obtain comparisons between them we will consider the model selection criteria such as the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). Besides, the comparisons between the fitted models will be based on the log-likelihood ratio test.

As far as the Pearson chi-squared X^2 and the log-likelihood ratio test G^2 are concerned, their statistics are given below:

$$X^2 = \sum_{r=1}^{2^p} \frac{(O(r) - E(r))^2}{E(r)}$$

$$G^2 = 2\sum_{r=1}^{2^p} O(r) \ln \frac{O(r)}{E(r)}$$

where r represents a response pattern and $O(r)$ and $E(r)$ represent the observed and expected frequencies of the response pattern r. Under the assumption that the model holds, both statistics are distributed approximately as

χ^2 with degrees of freedom equal to the number of different response patterns minus the number of independent parameters minus one. Thus, in the case of an unconstrained latent trait model, the corresponding degrees of freedom are $2^p - p(q+1) - 1$, whereas in the case of constrained latent trait models these degrees are reduced according to the number of the different parameters that are left. However, such an approximation of the above statistics is valid only if the sample size n is much bigger than the total number of the distinct response patterns, that is, 2^p, since the observed and the expected frequencies will thus be reasonably large. As a result, we should expect that the approximation of the Pearson chi-squared and the log-likelihood ratio test statistic by the χ^2 distribution is valid only when the expected frequencies are greater than 5. In other cases where the expected frequencies are less than 5 the grouping of the response patterns with expected frequencies less than 5 can be a solution.

When we are interested in comparing two different fitted latent trait models the model selection criteria AIC and BIC whose mathematical formulas are given below will be used:

$$AIC = -2\{\log L\left(\hat{\boldsymbol{\theta}}\right)\} + 2m$$

$$BIC = -2\{\log L\left(\hat{\boldsymbol{\theta}}\right)\} + \log(N) \times m$$

where $\log L\left(\hat{\boldsymbol{\theta}}\right)$ is the value of the log-likelihood at the ML estimate $\hat{\boldsymbol{\theta}}$, m is the number of the model parameters and N (number of individuals) the sample size. In fact, the model with the smallest values for these criteria is regarded as being the "best".

Apart from these two measures, it is possible to consider the likelihood ratio test which is approximated by the χ^2 distribution with degrees of freedom equal to the difference between the number of parameters of the models under comparison. However, this test should not be trusted in cases where the number of the possible response patterns is not much smaller than the sample size since the approximation by the χ^2 distribution would not then be valid. These three measures for comparing different models is useful for the determination of the number of factors required and the checking of equality or fixed parameter assumptions.

Chapter 3

Estimation Procedure-Optimization Algorithms

3.1 Introduction

The unknown parameters in the latent trait models can be estimated in terms of various estimation procedures common in applied statistics. Under the SEM approach we find estimation procedures based on weighted least squares (that are applicable in the presence of several latent variables) such as Christofferson's Method (1975) and Muthén's Method (1978). These methods are based on the assumption that most of the relevant information in the sample data is contained in the first- and second-order margins. However, under the IRT approach the estimation procedure that is usually adopted is the Maximum Likelihood Estimation Method. In particular, in this dissertation we are considering maximum likelihood estimation by means of iterative procedures based on Quasi Newton methods and the Estimation-Maximization (EM) algorithm (Dempster, Laird and Rubin, 1977). Bock and Aitkin (1981) were the first to use the EM method in the area of latent variable models and they suggested that the integral in the likelihood could be evaluated by the Gauss Hermite quadrature method. However, the impairments that are usually traced in the EM algorithm lead us adopt techniques that overcome them, such as the Quasi Newton methods. However, some desirable properties of the EM algorithm encourage us to adopt it in the iterative procedure.

This chapter is organized as follows. Firstly, we introduce the EM algorithm. Secondly, we present the methods of speeding up the convergence rate of the EM algorithm, such as the EM Gradient algorithm and the Quasi - Newton methods. Finally, we present the function `ltm.con` that has been developed in the S - language in order to fit latent trait models under the imposition of parameter constraints.

3.2 The Expectation-Maximization Algorithm

3.2.1 Introduction

There are two reasons for considering the EM algorithm in order to carry out maximum likelihood estimation. Firstly, the EM algorithm is generally useful to maximize certain complicated likelihood functions. Secondly, the EM algorithm takes into account the notion of missing data, which is common in latent variable models. That is, the latent variables are not in practice observable and should be regarded as missing quantities in a theoretical sense.

The EM algorithm consists of two steps. The E or Expectation step in which the missing data are filled in and the M or Maximization step in which the parameters are estimated after the missing data have been reconstructed. One of the advantages of the EM algorithm is its numerical stability. The EM algorithm leads to a steady increase in the likelihood of the observed data. Thus, the EM algorithm avoids wildly overshooting or undershooting the maximum of the likelihood along its current direction of search. Besides this desirable feature, the EM handles parameter constraints (parameters are constrained to lie in a pre specified range or to equal a fixed value) gracefully. Constraint satisfaction is by definition built into the solution of the M step. In contrast, competing methods of maximization must incorporate special techniques to cope with such parameter constraints.

A negative feature of the EM algorithm is its often excruciatingly slow convergence rate in a neighborhood of the optimal point. This rate directly reflects the amount of missing data in a problem. Under fairly mild assumptions, the EM algorithm is guaranteed to converge to a stationary point of the likelihood function. In some very contrived examples, it converges to a saddle point, but this rarely happens in practice. Convergence to a local maximum is more likely to occur. The global maximum can usually be reached by starting the parameters at good but suboptimal estimates such as method-of-moment estimates or by choosing multiple random starting points. In general, almost all maximum likelihood algorithms have trouble distinguishing global from local maximum points.

3.2.2 General Definition of the EM Algorithm

The EM algorithm distinguishes between the observed, incomplete data $\mathbf{Y}$ and the unobserved, complete data $\mathbf{X}$ of a statistical experiment. Some function $t(\mathbf{X}) = \mathbf{Y}$ collapses $\mathbf{X}$ onto $\mathbf{Y}$. For instance, if we represent $\mathbf{X}$ as $(\mathbf{Y}, \mathbf{Z})$, with $\mathbf{Z}$ as the missing data, then t is simply projection onto the $\mathbf{Y}$-component of $\mathbf{X}$. It should be stressed that the missing data can consist of more than just observations missing in the ordinary sense. In fact, the definition of $\mathbf{X}$ is left up to the intuition and cleverness of the statistician. The general idea is to choose $\mathbf{X}$ so that maximum likelihood becomes trivial for the complete data.

The complete data are assumed to have a probability density $f(\mathbf{X} \mid \boldsymbol{\theta})$ that is a function of a parameter vector $\boldsymbol{\theta}$ as well as of $\mathbf{X}$. In the E step of the EM algorithm, we calculate the conditional expectation, which is called the objective function

$$Q(\boldsymbol{\theta} \mid \boldsymbol{\theta}_n) \quad = \quad E\left[\ln f(\mathbf{X} \mid \boldsymbol{\theta}) \mid \mathbf{Y}, \boldsymbol{\theta}_n\right]$$

Here $\boldsymbol{\theta}_n$ is the current estimated value of $\boldsymbol{\theta}$. In the M step, we maximize $Q(\boldsymbol{\theta} \mid \boldsymbol{\theta}_n)$ with respect to $\boldsymbol{\theta}$. This yields the new parameter estimate $\boldsymbol{\theta}_{n+1}$, which satisfies for all θ

$$Q(\boldsymbol{\theta}_{n+1} \mid \boldsymbol{\theta}_n) \quad = \quad Q(\boldsymbol{\theta} \mid \boldsymbol{\theta}_n)$$

and we repeat this two-step process until convergence occurs. Note that $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_{n+1}$ play fundamentally different roles in $Q(\boldsymbol{\theta} \mid \boldsymbol{\theta}_n)$.

3.2.3 EM Gradient Algorithm

As we have already remarked one of the main reasons for using the EM algorithm as a means to obtain maximum likelihood estimates is its appropriateness in maximizing complicated functions. However, there are cases where the M step of the EM algorithm cannot always be solved exactly. In such cases one can approximately maximize the E-step function $Q(\boldsymbol{\theta} \mid \boldsymbol{\theta}_n)$ by one step of Newton's method. The EM gradient algorithm (Lange, 1995) iterates according to

$$\boldsymbol{\theta}_{n+1} = \boldsymbol{\theta}_n - \nabla^2 Q(\boldsymbol{\theta}_n \mid \boldsymbol{\theta}_n)^{-1} \nabla Q(\boldsymbol{\theta}_n \mid \boldsymbol{\theta}_n)^T = \boldsymbol{\theta}_n - \nabla^2 Q(\boldsymbol{\theta}_n \mid \boldsymbol{\theta}_n)^{-1} \nabla L(\boldsymbol{\theta}_n)^T$$

where $\nabla Q(\boldsymbol{\theta} \mid \boldsymbol{\theta}_n)$ and $\nabla^2 Q(\boldsymbol{\theta} \mid \boldsymbol{\theta}_n)$ indicate the first and second differentials of $Q(\boldsymbol{\theta} \mid \boldsymbol{\theta}_n)$ with respect to its left variable $\boldsymbol{\theta}$. The substitution of the score $\nabla L(\boldsymbol{\theta}_n)^T$ for $\nabla Q(\boldsymbol{\theta}_n \mid \boldsymbol{\theta}_n)$ is valid because $L(\boldsymbol{\theta}) - Q(\boldsymbol{\theta} \mid \boldsymbol{\theta}_n)$ attains its minimum at $\boldsymbol{\theta} = \boldsymbol{\theta}_n$. The EM gradient algorithm and the EM algorithm enjoy the same rate of convergence approaching the maximum likelihood point $\hat{\boldsymbol{\theta}}$. Furthermore, in the vicinity of $\hat{\boldsymbol{\theta}}$, the EM gradient algorithm also satisfies the ascent condition $L(\boldsymbol{\theta}_{n+1}) < L(\boldsymbol{\theta}_n)$ (Lange, 1995).

In practice, an alternative to the iterative step of the EM gradient algorithm replaces $\nabla^2 Q(\boldsymbol{\theta}_n \mid \boldsymbol{\theta}_n)^{-1}$, that is the inverse of the Hessian matrix $\mathbf{H}$, with $E(\mathbf{H}) = -\mathbf{I}$, where $\mathbf{I}$ is the Fisher Information matrix, the expected value of the minus second derivatives. Therefore, the EM gradient algorithm iterates according to

$$\boldsymbol{\theta}_{n+1} = \boldsymbol{\theta}_n + \mathrm{I}^{-1} \nabla L(\boldsymbol{\theta}_n)^T$$

similarly with the Fisher Scoring algorithm.

3.2.4 The EM Gradient Algorithm in Latent Trait Models

The maximization of the log-likelihood (or the minimization of the minus log-likelihood), in the latent trait models, is done by the EM algorithm. The steps of the algorithm in the case where equality and fixed value constraints are imposed on the parameters are defined as follows:

Step 1 Choose initial estimates for the model parameters α_{ij}.

Step 2 E - Step. Compute the values of r_{it} and N_t from (2.19) and (2.20).

Step 3 Obtain improved estimates of the parameters α_{ij} by solving the non-linear maximum likelihood equations (2.21) for $l = 0, 1$ and $i = 1, 2, \ldots, p$ using Fisher Scoring, treating r_{it} and N_t as given numbers.

Step 4 Return to Step 2 and continue until convergence is attained.

In the case where fixed value parameter constraints are imposed we need to intervene in the EM algorithm. In Step 3 improved estimates of all the parameters are obtained. Some of these estimated parameters are assumed to take a pre-defined fixed value. Thus, these parameters after Step 3 are set equal to their pre-defined value. This procedure is followed in each iteration until convergence.

3.3 Quasi-Newton Methods

3.3.1 Introduction

As we have already remarked one of the main drawbacks of the EM algorithm is its excruciatingly slow convergence. In contrast, Newton's method enjoys exceptionally quick convergence in a neighborhood of the maximum point. This suggests amending the EM algorithm so that it resembles Newton's method. Because the EM algorithm typically performs well far from the maximum likelihood point, hybrid algorithms that begin as pure EM and gradually make the transition to Newton's method are apt to perform best.

In adopting the Gradient EM algorithm in order to improve the convergence rate of the pure EM algorithm requires the computation of the Hessian matrix at each iteration and finding a solution for $\nabla^2 Q\left(\boldsymbol{\theta}_n|\boldsymbol{\theta}_n\right)^{-1} \nabla L\left(\boldsymbol{\theta}_n\right)$ at each iteration. Such a procedure sounds very time consuming and thus it is necessary to avoid computing the Hessian matrix and use other matrices instead of the inverse Hessian. Such methods that try to mimic Newton's method without directly calculating the Hessian matrix are called Quasi-Newton methods. In the literature there are many suggestions concerning the acceleration of the convergence of the EM algorithm. Among these we find Jamshidian and Jennrich (1997), who propose two accelerated versions of the EM algorithm based on Quasi-Newton methods for solving equations and minimizing functions and Lange (1995).

In this section, we will represent the BFGS minimization algorithm.

3.3.2 BFGS Algorithm

General Definition. The general BFGS algorithm is described as follows:

- given the current point θ_0 and an approximate Hessian $\mathbf{A_0}$, compute the new search direction $-\mathbf{A}_0^{-1} \bigtriangledown f(\theta_0)$
- find a better point θ_1 in this direction, and
- update $\mathbf{A}_0$ to a new approximation $\mathbf{A}_1$, repeating until convergence.

The algebraic form of the BFGS update of $\mathbf{A}_0$ is:

$$\mathbf{A}_1 = \mathbf{A}_0 + \frac{1}{y^{'}s}yy^{'} - \frac{1}{s^{'}\mathbf{A}_0 s}\mathbf{A}_0 ss^{'}\mathbf{A}_0 \tag{3.1}$$

where $y = \bigtriangledown f(\boldsymbol{\theta}_1) - \bigtriangledown f(\boldsymbol{\theta}_0)$ and $s = \boldsymbol{\theta}_1 - \boldsymbol{\theta}_0$. The difference $\mathbf{A}_1 - \mathbf{A}_0$ is the sum of two rank 1 matrices, and is thus generally a rank 2 matrix. From (3.1) we have that $\mathbf{A}_1$ is positive definite if $\mathbf{A}_0$ is positive definite and $s^{'}y \succ 0$. Another property of the $\mathbf{A}_1$ BFGS update of $\mathbf{A}_0$ is that it satisfies

$$\bigtriangledown f(\boldsymbol{\theta}_1) - \bigtriangledown f(\boldsymbol{\theta}_0) = \mathbf{A}_1 (\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0) \tag{3.2}$$

Assuming a first order Taylor expansion of the first order derivative $\bigtriangledown f(\boldsymbol{\theta})$ around the current point $\boldsymbol{\theta}_1$ implies that the Hessian $\bigtriangledown^2 f(\boldsymbol{\theta})$ also satisfies (3.2).

An algebraically equivalent update to (3.1) is given by

$$\mathbf{A}_1^{-1} \quad = \quad \mathbf{A}_0^{-1} + \frac{1}{y^{'}s}ss^{'} - \frac{1}{y^{'}\mathbf{A}_0^{-1}yy^{'}\mathbf{A}_0^{-1} + \left(y^{'}\mathbf{A}_0^{-1}y\right)uu^{'}} \tag{3.3}$$

where $u = \frac{1}{y^{'}s} - \frac{1}{y^{'}\mathbf{A}_0^{-1}y}\mathbf{A}_0^{-1}y$.

Since Newton Raphson is known for its good convergence properties once it gets close to the solution, it would be desirable for the BFGS updates $\mathbf{A}_1$

to approximate $\nabla^2 f(\boldsymbol{\theta}_1)$ as the algorithm converges. This is the case if the true function is a positive definite quadratic.

However, accuracy of the $\mathbf{A}_1$ matrix at convergence is not guaranteed and thus this approximate Hessian should not be used to obtain the standard errors of the parameter estimates. Indeed, the Hessian should be calculated after the BFGS has converged.

In addition, caution is needed in the selection of the initial positive definite matrix and the choice of the new point $\boldsymbol{\theta}_1$. In fact, in the BFGS algorithm the choice of the initial hessian matrix is of paramount importance since it must be scaled appropriately. If it differs by several orders of magnitude from the correct hessian, it will tend to create difficulties in the search for a better point at each iteration, since the initial step based on this matrix will be scaled inappropriately. Also, it may take many iterations of the BFGS updates to correct the poor initial scaling. Thus as an initial matrix, a diagonal matrix is usually chosen with elements the $max(|-L(\boldsymbol{\theta})|, 1)$ in the main diagonal, while for the choice of the new point $\boldsymbol{\theta}_{n+1}$ the backtracking methodology, that will be presented in the next section, is used.

In the case where fixed value parameter constraints are imposed we need to intervene in the BFGS algorithm. In Step 2 improved estimates of all the parameters are obtained. Some of these estimated parameters are assumed to take a pre-defined fixed value. Thus, these parameters after Step 2 are set equal to their pre-defined value. This procedure is followed in each iteration until convergence.

3.3.3 Backtracking

As we have already stated the BFGS algorithm belongs to the family of the Quasi-Newton methods that try to mimic the Newton Raphson without directly calculating the Hessian matrix. Thus, some problems that usually occur in the Newton-Raphson can be traced in the BFGS algorithm as well. Thus, for the BFGS algorithm to be successful it is necessary that the direction $-A_0^{-1} \nabla f(\boldsymbol{\theta}_0)$ is descent for f, that is, it leads f to decrease. Otherwise the algorithm diverges.

However, the divergence problem of the BFGS algorithm can be fixed easily. Thus, instead of taking fixed steps at each iteration another methodology can be used, which is known as backtracking. The following step of the Newton-Raphson algorithm in which the updates are obtained:

$$\boldsymbol{\theta}_{n+1} = \boldsymbol{\theta}_n - \left[\nabla^2 f(\boldsymbol{\theta}_0)\right]^{-1} \nabla f(\boldsymbol{\theta}_0)$$

can be modified and take the following form:

$$\boldsymbol{\theta}_{n+1} = \boldsymbol{\theta}_n - \lambda \left[\nabla^2 f(\boldsymbol{\theta}_0)\right]^{-1} \nabla f(\boldsymbol{\theta}_0) \tag{3.4}$$

where the multiplier λ needs to be determined. The backtracking procedure that is adopted consists of the following steps:

Step 1. Compute the full Newton-Raphson step, corresponding to $\lambda = 1$ in (3.4). If $f(\boldsymbol{\theta}_{n+1}) < f(\boldsymbol{\theta}_n)$ then keep the new point and repeat.

Step 2. Backtracking step. If $f(\boldsymbol{\theta}_{n+1}) \geq f(\boldsymbol{\theta}_n)$, then reject the new point and backtrack towards $\boldsymbol{\theta}_n$ by computing $\boldsymbol{\theta}_{n+1}$ for values $\lambda < 1$, until a better point is found.

However, satisfaction of Step 1 does not imply that the algorithm will converge. Thus, Dennis and Schnabel (1983) have recommended the following stronger condition that λ has to satisfy:

$$f\left(\boldsymbol{\theta}_{n+1}(\lambda)\right) < f\left(\boldsymbol{\theta}_{n}\right) + 10^{-4}\left(\boldsymbol{\theta}(\lambda)_{n+1} - (\boldsymbol{\theta}_{n})\right)' \nabla f\left(\boldsymbol{\theta}_{n}\right) \qquad (3.5)$$

This condition is always satisfied if $\boldsymbol{\theta}_{n+1} - \boldsymbol{\theta}_{n}$ is a descent direction. Then in order to choose a sequence of λ values for the backtracking step we use a strategy that reduces the step length by a fixed fraction δ each time. Thus, we start with $\lambda = 1$ and if $f\left(\boldsymbol{\theta}_{n+1}(1)\right)$ does not improve on $f\left(\boldsymbol{\theta}_{n}\right)$, we try $\boldsymbol{\theta}_{n+1}(\delta)$. If $f\left(\boldsymbol{\theta}_{n+1}(\delta)\right)$ does not improve on $f\left(\boldsymbol{\theta}_{n}\right)$, we try $\boldsymbol{\theta}_{n+1}(\delta^2)$, etc. Usually δ is chosen to equal 1/2 and in this case the backtracking methodology is called step halving. In addition, smaller values of δ can be used which sometimes lead to faster convergence.

3.4 Description of the `ltm.con` function

In this section we will present the function `ltm.con` that has been developed in the S-language in order to fit "constrained" latent trait models. That is, this function is designed to fit models under the imposition either equality and fixed value constraints or equality constraints only. The estimation procedure that we have chosen is the maximum likelihood estimation which is implemented by means of the BFGS algorithm and the EM Gradient algorithm that lead to a relatively quick convergence of the likelihood, according to the theory described in the previous sections. In this function the maximum likelihood estimates for the log-likelihood are obtained by maximizing the log-likelihood during the iterations of the EM algorithm and by minimizing the minus log-likelihood during the iterations of the BFGS algorithm.

The function `ltm.con` consists mainly of two algorithms, the EM and the BFGS algorithm. As we have already stated the reason for adopting the BFGS algorithm is the speeding-up of the convergence of the likelihood in the neighborhood of the maximum, while the EM algorithm is used in order to lead the likelihood quickly near to the neighborhood of the maximum. Thus, we have decided to implement firstly the EM algorithm for a pre-specified number of iterations and then the BFGS algorithm until the convergence of the likelihood. The number of iterations for the EM algorithm are dependent on the number of the latent variables, the number of the constraints imposed on the parameters and the starting values. As far as the starting values are concerned the user is given the option either to specify the starting values of the parameters or let them be randomly chosen from the normal distribution with pre-specified mean and variance, when there is no prior information available. For randomly chosen initial values, many latent variables (two or more) and few parameter constraints it has been proven (due to simulations) that the EM algorithm may need 70-100 iterations in order to get near to the neighborhood of the maximum.

After the iterative procedure based on the EM algorithm is terminated, the values of the last iteration of the EM algorithm constitute the initial values for the BFGS algorithm. In particular, the initial values of the BFGS algorithm consist of the estimated parameters obtained by the last iteration of the EM algorithm and the inverse of the Fisher Information matrix (or the minus Hessian matrix), which is calculated according to (2.32) at the obtained estimated parameters. This initial Hessian matrix is positive definite since such an assumption is checked during the EM iterations. In the cases where the Hessian matrix is not positive definite it is modified by adding to the main diagonal a large enough value to make it positive definite. Apart from the inverse of the Fisher Information matrix of the last iteration of the EM algorithm the user is given the chance to use as initial hessian matrix a diagonal matrix with elements $max\left(\left|-L\left(\boldsymbol{\alpha}\right)\right|, 1\right)$ on the diagonal. In fact, the BFGS algorithm is a function implementing the update (3.3), together with the backtracking algorithm.

For the search at each iteration, the BFGS algorithm simply starts by computing

$$\boldsymbol{\theta}_{n+1} = \boldsymbol{\theta}_n + \mathrm{I}^{-1} \bigtriangledown L\left(\boldsymbol{\theta}_n\right)^T$$

and if that point does not satisfy (3.5), the algorithm backtracks towards $\boldsymbol{\theta}_n$ until a point satisfying (3.5) is found. This is done using the backtracking methodology presented in Section 3.3.3. During the backtracking the condition $s'y > 0$ is checked in order to obtain a positive definite update. The iterative procedure of the BFGS algorithm continues until the convergence criteria are satisfied. The primary convergence criterion stops the iteration when the maximum of the absolute value of the first order partial derivatives is less than a pre-specified value which is set by default equal to 10^{-5}. In the case where fixed value parameter constraints are imposed, the first order partial derivatives of the parameters that are assumed to equal a pre-defined fixed value are set equal to zero. According to a second convergence criterion, the iteration will also stop if the relative change in the parameters in the full Newton step is less than a pre-specified value which is set by default equal to 10^{-8}. This is sometimes needed, because due to limitations on numeric accuracy, it is not always possible to attain the specified criterion on the gradient, especially when finite differences are used to numerically approximate the gradient.

The virtue of this general approach is that as the algorithm approaches a solution and the $-\mathbf{I}^{-1}$ matrix approaches the inverse Hessian, the behavior of the algorithm becomes nearly identical to the Newton-Raphson iteration, and converges rapidly.

Chapter 4

Applications

4.1 Introduction

In this chapter we will illustrate the fitting procedure of "constrained" latent trait models through two examples derived from Brooke et al. (1992) and Bock and Lieberman (1970) respectively. The last example has been analyzed in Bartholomew *et. al.* (2002, Chapter 7), without assuming that the imposition of parameter constraints is needed. Such an assumption will be investigated in this chapter. Thus, for each of the above two examples we will firstly represent a latent trait model fitted to the data without imposing any parameter constraints. Secondly, after observing the estimated parameters we will impose equality constraints between some parameters in order to check whether the equality assumption holds or not. The estimation of the latent trait models without the imposition of constraints will be obtained by means of the function ltm (Rizopoulos, 2003) while the estimation of the restricted model will be obtained by means of the function ltm.con which has been presented in Chapter 3.

The two examples are given for illustrative reasons, while the imposition of equality parameter constraints should be done after prior information is available (Section 1.2.2).

4.2 Latent Trait Model for Race data

In this section we will illustrate the use of the imposition of constraints in latent trait models through an example based on data from the British Social Attitudes Survey in 1991 given in Brooke *et. al.* (1992). According to this survey four questions relevant to the racial attitudes were addressed to 1268 individuals and their replies have been recorded. The questions were:

- Thinking of black people - that is people whose families were originally from the West Indies or Africa - who now live in Britain. Do you think there is a lot of prejudice against them in Britain nowadays, or hardly any?
- Do you think most white people in Britain would mind or not if a suitably qualified person of Asian origin was appointed as their boss?
- And you personally? Would you mind or not?
- Do you think that most white people in Britain would mind if one of their close relatives was to marry a person of Asian origin?

Each variable is binary, coded so that "1" denotes the answer indicating presence or expression of racial prejudice, and "0" denotes the absence of such a prejudice.

We are interested in checking whether the items selected are the indicator variables of the trait for which they have been designed or not. In addition, we are interested in drawing conclusions for the items in terms of the estimated parameters. The analysis of the latent trait model is organized as follows. Firstly, a descriptive analysis of the Race data will be performed. Secondly, the "unconstrained" latent trait model will be fitted to the data and its adequacy will be estimated in terms of goodness-of-fit measures. Thirdly, after observing the estimated parameters we will explore whether the imposition of equality constraints between some parameters is necessary or not. Finally, we will compute the factor scores for the individuals on the latent dimension, for the model that fits adequately the data.

4.2.1 Descriptive Analysis

As far as the underlying research is concerned, four items have been constructed in order to measure the racial prejudice of British people and their answering them positively or not has been recorded. At this point, we will describe the data by calculating frequencies and percentages and by exploring the associations between the items.

The answers of the individuals to the four items are represented by the response pattern of each individual and all the possible response patterns are 2^4, that is, 16. The frequencies of these response patterns are given in Table 4.1. From Table 4.1 we observe that there are not any response patterns with zero frequency, while 6 out of 16 response patterns have frequencies ≤ 5.

The percentages of individuals giving positive or negative answers to the four items are summarized in Table 4.2.

Item 1	Item 2	Item 3	Item 4	Observed Frequency
0	0	0	0	26
1	0	0	0	20
0	1	0	0	5
1	1	0	0	3
0	0	1	0	98
1	0	1	0	45
0	1	1	0	158
1	1	1	0	131
0	0	0	1	6
1	0	0	1	5
0	1	0	1	2
1	1	0	1	2
0	0	1	1	36
1	0	1	1	21
0	1	1	1	331
1	1	1	1	379

Table 4.1: Observed frequencies of response patterns, Race data

	Response 1	Response 0
Item 1	47.79	52.21
Item 2	79.73	20.27
Item 3	94.56	5.44
Item 4	61.67	38.33

Table 4.2: Percentages of expressing or not racial prejudice for the observed items, Race data

We observe that items 2 and 3, that represent the prejudice of British people and the respondents' prejudice against bosses with Asian origin, have the greatest percentages of expressing racial prejudice.

Before performing a latent trait analysis, we need to explore associations between pairs of variables which might suggest the existence of one or more common underlying factors. In the case of binary variables the only way to explore the pairwise associations between the variables is to construct 2×2 contingency tables and check for the statistical significance of the displayed associations. In Table 4.3 the p-values of the pairwise associations of the variables are given.

Item	Item	p-value
1	2	< 0.001
1	3	0.539
1	4	< 0.001
2	3	< 0.001
2	4	< 0.001
3	4	< 0.001

Table 4.3: Pairwise associations between observed variables, Race data

From Table 4.3 we observe that there is association between the pairs of all the variables except from the pair "Prejudice against Black People" and "Respondent's Prejudice against Boss with Asian Origin". This conclusion suggests that it would be worth asking whether these associations can be attributed to one or more common factors. This is what a latent trait model enables us to do. Since there are four items and in factor analysis it is common practice to use one factor every three items, we begin our analysis by fitting an one-factor model to the four items. Thus, if we can identify common factors, we may then wish to go on to compute scores for individuals on the latent dimensions.

4.2.2 One-Factor Unconstrained LTM

In this subsection we will assume that the associations between the four items of the test can be accounted for by only one common factor. Thus we will assume the following latent trait model:

$$\text{logit}\,(\pi_1) = \alpha_{10} + \alpha_{11} y_1$$

$$\begin{aligned}
\text{logit}\,(\pi_2) &= \alpha_{20} + \alpha_{21} y_1 \\
\text{logit}\,(\pi_3) &= \alpha_{30} + \alpha_{31} y_1 \\
\text{logit}\,(\pi_4) &= \alpha_{40} + \alpha_{41} y_1
\end{aligned}$$

or equivalently

$$\pi_i\,(y) = \frac{exp\,(\alpha_{i0} + \alpha_{i1} y)}{1 + exp\,(\alpha_{i0} + \alpha_{i1} y)}, \; i = 1, 2, 3, 4$$

where α_{i0} denotes the difficulty parameter, α_{i1} denotes the discrimination parameter and y_1 represents the latent variable that measures the racial prejudice of the individuals of the population under study. The assumptions of that model are listed below:

1. Conditional independence. That is, the responses to the 4 observed items must be independent conditional on the latent variable. This means that the latent variable accounts for all the associations among the observed items. Since the latent variable is unobserved, the assumption of conditional independence can only be tested indirectly by checking whether the model fits the data.

2. The link function: $\text{logit}\pi_i\,(y) = \alpha_{i0} + \alpha_{i1} y$, where $P\,(x_i = 1|y) = \pi_i\,(y)$

3. The latent variable comes from the standard normal distribution. That is $y \sim N\,(0, 1)$.

Estimation of the unconstrained LTM

The model that we have fitted can be estimated by means of the maximum likelihood method. The parameter estimates and their standard errors are given in Table 4.4:

Item	$\hat{\alpha_{i0}}$	se ($\hat{\alpha_{i0}}$)	$\hat{\alpha_{i1}}$	se ($\hat{\alpha_{i1}}$)	st ($\hat{\alpha_{i0}}$)	$\hat{\pi}_i$ (0)
1	-0.09	0.0582	0.38	0.0822	0.355	0.478
2	5.63	4.8567	6.30	5.3034	0.988	0.996
3	4.87	0.5555	2.45	0.4159	0.926	0.992
4	0.64	0.0849	1.33	0.1846	0.799	0.655

Table 4.4: Parameter estimates and standard errors for the unconstrained one-factor LTM, Race data

In the first column of Table 4.4, we give the estimated difficulty parameters $\hat{\alpha_{i0}}$ and in the second column their standard errors. In the third column, we give the estimated discrimination parameters $\hat{\alpha_{i1}}$ and in the fourth column their standard errors. Then, in the fifth column the standardized values of the factor loadings are given and in the sixth column the probabilities that the median individual will respond positively to items $1-4$ are given.

In Table 4.4, the estimated discrimination parameters, $\hat{\alpha_{i1}}$, that is the estimated factor loadings, are large (at least the last three). This implies that there is an underlying factor that is common to all items. In addition, this implies that the characteristic curves of these items are very steep. However, it worths mentioning that the standard errors of the estimated parameters of the second item are greater than those of the other items. Taking the loadings in their face value, it appears that the second item is the best discriminator between for and against attitude while the first item is the worst discriminator. The estimates of the probabilities of the median individual to respond positively to the four items show that the four items are of varying difficulty but relatively easy. In addition, the estimated probabilities of the median individual lead to conclusions similar to those drawn from the descriptive analysis (Table 4.2). The standardized values of the factor loadings can be interpreted as correlation coefficients and thus only the second and the third items have a close link to the common factor. The standardized values of the factor loadings are given by the following formula:

$$st\alpha_{ij} = \frac{\alpha_{ij}}{\sqrt{\sum_{j=1}^{q} \alpha_{ij}^2 + 1}}$$

After having estimated the unknown parameters of the above latent trait model we will proceed to the exploration of the imposition of equality parameter constraints. That is, we may be interested whether the weights of one item are exactly the same to the weights of another or other items. Thus in Section 4.2.3 we will fit the one-factor latent trait model after the imposition of equality constraints between the parameters of the model.

Goodness-of-fit

In order to assess the goodness-of-fit of the model that we have fitted we will use the log-likelihood ratio test G^2, the Pearson chi-squared goodness-of-fit test X^2 and the selection criteria AIC and BIC that have been defined in Chapter 2. In Table 4.5, the values of these measures are given.

Test	value	d.f.	p-value
Log-Likelihood Ratio Test G^2	6.731	7	0.4574
Pearson Chi-Squared X^2	6.836	7	0.4462
AIC	4941.998		
BIC	4994.25		

Table 4.5: Goodness-of-fit measures for the unconstrained one-factor model, Race data

Both measures do indicate a good fit of the one-factor latent trait model to the data. This result is easily verified by the small discrepancies between the observed and expected frequencies in the response patterns. After having

Item				Observed	Expected
1	2	3	4	Frequency	Frequency
0	0	0	0	26	32.573
1	0	0	0	20	14.771
0	1	0	0	5	3.841
1	1	0	0	3	2.678
0	0	1	0	98	90.664
1	0	1	0	45	51.402
0	1	1	0	158	153.947
1	1	1	0	131	135.916
0	0	0	1	6	6.361
1	0	0	1	5	3.313
0	1	0	1	2	3.201
1	1	0	1	2	2.488
0	0	1	1	36	35.457
1	0	1	1	21	21.598
0	1	1	1	331	335.946
1	1	1	1	379	373.829

Table 4.6: Observed and expected frequencies of the response patterns, Race data

fitted the one-factor unconstrained latent trait model, it is useful to represent graphically the relationship between examinees' item performance and their racial prejudice by means of the item characteristic curves (ICC).

From Figure 4.1 it is evident that as the factor's values increase the probability of expressing racial prejudice increases. In particular, the rate of increase is more rapid for the third item than the others, while for the first item this increase is the slowest.

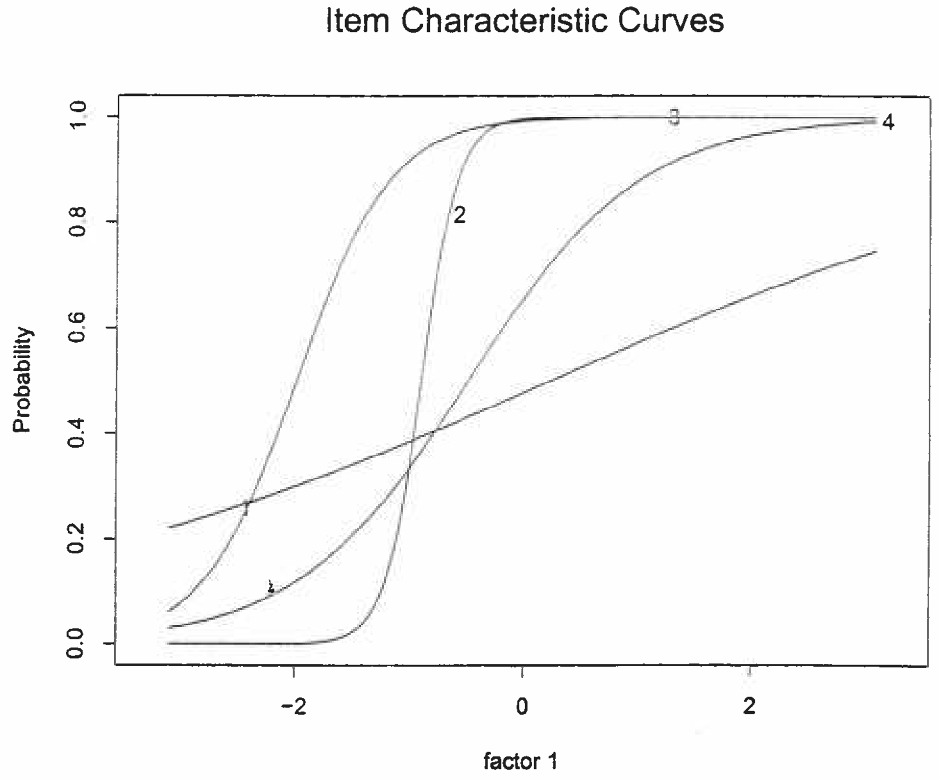

Figure 4.1: Item Characteristic Curves for the four items of the unconstrained latent trait model, Race data

4.2.3 One-Factor constrained LTM

In this section we will explore the imposition of parameter constraints in the one-factor latent trait model presented in Section 4.2.2. In Table 4.4 we observed that the estimated parameters of the one-factor latent trait model for the second and third item have large values. Thus we will assume that the discrimination parameters of the second item are equal to the discrimination parameters of the third item and explore the validity of such an assumption. In other words, we will fit the following latent trait model:

$$\begin{aligned}
\text{logit}\,(\pi_1) &= \alpha_{10} + \alpha_{11} y_1 \\
\text{logit}\,(\pi_2) &= \alpha_{20} + \alpha_{21} y_1 \\
\text{logit}\,(\pi_3) &= \alpha_{30} + \alpha_{31} y_1 \\
\text{logit}\,(\pi_4) &= \alpha_{40} + \alpha_{41} y_1
\end{aligned}$$

where we will assume that $\alpha_{21} = \alpha_{31}$.

The assumptions of the one-factor latent trait model given in Section 4.2.2 need also to be satisfied in the one-factor latent trait model under the imposition of constraints.

Estimation of the constrained LTM

The model that we have fitted can be estimated by means of the maximum likelihood method. The parameter estimates and their standard errors are given in Table 4.7.

In Table 4.7, we observe that the standard errors of the estimated difficulty and discrimination parameters of the second item have been reduced

Item	$\hat{\alpha}_{i0}$	se $(\hat{\alpha}_{i0})$	$\hat{\alpha}_{i1}$	se $(\hat{\alpha}_{i1})$	st $(\hat{\alpha}_{i0})$	$\hat{\pi}_i(0)$
1	-0.09	0.0034	0.38	0.0071	0.354	0.477
2	3.07	0.2460	3.21	0.4096	0.955	0.956
3	5.90	1.0796	3.21	0.5401	0.955	0.997
4	0.67	0.0079	1.49	0.0310	0.831	0.662

Table 4.7: Parameter estimates and standard errors for the constrained one-factor LTM, Race data

and are now in similar magnitude to the standard errors of the other estimates.

Goodness-of-fit

In order to assess the goodness-of-fit of the model that we have fitted we will use the log-likelihood ratio test G^2, the Pearson chi-squared goodness-of-fit test X^2 and the selection criteria AIC and BIC whose values are given in Table 4.8.

Test	value	d.f.	p-value
Log-Likelihood Ratio Test G^2	9.886	6	0.129
Pearson Chi-Squared X^2	10.402	6	0.109
AIC	4943.153		
BIC	5029.186		

Table 4.8: Goodness-of-fit measures for the constrained one-factor model, Race data

According to Table 4.8, both measures do indicate a good fit of the one-factor latent trait model to the data, while the values of the AIC and BIC of the constrained model are larger than those of the unconstrained one. This implies that the unconstrained model is preferable. The adequacy of the constrained model is easily verified by the small discrepancies between the observed and expected frequencies in response patterns in Table 4.9.

Item				Observed	Expected	Total	Component
1	2	3	4	Frequency	Frequency	Score	Score
0	0	0	0	26	34.966	0	0.000
1	0	0	0	20	15.413	1	0.379
0	0	0	1	6	4.864	1	1.494
0	0	1	0	98	84.617	1	3.214
0	1	0	0	5	4.983	1	3.214
1	0	0	1	5	2.479	2	1.873
1	0	1	0	45	50.255	2	3.593
1	1	0	0	3	2.960	2	3.593
0	0	1	1	36	37.419	2	4.708
0	1	0	1	2	2.204	2	4.708
0	1	1	0	158	158.248	2	6.428
1	0	1	1	21	25.638	3	5.087
1	1	0	1	2	1.510	3	5.087
1	1	1	0	131	134.056	3	6.807
0	1	1	1	331	334.680	3	7.922
1	1	1	1	379	373.710	4	8.301

Table 4.9: Observed, expected frequencies and scores of the response patterns for the constrained one-factor LTM, Race data

In addition, it is interesting to check whether the latent trait model fitted to the data after the imposition of constraints, is "better" than the one where no constraints were imposed or not. This will be done by means of the likelihood ratio test.

Model	Log-Likelihood	Likelihood Ratio Test Value	d.f.	p-value
Constraints	-2464.577			
No Constraints	-2462.999	3.156	1	0.0756

Table 4.10: Comparison between the constrained and the unconstrained model, Race data

The high p-value of the likelihood ratio test in $\alpha = 5\%$ implies that the two models do not differ and that the most parsimonious, that is the constrained model, is preferable contrary to the conclusions reached due to the AIC and BIC values. Thus, we conclude that the imposition of constraints in the above model has improved the fit to the data. The validity of this conclusion and the approximation of the likelihood-ratio test statistic by the χ^2 distribution is strengthened by the fact that in this example there are not any response patterns with zero frequencies.

Our purpose after fitting latent variable models is to obtain information about the latent variables through the posterior distribution $h(\mathbf{y} \mid \mathbf{x})$ (Posterior Analysis). In Section 2.6 we showed that this distribution depends on $\mathbf{x}$ only through the q-variate sufficient statistic $\mathbf{X} = \mathbf{Ax}$, where $\mathbf{A} = \alpha_{ij}$. Thus, for the constrained one-factor latent trait model that we have fitted we can locate the individuals on the latent dimension and thus assign to each individual a score with respect to their racial prejudice. These score are given in the last column of Table 4.9 next to the column of "Total Score" that gives the number of positive responses to each response pattern. Thus we observe that the rankings given by the total and the component score are not close. For example, the response patterns 0100 and 0010 have a component score which is substantially higher than the other cases where there was only one positive response. This is because the items 2 and 3 have a much higher weight than the others.

From Figure 4.2 it is evident that as the factor values increase the probability of expressing racial prejudice increases. In particular, the rate of increase is more rapid for the second and third item than the others, while for the first item this increase is the slowest. In addition, the item characteristic curves for the second and the third item have parallel evolutions due

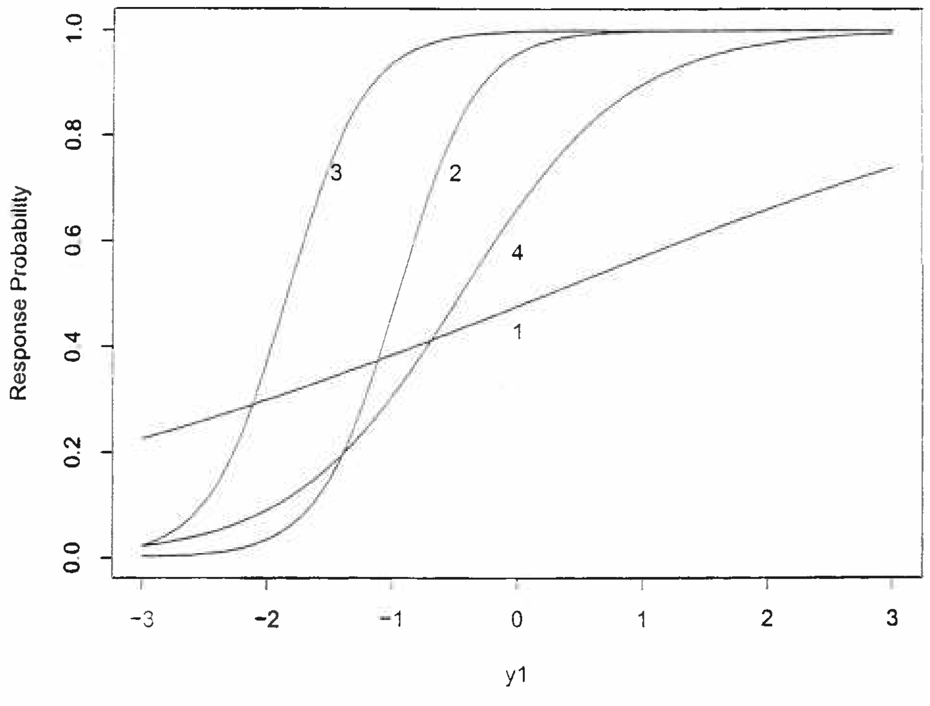

Figure 4.2: Item Characteristic Curves for the four items of the constrained latent trait model, Race data

to the fact that their discrimination parameters have been set equal.

4.3 Latent Trait Model for Law School Admission Test data

In this section we will illustrate the use of the imposition of constraints in latent trait models through an example based on part of data from Law School Admission Test (LSAT) given in Bock and Lieberman (1970). The LSAT is well known in educational testing for measuring ability traits. This part of the LSAT data that we are going to analyze is designed to measure a single latent ability scale and consists of five items taken by 1000 individuals. We are interested in checking whether the items selected are the indicator variables of the ability for which they have been designed to measure. In addition, we are interested in drawing conclusions for the items in terms of the estimated parameters.

The analysis of the latent trait model is organized as follows. Firstly, a descriptive analysis of the LSAT data will be performed. Secondly, the unconstrained latent trait model will be fitted to the data and its adequacy will be estimated in terms of goodness-of-fit measures.

Thus, in this section we will firstly represent a latent trait model fitted to the LSAT data without imposing any parameter constraints. Secondly, after observing the estimated parameters we will impose equality constraints between some parameters in order to check whether the equality assumption holds or not. Thirdly, after observing the estimated parameters we will explore whether the imposition of equality constraints between some parameters is necessary or not. Finally, we will compute the factor scores for the individuals on the latent dimensions for the model that fits adequately the data.

4.3.1 Descriptive Analysis

As far as the underlying research is concerned, five items have been constructed in order to measure ability that the test-takers possess. At this point, we will describe the data by calculating frequencies and percentages and by exploring the association between the items.

The answers of the individuals to the four items are represented by the response pattern of each individual and all the possible response patterns are 2^5, that is, 32. The frequencies of these response patterns are given in Table 4.11.

The percentages of individuals giving correct or wrong answers to the five items are summarized in Table 4.12.

We observe that item 3 appears to be the most difficult. Before performing a latent trait analysis, we need to explore the associations between pairs of variables which might suggest the existence of one or more common underlying factors. In Table 4.13 the p-values of the pairwise associations of the variables are given.

From Table 4.13 we observe that the majority of the pairs of the variables are highly correlated. After fitting a latent trait model we can check whether these associations can be attributed to one common factor. Thus, if we can identify such a common factor, we may then wish to go on to compute scores for individuals on the latent dimension.

Item					Observed
1	2	3	4	5	Frequency
0	0	0	0	0	3
1	0	0	0	0	10
0	1	0	0	0	1
1	1	0	0	0	16
0	0	1	0	0	1
1	0	1	0	0	3
0	1	1	0	0	0
1	1	1	0	0	11
0	0	0	1	0	2
1	0	0	1	0	14
0	1	0	1	0	0
1	1	0	1	0	21
0	0	1	1	0	3
1	0	1	1	0	15
0	1	1	1	0	2
1	1	1	1	0	28
0	0	0	0	1	6
1	0	0	0	1	29
0	1	0	0	1	8
1	1	0	0	1	56
0	0	1	0	1	1
1	0	1	0	1	28
0	1	1	0	1	3
1	1	1	0	1	61
0	0	0	1	1	11
1	0	0	1	1	81
0	1	0	1	1	16
1	1	0	1	1	173
0	0	1	1	1	4
1	0	1	1	1	80
0	1	1	1	1	15
1	1	1	1	1	298

Table 4.11: Observed frequencies of response patterns, LSAT data

	Response 1	Response 0
Item 1	92.4	7.6
Item 2	70.9	29.1
Item 3	55.3	44.7
Item 4	76.3	23.7
Item 5	87.0	13.0

Table 4.12: Percentages of answering correctly or wrongly to the observed items, LSAT data

4.3.2 One-Factor Unconstrained LTM

In this section we will assume that the associations between the five items of the test can be accounted for by only one common factor. Thus we will assume the following latent trait model:

$$\begin{aligned}
\text{logit}\,(\pi_1) &= \alpha_{10} + \alpha_{11} y_1 \\
\text{logit}\,(\pi_2) &= \alpha_{20} + \alpha_{21} y_1 \\
\text{logit}\,(\pi_3) &= \alpha_{30} + \alpha_{31} y_1 \\
\text{logit}\,(\pi_4) &= \alpha_{40} + \alpha_{41} y_1 \\
\text{logit}\,(\pi_5) &= \alpha_{50} + \alpha_{51} y_1
\end{aligned}$$

or equivalently

$$\pi_i\,(y) = \frac{exp\,(\alpha_{i0} + \alpha_{i1} y)}{1 + exp\,(\alpha_{i0} + \alpha_{i1} y)},\ i = 1, 2, 3, 4, 5$$

where α_{i0} denotes the difficulty parameter, α_{i1} denotes the discrimination parameter and y_1 represents the latent variable that measures the ability of the individuals of the population under study. The assumptions of that model are common to those given in Section 4.2.2.

Estimation of the unconstrained LTM

The model that we have fitted can be estimated by means of the maximum likelihood method. The parameter estimates and their standard errors are given in Table 4.14.

In the first column of the above table, we give the estimated difficulty parameters $\hat{\alpha_{i0}}$ and in the second column their standard errors. In the third column, we give the estimated discrimination parameters $\hat{\alpha_{i1}}$ and in the fourth

Item	Item	p-value
1	2	0.028
1	3	0.003
1	4	0.208
1	5	0.565
2	3	$<$ 0.001
2	4	0.059
2	5	0.009
3	4	0.001
3	5	0.113
4	5	0.002

Table 4.13: Pairwise associations between the observed variables. LSAT data

Item	$\hat{\alpha_{i0}}$	se $(\hat{\alpha_{i0}})$	$\hat{\alpha_{i1}}$	se $(\hat{\alpha_{i1}})$	st $(\hat{\alpha_{i0}})$	$\hat{\pi}_i(0)$
1	2.77	0.2058	0.83	0.2582	0.64	0.94
2	0.99	0.0900	0.72	0.1867	0.58	0.73
3	0.25	0.0763	0.89	0.2327	0.66	0.56
4	1.29	0.0990	0.69	0.1852	0.57	0.78
5	2.05	0.1354	0.66	0.2099	0.55	0.89

Table 4.14: Parameter estimates and standard errors for the unconstrained one-factor LTM, LSAT data

column their standard errors. Then, in the fifth column the standardized values of the factor loadings are given and in the sixth column the probabilities that the median individual will respond positively to items $1-5$ are given.

In Table 4.14, the estimated discrimination parameters, $\hat{\alpha_{i1}}$, that is the estimated factor loadings, are all positive and of similar magnitude with similar standard errors. This implies that all five items have similar discriminating power and so a similar weight is applied to each response. The estimates of the probability of the median individual to respond positively show that the five items are of varying difficulty but relatively easy. In addition, the estimated probabilities of the median individual lead to conclusions similar to those drawn in the descriptive analysis (Table 4.12). The standardized values of the factor loadings can be interpreted as correlation coefficients and thus all the five items have a close link to the common factor.

After having estimated the unknown parameters of the above latent trait model we will proceed to the exploration of the imposition of equality parameter constraints. In fact, the similarity observed in the above estimated parameters may encourage us to impose equality parameter constraints. That is, we may be interested whether the weights of some of the items are exactly the same. Thus in Section 4.3.3 we will fit the one-factor latent trait model after the imposition of equality constraints between the parameters of the model.

Goodness-of-fit

In order to assess the goodness-of-fit of the model that we have fitted we will use the log-likelihood ratio test G^2, the Pearson chi-squared goodness-of-fit test X^2 and the selection criteria AIC and BIC that have been defined in Chapter 2. In Table 4.15, the values of these measures are given.

Test	value	d.f.	p-value
Log-Likelihood Ratio Test G^2	17.55	21	0.5527
Pearson Chi-Squared X^2	18.143	21	0.6399
AIC	4953.333		
BIC	5018.506		

Table 4.15: Goodness-of-fit measures for the unconstrained one-factor model, LSAT data

Both measures do indicate a good fit of the one-factor latent trait model to the data. This result is easily verified by the small discrepancies between the observed and expected frequencies in response patterns.

After having fitted the one-factor unconstrained latent trait model, it is useful to represent graphically the relationship between examinees' item performance and the trait, that we wish to measure, by means of the item characteristic curves (ICC). From Figure 4.3 it is evident that as the factor's values increase the probability of responding correctly increases. In particular, all the items have similar rates of increase.

4.3.3 One-Factor constrained LTM

In this section we will explore the imposition of parameter constraints in the one-factor latent trait model presented in Section 4.3.2. In Table 4.14 we observed that the estimated parameters of the one-factor latent trait model

Item					Observed	Expected
1	2	3	4	5	Frequency	Frequency
0	0	0	0	0	3	2.274
1	0	0	0	0	10	9.476
0	1	0	0	0	1	1.840
1	1	0	0	0	16	11.255
0	0	1	0	0	1	0.696
1	0	1	0	0	3	4.657
0	1	1	0	0	0	0.852
1	1	1	0	0	11	8.446
0	0	0	1	0	2	2.596
1	0	0	1	0	14	15.589
0	1	0	1	0	0	2.891
1	1	0	1	0	21	25.652
0	0	1	1	0	3	1.178
1	0	1	1	0	15	11.462
0	1	1	1	0	2	2.000
1	1	1	1	0	28	29.136
0	0	0	0	1	6	5.863
1	0	0	0	1	29	34.617
0	1	0	0	1	8	6.434
1	1	0	0	1	56	56.111
0	0	1	0	1	1	2.613
1	0	1	0	1	28	24.983
0	1	1	0	1	3	4.369
1	1	1	0	1	61	62.517
0	0	0	1	1	11	8.945
1	0	0	1	1	81	76.566
0	1	0	1	1	16	13.579
1	1	0	1	1	173	173.311
0	0	1	1	1	4	5.953
1	0	1	1	1	80	83.535
0	1	1	1	1	15	13.915
1	1	1	1	1	298	296.678

Table 4.16: Observed and expected frequencies of the response patterns, LSAT data

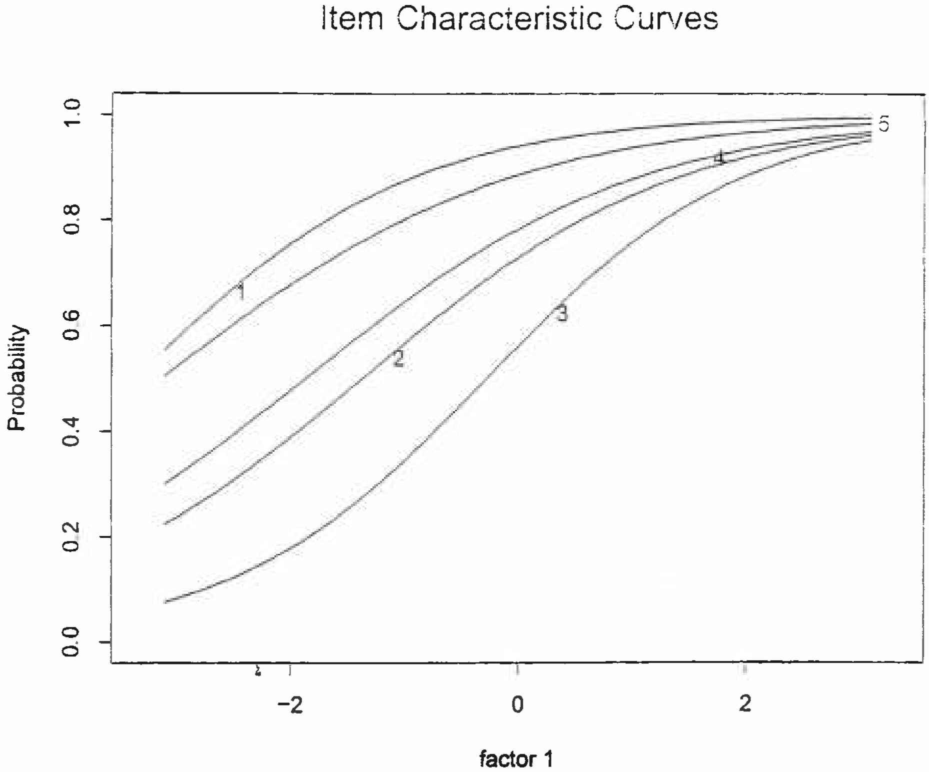

Figure 4.3: Item Characteristic Curves for the five items of the unconstrained latent trait model, LSAT data

of all the items are positive and of the same magnitude. Thus we will assume that the discrimination parameters of all the items are equal and explore the validity of such an assumption. In other words, we will fit the following latent trait model:

$$\begin{aligned}
\text{logit}\,(\pi_1) &= \alpha_{10} + \alpha_{11} y_1 \\
\text{logit}\,(\pi_2) &= \alpha_{20} + \alpha_{21} y_1 \\
\text{logit}\,(\pi_3) &= \alpha_{30} + \alpha_{31} y_1 \\
\text{logit}\,(\pi_4) &= \alpha_{40} + \alpha_{41} y_1 \\
\text{logit}\,(\pi_5) &= \alpha_{50} + \alpha_{51} y_1
\end{aligned}$$

where we will assume that $\alpha_{11} = \alpha_{21} = \alpha_{31} = \alpha_{41} = \alpha_{51}$. Such a model is known as the Rasch model.

The assumptions of the one-factor latent trait model given in Section 4.3.2 need also to be satisfied in the one-factor latent trait model under the imposition of constraints.

Estimation of the constrained LTM

The model that we have fitted can be estimated by means of the maximum likelihood method. The parameter estimates and their standard errors are given in Table 4.17. In Table 4.17, we observe that the standard errors of the estimated difficulty and discrimination parameters of all the items have been reduced and are now in similar magnitude to the standard errors of the other estimates.

Item	$\hat{\alpha}_{i0}$	se $(\hat{\alpha}_{i0})$	$\hat{\alpha}_{i1}$	se $(\hat{\alpha}_{i1})$	st $(\hat{\alpha}_{i0})$	$\hat{\pi}_i(0)$
1	2.73	0.0374	0.76	0.0635	0.605	0.939
2	0.99	0.0086	0.76	0.0388	0.605	0.731
3	0.24	0.0053	0.76	0.0360	0.605	0.560
4	1.31	0.0113	0.76	0.0415	0.605	0.790
5	2.10	0.0222	0.76	0.0502	0.605	0.891

Table 4.17: Parameter estimates and standard errors for the constrained one-factor LTM, LSAT data

Goodness-of-fit

In order to assess the goodness-of-fit of the model that we have fitted we will use the log-likelihood ratio test G^2, the Pearson chi-squared goodness-of-fit test X^2 and the selection criteria AIC and BIC whose values are given in Table 4.18. According to Table 4.18, both measures do indicate a good fit of the one-factor latent trait model to the data and the values of the AIC and BIC of the constrained model are smaller than those of the unconstrained one. This implies that the constrained model is preferable to the unconstrained model. The adequacy of this model is easily verified by the small discrepancies between the observed and expected frequencies in response patterns in Table 4.19. Our purpose after fitting latent variable models is to obtain information about the latent variables through the posterior distribution $h\left(\mathbf{y} \mid \mathbf{x}\right)$ (Posterior Analysis). In Section 2.6 we showed that this distribution depends on $\mathbf{x}$ only through the q-variate sufficient statistic $\mathbf{X} = \mathbf{A}\mathbf{x}$, where $\mathbf{A} = [\alpha_{ij}]$ has as element the factor loading. Thus, for the constrained one-factor latent trait model that we have fitted we can locate the individuals on the latent dimension and thus assign to each individual a score with respect to their trait that we wish to measure. These score are given in the last column of Table

Test	value	d.f.	p-value
Log-Likelihood Ratio Test G^2	21.824	25	0.647
Pearson Chi-Squared X^2	18.345	25	0.827
AIC	4945.902		
BIC	5016.795		

Table 4.18: Goodness-of-fit measures for the constrained one-factor model, LSAT data

Item					Observed	Expected	Total	Component
1	2	3	4	5	Frequency	Frequency	Score	Score
0	0	0	0	0	3	2.364	0	0.000
1	0	0	0	0	10	10.273	1	0.755
0	1	0	0	0	1	1.819	1	0.755
0	0	1	0	0	1	0.852	1	0.755
0	0	0	1	0	2	2.474	1	0.755
0	0	0	0	1	6	5.468	1	0.755
1	1	0	0	0	16	11.391	2	1.510
1	0	1	0	0	3	5.334	2	1.510
0	1	1	0	0	0	0.944	2	1.510
1	0	0	1	0	14	15.498	2	1.510
0	1	0	1	0	0	2.744	2	1.510
0	0	1	1	0	3	1.285	2	1.510
1	0	0	0	1	29	34.249	2	1.510
0	1	0	0	1	8	6.063	2	1.510
0	0	1	0	1	1	2.839	2	1.510
0	0	0	1	1	11	8.249	2	1.510
1	1	1	0	0	11	8.592	3	2.265
1	1	0	1	0	21	24.965	3	2.265
1	0	1	1	0	15	11.690	3	2.265
0	1	1	1	0	2	2.070	3	2.265
1	1	0	0	1	56	55.171	3	2.265
1	0	1	0	1	28	25.834	3	2.265
0	1	1	0	1	3	4.574	3	2.265
1	0	0	1	1	81	75.060	3	2.265
0	1	0	1	1	16	13.288	3	2.265
0	0	1	1	1	4	6.222	3	2.265
1	1	1	1	0	28	27.709	4	3.020
1	1	1	0	1	61	61.235	4	3.020
1	1	0	1	1	173	177.918	4	3.020
1	0	1	1	1	80	83.310	4	3.020
0	1	1	1	1	15	14.749	4	3.020
1	1	1	1	1	298	295.767	5	3.776

Table 4.19: Observed, expected frequencies and scores of the response patterns for the constrained one-factor LTM, LSAT data

4.19 next to the column of "Total Score" that gives the number of positive responses to each response pattern. Thus we observe that the rankings given by the total and the component score are consistent but not very close.

In addition, it is interesting to check whether the latent trait model fitted to the data after the imposition of constraints, is "better" than the one where no constraints were imposed or not. This will be done by means of the likelihood ratio test. The high p-value of the likelihood ratio test in $\alpha = 5\%$ implies that the two models do not differ and that the most parsimonious, that is the constrained model, is preferable. Thus, we conclude that the imposition of constraints in the above model has improved the fit to the data.

From Figure 4.4 it is evident that as the factor values increase the probability of responding correctly increases. In particular, the rate of increase in all the items is similar and seem to have parallel evolutions due to the fact that their discrimination parameters have been set equal.

Model	Log-Likelihood	Likelihood Ratio Test Value	d.f.	p-value
Constraints	-2466.951			
No Constraints	-2466.667	0.542	4	0.969

Table 4.20: Comparison between the constrained and the unconstrained model

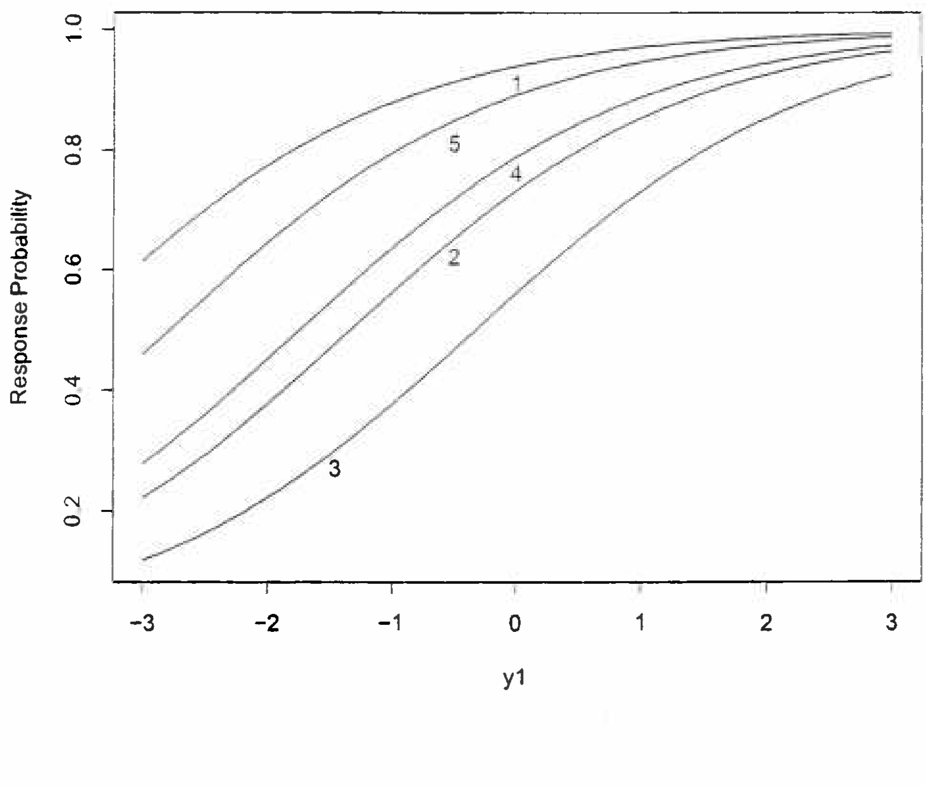

Figure 4.4: Item Characteristic Curves for the five items of the constrained latent trait model, LSAT data

Bibliography

[1] **Bartholomew, D.J., Steele, F., Moustaki, I. and Galbraith, J (2002).** *The Analysis and Interpretation of Multivariate Data for Social Scientists*, Chapman and Hall / CRC

[2] **Bartholomew, D.J. and Knott, M. (1999).** *Latent Variable Models and Factor Analysis*, Second Edition, Kendall's Library of Statistics

[3] Bentler, P. M. and Weeks, D. G. (1980). Linear structural equations with latent variables, *Psychometrika*, 45, 289-308

[4] Bentler, **P. M. and Lee, S. Y. (1983).** Covariance structures under polynomial constraints: Applications to correlation and Alpha-type structural models, *Journal of Educational Statistics*, 8, 207-222

[5] Byrne, Barbara **M. (1998).** *Structural Equation Modeling with LISREL, PRELIS, and SIMPLIS: Basic Concepts, Applications, and Programming*, Mahwah, N.J.: Lawrence Erlbaum Associates

[6] **Bock, R.**D. and **Aitkin, M.** (1981). Marginal maximum likelihood estimation of item parameters: application of an EM algorithm, *Psychometrika*, 46, 443-459

[7] **Bock, R.D. and Lieberman, M. (1970)**. Fitting a response curve model for dichotomously scored items, *Psychometrika*, 35, 179-198

[8] **Christofferson, A.** (1975). Factor analysis of dichotomized variables, *Psychometrika*, 40(1), 5-31

[9] Clogg, C. C. and Goodman, L. A. **(1985)**. Simultaneous latent structure analysis in several groups, In N. B. Tuma (Ed.), *Sociological Methodology*, 81-110, San Francisco: Jossey-Bass

[10] **Dempster, A.P., Laird,** N.M. and Rubin, D.B. **(1977)**. Maximum likelihood from incomplete data via the EM algorithm (with discussion), *Journal of the Royal Statistical Society B*, 39, 1-38

[11] **Everitt, B.S. (1984)**. *An Introduction to Latent Variable Models*, Monographs on Statistics and Applied Probability, Chapman and Hall

[12] **Fisher, H. G.** (1983). Logistic Latent Trait Models with linear constraints, *Psychometrika*, 48, 3-26

[13] Goodman, **L. A. (1974a)**. Explonatory latent structure avalysis using both identifiable and unidentifiable models, *Biometrika*, 61, 215-231

[14] Goodman, **L. A. (1974b)**. The analysis of systems of qualitative variables when some of the variables are unobservable. Part I: A modified latent structure approach, *American Journal of Sociology*, 79, 1179-1259

[15] **Hambleton, R. and Swaminathan, H. (1985).** *Item Response Theory: Principles and Applications*, Kluwer Academic Publishers Group

[16] **Hambleton, R., Swaminathan**, H. and Rogers, **J. H. (1991).** *Foundamentals of Item Response Theory*, Sage Publications

[17] **Hämmerlin G. and Hoffmann K-H.** (1991). *Numerical Mathematics*, Springer-Verlag, New York

[18] **Holland, P.W. (1981).** When are item response models consistent with observed data?. *Psychometrika*. 46. 79-92

[19] Holland, P.W. and Rosenbaum, **P.R. (1985).** *Conditional association and unidimensionality in monotone latent variable models*, Technical Report, Educational Testing Service, Princeton, NJ

[20] **Jamshidian, M.** and Jennrich, RI. (1995). Acceleration of the EM algorithm by using quasi-Newton methods. *Journal of the Royal Statistical Society B*, 59, 569-587

[21] Jöres**kog, K. G. and Sörbom, D. (1989).** *LISREL 7: User' s reference guide*, Chicago: Scientific Software Inc

[22] Jöreskog, K. G. and **Sörbom, D. (1993b).** *LISREL 8: User' s reference guide*, Chicago: Scientific Software International

[23] **Kenny, D. A. and Cohen, S. H. (1980).** A reexamination of selection and growth processes in the nonequivalent control group design,

In K. F. Scheussler (Ed.), *Sociological methodology, 1980*, 290-313, San Francisco, CA: Jossey-Bass

[24] **Körner,** T.W. (1988). *Fourier Analysis*, Cambridge University Press, Cambridge

[25] **Lange, K. (1995)**. A gradient algorithm locally equivalent to the EM algorithm, *Journal of the Royal Statistical Society B*, 57, 425-437

[26] **Lange, K. (1995)**. A quasi-Newton acceleration of the EM algorithm, *Statistica Sinica*, 5, 1-18

[27] **Lee, S. Y. (1980)**. Estimation of covariance structure models with parameters subject to functional restraints, *Psychometrika*, 45, 309-324

[28] **Lee, S. Y. and Tsui, K. L. (1982)**. Covariance structure analysis in several populations, *Psychometrika*, 47, 297-308

[29] **Little R.J.A.** and Rubin D.B. **(1987)**. *Statistical Analysis with missing data*, Wiley, New York

[30] **Lord, F. M. (1952)**. A theory of test scores, *Psychometric Monograph*, No. 7, Psychometric Society

[31] **Lord, F. M. (1953a)**. An application of confidence intervals and of maximum likelihood to the estimation of an examinee's ability, *Psychometrika*, 18, 57-75

[32] Lord, **F. M. (1953b)**. The relation of test score to the trait underlying the test, *Educational and Psycological Measurement*, 13, 517-548

[33] **Lord, F. M. (1968).** An analysis of the Verbal Scholastic Aptitude Test using Birnbaum's three-parameter logistic model, *Educational and Psycological Measurement*, 28, 989-1020

[34] **Lord, F. M. and Novick, M.R. (1968).** *Statistical theories of mental test scores*, Reading Mass: Addison-Wesley

[35] **Lord, F. M. (1980).** *Applications of item response theory to practical testing problems*, Hillsdale, NJ: Erlbaum

[36] **McDonald, R.P. (1980).** A simple comprehensive model for the analysis of covariance structures: Some remarks on applications, *British Journal of Mathematical and Statistical Psychology*, 33, 161-183

[37] **McLachlan, G.J. and Krishnan, T. (1997).** *The EM algorithm and Extensions*, Wiley, New York

[38] **Mooijaart, Ab. and Peter G.M. Van Der Heijden (1992).** The EM Algorithm for Latent Class Analysis with equality constraints, *Psychometrika*, 57, 261-269

[39] Moustaki, I. and Knott, M. **(2000).** Generalized Latent Trait Models, *Psychometrika*, 65, 391-411

[40] **Muthén, B. (1978).** Contributions to factor analysis of dichotomous variables, *Psychometrika*, 43, 551-560

[41] **Powell, M.J.D.** (1981). *Approximation Theory and Methods*, Cambridge University Press, Cambridge

[42] **Press, W.H., Teukolsky, S.A., Vetterling, W.T. and Flannery, B.P. (1992).** *Numerical Recipes in Fortran: The art of Scientific Computing*, 2nd ed., Cambridge University Press, Cambridge

[43] **Rindskopf, D. (1983).** Parameterizing inequality constraints on unique variances in linear structural models, *Psychometrika*, 48, 73-83

[44] **Rindskopf, D. (1984).** Using phantom and imaginary latent variables to parameterise constraints in linear structural models, *Psychometrika*, 49, 37-47

[45] **Rizopoulos, D. (2003).** *Generalized Latent Trait Models: Interaction effects among latent variables*, M.Sc. Thesis, Dept of Statistics, Athens University of Economics and Business

[46] **Rosenbaum, P.R. (1984).** Testing the conditional independence and monotonicity assumptions of item response theory, *Psychometrika*, 49, 425-435

[47] **Stout, W. (1987).** A non parametric approach for assessing latent trait unidimensionality, *Psychometrika*, 52, 589-617

[48] **Stout, W. (1990).** A new item response theory modelling approach with applications to unidimensionality assessment and ability estimation, *Psychometrika*, 55, 293-325

[49] **van de Pol, F. J. R. and Langeheine, R.** (1990). Mixed markov latent class models, In C. C. Clogg (Ed.), *Sociological methodology, 1990, 213-247*, Oxford: American Sociological Association

[50] **Wright, B. D. and Stone, M. H. (1979)**. *Best test design*, Chicago: Mesa Press

Δωρεά